CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Analysis of Czech Trade Structure Using the Zipf's Law

Bachelor's thesis

Author: Bc. Rufat Asadli

Study program: Economics and Finance

Supervisor: RNDr. Michal Červinka, Ph.D.

Year of defense: 2022

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, April 28, 2022

Rufat Asadli

Abstract

The Revealed Comparative Advantage (RCA) has been a cornerstone for cross-country trade analyses. Corresponding indices that quantify this phenomenon have long been subject to studies testing how far they are distributed from normal distribution. This thesis analyzes RCA indices alongside the Zipf's Law, which asserts that data points are linearly distributed with a negative relationship to their ranks. Trade data, covering 1998-2020, from the Czech Republic are initially utilized in a particular model. In this case, it is documented that RCA indices largely follow the Zipf's Law in log-log scale. Regression coefficients from this model, specifying exponent parameters of the distribution, are employed alongside a set of economic indicators in an advanced time series model, namely Vector Autoregression (VAR), to derive the economic inference of so-called power law exponents. The model reveals certain economic interpretations, including the fact that economic shocks might reduce comparative advantage of a country over a commodity.

JEL Classification C22, F10, F14, F17, F47
Keywords RCA Indices, Zipf's Law, Trade, Exports, The Czech Republic, OLS, VAR
Author's e-mail `rufat.aaa@gmail.com`
Supervisor's e-mail `michal.cervinka@fsv.cuni.cz`

Abstrakt

Index odhalené komparativní výhody (Revealed Comparative Advantage, RCA) je považován za základ analýz obchodu mezi zeměmi. Příslušné indexy jsou předmětem mnoha studií, které zkoumají nakolik je jejich empirické rozdělení daleko od normálního rozdělení. Tato práce zkoumá RCA indexy z pohledu Zipfova zákonu, který je založen na hypotéze, že data jsou rozdělena v klesajícím lineárním vztahu vzhledem k jejich ranku. Data o obchodu České republiky v letech 1998-2020 jsou nejprve použity pro odhad modelu, kdy jsou RCA indexy prokazatelně rozdělené v souladu se Zipfovým zákonem v log-log škále. Příslušné regresní koeficienty jsou následně použity společně s dalšími ekonomickými indikátory jako vstupy pro pokročilý ekonometrický model vektorové autoregrese (VAR) za účelem zkoumání vlivu takzvaných power law exponentů. Z odhadnutého modelu plynou ekonomické interpretace výsledků zahrnující fakt, že ekonomické šoky mohou snižovat komparativní výhodu země u příslušné komodity.

Klasifikace JEL	C22, F10, F14, F17, F47
Klicova slova	RCA indexy, Zipfův zákon, obchod, vývoz, Česká republika, OLS, VAR
E-mail autora	`rufat.aaa@gmail.com`
E-mail vedoucího práce	`michal.cervinka@fsv.cuni.cz`

Acknowledgments

I am wholeheartedly thankful to my supervisor RNDr. Michal Červinka, Ph.D. for his patient guidance, continuous support and comments that improved not only this thesis, but also my performance throughout my studies.

Further, special gratefulness to my family and friends, who as well persistently supported me in all difficult times. Without them, I would not have been successful at all.

Lastly, I would like to express my gratitude to Prof. Nizami Shikhaliyev, who carefully guided me when learning the deepest mathematical concepts during high school.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Asadli, Rufat: *Analysis of Czech Trade Structure Using the Zipf's Law.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2022, pages 47. Advisor: RNDr. Michal Červinka, Ph.D.

Contents

1 Introduction

Trade, from sufficiently ancient times, has admittedly been of an utmost importance for any civilization such that its complex structure is being explored for a long time. That being mentioned, scholars have long thought over the development of viable concepts that could provide a sustainable understanding of trade flows. Admittedly, such frameworks make it straightforward to comprehend the rationale behind incentives of two or more parties to engage in trade. Therefore, in today's ever-changing set of economies, it is necessary to update current views on trade with authentic empirical tools that are, indeed, captivating the heart of this study.

Unsurprisingly, trade, as a sector contributing to the economic prosperity, plays a crucial role for the main actor of this thesis – the Czech Republic – as well. As a member state of the European Union (EU), Ministry of Industry and Trade within the Czech Republic supervises the country's position both in the internal market of the EU and international trade arena, along with implementing the Czech and EU legislation and co-operating with other respective financial bodies of the Czech Republic. Although the COVID-19 breakout obscured the acceleration of international trade to a significant extent, the World Bank data show that exports of goods and services had a volume of almost three-fourths of the Czech GDP back in 2020. Nevertheless, a recent publication from Ministry of Industry and Trade mentions that according to the preliminary data from Czech Statistical Office (CZSO, 2021), the size of both exports and imports has reached a record volume of CZK 4 billion last year. In contrast to those figures of 2019, exports and imports have shown an annual increase of 13% and 6.6%, respectively, in 2020; meanwhile, these figures positioned at the corresponding 19.2% and 11% last year. It merely shows that the Czech Republic, much like any other economy once suffering from the pandemic, is gradually gaining the pace it used to have not-so-long before and thus, positive indicators are expected for the Czech economy in 2022, as delivered by Ministry of Industry and Trade as well.

