STATISTIKA

STATISTICS AND ECONOMY JOURNAL

VOL. **103** (2) 2023

CZECH STATISTICAL OFFICE

CONTENTS

ANALYSES

CONSULTATION

INFORMATION

About Statistika

The journal of Statistika has been published by the Czech Statistical Office since 1964. Its aim is to create a platform enabling national statistical and research institutions to present the progress and results of complex analyses in the economic, environmental, and social spheres. Its mission is to promote the official statistics as a tool supporting the decision making at the level of international organizations, central and local authorities, as well as businesses. We contribute to the world debate and efforts in strengthening the bridge between theory and practice of the official statistics. Statistika is professional double-blind peer reviewed open access journal included in the citation database of peer-reviewed literature **Scopus** (since 2015), in the **Web of Science** *Emerging Sources Citation Index* (since 2016), and also in other international databases of scientific journals. Since 2011, Statistika has been published quarterly in English only.

Publisher

The Czech Statistical Office is an official national statistical institution of the Czech Republic. The Office's main goal, as the coordinator of the State Statistical Service, consists in the acquisition of data and the subsequent production of statistical information on social, economic, demographic, and environmental development of the state. Based on the data acquired, the Czech Statistical Office produces a reliable and consistent image of the current society and its developments satisfying various needs of potential users.

Contact us

Journal of Statistika | Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic
E-mail: statistika.journal@czso.cz | Web: www.czso.cz/statistika_journal

Economic Performance of Czech Enterprises under the Control of Subjects from War-Affected European Countries and Their Comparison with Western and V4 Controlling Countries

Jan Zeman[1] | *Czech Statistical Office, Prague, Czech Republic*

Received 21.7.2022, **Accepted** (reviewed) 5.1.2023, **Published** 16.6.2023

Abstract

The article deals with the evaluation of the economic performance of domestic enterprises under foreign control, especially by subjects from Eastern European countries currently involved in the war conflict in Ukraine, through several key indicators monitored by Inward Foreign Affiliates Statistics (IFATS) and compiled by the Czech Statistical Office (CZSO). The war conflict in Ukraine raises, among other, questions about the effects of sanctions against aggressor countries. IFATS could serve as one possible source for evaluation of their impact with new data coming in the future. In the article, selected data from the IFATS database will be compared for the economy as a whole and according to the specific prevailing economic activities as well. Although the contribution of enterprises controlled by subjects from both aggressor countries and Ukraine to the Czech economy is quite marginal, the latter ones are among the most represented in terms of their number that is still growing. However, their economic importance grows much slower.

Keywords	***DOI***	***JEL code***
Foreign affiliates, foreign control, business statistics, economic performance indicators, ultimate controlling institutional unit	*https://doi.org/10.54694/stat.2022.32*	*L25, M21*

[1] Czech Statistical Office, Na padesátém 81, 100 82 Prague 10, Czech Republic. E-mail: jan.zeman@czso.cz, phone: (+420)274052039. Also Prague University of Economics and Business, Prague, Czech Republic.

INTRODUCTION

The recent emergence of war conflict in Ukraine, launched by the Russian Federation with the support of Belarus, has provoked discussions, not only in Europe, on the imposition of sanctions and other restrictions against these countries. Among them, to large extent, there are also sanctions having an economic impact on specific persons and enterprises both in the territory of aggressors and in the Czech Republic, where many enterprises owned and controlled by economic entities from the above countries operate.

Economic sanctions can have an impact primarily on employment, but also on the import and export of goods and services, which goes hand in hand with foreign investments that play an important role in the Czech economy. This is also evidenced by data of inward foreign affiliates statistics (hereinafter IFATS), which monitors selected variables (indicators) on domestic enterprises broken down by country of residence of the ultimate owner and by prevailing business activity. Thus, as one of the tools for monitoring and measuring globalization, IFATS makes it possible to compare the behaviour and especially the economic performance and the degree of influence of domestic enterprises under the control of domestic and foreign owners in the economy. Thanks to this fact, in the context of the current situation in Eastern Europe, it will be possible to monitor and evaluate the possible and real effects of the above-mentioned sanctions on enterprises based in the Czech Republic. The main goal of the article is the evaluation of economic performance of domestic enterprises under the control of subjects from the European countries involved in the war conflict. For clearer setting in the framework of the issue, the most important countries for the Czech economy as well as members of the Visegrad group (V4), being the closest partners with common interests within the European integration, are also selected for this analysis. Such evaluation can then serve as a starting point for evaluating the impact of sanctions on aggressor countries, or other phenomena with new data available.

Although the number of domestic (non-financial) enterprises, directly or indirectly controlled by foreign entities, is very low (approximately 2% of all domestic enterprises), their economic importance for the Czech Republic is indisputably great. One of the greatest values that foreign owners bring to the Czech economy can be considered knowledge in the form of new technologies, production and business knowledge and procedures, work organization and management (Šmíd and Lojka, 2015). In this way, they also contribute to increasing of the productivity of the domestic economy. Their primary goal is to appreciate their capital. Therefore, it is not surprising that they sent their foreign ultimate owners dividends of CZK 232 billion in 2020 representing 57% of the total value of dividends paid by Czech (non-financial) enterprises. Foreign-controlled enterprises invested almost CZK 250 billion in the Czech Republic in 2020 (latest available data) representing the share of 35.5% of total gross investment in tangible goods (hereinafter investments), generated value added at factor cost (hereinafter value added) of CZK 1.26 trillion (share of 42.3%) and employed more than 1 million persons (28.5%). The year 2020 was affected by the global coronavirus pandemic. For that reason, between 2019 and 2020 there was a decrease in value of investments from 38.9% and value added from 42.7%. The relative share of number of persons employed did not change significantly. However, each industry was affected by the pandemic with different force. Foreign-controlled enterprises as a whole represent an important part of the Czech economy, but in the case of finer geographical breakdown according to where the actual control of the enterprise is carried out from, their performance does not always necessarily correlate with their number. Foreign investors are important especially in larger enterprises whose economic importance far exceeds their number (Ernest, 2014).

The first chapter introduces FATS as one of the official statistics focused on the issue of globalization and its subset, IFATS, which monitors domestic enterprises under the control by non-resident subjects. Besides the methodology of IFATS, overall findings both at national and international level are presented. They are based on the comparison of values of several key variables which make part of the CZSO database that has been published since 2020. The second chapter presents a closer look at the development of values of key variables regarding the prevailing economic activities according to sections and divisions of CZ-NACE Rev. 2 classification in order to find out where the countries concentrated their interest and assess the economic performance in more detail.

1 FOREIGN AFFILIATES STATISTICS

Foreign affiliates statistics (FATS) as relatively new statistics, compiled in accordance with the Regulation (EC) No 716/2007 of the European Parliament and of the Council on Community Statistics on the structure and activity of foreign affiliates (European Union, 2007), focuses on the monitoring of the impact of the globalization. It measures the degree of influence of foreign-controlled enterprises in a specific economy and quantifies who controls such enterprises. Therefore, it allows to compare the behaviour and performance of enterprises under domestic (national) and foreign control through a set of variables. European Commission (2012) provides the definition of foreign affiliate. It shall mean an enterprise resident in the compiling country over which an institutional unit not resident in the compiling country has control, or an enterprise not resident in the compiling country over which an institutional unit resident in the compiling country has control. Based on such definition, inward and outward FATS are distinguished.

1.1 Inward foreign affiliates statistics

From the Czech Republic's point of view, IFATS, compiled by the CZSO, monitors performance of enterprises domiciled in the Czech Republic that are controlled by any of the subjects domiciled abroad. Outward foreign affiliates statistics (OFATS), compiled by the Czech National Bank, describes the performance of affiliates residing outside the Czech Republic being under the control of subjects residing in the Czech Republic. Apart from many other statistics, IFATS does not relay on its own survey, but it relies on the data on structural business statistics (SBS) and research and development statistics (R&D). It describes the same part of economy as SBS, but it represents its extension in terms of the division of enterprises according to the origin of subject that controls (directly or indirectly) and manages such enterprises. Control is realized by holding more than a half of the voting rights or shares, or the largest share of voting rights and shares in cases of effective minority control, when none of the subjects holds more than a half (European Commission, 2012).

There are 12 variables overarching IFATS as subset of SBS, 10 of which are compiled annually and 2 (R&D indicators) are compiled biennially in accordance with the Regulation (EC) No 250/2009 implementing Regulation (EC) No 295/2008 of the European Parliament and of the Council as regards, inter alia, the definitions of characteristics (European Commission, 2009). This analysis is based on the following ones:

- Number of enterprises;
- Value added at factor cost (hereinafter also referred as value added);
- Gross investment in tangible goods (hereinafter also referred as investments);
- Number of persons employed.

IFATS indicators are compiled from annual business statistics data, therefore they are not in methodological consistency with macroeconomic data, e.g. national accounts (CZSO, 2022). On the other hand, they are fully internationally methodologically comparable. Methods ensuring the protection of confidential data are applied to the published data in accordance with Act No 89/1995 Coll., on the State Statistical Service, as amended, and internal regulations of the CZSO.

1.2 Methodology of IFATS

Ultimate ownership represents the main core of IFATS and is based on so-called UCI concept (Ultimate controlling institutional unit of a foreign affiliate). European Commission (2012) provides the definition of UCI. It shall mean the institutional unit, proceeding up a foreign affiliate's chain of control, which is not controlled by another institutional unit. Identification of UCI (legal entity or natural person) is the key factor to determine the country from which the control of a particular enterprise is carried out. UCI concept relies on the concept of residency, i.e. the place of residency of the UCI should be the country of registration in the case of legal entities and the country of residence in the case of natural persons acting as UCIs (not the country according to the nationality of the natural person, although they are often the same). There are several types of entities that cannot be assigned as UCI. They include countries, regions, cities or municipalities, trust funds or trust funds administrators, foundations and so-called empty shells. Empty shells, most often set up for tax or legal reasons, are represented by heads of groups of enterprises that are purely formal and serve only to optimize a group structure where no real business activity takes place. Therefore, such a head cannot be identified as UCI. Empty shells usually lack substantial physical presence and significant business activities. They are very often located in tax havens and offshore financial centers (e.g. Cyprus, Seychelles, etc.). Determination of UCIs requires step-by-step analysis of control relationships up the ownership chain in the enterprise groups using appropriate data sources such as Business register, EuroGroups register, annual reports or sources available on the Internet. There is no uniform procedure in the UCI identification, so a significant degree of freedom is given to the Member States in this respect.

Population of IFATS consists of subjects belonging to institutional sectors S.11 – Non-financial corporations, S.12 – Financial corporations, and S.14 – Households (S.142 Own-account workers). However, there is a difference between these groups of subjects from the perspective of ultimate ownership as own-account workers are not controlled or owned by anybody in the control chain of enterprises and therefore cannot be considered foreign-controlled. The decisive factor is the location where they have their business registered. Therefore, they are automatically assigned being under domestic control.

IFATS is compiled in accordance with FATS Recommendations Manual (European Commission, 2012) that defines the way in which data is broken down according to the country of residence of ultimate owners (geographical breakdown set by ISO 3166-1 standard) and according to the prevailing economic activity (activity breakdown set by CZ-NACE classification). As well as SBS, IFATS includes mainly market activities classified into sections B to N (except for K) and division S95. The data for section K is also collected by IFATS, but due to the specifics of financial entities (typically different accounting methods), the range of variables compiled is much more limited.

1.3 Overall findings

Foreign control of enterprises in the Czech Republic originates mainly from EU Member States (66% of all foreign-controlled enterprises in 2020). In 2020 such enterprises generated almost 27% of total domestic value added, employed almost a quarter of all persons employed and realized 24% of all investments. The greatest number of enterprises is controlled from Germany, as the most important trading partner of the Czech Republic. Control from the territory of our western neighbour is carried out in approximately 16% of enterprises owned by foreign subjects. These enterprises generated CZK 443 billion of value added (35.1% share in all foreign-controlled enterprises), employed over 346 thousand persons (32.8%) and invested CZK 90.5 billion (36.6%).

Other important foreign investors include the USA (13% share in value added and number of persons employed, in investments almost 9%). It is followed by France, whose control helped domestic enterprises to shares around 6% in terms of the indicators monitored. Regarding the number of foreign-controlled enterprises (almost 20 thousand), Eastern and Southeastern European countries come to the forefront. Specifically, it concerns Ukraine with 7.5% share. When looking at the key indicators, it significantly falls behind (0.21% share in value added, 0.37% in investments and 0.57% in number of persons employed). In case of the Russian Federation, whose subjects controlled 4.6% of Czech enterprises under foreign control, the values of other indicators are only slightly higher, as indicated in Table 1. It shows that enterprises controlled from Belarus make up insignificant part of IFATS with shares in the order of 1 to 2 hundredths of a percentage point. If we added up the shares of the Russian Federation and Belarus, it would not even reach 1% of the total in the case of investments and number of persons employed. In the case of value added, the common share slightly exceeds 1%.

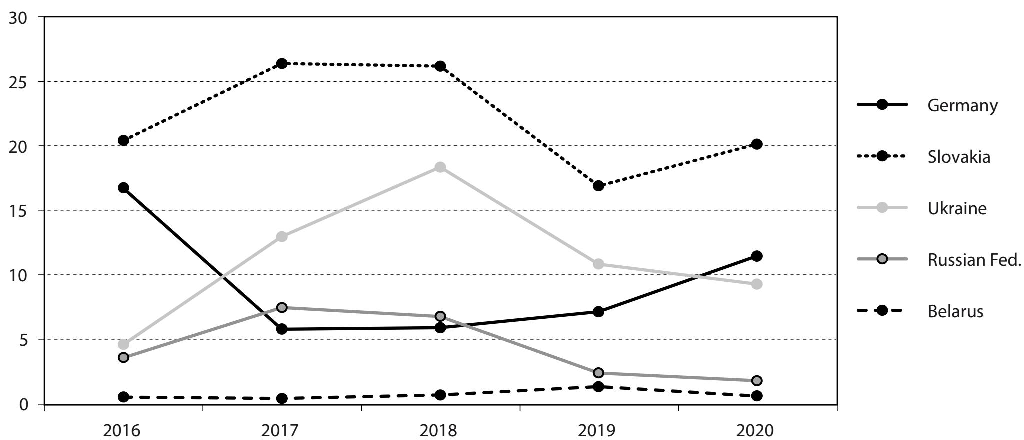
Table 1 Share of foreign-controlled enterprises by country of UCI and key IFATS variables in all foreign-controlled enterprises (in %)

Reference year	Variable	Country of UCI								
		Germany	USA	France	Slovakia	Poland	Hungary	Ukraine	Russian Fed.	Belarus
2018	VAFC	36.05	13.15	5.99	2.23	2.24	0.20	0.13	0.87	0.01
	GITG	36.50	9.09	5.99	1.84	3.20	0.44	0.11	0.65	0.00
	NPE	32.86	13.01	5.90	3.91	2.00	0.13	0.34	0.88	0.01
2019	VAFC	35.24	13.94	5.95	2.38	1.76	0.27	0.24	0.96	0.01
	GITG	39.91	8.46	6.34	2.20	2.59	0.44	0.20	0.87	0.01
	NPE	32.78	12.88	5.94	3.84	2.03	0.18	0.60	0.89	0.01
2020	VAFC	35.13	13.67	6.04	2.40	1.25	0.79	0.21	1.11	0.02
	GITG	36.58	8.71	6.72	1.92	3.02	0.43	0.37	0.93	0.01
	NPE	32.80	12.54	6.02	3.76	2.08	0.50	0.57	0.97	0.01

Note: VAFC – Value Added at Factor Cost, GITG – Gross Investment in Tangible Goods, NPE – Number of Persons Employed.
Source: Own construction based on CZSO's data

As already mentioned above, it is not appropriate to evaluate the importance of foreign owners according to the number of enterprises they control, since many enterprises with very little economic importance operate in the Czech Republic. An alternative option may be rather monitoring the development trend of the number of newly established and defunct enterprises. It shows how the level of scope of individual countries develops in our economy. Based on such characteristic, we find the share of number of enterprises controlled by Ukrainian and Slovak subjects slowly but surely increasing at the expense of German subjects, whose representation in our country is slowly decreasing over time (Zeman, 2022). On the contrary, in the case of the Russian Federation and Belarus the shares do not increase significantly and maintain rather stable trend. Figure 1 depicts the number of newly established enterprises under

Figure 1 Share of newly established enterprises under foreign control of selected countries in all newly established foreign-controlled enterprises (in %)



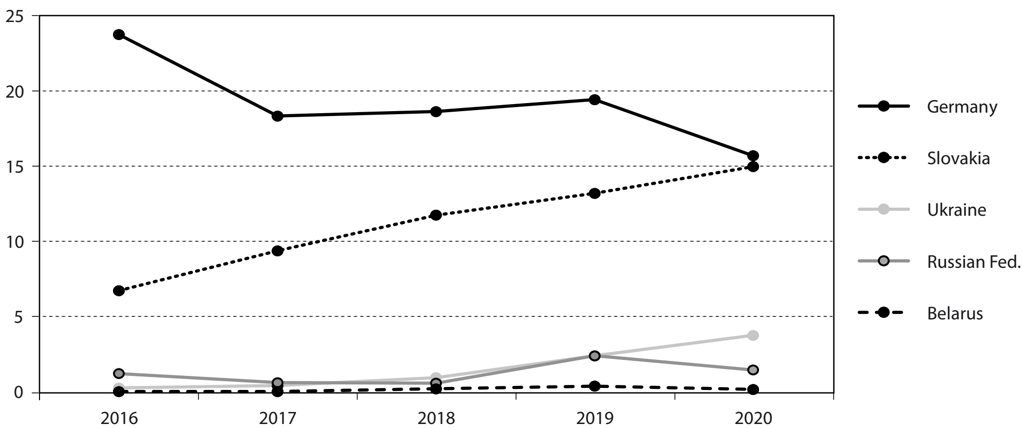
Source: Own construction

the control of selected countries, i.e. those that did not yet exist in the previous reference year. It is evident that enterprises controlled from Slovakia dominate over enterprises controlled from Germany and Ukraine.

Figure 2 shows the number of enterprises that have ceased to exist in the given period (e.g. value in 2020 represents enterprises defunct between 2019 and 2020 and thus did not enter the IFATS population in 2020). We can see the most of defunct enterprises were controlled from Germany and further from Slovakia with constantly increasing trend. Ukraine and the Russian Federation follow with greater distance. The net change in the number of newly established enterprises, i.e. after adjustment by defunct enterprises, we find there has been a decrease in the number of enterprises controlled by German subjects, while the ones controlled by Slovak and Ukrainian subjects experienced an increase.

However, the resulting number of foreign-controlled enterprises is determined not only by the number of newly established and the defunct ones. Another factor affecting this indicator is year-on-year change of UCI, either between the Czech and foreign one, or between two foreign ones. In this respect, the data speaks in favour of Ukraine, as between 2019 and 2020 there was a change of UCI towards Ukraine in 147 cases and in the opposite direction in 78 cases. This results in 69 more enterprises controlled from Ukraine being the most of all countries. For comparison, in the case of Germany there was a net increase of –31 enterprises, in the case of Slovakia 3 enterprises, in the case of Belarus 2 enterprises and in the case of the Russian Federation 18 enterprises.

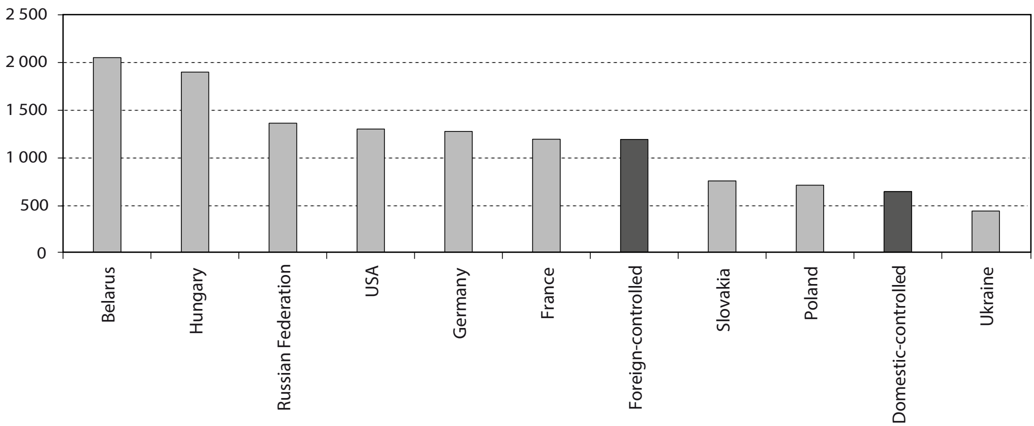
Figure 2 Share of defunct enterprises under foreign control of selected countries in all defunct foreign-controlled enterprises (in %)



Source: Own construction

Labour productivity is also related to the economic performance of enterprises. It is not a part of the official set of IFATS indicators, but can be calculated as a share of value added in the number of persons employed. Figure 3 provides labour productivity by UCI countries in 2020. It also includes values calculated for the entire population of foreign-controlled and domestic-controlled enterprises. Quite interesting finding is that Belarus ranked highest with its productivity of more than CZK 2 million. Its high labour productivity was driven mainly by manufacture of wearing apparel (division C14 of CZ-NACE) and wholesale trade, except of motor vehicles and motorcycles (G46). Chapter 2 provides more detailed information on performance in CZ-NACE sections. Belarus is followed by Hungary as the highest ranked country from the V4 group, where 1 person employed generated value added of CZK 1.9 million. This ranking was the same as in 2018. In 2019, Hungary ranked first ahead of Belarus. Such results suggest a long-term trend with minimal productivity differences that fluctuate only slightly. As in 2019, third place from the group of the aforementioned countries belongs to the Russian Federation with CZK 1.37 million and so it surpasses even the USA, Germany and France. Direct comparison of labour productivity between domestic- and foreign-controlled enterprises sounds more favourable for the latter one resulting in the difference of CZK 546 thousand. Slovakia and Poland ranked behind the foreign-controlled enterprises as a whole with productivity of CZK 761 and 717 thousand, respectively. One of the lowest labour productivity among foreign-controlled enterprises falls to Ukraine (only CZK 445 thousand). Such finding confirms that although there are many enterprises controlled by Ukrainian subjects in the Czech Republic (approximately 1.5 thousand), these are mostly small and insignificant for Czech economy, which corresponds to the generated value added.

Figure 3 Labour productivity by country of UCI (2020, thousands CZK)

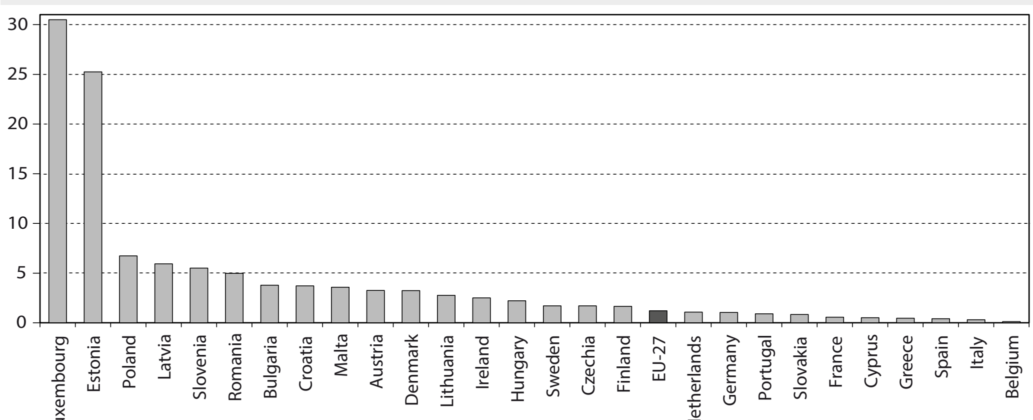


Source: Own construction based on CZSO's data

1.4 International comparison within European Union

At national level, we can observe very similar trend in the development of key IFATS indicators over a longer period. Eurostat publishes data in its database at relatively long interval of approximately 30 months after the end of reference year (Eurostat, 2022). The share of the number of foreign-controlled enterprises is provided in Figure 4. It is comparable across most of the EU Member States and in 2019

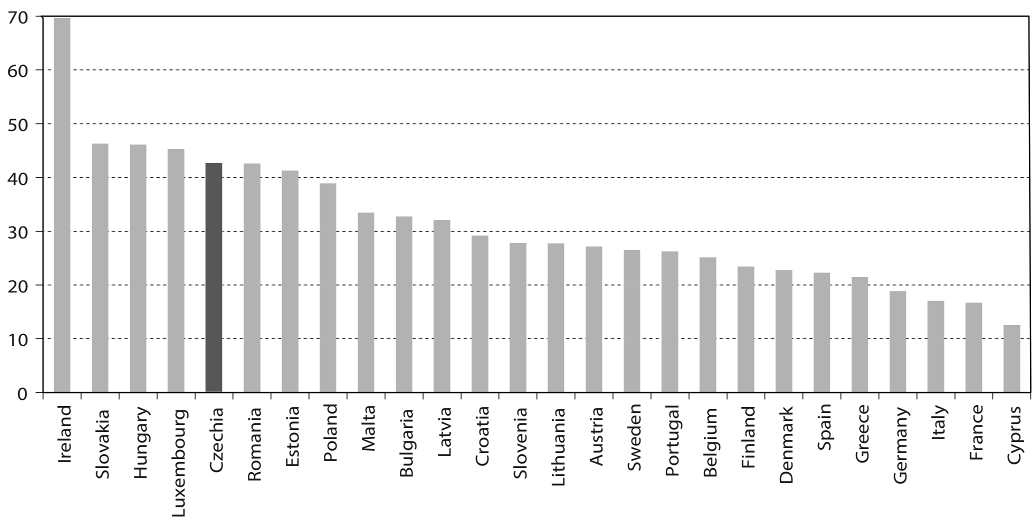
Figure 4 Share of number of foreign-controlled enterprises in total number of enterprises in EU countries (2019, in %)



Source: Own construction based on Eurostat's data (latest data available) ranged from 1 to 4%. The Czech Republic reached the share of 1.7%. The average for EU-27 Member States was 1.21%. Data for the United Kingdom has no longer been part of the publication outputs since the reference year 2018. Luxembourg with its share of more than 30% and Estonia with its share over 25% achieved exceptionally high values compared to the other countries. Poland with its share of 6.7% closes the top three with a greater distance. Countries mostly of Eastern and South-Eastern Europe follow with similar results. At the opposite end there are Southern European countries (Portugal, Spain, Italy and Cyprus), but our eastern neighbour Slovakia (0.83%) and our most important trade partners Germany and France (around 1%) as well.

The real importance of foreign-controlled enterprises should be measured mainly by value added generated by such enterprises in the economy. In this respect, the Czech Republic belongs into the group of Member States with unquestionably important role of foreign-controlled enterprises. This is evidenced by the fact that they generated almost 43% of value added, which ranks the Czech Republic in fifth place among EU Member States. Even greater value added was generated by foreign-controlled enterprises in Slovakia (46.3%) and Hungary (43.1%), which ranks them the second and third place, respectively. However, foreign-controlled enterprises play the most important role in Ireland (almost 70%). Figure 5 summarizes the data. The Netherlands is not part of the figure, as data on value added has been subjected to statistical confidentiality and could not be published. This fact also results in the absence of total EU-27 aggregate. In previous years the total share for EU-28 reached around 25% (Zeman, 2020a), so with the exclusion of the United Kingdom (27.5% in 2017) we can expect the overall share slightly lower.

Figure 5 Share of value added at factor cost generated by foreign-controlled enterprises in total value added at factor cost generated by enterprises in EU countries (2019, in %)



Source: Own construction based on Eurostat's data

The amount of investments and number of persons employed by foreign-controlled enterprises in individual Member States quite strongly correlate with the value added generated as IFATS data suggests. Therefore, it is not surprising that the highest share of investments was achieved by Ireland again (more than 80%). Almost 39% share of investments and 28.4% share of number of persons employed was recorded in the Czech enterprises under foreign control. Very similar shares were achieved by other V4 countries. Slovakia: 42.6% in investments and 28% in number of persons employed; Poland: 40.2%

in investments and 31.8% in number of persons employed; Hungary: 40.2% in investments and 25.1% in number of persons employed. Regarding Germany and France, the shares reached 14–15% in case of investments and 12–13% in case of number of persons employed.

2 CHARACTERISTICS OF ECONOMIC PERFORMANCE BY PREVAILING ECONOMIC ACTIVITY

The importance of foreign-controlled enterprises depends on the industry according to their prevailing economic activity (Zeman, 2020b). Manufacturing industry (section C of CZ-NACE classification) is the one, where foreign-controlled enterprises dominate for a long time. It received the most investments from abroad in 2020 as well (61.6% share of all investments by foreign-controlled enterprises in this section). The greatest share of value added was generated (59.4%) and the most of persons were employed (44.9%) here. The manufacture of motor vehicles, trailers and semi-trailers (division C29 of CZ-NACE) clearly dominates within the manufacturing industry with 27.7% of foreign-controlled enterprises' value added. Manufacturing industry is followed by information and communication (section J) with 44.7% of value added, 32.7% of investments and 35.3% of persons employed and then by wholesale and retail trade; repair of motor vehicles and motorcycles (section G) with 41.2% of value added, 32.6% of investments and 26.6% of persons employed.

If we look at foreign control in individual CZ-NACE sections by countries of UCI, we find it concentrated in section C especially. For example, German, French and American owners centralized here the highest share of investments, employed the most of persons and their enterprises generated the most of value added. With the exception of value added, the same applies to enterprises under the control of Polish subjects (they generated the biggest value added in section G). However, if we were to express the data in the form of shares in all foreign-controlled enterprises, we would find out that enterprises of German subjects generated over half a share of value added (54.8%) and invested almost 77% in transportation and storage (section H). Enterprises of French UCIs generated the biggest share of value added (29.3%) in water supply; sewerage, waste management and remediation activities (section E) and employed 31% of persons. However, they invested the most (24.8%) in administrative and support service activities (section N). American subjects invested primarily in repair of computers and personal and household goods (division S95). Regarding the rest of key IFATS indicators, they achieved the highest shares in section J (34.7% of value added, 37.3% of persons employed).

V4 countries spread their business activities in the Czech Republic rather unevenly. Enterprises under the control of Slovak entities had the highest rate of representation, more than a quarter, in construction (section F). The same can be said about Hungarian subjects, but their share reached only slightly over 4%. Polish subjects controlled a fifth of foreign-controlled enterprises operating in mining and quarrying (section B). In terms of value added and number of persons employed, enterprises under the control from Slovakia reached the highest representation in section N (10.4% and 17.7%). A significant part of investments was brought in division S95 (27.4%). It is also worth mentioning that the significant share of investments in the construction sector (18.2%) was recorded. Approximately one-tenth of value added generated and 4.3% of persons employed in electricity, gas, steam and air conditioning supply (section D) belonged to Hungary-controlled enterprises. They achieved around 5% share of investments in professional, scientific and technical activities (section M). Enterprises controlled by Polish entities no longer have any significant representation among foreign-controlled enterprises at the level of CZ-NACE sections, but at the level of divisions, they dominate in manufacture of coke and refined petroleum products (division C19). Their importance is also reflected in manufacture of chemicals and chemical products (division C20), where they reached almost 52% share in investments and almost 30% in number of persons employed.

Although geographically Ukraine, the Russian Federation and Belarus represent countries close to each other, we cannot talk about closeness from the perspective of prevailing economic activity of the enterprises controlled by them in the territory of the Czech Republic (Zeman, 2022). There are few dozens of Belarus-controlled enterprises in the Czech Republic. They participated mostly in section H of CZ-NACE in terms of their number (0.7%). In case of the other key indicators, the shares are negligible as well (around 0.1%). As shown in Table 2, regarding the structure, they spread their activities mainly in section G, where they generated the most of value added (85.7%), investments (78.4%) and employed the most of people (42.9%). However, the values themselves are very low, compared to other UCI countries. Enterprises under the control of subjects from the Russian Federation accounted for 11% of their number in accommodation and food service activities (section I). In the same section, they contributed almost 5% to investments and a little over 4% to the number of persons employed. 14.1% of all foreign-controlled enterprises participated directly in the accommodation division (I55). More than 5% of value added of all foreign-controlled enterprises was generated in section J. It corresponds in particular to the division of publishing activities (J58), where such enterprises generated almost 57% of section J's total value added and achieved the highest labour productivity. Although the highest share of investments belongs to section I (almost 5%), in terms of more detailed breakdown they belong to activities of head offices; management consultancy activities (division M70; 10.6%) and manufacture of computer, electronic and optical products (division C26; 9.2%).

Table 2 Structure of key IFATS variables by selected CZ-NACE sections (2020, in %)

UCI country	CZ-NACE	ENT	VAFC	GITG	NPE
Ukraine	C	10.0	40.1	46.5	66.9
	G	27.5	6.5	8.5	10.9
	J	6.8	41.9	2.9	2.3
	L	23.0	2.2	30.5	2.2
	M	13.6	2.7	5.3	3.4
	N	4.1	0.2	0.1	2.1
Russian Federation	C	9.4	5.6	3.3	6.0
	G	19.9	14.8	13.6	8.8
	J	3.2	1.7	0.5	0.8
	L	10.8	7.5	49.7	3.6
	M	20.1	14.7	10.8	10.2
	N	10.5	34.8	6.4	44.9
Belarus	C	3.0	1.1	0.0	0.9
	G	47.8	85.7	78.4	42.9
	J	7.5	0.8	0.0	4.9
	L	11.9	1.5	9.8	5.6
	M	9.0	1.5	11.1	2.0
	N	6.0	0.3	0.4	17.1

Note: ENT – Number of enterprises, VAFC – Value Added at Factor Cost, GITG – Gross Investment in Tangible Goods, NPE – Number of Persons Employed.
Source: Own construction based on CZSO's data

Enterprises controlled by Ukrainian entities, whose number is one of the highest in the Czech Republic, were concentrated quite evenly in several CZ-NACE sections with more than a fifth of them having construction as their prevailing business activity. Taking a closer look at the individual sections, we can see their representation was highest especially in services to buildings and landscape activities (division N81; 41.3%). Regarding investments, the highest share belongs to section N as well, but in employment activities (N78; 14%). In terms of value added and number of persons employed, Ukrainian enterprises bring their importance in other professional, scientific and technical activities (M74) with almost a fifth of persons employed (17.4%) and 22.2% of value added generated. However, in the other divisions and entire sections, with the exception of construction, Ukraine-controlled enterprises play rather marginal role. Nor can we speak of significance in terms of labour productivity, although in a few divisions one can identify productivity an order of magnitude higher than in others. Such divisions include in particular information service activities (J63) and scientific research and development (M72). A comprehensive overview of individual representations of enterprises under control of Ukraine, the Russian Federation and Belarus is presented in Table 3. Some values could not be published, as they are result of calculation operations with values subject to statistical confidentiality.

Table 3 Share of foreign-controlled enterprises by country of UCI and CZ-NACE sections in all foreign-controlled enterprises (2020, in %)

UCI country	CZ-NACE	ENT	VAFC	GITG	NPE
Ukraine	Total	7.48	0.21	0.37	0.56
	B	2.38	i.d.	i.d.	i.d.
	C	3.93	0.02	0.02	0.06
	E	6.13	0.30	0.10	0.88
	F	20.27	1.09	1.49	3.79
	G	5.39	0.18	0.44	0.28
	H	7.81	0.29	0.13	0.54
	I	7.93	0.96	1.92	0.91
	J	4.08	0.04	0.06	0.09
	L	5.68	0.53	2.04	3.44
	M	11.36	0.69	3.19	1.33
	N	14.99	1.93	0.42	4.05
Russian Federation	Total	4.58	1.11	0.93	0.97
	C	2.55	0.86	0.82	1.19
	D	0.50	i.d.	i.d.	i.d.
	E	1.89	i.d.	i.d.	i.d.
	F	3.74	0.27	1.03	0.37
	G	4.55	0.42	0.69	0.58
	H	2.69	1.24	0.13	0.91
	I	10.99	1.37	4.79	4.31
	J	5.27	5.33	0.89	0.42
	L	7.42	0.82	3.14	3.68
	M	4.73	0.66	3.93	0.76
	N	3.55	0.06	0.02	0.32
	S95	2.22	i.d.	i.d.	i.d.
Belarus	Total	0.34	0.02	0.01	0.01
	C	0.06	i.d.	i.d.	0.00
	F	0.30	0.00	0.00	0.01
	G	0.58	0.12	0.04	0.03
	H	0.67	i.d.	i.d.	0.07
	I	0.18	i.d.	i.d.	i.d.
	J	0.43	0.00	0.00	0.01
	L	0.28	0.01	0.01	0.13
	M	0.23	0.01	0.05	0.01
	N	0.38	0.00	0.00	0.04

Note: ENT – Number of enterprises, VAFC – Value Added at Factor Cost, GITG – Gross Investment in Tangible Goods, NPE – Number of Persons Employed, i.d. – individual data (confidential).
Source: Own construction based on CZSO's data

CONCLUSION

Foreign-controlled enterprises represent a group of enterprises that play very important role in the Czech economy in terms of maintaining and developing economic growth. Inward foreign affiliates statistics allows to monitor several key indicators and evaluate behaviour and performance of enterprises according to the country from which the control is carried out. Currently, the focus is mainly on the countries involved in the war conflict in Ukraine and the related possible impact of sanctions on the Czech economy. Based on IFATS results, enterprises controlled by the Russian Federation and Belarus represent rather marginal part of economic performance as the value added of these two countries accounts for approximately 1.13% generated by foreign-controlled enterprises and about 1% of investments and number of persons employed as well. Above-average labour productivity was identified for both countries, but it is driven by only a few enterprises. Ukraine-controlled enterprises are abundantly represented in the Czech Republic, but their economic importance is not very significant, either. However, the number of these enterprises gradually increases every year (along with Slovak ones), especially at the expense of our most important trading partner, Germany. It will be appropriate to monitor whether such a trend will continue in the following years (with regard to the development of the war conflict) and how the development of value added and investments will look like. In the same way, possible further restrictions can be expected due to the predicted growing wave of coronavirus pandemic that can affect the behaviour of not only enterprises, but also consumers. Any restrictions imposed on the subjects from the Russian Federation could affect, although not as strongly, industry of accommodation and food service activities, where they invested the most and employed most of persons (in relation to all foreign-controlled enterprises). Information and communication activities could be another industry prone to sanctions' impacts as these enterprises generated the highest share of value added.

References

CZSO. (2022). *Cross-sectional business statistics: Statistics on foreign affiliates 2022* [online]. Prague: Czech Statistical Office. [cit. 27.6.2022]. <https://www.czso.cz/csu/czso/statistics-on-foreign-affiliates>.

ERNEST, J. (2014). Český průmysl je „náš" jen z poloviny. *Statistika&My*, 10 (October): 18–19.

EUROPEAN COMMISSION. (2009). *Commission Regulation (EC) No 250/2009 of 11 March 2009* [online]. Brussels: European Commission. [cit. 8.7.2022]. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02009R0250-20140520&from=CS>.

EUROPEAN COMMISSION. (2012). *Foreign AffiliaTes Statistics (FATS) Recommendations Manual* [online]. Luxembourg: Publications Office of the European Union. <http://doi.org/10.2785/32696>.

EUROPEAN UNION. (2007). Regulation (EC) No 716/2007 of the European Parliament and of the Council of 20 June 2007 on Community statistics on the structure and activity of foreign affiliates. *Official Journal of the European Union*, 50(L 171): 17–31.

EUROSTAT. (2022). *Globalisation in business statistics: Foreign affiliates 2022* [online]. Luxembourg: Eurostat. [cit. 28.6.2022]. <https://ec.europa.eu/eurostat/web/economic-globalisation/globalisation-in-business-statistics/foreign-affiliates>.

ŠMÍD, R., LOJKA, J. (2015). Kdo tahá za nitky české ekonomiky? *Statistika&My*, 3 (March): 21–24.

ZEMAN, J. (2020a). Nadnárodní podniky jsou v české ekonomice nepostradatelnými hráči. *Statistika&My*, 1 (January): 12–13.

ZEMAN, J. (2020b). Jak podniky pod zahraniční kontrolou ovlivňují různá odvětví ekonomiky. *Statistika&My*, 11–12 (November–December): 44–46.

ZEMAN, J. (2022). Vliv zemí východní Evropy na českou ekonomiku není velký. *Statistika&My*, 4 (April): 16–17.

Inequality of Opportunity in Education: the Visegrad Countries Case

Iveta Pauhofová[1] | *Slovak Academy of Sciences, Bratislava, Slovak Republic*
Zulfiya Ibragimova[2] | *Ufa University of Science and Technology, Ufa, Russian Federation*
Marina Frants[3] | *Ufa University of Science and Technology, Ufa, Russian Federation*

Received 19.12.2022, **Accepted** (reviewed) 20.1.2023, **Published** 16.6.2023

Abstract

In the run-up to launching the «Next Generation EU» large-scale recovery plan aimed at overcoming the Covid-19 pandemic consequences it becomes important to analyse the pre-Covid development problems, including those in the field of education. The purpose of this work is to study the dynamics of inequality in educational achievements and opportunities in the Visegrad Group countries based on the PISA data from 2003–2018 period. The results obtained by our research team suggest that Hungary has the highest level of inequality of opportunity among the Visegrad Group countries, followed by Slovakia. Meanwhile, individual factors contributing to the overall level of inequality of opportunity have both features common to all countries and unique features.

Keywords	***DOI***	***JEL code***
Inequality of opportunity in education, individual achievements, circumstance factor, effort factor, PISA, the Visegrad countries	*https://doi.org/10.54694/stat.2022.57*	*I24, D63*

INTRODUCTION

The Covid-19 pandemic has turned out to be a serious challenge to social and economic well-being of people globally, including the EU countries. In response to the challenge, a «reset» plan was adopted, dubbed «Next Generation EU» (*Europe's moment,* EC, 2020). Its main goal is to help the EU Member States make up for the economic and social damage caused by the pandemic and ensure a better future for the oncoming generation of Europeans. The «reset» plan implies subsidies to be allocated to specific

[1] Institute of Economic Research, Slovak Academy of Sciences, Sq. Šancová 56, 811 05 Bratislava, Slovak Republic. E-mail: ipauhofova@yahoo.com.
[2] Institute of Economics, Finance and Business, Ufa University of Science and Technology, 32 Zaki Validi, 450076 Ufa, Russian Federation. Corresponding author: e-mail: badertdinova@mail.ru, phone: (+7)3472299608.
[3] Institute of Economics and Management, Ufa University of Science and Technology, 32 Zaki Validi, 450076 Ufa, Russian Federation. E-mail: tan-marina@mail.ru.

reform and investment programs developed by recipient countries. This refers to complex social, economic, and environmental transformations, including the climate warming control, health care development, support for backward regions, and so on.

Prior to making out national plans of reforms it is necessary to thoroughly analyse the dynamics and current state of things in all public spheres. National education system is a key element that determines development prospects for the entire set of social and economic relations, therefore, the study of education trends and underlying factors is of great importance for understanding and long-term forecasting of social development.

Our study is focused on education development trends in the Visegrad Group countries, applying the theory of equal opportunities. The paper is based on data by the Program for International Student Assessment of educational achievements (PISA).

The paper briefly discusses main ideas of the theory and the impact of educational opportunities on setting further economic and social inequality at later stages of life. A review of works on inequality of opportunity in the Visegrad countries is given as well as description of the study's calculation method and information base. The last section describes the results of assessing inequality of educational achievements and opportunities in the countries under review and the contribution of certain groups of circumstance factors thereto. This study continues the series of articles written by the authors and covering the theory and practice of measuring inequality of opportunity (Ibragimova and Frants, 2020; Pauhofova et al., 2020; Ibragimova and Frants, 2021a; Ibragimova and Frants, 2021b).

1 THEORETICAL OVERVIEW

1.1 Theoretical background of the inequality of opportunity problem studies

The work by Rawls (1971) on distribution of rights and resources between society members is the seminal research on inequality of opportunity. Based on the idea that individuals should be responsible for their life views and respective personal objectives, aspirations and behavior, the following concept of social justice was substantiated: if rights and resources are equally distributed among members of society, then differences in their well-being, resulting from different views on life and, hence, different behavior, refer to personal responsibility and cannot be deemed unfair. Further on, works by Dworkin (1981a, 1981b), Arneson (1989), Cohen (1989), Roemer (1993) proposed to replace the inequality of «resources» with the inequality of «opportunity». To the forefront came the idea that any individual should be responsible for what he/she controls. Therefore, the differences in wealth inequality caused by factors entirely dependent on an individual (effort factors) are just and should not be compensated (the natural reward principle). On the contrary, wealth differences caused by factors outside an individual's control (circumstance factors) are unfair and subject to compensation in a just society (the compensation principle).

The circumstance factor classification proposed by Barros (Barros et al., 2009) singles out the category of basic opportunities or conditions, which include circumstances present at the early stages of an individual's formation and having a significant impact on his/her chances to succeed in life. Providing all children with the equal opportunity to receive quality school education falls under the category of basic opportunities – since schooling takes place earlier in life and has a significant impact on the later life path: inequality in getting a school education produces inequality in access to higher levels of education. Further on, the highest level of education attained makes a major determinant of other individual achievements later in life, including the economic (e.g., earnings, access to quality jobs), political and social ones.