Since the above facts show how important trade is for the Czech Republic, this thesis reminds the essence of digging deeper and exploring the nature of trade via useful empirical tools. Hence, a concept called Revealed Comparative Advantage (RCA), which is a cornerstone in trade theory, is covered through its own corresponding index for empirical purposes. In essence, this work contributes to the existing literature by bringing the RCA index together with an empirical law, namely the Zipf's Law. The reasoning behind this law suggests that certain data sets follow a rank-size distribution that bears a negative linear relationship, stemming from log-transformation of $y = Cx^{-\alpha}$ $(C > 0, C \in \mathbb{R})$. In this association between arbitrary y (rank) and x (size), the Zipf's Law hypothesizes that power law exponent α is equal to 1. That is why the variables are said to have a negative linear relationship in log-log scale as well in econometric analyses. Thus, the main aim of this thesis is to test how far RCA indices are distributed from a straight line with a negative unit slope. To perform this analysis, trade data, covering a 23-year period (1998-2020), are utilized through a specific model that is inspired from a previous study. This model is expected to return regression coefficients that are, in turn, slopes as well. The coefficients are referred as power law exponents too, which will be compared to the unit value that is proposed by the Zipf's Law. Results from the first phase are further employed within a time series model, that is Vector Autoregression (VAR), to explore the economic meaning of so-called power law exponents along with a group of economic indicators. This thesis has an academic significance, just as it is the second such work that examines the aforementioned problem which is, at the same time, analyzed with the data from the Czech Republic for the first time ever.

This thesis is structured as follows. The research commences with an in-depth overview of the existing literature devoted to the previous analyses of trade flows, trade structure of the Czech Republic, Zipf's Law, and Revealed Comparative Advantage. The next chapter gives its readers a clearer insight about the systematic structure of trade flows among the Czech Republic and other respective partners across the world. In Chapter 4, this thesis is mainly backed with the relevant theoretical narrative that focuses on detailed historical exploration of both RCA and Zipf's Law. Once the theoretical foundations are established, the following chapter covers all the variables, adding up to the whole data set, and model specifications that further establish the empirical part. Furthermore, descriptive statistics, derived results and, possibly, some limitations are discussed in Chapter 6. At the end, the very last chapter concludes the work with a final review of contributions made to the field by both quantitative and qualitative outcomes.

2 Literature Review

This part of the thesis is dedicated to studies that have been carried out previously and provides an examination of trade structure in the Czech Republic as a whole. Besides, it also stresses and compares other worth-mentioning economic frameworks, including RCA, that have been priorly used to analyze the very same problem. More importantly, a throwback to the initial introduction of the Zipf's Law is mentioned and its association with trade theory is clarified as well.

It is worth pointing out that there are various studies devoted to the analysis of trade in the Czech Republic, which has been a subject to not only individual investigations, but also regional cases including the EU, Central Eastern Europe (CEE), and overseas relations (e.g., the USA, China, etc.) as well. One of such works is published by Fitzová and Žídek (2015) who describe historical transition from Czechoslovak to Czech economy in terms of exports and imports structures. In their work, they show that trade in Czechoslovakia had a relatively small effect on the whole economy such that its turnover to GDP was only around 40%. Moreover, majority of trade flows was centralized around the Eastern Bloc, while relations with the West were quite limited. Nevertheless, the whole scheme drastically changes after 1989, covering three stages as stated by Janda and Munich (2004). To exemplify, they describe the situation as a transition from traditional export markets to a more liberally structured form, resulting in trade balance and also lower trade deficit. As a matter of fact, the exchange rate had been stabilized at 30 CZK/USD until 1997, embracing a protective measure for the Czech market. However, the Czech koruna started to appreciate through rather volatile trend after the period of accession to the EU (Tichý, 2007).