Thus, the inequality in school education opportunities is translated further into the inequality of economic, political, social opportunities and achievements. Therefore, the inequality of education opportunities in schooling sets barriers to economic and social mobility in the perspective and preserves socio-economic stratification. Besides, it hinders individuals from fully realizing their potential and thus slows down social development. For these reasons, studying the inequality of education opportunities, including those at the school level, is of paramount importance for the understanding and long-term forecasting of social processes.

1.2 Review of works on inequality of opportunity in the Visegrad Group countries

Practical application of the equal opportunity theory to the study of inequality required some clarification and mathematical formalization of the ideas described. Roemer (1993, 1998) and Van de Gaer (1993) made a huge contribution thereto, having laid basis for measuring inequality of opportunity. A vast amount of research has been published by now that proposes different estimation methods and measures inequality of opportunity across different countries and regions in respect of various benefits that are meaningful to all or the majority. Of course, most of the works are devoted to inequality of opportunity referring to income. Research on inequality of opportunity in education is less numerous, but it is also being actively developed. In this review, we will limit ourselves to works that provide estimates of inequality of opportunity in all four countries of the Visegrad Group, first in terms of income, then in terms of education.

Table 1 Inequality of opportunity input to inequality of income across the Visegrad Group countries (in %)

Source	Assessment technique	Inequality index	Information basis	Circumstance factors	Income measure	Outcome, %
Transition report (2016–17)	Parametric, ex-ante	Gini	LITS III, 2015–2016	Parent's education and membership in the Communist Party, birthplace, gender, ethnicity	Average monthly personal earned income	HUN: 40.3 SVK: 46.4 POL: 36.8 CZE: 41.9
Ibragimova and Frants (2021)	Parametric, ex-ante ex-post	Gini, GE(0)	LITS III, 2015–2016	Parent's education, birthplace, gender, ethnicity	Average monthly personal earned income	Gini HUN: 37.7 SVK: 36.2 POL: - CZE: 38.75 GE(0) HUN: 13.9 SVK: 12.79 POL: - CZE: 14.45
Marrero and Rodriguez (2012)	Parametric, ex-ante	GE(0)	EU-SILC, 2005	Parent's education and family wealth, father's professional status, birth country	Annual equivalent disposable income	HUN: 11.6 SVK: 3.6 POL: 10.27 CZE: 5.85
Brzeziński (2015)	Parametric, ex-ante	GE(0)	EU-SILC, 2004, 2010	Parent's education, father's professional status, birth country	Annual equivalent disposable income	2004: 00.00 HUN: 11 SVK: 2.5 POL: 11 CZE: 6 2010: 00.00 HUN: 15 SVK: 7.5 POL: 10 CZE: 9
Checchi, Peragine, Serlenga (2010)	Parametric, ex-ante	GE(0)	EU-SILC, 2005	Parent's education, father's professional status, birth country, birthplace, gender	Personal earned income after tax	HUN: 10.1 SVK: 13.3 POL: 9.1 CZE: 10.7

Source: Author's computations

As seen from Table 1, estimates of the inequality of income opportunity in the Visegrad countries often turn out to be close when the circumstances include the location factor (type of community – birth place of an individual). Unless the location factor is taken into account, the inequality of opportunity in terms of income in Slovakia and the Czech Republic turns out to be noticeably lower than in Poland. Table 1 also shows that the measure of inequality used has a significant impact on the assessment outcome – the contribution of inequality of opportunity to income inequal ity is much higher when using the Gini index as compared to the Theil L-index (GE0). This point is reviewed in more detail in our paper (Ibragimova and Frants, 2021b).

Educational achievements and opportunities are given a significant share of studies in official publications under the PISA project (OECD, UNESCO, 2003; OECD, 2004; OECD, 2007; OECD, 2010; OECD, 2013b; OECD, 2016; OECD, 2020) for all years, and the focus is made on cross-country differences, but not on change dynamics over time in each individual country. Results of assessing the impact of economic, social and cultural status of students' families on educational achievements in the Visegrad countries are shown in Table 2.

Table 2 Inequality of opportunity input to inequality of educational achievements across the Visegrad Group countries, based on the PISA data

Publication	Assessment technique	Circumstance factors	Outcome*		
Official publications by PISA	Percentage of explained variance	ESCS index (index of economic, social and cultural status	**2003** HUN: 27, –, – SVK: 22, –, – POL: 17, –, – CZE: 19, –, –	**2006** HUN: –, –, 21 SVK: –, –, 19 POL: –, –, 14 CZE: –, –, 16	**2009** HUN: –, 26, – SVK: –, 15, – POL: –, 15, – CZE: –, 12, –
			2012 HUN: 23, 20, 22 SVK: 25, 24, 26 POL: 17, 13, 14 CZE: 16, 15, 14	**2015** HUN: –, –, 21 SVK: –, –, 16 POL: –, –, 13 CZE: –, –, 19	**2018** HUN: –, 19, – SVK: –, 17, – POL: –, 12, – CZE: –, 16, –
Ferreira and Gignoux (2014)	Parametric, ex–ante	Parent's educational, professional and migration status, native tongue, number of books at home, material wealth, cultural resources, school locality type, gender	**2006** HUN: 33, 34, 33 SVK: 32, 29, 30 POL: 24, 27, 24 CZE: 27, 30, 28		
Luongo (2015)	Parametric, ex–ante	Parent's education and professional status, gender	**2003** HUN: 19, 18, 15 SVK: 16, 16, 16 POL: 12, 16, 12 CZE: 12, 12, 10	**2006** HUN: 16, 18, 14 SVK: 1, 15, 13 POL: 11, 15, 12 CZE: 9, 11, 8	
			2009 HUN: 19, 24, 18 SVK: 10, 18, 10 POL: 13, 20, 12 CZE: 9, 15, 8	**2012** HUN: 15, 19, 14 SVK: 17, 21, 1 POL: 13, 17, 12 CZE: 10, 13, 9	
Vega (2013)	Parametric	Native tongue, family composition, ESCS index, classroom atmosphere, gender	**2009** HUN: –, 34, – SVK: –, 30, – POL: –, 27, – CZE: –, 20, –		

Note: * the first measurement refers to mathematical literacy, the second – reading literacy, the third – natural science literacy.
Source: Author's computations

The methodological approach to assessing inequality of opportunity in educational achievements in keeping with the «spirit and letter» of the equal opportunity theory was proposed by Ferreira and Gignoux (2014) and tested on PISA-2006 data. Later on, using the same methodology, Luongo (2015) measured the inequality of educational opportunities in different countries over time based on the PISA 2003, 2006, 2009, 2012 data. A slightly different methodology, also based on regression analysis, was used by Vega (2013) to assess inequality of opportunity in terms of reading literacy based on the PISA-2009 data. The estimates obtained in these works for the Visegrad countries are also presented in Table 2.

As shown in Table 2, the inequality of opportunity analysis always includes family background factors and gender as uncontrollable individual characteristics. Locality and schooling factors are taken into account «optionally», but their inclusion leads to a significant rise of the assessment value. Comparison of inequality of opportunity in different countries shows that its level in Hungary and Slovakia is higher than in Poland and the Czech Republic.

2 OBJECTIVE, INFORMATION BASE AND RESEARCH METHODOLOGY

Our work is intended to study the dynamics and make a comparative analysis of the inequality of educational achievements and opportunities in the Visegrad Group countries based on the data from the Program for International Student Assessment (PISA) project.

The PISA project aims at studying educational achievement of students aged 15. The data on student educational achievements has been collected for every three year period since 2000. The Visegrad countries have participated in all waves of the project. In addition to per se testing of students, the project accumulates contextual data as per the model structuring educational performance factors (OECD, 2013a). All the project wave data and questionnaires are freely available on the OECD website (OECD, n.d.)

Table 3 Variables used in the study as circumstance factors

Short name in DB	Item	Type	Derivation method
Family background parameters			
HOMEPOS	Household property	Continuous	The variable derived from answers to questions on family's home appliances and computer machines, cars, books, musical instruments, and housing peculiarities.
CULTPOSS	Cultural resources of the family	Continuous	The variable computed from answers to questions about family's classical literature, books on arts and music, pieces of art and musical instruments.
HEDRES	Family educational resources	Continuous	The variable computed from answers to questions about available learning conditions, computer, educational literature in the family.
HISEI	Parents professional status	Continuous	The variable computed from answers to questions about parent employment and work details.
HESCED	Parents education level	Discrete	The variable computed on the basis of answers to questions about parent education and the ISCED 1997 educational level classification.
Location parameters			
SC004Q001	Type of school location	Discrete	Question on the type of community where the school is located (answering options: village, little town, medium town, major city, megalopolis).
Individual peculiarities			
ST004Q001	Student's gender	Discrete	Male/female

Note: The variables names may differ in various year DBs, the table has them as per the 2018 data base.
Source: Author's computations

When selecting the variables engaged in the analysis as circumstance factors, we sought to include various kinds of circumstances: family background factors, location characteristics, and individual peculiarities. We also aimed at avoiding any variables risking a significant subjective component (e.g., assessment of teaching quality by students). Besides, we limited ourselves to the variables present in all the effected waves to ensure comparability of results from year to year. The analysis-engaged circumstance factors are shown in Table 3.

As follows from Table 3, there is a certain bias towards the family background parameters. However, this gives additional interest to the study – it allows a more subtle analysis of which exactly family peculiarities have greater impact in terms of influencing a child's educational achievements. Various mechanisms of family background influencing the academic performance have been identified on the theoretical level. Firstly, the economic factor is singled out: in richer families a child is provided with better support (primarily in terms of nutrition and medical care) in prenatal life, childbirth, and the early years, which is crucial for setting cognitive abilities. In addition, better-off population segments bear lower relative expenses on education, which ensures a better material support for the educational process. Secondly, the cultural and educational factor does also matter: it is believed that the education-biased environment is in itself inherent to the culture of wealthy and well-educated segments of the population, so children with a good family background adapt more easily. Thirdly, there is also a psychological factor – it is believed that the parents education serves as a "reference point" for the child – if a person does not reach his/her parent educational level, then this is perceived as a social failure. The desire to avoid such failure makes a strong motivation to study.

Since the relation of educational achievement to continuous variables (HOMEPOS, CULTPOSS, HEDRES, HISEI) may be of not linear type, these indices were discretized. HOMEPOS, CULTPOSS, HISEI were categorized into three levels: low (less than the lower quartile), medium (from the lower to the upper quartile), high (above the upper quartile of distribution). When evaluating regressions, the medium level was taken as the baseline. HEDRES was categorized into two levels: low (below the median), high (equal to and above the median), the low level was taken as the baseline in the regression.

Due to rare occurrence of the HESCED variable categories, they were combined. The calculations covered three levels of education: ISCED3B_C and below, ISCED3A_4_5B, ISCED5A and above. The ISCED3A_4_5B category was used as the basis.

The type of school location area and schoolchildren gender were used without any modification. Medium-size town and male gender were used as base categories.

The PISA data set has a number of features essential for its statistical processing. These features are detailed in OECD (2009), while in this study we just enumerate briefly those which matter herein:

1. In the PISA database, an individual's educational performance in each area of literacy is described not by a single variable, but by an entire set of variables. Based on the respondents' answers to the test items and on the difficulty thereof, the characteristics of distribution of unobserved individual literacy levels are estimated using IRT-theory models, after which ten (or five in earlier waves of the project) random values are extracted from this distribution. This results in the necessity to make the educational achievement-related calculations several times, with subsequent averaging the results.

2. The PISA study used a two-stage sampling design. In the first stage, an educational institution was randomly selected, whereas in the second stage it was students who were selected. That sampling design results in the conventional methods for estimating standard errors and confidence intervals being incorrect. In this connection, a specific procedure of balanced repeated replication with Fay's adjustment is used to estimate variation statistics, and the data array contains special variables – weights needed to implement this procedure.

These features are quite complex from the calculation point of view, but up to date some specialized packages have already been developed and are freely available, which greatly simplify the implementation of processes described when working with the PISA data. We have applied REPEST module developed and maintained by Francesco Avvisati and Francois Keslair (Avvisati and Keslair, 2014) and recommended on Official Cite of PISA.[4]

The PISA database has an immense advantage of significant volume of observations and little number of gaps in the data. The table shows both the total number of tested schoolchildren by all the project waves and the number of observations remaining after deleting data with gaps in the variables listed in Table 4.

Table 4 Volume of observations

Year	2003	2006	2009	2012	2015	2018
HUN						
Total number of observations	4 765	4 490	4 605	4 810	5 658	5 132
Sample volume after deleting observations with gaps	4 606	4 369	4 546	4 653	5 226	5 025
SVK						
Total number of observations	7 346	4 731	4 555	4 678	6 350	5 965
Sample volume after deleting observations with gaps	7 308	4 703	4 524	4 606	6 036	5 600
POL						
Total number of observations	4 383	5 547	4 917	4 607	4 478	5 625
Sample volume after deleting observations with gaps	4 383	5 469	4 807	4 438	4 319	5 504
CZE						
Total number of observations	6 320	5 932	6 064	5 327	6 894	7 019
Sample volume after deleting observations with gaps	6 042	5 656	5 653	4 501	6 529	6 650

Source: Author's computations

As can be seen from Table 4, the loss due to deleting observations caused by data gaps is negligible.

The technique used herein was first proposed in the paper by (Ferreira and Gignoux, 2014). It was chosen due to its having been developed to handle peculiarities of the PISA data. This technique is based on the ex-ante criterion of equality of opportunity and the parametric approach to describing the relation between achievement and circumstance factors. The parametric approach implies the use of particular equation that describes the dependence of achievement on uncontrollable circumstances. The work (Ferreira and Gignoux, 2014) uses a linear equation and the estimation of regression coefficients by the ordinary method of least squares, and we also follow this strategy. Predicted values $\hat{y}_i$, derived from the regression, represent conditional average of an achievement for a fixed set of circumstance factor values. According to the ex-ante criterion, in case the equality of opportunity is achieved, there should be no variation in the distribution of $\bar{y}_i$. The use of variance as a measure of inequality leads to the situation

[4] How to prepare and analyse the PISA database (OECD): <*https://www.oecd.org/pisa/data/httpoecdorgpisadatabase-instructions.htm*>.

when the determination coefficient R^2, which is the ratio of predicted values variance to the actual values variance, can be interpreted as a relative measure of inequality of opportunity, reflecting the contribution of inequality of opportunity to inequality of achievement.

The relative importance of circumstance factors was estimated using Shapley decomposition, which comes to be the most universal procedure to solve this task. A detailed description of this method of factorial decomposition is given in (Shorrocks, 2012). The Shapley decomposition has many attractive features, however, it has one significant drawback – it is computationally expensive in case the factors are numerous. In this connection, we have pooled the model factors into a few groups (household material wealth, cultural, educational resources; educational and professional status of parents; type of area where the school is located; gender of the student) and have done the decomposition by these groups.

3 RESULTS AND DISCUSSION

Means (M) and standard deviations (SD) of student educational achievements in three subjects – math, reading and science – are shown in Table 5.

Table 5 Educational achievements

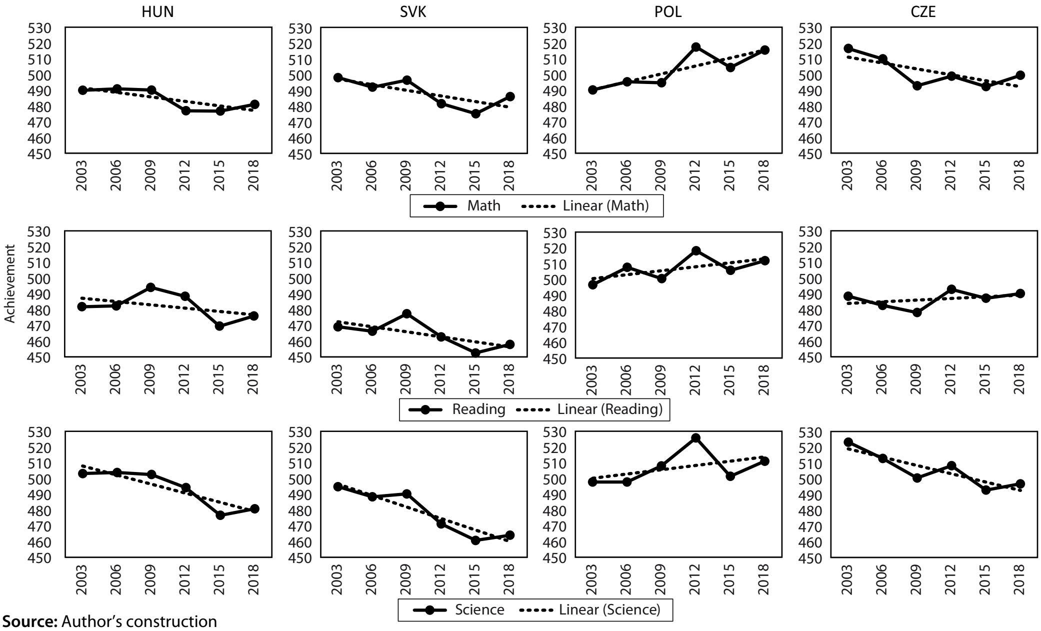
	2003		2006		2009		2012		2015		2018	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
HUN												
Math	490.01	93.51	490.93	91.01	490.17	92.07	477.04	93.62	476.83	93.79	481.08	91.14
Reading	481.87	91.98	482.37	94.35	494.17	90.18	488.46	91.85	469.52	97.03	475.98	97.58
Science	503.27	97.27	503.93	88.17	502.64	86.49	494.30	90.15	476.74	96.33	480.91	93.85
SVK												
Math	498.18	93.31	492.10	94.53	496.68	96.09	481.64	100.84	475.23	95.43	486.16	99.57
Reading	469.16	92.52	466.35	105.07	477.44	90.19	462.77	104.28	452.51	104.23	457.98	100.33
Science	494.86	102.19	488.43	93.14	490.26	95.43	471.19	101.18	460.77	98.93	464.05	95.83
POL												
Math	490.24	90.24	495.42	86.52	494.80	88.35	517.50	90.37	504.47	87.64	515.65	90.09
Reading	496.60	95.92	507.64	100.22	500.47	89.16	518.18	87.29	505.69	89.58	511.85	97.33
Science	497.78	102.40	497.80	89.88	508.07	86.87	525.82	86.35	501.43	90.79	511.03	91.51
CZE												
Math	516.45	95.94	509.86	103.17	492.81	93.19	498.95	94.94	492.32	90.68	499.46	93.16
Reading	488.54	95.51	482.71	111.26	478.18	92.28	492.89	88.70	487.25	100.45	490.22	97.33
Science	523.25	100.64	512.86	98.45	500.49	97.33	508.30	90.58	492.83	92,.6	496.79	94.49

Note: M – mean grade, SD – standard deviation.
Source: Author's computations

As follows from Figure 1, the dynamics of mathematical literacy in the Visegrad Group countries differs significantly. The Czech Republic, which was the undisputed leader earlier in the period under review, has lost the leading position to Poland, which, being an outsider at the beginning, took the lead due to a steady positive trend in terms of mathematical literacy. Slovakia and Hungary show a negative trend in mathematical literacy, due to which their lagging behind the leaders increases.

Through the entire period under review the Polish schoolchildren show the highest level of reading literacy with a tendency to improve, due to which the gap from the rest of the Visegrad Four countries is only increasing. Reading achievements by the Czech and Hungarian schoolchildren were close in 2003–2012, then Hungary began to lag behind the Czech Republic. Slovakia was a sheer outsider at the beginning of the period under review, and its falling behind the leaders only intensified due to negative trends in reading literacy among schoolchildren.

Figure 1 Literacy dynamics in the Visegrad Group countries



Source: Author's construction

As follows from Figure 1, there is a steady positive trend in science literacy in Poland only, which allowed this country to move from an outsider to a leader. The Czech Republic, which was in the lead at the beginning, showed a negative trend, due to which it lost its leading position and moved to the second place. Through the entire period under review, the educational achievements of the Slovak schoolchildren in science literacy have been lower than those of students from other countries, and the negative dynamics leads to further falling behind.

The educational literature accounts the steady progress in educational achievements by Polish schoolchildren for the consistent implementation of education development program, in the first place (Csapó et al., 2019), and secondly, for the high level of government spending on education (Pelle and Kuruczleki, 2016).

The influence of uncontrolled circumstance factors on educational achievements of schoolchildren was studied using regression analysis. Because of the large number of estimated regression models, it is not possible to show all the tables in their entirety. Tables A1–A3 in the Appendix show the results of regression analysis based on the 2018 data. As follows from the results, in most cases, the factors involved do demonstrate the expected impact on students' educational achievements. A low level of Household property produces a meaningful negative effect on educational achievements by students, whereas a high level thereof has a significant positive effect, and that refers to all types of literacy. In Hungary and Poland, a high level of Cultural resources of the family improves – while a low level thereof deteriorates – the educational results of schoolchildren, but in Slovakia and the Czech Republic this factor does not seem to play a significant role. A high level of home educational resources has a meaningfully positive effect on the educational achievements of students in Hungary and the Czech Republic, to a lesser extent in Poland, and is virtually insignificant in Slovakia. A low professional and educational status of parents has an essentially negative impact on educational performance in all countries. In Slovakia, Poland and the Czech Republic, a high level of parental education facilitates improving the educational performance. The rural location of schools significantly impairs educational achievements in all countries. The size of cities does matter in Hungary and Slovakia, while in Poland and the Czech Republic this factor is insignificant. The gender of students is meaningful in terms of reading literacy – in all countries girls demonstrate better reading results than boys, with other conditions being equal. Achievements in mathematics are better for boys than for girls in Hungary and the Czech Republic, while in Slovakia and Poland the gender factor is insignificant. In Hungary only, boys perform better than girls in natural science, other conditions being the same.

The gender gap in reading literacy occurs nearly worldwide. The PISA official publications assign the existence of this gap to reading habits and the maturity of teaching strategies for handling texts (OECD, 2019). Besides, there are works explaining this phenomenon with differences in the socialization of boys and girls at home and at school (see, for example, Hadjar et al., 2014).

The dynamics of inequality of opportunity in respect of educational achievements of students is shown in Table 6.

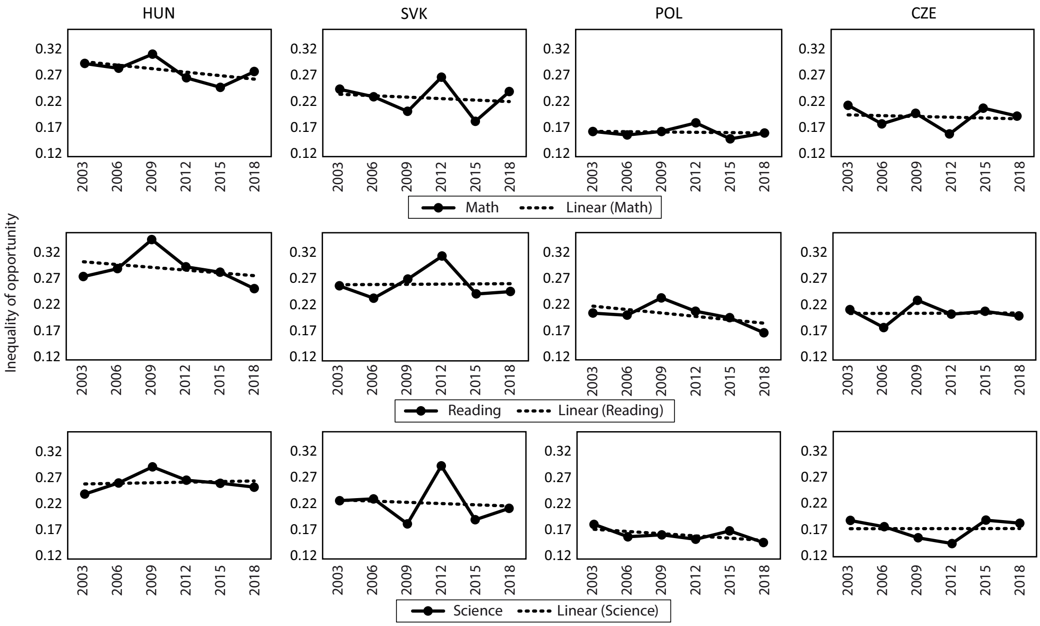
Table 6 Inequality of opportunity as for educational achievements by schoolchildren in Visegrad Group countries

		2003	2006	2009	2012	2015	2018
HUN	Math	0.2947	0.2855	0.3126	0.2672	0.2490	0.2792
	Reading	0.2759	0.2911	0.3467	0.2946	0.2843	0.2528
	Science	0.2407	0.2621	0.2930	0.2674	0.2615	0.2541
SVK	Math	0.2454	0.2310	0.2030	0.2685	0.1839	0.2409
	Reading	0.2583	0.2347	0.2713	0.3154	0.2428	0.2474
	Science	0.2280	0.2313	0.1835	0.2946	0.1914	0.2132
POL	Math	0.1646	0.1579	0.1646	0.1812	0.1504	0.1617
	Reading	0.2059	0.2018	0.2349	0.2095	0.1969	0.1681
	Science	0.1827	0.1589	0.1625	0.1541	0.1702	0.1478
CZE	Math	0.2147	0.1791	0.1993	0.1598	0.2092	0.1940
	Reading	0.2129	0.1786	0.2307	0.2042	0.2096	0.2008
	Science	0.1897	0.1776	0.1564	0.1454	0.1903	0.1844

Source: Author's computations

As can be seen from Figure 2, Hungary has the highest level of inequality of opportunity among the Visegrad Group countries, followed by Slovakia. This conclusion is in good agreement with the calculations made by other researchers (see Table 2). In terms of reading and science literacy, the levels of inequality of opportunity in the Czech Republic and Poland are close, but in terms of mathematical literacy, the inequality of opportunity in the Czech Republic is higher than in Poland. In most cases, the highest level of inequality of opportunity occurs in reading literacy. Maybe this is due to the fact that innate abilities, which are not taken into account in this work, are more important for mathematical and natural science literacy, while reading literacy is largely shaped by social environment, primarily the family background. There are no clear trends of increasing or decreasing inequality of opportunity in any of the countries reviewed.

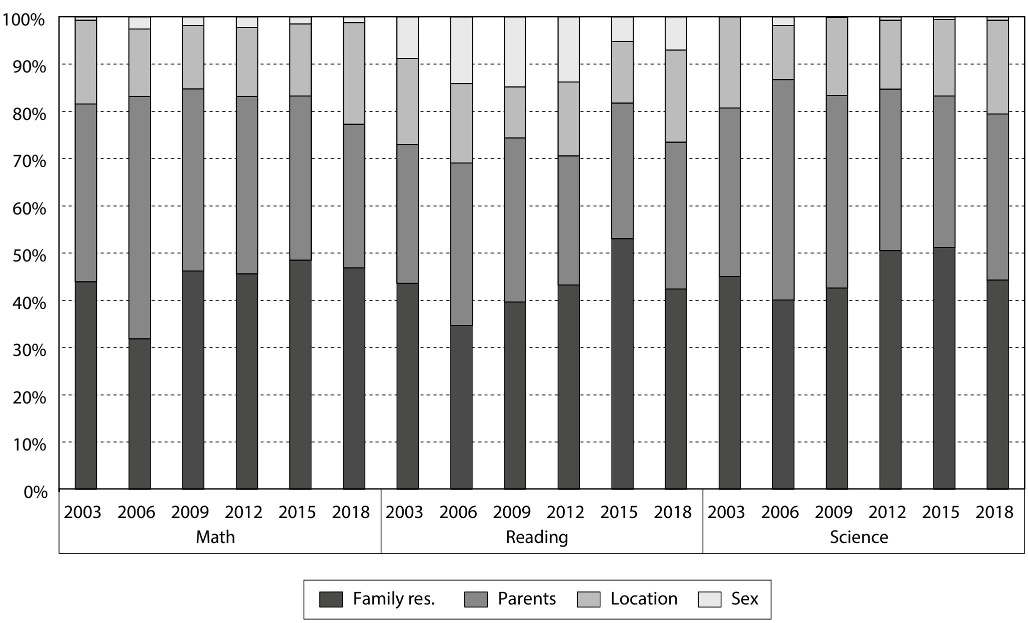
Figure 2 Dynamics of inequality of opportunity in the Visegrad Group countries



Source: Author's construction

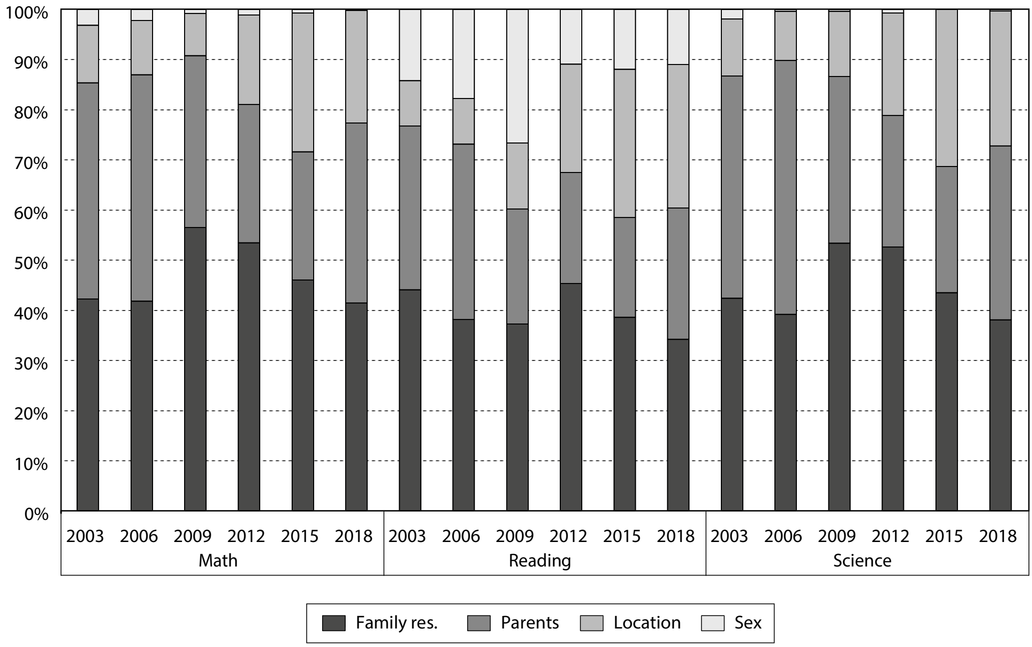
The input of individual factors to the overall level of inequality of opportunity is shown in Figures 3–6.

Figure 3 Input of individual circumstance factors to inequality of educational opportunities in Hungary



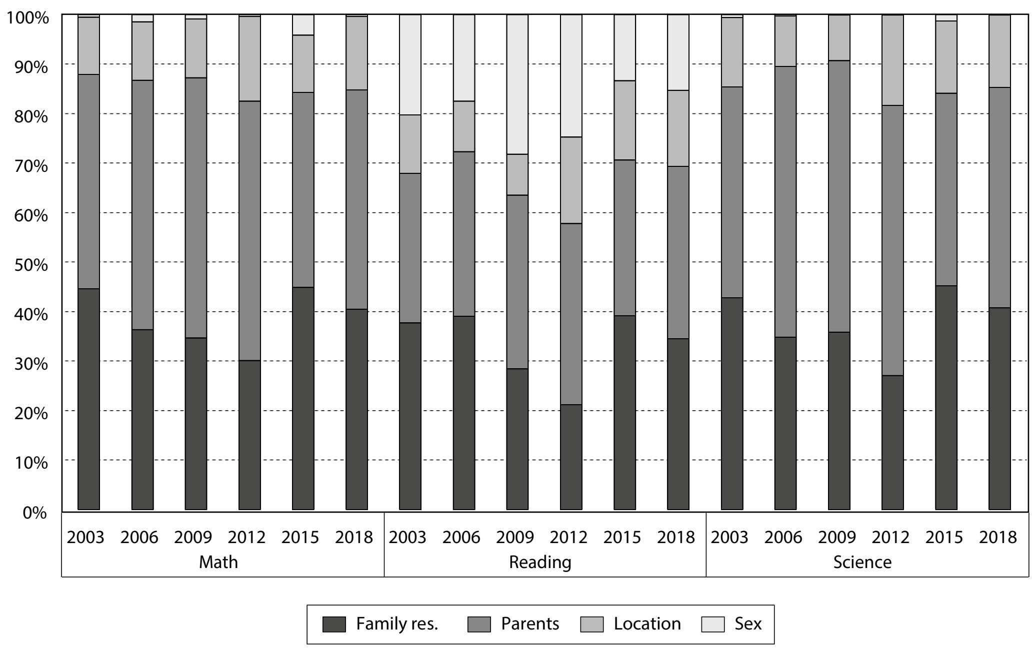
Source: Author's construction

Figure 4 Input of individual circumstance factors to inequality of educational opportunities in Slovakia



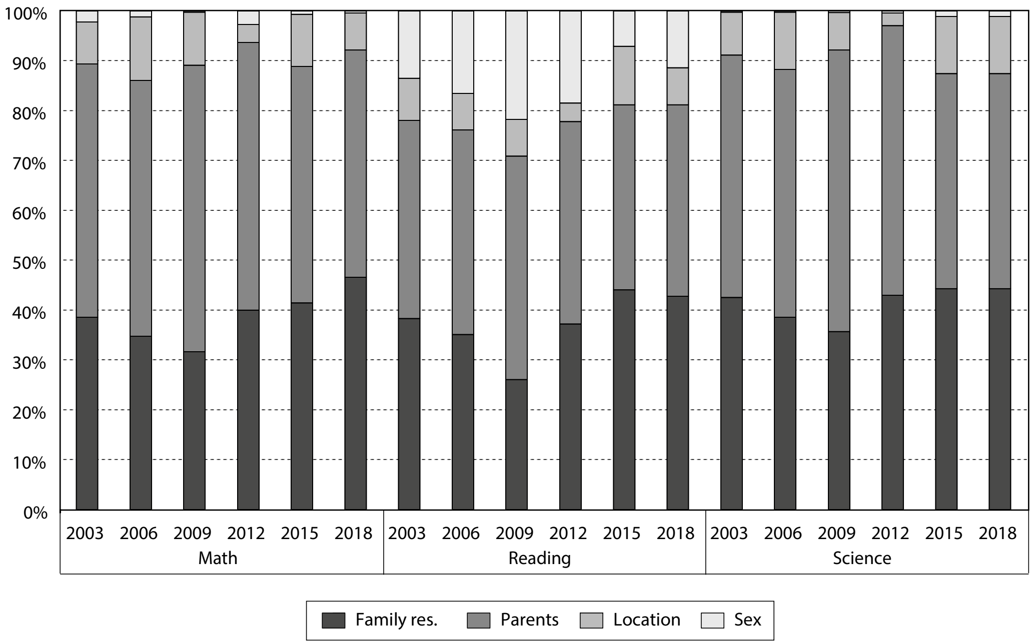
Source: Author's construction

Figure 5 Input of individual circumstance factors to inequality of educational opportunities in Poland



Source: Author's construction

Figure 6 Input of individual circumstance factors to inequality of educational opportunities in the Czech Republic



Source: Author's construction

As follows from Figures 3–6, the input of individual factors to the overall level of inequality of opportunity has both the features common for all countries and unique features. In all countries reviewed, student gender contributes significantly to inequality of opportunity in terms of reading literacy and does not matter much for the math and science literacy. Regression analysis results show that girls demonstrate better achievements in terms of reading literacy compared to boys.

The role of location factor has an express increasing tendency in relation to all types of literacy in Slovakia. Other countries do not show such trend. The role of location factor is higher in Hungary compared to Poland and the Czech Republic.

In all countries, the impact of family background is great. This can be partially explained by the fact that family background parameters are more fully taken into consideration herein. At the same time, all countries show considerable importance of both components – family material resources, and the parent educational and professional status. However, in Slovakia and Hungary, the role of material component prevails, while in the Czech Republic and Poland more significant is the contribution by non-material factors – education and professional status of parents.

CONCLUSION

This paper uses the PISA database to analyse the inequality of educational achievement and opportunity in the countries of the Visegrad Group. The following parameters were taken into account as circumstance factors: educational and professional status of parents, family material, cultural and educational resources, location of school and gender of students. The methodology used can be characterized as a parametric one based on the ex-ante approach.

The 2003–2018 educational performance by schoolchildren was found to tend towards worsening in Hungary, the Czech Republic and Slovakia, which points to the need for an in-depth analysis of reasons behind this deterioration and for adopting strategies to improve national systems of education. Otherwise, in the long run, these processes will result in lower competitiveness and economic potential.

It was also found that the inequality of educational opportunities is noticeably higher in Hungary and Slovakia than in Poland and the Czech Republic. All countries showed a slight trend towards decreasing the inequality of opportunity. This can be accounted for general long-term opportunity-equalizing trends, like growing levels of education, urbanization, a more even localization of educational establishments across the countries, reduced size of families, development and spread of digital technologies overall and particularly in education.

The remaining inequality of educational opportunities necessitates monitoring the situation and making efforts to reduce it due to the fact that schooling provides initial levels of education, so inequality in schooling produces inequality in access to higher educational levels. Further on, since the level of education achieved becomes the most important determinant of other individual achievements, the inequality of educational opportunities carries over to the inequality of economic, political, social opportunities and achievements, setting up barriers to economic and social mobility and preserving socio-economic stratification. This is why monitoring the dynamics of inequality of opportunity in education is of paramount importance for the understanding and long-term forecasting of social processes.

ACKNOWLEDGMENTS

The reported study was funded by the Slovak project 1/0034/23 VEGA Methodological and Empirical Aspects of Subjective Welfare Measurement. The authors would like to thank the reviewers for useful and meaningful comments, which helped improve the paper's final version to a great extent.

References

ARNESON, R. (1989). Equality and equal opportunity of welfare. *Philosophical Studies*, 56: 77–93.

AVVISATI, F., KESLAIR, F. (2014). REPEST: Stata module to run estimations with weighted replicate samples and plausible values. *Statistical Software Components S457918*, Boston College Department of Economics.

BARROS, R. P., FERREIRA, F. H. G., MOLINAS, V. J. R., CHANDUVI, J. S. (2009). *Measuring Inequality of Opportunities in Latin America and the Caribbean*. World Bank Publications.

BRZENZIŃSKI, M. (2015). Inequality of opportunity in Europe before and after the Great Recession. In: *Working Paper No 2/2015 (150),* University of Warsaw.

CHECCHI, D., PERAGINE, V., SERLENGA, L. (2010). Fair and Unfair Income Inequalities in Europe. *IZA Discussion Papers 5025*, Institute of Labor Economics.

COHEN, G. (1989). On the currency of egalitarian justice. *Ethics*, 99(4): 906–944.

CSAPÓ, B., FEJES, J. B., KINYÓ, L., TÓTH, E. (2019). Educational Achievement in Social and International Contexts. In: KOLOSI, T., TÓTH, I. G. (eds.) *Social Riport 2019*, Budapest: TÁRKI, 217–236.

DE LA VEGA, C., LEKUONA, A. (2013). *Analyzing Inequality of Opportunity in Educational Achievements*. Paper presented at the Fifth Meeting of the Society for the Study of Economic Inequality (ECINEQ), Bari, Italy.

DWORKIN, R. (1981a). What is equality. Part 1: Equality of welfare. *Philosophy and Public Affairs*, 10: 185–246.

DWORKIN, R. (1981b). What is equality. Part 2: Equality of resources. *Philosophy and Public Affairs*, 10: 283–345.

EBRD. (2016). *Transition Report* [online]. London: EBRD. [cit. 7.5.2020]. <http://2016.tr-ebrd.com/inequality-ofopportunity>.

EUROPEAN COMMISSION. (2020). *Europe's moment. Repair and Prepare for the Next Generation. Communication from the European Commission to the European Parliament*, *the European Council, the Council, the Economic and Social Committee and the Committee of the Regions*. COM 456 final.

FERREIRA, F., GIGNOUX, J. (2014). The Measurement of Educational Inequality: Achievement and Opportunity [online]. *World Bank Economic Review*, 28: 210–246. <https://doi.org/10.1093/wber/lht004>.

FLEURBAEY, M. (2008). *Fairness, Responsibility, and Welfare*. London: Oxford University Press.

HADJAR, A., KROLAK-SCHWERDT, S., PRIEM, K., GLOCK, S. (2014). Gender and educational achievement. *Educational Research*, 56(2): 117–125.

IBRAGIMOVA, Z. F., FRANTS, M. V. (2020). Inequality of opportunities: Interpretation, methods and problems of estimation [online]. *St. Petersburg University Journal of Economic Studies,* 36(4): 624–652. <https://doi.org/10.21638/spbu05.2020.404>.

IBRAGIMOVA, Z. F., FRANTS, M. V. (2021a). Inequality of Educational Opportunity in Soviet and Post-Soviet Russia: an Empirical Analysis [online]. *Voprosy obrazovaniya/Educational Studies Moscow*, 2: 43–62. <https://doi.org/10.17323/1814-9545-2021-2-43-62>.

IBRAGIMOVA, Z. F., FRANTS, M. V. (2021b). Measuring Inequality of Opportunity: Does Inequality Index Matter? [online]. *Statistika: Statistics and Economy Journal*, 101(1): 66–90. <ae45bbfd-a734-4ce8-bbd6-bff26a7fe7bf(czso.cz)>.

LUONGO, P. (2015). Inequality of opportunity in educational achievements: Cross-country and intertemporal comparisons. *WIDER Working Paper 2015/043*.

MARRERO, G., RODRÍGUEZ, J. G. (2012). Inequality of Opportunity in Europe [online]. *Review of Income and Wealth*, 58(4): 597–621. <https://doi.org/10.1111/roiw.2012.58.issue-4>.

OECD (n.d.). *Programme for International Students Assessment* [online]. Paris: Organisation for Economic Co-operation and Development. <http://www.oecd.org/pisa>.

OECD. (2004). *Learning for Tomorrow's World First Results from PISA 2003*. Paris: Organisation for Economic Co-operation and Development.

OECD. (2007). *PISA 2006: science competencies for tomorrow's world. Volume 1 – Analysis*. Paris: Organisation for Economic Co-operation and Development.

OECD. (2010). *PISA 2009 Results: overcoming social background – equity in learning opportunities and outcomes (Volume ii)*. Paris: Organisation for Economic Co-operation and Development.

OECD. (2013a). *PISA-2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing.

OECD. (2013b). *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed (Volume II)*: Paris: OECD Publishing.

OECD. (2009). *PISA Data Analysis Manual: SPSS*. Paris: OECD Publishing.

OECD. (2016). *Socio-economic status, student performance and students' attitudes towards science, PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.

OECD. (2019). *Students' socio-economic status and performance, PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD Publishing.

OECD, UNESCO. (2003). *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris: OECD Publishing.

PAUHOFOVA, I., BUKHARBAEVA, L., IBRAGIMOVA, Z., FRANTS, M. (2020). Measuring of inequality of opportunity: parametric approach. *Ekonomický časopis*, 68(5): 455–476.

PELLE, A., KURUCZLEKI, E. (2016). Education policies and performance of the Visegrad countries in light of their OECD membership: a comparative study, Yearbook of the Institute of East-Central Europe, Special Issue. In: VISVIZI, A. (eds.) *Re-thinking the OECD's role in global governance: members, policies, influence*, 14(4): 173–207.

RAWLS, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

ROEMER, J. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy and Public Affairs*, 22(2): 146–166.

ROEMER, J. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

SHORROCKS, A. F. (2012). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value [online]. *The Journal of Economic Inequality*, 11(1): 99–126. <https://doi.org/10.1007/s10888-011-9214-z>.

VAN DE GAER, D. (1993). *Equality of opportunity and investment in human capital*. Ph.D. Dissertation, Catholic University of Leuven.