Meanwhile, by the time domestic demand was increasing, imports went up by over 20%, whereas the growth rate of exports was a mere 4%. Yet, facing a fine growth in FDI, production performance in terms of exports improved as well and further led to a larger GDP; in fact, the per capita inward FDI is in line with the projected convergence in exports figures. Likewise, Jakab et al. (2001) suggest that the Czech Republic experienced a faster convergence of exports than imports relative to the EU in the CEE region in the late 1990's. In cumulative figures, the Czech Republic enjoyed a volume of US$ 21.68 billion FDI after the well-known transition period (Rovna, 2011). In modern period, a relatively recent data from the World Bank (2019) exhibit that the Czech Republic experienced a positive trade balance of almost US$ 20 billion during the pre-pandemic times. Even though there is a negative trade growth of -2% approximately, both exports and imports of goods and services were as large as about 70% of GDP. The difference between service exports and imports was about US$ 5 billion in Balance of Payments (BoP). Noteworthily, Germany is the biggest bidirectional trade partner of the Czech Republic and it is followed by other EU member states such as Slovakia, Poland, and Italy. In this context, Šimǎkovǎ (2014) considers an analysis of the Czech Republic with its eight biggest trade partners through the J-curve method and examines the effect of exchange rate volatility on trade flows.

The previous moderate volume of works briefly covered the economic review of the Czech Republic. Now the main interest shifts to examples of specific cases where the concept of RCA is employed. In practice, this phenomenon has been of interest in the analysis of numerous countries, including the Czech Republic. The RCA index, after its appearance in economics (Balassa, 1965), has been subject to several reports by world-recognized organizations – World Bank, OECD, and UN Industrial Development Organization (UNIDO), for example. Since it is a well-studied field in economics, there are roughly 7 versions of comparative advantage based indices that have been practiced in real-world cases (Bebek, 2017). To illustrate, there can be named numerous publications which compare these indices from distinct standpoints (Aquino, 1981; Soete & Wyatt, 1983; Hinloopen & van Marrewijk, 2008). Specifically, Batra and Khan (2005) focus on the investigation of RCA in China and India, calculating the index for over 4000 commodities. Subsequently, Ma (2013) takes ASEAN-China trade flows into account, once again with symmetric RCA indices. A relatively older article from Kaitila (1999) makes use of RCA to describe trade flows between the EU and a couple of CEE countries: the Czech Republic and Hungary. A rather recent article by Kuzmenko et al. (2022) addresses how the Czech Republic performs in agricultural sector using a pool of empirical indices such as RCA Index, Lafay Index (Lafay, 1992), and Grubel-Llyod Index (Grubel & Lloyd, 1971). Although Nowak et al. (2020) show that none of the more recent members developed their competitiveness in agriculture to be compatible with old member states of the EU, the Czech Republic outperformed top agriculture actors (e.g., the Netherlands, Denmark, Spain) in production of certain commodities. While on the other hand, there is lack of comparative advantage in products like meat and meat preparations, vegetables and fruits, and animal oils and fats.

When it comes to the so-called Zipf's Law, it has been introduced by Zipf (1949) as an empirical law, which proposed that rank-frequency distribution constitutes a negative relationship. Although its primary application was introduced in linguistics by the author himself, the follow-up research revealed that it has, much like some other power laws, certain implications in other fields too (Simon, 1955). Certainly, its mathematical roots, which suffice to create linkages to economics and finance, originate from rationale behind the Zeta function. This law brings new insights each time as more works are contributing to the field such that there are well-exhibited examples of distributions of cities, firm-sizes, salaries (Samuels, 1965; Reed, 2001; Fujiwara et al., 2004). The research, conducted by Gabaix (1999), reveals that the relationship between countries and their populations follow up a fit which is strikingly similar to the one proposed by the Zipf's Law. The slope of this fit is estimated to be -1 approximately, which is also the power law exponent with opposite sign. Unbiased estimations for the power law have been developed by Gabaix and Ibragimov (2011) with a basic yet effective OLS model. Gabaix (2009), once again, in his paper opens a new outlook for the power laws, which, in fact, align with patterns in stock market returns, international trade, and income and wealth. In trade theory, although there are not so many example works, Josic and Nikic (2013) scrutinize the linkage between Tinbergen's Gravity Model and Zipf's Law. Besides, there is only a single study (Hinloopen & van Marrewijk, 2006), which is carried out to explore RCA indices with the Zipf's Law. Their work with massive volume of data reveals that RCA indices nicely align with the Zipf's Law, depending on the choice of sample country.

The previously mentioned analyses, additionally those which are subsequently referenced, clearly contribute to the context of this thesis. Nevertheless, neither of the literature above adequately questioned the puzzle between RCA indices and Zipf's Law, and there is obviously a sufficient opportunity for improvements. Therefore, this thesis is the second work in the field in general, but the first ever work to conduct a country-specific investigation of RCA indices with regard to the Zipf's Law. Thus, this study might be inspirational for the future research in this field.

3 International Trade in Czech Economy

This chapter broadens the previously established knowledge of place of the Czech Republic in international trade with respect to its regional and global partners. In other words, it covers shape of trade relations of the Czech economy. Last but not least, association of trade with several economic variables is captured as well, which will later be useful.

3.1 Trade with Partners

The Czech Republic is an open emerging economy that had a satisfactory economic expansion around 2.5% on average throughout the previous decade, in accordance with the World Bank database. International trade covers an integral part of the Czech economy, which has become a member of the General Agreement on Tariffs and Trade (GATT) and World Trade Organization (WTO) in 1993 and 1995, respectively; the partnership in international organizations undoubtedly encouraged the pace of involvement in trade and thus, turned the Czech Republic into an export-driven economy. More importantly, a more significant milestone was when the Czech Republic realized the importance of upgrading the integration process into European continent. Accession to the EU officially took place as a part of enlargement process with 10 countries in total as of May 1, 2004. In general, such an event was so remarkable for its agenda, just as former Czechoslovakia had completed most of its trade with the Eastern Bloc; thus the integration into the West could then mean a faster economic convergence for the Czech Republic. In practice, it meets a couple of conditions of the Maastricht Convergence criteria, which is a criterion examining the compatibility with the EU-level economies (European Commission, 2020). Yet the Czech Republic has been reluctant to adopt Euro as the national currency ever since becoming an official EU member. Even though Optimum Currency Area (OCA), as defended by Mundell (1961), offers an economic efficiency maximization with a single currency, there has long been “euroskepticism” in the Czech Republic. Dědek (2004) foresaw potential unpredictable risks of embracing Euro for the Czech Republic and pointed out risks such as correspondence of catch-up effects between productivity of labor and stabilized price levels. The unwilling approach towards Euro has rooted from the political economy of the country as well, in addition to structure of economic relationships with neighboring fellow member states. For instance, Dandashly and Verdun (2015) argue that it was favorable for Slovakia to accept Euro to tie-up with Germany, whereas the Czechs were indifferent on this matter. It is largely because Germany was already the biggest trade partner of the Czech Republic. A decade ago, Germany accounted for almost one-third of foreign trade turnover of the Czech Republic (CNB, 2011). During these times around half of total exports to Germany were of machinery and transport equipment, constituting of EUR 30 billion after the 2007-2009 downturn. Recent data from the UN Comtrade indicate that the volume of exports to Germany has reached approximately EUR 65 billion as of 2021, surging notably from EUR 55 billion after the pandemic. As a matter of fact, machinery share the main attention with a size of EUR 15 billion along with electrical equipment and vehicles such as tramway. Subsequently, roughly 8% of trade turnover has been completed with Slovakia, counting for over EUR 15 billion in 2021. Much like Germany, Slovakia imports machinery and other equipment primarily from the Czech Republic. In a similar fashion, the export structure is followed by goods such as mineral fuels, iron, steel, pharmaceuticals, and furniture. Further there comes Poland, France, and Austria with trade shares ranging between 4.5%-6.5%. Not surprisingly, the structure of exports to these member states is almost unchanged, merely signifying the fields of expertise of the Czech industry. Other than the fellow EU members, which cover 88% of export volume, the United States ranks 10th as one of the main trade partners of the Czech Republic (International Trade Administration, 2022). Just as the pandemic slowed down with mild effects, exports to the U.S. increased to slightly more than US$ 5 billion in the last year. Although the COVID-19 breakout halted trade of services such as transport and travel, the adverse effects were mainly recovered by services of financial and IT sectors together with automotive and machinery. The magnitude of the U.S. FDI in the Czech economy weighted the same as trade value, while the Czech FDI in the U.S. was a modest US$ 212 million as of 2020. In overseas trade, China is the next prominent partner with a share of a tiny 1.4% or, equivalently, EUR 3 billion within total turnover. In contrast, China has a bigger role in the Czech market with a huge trade surplus that has further widened due to monetary interventions of the CNB (De Castro et al., 2017). However, the value added records outweigh the gross trade deficit, mainly because of future re-export of Chinese imports. Last but not least, China is followed by Turkey and Russia, which seize a little share slightly above 1% only.