APPENDIX

Table A1 Results of the OLS-regression of the educational achievement on circumstance-factors (Math)

Independent variables	HUN	SVK	POL	CZE
Household property				
Low	–32.64***	–54.03***	–21.25***	–33.25***
High	14.90***	25.38***	12.54***	16.29***
Cultural resources of the family				
Low	–16.48***	11.24***	–19.47***	–1.64
High	13.02***	–3.87	6.37	4.74
Family educational resources				
High	10.79***	1.80	5.99	19.37***
Parents professional status				
Low	–27.69***	–34.40***	–9.96***	–19.88***
High	8.64*	–7.44**	–0.82	5.87*
Parents educational level				
ISCED3B_C and below	–19.27***	–40.61***	–16.18***	–23.08***
ISCED5A and above	0.54	15.62***	31.09***	22.16***
Type of school location				
Village	–65.67***	–35.26***	–13.01***	–34.17***
Little town	–19.81**	–12.07*	–4.49	–9.07*
Major city	6.34	23.37***	14.13**	9.29
Megalopolis	23.53***	–	8.09	5.84
Gender				
Female	–12.58***	–3.13	–3.9	–7.32**
Constant				
Const.	494.26***	504.84***	514.31***	488.22***
R^2	0.2792	0.2409	0.1617	0.1940

Notes: *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations

Table A2 Results of the OLS-regression of the educational achievement on circumstance-factors (Reading)

Independent variables	HUN	SVK	POL	CZE
Household property				
Low	–28.63***	–49.71***	–16.86***	–28.03***
High	11.29***	15.73***	5.43	14.11**
Cultural resources of the family				
Low	–20.78***	4.07	–24.81***	–5.09
High	14.21***	–0.22	9.77**	5.50
Family educational resources				
High	7.78**	3.8	9.15***	23.04***
Parents professional status				
Low	–32.25***	–34.75***	–12.15***	–20.14***
High	9.63**	9.71***	–7.21*	6.21*
Parents educational level				
ISCED3B_C and below	–22.11***	–34.75***	–17.16***	–23.45***
ISCED5A and above	–3.50	9.71***	30.10***	17.39***
Type of school location				
Village	–63.70***	–47.85***	–20.79***	–35.42****
Little town	–20.24*	–18.98***	–11.14*	–8.12
Major city	1.85	28.47***	11.25	9.47
Megalopolis	22.96***	–	–3.22	9.32
Gender				
Female	22.42***	33.90***	29.49***	29.16***
Constant				
Const.	478.57***	461.99***	500.07***	469.21***
R^2	0.2528	0.2478	0.1681	0.2008

Notes: *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations

Table A3 Results of the OLS-regression of the educational achievement on circumstance-factors (Science)

Independent variables	HUN	SVK	POL	CZE
Household property				
Low	–26.49***	–47.16***	–17.23***	–31.60***
High	10.15***	17.84***	7.03*	18.70***
Cultural resources of the family				
Low	–22.16***	7.88*	–24.44***	–2.47
High	15.13***	–2.45	9.65**	0.32
Family educational resources				
High	8.63**	2.76	6.22*	16.83***
Parents professional status				
Low	–31.66***	–31.02***	–12.32***	–21.95***
High	8.64**	–5.88*	–5.46	5.76*
Parents educational level				
ISCED3B_C and below	–26.72***	–33.45***	–17.64***	–25.03***
ISCED5A and above	–2.84	14.47***	29.67***	19.87***
Type of school location				
Village	–53.39***	–39.90***	–14.41***	–29.81***
Little town	–18.19*	–13.63**	–4.97	–6.68
Major city	3.21	23.60***	12.35*	10.00
Megalopolis	24.64***	–	–6.60	13.35
Gender				
Female	–10.21***	5.94*	–2.65	–2.01
Constant				
Const.	498.38***	477.49***	513.55***	494.53***
R^2	0.2541	0.2132	0.1478	0.1844

Notes: *** p<0.01, ** p<0.05, * p<0.1.
Source: Author's calculations

Generations on the Labour Market – an Analysis of Employed Persons in Terms of Age and Sex

Wiesława Gierańczyk[1] | *Statistical Office in Bydgoszcz, Bydgoszcz, Poland*
Milena Krajewska[2] | *Statistical Office in Bydgoszcz, Bydgoszcz, Poland*

Received 17.11.2022 (revision received 6.3.2023), **Accepted** (reviewed) 15.3.2023, **Published** 16.6.2023

Abstract

The pace of change in contemporary reality creates a need for fresh data to diagnose the impact of this change on the labour market. Opportunities to meet these expectations are provided by data resources collected in registers and administrative systems directly or indirectly related to employment. Experimental work carried out on the adaptation of data from these sources for statistical purposes shows that, for instance, data on employed persons calculated on the basis of the resources of the Social Insurance Institution (ZUS) and the Agricultural Social Insurance Fund (KRUS) allow analyses to be carried out taking into account, i.a., the generations. Currently, there are four generations on the labour market in Poland: the Baby Boomers, X, Y and Z, which differ in terms of their value system, motivation to work, expectations and loyalty to employers.

Keywords	***DOI***	***JEL code***
Statistics, labour market, generations on the labour market	*https://doi.org/10.54694/stat.2022.54*	*C10, E24*

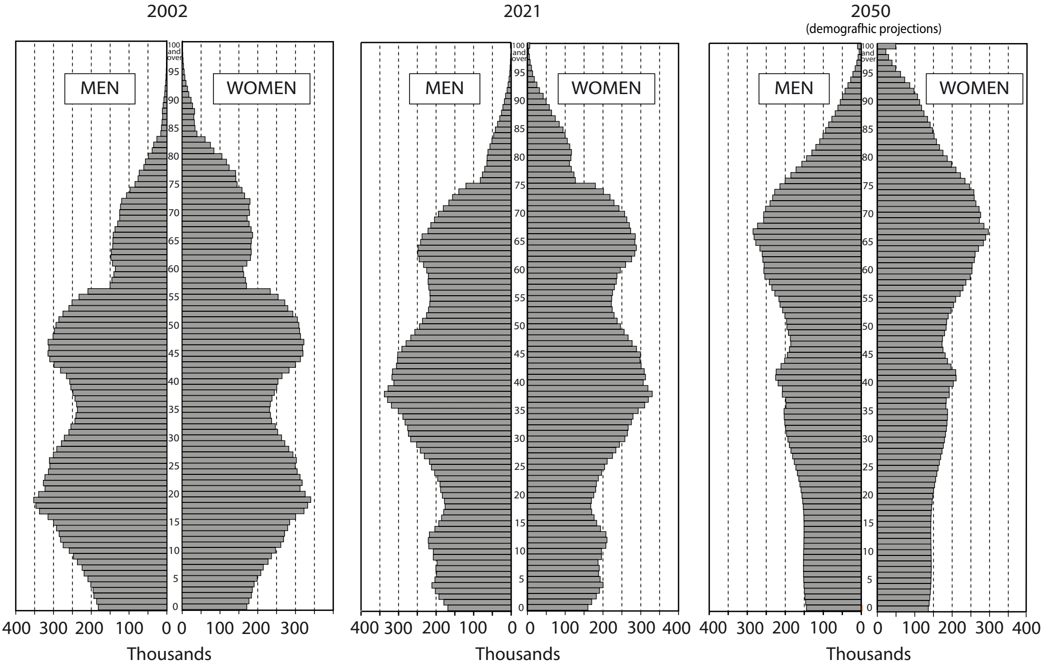
INTRODUCTION

The current labour market in Poland is undergoing constant transformations, which results, i.a., from demographic changes. Increasing life expectancy, decreasing fertility rates, increasing average age of women giving birth to children, decreasing mortality rates, as well as increasing share of elderly people in the total population result in population ageing. Ageing is a characteristic feature of developed countries – in Poland this process has been observed since the early 1990s (Ciura and Szymczak, 2012). The shape, shown in Figure 1, of the population pyramids clearly confirms the process of population

[1] Statistical Office in Bydgoszcz, Stanisława Konarskiego 1–3, 85-066 Bydgoszcz, Poland. E-mail: w.gieranczyk@stat.gov.pl, phone: (+48)533304739.
[2] Statistical Office in Bydgoszcz, Stanisława Konarskiego 1–3, 85-066 Bydgoszcz, Poland. E-mail: m.krajewska@stat.gov.pl, phone: (+48)532089849.

ageing in Poland. In the perspective to 2050, the pyramid will assume a shape with a clearly narrowed base – the regressive type of the pyramid. This type is characterised by a low share of younger age groups in the total population and a progressive ageing of the population.

Figure 1 Population by sex and age (as at 31st December)



Source: Own construction

Demographic changes, their scale and pace are reflected in the size and age structure of the labour force. Therefore, as the population ages, it becomes important to make the best possible use of the potential of people belonging to different generations who are capable to work. The coexistence of two or three generations on the labour market is a natural phenomenon. Presently, the labour force in Poland is made up of four or even five generations.

1 GENERATIONS ON THE LABOUR MARKET

Attempts to describe generations scientifically date as far back as the first half of the 19th century, when August Comte reflected on the relationship between the rate of generational change and the mechanisms of civilisational progress (Jaeger, 1977). Currently, a generation is most often defined as a group of people of similar age, living at the same time, influenced by the same events, and thus with similar perceptions of reality, with similar expectations of work and even similar ways of realising their needs (Kopertyńska and Kmiotek, 2014). In a way, this concept coincides with the ‚cohort' in demography, which refers to a group of people distinguished by an accepted time category, e.g. those born in a particular decade, graduates of the same year, employees retiring during a particular period, etc. (Kudela-Świątek, Saryusz-Wolska). From this perspective, distinguishing the boundaries of generations is a matter of accepted convention. Hence, in the literature, one can find classifications of generations that differ in their temporal scope as well as in their nomenclature. In this article, in order to show the economic activity of generational groups, the following are distinguished:

- the generation of people aged 55 and over, i.e. the Baby Boomers and Traditionalists (Lain-Kennedy, 2007);
- Generation X, which includes people aged 40–54 years;
- Generation Y, people aged 25–39 years, and
- Generation Z, which includes people aged 15–24 years.[3]

It should be emphasised that this is a conventional division of generations adopted by the authors, which does not directly reflect the replacement of generations (i.e. every 20–25 years or so), but refers to people born at a similar time and shaped by similar events and experiences. Research by a number of authors (Borczyk, Chmiel, Czarnecka-Stańczak et al., 2011) shows that the simultaneous existence of several generations of employees, especially those so different, affects the current labour market situation to such an extent that multigenerational management is beginning to be an important element in human resource management (Hysa, 2016).

2 DATA SOURCES

Employed persons are a category of economically active persons characterised by being in work. In the Polish official statistics there are two approaches to measuring the employed population. The first approach is based on collecting data on those employed in the national economy with the use of a survey of national economy entities. In this survey, the definition of employed persons is close to the perspective of understanding work that is formalised by provisions of law, i.e. as persons performing work that brings them earnings (in the form of remuneration for work) or income. In this survey, employed persons include those working under an employment contract and those who conduct business activities, and exclude those working under civil law contracts.

The second approach to measuring the population of employed persons, based on their definition prepared by the International Labour Organization, is applied in the Labour Force Survey (LFS). In this survey, people aged 15–89 years who met certain criteria during the reference week are considered to be employed. In this case, the employed population is determined on the basis of the respondent's declaration with regard to his or her main workplace. In the Labour Force Survey, the basis for qualifying a person as employed is the performance of work, also without a contract and undeclared work, as well as helping unpaid in a family business. Consequently, there may be persons counted as employed in the Labour Force Survey who are not employed according to the Survey on employment in the national economy.

In this article, data describing population of employed persons are presented according to the first approach. However, the Survey on employment in the national economy, which is based on statistical reports submitted by national economy entities, does not provide data on the age of employed persons, in an aggregation that allows for combining employed persons into any age groups. Among other things, these needs of data users were taken into account when looking for other sources of data on employed persons. Experimental work in this area, which has been carried out in official statistics in Poland for many years, showed that to a large extent these expectations could be met by data related directly or indirectly to employment in registers and administrative systems of, i.a., institutions responsible for social insurance. This work made it possible to diagnose that, although the administrators of registers and administrative systems collect data for their own purposes, they can be the basis for calculating data on employed persons corresponding, to a very large extent, in terms of definition and subject matter, to the data currently released by official statistics, collected from the Survey on employment in the national economy. As the institutions responsible for social insurance have information on all insured

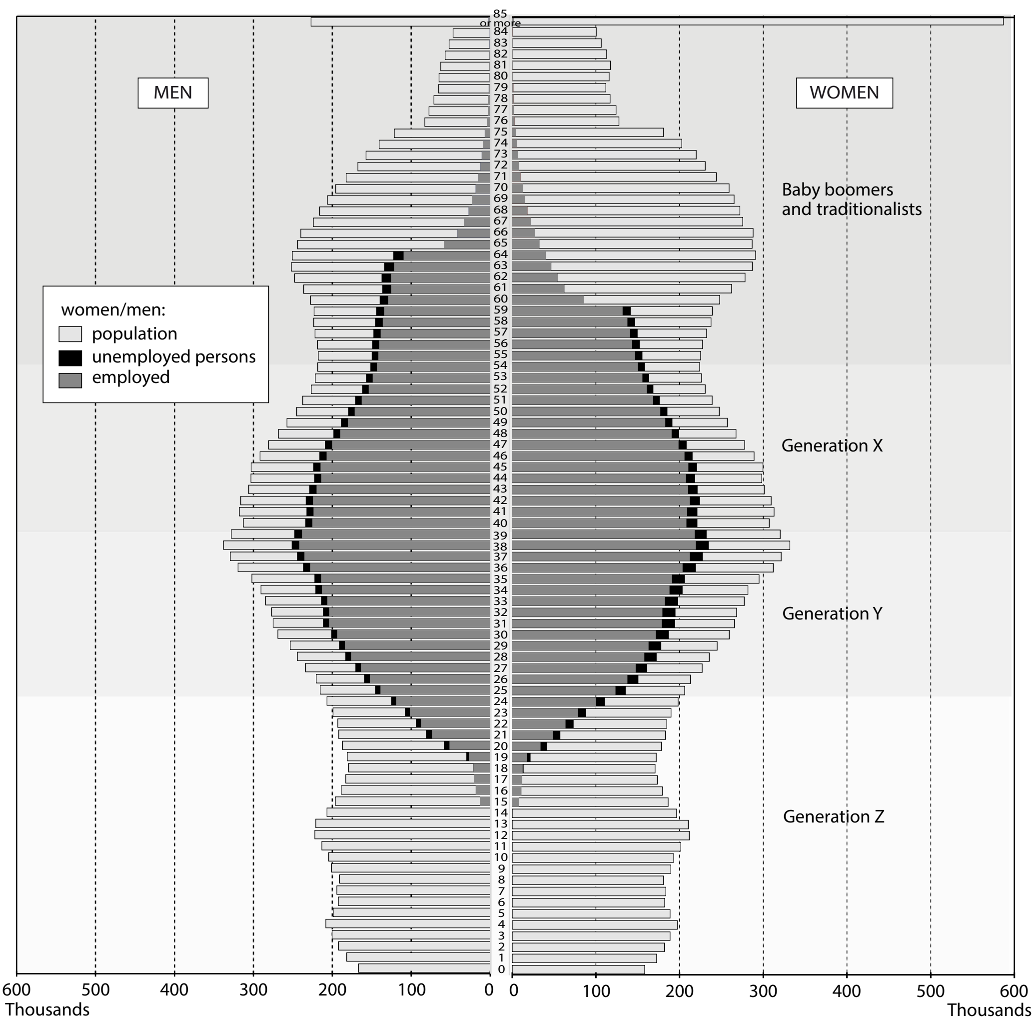
[3] The main aim of the article is to analyse employed persons by generational group. According to the currently binding law, one can be an employed person from the age of 15.

persons in respect of their work, the methodology developed on the basis of this source makes it possible to prepare data on all employed persons. Due to the fact that these are micro-data, they can be aggregated by demographic characteristics, including age, and, therefore, also by generation.

3 RESULTS

The correlations between the age and sex structure of employed persons, unemployed persons and population by generation presented in Figure 2 confirm the demographic changes taking place. Such an approach, despite the fact that it refers to a specific point in time (in this study, the data are as at 31st December 2021), makes it possible to retrospectively and prospectively assess the directions and pace of demographic phenomena and processes and the related economic activity of individual generations.

Figure 2 Employed persons, unemployed persons and population by sex and generation in 2021 (as at 31st December)



Source: Own construction

As defined, generations differ in terms of education, skills, qualifications, but also in their attitudes towards work (Chłoń-Domińczak, Hausner et al., 2021), which can be a major challenge for employers. This article focuses on the economic activity of generational groups, which is presented as the share of employed persons in the total population of a generation. This is one of the simpler measures for describing the situation on the labour market (Cierniak-Piotrowska, Dąbrowska, Stelmach, 2022). However, when interpreting these shares, it is important to be aware that the variation in the number of employed persons across generations results, inter alia, from the natural biological and occupational cycle of the population.

The analysis shows that in December 2021, all employed persons accounted for 46.7% of the population aged 15 and over, with the share of 51.8% among men and 42.0% among women. The highest share of employed persons in the total population was among generations X and Y – employed persons accounted for around 70% of each generation. The share was much lower among the generation of Baby Boomers and Traditionalists and the youngest Generation Z. In both generations, almost every fourth person was employed. The low labour market participation of these two generations is due, inter alia, to the fact that generation Z is made up of young people (15–24 years). This group therefore comprises learners, including students, as well as people who are just entering the labour market. The Baby Boomers, on the other hand, include people who have reached retirement age, so naturally there is a decrease in the number of representatives of this generation on the labour market.

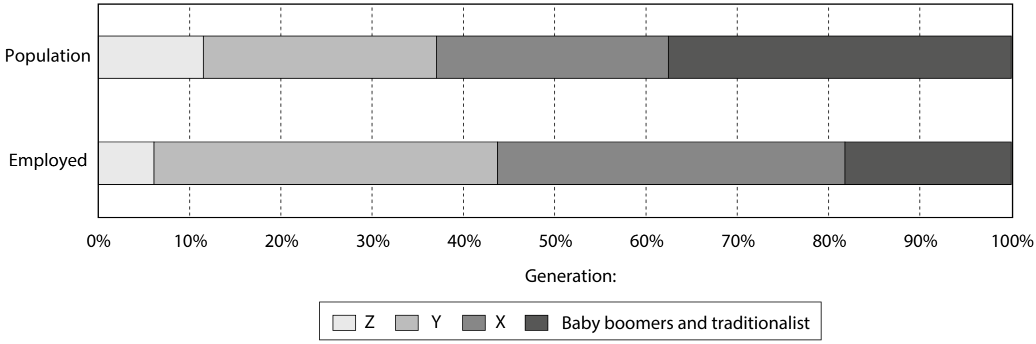
Despite the different commitment of each generation to work, and as a result of biological and occupational factors, the share of employed persons in the total population in each generation was higher among men than among women. Male and female shares in Generation X were closest to each other. In this generation, 70.2% of men and 69.7% of women were employed. Generation X had the highest female share among all generations in Poland. In contrast, the share of employed men was highest in Generation Y – 71.7%. In this generation, women's economic activity was approximately 6 percentage points lower than men's. Almost twice as high a share for men as for women characterised the population aged 55 and over. The Baby Boomers and Traditionalists are already largely exiting the labour market, especially women, for whom the retirement age is 60, five years less than for men. Therefore, of this generation, every sixth woman was in employment, but almost every third man. In Generation Z, which is entering the labour market, every fourth man and every fifth woman were in employment.

As can be seen from the data presented in Figure 3, members of the youngest and oldest generations accounted for a larger share in the population structure than in the structure of employed persons, while members of generations X and Y dominated in the structure of employed persons. Employed persons from these generations accounted for approximately three quarters of the total number of employed persons in Poland, while their share in the population aged 15 and over was just over 50%.

At the end of 2021, the Traditionalist and Baby Boomer generations were still active on the Polish labour market. Persons from these generations are mainly those born during the post-World War II baby boom. In line with the natural biological and occupational cycle associated primarily with reaching the retirement age, members of these generations are decreasing in numbers on the labour market with each passing year.

It is more difficult for this generation to be active on the labour market, as most of their work experience was before the introduction of widespread digitisation. In adjusting to the expectations of the labour market, they often struggle with barriers that limit their career opportunities. From the perspective of the labour market, this generation is characterised by being hard-working and consistent by building their careers step by step. For this generation, a sense of stability is very important, and, therefore, the members of this generation are afraid of losing their posts. As the results of the experimental study show (Figure 4), the relatively highest share of employed persons belonging to this generation was recorded in the Real estate activities section, and in the Electricity, gas, steam and air conditioning supply section.

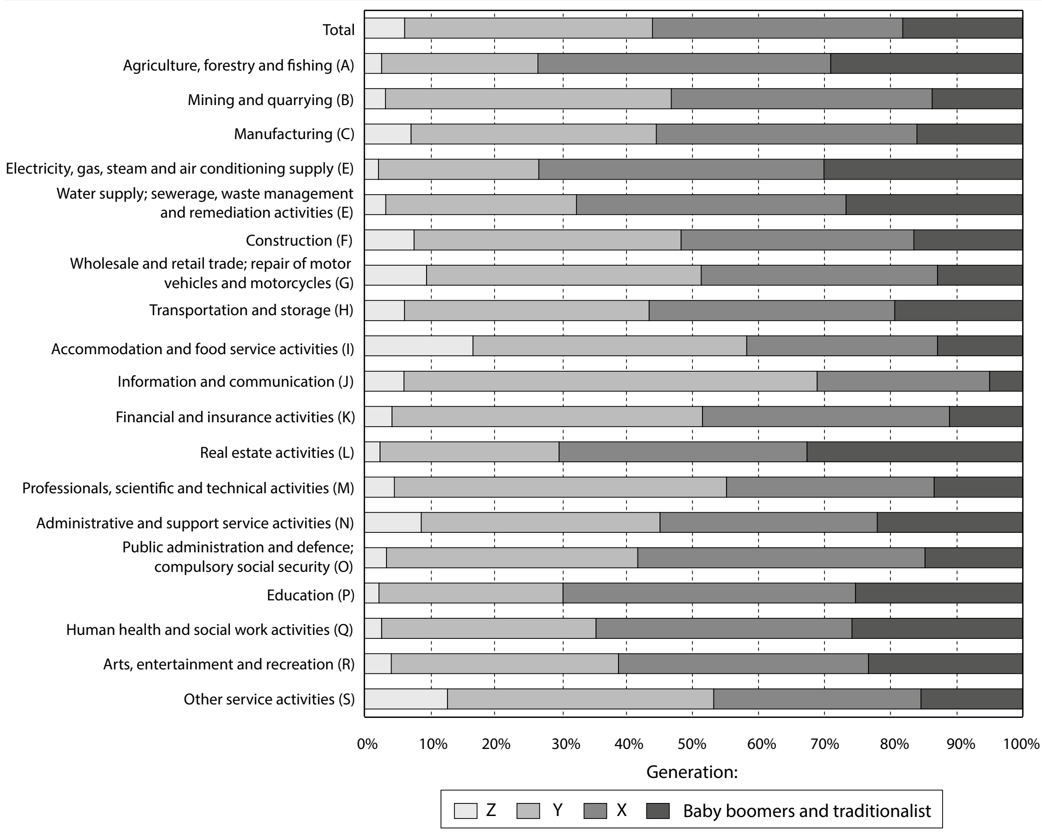
Figure 3 Structure of population and employed persons by generation in 2021 (as at 31st December)



Source: Own construction

Work for this generation is important; it gives the people belonging to this generation the feeling they are needed. For this reason, they remain economically active for a long time. This is evidenced by data – in December 2021, around a quarter of the Traditionalists and Baby Boomers were employed.

Figure 4 Structure of employed persons by generation and NACE/PKD (Polish classification of activities) section in 2021 (as at 31st December)



Source: Own construction

In this generation group, there is the biggest difference in the share of employed persons in the total population between men and women (13.2 percentage points), which is largely related to a lower by 5 years retirement age for women. In spatial terms, it can be seen that residents of areas surrounding Poland's largest cities, especially around Warsaw and capitals of provinces, have remained most active.

Generation X is the mature generation of professionally active people who were brought up in times of crisis, political and income changes (Kotler, 2005). For people of this generation, it is important to have a career and to be as committed to work as possible. They prefer quiet and secure work that does not require new challenges. The highest share of employed persons (Figure 4) belonging to this generation was recorded in sections: Agriculture, forestry and fishing; Education as well as Public administration and defence; compulsory social security. Employed persons from Generation X appreciate stability. Their humility and respect for their work make them conscientious employees. For them, professional success means hard work leading to a goal. Therefore, they are able to dedicate themselves to work, performing their duties diligently and subordinating their private lives to work (Boni, 2011; Rusak, 2013, 2014). The time in which they started their careers (fixed-term contracts, unemployment, etc.) strengthed the atmosphere of insecurity and fear related to the possibility of losing their jobs (Hysa, 2016). The share of employed persons for both men and women, oscillated around 70% in this generation. The generation was most active in large cities such as Warsaw, Łódź, Poznań and their surrounding areas. The distribution of male and female economic activity in the smallest administrative units in Poland, i.e. communes (*gminas*), was similar – the correlation coefficient was 0.829.

At the end of 2021, a group of similar size was made up byof employed persons from Generation Y. Together with Generation X, they accounted for almost three quarters of employed persons in Poland. Despite a longer time before retirement age, at the end of 2021, the share of employed persons in the populations among Generation Y was approximately 1 percentage point lower than for Generation X. The difference in the share between men and women was also larger, approximately 6 percentage points. The share of employed men from Generation Y was the highest of all generations and indicated that at the end of 2021 almost three quarters of men from this generation were employed. The economic activity of women belonging to Generation Y was almost 4 percentage points lower than in the case of women from Generation X. This may be due to higher procreative activity – in 2021, around four fifths of live births are children of mothers from Generation Y. Research shows that, Generation Y's attitudes to work and professional life have been significantly influenced by the fast pace of life, globalisation and widespread access to the Internet (Dolot, 2018; Smolbik-Jęczmień, 2013) – members of this generation place great emphasis on work-life balance (Ng, Schweitzer, Lyons et al., 2010 Twenge, 2010; Stosik and Lesniewska, 2015; Grobelna and Tokarz-Kocik, 2016), and consequently seek work that gives professional fulfilment and is valuable in its own right (Yang and Guy, 2006; Stachowska, 2012). A trait attributed to this generation by researchers is mobility both in reference to careers and in psychological terms (Lyons et al., 2012; Mazur-Wierzbicka, 2015). People from this generation move easily from city to city, from company to company, from country to country, which is due to easy travelling as a result of open borders, but also the possibility to go abroad for scholarships, on-the-job training, traineeships). The approach to the so-called employee loyalty is also different from that of older generations – career building and long-term commitment to the organisation are not important to them. However, if the job satisfies their needs and meets their expectations, they are able to be committed to their role.

Members of Generation Y not only want to work in the chosen profession, but at the same time in a company that is able to guarantee them the possibility of self-realisation and a good atmosphere – the more a company pays attention to this, the better loyalty it receives from them. When taking up a job, they are guided by the company's prestige and, above all, the sector. As can be seen from the data presented in Figure 4, the sections characterised by a higher than average share of Generation Y at the end of 2021 were in particular: Information and communication; Professional, scientific and technical activities and Financial and insurance activities. The economic activity of Generation Y across communes is strongly positively correlated with the economic activity of Generation X (r = 0.888). This means that Generation Y is characterised by the biggest share of employed persons in total populations of large cities and communes surrounding the largest cities, especially around Warsaw, Łódź, Poznań, Wrocław or Kraków.

An approach to work is very close to the mentality of Generation Y is found in Generation Z. Generation Z is the first generation to come into a world dominated by technology. People from this generation do not know a world without the Internet, telephone or computer. They demonstrate a freedom in using modern technology. Most young people from Generation Z cannot imagine a world without the Internet, but are able to function in parallel in both the real and virtual worlds, as well as to move smoothly from one to the other (Czyczerska, Ławnik, Szlenk-Czyczerska, 2020). It is also distinguished by its need for immediate satisfaction. What is a threat to older generations, for this generation is a field for experimentation and an object of fascination. In a sense, it is a generation of extremes (Żarczyńska-Dobiesz and Chomątowska, 2014). They do not pay attention to the stability of employment, they escape from routine, and they seek diversity. They want to try out new ways of working, go for traineeship abroad, change and improve established processes. They are characterised by their knowledge of foreign languages and mobility, hence they view the labour market from a global perspective. They are eager to deepen their knowledge in tertiary education – in 2020, more than half of the members of this generation studied. Consequently, they have the opportunity to look for work not only in Poland, but all over the world (Czyczerska, Ławnik, Szlenk-Czyczerska, 2020). But at the end of 2021, around 25% of the members of this generation were employed. As with other generations, economic activity was higher for men than for women. Among people aged 15–24 years, every fourth man and every fifth woman were employed. This generation's engagement in work was significantly different than in the other generations. The highest positive correlation was with the spatial distribution of the share of employed persons in Generation Y – r = 0.544. On the map of Poland, the highest share of employed persons in Generation Z was found in communes surrounding Poznań.

CONCLUSION

The contemporary labour market in Poland is undergoing a constant transformation, which is mainly related to the population ageing. As presented in the above analysis, there are currently four generations on the labour market in Poland, and even five with the oldest Traditionalists. Each generation is distinguished by different beliefs, characteristics, interests and expectations. The mental complexity of the labour force indicates that, with the growing need to activate potential labour force, openness to diversity is becoming essential. The success of today's businesses will increasingly depend on their ability to appreciate and use each generation. It will therefore be necessary to develop an organisational culture based on mutual respect, which will enable all employees to exchange experiences, learn from each other (without division into young and old). The most important thing is to create a platform for mutual cooperation between members of all generations. People from different generations should not compete with each other, but complement each other on the labour market. Such an approach offers an opportunity to increase the utilisation of the potential of all generations as employees. With so many areas of diversity in the labour force, human resource management on the labour market is and will be a major challenge, the solution to which can at least partially be supported by statistics showing the different generational groups on the labour market.

References

BERŁOWSKI, P. (2012). *Blaski i cienie współpracy X, Y i Z, czyli zarządzanie wielopokoleniowością na przykładzie Goodyear w Polsce* [online]. Bydgoszcz: Poland [cit. 10.10.2022]. <https://doi.org/10.26366/PTE.ZG.2021.197>.

BONI, M. et al. (2011). *Młodzi 2011*. Kancelaria Prezesa Rady Ministrów.

BORCZYK, A., CHMIEL, M., CZARNECKA-STAŃCZAK, W. et al. (2011). Sukces przez zaangażowanie. *Raport Magazynu ThinkTank*: 7–20.

CHŁOŃ-DOMIŃCZAK, A., HAUSNER, J. (2021). *Pokolenia – solidarni w rozwoju*. Fundacja Gospodarki i Administracji Publicznej.

CIERNIAK-PIOTROWSKA, M., DĄBROWSKA, A., STELMACH, K. (2022). *Population. Size, structure and vital statistics in Poland*. GUS.

CIURA, G., SZYMAŃCZAK, J. (2012). *Starzenie się społeczeństwa polskiego*. Biuro Analiz Sejmowych.

CZYCZERSKA, K. M., ŁAWNIK, A., SZLENK-CZYCZERSKA, E. (2020). Współczesny rynek pracy w Polsce a generacja. Różnice między pokoleniami X, Y i Z [online]. *Rozprawy Społeczne*, 14(3): 102–125. <https://doi.org/10.29316/rs/125693>.

DOLOT, A. (2018). Proces poszukiwania pracy przez młode pokolenie – wybrane zagadnienia – wyniki badań empirycznych. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, 359: 284–299.

GROBELNA, A., TOKARZ-KOCIK, A. (2016). Work-life Balance and Its Importance for the Work Process in Hospitality Industry. A Perspective of Generation Y Employees. In: PRIMORAC, Z., BUSSOLI, C., RECKER, N. (eds.) *Economic and Social Development*, Split: University of Split and University of North Split.

HYSA, B. (2016). Zarządzanie różnorodnością pokoleniową. *Zeszyty naukowe Politechniki Śląskiej. Seria: Organizacja i zarządzanie*, 97: 385–398.

JAEGER, H. (1977). *Generationen in der Geschichte. Überlegungen zu einer umstrittenen Konzeption*. Geschichte und Gesellschaft.

KOPERTYŃSKA, M. W., KMIOTEK, K. (2014). Budowanie zaangażowania pracowników pokolenia Y [online]. *Prace Naukowe UE we Wrocławiu*, 358: 25–38. <https://doi.org/10.15611/pn.2014.358.03>.

KOTLER, P. (2005). *Marketing*. REBIS.

KUDELA-ŚWIĄTEK, W., SARYUSZ-WOLSKA, M. (2022). *Pokolenie* [online]. Berlin: Centrum Badań Historycznych Polskiej Akademii Nauk w Berlinie. [cit. 18.10.2022]. <https://cbh.pan.pl/en/pokolenie>.

LAIN-KENNEDY, J. (2007). *Job Interviews for Dummies*. Hoboken NJ.

LYONS, S. T., SCHWEITZER, L., NG, E.S., KURON, L. K. (2012). Comparing Apples to Apples: a Qualitative Investigation of Career Mobility Patterns Across Four Generations. *Career Development International*, 17(4): 333–357. <https://doi.org/10.1108/13620431211255824>.

MAZUR-WIERZBICKA, E. (2015). Kompetencje pokolenia Y: Wybrane aspekty. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 39/3: 307–320.

NG, E. S., SCHWEITZER, L. W., LYONS, S. T. (2010). New Generation, Great Expectations: a Field Study of the Millennial Generation. *Journal of Business Psychology*, 25(2): 281–292. <https://doi.org/10.1007/s10869-010-9159-4>.

RUSAK, P. (2013). X, Y, Z: pokoleniowa bitwa biurowa. *Przewodnik Pracodawcy 2013/2014*.

SMOLBIK-JĘCZMIEŃ, A. (2013). Podejście do pracy i kariery zawodowej wśród przedstawicieli generacji X i Y – podobieństwa i różnice. *Nauki o zarządzaniu*, 1(14): 89–97.

STACHOWSKA, S. (2012). Oczekiwania przedstawicieli pokolenia Y wobec pracy i pracodawcy. *Zarządzanie Zasobami Ludzkimi*, 2(85): 33–56.

STOSIK, A., LEŚNIEWSKA, A. (2015). Innowacyjne rozwiązania w zarządzaniu zasobami ludzkimi – perspektywa adaptacji i różnic pokoleniowych. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 39: 183–196.

TWENGE, J. M. (2010). A Review of the Empirical Evidence on Generational Differences in Work Attitudes [online]. *Journal of Business and Psychology*, 25(2): 201–210. <http://dx.doi.org/10.1007/s10869-010-9165-6>.

YANG, S., GUY, M. E. (2006). GenXers versus Boomers: Work motivators and Management Implications [online]. *Public Performance and Management Review*, 29(3): 267–284. <http://dx.doi.org/10.2753/PMR1530-9576290302>.

ŻARCZYŃSKA-DOBIESZ, A., CHOMĄTOWSKA, B. (2014). Pokolenie „Z" na rynku pracy – wyzwania dla zarządzania zasobami ludzkimi. Sukces w zarządzaniu kadrami. Różnorodność w zarządzaniu kapitałem ludzkim – podejścia, metody, narzędzia [online]. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 350: 405–415. <http://dx.doi.org/10.15611/pn.2014.350.36>.

Role of Institutional Quality in Trade Openness and Economic Growth Nexus: Empirical Evidence from India

Suadat Hussain Wani[1] | *University of Kashmir, Srinagar, India*
Effat Yasmin[2] | *University of Kashmir, Srinagar, India*
Mohammed Ayub Soudager[3] | *University of Kashmir, Srinagar, India*

Received 17.1.2023 (revision received 28.2.2023), **Accepted** (reviewed) 9.3.2023, **Published** 16.6.2023

Abstract

In changing context of the present-day world, trade openness has a crucial role to play in economic development of different countries. Besides other factors, institutional quality plays a vibrant role in achieving a high growth rate. The objective of the present study is to understand how institutional quality influences economic growth and trade openness in India. To achieve the objectives of the study, Autoregressive Distributed Lag bound testing approach has been used. The findings show that there exists long-run relationship between the variables used in this study. From the findings, it can be concluded that total trade has a negative impact, whereas export enhances economic growth in the country. The results also show that improvement in institutional quality has a positive impact on economic growth. Thus the findings suggest that the country needs to adopt policies that can improve the quality of institutions and can enhance the formation of physical and human capital.

Keywords	***DOI***	***JEL code***
Institutional quality, economic growth, human capital, physical capital, trade openness, ARDL	*https://doi.org/10.54694/stat.2023.2*	*F00, C00, F43, O4*

INTRODUCTION

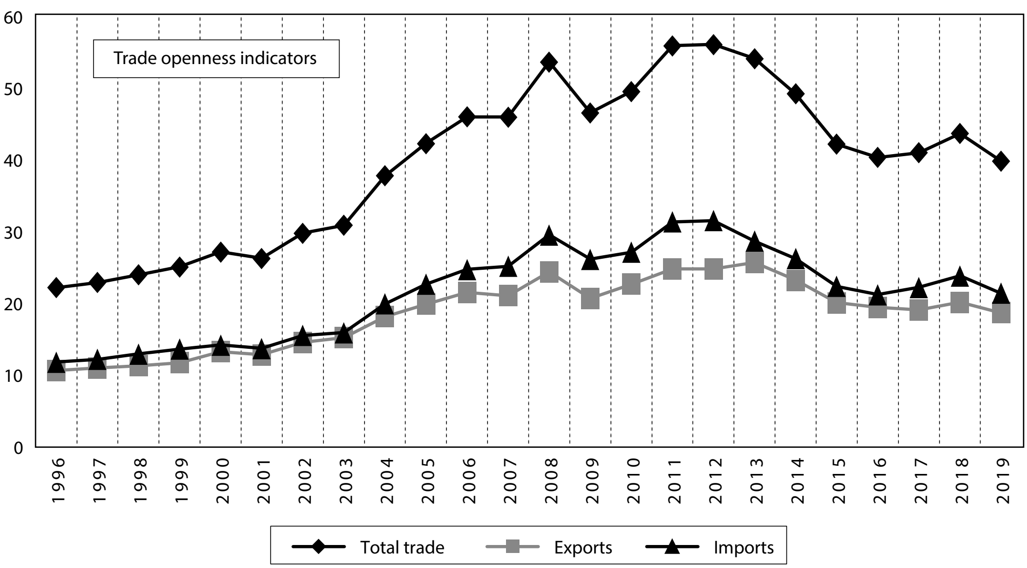
In recent times, increased attention has been paid to examine the imprints of trade openness on the level of economic growth, particularly in case of emerging countries. This increased attention is due to the fact that the integration of different countries in world economy has increased over the last few decades.

[1] Department of Economics, University of Kashmir, Dargah Hazratbal, Srinagar, 190006, Jammu and Kashmir, India. E-mail: suadat.scholar@kashmiruniversity.net.
[2] Department of Economics, University of Kashmir, Dargah Hazratbal, Srinagar, 190006, Jammu and Kashmir, India. E-mail: fgulwani@gmail.com.
[3] Department of Economics,University of Kashmir, Dargah Hazratbal, Srinagar, 190006, Jammu and Kashmir, India. E-mail: muhammadayub@kashmiruniversity.ac.in.

Further, given the changing economic scenario around the world, particularly since the 1990s, trade integration is considered a necessary condition for transaction from closed to open economy (see Zahonogo, 2016). A look at the theoretical aspect indicates that high degree of openness encourages the transmission of new technology and enables technological advancement and innovations, which augments productive knowledge and hence boost employment and real earnings (see Grossman and Helpman, 1991). This theoretical justification and partly disappointed performance of import substitution strategy have motivated many developing economies to switch to more liberalized trade regimes (see Udeagha and Ngepah, 2021; Akpan and Atan, 2016; Chatterji, Mohan and Dastidar, 2014). However, despite theoretical connection, the available empirical findings (see Rivera-Batiz and Romer 1991) point towards the group of the model, which contend that trade openness may either improve or retard international economic growth. If countries have divergent factor endowment, then (see Young, 1991; Kind, 2002) individual countries may face adverse impact even though economic integration increases global economic growth. Another strand of literature highlights negative link of economic growth with trade openness (see Akinlo and Okunlola, 2021; Sarkar, 2005; Rigobon and Rodrik, 2004; Batra, 1993; Leamer, 1995; Vamvakids, 2002; Kim, 2011).

This study is induced by the fact that India decided to open the domestic market to outside world in early 1990s to enhance the process of economic development. A look at the economic history of the country shows that there have been two major policy regimes that have contributed to its economic progress (see Aggarwal and Kumar, 2012). The state-led growth model was established Between 1950–80, state-led growth model was adopted in which the public sector was given a key role in the process of economic development. But starting in 1980, the nation began transition to a more liberal and open system. From the middle of the 1980s, this movement toward market-led growth intensified, and from 1991 onwards, more extensive and systemic liberalization measures were implemented (Chatterji, Mohan and Dastidar, 2014). The graph of the various trade openness indicators used in the current study is presented in Figure 1.

Figure 1 Evolution of trade in India



Note: X-axis represents time, and Y-axis represents the value of exports, imports, and total trade as percentage of GDP.
Source: Author's calculation based on data from WDI using EViews 10

Several empirical studies have highlighted poor institutional quality as one of the reasons for the negative impact of trade openness on economic growth. These findings indicate that trade openness accompanied by high institutional quality has a strong growth effect. In case of advanced countries, institutional quality tends to accelerate trade benefits (see Nguyen and Nguyen, 2018). Similarly, Hall and Jones (1999) contended that the difference in output per labor across nations is widely attributed to institutional quality. Few studies have emphasized the importance of institutions in fostering economic growth (see Subramanian and Trebbi, 2004; Matthew and Adegboye, 2014). To effectively coordinate efforts to achieve economic growth, good governance, including the rule of law, property rights, control of corruption, accountability, and regulatory quality is crucial. Given this background, the present study attempts to demonstrate what role institutional quality plays in the link between trade openness and economic growth in case of India. This is crucial as one finds a gap in the reviewed literature on economic growth in India. In the available literature, there has not yet been a thorough examination of the role that institutional quality plays in trade openness and economic growth.

The present study enriches the current literature in two significant ways. First, all three measures of trade openness, which include exports, imports, and total trade, are used. As highlighted in Figure 1, the magnitude of exports and imports is different, which may offer different options for policy formulation. Second, in case of India, the role of institutions in the link between openness and economic growth is missing in the available literature. The present study attempts to make significant contribution in this direction. To achieve objectives of the study, an extended Cobb-Douglas production function has been used. It applies Autoregressive Distributed Lag (ARDL) procedure which is relatively a new cointegration technique (see Pesaran et al., 2001). The structure of the present study is as follows: review of the literature is discussed in Section 1, followed by methodology in Section 2. Empirical results are elaborated in Section 3 and main findings and policy implications are discussed in final Conclusion.

1 REVIEW OF LITERATURE

1.1 Trade openness and economic growth

In recent times, the link between trade openness and economic growth has received enough attention. However, no common consensus has been reached on this link between the two.

Regarding studies on India, Barua and Chakraborty (2006) are of the view that trade openness has enhanced consumer surplus but has reduced industrial concentration and the surplus of producers. Similarly, Topalova and Khandelwal (2004, 2011) conclude that the productivity of firms improves with an increase in trade openness. In case of China and India, Marelli and Signorelli (2011) argue that there is positive link between trade openness and economic growth. Similarly, Mercan et al. (2013) argue that in case of Brazil, China, India, Russia and Turkey, trade openness augments economic growth and these countries need to formulate policies to enhance exports. Besides, Chatterji, Mohan, and Distidar (2014) are of the view that in increase in trade volume have enhanced economic growth in India over the years. However, (see Hye and Lau, 2015; Sengupta, 2020) have concluded that in the long run, there is a negative link between trade openness and economic growth in case of India.

Regarding other countries, Kind (2002) argue that there is ambiguous effect of trade openness on economic growth given the different size of domestic markets in different countries. Yanikkay (2003) argue that through several channels like transfer of technology, comparative advantage, and economies of scale, trade augments growth and development of different countries. However, the study concludes that under certain conditions, trade restrictions can promote economic growth particularly in developing countries. Besides, Ved and Sudesh (2007) have concluded that higher level of trade openness increases economic growth.

In comparison, Hye (2012) has concluded that in case of Pakistan, human and physical capital enhances economic growth. In the case of Australia, Singh (2011) is of the view that exports enhance economic growth, whereas imports have a negative link with economic growth of the country. Similarly, Adhikary (2011) have concluded that trade openness has a negative link with economic growth. Xie et al. (2018) examined causal link between trade openness, financial development and economic growth in case of china. The study finds that financial development and trade openness promote economic growth. However, increase in economic growth leads to decline in trade openness.

1.2 Economic growth and institutional quality

In the available literature, trade openness has been widely acknowledged as the determinant of economic growth; however, another parallel strand of literature has documented non-economic factors, such as institutional quality, as an important determinant of economic growth. As Rodrik (1999) stated, economies with defective or weak institutions are less able to react appropriately to external shocks, such as openness shocks, which can lead to a long-term slowdown of economic growth. Institutional quality and its interaction with trade openness are considered significant in determining economic growth (see Stensnes, 2006; Akpan and Atan, 2016; Doan, 2019). According to empirical findings, institutions are critical for the success of economic reforms in developing countries. The failure of trade reforms to enhance international trade and growth in these countries is attributed to low institutional quality (see Addison and Baliamoune-Lutz, 2006; Kraay, 2003). In another study, Rodrik, Subramanian and Trebbi (2002) concluded that Institutions are better predictors of economic growth than trade openness. In the absence of quality institutions, trade openness either retards or has a modest effect on economic growth.

1.3 Capital formation and economic growth

The availability of capital, natural resources, and common elements of production are not distributed equally among all economies in the world. The majority of the growth hypothesis has been based on this unevenness. Neoclassical synthesis states that high levels of capital formation result in higher productivity, which fosters economic growth (see Aslan and Altinoz, 2021). Hye and Lau (2015) argue that physical and human capital is mainly related to economic growth. Numerous studies have scrutinized the connection between capital formation and economic growth in various nations. According to these empirical findings, as compared to physical capital, development of human capital, which includes literacy, skill acquisition, access to healthcare, and experience, enriches economic growth (see Todaro, 2002; Ali et al., 2012; Mahmood et al., 2014; Munir and Arshid, 2018). Furthermore, the developing countries rated low in the inequality-adjusted Human development index were found to benefit more from capital formation than developed countries (Ahumada et al., 2020). Therefore, governments in these countries should spend more on providing social provisions. Baker (2011) found a positive association between capital accumulation and growth nexus in the case of Nigeria. His findings suggested more government involvement in encouraging savings to foster the environment of investment and promote economic growth.