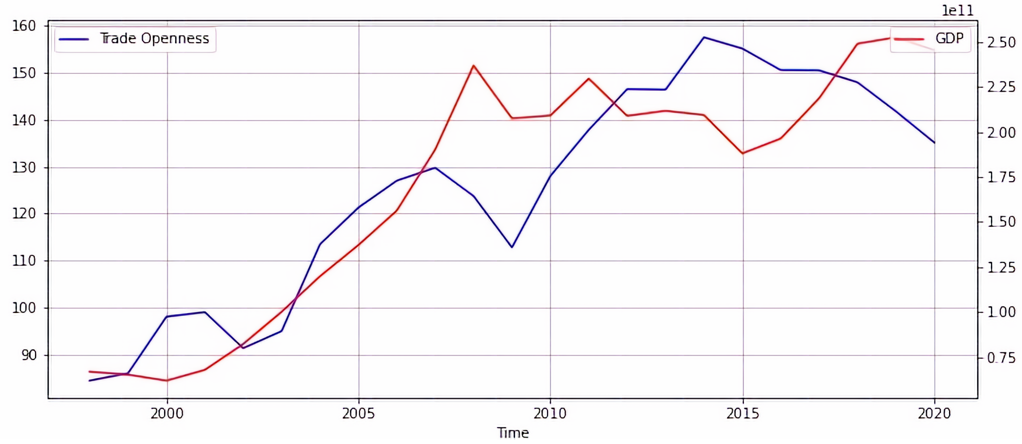
3.2 Trade and Macroeconomic Variables

This subsection serves more of an analytically descriptive role by evoking the association between trade and vital economic indicators which are GDP, FDI, exchange rate, and inflation in the Czech Republic. To scale trade in physical quantities, trade openness (ratio of exports plus imports over GDP) is being applied. Last but not least, prior research is supported by real-world data visualizations that are constructed for the course of this thesis. Relatedly, the annual data for these depictions are collected from World Bank's World Development Indicator (WDI). It should be noted that these plots comment on development of indicators only and does not assert any decisive linkages between variables, which should be and are further analyzed through proper econometric methodologies.

Trade Openness and GDP

Gross Domestic Product is often considered the utmost representation of economy of any country. In one of the basic approaches, GDP is defined as a sum of Consumption, Investment, Government Expenditures, and Net Exports. Thus, it is obvious to infer a relationship between trade and GDP to some, for now, unknown effect to an extent. The effect, to say, has been ambiguous in some of the literature. On the one hand, Frankel and Romer (1999) proposed a positive yet tiny association between trade openness and GDP. In contrast to well-known exogenous growth models, specifically Solow's (1957) model, endogenous growth models defend a positive relationship between these two indicators (Fatima et al., 2020). To exemplify, Romer (1990) finds trade openness to be a technological and innovative driver of the economic growth, as a result of cross-country study covering 90 economies. Another paper by Sachs et al. (1995) investigates more than 100 countries and suggests that outward-oriented states perform much better in terms of GDP growth. Yet, as mentioned before, the vague nature of this linkage is reflected as a difficult milestone in some studies, such as that of Greenaway et al. (2002), which point out a positive relationship in the presence of lagged values; whereas some others stress the same situation on some conditions such as high capital accumulation and per capita income, for instance (Fetahi-Vehapi et al., 2015). A very brief case study of this thesis documents trade openness and GDP (in hundred bln) data of the Czech Republic and compares them in a single graph as below, trying to unearth a possible nexus. It is noteworthy that GDP mostly follows a trend in the opposite direction of trade openness, although both of them similarly increase around 2005.

Figure 1: Comparison of Czech GDP and Trade Openness (1998-2020)

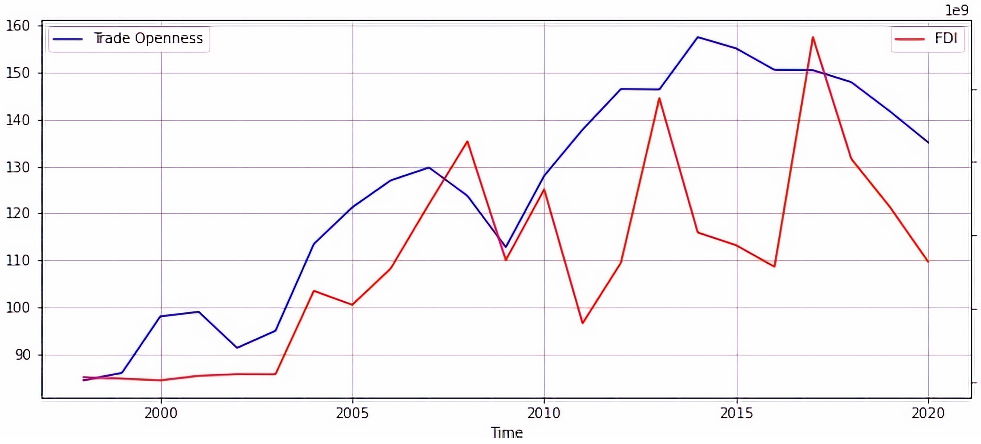


Trade Openness and FDI

Foreign Direct Investment is another prominent economic phenomenon indicating cross-border investment flows which create strong links among international economies and thus stimulates economic integration, as defined by OECD. Unlike GDP, FDI is apparently connected to trade openness in a less complicated manner. Yet, there is still a number of studies that portray a negative association, which is more or less unusual, depending on the choice of country. Prior research by Zaman et al. (2018) prevails that high trade openness significantly increases FDI both in short-/ and long-run, as suggested by their study on Iran, India, and Pakistan. Another study, that focuses on Vietnamese economy through a VAR model, similarly concludes that trade openness has a huge positive impact on FDI (Lien, 2021). Akin to these studies, Liargovas and Skandalis (2012) demonstrate that trade openness can significantly explain FDI with a positive trend in developing economies across the world, mainly from Asia and Eastern Europe, as a result of panel data analysis.

On the other hand, Brun and Gnangnon (2017) illustrate a negative relationship for Least Developed Countries (LDC), whereas a positive impact is visible for non-LDCs. For this thesis, the following Figure 2 depicts how trade openness and FDI (in bln) evolve with respect to each other over time in the Czech Republic. This comparison is not totally transparent, since FDI experiences a rather volatile development with ups and downs, specifically, after the Global Financial Crisis around 2007. Yet, all in all, it is possible to reveal periods when they both run parallel with similar growths or reductions.

Figure 2: Comparison of Czech FDI and Trade Openness (1998-2020)

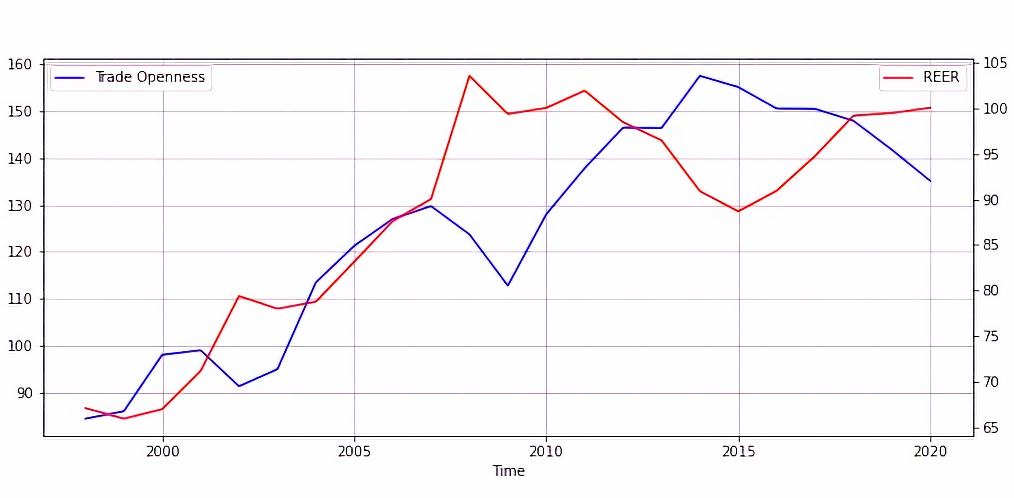


Trade Openness and Exchange Rate

The World Bank conceptualizes the Real Effective Exchange Rate (REER) as follows: *"Real effective exchange rate is the nominal effective exchange rate (a measure of the value of a currency against a weighted average of several foreign currencies) divided by a price deflator or index of costs."*

This metric is an index that is normally fixed at 100 and any value above this border translates into appreciation of the national currency. Remarkably, previous works affirm an overlapping agreement regarding the correspondence between the REER and Trade Openness in a less complicated manner. With respect to trade theory, classical models by authors (Dornbusch, 1974) long asserted that trade liberalization overshadows protectionist measures and thus causes depreciation of the REER. Following the same logic, Balassa (1975) explains that a cutoff in import tariffs stirs up demand for imports; that, indeed, simply depreciates the REER. Gantman and Dabós (2018) support the existent hypotheses and conclude that trade openness has a negative impact on the REER through robust empirical results. Likewise, a rather older paper by Hau (2002) holds the same view such that openness is inversely related to exchange rate and, therefore, open economies designate more flexible price adjustments and exchange rate regimes. Lastly, the situation for the Czech Republic is illustrated hereunder by the respective graph. Among all comparisons, the association between trade openness and REER might be the clearest in a sense that they are heading to opposite directions almost at any point of time. Such an insight is noteworthily the most obvious around 2010 and 2015. At the end, around 2018, most probably, REER experiences a tiny increase to above 100, whereas trade openness drops to approximately 140.