A plethora of literature discussed above supports the positive relation between trade openness and economic growth (see Yanikkay, 2003; Ved and Sudesh, 2007; Klasra, 2011). However (see Hye, 2012; Adhikary, 2011; Hye and Lau, 2015; Sengupta, 2020), there is negative relation between trade openness and economic growth. Thus it is not surprising that the debate on the connection of trade openness with economic growth continues to crawl and needs further empirical studies to contribute to the validation of available literature. Further from the studies (see Rodrik, 1999; Rodrik, Subramanian and Trebbi, 2002; Kraay, 2003; Addison and Baliamoune-Lutz, 2006; Stensnes, 2006; Akpan and Atan, 2016; Doan, 2019), it can be concluded that institutional quality, as well as its interaction with trade openness, plays a dynamic role in the economic growth of a country. Available literature in the case of India (Barua and Chakraborty, 2006; Topalova and Khandelwal, 2004 and 2011; Marelli and Signorelli, 2011; Mercan et al., 2013; Chatterji, Mohan and Distidar, 2014) has ignored or overlooked the role of institutional quality in economic growth. As a contribution to available literature, the present study aims to include role of institutional quality in the link between trade openness and economic growth in Indian context.

2 METHODOLOGY

2.1 Collection of data and transformation

In the present study, annual data from 1996–2019 has been applied in case of India. Given the availability of data for institutional quality, 1996 has been taken as starting year. Following the available literature discussed above, secondary school enrollment (% gross) is used as a proxy for Human Capital (HC), real gross fixed capital formation for Physical Capital (PC), and real GDP for economic growth. The institutional quality index is developed using the WGI dataset, which provides data for six dimensions[4] to capture institutional quality. Each dimension falls within the range of –2.5 and +2.5. Following Raychaudhuri and Haldar (2009), an institutional quality index is constructed to achieve the objectives. The required data for selected variables was extracted from WDI (World Development Indicators) and WGI (World Governance Indicators, World Bank, 2021).

2.1.1 Trade openness index

The available literature indicates that various proxies for trade openness like exports, imports and total trade as a percentage of GDP have been used to examine the impact of trade openness on economic growth of different countries. Each of these measures captures a different aspect of trade openness. Grossman and Helpman (1989) argue that trade openness leads to the production process in a country according to its comparative advantage. Similarly, exports as a percentage of GDP are used as a proxy for openness to capture the length of trade openness related to scale economies. Besides, to measure the level of international competition in the domestic market, imports as a percentage of GDP are used as a proxy for trade openness. Further, the share of total trade as a per cent of GDP provides a representation of technological spillover due to trade liberalisation measures by a particular country (Hye and Lau, 2015). In the present study, given the importance of each indicator, all three proxies of trade openness have been included.

2.2 Theoretical framework and model specification

Following Mankiw et al. (1992), Omoke and Opuala–Charles (2021), and Shahbaz (2012), the Cobb-Douglas production function in period t is given as:

$$Y = A(t)L(t)^{\propto} C(t)^{1-\propto} \text{ where } 0 < \propto < 1, \tag{1}$$

where Y represents real Gross Domestic Product (GDP), A stands for technological progress, L for labor, and C for capital stock. In the present study, this production function is extended by assuming that the technological progress is determined by trade openness, institutional quality and capital formation which include both human and physical capital formation. This is given as:

$$A(t) = \varphi.TO(t)^{\beta}\ IQI(t)^{\delta}\ HC(t)^{\sigma}\ PC(t)^{1-\sigma}, \tag{2}$$

where φ is the time-invariant constant, TO represents trade openness, IQI stands for institutional quality index, HC for human capital and PC for physical capital. The extended log-linear form of Formula (2) is given as:

---

[4] Control of corruption (CC), Government effectiveness (GE), Political stability and absence of violence/terrorism (PA), Rule of law (RL), Regulatory quality (RQ), and Voice and accountability (VA).

$$LnY_t = \beta_0 + \beta_1 LnTO_t + \beta_2 IQI_t + \beta_3 LnHC_t + \beta_4 LnPC_t + \varepsilon_t, \quad (3)$$

where L_nY_t represents real gross domestic product, L_nTO_t trade openness, IQI_t institutional quality index, L_nHC_t denotes human capital, L_nPC_t represents physical capital, and ε_t represents ordinary disturbance term. In the available literature, various proxies for trade openness as discussed above have been put to use to dig into the link of trade openness with economic growth in different countries. It is important to mention here that a low value of these variables shows a high degree of policy intervention, given the fact that each variable captures different aspect of openness. Thus given the importance of each of these proxies of trade openness, the present study uses all three indicators. Following Duodu and Baidoo (2020), Formula (3) is extended to incorporate the interaction of trade openness and institutional quality as given in Formula (4):

$$LnY_t = \beta_0 + \beta_1 LnTO_t + \beta_2 IQI_t + \beta_3 LnHC_t + \beta_4 LnPC_t + \beta_5 \left(LnTO_t * IQI_t\right) + \varepsilon_t. \quad (4)$$

Theoretically, all the variables, which include trade openness, human and physical capital, and institutional quality, are expected to enhance the country's economic growth. ($LnTO_t * IQI_t$) is an interaction term that captures the combined effect of institutional quality and trade openness.

2.3 Estimation framework

In recent years, different cointegration models have been developed for non-stationary variables. Among various techniques, the "Autoregressive Distributed Lagged Model (ARDL)" advanced by Pesaran et al. (2001) works well. This technique aims to inquire into the stable long-run stationary relationship between non-stationary variables.[5] The technique of cointegration is used to model the log-linear specifications in Formulas (4) and (5) as follows:

$$\begin{aligned} LnY_t = {} & \propto_0 + \sum_{i=0}^{n} \propto_{1i} LnY_{t-i} + \sum_{i=0}^{n} \propto_{2i} LnTO_{t-i} + \sum_{i=0}^{n} \propto_{3i} IQI_{t-i} + \sum_{i=0}^{n} \propto_{4i} LnHC_{t-i} + \sum_{i=0}^{n} \propto_{5i} PC_{t-i} \\ & + \beta_1 LnY_{t-1} + \beta_2 LnTO_{t-1} + \beta_3 IQI_{t-1} + \beta_4 LnHC_{t-1} + \beta_5 PC_{t-1} + \epsilon_t, \end{aligned} \quad (5)$$

$$\begin{aligned} LnY_t = {} & \propto_0 + \sum_{i=0}^{n} \propto_{1i} LnY_{t-i} + \sum_{i=0}^{n} \propto_{2i} LnTO_{t-i} + \sum_{i=0}^{n} \propto_{3i} LnIQI_{t-i} + \sum_{i=0}^{n} \propto_{4i} LnHC_{t-i} + \sum_{i=0}^{n} \propto_{5i} PC_{t-i} \\ & + \sum_{i=0}^{n} \propto_{6i} LnTO * IQI_{t-i} + \beta_1 LnY_{t-1} + \beta_2 LnTO_{t-1} + \beta_3 LnIQI_{t-1} + \beta_4 LnHC_{t-1} \\ & + \beta_5 PC_{t-1} + \beta_6 (LnTO * IQI_{t-1}) + \epsilon_t. \end{aligned} \quad (6)$$

The parameters ($\propto_0 - \propto_6$) capture short-run relationships and ($\beta_1 - \beta_6$) measure long run relationship among the variables. The ARDL framework has been used to estimate Formulas (5) and (6). After cointegration, Error Correction Model (ECM) has been estimated. This model provides information regarding the rate of adjustment and aids in the study of short-run dynamics. Thus following ECM models have been specified:

[5] This method has many advantages: first, it can apply irrespective of whether the regressors are integrated of order one or order zero or mutually (Pesaran et al., 2001); second, the ARDL model is free from serial correlation and endogeneity problems; finally, a dynamic error correction model (ECM) can be derived from ARDL through a simple linear transformation.

$$\mathrm{LnY_t} = \propto_0 + \sum_{i=0}^{n} \propto_{1i} \mathrm{LnY}_{t-i} + \sum_{i=0}^{n} \propto_{2i} \mathrm{LnTO}_{t-i} + \sum_{i=0}^{n} \propto_{3i} \mathrm{IQI}_{t-i} + \sum_{i=0}^{n} \propto_{4i} \mathrm{LnHC}_{t-i} + \sum_{i=0}^{n} \propto_{5i} \mathrm{PC}_{t-i} + \lambda_1 \mathrm{ECM}_{t-1} + \epsilon_t, \quad (7)$$

$$\mathrm{LnY_t} = \propto_0 + \sum_{i=0}^{n} \propto_{1i} \mathrm{LnY}_{t-i} + \sum_{i=0}^{n} \propto_{2i} \mathrm{LnTO}_{t-i} + \sum_{i=0}^{n} \propto_{3i} \mathrm{LnIQI}_{t-i} + \sum_{i=0}^{n} \propto_{4i} \mathrm{LnHC}_{t-i} + \sum_{i=0}^{n} \propto_{5i} \mathrm{PC}_{t-i} + \sum_{i=0}^{n} \propto_{6i} \mathrm{LnTO} * \mathrm{IQI}_{t-i} \lambda_1 \mathrm{ECM}_{t-1} + \epsilon_t. \quad (8)$$

In Formulas (7) and (8), ECM indicates speed of adjustment.

3 EMPIRICAL RESULTS

3.1 Unit root test

The present study make use of ADF (Augmented Dickey-Fuller) and PP (Phillips-Perron) to check the stationery nature of the selected variables. Both tests assume the unit root problem under the null hypothesis. Table 1 shows the result of both these tests. It is important to note here that at first difference all the variables are stationary. Following (Tahir and Hayat, 2020; Omoke and Opuala–Charles, 2021), the ARDL approach seems to be most suitable for the present study given the nature of selected variables.

Table 1 Share of positive answers to job search questions and item-response probabilities

Variables	ADF test		PP test	
	Level	First difference	Level	First difference
LnY	0.58	–3.69**	0.79	–4.11***
LnHC	–0.34	–4.24***	–0.38	–4.26***
LnGCF	–1.54	–4.33***	–1.62	–4.37***
LnTO	–1.92	–3.74**	–1.87	–3.81***
LnExp	–1.93	–3.77**	–1.93	–3.87***
LnImp	–1.84	–3.41**	–1.79	–3.38**
IQI	–1.82	–3.12**	–1.98	–4.47***

Note: ***, **, * represent 1%, 5% and 10% level of significance.
Source: Author's calculation

3.2 Results of Bound test

Table 2 presents the calculated f-statistics from the ARDL cointegration test and compares it with critical values from Narayan (2005). The results show that the alternative hypothesis of cointegration is accepted in all specifications and null hypothesis of no cointegration is rejected. This indicates long run causal relationship between economic growth, trade openness, human and physical capital formation and institutional quality in India. The table shows lower and upper bound values only in case of first specification.[6]

[6] Additional material is available from the author: suadat.scholar@kashmiruniversity.net.

Table 2 Bounding test

Dependent variables	F-test	Optimal lag length	Decision
D(LnY) D(LnPC) D(LnHC) D(LnTO)	10.05	(3, 1, 3, 3)	Co-integrated
D(LnY) D(LnPC) D(LnHC) D(LnTO) D(IQI) D(LnTO_IQI)	13.48	(2, 0, 2, 2, 2)	Co-integrated
D(LnY) D(LnHC) D(LnExp) D(LnPC) D(LnExp_IQI) D(IQI)	5.65	(2, 1, 2, 2, 2, 2)	Co-integrated
D(LnY) D(LnPC) D(LnHC) D(LnExp) D(LnImp_IQI) D(IQI)	11.13	(2, 1, 2, 0, 1)	Co-integrated
Critical values	Lower bound 1(0)		Upper bound 1(1)
1%	3.65		4.66
5%	2.79		3.67
10%	2.37		3.2

Source: Author's calculation

3.3 Long and short-run results

Specifications 1 and 2 in Table 3 present the ARDL estimates from Formula (6). In specifications 1 and 2, total trade as percentage of GDP is taken as a measure of trade openness. The estimated long-run coefficients are presented in panel A. In contradiction to theoretical and empirical findings (see Romer, 1990; Ynikkaya, 2003; Wacziarg et al., 2008; Dash, 2009; Marelli et al., 2011), trade openness shows negative relationship with economic growth. The results indicate that 1 percent increase in trade openness leads to 0.07 percent decline in economic growth. Earlier findings (see Kind, 2002; Kim, 2011; Lawal et al., 2016) are in line with the findings of present study. In case of less developed countries, Kim (2011) found negative relationship between trade openness and economic growth. Moreover, in the case of Pakistan and India, Hye (2012) and Hye and Lau (2015) found inverse relation between trade openness and economic growth. However, in terms of magnitude, the results of the present study reveal that impact of trade openness on economic growth is lowest compared to other variables. Hye and Lau (2015) show that a 1 percent increase in trade openness leads to a 0.301 percent decline in economic growth in the long run in the case of India. However, the results in the present study indicate that a 1 percent increase in trade openness leads to a 0.07 percent decline in economic growth in the long run. Comparing the results in these studies indicate that in the long run, trade openness would reap its benefits and help in the long-run development of the country. Available literature, (see Batra, 1992; Batra and Slottje, 1993; Vamvakidis, 2002) point out that if trade liberalization is not managed in a proper way in a particular country, it can adversely impact economic performance of that country.

In line with the theoretical justification and findings of Wani (2022), the results reveal a direct link between HC and PC with economic growth. The results show that 1 percent increase in physical capital enhances economic growth by 0.11 percent in the country. Thus domestic investment shows positive and cogent effect on economic performance. The results of the present study align with findings of Barro (2003), and Tahir and Hayat (2020). Besides, human capital has a positive impact on economic growth of the country in long run. However, in contrast to the theoretical background, short-run results indicate the negative impact of human capital on economic growth. Nevertheless, it is hard to find an explanation for this relationship; one possible reason may be the insufficient quality of education to produce the required skill to enhance economic performance particularly in short run (Altinok, 2007). Further, Tahir and Azid (2015) argue that human capital may enhance economic growth in the future due to the nonlinear impact of human capital on economic growth. Available literature (see Hanushek and Woessmann, 2010; Pelinescu, 2015) also found a negative relationship between human capital and economic growth. Besides, the findings indicate that 1 percent improvement in institutional quality enhances economic growth by 0.52 percent (specification 2). Thus it can be concluded that policymakers need to pay special attention to improve the quality of institutions. The results accord with the findings of Stensnes (2006), Akpan and Atan (2016), Doan (2019), and Duodu and Baidoo (2020).

Panel B in Table 3 present the ARDL estimates from Formula (7). The results specify that in contrast to theoretical justification, the findings disclose that human capital has an adverse effect on economic growth. One possible reason may be insufficient quality of education necessary to deliver appropriate skills that can improve economic performance in the short run. The specification extended Formula (6) by incorporating an interaction term to capture role of institutions in trade openness – economic growth nexus in case of India. The coefficient of the interactive term is negative in all specifications. The results match the findings of Stensnes (2006), who concludes that as long as increased trade openness exposes an economy to stronger external shocks, it will adversely impact the growth of those economies with weak institutional capacity. According to Rodrik (1999), external shocks will negatively impact the economic growth of those countries that lack the proper institutional capacity to respond to them. The interaction term is positive and significant when lag is taken into consideration, given that trade policy and improvement in institutional quality impact the economic conditions after a span of time.

Table 3 Long run and short run results

Variables	Coefficients	T-statistic	Coefficients	T-statistic
Panel A: long-run coefficients	Specification 1		Specification 2	
LnPC	0.11**	2.65	0.08*	3.49
LnHC	0.15	1.52	0.18**	2.67
LnTO	–0.07*	–2.01		
IQI			0.52**	2.85
LnTO_IQI			–0.12**	–2.45
C	0.06***	24.4	0.06***	30.16
Panel B: Short-run coefficients				
LnPC	0.14***	8.98		
LnHC	–0.21***	–4.58	–0.08*	–1.82
LnHC(-1)	–0.58***	7.30	–0.25***	6.25
LnTO	–0.07***	–4.50		
LnTO(-1)	0.05**	2.58		
IQI			0.45***	5.46
(LnTO_IQI)			–0.10***	4.55
LnTO_IQI(-1)			0.06**	2.49
ECM(-1)	–1.65***	–9.15	–1.39***	–11.46
Adj. R: 0.93			0.94	

Note: ***, **, * stands for 1%, 5%, and 10% level of significance.
Source: Author's calculation

Table 3				Continuation
Variables	**Coefficients**	**T-statistic**	**Coefficients**	**T-statistic**
Panel C: long-run coefficients	Specification 3		Specification 4	
LnPC	0.12**	1.73	0.05	1.40
LnHC	0.08	0.66	0.22*	2.00
LnExp	0.01	0.50		
IQI	3.91**	3.02	0.28**	3.02
D(LnExp_IQI)	–0.14**	2.93		
D(LnImp_IQI)			0.005	0.85
C	0.06***	11.34	0.05***	5.23
Panel D: short-run coefficients				
LnPC	0.18***	17.83	0.11***	7.89
LnHC	–0.17***	–5.72	–0.09**	–2.15
LnHC(-1)	–0.11**	–3.52	–0.27***	6.65
LnExp	0.06***	3.20		
LnExp(-1)				
IQI	2.66***	6.20	0.45***	5.46
LnExp_IQI	–0.09**	–6.01	–0.10***	4.55
LnImp_IQI	–0.14***	9.01	–0.08***	–7.95
ECM(-1)	–1.19***	-9.95	-1.16***	-10.01
Adj. R: 0.97			0.93	

Note: ***, **, * stands for 1%, 5%, and 10% level of significance.
Source: Author's calculation

These results highlight the importance of building quality institutions as part of strategy to benefit from trade openness. Finally, the coefficient of ECM (–1) is negative and significant in all specification. This shows rate of change from short-run disequilibrium to established long-run equilibrium over a year. The results are in line with the findings of Wani (2022).

In Panel (C) and (D) of Table 3, trade openness is decomposed into two components.[7] Panel (C) shows long run estimation and panel (D) short run estimation. Interestingly the results show that exports positively contribute to the economic growth of India. The results show that 1 percent increase in exports enhances economic growth by 0.6 percent in long run. These findings lead to the conclusion that economic growth in India is export-led. In addition, the results reveal that total trade negatively impacts economic growth in the country. One possible reason may be the magnitude and composition of total trade between India and the outside world. A look at the total trade volume indicates that India imports more than exports and faces a trade deficit, as seen in Figure 1. From trade composition, it is clear that the import of capital goods has increased slightly from 19.23 percent in 1996 to 22.75 percent in 2019. On the other hand, the import of consumer goods, intermediate goods, and raw materials increased from 75.02 percent in 1996 to 76.8 percent in 2019. It is important to remark here that the availability of capital goods

[7] Export of goods and services as a percentage of GDP and import of goods and services as a percentage of GDP.

enhances the economic growth process in a country over the years. Thus from the data available, it is clear that given the low share of capital goods in the import basket, imports, as well as total trade, have an unfavorable impact on economic growth of the country. On the other hand, exports, irrespective of the composition, enhance economic growth in a particular country, given that they enhance aggregate demand. In the export basket, the chunk of raw materials and intermediate goods used in the production process has declined from 54.42 percent in 1996 to 38.25 percent in 2019, which is a healthy sign for the economy. At the same time, the share of consumer goods and capital goods has increased from 43.81 percent in 1996 to 61.67 percent in 2019, which shows the enhanced capacity of the economy to export to the outside world (data source: WITS World Bank, 2022). The coefficient of ECT (–1) is negative in all specifications, which shows that short-run disequilibrium may converge to the established long-run equilibrium.

3.4 Diagnostics testing

The results of some selected diagnostic tests which include serial correlation and heteroscedasticity are presented in Table 4. The results confirm that there is no such problem of heteroscedasticity and serial correlation. Moreover, the functional form of the model is confirmed by the Ramsey test. Finally, the stability of the model is confirmed by plots of CUSUM and CUSUMQ as shown in Figures A1 to A4 in the Appendix.

Table 4 Diagnostic checking

Diagnostics	F-statistic	F-statistic	F-statistic	F-statistic
LM test	0.004 (0.88)	0.02 (0.77)	0.35 (0.13)	0.13 (0.57)
ARCH	0.43 (0.49)	0.05 (0.81)	0.00 (0.97)	1.00 (0.30)
Breusch-Pagan-Godfrey	1.29 (0.32)	0.32 (0.86)	0.63 (0.52)	1.98 (0.17)
Normality	0.72 (0.69)	0.26 (0.87)	0.67 (0.71)	1.40 (0.49)
Ramsey test	4.18 (0.09)	0.26 (0.62)	0.00 (0.96)	2.73 (0.13)

Note: P-value in parenthesis.
Source: Author's calculation

3.5 Future research guidelines

In future studies, the time dimension can be increased. Similarly, the researchers can use all six dimensions individually from WGI to analyze their effect on the process of economic growth in the case of India. Further, given the present findings, export-led growth can be tested in the future.

CONCLUSION

The present study explores the association between trade openness and economic growth in India from 1996–2019, incorporating other factors, which include capital formation and institutional quality. The study employs three indicators of trade openness which include total trade, exports and imports as percentage of GDP. ARDL provides evidence of long run relationship among the selected variables. The estimation suggests that economic growth in the country has a negative link with trade openness. However, decomposing total trade into exports and imports yield interesting results. The estimation confirms that the growth of exports has a significant and positive impact on the economic growth of the country. In addition, the estimation reveals that institutions play an important role in economic development of the country both in short run and long run. Interestingly further empirical analysis through the introduction of interaction term of institutional quality and trade openness shows negative impact on economic growth of India. This indicates that trade openness accompanied by weak institutional quality may hamper the process of benefiting from opening the economy to outside world.

These finding highlight the need for number of policy consideration in case of India. Among others, the findings in the present study highlight the need to take necessary policy initiatives and introduce trade reforms to expand exports. Given that India has opened up its economy particularly from 1991 to outside world, the findings suggest benefits from trade openness will depend on the quality of institutions. Thus there is need to improve quality of institutions which will help to channel the benefits of trade openness into economic growth of the country. Similarly, in case of capital formation, the government of India needs to pay special attention to increase expenditure to enhance the skill level of available labor force. In addition, the economy of the country is benefiting from domestic investment, which is seen to have a positive impact on economic growth. Thus the need of the hour is to introduce various policy initiatives for both foreign and domestic investors.

ACKNOWLEDGEMENTS

The author Dr. Suadat Hussain Wani is the awardee of the ICSSR Post-Doctoral Fellowship. This paper is largely the outcome of the Post-Doctoral Fellowship sponsored by the Indian Council of Social Science Research (ICSSR). However, the responsibility for the facts stated, opinions expressed, and the conclusions drawn are entirely of the author.

References

ADDISON, T., BALIAMOUNE-LUTZ, M. (2006). Economic reform when institutional quality is weak: the case of the Maghreb [online]. *Journal of Policy Modeling*, 28(9): 1029–1043. <https://doi.org/10.1016/j.jpolmod.2006.04.004>.

ADHIKARY, B. K. (2011). FDI, trade openness, capital formation, and economic growth in Bangladesh: a linkage analysis. *International Journal of Business and Management,* 6(1): 16–28.

AGGARWAL, A., KUMAR, N. (2012). *Structural change, industrialisation, and poverty reduction: the case of India*. South and South-West Asia Development Paper, 1206, ESCAP.

AKCAY, S., DEMIRHAN, E. (2005). The causal relationship between openness and economic growth: Evidence from selected MENA countries. *Yapi Kredi Economic Review*, 16(2): 77–83.

AKINLO, T., OKUNLOLA, C. O. (2021). Trade openness, institutions and economic growth in Sub-Saharan Africa [online]. *Jurnal Perspektif Pembiayaan Dan Pembangunan Daerah*, 8(6): 541–560. <https://doi.org/10.22437/ppd.v8i6.10653>.

AKPAN, U. F., ATAN, J. A. (2016). Relationship between trade openness, institutions and economic growth in Sub-Saharan Africa: a further look at the evidence. *British Journal of Economics, Management & Trade*, 15(1): 1–20.

ALTINOK, N. (2007). *Human capital quality and economic growth* [online]. HALSHS. <https://halshs.archives-ouvertes.fr/halshs-00132531v2>.

AMNA INTISAR, R., YASEEN, M. R., KOUSAR, R., USMAN, M., MAKHDUM, M. S. A. (2020). Impact of trade openness and human capital on economic growth: a comparative investigation of Asian countries. *Sustainability*, 12(7): 2930.

BAL, D. P., DASH, D. P., SUBHASISH, B. (2016). The effects of capital formation on economic growth in India: Evidence from ARDL-bound testing approach [online]. *Global Business Review*, 17(6): 1388–1400. <https://doi.org/10.1177/0972150916660403>.

BALASSA, B. (1983). Policy responses to external shocks in sub-Saharan African countries [online]. *Journal of policy Modeling*, 5(1): 75–105. <https://doi.org/10.1016/0161-8938(83)90023-6>.

BARRO, R. J. (2003). Determinants of economic growth in a panel of countries. *Annals of economics and finance*, 4: 231–274.

BARUA, A., CHAKRABORTY, D. (2006). Liberalisation, trade and industrial performance: an empirical analysis for India. *The 2nd APEA Conference*, Center for International Economics, University of Washington, Seattle.

BATRA, R. (1992). The fallacy of free trade. *Review of International Economics*, 1(1): 19–31.

BATRA, R., SLOTTJE, D. J. (1993). Trade policy and poverty in the United States: Theory and evidence, 1947–1990. *Review of International Economics*, 1(3): 189–208.

CHATTERJI, M., MOHAN, S., DASTIDAR, S. G. (2014). Relationship Between Trade Openness and Economic Growth of India: a Time Series Analysis. *Journal of Academic Research in Economics*, 6(1).

DASH, R. K. (2009). Revisited export-led growth hypothesis: an empirical study on India [online]. *South Asia Economic Journal*, 10(2): 305–324. <https://doi.org/10.1177/139156140901000203>.

DOAN, H. Q. (2019). Trade, institutional quality and income: Empirical evidence for sub-Saharan Africa [online]. *Economies*, 7(2): 48. <https://doi.org/10.3390/economies7020048>.

DOLLAR, D. (1992). Outward-oriented developing economies really do grow more rapidly: evidence from 95 LDCs, 1976–1985. *Economic development and cultural change*, 40(3): 523–544.

DOLLAR, D., KRAAY, A. (2003). Institutions, trade, and growth. *Journal of monetary economics*, 50(1): 133–162.

DUODU, E., BAIDOO, S. T. (2020). How does quality of institutions affect the impact of trade openness on economic growth of Ghana? [online]. *Cogent Economics & Finance*, 8(1). <https://doi.org/10.1080/23322039.2020.1812258>.

FRANKEL, J. A., ROMER, D. (1999). Does trade cause growth? *The American Economic Review*, 89(3): 379–399.

GROSSMAN, G. M., HELPMAN, E. (1989). Growth and welfare in a small open economy. *NBER working paper*, w2970.

HADHEK, Z., MRAD, F. (2015). Trade openness, institutions and economic growth. *European Journal of Economics, Finance and Administrative Sciences*, 75: 96–104.

HAIDER, S., GANAIE, A. A., KAMAIAH, B. (2018). Total factor productivity and openness in Indian economy: 1970–2011 [online]. *Foreign Trade Review,* 54(1): 46–57. <https://doi.org/10.1177/0015732518810835>.

HANUSHEK, E. A., WOESSMANN, L. (2010). Education and economic growth. *Economics of education*, 60: 67.

HARRISON, A. (1996). Openness and growth: a time-series, cross-country analysis for developing countries. *Journal of development Economics*, 48(2): 419–447.

HYE, Q. M. A. (2012). Long term effect of trade openness on economic growth in case of Pakistan [online]. *Quality & Quantity*, 46(4): 1137–1149. <https://doi.org/10.1007/s11135-011-9612-0>.

HYE, Q. M. A., LAU, W. Y. (2015). Trade openness and economic growth: empirical evidence from India. *Journal of Business Economics and Management*, 16(1): 188–205.

INTISAR, R., YASEEN, M. R., KOUSAR, R. et al. (2020). Impact of trade openness and human capital on economic growth: a comparative investigation of Asian countries. *Sustainability*, 12(7): 1–19.

IYOHA, M., OKIM, A. (2017). The impact of trade on economic growth in ECOWAS countries: Evidence from panel data. *CBN Journal of Applied Statistics*, 8(1): 23–49.

KARTAL, Z., ZHUMASHEVA, A., ACAROGLU, H. (2017). The effect of human capital on economic growth: a time setries analysis for Turkey. In: *Regional Studies on Economic Growth, Financial Economics and Management*, Springer, Cham, 175–191.

KEHO, Y. (2017). The impact of trade openness on economic growth: the case of Cote d'Ivoire [online]. *Cogent Economics & Finance*, 5(1). <https://doi.org/10.1080/23322039.2017.1332820>.

KIM, D. H. (2011). Trade, growth and income. *The Journal of International Trade & Economic Development*, 20(5): 677–709.

KIND, H. J. (2002). Endogenous growth and trade liberalization between small and large countries. *Review of International Economics*, 10(1): 151–165.

KLASRA, M. A. (2011). Foreign direct investment, trade openness and economic growth in Pakistan and Turkey: an investigation using bounds test. *Quality & Quantity*, 45(1): 223–231.

KONG, Q. et al. (2021). Trade openness and economic growth quality of China: Empirical analysis using ARDL model [online]. *Finance Research Letters*, 38. <https://doi.org/10.1016/j.frl.2020.101488>.

KYRRE, S. (2006). *Trade Openness and Economic Growth: Do Institutions Matter?* Norwegian Institute of International Affairs (NIIA), 702.

LAWAL, A. I., NWANJI, T. I., ASALEYE, A., AHMED, V. (2016). Economic growth, financial development and trade openness in Nigeria: an application of the ARDL bound testing approach [online]. *Cogent Economics & Finance*, 4(1), <http://dx.doi.org/10.1080/23322039.2016.1258810>.

LEAMER, E. E. (1995). A trade economist's view of US wages and globalisation. In: *Brooking Conference Proceedings*, Brookings.

MALEFANE, M. R., ODHIAMBO, N. M. (2021). Trade openness and economic growth: empirical evidence from Lesotho [online]. *Global Business Review*, 22(5): 1103–1119. <https://doi.org/10.1177/0972150919830812>.

MARELLI, E., SIGNORELLI, M. (2011). China and India: Openness, trade and effects on economic growth. *The European Journal of comparative economics*, 8(1): 129.

MATTHEW, O., ADEGBOYE, F. B. (2014). Trade openness, institutions and economic growth in Sub-Saharan Africa (SSA). *Developing Country Studies*, 4(8): 18–30.

MERCAN, M., GOCER, I., BULUT, S., DAM, M. (2013). The effect of openness on economic growth for BRIC-T countries: Panel data analysis. *Eurasian Journal of business and economics*, 6(11): 1–14.

OMOKE, P. C., OPUALA−CHARLES, S. (2021). Trade openness and economic growth nexus: Exploring the role of institutional quality in Nigeria [online]. *Cogent Economics & Finance*, 9(1). <https://doi.org/10.1080/23322039.2020.1868686>.

PELINESCU, E. (2015). The impact of human capital on economic growth. *Procedia Economics and Finance*, 22: 184–190.

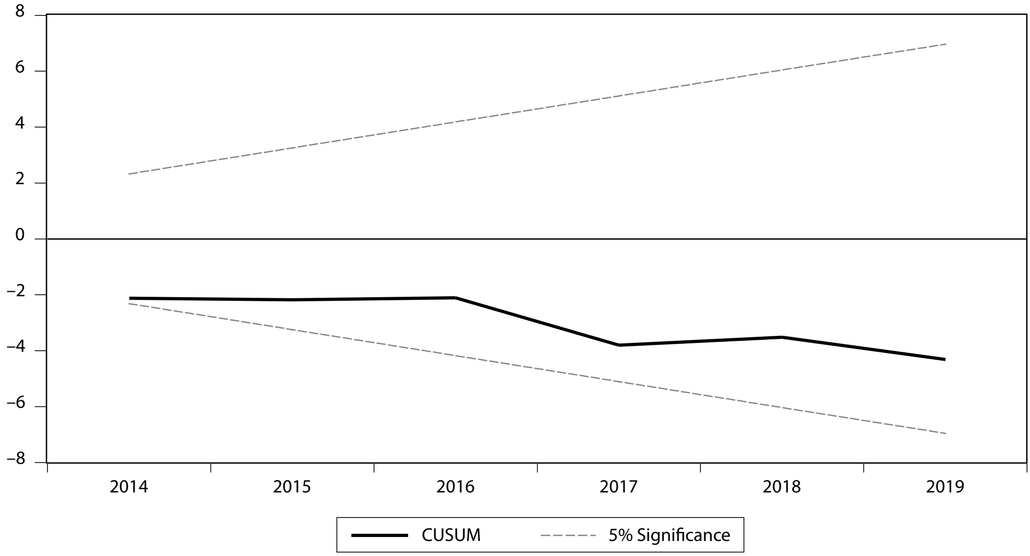
PESARAN, M. H., SHIN, Y., SMITH, R. J. (2001). Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics*, 16(3): 289–326.

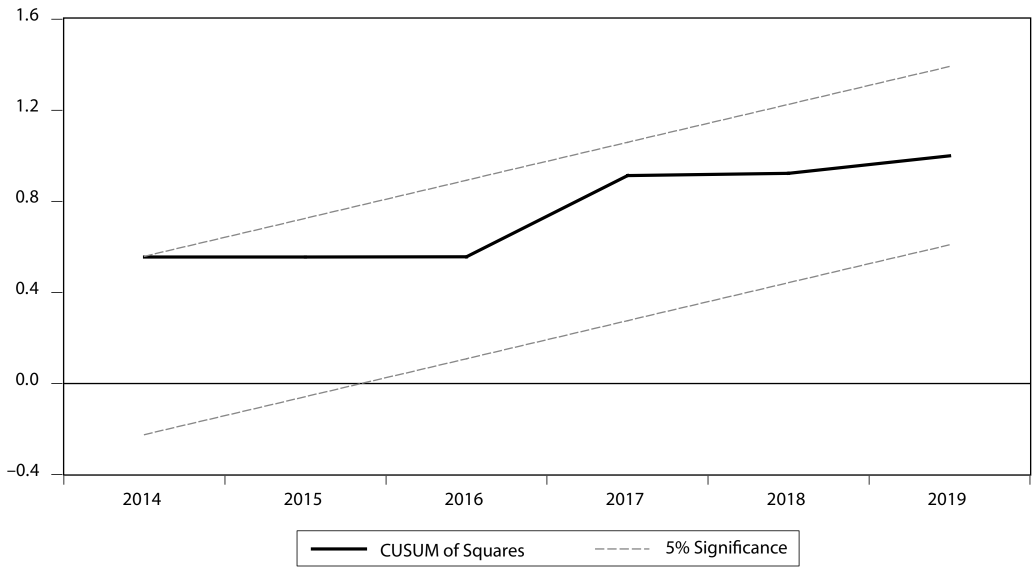
RAYCHAUDHURI, A., HALDAR, S. K. (2009). An investigation into the inter-district disparity in West Bengal, 1991–2005. *Economic and political weekly*, 258–263.

RIGOBON, R., RODRIK, N. (2004). *The Determinants of Institutions, Openness, and Income*. MIT, Mimeo.
RIVERA-BATIZ, L. A., ROMER, P. M. (1991). Economic integration and endogenous growth. *The Quarterly Journal of Economics*, 106(2): 531–555.
RODRIK, D. (1999). Where did all the growth go? External shocks, social conflict, and growth collapses. *Journal of economic growth*, 4(4): 385–412.
RODRIK, D., SUBRAMANIAN, A., TREBBI, F. (2004). Institutions rule: the primacy of institutions over geography and integration in economic development. *Journal of economic growth*, 9(2): 131–165.
ROMER, P. M. (1990). The problem of development: a conference of the Institute for the study of free enterprise systems. *Journal of Political Economy*, 98(1): 1–11.
SARKAR, P. (2005). Is there any impact of trade liberalisation on growth? Experiences of India and Korea. *Revue Tiers Monde*.
SARKAR, P. (2008). Trade openness and growth: Is there any link? *Journal of economic issues*, 42(3): 763–785.
SENGUPTA, S. (2020). How trade openness influenced economic growth in India: an empirical investigation. *Indian Journal of Economics and Development*, 8: 1–14.
SHAHBAZ, M. (2012). Does trade openness affect long run growth? Cointegration, causality and forecast error variance decomposition tests for Pakistan. *Economic Modelling*, 29(6): 2325–2339.
SINGH, T. (2011). International trade and economic growth nexus in Australia: a robust evidence from time-series estimators. *The World Economy*, 34(8): 1348–1394.
SU, D. T., NGUYEN, P. C., CHRISTOPHE, S. (2019). Impact of foreign direct investment, trade openness and economic institutions on growth in emerging countries: the case of Vietnam [online]. *Journal of International Studies,* 12(3): 243–264. <https://doi.org/10.14254/2071-8330.2019/12-3/20>.
SUBRAMANIAN, A., TREBBI, F., RODRIK, D. (2002). Institutions rule: the primacy of institutions over integration and geography in economic development. *IMF Working Papers, 189.*
TAHIR, M., AZID, T. (2015). The relationship between international trade openness and economic growth in the developing economies: Some new dimensions. *Journal of Chinese Economic and Foreign Trade Studies*.
TAHIR, M., HAYAT, A. (2020). Does international trade promote economic growth? An evidence from Brunei Darussalam [online]. *Journal of Chinese Economic and Foreign Trade Studies*. <https://doi.org/10.1108/JCEFTS-03-2020-0010>.
TOPALOVA, P., KHANDELWAL, A. (2011). Trade liberalisation and firm productivity: the case of India. *Review of economics and statistics*, 93(3): 995–1009.
UDEAGHA, M. C., NGEPAH, N. (2021). The asymmetric effect of trade openness on economic growth in South Africa: a nonlinear ARDL approach [online]. *Economic Change and Restructuring*, 54(2): 491–540. <https://doi.org/10.1007/s10644-020-09285-6>.
VAMVAKIDIS, A. (2002). How robust is the growth-openness connection? Historical evidence. *Journal of Economic Growth*, 7(1): 57–80.
VED, P., SUDESH, P. (2007). An empirical investigation of the casual: relationship between openness and economic growth in India. *Asian Economic Review*, 49(3): 485–494.
WACZIARG et al. (2008) Trade liberalisation and growth: New evidence. *The World Bank Economic Review,* 22(2): 187–231.
WANI, S. (2022). Trade Openness, Capital Formation, and Economic Growth: Empirical Evidence from India. *Eurasian Journal of Business and Economics*, 15(29): 35–49.
WEI, S. J. (2000). *Natural openness and good government* [online]. Working paper 7765. <https://www.nber.org/system/files/working_papers/w7765/w7765.pdf>.
WINTERS, L. A. (2004). Trade liberalisation and economic performance: an overview. *The Economic Journal*, 114(493): F4–F21.
XIE, H., CAI, Y., SAM, C. Y., CHANG, T. (2018). Revisit Financial Development, Trade Openness and Economic Growth Nexus in China Using a New Developed Bootstrap ARDL Test. *Economic Computation & Economic Cybernetics Studies & Research*, 52(4).
YANIKKAYA, H. (2003). Trade openness and economic growth: a cross-country empirical investigation [online]. *Journal of Development Economics,* 72(1): 57–89. <https://doi.org/10.1016/s0304-3878(03)00068-3>.
YOUNG, A. (1991). Learning by doing and the dynamic effects of international trade. *The Quarterly Journal of Economics*, 106(2): 369–405.
ZAHONOGO, P. (2016). Trade and economic growth in developing countries: Evidence from sub-Saharan Africa. *Journal of African Trade*, 3(1–2): 41–56.