Figure 3: Comparison of Czech REER and Trade Openness (1998-2020)

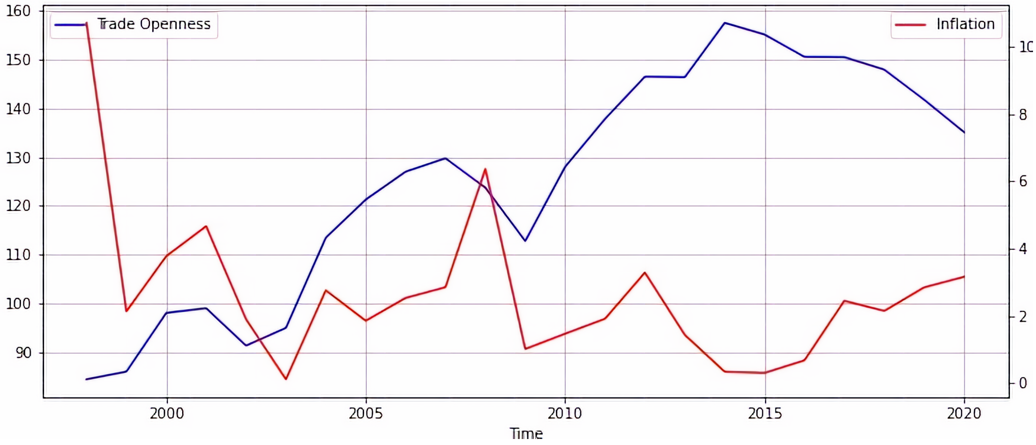


Trade Openness and Inflation

Inflation is admittedly a non-negligible indicator of stability for any economy, just as it gauges how expensive goods and services become over some period. Although several scholars reach consensus on a negative relationship between trade openness and inflation, the studies over this bidirectional association are open to debate, mainly due to choice of sample country. Firstly, Romer (1993) hypothesizes an inverse relationship that stems from stronger diversification that lowers chances of price shocks and thus aggregate inflation. Similarly, in their analysis of OECD countries, Bowdler and Nunziata (2006) reveal a negative relationship between trade openness and inflation. The main reasons for that are insufficient incentives for allowing expansionary policies and fierce competition that prevents firms from pushing prices up, as supported by the probit regressions for OECD data.

Nevertheless, Munir and Kiani (2011) reject Romer's conjecture and empirically demonstrate how trade openness has a positive impact on inflation in Pakistan. From a different point of view, Binici et al. (2012) test Romer's hypothesis and claim that there is not a statistically significant relationship between these two variables. The corresponding Figure 4 is added below to describe a mere visualization for this thesis. It can be commented that both indicators follow the same pattern occasionally over the timeframe, or around 2000, 2005, and 2010, to be exact. Yet, a possible negative relationship is the most obvious starting from 2012, when trade openness reaches a spike slightly above 150. Meanwhile inflation goes down considerably until 2016, before gradually rising to about 3%, whereas the former decreased to a rate between 130-140.

Figure 4: Comparison of Czech Inflation and Trade Openness (1998-2020)



4 Theoretical Background

This chapter dives into an in-depth investigation of contributions made by several trade concepts, especially that of RCA, throughout the history. Furthermore, it stresses the mathematical convenience, without diverging too much from economic core, behind the Zipf's Law as well.

4.1 Ricardian Model of Trade and Beyond

When analyzing bilateral trade flows between any countries, it is quite essential to better grasp how relations are being shaped. In that sense, a view of the historical development of trade concepts is quite helpful. Therefore, the throwback to the history of trade can commence with the theory of comparative advantage. Being introduced by Ricardo (1817), this concept is one of the building blocks of early economics. In general, it expands the primal understanding of scarce resources in terms of production and makes logical implementations to decision-making incentives of nations while trading back and forth with each other. In his vital work, Ricardo's classical theory of comparative advantage asks fundamental questions regarding why nations should trade at all and how basic factors (labor and capital) make these nations interdependent and eventually drives the whole system to an engagement in trade. Although his theory was exposed to some constraints (static labor and capital, perfect competition, etc.), it elegantly pioneered in providing initial yet intuitive insights in International Economics. This model, in its core, carries out the idea that if a country has (relatively) more productivity in particular industries, then it is better-off when it produces and then exports more of these industries (Leromain & Orefice, 2014). The official definition from UN Conference on Trade and Development (UNCTAD) similarly indicates that trade between nations is dominantly governed by distinct relative features in their productivity, definitely in accordance with the RCA index.