APPENDIX

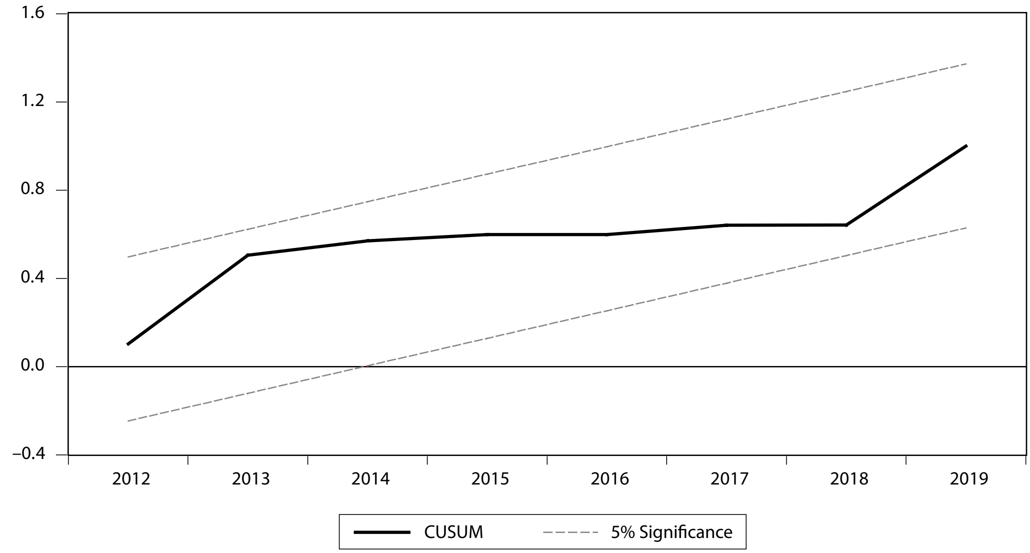
Figure A1 CUSUM and CUSUMSQ (Specification 1)

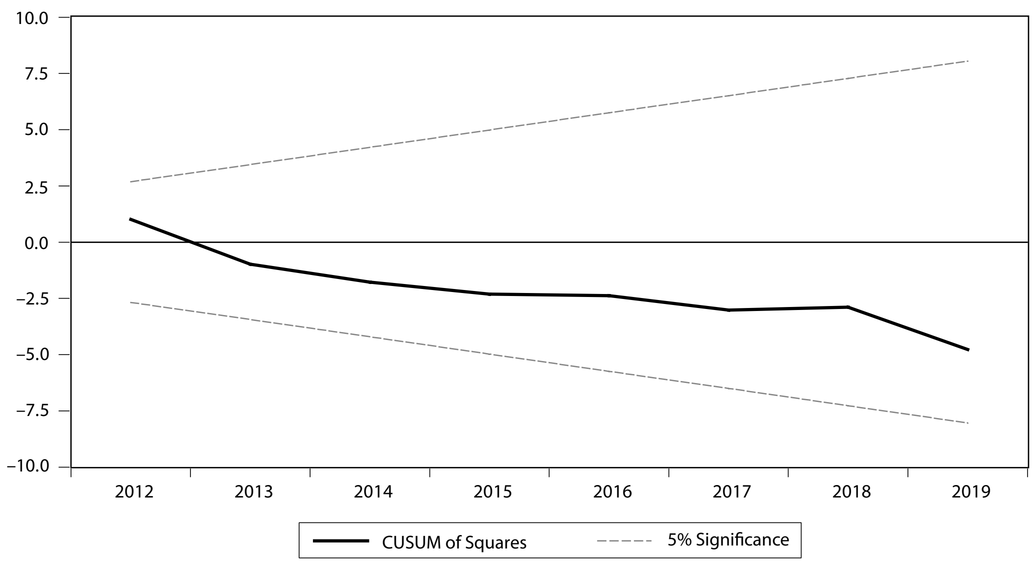




Source: Author's calculation

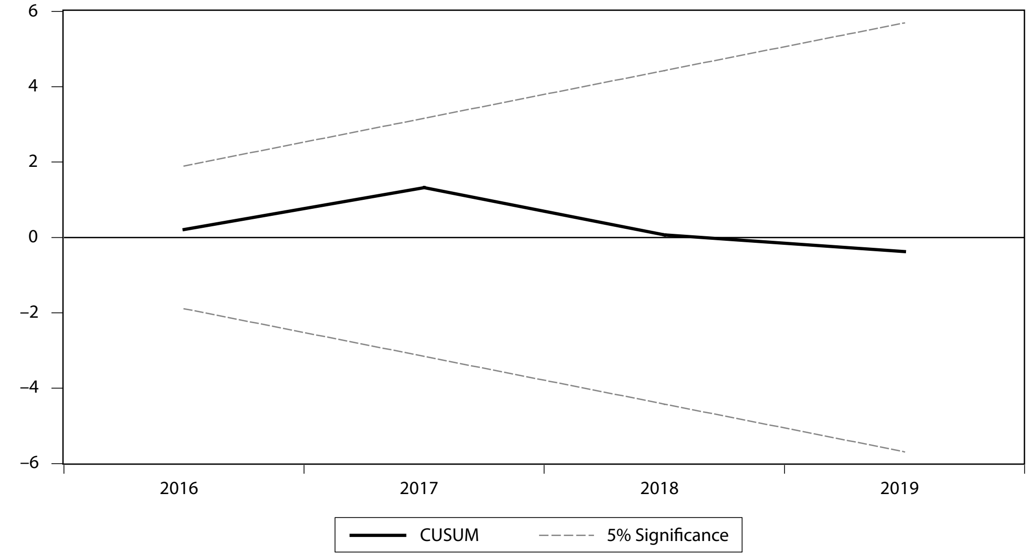
Figure A2 CUSUM and CUSUMSQ (Specification 2)

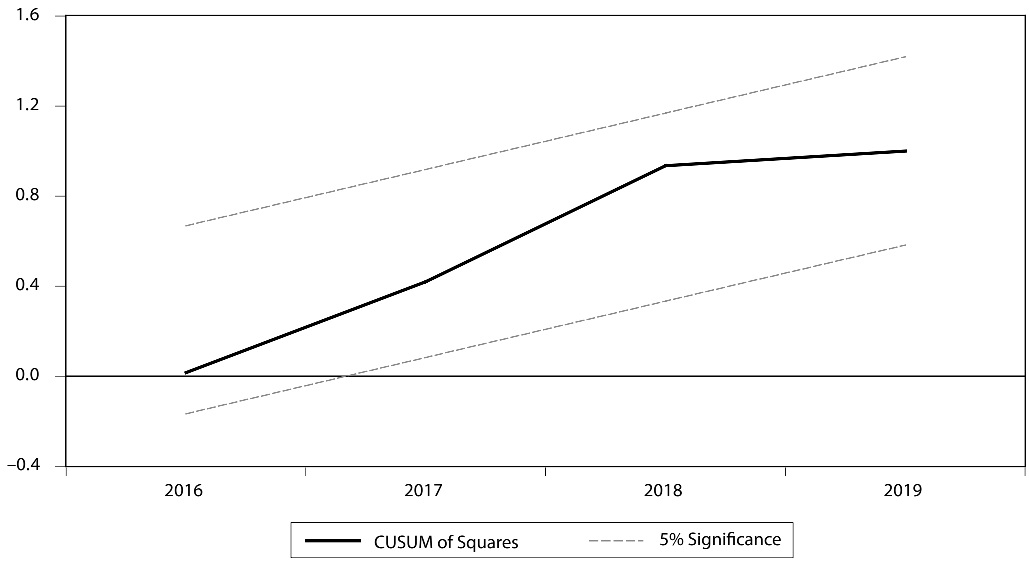




Source: Author's calculation

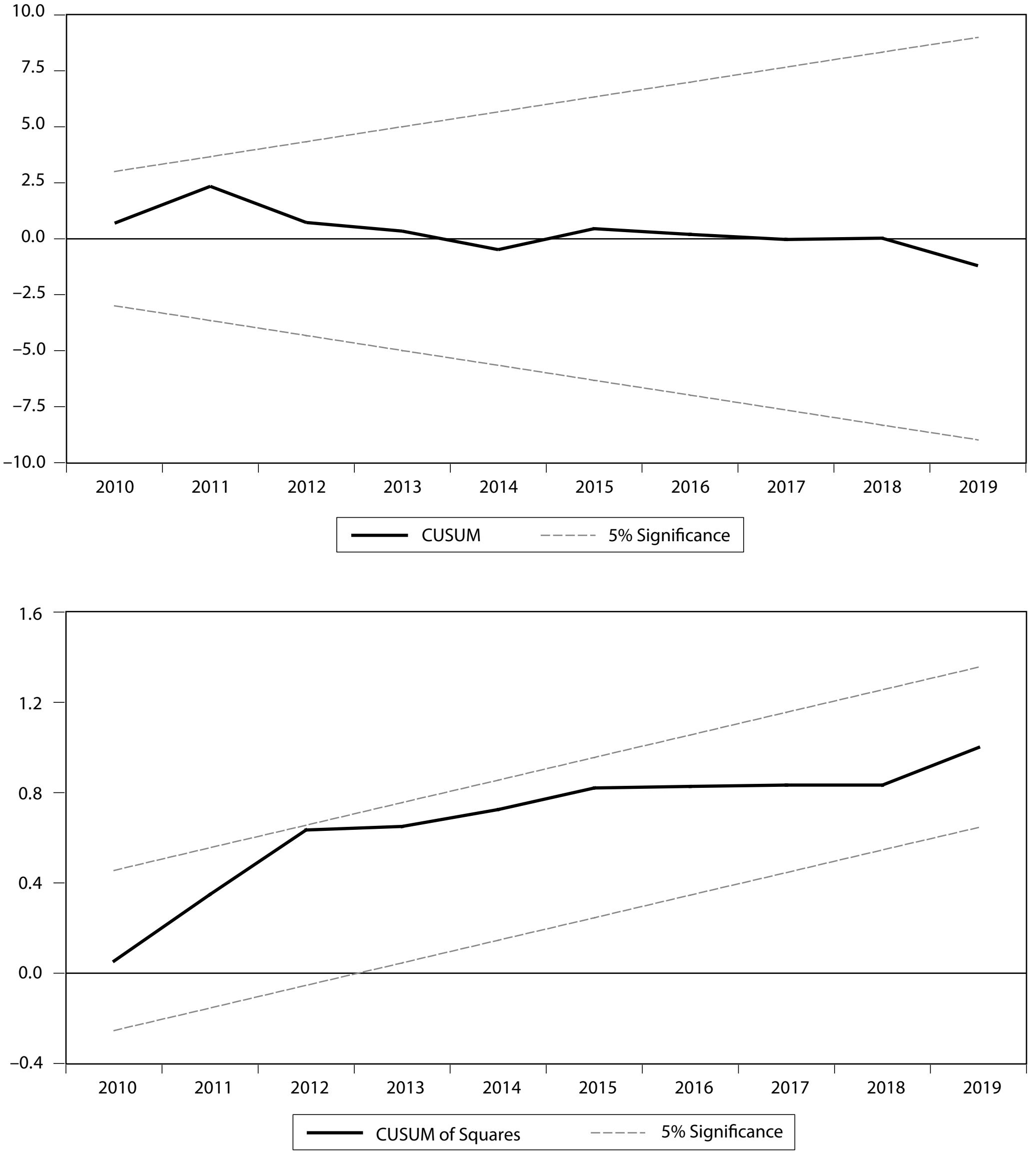
Figure A3 CUSUM and CUSUMSQ (Specification 3)





Source: Author's calculation

Figure A4 CUSUM and CUSUMSQ (Specification 4)



Source: Author's calculation

Hedonic Price Methods and Real Estate Price Index: an Explanatory Study for Apartments Market in Belo Horizonte, Brazil, from 2004 to 2015

Luiz Andrés Ribeiro Paixão[1] | *Brazilian Institute of Geography and Statistics, Rio de Janeiro, Brazil*

Received 30.9.2022, **Accepted** (reviewed) 16.11.2022, **Published** 16.6.2023

Abstract

Brazil does not have an official real estate price index yet. A 2011 presidential decree stipulated for the Brazilian Institute of Geography and Statistics (IBGE) the assignment to create and spread a real estate price index for Brazil. In this paper, we test some different hedonic model methods to estimate quarterly price indices for apartments in Belo Horizonte, Brazil, from January 2005 to December 2015. Our goals are: i) to measure and compare the different hedonic methods; ii) to present some results that will contribute to the discussion about the development of an official real estate price index in Brazil. The empirical results corroborate the idea of intense apartment prices valuation in Belo Horizonte, mainly between 2007 and 2011, when the annual price index growth remained above 20%. These results cast light on the potential use of both hedonic methods and administrative data base to construct an official real estate price index for Brazil.

Keywords	***DOI***	***JEL code***
Price indices, hedonic price model, housing market	*https://doi.org/10.54694/stat.2022.46*	*C43, E31, R31*

INTRODUCTION

Subprime crises turned attention around the world to the real estate price dynamics question. In Brazil the recent large valuation of real estate price adds more attention to the subject. Academics, news, real estate agents, Brazilians government agencies and statistical institutes started to discuss the importance

[1] Brazilian Institute of Geography and Statistics (IBGE), Avenida República do Chile 500, 5o andar Rio de Janeiro, Brazil. E-mail: luiz.paixao@ibge.gov.br, phone: (+55)2121420194.

of having an adequate measure of real estate price over time. The Federal Government Decree number 7 565, dated 21 September, 2015, established IBGE (Brazilian Institute of Geography and Statistics) as the responsible entity to create and disseminate an official real estate price index for Brazil. Since then, IBGE and others governmental agencies have implemented studies regarding the database and methodology to construct a future Brazilian official real estate price index (Nadalin and Furtado, 2011; Santos and Salazar, 2011).

The Brazilian academy has been studying real estate price index and its application in Brazilian context. In recent years, Rozenbaum (2009), Paixão (2015), and Simões (2017) are examples of doctoral thesis related to this subject. Rozenbaum (2009) used administrative data to construct a hedonic quality adjusted price index for the city of Rio de Janeiro. Simões (2017) also measured the hedonic quality adjusted price index for Rio de Janeiro, using real estate agencies data, while Paixão (2015), also used administrative data to construct a hedonic quality adjusted price index for Belo Horizonte's city.

Some researches in this field published in Brazilian academic journals. Gonzalez (1997) estimated a simple time-dummy hedonic model to construct price index for apartments rents in Porto Alegre. The same approach was applied by Rozenbaum and Macedo-Soares (2007), to estimated real estate valuation in Rio de Janeiro's district of Barra da Tijuca, and by Paixão (2015a), to estimate real estate price indices for Belo Horizonte. Albuquerque et al. (2018) used repeated sales method to construct an index for the city of Brasilia, the capital of Brazil.

Some University's agencies like the Institute of Economic Research Foundation (FIPE), from University of São Paulo (USP), and the Institute of Economic, Administrative and Accounting Research (IPEAD), from Federal University of Minas Gerais (UFMG), released real estate price indices using stratified median methods. The widespread FIPE-ZAP real estate index is calculated by FIPE from the real estate advertised data collected in ZAP's web site platform. The Brazilian Central Bank also estimated and published a monthly stratified median real estate price index, constructed from the real estate loans data, called Financed Residential Properties index (IVG-R). Despite the importance and relevance of those indices for the society, government, academics and real estate agencies some gaps remain. None of those indices used the hedonic quality adjustment methodology, recognized as the best to deal with the nature of real estate's market data (Diewert, 2009; Hill et al., 2018). Besides that, only IPEAD uses administrative data which cover the whole transacted market.

In this paper we try to construct quarterly hedonics quality adjusted real estate prices indices for Belo Horizonte city using administrative data. For this task we will use the hedonic methods proposed by Hill's (2013), and Hill's et al. (2018). We will use the same methods used by Hill et al. (2018) in their analysis of Sidney and Tokyo markets, to produce comparable results.

The last part of the paper is structured as follows. The next section explains the different hedonic price methods used to measure the quality adjusted housing price indices. The second section is focused on the database and introduces the Brazilian city of Belo Horizonte. The objective of the third section is to test the several hedonic methods in Belo Horizonte's real estate market database. Finally, our main results are summarized in the conclusion.

1 HEDONIC QUALITY ADJUSTED REAL ESTATE PRICE METHODS

1.1 The hedonic price model

Estimating housing price indices is a complex task. Housing is a type of complex good (or service), that is, a good where each unit or model differs from the others in qualitative terms. A complex good can be described as a bundle of many characteristics (or attributes), so each unit or model is a peculiar bundle of attributes. Computing index prices for complex good necessarily means controlling the change in the good price by the change in composition of its characteristics.

The hedonic price model establishes a functional relationship between the price of the good and its characteristics. In a hedonic perspective a good is a basket of characteristics Z, as represented below:

$$Z = Z\left(z_{1,} z_2, \ldots, z_n\right). \quad (1)$$

The price of a good follows a hedonic function as describes in Formula (2):

$$P = P\left(z_{1,} z_2, \ldots, z_n\right). \quad (2)$$

Although only the price of the good can be observed in the market, the hedonic function establishes that the price of a good is determined by the composition of basket of characteristics. Therefore, each attribute (i) has an unobserved price (implicit price) that is represented by the first derivative of the hedonic function with respect to i.

$$p_i = \frac{\partial P}{\partial x_i}. \quad (3)$$

The seminal paper of Rosen (1974) validated the hedonic price model in theoretical terms. Empirically, Waugh (1928) was pioneer in apply hedonic regression in vegetables market study. Court (1939) used hedonic price regression to construct automobile price indices. Griliches (1958, 1961) constructed hedonic quality adjusted price indices for fertilizers and automobile markets respectively. From Griliches contributions, the application of hedonic model widespread in the academic world, covering many types of different goods and services like computers, refrigerators, fruits, musical instruments, paints etc. However, it was in the real estate market that the hedonic approach achieved its largest projection.

1.2 The hedonic quality adjusted price indexes: the real estate case

Griliches (1971) argues that a complex good's price change can be divided in two dimensions. The first is the observed price change of the good in the market. The second is the unobserved price change of the basket of characteristics. To estimate the unobserved price change, it is necessary to use the hedonic model regression as a quality adjusted factor. Discounting the price change of the attributes bundle from the observed price of the good results in a "pure" estimate of a complex good price change.

There are several ways how to construct a quality adjusted price index from the hedonic methodology. Court (1939) and Griliches (1961) already advanced some questions, like the possibility of using both cross-section regressions or time-dummy approaches. Tripllet (2004) created a taxonomy of the several hedonic methods used to compute quality adjusted price indices for technological goods. Hill (2013) applied this Tripllet's taxonomy to the housing market case.

Hill et al. (2018) using a Hill (2013) approach compiled the hedonic methods used by the European national statistics institutes. The first category embraces all indices which requires cross-section regressions models and, as a result, involves data imputation. The second category is based on time-dummy regressions.

1.2.1 Imputation approach

1.2.1.1 Repricing model

The first imputation method described by Hill et al. (2018) is the repricing method. Like Hill et al. (2018) in this study we adopted quarterly price indices as default and hedonic quality adjusted price indices could be constructed for any period. Defining the base period is the first task in the repricing model. Then, a hedonic regression is estimated for this period. The price implicit estimated in hedonic regression is used to impute prices for each subsequent quarterly. The base period can be fixed or be updated at regular time intervals.

Hill et al. (2018) recommends estimating one regression for the whole base year:

$$\ln(p)_{(1,q),h} = \sum_{c=1}^{C} \beta_{1,c} z_{(1,q),h,c} + \varepsilon_{(1,q),h} \,, \tag{4}$$

where ln(*p*) is the natural logarithmic of housing price in the base year (1), *q* is the quarterly of the sale, *h* is the dwelling sold and *c* is each characteristic of the dwelling.

Then, the implicit prices estimated ($\hat{\beta}$) are used to estimated prices for each subsequent quarterly. The repricing method is, therefore, a sort of Laspeyres index. The quality adjustment factor $QAF_{(t,q-1),(t,q)}$ is defined as the ratio of the imputed prices for adjacent quarters, q and $q-1$ for example.

$$QAF_{(t,q-1),(t,q)} = \frac{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q),c}\right)}{\exp\left(\sum_{c=1}^{C} \hat{\beta}_{1,c} \bar{z}_{(t,q-1),c}\right)}. \tag{5}$$

To construct the repricing method price index (RP) a quality unadjusted price index ($QUPI_{(t,q),(tq-1)}$) defined as a ratio between geometric mean prices ($\tilde{p}$) of adjacent quarters is calculated as follows:

$$QUPI_{(t,q),(tq-1)} = \frac{\tilde{p}_{(t,q)}}{\tilde{p}_{(t,q-1)}} \,. \tag{6}$$

Finally, RP is the ratio between quality unadjusted price factor and quality adjusted price factor.

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \frac{QUPI_{(t,q),(t,q-1)}}{QAF_{(t,q),(t,q-1)}} \,. \tag{7}$$

The main attractive feature of RP relies on the fact that it is not regression intensive. In the end, it requires only one regression (Hill et al., 2018). However, to achieve good results, Hill et al. (2018: 224) suggested that "the base year under the repricing method should be updated at regular time intervals". Italy and Luxembourg national statistics institutes are benchmark examples since both updated the base every year (Hill et al., 2018).

1.2.1.2 Average characteristic method

The average characteristic method (AC) requires, as any method, a definition of a base period. After that, the average characteristic of the dwellings sold in the base period are computed. The next step consists in estimating hedonic regressions for each subsequent period, quarterly in our case. Then the imputed prices are calculated, applying the estimated quarterly implicit prices on the average characteristics of the base period.

Following Hill et al. (2018), the European national statistical institutes calculated the basket of average characteristics for a whole year (base year), $\bar{z}_{t-1}$. In this line, the European national statistical institutes adopted a Laspeyres version of AC[1]. The base is updated every year. Then, a hedonic regression is estimated for each quarter (q) of the following year (t):

$$\ln(p)_{(t,q),h} = \sum_{c=1}^{C} \beta_{t,c} z_{(t,q),h,c} + \varepsilon_{(t,q),h}. \tag{8}$$

The quality adjusted price index estimated by AC is given as follows:

$$\frac{P_{(t,q)}}{P_{(t,q-1)}} = \frac{\exp(\sum_{c=1}^{C} \hat{\beta}_{(t,q),c} \bar{z}_{t-1,c})}{\exp(\sum_{c=1}^{C} \hat{\beta}_{(t,q-1),c} \bar{z}_{t-1,c})} \,. \tag{9}$$

1.2.1.3 Hedonic imputation method

Real estate is a threshold situation of complex goods. Each unity of real estate differs from the other. Added to this, the set of dwellings sale in one period differs from the set of dwellings sale in other periods. Therefore, it is not possible to construct a basket of dwellings to follow over time. The hedonic imputation method is a way to estimate the price of each dwelling sold in t would have in another period, $t + 1$ for example. According to Hill et al. (2018: 225–6): once a hedonic model has been estimated, it allows one to ask counterfactual questions such as what a particular dwelling actually sold in say period t would have sold for instead in period $t + 1$.

Like the AC, one regression as (8) is estimated for each period. The regression in $t + 1$ is used to impute the price in $t + 1$ for each observed transacted dwelling in t. Likewise, the regression in t is used to impute the price in t for each observed transacted dwelling in $t + 1$. To construct the index, Hill (2013) recommended to use the regression estimated price in t instead of the observed price for each observed transacted dwelling in t. Such procedure is known as double imputation. From the hedonic imputation method geometric Laspeyres (GL), geometric Paasche (GP) and Tornqvist prices indices can be extracted.

Few European national statistical institutes use hedonic imputation methods. Hill et al. (2018) follow the German version of double imputation Tornqvist (DIT). From a set of regressions like (8) the GL, GP and DIT are estimated as follows:

$$GL = \frac{exp\sum_{c=1}^{C}\hat{\beta}_{(t,q)}\bar{z}_{(t-1,c)}}{exp\sum_{c=1}^{C}\hat{\beta}_{(t,q-1)}\bar{z}_{(t-1,c)}}, \tag{10}$$

$$GP = \frac{exp\sum_{c=1}^{C}\hat{\beta}_{(t,q)}\bar{z}_{(t,c)}}{exp\sum_{c=1}^{C}\hat{\beta}_{(t,q-1)}\bar{z}_{(t,c)}}, \tag{11}$$

$$DIT = \sqrt{GL * GP}. \tag{12}$$

1.2.2 Time-dummy approach

1.2.2.1 Simple time-dummy

The time-dummy approach consists in constructing price indices from the estimated parameter of a dummy time variable in a hedonic regression. Usually, the first period in the series is used as the base. The simple time-dummy model (TD) requires only one regression and it is the simples and most intuitive hedonic method. The typical TD regression is illustrated as follows:

$$\ln(p)_{(t,q),h} = \sum_{c=1}^{C}\beta_{t,c}z_{(t,q),h,c} + \gamma_{t,q}D_{t,q} + \varepsilon_{(t,q),h}, \tag{13}$$

where $D_{t,q}$ is a time dummy for each period and $\gamma_{t,q}$ is the prince index estimated for the period .

Despite its simplicity, there are some pitfalls in using TD (Hill, 2013). First, the TD does not allow the implicit price changes over time. As a result, the longer the series, the worst will the TD prince index estimations be. For national statistical institutes, TD is not recommended because it does not follow the temporal fixity criterium, as defined by Hill (2004). According to this criterium, once an index has already been disseminated by the national statistical institute it should remain unchanged when new data becomes available. Using the single regression TD approach, when new data is added, a new estimation of (13) changes all parameters $\gamma_{t,q}$ previously disseminated.

1.2.2.2 Rolling time-dummy method

The hedonic rolling time dummy method (RTD) consists in estimated hedonics time-dummy regressions for subperiods instead of only one regression for the whole period. The limiting case occurs when a regression is estimated for each pair of adjacent periods. Although when data points are scarce it is recommended to estimate a regression including more than one subperiod.

France and Portugal, for example, estimated price index from RTD with 2 quarter windows. In other words, both countries are using an adjacent period RTD. Other countries like Cyprus and Croatia estimated a 4 quarter windows RTD. The RTD price index is calculated from an RTD regression like (13) as showed below:

$$\frac{P_{q+1}}{P_q} = \frac{\exp\left(\hat{\gamma}_{q+1}\right)}{\exp\left(\hat{\gamma}_q\right)}. \quad (14)$$

2 THE DATA

2.1 Belo Horizonte, Brazil: an overview

Belo Horizonte, the capital of the State of Minas Gerais, is an important economic, politic and cultural center in Brazil. According to 2010 Brazilian Census (IBGE, 2010), Belo Horizonte had a population of almost 2.4 million and was the 6th most populous city in Brazil. The Metropolitan Area (MA) of Belo Horizonte, in turn, had a population of 5.4 million and was the 3th most populous MA in Brazil.

Belo Horizonte was a planned city, conceived to replace Ouro Preto as the capital of Minas Gerais, and was founded in 1897. Nowadays, the planned area corresponds to downtown and its nearby districts bounded by Contorno Avenue. Like Aguiar et al. (2014, 119) resumes: this planning created a center-periphery radial model for the city, which concentrated urban services and urban infrastructure in particular areas, and reinforced social disparities.

From an administrative point of view, the space of Belo Horizonte is divided into Districts (487), Planning Units (82), and Regionals (9). Following Villaça (1998), historically, the Central-South city's Regional (Regional Centro-Sul) concentrated the elite's neighborhoods. Nowadays, some bordering Central-South Regional districts in the West Regional (Regional Oeste) are also occupied by Belo Horizonte's elite. Nonetheless, a few elite's districts are in Pampulha Regional (Regional Pampulha), in the north of the city, surrounding Pampulha's lagoon.

2.2 Database

In Brazil all real estate transactions are subjected to the Real Estate Transfer Tax[2] (RETT) and its collection is in charge of municipalities. We used Belo Horizonte municipality's RETT as our dataset, covering the period from 2004 to 2015, collected by IPEAD/UFMG. The RETT dataset contains the value of transaction, type of building, area, age, quality of building finishing material, zoning and location (district). The type of building includes apartments, houses and commercial real estate. In this paper we analyzed only apartments market. Further we shall expand the analysis for house and commercial real estate markets.

Tables 1, 2 and 3 resume the data. There were 266 529 observations in our dataset for the whole period. The mean apartment price was around R$ 218 149 (approximately U$ 57 000) and the standard deviation was 235 011, indicating a high dispersion of this variable. Apartments sold in Belo Horizonte were fairly big and new, the mean area and age was 120 m^2 and 13 years, respectively. Most of apartments were

[2] Imposto de Transmissão Imobiliária Inter-Vivos (ITBI).

classified as normal in terms of quality of building finishing material and this variable classified the quality into 5 categories. Ordering from the top there were the following categories: luxury, high, normal, low and popular. Most observations where located in the Center-South (Centro-Sul) and West (Oeste) Regionals.

Table 1 Descriptive statistics for apartments in municipality of Belo Horizonte: 2004–2015

Year	Observation	Value (reais)			Area (m²)			Age (years)		
		Mean	Median	Standard deviation	Mean	Median	Standard deviation	Mean	Median	Standard deviation
2004	17 767	90 682	60 122	91 427	122.0	102.9	70.6	12.4	8.0	11.7
2005	39 606	93 363	61 746	99 219	118.7	99.1	68.9	13.7	10.0	12.0
2006	37 614	107 572	70 196	117 326	120.4	102.1	69.2	14.3	10.0	12.2
2007	19 664	125 848	83 000	129 249	120.9	103.2	68.4	14.3	10.0	12.2
2008	19 224	155 117	100 000	156 457	121.4	101.7	70.6	14.5	10.0	12.5
2009	18 272	186 106	130 000	172 571	118.4	98.9	69.8	14.0	10.0	12.8
2010	21 177	227 110	160 764	201 441	114.4	94.4	68.5	12.1	8.0	13.1
2011	19 257	297 875	220 000	247 637	118.9	98.3	70.3	11.7	7.0	13.3
2012	18 408	362 735	274 412	284 955	123.8	107.0	70.9	11.8	6.0	13.5
2013	20 364	389 815	303 982	283 168	120.5	104.4	67.9	11.1	4.0	13.7
2014	19 516	418 895	330 000	291 744	120.6	105.7	66.9	10.1	2.0	13.3
2015	15 727	420 250	334 000	290 847	117.0	102.1	64.9	10.9	3.0	13.8
2004–2015	266 596	218 149	140 000	235 011	119.7	101.4	69.0	12.8	8.0	12.8

Source: IPEAD/UFMG, author's calculation

Table 2 Distribution of quality building finishing materials: Belo Horizonte, 2004–2015

Quality of building finishing materials	Mean	Standard deviation
Popular	0.02	0.15
Low	0.21	0.41
Normal	0.59	0.49
High	0.16	0.37
Luxury	0.02	0.13

Source: IPEAD/UFMG, author's calculation

Table 3 Apartments transactions – mean for regional – Belo Horizonte: 2004–2015

Regional	Mean	Standard deviation
Centro-Sul	0.21	0.31
Leste	0.09	0.28
Nordeste	0.10	0.29
Noroeste	0.11	0.31
Oeste	0.21	0.40
Pampulha	0.15	0.35
Venda Nova	0.05	0.21
Barreiro	0.04	0.18
Norte	0.04	0.18

Source: IPEAD/UFMG, author's calculation

3 QUALITY ADJUSTED PRICE INDICES FOR BELO HORIZONTE, BRAZIL

3.1 Hedonic prices indices for Belo Horizonte, Brazil

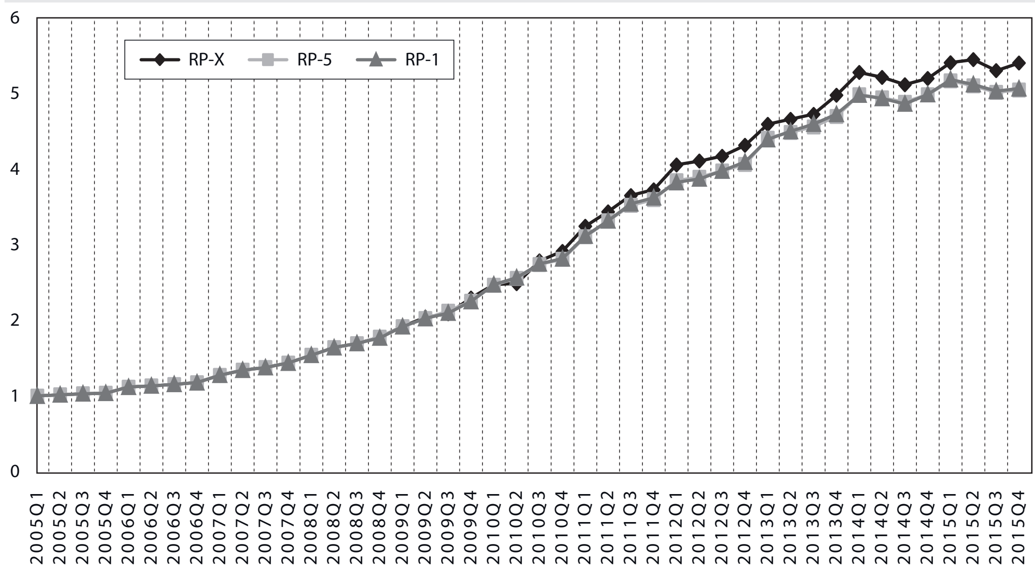
We estimated quarterly housing price indices for Belo Horizonte using the various methods discussed in Sections 1.2.1 and 1.2.2 and an UP's stratified median method (MIX-UP).[3] Since some methods were based on previous regression or mean characteristics, the 2004 data was used to compute reference baskets used in 2005. For this reason, results are represented for years 2005–2015. Following Hill et al. (2018) we estimated three forms of repricing model (RP): i) (RP-X, which uses shadow price from 2004 (no updating base year); ii) RP-5, which updates the shadows prices every five years; iii) RP-1, which updates the shadows prices annually.

The average characteristic indices (AC) were estimated with a base update every year. In AC case, the base is the one year lagged average characteristics. The double imputation indexes were calculated estimating, for some quarter set of observations, counterfactual housing's basket for previous and posterior periods. The double imputation Laspeyres (DIL), Paasche (DIP) and Tornqvist (DIT) were estimated as presented in Section 1.2.1.3. The rolling time dummy indices were estimated for 2 (RTD2) and 4 (RTD4) quarters window. Finally, we estimated an UP stratified median (MIX-UP) index, to compare the quality adjusted hedonic housing price index with a simpler and more intuitive price index. Table 6 resumes the quarter housing price indices estimated.

From the chosen period (2005–2015), the large appreciation of housing prices in Brazil was supported by estimated indices. Even so, the magnitude of the appreciation differs between methods. Using the DIP index, Belo Horizonte's house prices rose 383.6% in contrast with DIL which shows an increase of 435.7%.

RIP-X, which uses the shadow price from 2004, was apartheid from RP-5, which changes the base every five years, and RP-1, which updates shadow prices every year (Figure 1). From our Belo Horizonte's

Figure 1 Repricing Indices for apartments in Belo Horizonte (2005Q1 = 100)



Source: IPEAD/UFMG, author's calculation

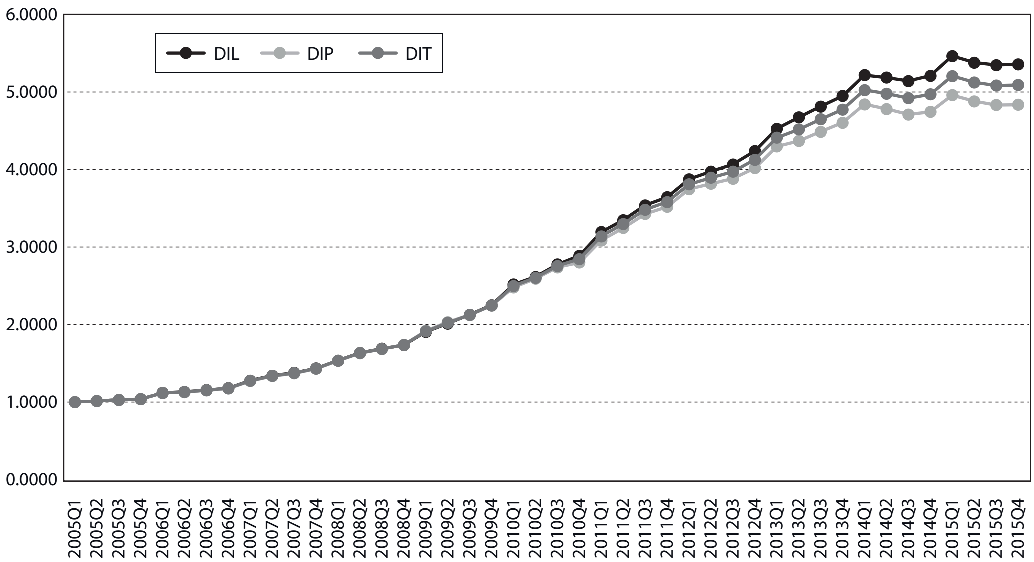
3 The control variables of hedonic price regression are resumed in Tables A1, A2, A3 and A4 in the Annex. We illustrated the results for control variables using a time-dummy regression (Formula 13) for the whole period in Table A5, also in the Annex. The results for the other regressions are available from the author.

database, the RIP-X seems not be an appropriate index due to its failure to control shadow prices change over time. RP-5 and RP-1 lines were close each other, highlighting the importance to update the base year from time to time. Since RP-1 is the more flexible RP index it will be used in the remainder of the paper.

The double imputation Laspeyres (DIL) and double imputation Paasche (DIP) showed evidence of drift (Figure 2), as Hill el al. (2018) also have noted for the Sydney data. For Belo Horizonte apartment's market, DIP estimated the smallest price variation while DIL estimated the biggest. In agreement with price index theory, DIL, as a Laspeyres index, tends to overestimate the price change and DIP, as a Paasche index, tends to underestimate it. Double imputation Tornqvist (DIT), in any case, does not exhibit a drift. DIT, as a Tornqvist-Geometric index, is a geometric mean of DIP and DIL. Bearing in mind that the Tornqvist indices are recognized as superlatives, the DIT becomes an attractive alternative method to compute housing index prices. Since DIP a DIL exhibited a drift behavior, the imputation methods will be reduced to DIT in the following analysis.

The rolling time dummies (RT) were estimated for 4 quarter window (RT4) and 2 quarter window (RT2) from our database. On the RT4 index the quarter base changes once a year, and on RT2 the base changes every quarter. RTs methods are attractive because the index corresponds to the estimated regression time dummy parameter. RT2 indices stayed above RT4 for the whole period. Both RT indices will be kept in the following analyses.

Figure 2 Double Imputation Indices for apartments in Belo Horizonte (2005Q1 = 100)

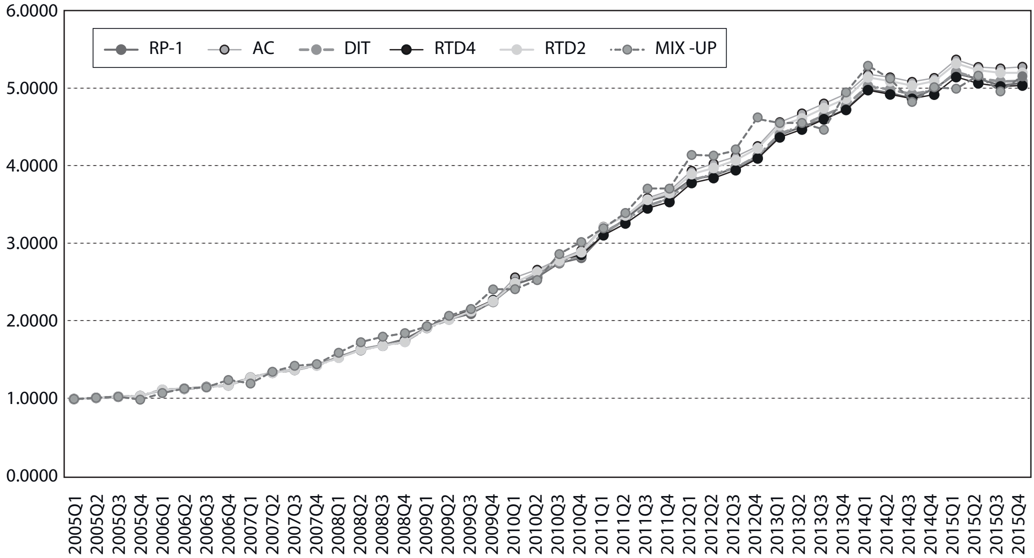


Source: IPEAD/UFMG, author's calculation

The Figure 3 compares the Belo Horizonte's housing price indices estimated by different hedonic methods and by stratified median.

DIT, RP1 and RT4 exhibited a very close behavior. AC and RT2 do the same, although the latter stayed above the former in the most recent quarters. Partly, the different behavior between AC and DIT is expected, since the first is a Laspeyres type of index and the second is a Tornqvist type. The MIX-UP line was more volatile than the other index lines due to the lack of characteristics control related to this method.

Figure 3 Estimate of Price Indices for apartments in Belo Horizonte (2005Q1 = 100)



Source: IPEAD/UFMG, author's calculation

Hill et al. (2018) recommended analyzing the volatility of indices in more detail. The authors present two volatility measures: the root mean squared error (RMSE) and mean absolute deviation (MAD). Also, they calculated the minimum (MIN) and maximum (MAX) value for each index. All these indicators are computed both on a year-by-year and quarter-by-quarter basis. The indicators volatility formulas are specified below:

$$RMSE = \sqrt{\frac{1}{T-1}\sum_{t-1}^{T-1}\left[ln\left(\frac{p_{(t+1,q)}}{p_{(t,q)}}\right)-\frac{1}{T-1}\left(\frac{p_{(T,q)}}{p_{(1,q)}}\right)\right]}, \quad (15)$$

$$MAD = \frac{1}{T-1}\sum_{t-1}^{T-1}\left|ln\left(\frac{p_{(t+1,q)}}{p_{(t,q)}}\right)-\frac{1}{T-1}ln\left(\frac{p_{(T,q)}}{p_{(1,q)}}\right)\right|, \quad (16)$$

$$MIN = Min_{1,\ldots,T-1}\left\{100\left[\frac{p_{(t+1,q)}}{p_{(t,q)}}\right]-1\right\}, \quad (17)$$

$$MAX = Max_{1,\ldots,T-1}\left\{100\left[\frac{p_{(t+1,q)}}{p_{(t,q)}}\right]-1\right\}. \quad (18)$$

The results are summarized in Table 4.

Table 4 Volatility of the House Price Indices in Belo Horizonte

	RP-X	RP-5	RP-1	AC	DIL	DIP	DIT	RTD4	RTD2	MIX-UP
Year-on-Year (Q1)										
RMSE	0.072	0.064	0.063	0.068	0.064	0.065	0.064	0.063	0.067	0.103
MAD	0.063	0.057	0.055	0.058	0.052	0.056	0.054	0.054	0.056	0.087
MIN	2.478	3.666	4.081	3.653	4.714	2.438	3.570	3.396	3.516	–5.605
MAX	31.089	28.572	29.046	32.496	32.032	29.225	30.621	29.619	30.013	32.880
Year-on-Year (Q2)										
RMSE	0.072	0.066	0.065	0.068	0.065	0.066	0.065	0.065	0.068	0.078
MAD	0.056	0.053	0.053	0.056	0.051	0.055	0.053	0.052	0.054	0.063
MIN	2.478	3.666	4.081	3.653	4.714	2.438	3.570	3.396	3.516	–5.605
MAX	31.089	28.572	29.046	32.496	32.032	29.225	30.621	29.619	30.013	32.880
Year-on-Year (Q3)										
RMSE	0.077	0.074	0.076	0.075	0.072	0.074	0.073	0.072	0.075	0.085
MAD	0.067	0.065	0.066	0.064	0.059	0.063	0.061	0.061	0.064	0.077
MIN	3.601	2.651	3.459	3.391	4.018	2.578	3.295	3.230	3.273	2.861
MAX	33.371	29.050	31.020	30.412	30.474	29.011	29.740	30.045	30.609	32.779
Year-on-Year (Q4)										
RMSE	0.075	0.075	0.076	0.077	0.075	0.077	0.076	0.075	0.078	0.090
MAD	0.065	0.065	0.066	0.065	0.062	0.065	0.064	0.062	0.065	0.080
MIN	3.961	1.250	1.535	2.799	2.847	1.961	2.403	2.360	2.115	1.320
MAX	29.141	28.151	28.631	29.721	29.755	29.023	29.388	30.185	30.102	30.513
Quater-on-Quarter										
RMSE	0.033	0.029	0.029	0.031	0.029	0.029	0.029	0.028	0.029	0.047
MAD	0.028	0.025	0.024	0.024	0.023	0.024	0.024	0.024	0.024	0.039
MIN	–2.665	–1.703	–1.703	–1.779	–1.531	–1.581	–1.556	–1.567	–1.661	–5.867
MAX	11.985	10.658	10.658	13.308	11.875	10.498	11.184	10.440	11.070	13.229

Source: IPEAD/UFMG, author's calculation

MIX-UP is more volatile than the other indices. As Hill et al. (2018) pointed out, it is expected for stratified median indices (like MIX-UP) to exhibit more volatility since they are not adjusted for changes in the quality of median over time. The hedonics quality adjusted indices exhibited relative low volatility, the magnitudes were between those which were estimated for Sidney and Tokyo by Hill et al. (2018). From volatility indicators perspectives our results suggest, for Belo Horizonte's housing market in 2005–2015, that the hedonic quality adjusted housing price indices were accurate, except for the MIX-UP cases.

3.2 Apartment price valuation in Belo Horizonte, Brazil: 2005–2015

We will illustrate the previous results measuring quarterly apartments prices rate of appreciation for Belo Horizonte. Table 5 summarizes the results for different methodologies.

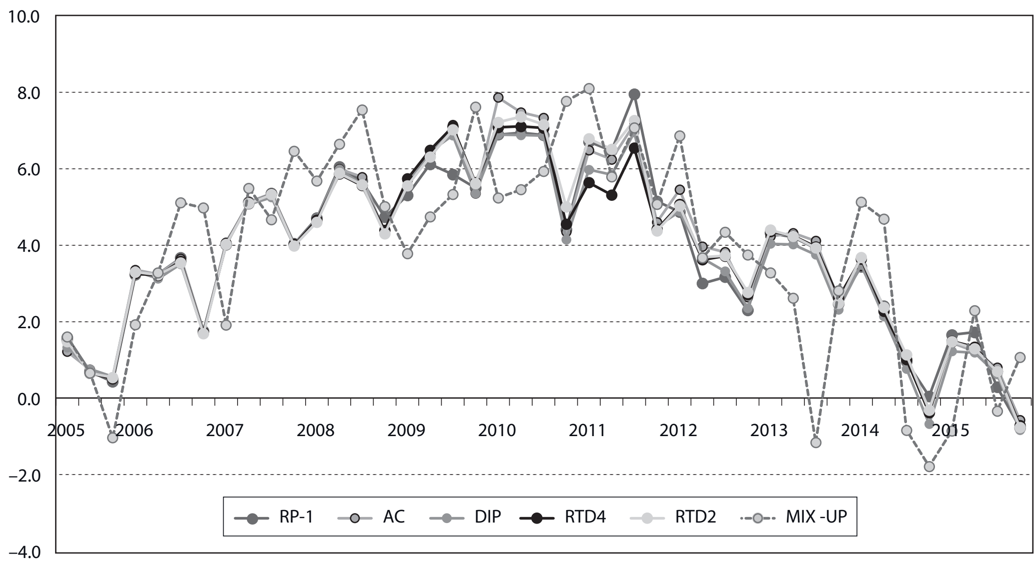
Table 5 Rate of appreciation (%) of apartment prices in Belo Horizonte: 2005–2015

Year	Quarter	RP-1	AC	DIP	RTD4	RTD2	MIX -UP
2005	Q2	1.59	1.22	1.32	1.51	1.48	1.60
2005	Q3	1.39	1.50	1.52	1.31	1.30	1.32
2005	Q4	0.59	0.85	0.92	0.88	0.97	–3.70
2006	Q1	7.86	7.86	7.65	7.73	7.75	8.52
2006	Q2	1.43	1.16	0.95	1.16	1.10	5.40
2006	Q3	1.84	2.04	2.03	1.98	1.90	1.53
2006	Q4	2.00	2.00	2.09	1.99	2.06	8.09
2007	Q1	8.33	8.25	8.25	8.30	8.24	–3.55
2007	Q2	5.14	5.23	4.95	5.08	5.07	12.59
2007	Q3	2.66	2.63	2.63	2.71	2.74	5.59
2007	Q4	4.33	4.27	4.37	4.24	4.15	1.49
2008	Q1	7.19	7.07	7.16	7.09	6.96	10.12
2008	Q2	6.64	6.65	6.44	6.28	6.56	8.50
2008	Q3	3.24	3.65	3.19	3.35	3.26	4.09
2008	Q4	4.34	2.71	3.33	3.57	3.12	2.53
2009	Q1	8.39	10.94	10.32	10.44	10.44	4.73
2009	Q2	5.66	5.59	5.80	5.57	5.49	7.02
2009	Q3	3.55	5.00	4.57	5.37	5.20	4.25
2009	Q4	7.31	5.47	5.70	5.97	6.15	11.70
2010	Q1	9.91	13.31	10.50	9.96	10.36	0.09
2010	Q2	3.68	3.87	4.53	5.45	5.62	4.89
2010	Q3	7.14	5.05	5.66	5.84	5.56	13.23
2010	Q4	2.46	4.13	2.29	2.38	3.85	5.38
2011	Q1	10.66	10.37	10.12	8.78	11.07	5.86
2011	Q2	6.45	4.35	5.27	4.87	4.76	6.13
2011	Q3	6.80	6.87	5.53	6.01	6.06	9.26
2011	Q4	2.25	2.63	2.69	2.38	2.34	0.02
2012	Q1	5.61	6.92	6.47	6.85	6.73	11.67
2012	Q2	1.19	2.40	1.88	1.70	2.04	–0.25
2012	Q3	2.75	2.21	1.68	2.66	2.49	1.97
2012	Q4	2.95	3.31	3.53	3.78	3.79	9.78
2013	Q1	7.30	7.24	6.97	6.67	6.96	–1.59
2013	Q2	2.44	2.46	1.63	2.32	2.01	0.02
2013	Q3	2.10	2.70	2.74	2.94	2.88	–1.91
2013	Q4	2.85	2.56	2.59	2.61	2.56	10.75
2014	Q1	5.39	5.14	5.16	5.38	5.61	6.94
2014	Q2	–0.92	–0.73	–1.23	–1.08	–0.95	–3.14
2014	Q3	–1.46	–1.12	–1.48	–1.14	–1.13	–5.87
2014	Q4	2.57	0.96	0.69	1.09	1.15	3.91
2015	Q1	3.93	4.60	4.55	4.59	4.50	–0.37
2015	Q2	–1.27	–1.78	–1.58	–1.57	–1.66	3.38
2015	Q3	–1.70	–0.32	–1.00	–0.81	–0.64	–3.90
2015	Q4	0.67	0.38	0.09	0.24	0.01	3.91

Source: IPEAD/UFMG, author's calculation

From the first quarter of 2007 to the first quarter of 2014 there was an intense apartment's price appreciation in Belo Horizonte. This situation is in line with Brazilian real estate outlook. This appreciation was contemporary with the expansion rate of housing credit in Brazil. Some institutional improvements like fiduciary alienation law's refinement in 2004 agreed with income growth and the decline interest rates helped the housing credit's growth (Aguiar, 2014). Cardoso and Leal (2009) highlighted the government politics and the restructuring (more market concentration) of real estate development's firm role in the real estate market expansion.

Figure 4 Apartment annual valuation rate (%) – triennial moving geometric average – Belo Horizonte: 2005–2015



Source: IPEAD/UFMG, author's calculation

Figure 4 shows the quarterly variation of Belo Horizonte's apartment prices. It's clear in the figure the great volatility of the median index (MIX-UP), as we have seen in the previous section. From 2005 to 2011 there was a significant housing prices growth path, from then there was a decline tendency. After the second quarter of 2014, the decline was more intense due to the Brazil's economic crises which began in this period.

CONCLUSION

Brazil does not have an official price index yet. In this paper we used a database from Belo Horizonte, a big Brazilian city, to test some hedonic quality adjusted price indices. The indices constructed were the same used by European Statistical Institutes as described by Hill et al. (2018). Our results suggested that the hedonic quality adjusted indices exhibited a good performance in volatility terms. However, some drift in double imputation Laspeyres and Paasche indices was detected. The former with a strong upper-ward bias relative to the other hedonic indices and the latter with a strong down-ward bias.

According to our analyses, the double imputation Tornqvis (DIT) and the repricing with an annual base update (RP-1) produced very similar magnitude's indices. The same could be said about the average characteristics (AC) and the 2-quarter rolling time dummy (RT2). However, the index price lines of the latter stayed above the former. The 4-quarter rolling time window, in its turn, exhibited an intermediate behavior, as compared with the previously listed indices.

In contrast to Hill et al. (2018) estimation for Sydney and Tokyo's evidence, the repricing method with no base update (RP-X) had an upper-ward bias relative to the other hedonic indices. The five years update base repricing method, as well as Hill et al. (2018) empirical evidence, exhibited a down-ward bias. Finally, our models result suggested that median indices are not the most appropriate to estimate housing price indices. The stratified median index (MIX-UP) used was more volatile than the hedonic indices. This is because this kind of index is imperfect in control for housing quality variation over time.

This paper emphasized the potentiality for constructed housing price indices in Brazil using hedonic quality adjusted price methods and for administrative data. The Real Estate Transfer Tax (RETT) emerges as a hopeful database once this tax is collected in the whole country. Further analysis could extend the hedonic price models to estimate price indices for the Brazilian smaller city context and its less frequent housing sales reality. The smaller number of observations imposes new challenges to estimated housing prices hedonic quality adjusted indices. In addition, future analysis could extend the types of real estate units, including houses and different types of commercial real estates.

References

AGUIAR, M. M., SIMÕES, R., GOLGHER, A. B. (2014). Housing Market Analysis using a hierarchical-spatial approach: the case of Belo Horizonte, Minas Gerais, Brazil [online]. *Regional Studies, Regional Science*, 1(1): 116–137. <https://doi.org/10.1080/21681376.2014.934391>

AGUIAR, M. M. (2014). Crédito e mercado imobiliário no Brasil: fatores institucionais e macroeconômicos (1964–2013). *XVI Seminário de Economia Mineria*, Diamantina, Anais, Diamantina: CEDEPLAR/UFMG.

ALBUQUERQUE, P., NADALIN, V., NETO, V., MONTENEGRO, M. (2018). Construção de índices de preços de imóveis para o Distrito Federal por meio de vendas repetidas e GWR [online]. *Nova Economia*, 28(1): 181–212. <https://dx.doi.org/10.1590/0103-6351/3366>.

CARDOSO, A. L., LEAL, J. A. (2009). Housing Markets in Brazil: recent trends and governmental responses to the 2008 crisis. *ISA International Housing Conference*, Glasgow: ISA.

COURT, A. (1939). *Hedonic price indexes with automotive examples*. The Dynamic of Automobile Demand, New York: General Motors, 98–119.

DIEWERT, W. E. (2009). The Paris OECD-IMF Workshop on Real Estate Price Indexes: conclusions and future directions. In: DIEWERT, E. E. et al. (eds.) *Price and Productivity Measurement*, Vol. 1 – Housing, Trafford Press.

GONZÁLEZ, M. A. S. (1997). Variação qualitativa e índice de preços na análise do comportamento recente dos aluguéis em Porto Alegre (1994–1997) [online]. *Análise Econômica*, 15(28). <https://doi.org/10.22456/2176-5456.10548>.

GOODMAN, A., THIBODEAU, T. (1995). Age-related heteroskedasticity in hedonic price equations. *Journal of Housing Research*, 6(1): 25–42.

GRILICHES, Z. (1961). Hedonic Price indexes for automobiles: an econometric of quality change. In: *The Price Statistics of the Federal Government*, General Series No 73, New York: Columbia Univ. Press for the National Bureau of Economic Research, 137–196.

GRILICHES, Z. (1971). Introduction: hedonic price indexes revisited. In: *Prices indexes and quality change: studies in new methods of measurement*, Cambridge: Harvard University Press.

GRILICHES, Z. (1958). The Demand for fertilizer: an econometric reinterpretation of a technical change [online]. *Journal of Farm Economics*, 40(3): 591–606. <https://doi.org/10.2307/1235370>.

HILL, R., MELSER, D. (2008). Hedonic imputation and the price index problem: an application to housing [online]. *Economic Inquiry*, 46(4): 593–609. <https://doi.org/10.1111/j.465-7295.2007.00110.x>.

HILL, R., SCHOLZ, M., SHIMIZU, C., STEURER, M. (2018). An evaluation of the methods countries to compute their official house prices indices [online]. *Economie et Statistique/ Economics and Statistics*, 500–501–502: 221–238. <https://doi.org/10.187/ecostat.2018500t.1953>.

HILL, R. (2004). Constructing price indexes across space and time: the case of the European Union [online]. *American Economic Review*, 4(5): 1379–1410. <https://doi.org/10.1257/0002828043052178>.

HILL, R. (2013). Price indexes for residential housing: a survey, evaluation and taxonomy [online]. *Journal of Economic Surveys*, 27(1): 879–914. <https://doi.org/10.1111/j.1467-6419.2012.00731.x>.

NADALIN, V., FURTADO, B. (2011). *Índices de preços para imóveis: uma revisão.* Boletim Regional, Urbano e Ambiental, 6: 21–32.

PAIXÃO, L. A. R. (2015a). *Índice de preços hedônicos para apartamentos: Belo Horizonte, 1995–2012*. Tese (Doutorado em Economia), Programa de Pós-Graduação em Economia, Universidade Federal do Rio de Janeiro.

PAIXÃO, L. A. R. (2015b). Índice de preços hedônicos para imóveis: uma análise para o município de Belo Horizonte [online]. *Economia Aplicada*, 19(1): 5–29. <https://doi.org/10.1590/1413-8050/ea36708>.
ROSEN, S. (1974). Hedonic price and implicit markets: products differentiation in pure competition [online]. *Journal of Political Economy*, 82(1): 35–55. <https://doi.org/10.2307/1830899>.
ROZENBAUM, S., MACEDO-SOARES, T. D. (2007). Proposta para a construção de um índice de preços de imóveis a partir dos lançamentos imobiliários de condomínios residenciais [online]. *Revista de Administração Pública*, 41(6). <https://doi.org/10.1590/S0034-76122007000600004>.
ROZENBAUM, S. (2009). *IMPA – Índice municipal de preço de apartamentos: proposta e metodologia*. Tese (Doutorado em Administração de Empresas), Programa de Pós-Graduação em Administração de Empresas, Pontifícia Universidade Católica do Rio de Janeiro.
SANTOS, F., SALAZAR, M. (2011). *Índice de preços imobiliários para o Brasil: estudos para discussão*. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística (IBGE).
SIMÕES, P. F. M. (2017). *Modelos longitudinais de efeitos mistos: aplicações ao cálculo de índices de preços de imóveis e à análise da perda não técnica de eletricidade*. Tese (Doutorado em Engenharia Elétrica), Programa de Pós-Graduação em Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.
TRIPLETT, J. (2004). *Handbook on hedonic and quality adjustments in price indexes: Special application to information technology products*. Paris: (OECD), Working Paper 2004/9.
VILLAÇA, F. (1988). *O espaço intraurbano no Brasil*. São Paulo: Studio Nobel.
WAUGH, F. (1928). Quality factors influencing vegetable prices [online]. *Journal of Farm Economics*, 10(2): 185–196. <https://doi.org/10.2307/1230278>.

ANNEX

Table A1 Belo Horizonte's apartment numeric variables

Variable	Description	Expected signal	Meaning
Area	Apartment's internal area	+	Consumers tend to prefer living in higher apartments
Age	Apartment's age	–	Proxy of depreciation
Age^2	Apartment's squared age	+	Repair and improvements reduce the depreciation's age effect. Also there is the vintage effect (Goodman and Thibodeau, 1995), when some old apartments are valued in the real estate market

Source: Own construction

Table A2 Belo Horizonte's apartment quality building finishing material dummy variable

Variable	Description	Expected signal
Luxury	Apartment's with the best kind of quality building finishing material classification	+
High	Apartment's with the second best kind of quality building finishing material classification	+
Normal	Apartment's with the mean kind of quality building finishing material classification	Basical category
Low	Apartment's with the second worst kind of quality building finishing material classification	–
Popular	Apartment's with the worst kind of quality finishing material classification	–

Source: Own construction

Table A3 Belo Horizonte's zoning dummy variables

Variable	Description	Expected signal	Meaning
ZAP	Zona de Adensamento Preferencial (preferential density zone)	Basic Category	Areas where the municipality encourages new bulding
ZPA	Zona de Proteção Ambiental (environmental protection zone)	–	Areas where new buildings are not allowed or encoraged due to natural or topographic conditions. Despite the new buildings restrictions, these areas are not valued in housing markets because of its lack of affordable natural conditions
ZAR	Zona de Adensamento Restrito (restricted density zone)	–	
ZA	Zona Adensada (dense zone)	+	High density areas where the municipality discourages new buildings. Commonly ZA's are in the most valuated Belo Horizonte's districts and it represents supply restrictions in a high demand context
ZE	Zona de Infraestrutura e Equipamentos Urbanos (urban's infrastructure equipment zone)	–	Areas with great urban equipment (like bus stations, cemeteries, waste treatment etc.). ZE represents poorly valued areas by the real estate's agents
ZEIS	Zona de Especial de Interesse Social (special social interest zone)	–	Original spontaneously occupations's areas (like informal slums), which were formalized by municipality. ZEIS represents poorly valued areas by the real estate's agents
ZHIP	Zona Hiper Central (over central zone)	+	Belo Horizonte's downtown. ZHIP represents valued land's location – where firms and families wish to be located
ZCBH	Zona Central de Belo Horizonte (Belo Horizonte's central zone)	+	ZCBH corresponds the districts besides Belo Horizonte downtown, into Contorno Avenue boundary. As ZHIP Its represents a valued land's location – where firms and families wish to be located
ZCVN	Zona Central de Venda Nova (Venda Nova's central zone)	+	Central area of Venda Nova's distanst suburb. ZCVN represents, in a minor magnitude, valued land's location on Belo Horizonte's extremely north location
ZCBA	Zona Central do Barreiro (Barreiro's central zone)	+	Central area of Barreiros's distant suburb. ZCBA represents, in a minor magnitude, valued land's location on Belo Horizonte's extremely west location

Source: Own construction

Table A4 Belo Horizonte's Planning Units (UP) dummy variables*

UP	Belo Horizonte's Regional
Sagrada Família	East
Floresta	East
Pompéia	East
Santa Efigênia	East
Santa Inês	East
Cabana	West
Jardim América	West
Barroca	West
Betânia	West
Buritis	West
Barro Preto	South-Center
Centro	South-Center
Savassi	South-Center
Prudente de Morais	South-Center
Serra	South-Center
São Bento	South-Center
Belvedere	South-Center
Anchieta	South-Center
Glória	Northwest
Padre Eustáquio	Northwest
Camargos	Northwest
PUC	Northwest
Abílio Machado	Northwest
Caiçara	Northwest
Pampulha	Pampulha
Santa Amélia	Pampulha
Ouro Preto	Pampulha
Jaraguá	Pampulha
Castelo	Pampulha
Cachoerinha	Northeast
Concórdia	Northeast
Cristiano Machado	Northeast
São Paulo	Northeast
Planalto	North
São Bernardo	North
Primeiro de Maio	North
Jaqueline/Tupi	North
Barreiro de Baixo	Barreiro
Cardoso	Barreiro
Europa	Venda Nova
Venda Nova	Venda Nova

Note: * Savassi is the basical category
Source: Own construction

Table A5 Hedonic price model for Belo Horizonte's apartment market: 2004–2015

TD regression				
Variable	**Estimated parameter**	**Standard deviation**	**t**	**P-value**
Constant	10.496	0.005	2 042.01	0.000
Area	0.007	0.000	705.04	0.000
Age	–0.018	0.000	–151.19	0.000
Age^2	0.000	0.000	53.89	0.000
Luxury	0.101	0.004	24.39	0.000
High	0.051	0.001	34.59	0.000
Low	–0.067	0.001	–50.69	0.000
Popular	–0.095	0.004	–25.79	0.000
ZPA	–0.054	0.009	–6.21	0.000
ZAR	–0.055	0.002	–33.83	0.000
ZA	0.139	0.002	60.23	0.000
ZHIP	0.072	0.006	12.74	0.000
ZCBH	0.443	0.004	123.34	0.000
ZCVN	0.062	0.020	3.18	0.002
ZCBA	0.200	0.013	15.62	0.000
ZE	–0.178	0.007	–25.5	0.000
ZEIS	–0.232	0.029	–10.61	0.000
Sagrada Família	–0.026	0.004	–6.367	0.000
Floresta	0.017	0.004	4.271	0.000
Pompéia	–0.130	0.007	–18.241	0.000
Santa Efigênia	–0.069	0.005	–14.192	0.000
Santa Inês	–0.090	0.007	–13.025	0.000
Cabana	–0.310	0.007	–47.263	0.000
Jardim América	–0.133	0.004	–35.616	0.000
Barroca	0.020	0.003	6.137	0.000
Betânia	–0.249	0.005	–46.833	0.000
Buritis	0.008	0.003	2.382	0.017
Barro Preto	–0.503	0.006	–81.412	0.000
Centro	–0.154	0.005	–30.452	0.000
Prudente de Morais	0.151	0.004	34.750	0.000
Serra	0.039	0.004	9.146	0.000
São Bento	0.205	0.008	26.220	0.000
Belvedere	0.450	0.010	43.427	0.000
Anchieta	0.141	0.003	41.652	0.000
Glória	–0.364	0.006	–64.096	0.000
Padre Eustáquio	–0.081	0.004	–20.465	0.000
Camargos	–0.323	0.005	–66.197	0.000
PUC	–0.117	0.005	–23.256	0.000
Abílio Machado	–0.244	0.006	–43.292	0.000
Caiçara	–0.062	0.004	–14.130	0.000
Pampulha	–0.186	0.009	–21.060	0.000
Santa Amélia	–0.189	0.004	–43.934	0.000
Ouro Preto	–0.093	0.004	–21.629	0.000
Jaraguá	–0.108	0.004	–26.144	0.000
Castelo	–0.139	0.004	–38.273	0.000
Cachoerinha	–0.233	0.005	–44.255	0.000
Concórdia	–0.202	0.008	–24.445	0.000
Cristiano Machado	–0.022	0.003	–6.313	0.000
São Paulo	–0.301	0.005	–62.567	0.000
Planalto	–0.279	0.005	–51.841	0.000
São Bernardo	–0.302	0.007	–46.206	0.000
Primeiro de Maio	–0.298	0.009	–32.703	0.000
Jaqueline/Tupi	–0.459	0.006	–82.517	0.000
Barreiro de Baixo	–0.348	0.004	–77.391	0.000
Cardoso	–0.498	0.006	–87.897	0.000
Europa	–0.309	0.006	–52.310	0.000
Venda Nova	–0.274	0.004	–63.328	0.000
Time fixed effect = yes				
Ajusted R2	0.924			
F	30 399.1			0.000

Source: IPEAD/UFMG, authors calculation

Determinants of Bilateral Agricultural Trade of SAARC Region: a Gravity Model Approach

Tariq Ahad Nengroo[1] | *University of Kashmir, Srinagar, India*
Imtiyaz Ahmad Shah[2] | *University of Kashmir, Srinagar, India*
Md. Sarafraz Equbal[3] | *University of Kashmir, Srinagar, India*

Received 7.9.2022 (revision received 24.10.2022), **Accepted** (reviewed) 28.11.2022, **Published** 16.6.2023

Abstract

The South Asian Association for Regional Cooperation (SAARC) region is an important player in the world agriculture trade. They have vast potential to strengthen their position in global agricultural trade due to theregion's opportunities to increase agricultural production combined with growing global demand. To discover the SAARC potential of agricultural trade patterns, the present paper examines the determinants of bilateral agricultural exports from 2000 to 2019. The gravity model was estimated by employing the Poisson Pseudo Maximum Likelihood (PMML) technique, including zero trade flows for panel data. The results confirm the positive and significant impact of exporter gross domestic product (GDP), importer GDP, Broder, common language, South Asian Free Trade Area (SAFTA), and India-Sri Lanka Free Trade Agreement (ISFTA) on bilateral agricultural trade inthe SAARC region. On the other hand, distance and development levels significantly negatively impact bilateral agricultural trade. Lastly, the study showed an insignificant impact of the bilateral exchange rate.

Keywords	***DOI***	***JEL code***
SAARC, agricultural trade, SAFTA, gravity model, ISFT	*https://doi.org/10.54694/stat.2022.40*	*C23, F15, F43*

INTRODUCTION

Globalisation has evolved dramatically after World War Second. This trend occurred due to the increased international trade and investment activities (Urata, 2002). This was evident becausethe global trade growth outpaced the global output growth (Feenstra, 1998). In 2018, the international merchandise trade climbed

[1] Department of Economics, University of Kashmir, Hazratbal, Srinagar, 190006, India.
[2] Department of Economics, University of Kashmir, Hazratbal, Srinagar, 190006, India.
[3] Department of Economics, University of Kashmir, Hazratbal, Srinagar, 190006, India. Corresponding author: e-mail: sarfraz.eco@gmail.com.

by 3.0% and the global gross domestic product (GDP) increased by 2.9%. Such a pace of globalisation has gradually increased the trend toward regionalism. Countries worldwide developed regional agreements, whether bilateral, regional, or multinational, to speed up trade and, therefore, their integration into the global economy (Joshi, 2010). Over the past several decades, both developed and developing countries have significantly lowered trade restrictions, bringing a paradigm shift in the international trade patterns. Such a shift in trade patterns has been attributed to how globalisation gained momentum and its impact on speeding up regionalism.

Regionalism is referred to when a group of countries form regional blocks on a regional basis (Carbaugh, 2006). It is being found that both developed and developing countries have shifted towards and formed different regional trading blocs to meet their developmental agendas. Asa result, regional trading agreements worldwide have increased rapidly. Keeping in view the success of major regional trading blocs like encouraged the Indian sub-continent to form a regional trading block, namely The South Asian Association for Regional Cooperation (SAARC), an association of 8 South Asian nations, of which Indiais one of the primary founding members apart from Sri Lanka, Bhutan, Bangladesh, Pakistan, Nepal, Maldives and Afghanistan which joined later in 2008. To promote trade and commerce within the region, these countries entered into a preferential trade agreement SAPTA (South Asian Preferential Trade Agreement) in 1992, which fully came into effect in 1993 as the first level of trade arrangement among SAARC members. Moving ahead for further integration, SAPTA was transformed into a South Asian Free Trade Area (SAFTA) in 2006. One of the most common features of south Asian countries is that they are primarily agrarian in nature. However, over the years,the contribution of the agriculture sector to the GDPs of these countries has declined, but still, a good chunk of the population derives employment from the agriculture sector. Regarding regional trading agreements, the agriculture sector has not been covered under these trading agreements till 2000. However, since the Doha Round of Development in 2001, agriculture has become part of many foreign trade agreements (FTA) negotiations.

Unlike developed countries, the agricultural sector exceeds most of the economic activity in developing countries. Thanks to their structural nature, agriculture contributes to economic development as a continuous process of improving the standard of living of the population. In fact, agriculture is the first economic activity without which life cannot subsist. It is also responsible for the provision of food and clothing for the population of other non-agricultural economies. Likewise, it's capable of supplying a large part of the production materials, such as capital, raw materials and human material for other economic sectors. Many economic indicators and criteria are used to judge the efficiency of the performance of the agricultural sector, which mainly depends on the value of GDP, the volume of production, investments and exports. In this context, agricultural exports are regarded as one of the main means of economic growth and sustainable development of the countries. They are seen as a crucial means of acquiring currency, stimulating agricultural investment, increasing the employment rate, reducing the number of the unemployed and eliminating the poverty rate.

So this backdrop, the present study will try to find out the impact of various determinantson the regional agriculture trade flow among SAARC. Such an analysis will benefit policy-related issues to promote agriculture trade in this region. To the authors' best knowledge, this research is the first attempt to employ the gravity model in determining the major determinants of bilateral agricultural exports of the SAARC region using the PPML econometric technique. For this purpose, a well-known gravity methodology will be employed. Unlike supply-side models such as the Ricardian and Heckscher-Ohlin, the gravity model of trade considers both supply and demand factors (GDP and population) as well as trade resistance factors (geographical distance, trade policies, uncertainty, and various bottlenecks) and trade preference factors (preferential trade agreements, monetary unions, political blocs, common language, and common borders) (Bacchetta et al., 2012; Benedictis and Vicarelli, 2004). Therefore, the research will help to identify the major determinants of bilateral agricultural exports of the SAARC region,which will allow this research to contribute to the existing literature with the necessary information during the decision-making processes of both public and private policymakers.

The present study has two-fold novel contributions to the analysis of bilateral agricultural exports of the SAARC region. The first is due to methodological concerns. In contrast to the previous studies of SAARC agricultural trade data, such as Dembatapitiya (2015), we apply Poisson Pseudo Maximum Likelihood (PPML) to generate our parameter estimates rather than the more conventional Ordinary Least Squares (OLS) technique. PPML has shown fewer bias estimators than conventional OLSas an estimation method of choice. The second novel contribution of the present research is the inclusion of both SAFTA and the India-Sri Lanka Free Trade Agreement (ISFTA) trade agreements as control variables on agricultural trade of the SAARC region.

The paper is structured as follows: current section introduces background, importance and novelty of the theme; first section presents the literature review; second section discusses methodological aspects of the article like the identification of variables through the gravity model, data and data sources and the PPML econometric technique for the gravity model; third section presents empirical results; and the final section concludes the paper with policy implications and future research gaps.

1 LITERATURE REVIEW

This part highlights the review of previous studies that have used the gravity model to examine the fundamental determinants of trade and its potential. International trade flows were studied using the gravity equation for the first time in 1962 by Nobel Prize-winning economist Jan Tinbergen. Using the data covering 18 countries in his first study in 1958, he found that the trade flow between the countries was proportionate to the product of an index of their economic sizes, and the factor of proportionality was dependent on the measures of trade resistance between them. Anderson (1979) was the first who tried to provide the theoretical underpinnings to the gravity equation based on the Armington (1969) assumption. He argued that the nation of origin differentiates goods, consumers have established preferences for all the differentiated goods assumption, and consumers have established preferences for all of the differentiated items. Later it was found that a multitude of international trade theories, such as the Ricardian model, Heckscher–Ohlin model, and new theories of economies of scale, monopolistic competition, and intra-industry trade, can be used to derive the gravitation equation Bergstrand (1985), Helpmann and Krugman (1985), Helpmann (1981), Alan (1995), and Anderson and Van Win coop (2003).

Srinivasan (1994) employed the gravity model to examine the impacts of SAPTA, and they found that smaller countries have more chances to get benefited from SAPTA. Rajapakse and Arunatilleke (1997), while examining the trade between Sri Lanka and its major trading partners through the gravity model approach, found that the abolition of restrictive trade policies has the ability to boost trade potential between Sri Lanka and its trading partners. Examining the trade among south Asian countries, Samaratunga et al. (2001), employing the gravity model, came up with the result that there is a potential for southAsian exports to expand and increase their volume. For the period 1996–2002, Shukar and Hassan (2001) used both panel and cross-sectional data to assess trade creation and trade diversion effects under the current SAFTA system by applying the gravity model to the panel data. According to a study, there was no indication of trade diversion between SAARC countries and other countries. Rahman, Shadat and Das (2006) studied additional regional trading blocs with an extended gravity model. SAPTA was proven to have considerable intra-regional trade creation, according to their findings. In contrast, Rodrguez-Delgado (2007) modified the gravity equation and found SAFTA's trade liberalisation programme to have limited effects on regional trade flows. Dayal et al. (2008) found that estimated trade is much higher than actual trade, indicating a huge potential for intra-regional trading in South Asia.

Using an augmented gravity model with trade costs as an additional variable, Banik and Gilbert (2008) investigated whether the presence of trade costs influences trade flows in the South Asian region. According to the report, South Asia has greater trade costs because of a lack of physical and service-related infrastructure, government regulation, port inefficiency, and corrupt customs officials. Jeevika's (2009) studies have shown that South Asia had only a moderate success in liberalising regional trade because of the leftover trade barriers, the lack of complementary production and consumption, and political friction between countries. This was demonstrated through sectoral gravity models of exports of five product categories related to food and agriculture: livestock, vegetables, processed foods, and manufactured goods. It was concluded in Moinuddin's (2013) study that lowering tariffs and non-tariff obstacles will have a positive impact on intra-bloc trade among South Asian economies.

Recently Kahaer and Buwajian (2020) used the gravity approach to determine the impact of international logisticsfrom the 22 countries of western and central Asia on the agricultural export growth of China during 2012–2019. The results of the study reveal that the population, GDP, mutual memberhood of the member countries and performance of international logistics significantly impacted the country's agricultural export growth. Similarly, González et al. (2018) confirm that population, GDP per capita, real exchange rate, and free trade agreements positively impact Nicaragua's agricultural exports, while distance significantly negative impacts agricultural exports. Fiankor, Haase and Brümmer (2021) applied the translog gravity model to determine the heterogenous effects of food standards on agricultural trade. The study confirms the negative impact of importer standards on agricultural trade flow. By using the two-step system generalised moment methods (two-step sys GMM) on agricultural exports on agricultural exports, Eshetu and Mehari (2020) confirms the GDP, road connectivity, exchange rate, domestic savings, tax revenue, and lagged agricultural exports as a major determinant of agricultural exports of Ethiopia. Bakari and Zidi (2021) found the positive impact of GDP in the agricultural sector, bank loans to the agricultural sector, agricultural imports, and imports of agricultural machinery on agricultural exports in the long run of Tunisia agricultural exports. On the other hand,the exploitation of agricultural land and domestic investment tothe agricultural sector harms Tunisia's agricultural exports.

2 METHODOLOGY AND DATA

The methodology of the paper is as; first, the determinants of bilateral agriculture of the SAARC region are identified using the gravity model of agricultural trade; second, the impact of identified determinants on agricultural exports is estimated using the PPML econometric technique and lastly, the methodology provides the data sources of all interesting variables.

To account for a comprehensive analysis and to examine the determinants of agricultural exports within the south Asian block. We have employed a widely celebrated gravity model technique derived from Newton's universal law of gravitation, which states that trade between two countries is directly related to the GDP of two countries and inversely determined by the distance between the countries. The model was first used by Tinbergen (1962), Pöyhönen (1963), and Pulliainen (1963) to explain bilateral trade. The theoretical model of gravity model is borrowed from Newton's law of gravity which assumes that the attraction between two bodies in the universe is directly proportional to their masses and inversely related to the distance between them. In international trade economics, the gravity model suggests that trade between two countries is proportional to their national incomes and inversely related to the distance between them. Therefore, the gravity model predicts that economically rich and closer countries trade more than developing countries. Braha et al. (2017) suggest that the gravity model efficiently explains a large proportion of international trade. Following Braha et al. (2017), the gravity equation of international trade is expressed as follows:

$$X_{ij,t} = \alpha GDP_{i,t}^{\beta 1} GDP_{j,t}^{\beta 2} Dis_{ij}^{\beta 3}\ ,\ \ \beta_1 > 0,\ \ \beta_2 > 0,\ \ \beta_3 < 0, \tag{1}$$

where $X_{ij,t}$ is the dependent variable representing the bilateral trade flow from country i to country j. The independent variables includes the GDP_i (economic size of the exporter country), GDP_j (economic size of the importer country), and Dis_{ij} (distance between the exporting and importing countries).

Traditionally, the gravity model is estimated through log-linearisation and is estimated through linear estimators like OLS. Taking the logarithm on both sides of Formula (1) as:

$$lnX_{ijt} = \alpha + \beta_1 lnGDP_{it} + \beta_2 lnGDP_{ij} + \ \beta_3 lnDis_{ij} + \varepsilon_{ijt}, \tag{2}$$

where β_1, β_2 and β_3 are elasticity coefficients showing the impact of export's GDP. Importer's GDP and bilateral distance on trade flow between countries.

However, Formula (2) is subject to two econometric issues which have received attention in recent methodological development. First, the original gravity model omits the multilateral resistance terms which are correlated to trade costs. Ignoring the multilateral trade costs could lead to biased estimators. To capture the trade costs a number of control variables are used as a proxy to capture the trade costs that are specific to our theme, including the binary variable, the border between trading countries, a binary variable of common language and others. The augmented gravity model with multilateral resistance variables is shown in Formula (3) as:

$$\begin{aligned} lnX_{ijt} &= \alpha + \beta_1 lnGDP_{it} + \beta_2 lnGDP_{ij} + \ \beta_3 lnDis_{ij} + \beta_4 lnExc_{ij} + \beta_5 lnGDPC_i + \beta_6 lnGDPCj + \beta_7 Lng_{ij} \\ &+ \beta_8 Bor_{ij} + \ \beta_9 SAFTA_{ij} + \ \beta_{10} ISFT_{ij} + \varepsilon_{ijt}. \end{aligned} \tag{3}$$

The second issue is the treatment of zero trade flow years between the trading countries. Ignoring the zero trade years may lead to another source of bias if ignored. The log-linearised model leads to a truncated dependent variable due to the non-existence of a natural log of zero observations. To address this issue, we apply Santos-Silva and Tenreyro's (2006) so-called PPML estimator. Not only does it capture the useful information contained in zero trade flows, but the PPML estimation technique is also a suitable method in the presence of heteroscedasticity. With a solid theoretical foundation and substantial empirical evidence, the PPML approach has been regarded as one of the most effective techniques for calculating gravity equations. The PPML equation of the augmented gravity model is as:

$$\begin{aligned} X_{ijt} &= \alpha + \beta_1 lnGDP_{it} + \beta_2 lnGDP_{ij} + \ \beta_3 lnDis_{ij} + \beta_4 lnExc_{ij} + \beta_5 lnGDPC_i + \beta_6 lnGDPCj + \beta_7 Lng_{ij} \\ &+ \beta_8 Bor_{ij} + \ \beta_9 SAFTA_{ij} + \ \beta_{10} ISFT_{ij} + \varepsilon_{ijt}, \end{aligned} \tag{4}$$

where $GDPC_i$ is the income-effecting variable measured by the GDP per capita of the exporting country, $GDPC_j$ is the income-effecting variable of importing country. Exc_{ij}, presents the bilateral exchange rate. Lng_{ij}, is a dummy variable showing whether exporting and importing country has a common primary language. The dummy variable, Bor_{ij}, if countries *i* and *j* share a common land border. $SAFTA_{ij}$ and $ISFT_{ij}$, are free trade agreements.

The fundamental premise of the model is that trade flow between countries increases with the increase in their GDPs and decreases with the distance between countries. So the GDP of both the countries that is exporting and importing countries is expected to have a positive impact on trade flows.GDP is taken as a proxy for income. The higher the income level, the higher thecountry's productive capacity, which means a greater amount of goods available for export. So, the coefficient of exporter GDP is expected to have a positive sign which means an increase in the goods available for exports. In the same way,a higher level of GDP of an importing country means a higher level of income, which means a higher absorptive capacity for importing country. The coefficient of importer GDP is also expected to have a positive sign (Harris and Matyas, 1998; Rahman, 2005; Jayasinge and Sarker, 2008). The distance, which is expressed in km, is the distance between two economic centres/capitals.

The distance, which is a proxy for transportation and other transaction costs,negatively impacts trade flows between countries. Therefore, it is expected to have a negative sign (Zorzoso and Lehman, 2000; Abraham and Hove, 2005; Rahman, 2005). The border, which we take as a dummy variable between country pairs. Countries which share borders carry the value1 and 0 otherwise. Common border reduces transaction costs, increasing trade volumes and is expected to have a positive sign. Language is another dummy variable expected to influence trade flows positively, therefore its coefficients are expected to have a positive sign. It takes the value 1 if counties have the same language and 0. It captures cultural characteristics and similarities/differences between two countries (Zorzoso and Lehman, 2000; Abraham and Hove, 2005; Rahman, 2005).

The coefficient of the per capita difference variable can be both positive or negative. The positive sign of this variable indicates that the trade pattern follows the H-O theory, which says countries that are similar trade less than those that are not. The negative coefficient of this variable indicates that the trade pattern follows the Linder demand hypothesis, which postulates that countries which are similar tend to trade more than otherwise.

The agricultural gravity trade model is frequently augmented by using the impact of the exchange rate. In our study annual exchange rate is determined by the export's currency units per unit of the importing currency. An increase in the exchange rate would be expected to devaluate the exporter currency, and exports would become cheaper. Therefore, the expected sign of the exchange rate is positive (Hatab et al., 2010).

The effects of trade liberalisation on agricultural exports are observed by using dummy variables of free trade agreements (FTAs). We incorporated two dummies, SAFTA and ISFTA, to observe the impact of free trade agreements in the SAARC region. SAFTA is a 2006 regional trade agreement among SAARC countries, and ISFT is a bilateral free trade agreement between India and Sri Lanka. Economists have debated the impacts of FTAs through trade creation and diversion effects. Trade diversion leads the inefficiency due to trade preferences for higher costs of member countries with FTA. On the other hand, the trade creation benefits of FTAs depend the initial structural conditions of member countries.

In order to estimate the panel gravity equation, the study uses the bilateral agricultural trade of seven SAARC countries, namely Bangladesh, Nepal, Sri Lanka, India, Pakistan, Bhutan and Maldives, for the time 2000–2019. Data on agriculture trade flows were extracted from the UNCOMTRADE database[4] at SITC code (0–12 and 4–27–28 define the agriculture sector under such classification, data in annual terms). Data regarding dummy variables like a common language, common border and distance are obtained from the CEPPI database.[5] SAFTA and *SFT* are self-constructed dummy variables. The data of macroeconomic variables of GDP, GDP per capita and exchange rate of importing and exporting countries are obtained from the World Bank database.[6] The description of variables is presented in Table 1.

3 EMPIRICAL RESULTS

In the presence of zero trade flows and heteroscedastic, OLS estimates will be inconsistent andbiased. To avoid such inconsistency and bias, the study estimates the gravity equation using the PPML and the OLS techniques. So here, we will discuss the results of the PPML estimation technique. We have

---

[4] *<https://comtrade.un.org>.*

[5] *<http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele.asp>.*

[6] *<https://data.worldbank.org>.*

Table 1 Description of variables

Dependent Variable	Description	Expected sign
lnX_{it}	Log of bilateral agricultural exports	
Independent variable		
$lnGDP_{it}$	Log of the gross domestic product of reporter country *i*	+
$lnGDP_{jt}$	Log of the gross domestic product of partner *c* country *ij*	+
$lnGDPC_{it}$	Log of GDP per capita of reporter country *i*	+/–
$lnGDP_{jt}$	Log of GDP per capita of partner country *j*	+/–
$lnDis_{ij}$	Log of Distance between country *i* and *j*	–
$lnExc_{ij}$	Log of relative exchange of ratio of export currency and import currency	–
Bor_{ij}	Dummy variable = 1 if country *i* and *j* have a common border; 0 otherwise	+
Lng_{ij}	Dummy variable = 1 if country *i* and *j* have a common ethnic language; 0 otherwise	+
$ISFT_{ij}$	Dummy variable = 1 if country *i* and *j* are members of ISFT; 0 otherwise	+
$SAFTA_{ij}$	Dummy variable = 1 for the year 2006; 0 otherwise	+/–

Source: Authors´ calculation

estimated two PPML models. First, we include only basic fundamental variables in the model second of PPML, we introduce trade agreements. SAFTA is a regional trade agreement, and ISFT is a bilateral free trade agreement between Sri Lanka and India.

Table 2 Estimates of the Gravity model

	OLS (1)	PPML (1)	PPML (2)
$lnGDP_{it}$	0.582*** (0.259)	0.683*** (0.0333)	0.534*** (0.0342)
$lnGDP_{jt}$	0.444 (0.258)	0.619*** (0.0339)	0.473*** (0.0378)
$lnDis_{ij}$	0.460 (0.887)	0.259* (0.129)	0.230* (0.145)
$lnExc_{ij}$	0.576 (0.701)	0.0381 (0.0929)	0.0489 (0.0727)
$lnGDPC_{it}$	0.0659 (0.675)	0.115 (0.0826)	0.200** (0.0757)
$lnGDPC_{jt}$	2.737*** (0.677)	0.271** (0.0926)	–0.360*** (0.0836)
Bor_{ij}	1.360 (1.792)	0.363* (0.142)	1.368*** (0.136)
Lng_{ij}	1.360 (2.871)	0.363* (0.156)	1.368*** (0.156)
$SAFTA_{ij}$	1.380* (0.666)		0.553*** (0.142)
$ISFT_{ij}$	0.787 (.696)		1.650*** (0.146)
Cons	–23.38* (10.01)	–9.894*** (1.592)	–2.558 (1.816)
N	280	280	280
R-sq	0.713	0.643	0.752

Notes: Standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001.
Source: Authors´ calculation

The PPML (2) confirms that the results of the baseline gravity model are persistent with the theoretical framework. The coefficient of GDP of exporter and importer countries are positive and significant. The coefficient of the exporter GDP suggests that a 1 percent increase in GDP will increase the agricultural trade flow by 0.53 per cent. Results also suggest that trade flow will increase with the increase in importing country's purchasing power. Importer country GDP coefficientsuggests that the increase 1 percent of the trade flow will improve by 0.47 per cent. Results indicate that the SAARC region's agricultural exports proportionally with an increase in the sizeof member countries. As expected, our results confirm that distance significantly negatively impacts agricultural exports. As expected, our results confirm the negative and significant impact of distance on agricultural exports. Increasing distance between the capital city of SAARC member countries proxies higher transport costs.

In addition to the classical variables, we adjust the basic model with the variables of GDP per capita as a proxy for the level of development and relative exchange rate. However, the estimates of the model find the negative and significant impact of the level of development on bilateral agricultural exports. This reason may be due to the fact negative elasticity of agricultural exports. The other variable of the relative exchange rate has an insignificant impact on bilateral agricultural exports. The reason may be that the consumption of agricultural goods is based on the customs and habits of people.

Results of the model, augmented with the effect of the common border and common language, confirm the validity ofthe theoretical foundation of the gravity model. The significant and positive coefficient of these variables depicts that SAARC agricultural exports are strongly influenced by transaction and transportation costs. Indeed results predict higher bilateral agricultural exports between countries that share a common language and broader.

The finding of the study confirms the agricultural export creation impact of trade agreements. SAFTA confirms the significant and positive impact on bilateral agricultural exports of the SAARC region. Many studies found that SAFTA has not led to any trade creation among member nations. However, the present study finds that in the case of agricultural trade, it has significantly improved or led to trade creation among member countries. Also, the bilateral free trade agreement ISFT between India and Srilanka has enhanced agricultural trade between these counties as its coefficient is positive and statistically significant.

CONCLUSION

The study investigates the nature of fundamental determinants of agricultural trade flow among SAARC nations through the gravity model approach. It utilises the econometric approach using PPML for bilateral agricultural exports, involving seven SAARC nations, excluding Afghanistan, for the period 2000–2019. Since most of the SAARC nations are agrarian, it is imperative to analyse the nature of trade flow among SAARC nations. The study reveals that the GDP of both exporting and importing countries positively impacts the agriculture trade within the SAARC region. The results indicate that increase in the region's GDP will enhance interregional agricultural trade, revealing a higher impact of member SAARC countries absorbing potential for enhancing bilateral agricultural trade. From the results, it can also be inferred that geographical distance impedes the trade between costs, resulting in higher trade costs. It indicates that geographical proximity and transport cost costs are the key drivers of agricultural exports. Such an outcome is further supported by the significant positive impact of a common language and common border on the bilateral trade of SAARC members. The results also reveal that SAFTA has been a trade-creating agreement and has improved the intra-regional agricultural trade in the SAARC region. Therefore, the present study prompts a need for deeper integration in South Asia to improve trade relations and tackle the poverty and unemployment this region has been facing for decades. Deeper integration will help in the development of better infrastructure, higher productivity and sustainable growth in GDP. The nature of agricultural commodities is such that they are mostly perishable and less durable. For such trade, distance plays an important role, and, therefore, steps must be taken to increase the connectivity channels and invest in roadinfrastructure is important in this region. Further special attention should be given to the competitiveness of farmers. Public investment in irrigation should be accompanied by direct support to farmers.

While our study examined the determinants of bilateral agricultural trade in the SAARC region beyond the traditional gravity model, it would be worthwhile to raise some of the limitations of the present study that should be incorporated into further research to improve our understanding of the central theme. Some of the issues are: first future research should use the sectoral GDP rather than the overall national income, second other factors, including tariffs, geopolitical concerns, import substitution policy and pricing that influence the bilateral trade in the SAARC region should be included in future research, thirdly, the authors recommend the future studies with a larger dataset about these variables and also comparison between different regions for better results and fewer errors. However, from this research point of view, it has some interesting findings that can help policymakers achieve a better view of the SAARC region's bilateral agricultural trade.

References

ABRAHAM, F., VAN HOVE, J. (2005). The rise of China: Prospects of regional trade policy. *Review of World Economics (Weltwirtschaftliches Archiv)*, 141(3): 486–509.

ABU HATAB, A., ROMSTAD, E., HUO, X. (2010). Determinants of Egyptian agricultural exports: a gravity model approach [online]. *Modern Economy*, 3: 134–143. <http://doi.org/10.4236/me.2010.13015>.

ANDERSON, J. (1979). A theoretical foundation for gravity equation. *American Economic Review*, 69: 106–16.

ANDERSON, J., WINCOOP, E. (2003). Gravity with gravitas: a solution to the border puzzle [online]. *American Economic Review*, 93: 170–192. <http://doi.org/10.1257/000282803321455214>.

ARMINGTON, P. S. (1969). A theory of demand for products distinguished by place of production. *International Monetary Fund Staff Papers,* 16: 159–178.

BACCHETTA, M., BEVERELLI, C., CADOT, O. FUGAZZA, M. GRETHER, J.-M., HELBLE, M. NICITA, A., PIERMARTINI, R. (2012). *A practical guide to trade policy analysis*. Geneva: World Trade Organization.

BAKARI, S., ZIDI, A. (2021). The determinants of agricultural exports: empirical validation for the case of Tunisia. *International Journal of Food and Agricultural Economics,* 9(3): 263–285.

BANIK, N., GILBERT, J. (2008). Trade cost and regional integration in South Asia. *Asian Development Bank Institute Working Paper*, 127.

BENEDICTIS, L. D., VICARELLI, C. (2004). Trade potentials in gravity panel data models. *ISAE Working Papers No 44*, Rome: ISTAT.

BERGSTRAND, J. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence [online]. *The Review of Economics and Statistics*, 67: 474–81. <https://doi.org/10.2307/1925976>.

BRAHA, K., QINETI, A., CUPÁK, A., LAZORCAKOVA, E. (2017). Determinants of Albanian agricultural export: the gravity model approach. *Agris on-line Papers in Economics and Informatics*, 8(1): 3–21.

CARBAUGH, R. J. (2012). *International Economics*. 14th Ed. Cengage Learning.

DAYAL, P., DAS, A. BANGA, R., IYENGAR, K., AHMED, S. (2008). *Qualification of benefits for economic cooperation in South Asia.* Asian Development Bank (ADB) and United Nations Conference on Trade and Development (UNCTAD), 4–75.

DEARDORFF, A. V. (1995). Determinants of bilateral trade: does gravity work in a neoclassical world? [online]. *NBER Working Papers 5377*, National Bureau of Economic Research, Inc. <https://ideas.repec.org/p/nbr/nberwo/5377.html>.

DEMBATAPITIYA, P. (2015). An analysis of determinants of agriculture trade intensities of South Asia. *Sri Lankan Journal of Agricultural Economics,* 14(1): 59–81.

ESHETU, F., MEHARE, A. (2020). Determinants of Ethiopian agricultural exports: a dynamic panel data analysis [online]. *Review of Market Integration*, 12(1–2): 70–94. <https://doi.org/10.1177/0974929220969272>.

FEENSTRA, R. C. (1998). Integration of trade and disintegration of production in the global economy [online]. *Journal of Economic Perspectives*, 12(4): 31–50. <https://doi.org/10.1257/jep.12.4.31>.

FIANKOR, D.-D. D., HAASE, O.-K., BRÜMMER, B. (2021), The heterogeneous effects of standards on agricultural trade flows [online]. *J, Agric. Econ.*, 72(1): 25–46. <https://doi.org/10.1111/1477-9552.12405>.

FRANKEL, J. A. (1997). *Regional trading blocs in the world economic system* [online]. Washington, D.C.: Institute for international economics. <http://dx.doi.org/10.7208/chicago/9780226260228.001.0001>.

GONZÁLEZ, M., CAMPANO, C., LÓPEZ, L. (2018). A gravity model of trade for Nicaraguan agricultural exports. *Cuadernos de Economía*, 37(1): 391–428.

HARRIS, M. N., MATYAS, L. (1998), The Econometrics of gravity models. *Melbourne Institute Working Paper,* No 5.
HELPMAN, E. (1981). International trade in the presence of product differentiation, economies of scale and monopolistic competition: a Chamberlin-Heckscher-Ohlin approach. *Journal of International Economics,* 11: 305–40.
HELPMAN, E., KRUGMAN, P. R. (1985). *Market structure and foreign trade. Increasing returns, imperfect competition, and the international economy.* Cambridge, MA: MIT Press.
JAYASINGHE, S., SARKER, R. (2008). Effects of regional trade agreements on trade in agrifood products: Evidence from gravity modeling using disaggregated data. *Review of Agricultural Economics*, 30: 61–81.
JOSHI, V. (2010). An econometric analysis of India-Sri Lanka free trade agreement. *IHEID Working Papers No 4,* The Graduate Institute of International Studies.
KAHAER, A., BUWAJIAN, A. (2021) An analysis of gravity model based on the impact of China's agricultural exports – a case study of Western and Central Asia along the economic corridor [online]. *Acta Agriculturae Scandinavica, Soil & Plant Science*, 71(6): 432–442. <https://doi.org/10.1080/09064710.2021.1910725>.
MOINUDDIN, M. (2017). Fulfilling the promises of South Asian integration: a gravity estimation [online]. *ADBI Working Paper 415*. <http://dx.doi.org/10.2139/ssrn.2246563>.
PÖYHÖNEN, P. (1963). A tentative model for the volume of trade between countries. *Weltwirtschaftliches Archive*, 90: 93–100.
PULLIAINEN, K. (1963). A world trade study: an econometric model of the pattern of the commodity flows in international trade in 1948–1960. *Ekonomiska Samfundets Tidskrift,* 16: 69–77.
RAHMAN, M. (2005). The determinants of Bangladesh's trade: Evidence from the generalised gravity model. *Working Paper No 3*, University of Sydney.
RAHMAN, M., SHADAT, W., D., NARAYAN, C. (2006). Trade potential in SAFTA: an application of augmented gravity model. *CPD Working Paper No 61*, Centre for Policy Dialogue (CPD).
RAJAPAKSE, P., ARUNATILAKE, N. (1996). Would a reduction in trade barriers promote intra-SAARC trade? *Institute of Policy Studies,* Colombo, 1–2.
RODRÍGUEZ-DELGADO, J. D. (2007). Safta: Living in a world of regional trade agreements [online]. *IMF Working Papers No 2007/023*, International Monetary Fund. <https://EconPapers.repec.org/RePEc:imf:imfwpa:2007/023>.
SAMARATUNGA, R. H. S, JAYASURIYA, S., PRENTICE, D. (2001). Impact of the enlargement of APEC on South Asia: an analysis of trade potential [online]. *South Asia Economic Journal*, 2(2): 171–201. <https://doi.org/10.1177/139156140100200201>.
SANTOS SILVA, M, C. J., TENREYRO, S. (2006). The log of gravity [online]. *The Review of Economics and Statistics,* 88(4): 641–658. <https://doi.org/10.1162/rest.88.4.641>.
SRINIVASAN, T. N. (1994). *Regional trading arrangements and beyond: exploring some option for South-Asia: theory, empirics and policy*. Report No IDP-142, South Asia Region, World Bank, 1–40.
SUKAR, A.-H., SEID, H. (2001). US exports and time-varying volatility of real exchange rate [online]. *Global Finance Journal*, 12: 109–119. <https://doi.org/10.1016/S1044-0283(01)00020-5>.
TINBERGEN, J. (1962). *Shaping the world economy, suggestion for an international economic policy*. New York: Twentieth Century Fund.
URATA, S. (2002). Globalisation and the growth in free trade agreements [online]. *Asia-Pacific Review*, 9(1): 20–32. <https://doi.org/10.1080/13439000220141569>.
WEERAHEWA, J. (2009). *Impact of trade facilitation measures and regional trade agreements on food and agricultural tarde in South Asia*. Working Papers 6909, Asia-Pacific research and training network on trade (ARTNET), an initiative of UNESCAP and IDRC.
ZARZOSO, I. M., LEHMAN, F. N. (2000). Augmented gravity model: an empirical application to MERCOSUR-European Union trade flows. *Journal of Applied Economics,* 6(2): 291–316.

Estimating Conditional Event Probabilities with Mixed Regressors: a Weighted Nearest Neighbour Approach

Mahfuza Khatun[1] | *Jahangirnagar University, Dhaka, Bangladesh*
Sikandar Siddiqui[2] | *Deloitte Audit Analytics, Frankfurt, Germany*

Received 29.9.2022, **Accepted** (reviewed) 21.11.2022**, Published** 16.6.2023

Abstract

The k-Nearest Neighbour method is a popular nonparametric technique for solving classification and regression problems without having to make potentially restrictive a priori assumptions about the functional form of the statistical relationship under investigation. The purpose of this paper was to demonstrate that the scope of this method can be extended in a way that enables the simultaneous consideration of continuous, ordered discrete, and unordered discrete explanatory variables. An exemplary application to a publicly available dataset demonstrated the feasibility of the proposed approach.

Keywords	***DOI***	***JEL code***
Qualitative choice models, nonparametric estimation	*https://doi.org/10.54694/stat.2022.45*	*C14, C25*

INTRODUCTION

The k-nearest neighbour method is a common statistical technique for solving classification problems (see, e.g., Fix and Hodges, 1951). In the following, the symbol Y either represents either a scalar, discrete random variable or a vector of random variables indicating the class to which a given statistical entity belongs. Moreover, the quantity Q represents a sample vector of explanatory variables to be used as inputs for the classification procedure. A sample consisting of N independent statistical entities $i = 1, \ldots, N$, each of which is characterized by a pair of realisations $\{y_i; q_i\}$ of Y and Q is also assumed to be given. Then the k-nearest neighbour procedure essentially consists of approximating the function representing the conditional expectation of Y for a given value Q^* of Q by:

- either calculating a locally weighted average of the y_i values of those entities where the associated realisations q_i of Q are among the k closest neighbours of Q^*,

[1] Jahangirnagar University, Savar, 1342 Dhaka, Bangladesh.
[2] Deloitte Audit Analytics GmbH, Europa-Allee 91, 60486 Frankfurt, Germany. Corresponding author: e-mail: siddiqui@web.de, phone: (+49)1718881667.

- or performing a locally weighted linear regression of Y on Q, in which only those entities in the sample are assigned a positive weight if they are among the k closest neighbours of Q^*.

In doing so, each of the k closest neighbours can either be assigned the same weight, or each of them separately can be assigned a weight that declines with increasing distance from Q^*. Both of these approaches have been nicely summarized, for example, by Chen, Härdle and Schulz (2004), and Cleveland and Loader (1995).

Typically, the k-nearest neighbour techniques requires all the conditioning variables included in Q to be continuous. This condition might easily be seen as too restrictive when there are either ordered discrete variables (like, e.g. school grades or credit ratings), or unordered discrete variables (e.g. information on the sectoral or geographical affiliation of a company). Hence, the need may arise to introduce a more general measure of the distance between two statistical entities in which all of these variable types can be included in an intuitively plausible manner. This is the main purpose of this paper. It hence proceeds as follows: Section 1 seeks to clarify how the distance between two statistical entities can be calculated if each of them is characterised by a specific realisation of such a mixture of continuous, ordered discrete and unordered discrete variables. The application of the proposed procedure in the context of binomial or multinomial classification problems is described in Section 2. This is also where a measure of the out-of-sample predictive accuracy and a selection rule for the number of neighbours, k, is given. Section 3 presents an application. Last section concludes.

1 MEASURING THE DISTANCE BETWEEN PAIRS OF ENTITIES

1.1 Starting point

Let there be a sample of N entities (e.g. companies) $i = 1, \ldots, N$, each of which is characterised by a tuple of characteristics $q_i = \{ x_i, v_i, z_i \}$. Here,

- x_i is an entity-specific realisation of a $(K_1 \times 1)$ vector X of continuous variables, (e.g. the total revenue of a firm in a given year, or its total assets at a given point in time, etc.),
- v_i is an entity-specific realisation of a $((K_2 - K_1) \times 1)$ vector V of ordered discrete variables (like, e.g., a firm's credit rating, or an analyst recommendation [“Strong Buy”, “Buy”, “Hold” or “Sell”]), numbered consecutively in steps of 1, and
- z_i is an entity-specific realisation of a $((K_3 - (K_2 + K_1)) \times 1)$ vector Z of unordered, discrete variables (like, e.g., a firm's ISO country-of-residence identifier or its NAICS industry classification code).

1.2 Pairwise distance based on continuous characteristics

The distance between two distinct entities i and j with respect to the values taken by X can be measured by Gower's (1971) measure:

$$d_x(i,j) = \sum_{s=1}^{K_1} \frac{\left| x_{i,s} - x_{j,s} \right|}{r(X_s)}, \tag{1}$$

where $x_{i,s}$ denotes the s'th element of the $(K_1 \times 1)$ vector x_i, and $r(X_s)$ is the range (sample maximum minus sample minimum) of the values of X_s observed.

In principle, the above Formula (1) is only one out of several distance measures that could be applied to measure the degree of dissimilarity prevailing between realisations of continuous random vectors; see, e.g., Weller-Fahy, Borghetti, and Sodemann (2015) for a theoretically substantiated overview. The motivation for choosing this particular measure is that it facilitates the aggregation of distance measures for the different variable types involved, as will be demonstrated below.

1.3 Pairwise distance based on ordered discrete characteristics

Likewise, the distance between i and j with respect to the values taken by the components of V can be measured by:

$$d_v(i,j) = \sum_{s=K_1+1}^{K_1+K_2} \frac{\left|v_{i,s} - v_{j,s}\right|}{m_s}, \tag{2}$$

where m_s is the number of possible realisations of the s-th ordered discrete variable.

A rationale for this way of proceeding can be given as follows: Assume that that variable V_s is a rating grade, measured in equally spaced steps of unit length from 1 (= best possible outcome) to m_s (= worst possible outcome). Then, the actual realisation taken by V_s can be assumed to be dependent on the value taken by a latent (=unobservable) variable as follows:

$$v_{i,s} = j \text{ if } v^*_{i,s} \in \left]\tfrac{j-1}{m_s}; \tfrac{j}{m_s}\right] \text{ with } j \in \{1, 2, ..., m_s\}. \tag{3}$$

The distance between the mid-points of two neighbouring intervals inside the range of is V_s^* is $1/m_s$.

1.4 Pairwise distance based on unordered discrete characteristics

Finally, the distance between i and j with respect to the values taken by the components of Z can be measured by their separateness, or lack of overlap (see, e.g., Stanfill and Waltz, 1986), based on the Hamming Distance (Hamming, 1950):

$$d_z(i,j) = \sum_{s=K_1+K_2+1}^{K_1+K_2+K_3} \frac{I\left(z_{i,s} \neq z_{j,s}\right)}{m_s}. \tag{4}$$

In Formula (4), $I(\cdot)$ denotes an indicator function that takes the value 1 if the condition in brackets holds true, and is set to zero otherwise, and m_s is the number of distinct possible realisations of the s-th unordered discrete variable.

The rationale underlying this way of proceeding can be explained as follows: the m_s possible, yet mutually exclusive, realisations of a single, discrete unordered variable of the Z_s can be recoded as a vector of m_s auxiliary, binary variables A_1 to $A_{m(s)}$. This technique of representing unordered categorical data, which is referred to as one-hot encoding, is explained in an exemplary manner in Table 1 (for the special case of $m_s = 4$).

Table 1 Representing unordered categorical data via one-hot encoding

	Values taken by the auxiliary variables A_1 to $A_{m(s)}$			
Variables taken by original variable *Z*	A_1	A_2	A_3	A_4
'Tinker'	1	0	0	0
'Tailor'	0	1	0	0
'Soldier'	0	0	1	0
'Spy'	0	0	0	1

Source: Own construction

The distance prevailing between two specific realisations $z_{i,s}$ and $z_{j,s}$ of Z_s with respect to the values $a_{i,g}$ and $a_{j,g}$ taken by the p-th of the m_s auxiliary variables can thus be measured using the same latent variable approach as sketched in (3), which comes down to:

$$\textit{Distance between } z_{i,s} \textit{ and } z_{j,s} \textit{ with regard to } A_g = \frac{I\left(a_{i,p} \neq a_{j,p}\right)}{2}. \tag{5}$$

The overall distance between $z_{i,s}$ and $z_{j,s}$ can then be calculated as the average of the distances (5), calculated across all values 1, …, m_s of the auxiliary variable index g:

$$\begin{aligned} &\textit{Distance between } z_{i,s} \textit{ and } z_{j,s} \\ &= \left(\frac{1}{m_s}\right) \cdot \sum_{g=1}^{m_s} \textit{Distance between } z_{i,s} \textit{ and } z_{j,s} \textit{ with regard to } A_g, \\ &= \left(\frac{1}{m_s}\right) \cdot \sum_{g=1}^{m_s} \frac{I\left(a_{i,g} \neq a_{j,g}\right)}{2}, \\ &= \left(\frac{1}{m_s}\right) I\left(z_{i,s} \neq z_{j,s}\right). \end{aligned} \tag{6}$$

The above normalisation will cause the distance between two different realisations of an unordered discrete variable (e.g. country of residence) to take the same value as the distance between two distinct yet immediately adjacent realisations of an ordered discrete variable having the same number of possible realisations. Using m_s in the denominator of Formula (4) will cause the average perceived distance between pairs of observations with different values of Z_s to shrink as the number of different realisations of this variable increases. This is well in line with intuition because *ceteris paribus*, the more fine-grained a classification scheme with discrete, unordered categories becomes, the greater the average pairwise similarity between two entities assigned to different classes will tend to become.

1.5 Overall pairwise distance between two entities

The overall distance score between two entities i and j can then be calculated by summing up the distance measures for the different types of variables given in (1), (2), and (4):

$$d(i,j) := d_x(i,j) + d_v(i,j) + d_z(i,j)\,. \tag{7}$$

2 ESTIMATION PROCEDURE AND GOODNESS-OF-FIT ASSESSMENT

2.1 Assigning estimated conditional probabilities to possible outcomes

The following setting applies to a situation where we have a discrete dependent variable Y with a finite number of possible values $g = \{1, 2, \dots G\}$ (which may, but need not, be ordered). Then, if we have an observed or hypothetical entity characterized by the tuple $q^* = \{\, x^*, v^*, z^* \,\}$, we can estimate the conditional probability $\Pr(Y = g \mid Q = q^*)$ using the k-nearest neighbour principle by applying the following sequence of steps:

(i) For each entity i in the sample, calculate the distance score with respect to the entity under consideration as:

$$d\left(i, *\right) = \sum_{s=1}^{K_1} \frac{\left|x_{i,s} - x_s^*\right|}{r\left(x_s\right)} + \sum_{s=K_1+1}^{K_1+K_2} \frac{\left|v_{i,s} - v_s^*\right|}{m_s} + \sum_{s=K_1+K_2+1}^{K_1+K_2+K_3} \frac{I\left(z_{i,s} \neq z_s^*\right)}{m_s}. \tag{8}$$

(ii) Sort the object-specific dissimilarity scores obtained in (i) in ascending order. Let $\tilde{d}_k(*)$ denote the k-th score in this ascending sequence, and $\tilde{d}_{k+1}(*)$ denote the smallest value of $d(i,*)$ that exceeds $\tilde{d}_k(*)$.

(iii) Assign to each object in the sample a weighing factor $\tilde{w}(i,*)$ that is calculated as follows:

$$\tilde{w}(i,*) := \begin{cases} 1-\left(\dfrac{d(i,*)}{\tilde{d}_{k+1}(*)}\right) if\ d(i,*) < \tilde{d}_{k+1}(*). \\ 0\ otherwise \end{cases} \tag{9}$$

(iv) Transform the outcomes of step (iii) into a normalized set of weighting factors $w(i,*)$ that sum up to 1:

$$w(i,*) = \frac{\tilde{w}(i,*)}{\sum_{j=1}^{N}\tilde{w}(j,*)}. \tag{10}$$

(v) Let , $\check{p}(g\,|\,q^*) := \max\left\{\min\left\{x'\hat{\beta};1\right\};0\right\}$, (11)

where $\hat{\beta}^* := argmin_{\tilde{\beta}} \sum_{i=1}^{N} w(i,*)\left[I(y_i = g) - x_i'\tilde{\beta}\right]^2$ denotes the locally linear least squares estimated of the conditional probability of $(Y = g)$ given $(q = q^*)$. If the dependent variable Y has only two possible values (i.e. $G = 2$, and $g \in \{1; 2\}$), the concluding estimates of the related conditional probabilities can be set to $\widehat{Pr}(Y=1 \mid q=q^*) = \check{p}(1\,|\,q^*)$ and $\widehat{Pr}(Y=2 \mid q=q^*) = 1 - \check{p}(1\,|\,q^*)$, respectively.

(vi) Whenever the dependent variable Y has more than two possible values (i.e. $G > 2$), it cannot be taken for granted that the estimated conditional probabilities $\check{p}(g\,|\,q^*)$ from step (v) sum up to unity. In this case, a normalisation needs to be applied to these provisional estimates, which amounts to setting the concluding estimate of $Pr(Y = g \mid q = q^*)$ to:

$$\widehat{Pr}(Y=g \mid q=q^*) = \frac{\check{p}(g\,|\,q^*)}{\sum_{c=1}^{G}\check{p}(c\,|\,q^*)}. \tag{12}$$

2.2 Out-of-sample goodness of fit assessment and choice of the number of neighbours

In order to assess the predictive reliability of the proposed method, and to choose a favourable value of number of neighbours, *k,* to be used, the leave-one-out log likelihood function can be used. It can be calculated as follows:

(1) For each observation in the sample, and for each of the possible values of *g,* calculate the quantity $\hat{p}_{-i}(g\,|\,q_i)$, i.e. the estimated conditional probability:

$$\widehat{Pr}(Y=g \mid q=q_i),$$

that is obtained by applying the above procedure to a sample from whichthe observation with index *i* has deliberately been omitted.

(2) Calculate the leave-one-out log likelihood function as:

$$\sum_{i=1}^{N}\sum_{g=1}^{G} I\left(y_i = g\right) \cdot \ln(\hat{p}_{-i}\left(g \mid q_i\right)).$$

The optimum value of k could then be equated with the one that maximizes the quantity given under point (2) above.

3 APPLICATION

3.1 Data

The dataset chosen for our sample application is the Automobile Data Set from 1985 Ward's Automotive Yearbook, donated by Jeffrey C. Schlimmer[3] and freely available online,[4] where also more detailed information on the informational content of the variables of interest can be found. It contains $N = 205$ rows and 26 columns, the first of which is a relative risk score compared to equally priced vehicles. Theoretically, it ranges from (–3 = very safe) to (3 = very risky) in steps of 1, but the best score found in the sample was (–2). From the remaining 25 columns, the 20 summarized in Table 1 were used as explanatory variables. Ten entities were removed from the dataset used for estimation due to missing values.

Table 2 1985 Automobile Data Set: variables in use

Name	Description
Risk score	Dependent variable, ranging frorm –3 (safest) to 3 (most risky) in steps of 1
Make	Brand name of manufacturer
Fuel-type	Diesel or gas
Aspiration	Standard or turbo
Num-of-doors	Two or four
Drive-wheels	4wd, front-wheel or rear-while drive
Engine-location	Front or rear
Wheel-base	Continuous
Length	Continuous
Width	Continuous
Height	Continuous
Curb-weight	Continuous
Num-of-cylinders	Integer
Engine-size	Continuous
Bore	Continuous
Stroke	Continuous
Compression-ratio	Continuous
Horsepower	Continuous
Peak-rpm	Continuous
City-mpg	Continuous
Highway-mpg	Continuous

Source: Automobile Data Set from 1985 Ward's Automotive Yearbook, donated by Jeffrey C. Schlimmer, available at: <*https://archive.ics.uci.edu/ml/datasets/automobile*>

[3] Jeffrey.Schlimmer@a.gp.cs.cmu.edu.

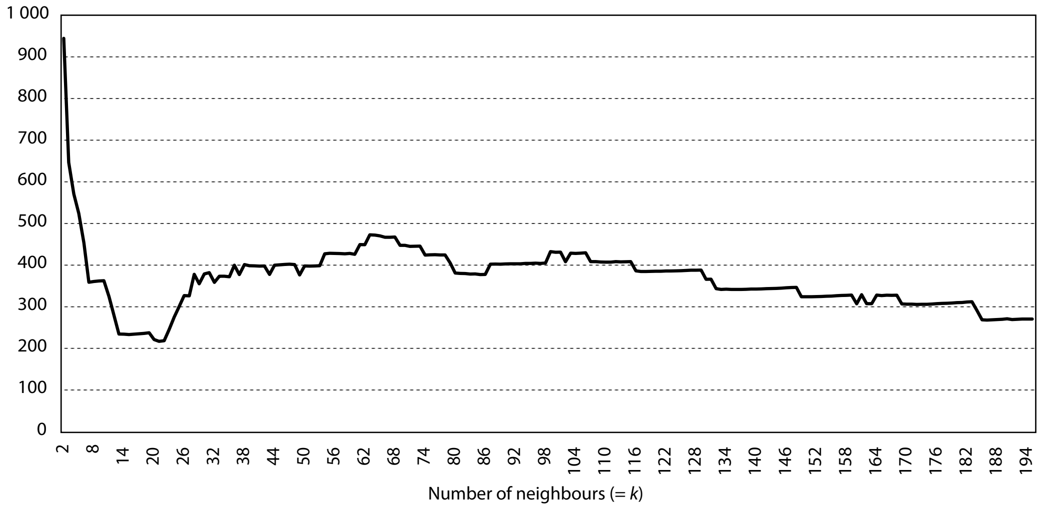
[4] <*https://archive.ics.uci.edu/ml/datasets/automobile*>.

In this example, only continuous, integer and unordered discrete explanatory variables have been used, but it should nevertheless be suitable for demonstrating the general usefulness of the proposed way of proceeding.

3.2 Results obtained with the for the *k*-Nearest Neighbour procedure

In order to avoid overfitting, the leave-one out approach sketched in Section 2.2 was employed. A related grid search procedure, the results of which are given in Figure 1, yielded an optimal value of *k* equal to 21.

Figure 1 Negative log-likelihood function for leave-one-out estimates for different values of *k*



Source: Own calculation

Assuming that the predicted value of the dependent variable is always the one to which the model has assigned the highest conditional probability, the comparison of actual and predicted values results from applying the *k*-Nearest Neighbour Procedure with $k = 21$ to the entire dataset (Table 3).

Table 3 *k*-Nearest Neighbour estimation results: actual vs. predicted outcomes

		Predicted						
	Indicator	**–2**	**–1**	**0**	**1**	**2**	**3**	**Σ**
Actual	–2	0.0000%	1.5385%	0.0000%	0.0000%	0.0000%	0.0000%	**1.5385%**
	–1	0.0000%	7.6923%	3.0769%	0.5128%	0.0000%	0.0000%	**11.2821%**
	0	0.0000%	0.5128%	31.2821%	1.0256%	0.0000%	0.0000%	**32.8205%**
	1	0.0000%	0.5128%	1.5385%	23.5897%	0.5128%	0.5128%	**26.6667%**
	2	0.0000%	0.0000%	1.0256%	3.0769%	11.7949%	0.0000%	**15.8974%**
	3	0.0000%	0.0000%	0.0000%	1.0256%	0.5128%	10.2564%	**11.7949%**
	Σ	**0.0000%**	**10.2564%**	**36.9231%**	**29.2308%**	**12.8205%**	**10.7692%**	
% correctly predicted = 84.6154%								

Source: Own calculation

The geometric mean of model-generated ex-post probability estimates for the actually observed values of Y, which is calculated as

$$\prod_{i=1}^{N}\prod_{g=1}^{G}\widehat{Pr}(Y=g \mid Q=q_i)^{I(y_i=g)}$$

equals 56.31%.

3.3 Comparison with an Ordered Probit Model

In order to assess the predictive performance of the proposed approach, we compared the nonparametric model advocated here to the more commonly used, parametric Ordered Probit model (see, e.g. Greene, 2003, chapter 21.8). Results obtained for the Ordered Probit are given in Table 4.

Table 4 Ordered Probit Model: actual vs. predicted outcomes

		Predicted						
	Indicator	**–2**	**–1**	**0**	**1**	**2**	**3**	**Σ**
Actual	–2	0.5128%	1.0256%	0.0000%	0.0000%	0.0000%	0.0000%	**1.5385%**
	–1	0.5128%	7.6923%	3.0769%	0.0000%	0.0000%	0.0000%	**11.2821%**
	0	0.0000%	0.5128%	29.2308%	3.0769%	0.0000%	0.0000%	**32.8205%**
	1	0.0000%	0.0000%	4.1026%	20.0000%	1.0256%	1.5385%	**26.6667%**
	2	0.0000%	0.0000%	0.5128%	5.6410%	6.1539%	3.5897%	**15.8974%**
	3	0.0000%	0.0000%	0.0000%	1.5385%	1.5385%	8.7180%	**11.7949%**
	Σ	**1.0256%**	**9.2308%**	**36.9231%**	**30.2564%**	**8.7180%**	**13.8462%**	
% correctly predicted = 72.3077%								

Source: Own calculation

In this case, we obtain a geometric mean of model-generated ex-post probability estimates for the actually observed values of Y that amounts to 50.41%.

CONCLUSION

The key advantage that nonparametric classification methods have over parametric models is that they do not require any assumptions about the form of the function that translates the values of the explanatory variables into conditional probabilities of the possible outcomes. Against this background, the purpose of this paper was to demonstrate that the scope of the standard k-Nearest Neighbour method can be extended in a way that enables the simultaneous consideration of continuous, ordered discrete, and unordered discrete explanatory variables. An exemplary application to a publicly available dataset demonstrated the feasibility of the proposed approach.

ACKNOWLEDGMENT

We are indebted to two anonymous referees and this journal's managing editor, whose helpful comments have greatly improved this paper.

The project underlying this publication was funded by the German Federal Ministry of Economic Affairs and Climate Action under project funding reference number 01MK21002G.

References

ALTMAN, N. S. (1991). *An Introduction to Kernel and Nearest Neighbor Nonparametric Regression*. Cornell University, Ithaca, NY14853: Biometrics Unit.

CHEN, Y., HÄRDLE, W. K., SCHULZ, R. (2004). *Prognose mit nichtparametrischen Verfahren*. Humboldt-Universität Berlin: Center for Applied Statistics and Economics.

CLEVELAND, W. S., LOADER, C. (1995). Smoothing by local regression: Principles and methods. In: HÄRDLE, W. K., SCHIMEK, M. G. (eds.) *Statistical Theory and Computational Aspects of Smoothing*, Proceedings of the COMPSTAT '94 Satellite Meeting held in Semmering, Austria, 27–28 August: 10–49.

FIX, E., HODGES, J. L. (1951). *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*. Randolph Field, Texas: USAF School of Aviation Medicine.

GOWER, J. C. (1971). A general coefficient of similarity and some of its properties [online]. *Biometrics,* 27: 623–637. <https://doi.org/10.2307/2528823>.

GREENE, W. H. (2005). *Econometric Analysis*. Upper Saddle River (New Jersey): Prentice Hall.

HAMMING, R. W. (1950). Error detecting and error correcting codes [online]. *Bell System Technical Journal*, 29(2): 147–160. <https://doi.org/10.1002/j.1538-7305.1950.tb00463.x>.

STANFILL, C., WALTZ, D. (1986). Toward memory-based reasoning [online]. *Communications of the ACM*, 29(12): 1213–1228. <https://doi.org/10.1145/7902.7906>.

STONE, C. J. (1977). Consistent Nonparametric Regression [online]. *Annals of Statistics,* 5(4): 595–645. <https://doi.org/10.1214/aos/1176343886>.

WELLER-FAHY, D., BORGHETTI, B. J., SODEMANN, A. A. (2015). Survey of Distance and Similarity Measures Used Within Network Intrusion Anomaly Detection. *IEEE Communication Surveys & Tutorials*, 17 (1): 70–91. <https://doi.org/10.1109/ COMST.2014.2336610>.

State Space Estimation: from Kalman Filter Back to Least Squares

Miroslav Plašil[1] | *Prague University of Economics and Business, Prague, Czech Republic*

Received 19.1.2023 (revision received 28.2.2023), **Accepted** (reviewed) 4.3.2023, **Published** 16.6.2023

Abstract

This note reviews a direct least squares estimation of a state space model and highlights its advantages over the standard Kalman filter in some applications. Although there is a close relationship between these two concepts, dual understanding of the estimation problem seems to be little appreciated by the mainstream econometric literature as well as applied researchers. Due to computational and theoretical advancements, the least squares estimation of a state space model has become a viable alternative in many fields, showing great potential in solving otherwise difficult problems. This note gathers and discusses some possible applications to illustrate the point and contribute to their wider use in practice.

Keywords	***DOI***	***JEL code***
Multi-objective least squares, State Space model, Kalman Filter	*https://doi.org/10.54694/stat.2023.3*	*C10, C51, C63*

INTRODUCTION

State space modelling has become the most popular approach to describing the behavior of dynamic systems. The state space representation of a model is very flexible and can subsume a wide class of models, including structural time series models, time-varying parameter models, dynamic factor models and many others. The range of applications varies from medicine, sociology and economics to national statistics. For instance, advanced temporal disaggregation methods used by many statistical institutes can be cast into a state space framework (Proietti, 2006).

Estimation of the unknown model quantities (or states) most often relies on the well-known Kalman filter recursions, which offer a computationally efficient solution to the underlying estimation problem (Kalman, 1960; Kalman and Bucy, 1961). However, one can argue that Kalman filter is just a logical extension of Gauss' original ideas on parameter estimation (Sorenson, 1970) and an elegant way to solve large – at the time infeasible – least squares problems (see Kollmann, 2013, or Andrle, 2014, among others).

The equivalence between Kalman filtering and 'large' least squares was well understood in the old days. Yet, it seems to have vanished from the radar screens of many applied researchers and statistical

[1] Faculty of Informatics and Statistics, Prague University of Economics and Business, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: plasil@vse.cz. Also the Czech National Bank, Prague, Czech Republic.

practitioners nowadays. This is quite unfortunate as alternative views at the estimation problem may equip researchers with an array of additional tools and help solve otherwise difficult problems. Advances in both computational power and theory make it now possible to benefit from these tools.

This note aims to return the equivalence between Kalman filtering and 'direct' least squares estimation of a state space model back into the spotlight and reviews why the latter can be of practical usefulness. It does not necessarily present any new results but offers another perspective on how to approach the estimation problem in question while trying to justify some of its benefits.

Unlike more traditional approaches that link the Kalman filter and the least squares estimator via the concept of the likelihood function (Chan and Jeliazkov, 2009; Aravkin et al., 2017), recursive least squares estimation (Pollock, 1999), or penalized least squares regression (Gomes, 1999), this note draws on the formulation of the multi-objective least squares problem and its neat solution. Although there is a close connection between this formulation and above-mentioned concepts (penalized least squares, in particular), the explicit reference to the multi-objective least squares is new in the literature to my utmost knowledge. It can be of some methodological interest, but above all this is of huge practical importance. Potential benefits are at least twofold: first, ordinary least squares and its extensions belong to the standard toolkit of many researchers. Reducing state space estimation to a common least squares problem thus invites to an active use of the least squares toolkit in the domain of state space estimation and modelling. The full scope of possibilities is yet to be fully appreciated by future research, but some of the state-of-the-art applications are discussed below to illustrate practical merits. Second, data augmentation approach used for solving multi-objective least squares enjoys a great flexibility in incorporating additional (cross) restrictions on the states which can greatly discipline high-level properties of the whole model. Data augmentation approach also provides a clear link to other statistical concepts, such as celebrated mixed estimation of Theil and Goldberger (1961) or more recent ridge regression approach to time-varying parameter estimation (Goulet Coulombe, 2020). Last but not the least, data augmentation approach enables to use standard regression routines available in common statistical environments (R, Matlab, Julia, Stata and others) to estimate state space models, thus avoiding a need for specific software or packages. More advanced users can obtain state estimates by using simple matrix multiplications that allow incorporating many practical extensions with ease.

1 A PRIMER ON THE LEAST SQUARES ESTIMATION AND STATE SPACE MODELLING

1.1 Least squares

For the sake of completeness and to set up the notation, it is helpful first to review the basics of the least squares estimation theory. Suppose that the data and parameters are related according to the linear (regression) model:

$$y = Xb + e, \tag{1}$$

where e are the measurement errors, y and X are formed by the observed data and b is a $q \times 1$ vector of parameters to be estimated. Good estimates, $\hat{b}$, should make the errors 'small' in some sense, i.e. they should make them close to zero as possible, $y - X\hat{b} \approx 0$.

The least squares meet this objective by minimizing the sum of squared errors, which is equivalent to minimizing squared Euclidean norm of the vector of errors, $||.||^2$:

$$\arg\min_b f = \arg\min_b \|Xb - y\|^2 = \arg\min_b \|e\|^2 = \left(e_1^2 + e_2^2 + \ldots + e_T^2\right) . \tag{2}$$

The minimization problem (2) is known to have (provided the matrix X has linearly independent columns) a closed-form solution (Boyd and Vandenberghe, 2018):

$$\hat{b} = (X'X)^{-1}X'y\,. \tag{3}$$

No specific distributional assumptions are needed to obtain the solution in (3) but these might be necessary for statistical inference. Under the standard assumptions of the ordinary linear regression model and Gaussian errors, the estimator $\hat{b}$ is normally distributed with mean b and variance $\sigma^2(X'X)^{-1}$:

$$\hat{b} = N(b,\sigma^2(X'X)^{-1}), \tag{4}$$

where σ^2 is the variance of the error term.

In some applications, the researchers may wish to work with several least squares objectives $f_1, f_2,\ldots, f_k$ associated with the same vector of parameters b, all of which should be minimized simultaneously. In this case, a standard solution for finding the values of the unknown parameter vector is to use a weighted sum of the objectives (also referred to as linear scalarization of the multi-objective optimization problem in some areas):

$$f = \lambda_1 f_1 + \ldots + \lambda_k f_k = \lambda_1 \|X_1 b - y_1\|^2 + \ldots + \lambda_k \|X_k b - y_k\|^2, \tag{5}$$

where $\lambda_1,\ldots,\lambda_k$ are positive constants representing the weights attached to individual objectives. The higher is the value of λ_i, the stronger is our desire for f_i to be minimized. In other words: constants λ_i determine a relative trade-off between individual objectives. Since scaling all the weights by any positive number does not change the solution of (5), it is possible to set λ_1 to 1 with no loss of generality.

The ingenious thing about multiple-objective least squares is that the objective (5) can be minimized using data augmentation approach (also known as *stacking*). It relies on stacking the matrices X_i and the corresponding vectors y_i below one another and solving a single (but bigger) least squares problem on the newly formed 'data set' (Boyd and Vandenberghe, 2018):

$$\underset{b}{\arg\min} \left\| \begin{pmatrix} \sqrt{\lambda_1} X_1 \\ \vdots \\ \sqrt{\lambda_k} X_k \end{pmatrix} b - \begin{pmatrix} \sqrt{\lambda_1} y_1 \\ \vdots \\ \sqrt{\lambda_k} y_k \end{pmatrix} \right\|^2 = \underset{b}{\arg\min} \|\tilde{X}b - \tilde{y}\|^2, \tag{6}$$

where $\tilde{X}$ and $\tilde{y}$ represent a stacked matrix and a stacked vector, respectively. One can therefore still make use of the Formulas (3) and (4) to derive a solution to (5). To illustrate the underlying machinery, let us take an example of the ridge regression (Hoerl and Kennard, 1970). It was originally developed as a tool to deal with highly correlated independent variables but is also heavily used as a regularization tool in many machine learning applications (Hastie, 2020). Ridge regression seeks to minimize the following objective function:

$$f = \|Xb - y\|^2 + \lambda \|b\|^2 = \|Xb - y\|^2 + \lambda \|I_q b - 0_q\|^2,$$

where I_q denotes identity matrix of order q and 0_q is a vector of q zeroes. Ridge regression modifies the objective of sum of squared errors associated with the traditional linear regression by adding a sum of squares of individual regression coefficients. The second objective serves as a penalty, which shrinks the coefficients towards zero. The trade-off between the goodness of fit and the size of coefficients is regulated through the hyperparameter λ.

Ridge regression naturally fits into the multi-objective least squares framework with $\lambda_1 = 1,\ \lambda_2 = \lambda,\ X_1 = X, y_1 = y, X_2 = I_q, y_2 = 0_q$. Stacking the matrices according to (6) and feeding them into the solution (3) one can obtain a closed-form expression for the estimates, $\hat{b}$:

$$\hat{b} = \left(\begin{pmatrix} X' & \sqrt{\lambda} I_q \end{pmatrix} \begin{pmatrix} X \\ \sqrt{\lambda} I_q \end{pmatrix} \right)^{-1} \begin{pmatrix} X' & \sqrt{\lambda} I_q \end{pmatrix} \begin{pmatrix} y \\ \sqrt{\lambda} 0_q \end{pmatrix} = \left(X'X + \lambda I_q \right)^{-1} X'y, \tag{7}$$

which is the canonical formula of ridge regression.

To close this section, we note that positive constants λ_i are usually scalars in the multi-objective least squares setting, but the least squares framework is flexible enough to account for a more complex system of weights where each observation or variable in the objective function can potentially conform to a different set of weights. If this is deemed necessary for the analysis, scalars λ_i in (6) can be replaced by the arbitrary square matrices W_i allowing for the implementation of fully general weighting schemes. In technical terms, this is a standard weighted least squares problem commonly encountered in practice. The choice of λ_i and W_i can be problem-dependent or they can sometimes be inferred from data (e.g. via cross validation, see Goulet Coulombe, 2020). In many applications, however, the weights are put equal to the reciprocal of the error variances (see below).

1.2 State space model

More than anything else, the state space form is an elegant and fully general representation of dynamic models associated with the underlying likelihood-based estimation procedure – the Kalman filter. Below we consider a simple linear state space model, which consists of two equations:

$$y_t = G_t \alpha_t + \varepsilon_t, \tag{8}$$

$$\alpha_t = F_t \alpha_{t-1} + \nu_t, \tag{9}$$

where an $n \times 1$ vector of observations, y_t, $t = 1,2, \dots, T$, depends on a $q \times 1$ vector of unknown states, α_t, which follow a Markovian process. System matrices G_t and F_t define the structure of the model and are assumed to be known. Interpretation of the states and system matrices is problem-dependent, but the states typically represent some unobserved latent variables (trends, cycles, factors) or time-varying regression parameters. In addition to defining the model structure, the matrix G_t can also contain observations of independent variables.

Observation equation (8) bears close similarities to the linear regression model. But unlike the traditional regression model, the states (quantities to be estimated) are not assumed to be constant and evolve over time. Transition equation (9) restricts their dynamics to follow an autoregressive process of order one.[2] The specification of the state space model is completed by specifying distributional assumptions on the error terms ε_t and ν_t. In the classical setting, they are assumed to be independently and identically distributed and Gaussian:

$$\begin{pmatrix} \varepsilon_t \\ \nu_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{pmatrix} \right). \tag{10}$$

[2] A random walk specification can be used in many applications. Note that we also allow for autoregressive processes of higher order since they can always be expressed as a large AR(1) model.

Formulas (8)–(10) form the basis for Kalman filter estimation, a recursive procedure for producing optimal estimates in the mean-square error sense. To spare space, we do not present Kalman filter formulas here, nor their derivation.[3] However, a few notes are in order.

First, estimates of the states can come in two flavors: *i*) estimates based on the information available up to time *t*, commonly referred to as the Kalman filter, and *ii*) estimates based on the whole data sample (i.e. on information up to time *T*), commonly referred to as the Kalman smoother. The direct 'large' least squares estimation of the states described below is equivalent to the Kalman smoother rather than the Kalman filter as it uses the whole data set. In theory, obtaining a series of filtered estimates via 'large' least squares would be possible but often impractical since one would always need to compute solution to a large least squares problem and store the estimate of the latest state once new observations have been added to the sample.

The second thing to observe is that due to its recursive nature, the Kalman filter/smoother needs to be initialized with some starting values to work. This consists in formulating a prior distribution of the initial state $\alpha_1 \sim N(x_0, D)$. The prior represents a best guess about the state and its covariance matrix before any observation has been collected. Strictly speaking, the least squares estimation procedure described below does not require the initialization step, but it might be helpful if there exist strong expert views on the proper starting values of the states. To maintain the closest link possible to the estimates produced by the Kalman smoother, we account for the starting values and associated uncertainty in the estimation procedure.

1.3 Estimating the state space model with the large least squares

To relate the state space model to a least squares problem it is useful to rewrite the Formulas (8)–(10) into a matrix form. Using $y = \left(y_1', y_2', \dots, y_T'\right)'$, $\alpha = \left(\alpha_1', \alpha_2', \dots, \alpha_T'\right)'$

$$G = \begin{bmatrix} G_1 & & \\ & \ddots & \\ & & G_T \end{bmatrix};\ H = \begin{bmatrix} I_q & & & & \\ -F_2 & I_q & & & \\ & -F_3 & I_q & & \\ & & \ddots & \ddots & \\ & & & -F_T & I_q \end{bmatrix},$$

$$\varepsilon \sim N\left(0, I_T \otimes \Omega_{11}\right),$$

$$v \sim N\left(0, S\right),$$

where $\otimes$ denotes Kronecker product and

$$S = \begin{bmatrix} D & & & & \\ & \Omega_{22} & & & \\ & & \Omega_{22} & & \\ & & & \ddots & \\ & & & & \Omega_{22} \end{bmatrix},$$

[3] The formulas are readily available elsewhere including a handful of online sources. For the textbook treatment, see Kim and Nelson (1999), for example.

the state space model (8)–(10) can be compactly rewritten as (see also Chan and Jeliazkov, 2009):

$$y = G\alpha + \varepsilon\,, \tag{11}$$

$$H\alpha = v\,. \tag{12}$$

Formulas (11) and (12) bear the structure of the standard linear regression model, and indeed, one can treat them as such. Least squares estimation of unknown states proceeds with minimizing the sum of squared errors in both equations (i.e. $y - G\hat{\alpha} \approx 0$ and $0 - H\hat{\alpha} \approx 0$). Since there does not generally exist a unique solution that would simultaneously minimize both objectives related to the same vector of states, there will be a trade-off between the two (regulated by their relative weights, λ_i):

$$\underset{\alpha}{\arg\min}\, f = \underset{\alpha}{\arg\min}\, \lambda_1 \|G\alpha - y\|^2 + \lambda_2 \|H\alpha - 0\|^2. \tag{13}$$

Due to the structure of the objective (13), its minimization can be treated as a multi-objective least squares problem and solved using a data augmentation approach (6). Inserting the stacked matrix into (3) while making no distributional assumptions, the solution to (13) is equal to:

$$\hat{\alpha} = \left(\left(\sqrt{\lambda_1} G' \sqrt{\lambda_2} H' \right) \begin{pmatrix} \sqrt{\lambda_1} G \\ \sqrt{\lambda_2} H \end{pmatrix} \right)^{-1} \left(\sqrt{\lambda_1} G' \quad \sqrt{\lambda_2} H' \right) \begin{pmatrix} \sqrt{\lambda_1} y \\ \sqrt{\lambda_2}\, 0 \end{pmatrix} = \left(\lambda_1 G' G + \lambda_2 H' H \right)^{-1} \lambda_1 G' y\,. \tag{14}$$

Formula (14) with the hyperparameters λ_1 and λ_2 left unspecified provides a general result that does not automatically coincide with the Kalman smoother as the latter uses a specific weighting scheme. However, it is worth noting that (14) is equivalent to the flexible least squares (FLS) estimator of Kalaba and Tesfatsion (1989), which was designed to handle time-varying parameter regression problems. Derivation in (14) thus establishes a clear formal relationship between the Kalman smoother and the FLS estimator. Kalman smoother simply represents a specific case of a more general FLS estimator – a thing that seems to be missed by the existing literature.

If the errors are assumed to be Gaussian, the weights λ_1, λ_2 can be replaced by the weighting matrices W_1 and W_2. In particular, it can be shown that optimal estimates are obtained by putting the weights equal to the inverse of the error variance in the measurement and the state equation, respectively, i.e. $W_1 = I_T \otimes \Omega_{11}^{-1}$ and $W_2 = S^{-1}$ (Aravkin et al., 2021). Plugging these weights into (14) one obtains:

$$\hat{\alpha} = \left(G' \left(I_T \otimes \Omega_{11}^{-1} \right) G + H' S^{-1} H \right)^{-1} G' \left(I_T \otimes \Omega_{11}^{-1} \right) y. \tag{15}$$

The expression (15) describing the solution to the weighted least squares problem provides exactly the same estimates of the unknown states as the Kalman smoother (see also Chan and Jeliazkov, 2009) and establishes formal equivalence between the two. The (normally distributed) estimator also has the same variance:[4]

$$Var\left(\hat{\alpha}\right) = \left(G' \left(I_T \otimes \Omega_{11}^{-1} \right) G + H' S^{-1} H \right)^{-1}. \tag{16}$$

[4] In multi-objective least squares framework, the variance of the estimator (16) can be obtained directly by applying Formula (4) on the stacked matrix $\tilde{X} = \begin{pmatrix} \sqrt{\lambda_1} G \\ \sqrt{\lambda_2} H \end{pmatrix}$ with the weights $\lambda_1 = I_T \otimes \Omega_{11}^{-1}$ and $\lambda_2 = S^{-1}$.

2 PRACTICAL USEFULNESS OF THE LEAST SQUARES APPROACH

Making a case for the least squares view on the state space model estimation should take nothing away from the Kalman filter brilliance. It still ranks among the most frequently used algorithms in statistical computing while giving rise to a myriad of useful extensions (e.g. to non-linear systems). There might be situations, however, in which the least squares approach can be preferable or even more natural to use. Since their formal review seems to be missing in the current literature, we try, at least, to point to some of them to provide practitioners with a brief outline of potential benefits and tools. Interest readers are, however, encouraged to dive in deeper into the references provided below.

2.1 Speed and stability of computations

The very motivation of Kalman for designing his celebrated recursive algorithm was to avoid a need to solve large-scale least squares problems. Their solution requires inverting (potentially very) large matrices in (15) which can be computationally expensive and oftentimes unstable. Against this backdrop, it might be surprising to argue in favor of the least squares approach. While it might be difficult to beat the Kalman filter in online applications, when it is necessary to update the solution every time new data arrive, this is not necessarily the case in static problems where all data are already available to the analyst.

An important observation to make is that the matrix to be inverted in (15) has a very specific structure: it is a symmetric positive definite block tridiagonal matrix (Aravkin et al., 2021). Recalling that matrix inversion has a close relation to solving a system of linear equations (Golub and Van Loan, 1996), the computation of (15) can be understood as a solution to symmetric block tridiagonal (SBT) system of linear equations.

Solving SBT systems has a long tradition (independent of the Kalman filter literature and largely unknown to applied econometricians) with many fast algorithms having been proposed and used in practice. Aravkin et al. (2021) showed that some of the popular recursive algorithms for solving SBT systems are actually equivalent to traditional Kalman filter recursions applied to state space estimation.[5] New methods for solving SBT systems are still being developed, which creates a potential for further speed-ups of the estimation process. Importantly, the solution to SBT systems also sheds some additional light on the stability guarantees of the Kalman filter algorithm and can serve as a basis for more flexible statistical modelling frameworks (see Aravkin et al., 2017, and Aravkin et al., 2021).

A whole range of new possibilities for the least squares approach to speed up the computations can also be found in Bayesian state space modeling and posterior sampling. Given that other parameters in the model are kept fixed, it is straightforward to show that the conditional posterior distribution of the states is Gaussian and its mean and variance coincide with (15) and (16), respectively (Chan and Jeliazkov, 2009). To obtain samples from the full posterior distribution of states, one can simply draw a random sample from the multivariate normal distribution fully specified by (15) and (16). This approach – unlike many alternative approaches – avoids the need to sample states recursively (Carter and Cohn, 1994). Using a Cholesky decomposition to sidestep the matrix inversion in (15), Chan and Jeliazkov (2009) report 20–40 % faster run times than the traditional posterior sampling of the states. The speed gains can grow even further with an efficient SBT solver.[6]

---

[5] Taking the opposite perspective, robust algorithms to solve SBT systems can potentially be understood as new (faster and more stable) implementations of the Kalman filter recursions.

[6] Some of the solvers are freely available in common statistical environments. If the speed of computations is not the main issue, standard linear regression routines implemented in these environments should work fine for a lot of practitioners.

2.2 Implementation of expert views into estimation process

In many instances, modeling real-world phenomena unavoidably relies on the vast experience of researchers. In macroeconomics, for example, expert prior views are an integral part of any serious model-based macroeconomic evaluation. If used with caution, transparent and informed judgements can help discipline high-level properties of the model and can greatly enhance its practical usefulness. Such properties include steady-state values of the variables and the states, frequency-domain features of the model, its impulse-response functions or its real-time revision properties. One way of incorporating expert views into the model is to express these views as a set of linear stochastic restrictions:

$$r = R\alpha + \epsilon, \tag{17}$$

where r and R is a vector and a matrix, respectively, which define a nature of the restriction on the state dynamics and ϵ is a vector of random disturbances which regulates the tightness of the restriction.

Noting that additional expert information contained in (17) has a form of the linear regression model (1) with properly chosen *artificial* data, we can effortlessly incorporate the set of restrictions (17) into the estimation process using the data augmentation approach (6), i.e., by stacking the artificial data in (17) below those observed in (11) and (12) and derive an explicit solution similarly to that shown in (14).

This approach can be seen as an application of the mixed-estimation ideas of Theil and Goldberger (1961) to state space modeling and allows taking on board relevant information (expert knowledge) which can hardly be incorporated into the model through observed data or a change in the model structure. We note that the implementation of stochastic restrictions into Kalman filtering is also possible due to Doran (1992). However, the least squares view on restrictions can perhaps be more intuitive and easier to implement for some practitioners. The upside of the least squares view is also methodological since it directly reveals the intimate relationship between observed data, expert prior views (i.e., *artificial* or extraneous data stacked below observed sample) and structural modeling (stochastic restrictions in a form of artificial data imply cross-restrictions on the model states, thus tying down the structure of the model).

2.3 Vast stock of knowledge accumulated in the least squares domain

Over the years, statisticians have accumulated much knowledge related to the least squares methodology suggesting a lot of its useful extensions and computational shortcuts. At present, the full potential of these theoretical advances has not yet been systematically explored within the state space context. While the exhaustive exploration of this area is far beyond the scope of this note, I set forth at least two practical examples to demonstrate how the least squares theory might help solve otherwise difficult tasks.

First, an iteratively reweighted least squares algorithm (IRLS) can be employed to approximate the minimization of arbitrary vector norm, thus making it possible to incorporate any user-defined loss function (and therefore error distributions) into state space modeling.[7] Although there might exist tailor-made algorithms to handle specific loss functions, IRSL represents a universal and computationally feasible tool that can handle general cases.

For example, the least squares solution to the optimization problem (13) can be made equivalent to minimizing the L1 norm. This can be achieved by multiplying each observation by the weight

[7] In the Bayesian context, this can be used to elicit different types of prior distributions. Although from a computational perspective, some stability issues can, in theory, be associated with the iteratively reweighted least squares procedure, it usually shows good practical performance. Recently, some new convergence guarantees have been found for the iteratively reweighted least squares, see e.g. Kümmerle et al. (2021). Note however that there also might exist problem-specific computational algorithms that are faster and more stable than IRLS for some vector norms.

$\sqrt{1/|\hat{e}_i|}$ where $\hat{e}_i$ is a regression residual. It comes from the fact that $\sum|\hat{e}_i| = \sum\left(\sqrt{\frac{1}{|\hat{e}_i|}}\hat{e}_i\right)^2$. Since the size of the residuals is not known a priori, the solution is obtained iteratively by setting all the weights to 1 in the initial step and alternating between the weighted least squares estimation of the states and the calculation of residuals.

Making use of a specific loss function and its properties can help solve many practical problems. For instance, imposing a L1 ('LASSO') penalty on the residuals in the observation equation (11) can result in the state space modeling robust to outliers, while the application of the L1 penalty to the transition equation (12) would force the signals to be piecewise linear. Moreover, it is straightforward to induce sparsity of time-varying signals (states) by considering an additional LASSO-type penalty on the sum of the absolute values of the states. This leads to a sparse LASSO-Kalman smoother analogous to the work of Angelosante et al. (2009).

Second, the least squares view on the state space estimation can be very beneficial within the area of the DSGE modeling – a workhorse modeling framework in modern macroeconomics (Christiano et al., 2018). Many DSGE models can be cast into the state space representation and its standard tools can be then applied. The need for the least squares machinery may arise in the domain of the stochastically singular DSGE models (i.e. models with more observables than economic shocks).[8] While no meaningful solution to such models can be found via the Kalman smoother, this represents a standard (under-determined) problem in the least squares domain, which is well-understood in linear algebra (see Andrle, 2014).

However, the role for the least squares does not end here. The estimation can also benefit from additional least squares restrictions on structural shocks in the DSGE model. True 'structural' or economic shocks in the DSGE models should be uncorrelated, but it is rarely true in empirical applications (Andrle, 2014). Using the multi-objective least squares framework (5) and penalizing solutions with correlated shocks, it is straightforward to retrieve structural shocks in line with the underlying economic theory. Moreover, the form of the solution (15) which directly expresses the states as a weighted average of observables can be useful for assessing frequency-domain properties of the underlying DSGE model as well as for the decomposition of unobserved quantities into observables (see e.g. Gomes, 2007). Such analyses are a vital part of the macroeconomic assessment in many central banks and international economic organizations.[9]

CONCLUSION

State space modeling and Kalman filtering belong to the standard toolkit of applied researchers in many scientific fields. The possibility of estimating unknown states as a large but standard linear regression problem has long been known but somewhat neglected in recent decades, partially due to huge success of the Kalman filter in many applications. However, computational advances made it possible to return the direct least squares estimation back into the spotlight and provide practitioners with a whole new range of possibilities. Least squares estimation of a state space model can lead to computational speed-ups in static applications and show practical improvements over traditional Kalman filtering while broadening our understanding of the latter.

---

[8] This is (or should be) a common situation in advanced macroeconomic models since the evolution of many economic variables is driven by only a few underlying economic forces (shocks), say a demand shock.

[9] In macroeconomics, the unknown states often represent important (directly unobserved) economic concepts such as an output gap. Decomposing contributions of observed variables to the estimation of unobserved states is thus very useful for policy and model assessment.

This note aimed at introducing the least squares estimation of a state space model to a wide audience of applied researchers. The ambition was to discuss new perspectives, demonstrate relationships between diverse statistical concepts and provide practitioners with a brief list of potential applications. The hope is that it can open the door to whole new promising avenues of research in the area of advanced state space modeling where the advantages of a (penalized) least squares solution can be thoroughly appreciated. Although this short note only covered a limited set of applications, the potential of using least squares methodology in the context of state space modelling is immense and yet to be fully realized in both theory and practice.

References

ANDRLE, M. (2014). *Estimating Structural Shocks with DSGE Models* [online]. <http://michalandrle.weebly.com/uploads/1/3/9/2/13921270/opfilter.pdf>.

ANGELOSANTE, D., ROUMELIOTIS, S. I., GIANNAKIS, G. B. (2009). Lasso-Kalman smoother for tracking sparse signals [online]. *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, 181–185. <https://doi.org/10.1109/ACSSC.2009.5470133>.

ARAVKIN, A. I., BURKE, J. V., BELL, B. M., PILLONETTO, G. (2021). Algorithms for Block Tridiagonal Systems: Stability Results for Generalized Kalman Smoothing [online]. *IFAC-PapersOnLine*, 54(7): 821–826. <https://doi.org/10.1016/j.ifacol.2021.08.463>.

ARAVKIN, A. I, BURKE, J. V., LJUNG, L., LOZANO, A., PILLONETTO, G. (2017). Generalized Kalman smoothing: Modeling and algorithms [online]. *Automatica*, 86: 63–86. <https://doi.org/10.1016/j.automatica.2017.08.011>.

BOYD, S., VANDENBERGHE, L. (2018). *Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares*. Cambridge University Press.

CHAN, J. C. C., JELIAZKOV, I. (2009). Efficient simulation and integrated likelihood estimation in state space models. *Mathematical Modelling and Numerical Optimisation*, 1(1/2): 101–120.

CARTER, C. K., KOHN, R. (1994). On Gibbs sampling for state space models [online]. *Biometrika*, 81(3): 541–553. <https://doi.org/10.1093/biomet/81.3.541>.

CHRISTIANO, L. J., EICHENBAUM, M. S., TRABANDT, M. (2018). On DSGE Models [online]. *Journal of Economic Perspectives*, 32(3): 113–40. <https://doi.org/10.1257/jep.32.3.113>.

DORAN, H. E. (1992). Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions [online]. *The Review of Economics and Statistics*, MIT Press, 74(3): 568–572. <https://doi.org/10.2307/2109505>.

GOLUB, G. H., VAN LOAN, CH. F. (1996). *Matrix Computations*. Johns Hopkins University Press.

GOMES, V. (1999). Three Equivalent Methods for Filtering Finite Nonstationary Time Series [online]. *Journal of Business & Economic Statistics*, 17(1), 109–116. <https://doi.org/10.1080/07350015.1999.10524800>.

GOMES, V. (2007). Wiener–Kolmogorov Filtering and Smoothing for Multivariate Series With State–Space Structure [online]. *Journal of Time Series Analysis*, 28(3): 361–385. <https://doi.org/10.1111/j.1467-9892.2006.00514.x>.

GOULET COULOMBE, P. (2020). *Time-Varying Parameters as Ridge Regressions* [online]. *arXiv.org*. <https://doi.org/10.48550/arXiv.2009.00401>.

HASTIE, T. (2020). Ridge Regularization: an Essential Concept in Data Science [online]. *Technometrics*, 62(4), 426–433. <https://doi.org/10.1080/00401706.2020.1791959>.

HOERL, A. E., KENNARD R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems [online]. *Technometrics*, 12(1): 55–67. <https://doi.org/10.2307/1267351>.

KALABA, R., TESFATSION, L. (1989). Time-varying linear regression via flexible least squares [online]. *Computers and Mathematics with Applications*, 17(8–9): 1215–1245. <https://doi.org/10.1016/0898-1221(89)90091-6>.

KALMAN, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems [online]. *Journal of Basic Engineering*. 82: 35–45. <http://dx.doi.org/10.1115/1.3662552>.

KALMAN, R. E., BUCY, R. S. (1961). New Results in Linear Filtering and Prediction Theory [online]. *Journal of Basic Engineering*, 83(1): 95–108. <https://doi.org/10.1115/1.3658902>.

KIM, C-J., NELSON, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications* [online]. The MIT Press. <https://doi.org/10.7551/mitpress/6444.001.0001>.

KOLLMANN, R. (2013). Estimating the state vector of linearized DSGE models without the Kalman filter [online]. *Economics Letters*, 120(1): 65–66. <https://doi.org/10.1016/j.econlet.2013.03.041>.

KÜMMERLE, C., VERDUN, C. M., STÖGER, D. (2021). Iteratively Reweighted Least Squares for Basis Pursuit with Global Linear Convergence Rate. In: BEYGELZIMER, A., DAUPHIN, Y., LIANG, P., VAUGHAN, J. W. (eds.) *Advances in Neural Information Processing Systems*.

POLLOCK, D. S. G. (1999). *Time-Series Analysis, Signal Processing and Dynamics*. London: The Academic Press.
PROIETTI, T. (2006). Temporal Disaggregation by State Space Methods: Dynamic Regression Methods Revisited [online]. *The Econometrics Journal*, 9(3): 357–372. <https://doi.org/10.1111/j.1368-423X.2006.00189.x>.
SORENSON, H. W. (1970). Least-squares estimation: from Gauss to Kalman [online]. *IEEE Spectrum*, 7 (7): 63–68. <https://doi.org/10.1109/MSPEC.1970.5213471>.
THEIL, H., GOLDBERGER, A. (1961). On Pure and Mixed Statistical Estimation in Economics [online]. *International Economic Review*, 2(1):65–78. <http://dx.doi.org/10.2307/2525589>.

A Simple Estimation of Parameters for Discrete Distributions from the Schröter Family

Friday I. Agu[1] | *Slovak Academy of Sciences, Bratislava, Slovakia*
Ján Mačutek[2] | *Constantine the Philosopher University in Nitra, Nitra, Slovakia*
Gábor Szűcs[3] | *Comenius University, Bratislava, Slovakia*

Received 12.9.2022, **Accepted** (reviewed) 3.10.2022, **Published** 16.6.2023

Abstract

One of the common challenges in actuarial mathematics is finding a model for the number of claims and claim severity. We focus on one of the suggested models, namely, on the Schröter family of discrete probability distributions. Furthermore, we introduce a simple and easy-to-compute parameter estimate for this family of distributions, which can be used especially as initial values in optimization algorithms that are needed to compute other estimates.

Keywords	***DOI***	***JEL code***
Discrete probability distributions, Schröter recursive formula, parameter estimation, aggregate claim distribution	*https://doi.org/10.54694/stat.2022.42*	*C46, C13, C25*

INTRODUCTION

In actuarial mathematics, there has been a significant active interest in modeling the number of claims and claim severities of a collective risk model for life and non-life insurance. A reliable model enables predictions that could help insurance companies set competitive prices for insurance portfolios and maintain adequate investment for the next operational year. Also, the insurance company could decide on the appropriate margins for the cost of each portfolio to account for future uncertainty.

[1] Mathematical Institute, Slovak Academy of Sciences, Štefánikova 49, 81743 Bratislava, Slovakia. E-mail: agu1@uniba.sk. Also Department of Statistics, University of Calabar, Calabar, Nigeria.

[2] Department of Mathematics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Tr. A. Hlinku 1, 949 01 Nitra, Slovakia. E-mail: jmacutek@yahoo.com. Also Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia.

[3] Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina F1, 842 48 Bratislava, Slovakia.

A “naïve” approach to obtaining accurate aggregate claims is to find an appropriate family of counting distributions and to fit them to the number of claims and claim severities separately. However, this approach fails more often than not. The distribution of aggregate claims is, in fact, a convolution of the distributions of the number of claims and the claim severity.

Therefore, one of the common problems in actuarial mathematics is modeling the aggregate claims distribution of a collective risk model, with the claim severity and the number of claims considered as discrete random variables. In the literature, distributions such as compound negative binomial and compound Poisson have been studied extensively and used to model the aggregate claims distribution based on the convolution approach. Also, the theoretical aspects of compound distributions are in Johnson et al. (2005). Bening and Korolev (2012) focus on the compound Poisson distributions and their applications in actuarial and financial mathematics. Wimmer and Altmann (1999) enumerate several examples of these distributions.

From a theoretical perspective, the convolution approach is unambiguous. However, when the number of claim events increases, the computation of the aggregate claim distribution using the convolution approach becomes difficult. This problem has led researchers to search for alternative methods for computing the aggregate claim distribution of collective risk models. In this paper, we focus on one of them, namely, the recursive method.

The recursive method (Sundt and Vernic, 2009) assumes that the claim severity distribution is discrete and can compute aggregate claims recursively when the number of claims increases. It does not involve computing several convolutions of the conditional distribution of the number of claim events and requires far less computer time.

Panjer (1981) introduced the famous recursive formula:

$$P_n = \left(a + \frac{b}{n}\right) P_{n-1}, \quad n = 1, 2, \ldots \tag{1}$$

where a and b are parameters (by definition, henceforth $P_n = 0$ for $n < 0$). It addressed the computational problems of the convolution approach. However, only a few distributions – binomial, Poisson, and negative binomial[4] – satisfy the formula. Therefore, several suggestions to generalize (1) appeared in the literature, for example, see Schröter (1990), and Sundt (1992).

1 SCHRÖTER DISTRIBUTION FAMILY

Schröter’s (1990) second-order recursive formula is:

$$P_n = \left(a + \frac{b}{n}\right) P_{n-1} + \frac{c}{n} P_{n-2}, \quad n = 1, 2, 3, \ldots \tag{2}$$

where a, b, and c are parameters. It is obvious that (1) is a special case of (2) for $c = 0$. In addition to the distributions given by the Panjer’s recursion (1), the distribution family defined by (2) contains also convolutions of the Poisson distribution and another distribution from the Panjer family. Thus, the Schröter family is more flexible and can better capture the behavior of the number of claims and their severity. It can also be applied within actuarial reserving approaches, such as e.g. the one proposed by Maciak et al. (2021).

Also, Schröter (1990) derived some basic properties of distributions given by (2). These distributions have the probability generating function:[5]

[4] The Panjer’s Formula (1) defines, in fact, the Katz family of discrete probability distributions, see Wimmer and Altmann (1999).

[5] Formulas (3)–(6) are true if $a \neq 0$. However, they become more complicated if $a = 0$.

$$G(s) = e^{-\frac{c}{a}(s-1)}\left(\frac{1-a}{1-as}\right)^{\frac{a(a+b)+c}{a^2}}, \tag{3}$$

the explicit expression for the probability mass function:

$$P_n = e^{\frac{c}{a}}(1-a)^{\frac{a(a+b)+c}{a^2}}\sum_{j=0}^{n}\binom{\frac{a(a+b)+c}{a^2}+j-1}{j}\frac{\left(-\frac{c}{a}\right)^{n-j}a^j}{(n-j)!}, \tag{4}$$

the mean:

$$\mu = \frac{a+b+c}{1-a}, \tag{5}$$

and the variance:

$$\sigma^2 = \frac{a+b+(2-a)c}{(1-a)^2}. \tag{6}$$

Several problems are related to the discrete probability distributions given by (2), which remain open. The parametric space of these distributions has not yet been specified. Although, Schröter (1990) presents some conditions that the parameters must satisfy. Another area that deserves a detailed investigation is parameter estimation.

2 PARAMETER ETIMATION

Due to a relatively complicated formula for the probability mass function (4), classical parameter estimation methods have no explicit results. For example, the maximum likelihood method results in a system of equations that has no explicit solutions and must be solved numerically. Luong and Garrido (1993) mention potential problems of such numerical solutions.

Therefore, Luong and Garrido (1993) suggested an estimation method specifically for recursively defined probability distributions, based on minimizing the quadratic distance (the recursion formula is considered a linear regression model). Luong and Doray (2002), and Doray and Haziza (2004) further elaborated on this idea. The minimum quadratic distance estimations have, under certain conditions, desirable properties (asymptotic normality, consistency, asymptotic efficiency). However, computations still require numerical methods and the use of software (they involve, e.g., matrix inversion).

We derive a simple, easy-to-compute estimation of parameters for distributions from the Schröter family. In Formulas (5) and (6), we replace the mean and variance with their empirical counterparts and the parameters with their estimates. That is, we have:

$$\bar{x} = \frac{\hat{a}+\hat{b}+\hat{c}}{1-\hat{a}}, \tag{7}$$

$$s^2 = \frac{\hat{a}+\hat{b}+(2-\hat{a})\hat{c}}{(1-\hat{a})^2}. \tag{8}$$

Next, denote N the sample size, f_0, f_1, f_2, ... the observed frequencies of values 0, 1, 2, …, and k the number for which the sum of three neighbouring frequencies $f_k + f_{k+1} + f_{k+2}$ attains its maximum. The empirical analogue of (2) for $n = k$ is the equation:

$$\hat{p}_k = \left(\hat{a} + \frac{\hat{b}}{k} \right) \hat{p}_{k-1} + \frac{\hat{c}}{k} \hat{p}_{k-2}, \tag{9}$$

where $\hat{p}_k$, $\hat{p}_{k-1}$, and $\hat{p}_{k-2}$ are empirical probabilities (i.e. $\hat{p}_k = \frac{f_k}{N}$).[6]

The parameters will be estimated by solving the system of Formulas (7), (8), and (9). From (7) we have:

$$\bar{x}(1-\hat{a}) - \hat{c} = \hat{a} + \hat{b}. \tag{10}$$

Using (10), Formula (9) can be rewritten as:

$$k\hat{p}_k = \left[\hat{a}(k-1+1) + \hat{b} \right] \hat{p}_{k-1} + \hat{c}\hat{p}_{k-2} = \hat{a}\left(k-1-\bar{x} \right) \hat{p}_{k-1} + \bar{x}\,\hat{p}_{k-1} + \hat{c}\left(\hat{p}_{k-2} - \hat{p}_{k-1} \right), \tag{11}$$

and from (11) it follows that:

$$\hat{c} = \frac{k\hat{p}_k - \hat{a}\left(k-1-\bar{x} \right) \hat{p}_{k-1} - \bar{x}\,\hat{p}_{k-1}}{\hat{p}_{k-2} - \hat{p}_{k-1}}. \tag{12}$$

Combining (8) and (10) gives:

$$s^2 (1-\hat{a}) = \bar{x} + \hat{c}, \tag{13}$$

and solving (12) and (13) yields the estimate:

$$\hat{a} = \frac{\left(s^2 - \bar{x} \right)\left(\hat{p}_{k-2} - \hat{p}_{k-1} \right) - k\hat{p}_k + \bar{x}\,\hat{p}_{k-1}}{s^2 \left(\hat{p}_{k-2} - \hat{p}_{k-1} \right) - \left(k-1-\bar{x} \right) \hat{p}_{k-1}}. \tag{14}$$

Using (10), (13), and (14) we obtain estimates of the other two parameters:

$$\hat{c} = s^2 (1-\hat{a}) - \bar{x}, \tag{15}$$

$$\hat{b} = \bar{x}(1-\hat{a}) - \hat{a} - \hat{c}. \tag{16}$$

3 NUMERICAL APPLICATIONS

3.1 Simulation study

We performed a simple simulation study to gain insight into the behavior of the estimates derived in Section 2.[7] We set parameter values as $a = 0.6$, $b = 2.6$, and $c = -1.1$ and generated 100 000 random numbers from this distribution. The generated numbers become the random sample from which we estimate the parameters. Furthermore, we repeat this procedure 10 000 times (that is, we obtained 10 000 estimates of each parameter). Table 1 and Figure 1 present the descriptive statistics of the estimated parameters.

[6] In principle, any trinity of neighbouring empirical probabilities can be used in Formula (9).

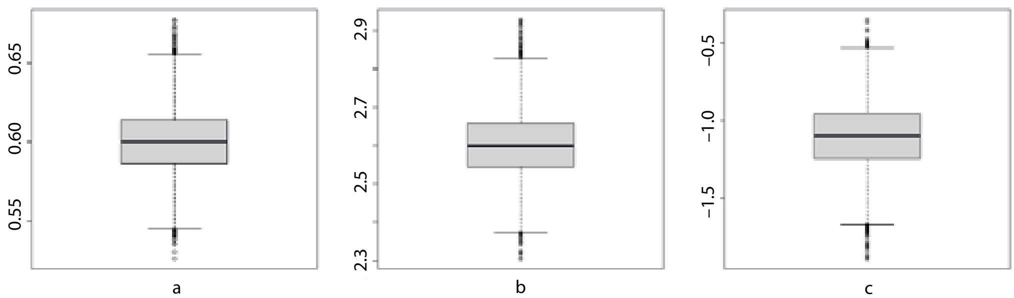
[7] All computations were performed in the statistical software environment R: <*www.r-project.org*>.

Table 1 The descriptive statistics of the estimates

	$\hat{a}$	$\hat{b}$	$\hat{c}$
Mean	0.60033	2.60136	–1.103495
Standard deviation	0.02063	0.08455	0.21174
Minimum	0.52549	2.30470	–1.89726
Maximum	0.67759	2.91813	–0.34615

Source: Own study

Figure 1 Boxplots of the estimates



Source: Own study

The means from Table 1 indicate that the estimates (at least for parameters a and b) could even be unbiased.[8] However, their standard deviations (especially for $\hat{c}$) seem not to converge to 0, leaving thus their consistency doubtful.

3.2 Car accident injuries

We applied the Schröter recursive relation (2) as a model for car accident injury data.[9] We consider the accidents that happened in the Olomouc region of the Czech Republic from January 1, 2021, to December 31, 2021, and all types of injuries (deadly, serious, and minor). In Table 2, we present the number of days with a particular number of injuries (for example, there were 40 days without an injury, 64 days with one injury, etc.).

Table 2 Car accident injuries in the Olomouc region from January 1, 2021, to December 31, 2021 (x – number of injuries, $f(x)$ – number of days with x injuries)

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
0	40	3	55	6	29	9	7
1	64	4	33	7	22	10	5
2	60	5	39	8	8	12	3

Source: Own study, using surveys from <*www.irozhlas.cz/nehody*>

---

[8] The values from Table 1 are quite stable when simulations are rerun under the same conditions, with the means and the standard deviations changing at the third decimal place at most.

[9] The data were created from surveys (the webpage of the Czech public radio broadcaster, accessed on 10 October 2022) available at: <*www.irozhlas.cz/nehody*>.

The model achieves a good fit in terms of the Pearson chi-square test (p = 0.1677; $\hat{a}$ = 0.451, $\hat{b}$ = 1.127, $\hat{c}$ = 0.254). We note that this p-value was computed using parameter estimates given by Formulas (14)–(16), and it can be improved if, for example, we use the minimum chi-square method to estimate the parameters (with our estimations serving as initial values in optimization algorithms).

CONCLUSION

The paper presents new parameter estimations for discrete probability distributions from the Schröter family. They are very easy to compute, as they are given explicitly and do not involve iteration algorithms, numerical optimization etc., which is the case of previously suggested estimations (Luong and Garido, 1993; Luong and Doray, 2002; Doray and Haziza, 2004). The new Formulas (14)–(16) for parameter estimation can thus be used to gain a preliminary insight into data, and especially as initial values in numerical procedures which are used in the abovementioned more sophisticated approaches.

ACKNOWLEDGEMENT

Supported by research projects VEGA 2/0096/21 (F. I. Agu and J. Mačutek), APVV-21-0216 (J. Mačutek), and the Operational Programme Integrated Infrastructure (OPII) for the project 313011BWH2: "InoCHF – Research and development in the field of innovative technologies in the management of patients with CHF", co-financed by the European Regional Development Fund (J. Mačutek).

References

BENING, V. E., KOROLEV, V. Y. (2012). *Generalized Poisson Models and their Applications in Insurance and Finance*. Berlin, Boston: de Gruyter.

DORAY, L. G., HAZIZA, A. (2004). Minimum distance inference for Sundt's distributions. In: ANTOCH, J. (eds.) *COMPSTAT 2004 – Proceedings in Computational Statistics*, Heidelberg: Physica-Verlag.

JOHNSON, N. L., KEMP, A. W., KOTZ, S. (2005). *Univariate Discrete Distributions*. Hoboken: Wiley.

LUONG, A., DORAY, L. G. (2002). General quadratic distance methods for discrete distributions are definable recursively [online]. *Insurance: Mathematics and Economy*, 30(2): 255–257. <https://doi.org/10.1016/S0167-6687(02)00100–2>.

LUONG, A., GARRIDO, J. (1993). Minimum quadratic distance estimation for a parametric family of discrete distributions defines recursively [online]. *Australian Journal of Statistics*, 35(1): 59–67. <https://doi.org/10.1111/j.1467-842X.1993.tb01312.x>.

MACIAK, M., OKHRIN, O., PEŠTA, M. (2021). Infinitely stochastic micro reserving [online]. *Insurance: Mathematics and Economics*, 100: 30–58. <https://doi.org/10.1016/j.insmatheco.2021.04.007>.

PANJER, H. H. (1981). Recursive evaluation of a family of compound distributions [online]. *ASTIN Bulletin*, 12(1): 22–26. <https://doi.org/10.1017/S0515036100006796>.

SCHRÖTER, K. J. (1990). On a family of counting distributions and recursions for related distributions [online]. *Scandinavian Actuarial Journal*, 2–3: 161–175. <https://doi.org/10.1080/03461238.1990.10413879>.

SUNDT, B. (1992). On some extensions of Panjer's class of counting distributions [online]. *Scandinavian Actuarial Journal,* 1990 (2–3): 161–175. <https://doi.org/10.2143/AST.22.1.2005127>.

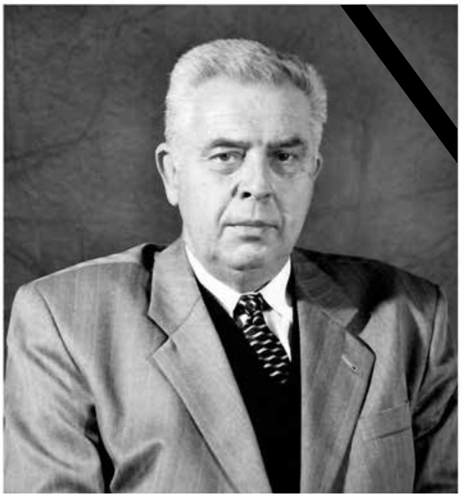
SUNDT, B., VERNIC, R. (2009). *Recursions for Convolutions and Compound Distributions with Insurance Applications*. Berlin, Heidelberg: Springer.

WIMMER, G., ALTMANN, G. (1999). *Thesaurus of Univariate Discrete Probability Distributions*. Essen: Stamm.

Remembering Professor Jan Seger (21.2.1938–9.12.2022)

Richard Hindls | *Prague University of Economics and Business, Prague, Czech Republic*
Prokop Závodský | *Prague University of Economics and Business, Prague, Czech Republic*

At the age of 84 years, Rector Emeritus of the Prague University of Economics and Business (VŠE) and Dean Emeritus of the Faculty of Informatics and Statistics of the same University passed away at the end of the year 2022. He was a great educator deceased who devoted virtually his entire professional life to work in the academic milieu; he educated a long row of successful graduates and participated in the development of his alma mater.



Jan Seger was born in Prague, where he also graduated from his secondary school studies (A-level examination in the year 1956). From 1956 to 1961, he studied at the Prague University of Economics and Business (VŠE) in the field of statistics. There he experienced reforms of studies, which were frequent there at that time – he was admitted to the Faculty of Statistics yet graduated from the Faculty of Political Economy; the original denomination of "graduated economist" was changed for those who graduated in the year 1966 to the "Ing." university degree. For two years (partly also filled with military service), he worked as a programmer in a computer technology company called Aritma Vokovice.

In 1963, he was admitted to the Department of Statistics of the University (VŠE) as an assistant (later lecturer), and he remained faithful to that workplace almost for half a century. In professional terms, he mainly specialised in analysing economic time series and statistical prognostic methods. In 1968, he defended his dissertation called *Possibilities of Application of Exponential Seasonal Smoothing for Analysis of Seasonal Time Series of Economic Indicators*; afterwards, he paid a four-month study visit to a university in Brussels. In 1977, he was appointed senior lecturer after he defended his habilitation thesis called *Possibilities of Construction of Extrapolative Forcasting of Economic Indicators*. In 1992, he was appointed Professor of Statistics. Prof. Seger was always evaluated in surveys of students as an excellent lecturer.

The scientific and publication activities of Jan Seger are enornous. From 1960s to 1980s, they mainly focused on the aforementioned analysis methods of time series and prognostic methods. Besides numerous contributions in journals and compendia (*Statistika*, *Statistická revue*, *Acta Oeconomica Pragensia* and others, even abroad), they were didactic texts, e.g. *Introduction to Analysis of Time Series* (1970 – with Josef Kozák), *Simple statistical methods in prognostics* (1975 – with Josef Kozák), S*olved problems from statistical prognostics* (1977 – with Jiří Žváček) and others. Since the beginning of 1980s, Jan Seger published several didactic texts for a basic course of statistics at the University (VŠE), which were popular with students for their conciseness and clarity. The first one was *Fundamentals of statistics for Business Administration Faculty* (1981, 249 pp.). Later a textbook, issued as a book, was already significantly larger – *Statistical Methods for Economists of Industry* (1988, 545 pp.). In 1990s, Jan Seger published another series of textbooks for a modernised course of fundamentals of statistics – *Statistical Methods in Economics* (1993 – with Richard Hindls), *Statistical Methods in the Market Economy* (1995 – with Richard Hindls), and *Statistics in the Economy* (1998 – with Richard Hindls and Stanislava Hronová). In subsequent years, Prof. Seger was a co-author of a successful textbook, *Statistics for Economists,* that was issued in 8 editions in total.

Jan Seger held at the University (VŠE) many important academic posts. Already in the second half of the 1980s, he was a Deputy Head of the Department of Statistics. Soon after the department was split during the reorganisation of the University (VŠE) in 1990, he was appointed head of a newly established Department of Economic Statistics (he headed it for the whole 16 years: 1990–2006). From 1991 to 1992, he was a Vice-Dean for education at the Faculty of Informatics and Statistics. Afterwards, as a Vice-Rector for education (1992–1994), he was instrumental in finishing a system of credit hours in studies, which had not been tried at Czech universities at that time yet. In December 1993 and again three years later, he was smoothly elected Rector of the University (VŠE) by the academic senate of the University (VŠE). As the eighth rector since the foundation of the University (VŠE) and the third one after the Velvet Revolution of 1989, Prof. Seger performed the function from 1994 to 2000. During the inaugural, in February 1994, for the first time in the history of the University (VŠE), there was a lecture by a rector about the importance of statistics for economic theory and practice. As a Rector, he managed to considerably stabilise the University and create foundations for its development on the brink of the millennium. For the years 2006–2010, Prof. Seger was elected Dean of the Faculty of Informatics and Statistics.

During the whole period of his work at the University (VŠE) since the 1960s, Jan Seger co-operated with the state statistical office (the office was repeatedly re-organised and re-named during that period) and its research institute; he participated in solving many research assignments and in technical publications. He also held the position of the Vice Chairman of the Czech Statistical Council. In 2014, Prof. Jan Seger became the second laureate awarded the Prize of the President of the Czech Statistical Office and of the Rector of the Prague University of Economics and Business (VŠE) for his life-long contribution to statistics.

Professor Seger had a great sense of the personal traits of colleagues and was a smart commentator of various situations and human characters. He was famous for his fondness for spending time in his cottage, where he relaxed and gained strength. At first, he used to go to the Šumava Mountains and, since the 1970s, to the Sedlčany Area, where he was also laid to eternal rest at the cemetery in Vojkov as he wished.

Tribute to his memory!

Conferences

The ***25th Applications of Mathematics and Statistics in Economics Conference (AMSE 2023)*** will take place **from 30th August to 3rd September 2023 in Rájecké Teplice** (hotel Diplomat), **Slovakia**. This year organized by the Faculty of Economics, Matej Bel University in Banská Bystrica. More at: <*https://www.amse-conference.eu*>.

The ***31st Interdisciplinary Information Management Talks Conference (IDIMT 2023)*** on "New Challenges for ICT and Management" will be held **during 6–8 September 2023 in Hradec Králové, Czechia**. More at: <*https://idimt.org*>.

The ***17th International Days of Statistics and Economics Conference (MSED 2023)*** will take place **during 7–9 September 2023 at the Prague University of Economics and Business, Czechia**. The aim of the conference is to present and discuss current problems of Statistics, Demography, Economics and Management. More at: <*http://msed.vse.cz*>.

The ***41st International Conference on Mathematical Methods in Economics (MME 2023)*** will be held **from 13th to 15th September 2023 in Prague, Czechia**. More at <*https://mme2023.vse.cz*>.

Papers
We publish articles focused at theoretical and applied statistics, mathematical and statistical methods, conception of official (state) statistics, statistical education, applied economics and econometrics, economic, social and environmental analyses, economic indicators, social and environmental issues in terms of statistics or economics, and regional development issues.

The journal of *Statistika* has following sections:

The *Analyses* section publishes complex and advanced analyses based on the official statistics data focused on economic, environmental, social and other topics. Papers shall have up to 12 000 words or up to 20 1.5-spaced pages.

Discussion brings the opportunity to openly discuss the current or more general statistical or economic issues, in short what the authors would like to contribute to the scientific debate. Contribution shall have up to 6 000 words or up to 10 1.5-spaced pages.

In the *Methodology* section we publish articles dealing with possible approaches and methods of researching and exploring social, economic, environmental and other phenomena or indicators. Articles shall have up to 12 000 words or up to 20 1.5-spaced pages.

Consultation contains papers focused primarily on new perspectives or innovative approaches in statistics or economics about which the authors would like to inform the professional public. Consultation shall have up to 6 000 words or up to 10 1.5-spaced pages.

Book Review evaluates selected titles of recent books from the official statistics field (published in the Czech Republic or abroad). Reviews shall have up to 600 words or 1–2 1.5-spaced pages.

The *Information* section includes informative (descriptive) texts, information on latest publications (issued not only by the Czech Statistical Office), or recent and upcoming scientific conferences. Recommended range of information is 6 000 words or up to 10 1.5-spaced pages.

Language
The submission language is English only. Authors are expected to refer to a native language speaker in case they are not sure of language quality of their papers.

Recommended paper structure
Title – Authors and Contacts – Abstract (max. 160 words) Keywords (max. 6 words / phrases) – Introduction – 1 Literature survey – 2 Methods – 3 Results – 4 Discussion –Conclusion – Acknowledgments – References – Annex (Appendix). Tables and Figures (for print at the end of the paper; for the review process shall be placed in the text).

Authors and contacts
Rudolf Novak,[1] Institution Name, City, Country
Jonathan Davis, Institution Name, City, Country
[1] Address. Corresponding author: e-mail: rudolf.novak @domainname.cz, phone: (+420) 111 222 333.

Main text format
Times 12 (main text), 1.5 spacing between lines. Page numbers in the lower right-hand corner. *Italics* can be used in the text if necessary. *Do not* use **bold** or <u>underline</u> in the text. Paper parts numbering: 1, 1.1, 1.2, etc.

Headings
1 FIRST-LEVEL HEADING (Times New Roman 12, bold)
1.1 Second-level heading (Times New Roman 12, bold)
1.1.1 Third-level heading (Times New Roman 12, bold italic)

Footnotes
Footnotes should be used sparingly. Do not use endnotes. Do not use footnotes for citing references.

References in the text
Place references in the text enclosing authors' names and the year of the reference, e.g. „…White (2009) points out that…", „…recent literature (Atkinson and Black, 2010a, 2010b, 2011; Chase et al., 2011: 12–14) conclude…". Note the use of alphabetical order. Between the names of two authors please insert „and", for more authors we recommend to put „et al.". Include page numbers if appropriate.

List of references
Arrange list of references alphabetically. Use the following reference styles: [book] HICKS, J. (1939). *Value and Capital: an Inquiry into Some Fundamental Principles of Economic Theory.* 1st Ed. Oxford: Clarendon Press. [chapter in an edited book] DASGUPTA, P. et al. (1999). Intergenerational Equity, Social Discount Rates and Global Warming. In: PORTNEY, P., WEYANT, J. (eds.) *Discounting and Intergenerational Equity,* Washington, D.C.: Resources for the Future. [on-line source] CZECH COAL. (2008). *Annual Report and Financial Statement 2007* [online]. Prague: Czech Coal. [cit. 20.9.2008]. <http://www.czechcoal.cz/cs/ur/zprava/ur2007cz.pdf>. [article in a journal] HRONOVÁ, S., HINDLS, R., ČABLA, A. (2011). Conjunctural Evolution of the Czech Economy. *Statistika: Statistics and Economy Journal,* 91(3): 4–17. [article in a journal with DOI]: Stewart, M. B. (2004). The Employment Effect of the National Minimum Wage [online]. *The Economic Journal,* 114(494): 110–116. <http://doi.org/10.1111/j.0013-0133.2003.0020.x>.
Please **add DOI numbers** to all articles where appropriate (prescribed format = link, see above).

Tables
Provide each table on a separate page. Indicate position of the table by placing in the text "<u>insert Table 1 about here</u>". Number tables in the order of appearance Table 1, Table 2, etc. Each table should be titled (e.g. Table 1 Self-explanatory title). Refer to tables using their numbers (e.g. see Table 1, Table A1 in the Annex). Try to break one large table into several smaller tables, whenever possible. Separate thousands with a space (e.g. 1 528 000) and decimal points with a dot (e.g. 1.0). Specify the data source below the tables.

Figures
Figure is any graphical object other than table. Attach each figure as a separate file. Indicate position of the figure by placing in the text "<u>insert Figure 1 about here</u>". Number figures in the order of appearance Figure 1, Figure 2, etc. Each figure should be titled (e.g. Figure 1 Self-explanatory title). Refer to figures using their numbers (e.g. see Figure 1, Figure A1 in the Annex).

Figures should be accompanied by the *.xls, *.xlsx table with the source data. Please provide cartograms in the vector format. Other graphic objects should be provided in *.tif, *.jpg, *.eps formats. Do not supply low-resolution files optimized for the screen use. Specify the source below the figures.

Formulas
Formulas should be prepared in formula editor in the same text format (Times 12) as the main text and numbered.

Paper submission
Please email your papers in *.doc, *.docx or *.pdf formats to statistika.journal@czso.cz. All papers are subject to double-blind peer review procedure. Articles for the review process are accepted continuously and may contain tables and figures in the text (for final graphical typesetting must be supplied separately as specified in the instructions above). Please be informed about our Publication ethics rules (i.e. Authors responsibilities) published at: *<http://www.czso.cz/statistika_journal>*.

Managing Editor: Jiří Novotný
Phone: (+420) 274 054 299 | **fax:** (+420) 274 052 133
E-mail: statistika.journal@czso.cz | **web:** www.czso.cz/statistika_journal
Address: Czech Statistical Office | Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscription price (4 issues yearly)
CZK 66 per copy + postage.
Printed copies can be bought at the Publications Shop of the Czech Statistical Office (CZK 66 per copy).
Address: Na padesátém 81 | 100 82 Prague 10 | Czech Republic

Subscriptions and orders
Czech Statistical Office | Information Services
Na padesátém 81 | 100 82 Prague 10 | Czech Republic
Phone: (+420) 274 052 733, (+420) 274 052 783
E-mail: objednavky@czso.cz

Design: Toman Design
Layout: Ondřej Pazdera
Typesetting: Václav Adam
Print: Czech Statistical Office

103rd year of the series of professional statistics and economy journals of the State Statistical Service in the Czech Republic: ***Statistika*** (since 1964), ***Statistika a kontrola*** (1962–1963), ***Statistický obzor*** (1931–1961) and ***Československý statistický věstník*** (1920–1930).

Published by the Czech Statistical Office
ISSN 1804-8765 (Online)
ISSN 0322-788X (Print)
Reg. MK CR E 4684