In the following years, such innovative leaps led to slightly more complex models like Heckscher-Ohlin by Eli Heckscher and Bertil Ohlin who proposed that endowment of production factors affects specialization in certain sectors such that countries export goods that they can abundantly produce with optimal performance (Deb & Hauk, 2017). Going back to the rationale behind it, the theory reminds that geographical position of nations had distributed scarce resources unevenly and thus creates commercial incentives (Leamer, 1995). The original model, containing 2 countries together with 2 products and 2 homogeneous production factors, is being called the "$2 \times 2 \times 2$" model in some literature as well. It is also worth pointing out that the H-O model have continuously been a part of further more developed frameworks; as such, Stolper and Samuelson (1941) demonstrated that changes in commodity prices can influence prices of factors of production both in a positive and negative way. For instance, if the world price of a capital-intensive commodity increases, price of rental rate increases too, whereas price of wage rate decreases (and vice versa). Consequently, a decade later Rybczinski (1955) assumed that the overall production of a certain good goes up, if amount of a particular factor of production within that good increases as well. Although the H-O model added up to the general improvement to a significant extent, there are indeed certain critiques arising from both theoretical assumptions and further empirical findings. First of all, the H-O model relies on a set of strong assumptions to create an economically cultivating system based on a production cycle of similar goods with similar factors through similar technological basis (Lancaster, 1957). At the same time, their production is characterized by similar production functions (Cobb-Douglas) with constant returns to scale (CRS) – a concept which implies that the resulting output is multiplied by an arbitrary k in case both factors of production are multiplied by the very same k. The dimension of this unrealistic approach intensifies by the property of perfect (internal) market too. Second of all, the critique of H-O was taken to another level by the time when Leontief (1954) proposed findings that built up a strong case against the model; in fact, his empirical results, referred as Leontief Paradox, demonstrated that U.S. exports were less capital intensive than its imports in 1947. In spite of Leamer's (1980) attempts to rejuvenate the theory once again, years later it was put forward that there is not a remarkable linkage between factor engagement in trade and factor supply (Bowen et al., 1987).

Much similar to Ricardian model, the H-O was an inter-industry framework – trade of commodities from different industries – that was unrealistic yet a seed of more relevant future models (Davis, 1995). Particularly, the trade theory entered a new orientation by the time Krugman's paper (1979) was published. Even though the core idea behind Krugman's innovation was a mature version of the H-O, Krugman (2009) also stresses that (internal) economies of scale comes together with imperfect competition, which was impossible to be captured by general-equilibrium models during Ohlin's times. Hence, Krugman basically introduced a simple and rigorous depiction of intra-industry trade in a well-determined general-equilibrium sphere (Neary, 2009). Nevertheless, it is also essential to recall how Krugman could rely on imperfect competition; he was clearly inspired by one of the early works of Dixit and Stiglitz (1977), who pioneered in development of Monopolistic Competition. Such a new scheme brought diversity into production together with increasing returns to scale (IRS), which, as a whole, explained "old" trade theory as a profitable specialization. It could then guide scholars to produce more sophisticated models such as New Economic Geography, which resemble real-world situations effectively. It is undeniable that the systematic analysis of frameworks from trade theory is really deep, yet it is of investigation that deviates from the main interest of this thesis. Therefore, the volume of baseline knowledge suffices to return back to the Revealed Comparative Advantage and respective indices. In the famous theory of comparative advantage, specialization is basically a more intense focus of a country on a specific production area in comparison to others (Laursen, 2015). Thus, the Revealed Comparative Advantage simply stands for a phenomenon describing a country that is mastering and then focusing on well-defined sector(s) in terms of exports. This mechanism is first put forward by Balassa (1965), being clearly inspired by Ricardian model of trade. After installation of such a mechanism into economics, Balassa and Noland (1989) utilize RCA for exploring situation for twenty commodity aggregates in Japan and the United States. In their work the authors provide evidence that Japan had a comparative advantage in unskilled-labor intensive products (e.g., textile, rubber, leather, etc.). While on the other hand, they detected comparative advantage in primary, physical-/and capital-intensive products in case of the U.S. In their research, they formulate RCA indices for a commodity i with physical export volume of X_{ij} as follows

$$RCA_{ij} = \frac{X_{ij} / \sum_i X_{ij}}{\sum_j X_{ij} / \sum_i \sum_j X_{ij}}, \tag{1}$$

where the denominator quantifies the worldwide share of exports of the given product i $\left(\sum_j X_{ij}\right)$ among all sectors across all countries $\left(\sum_i \sum_j X_{ij}\right)$. Meanwhile, the numerator describes percentage share of the commodity i with physical volume X_{ij} in total exports of country j $(\sum_i X_{ij})$. The country j is said to have a comparative advantage of the product i, if the respective RCA index is larger than 1. In other words, it means that the country has a greater share of the commodity within its internal market than its share of all goods in the world market of exports. In case of the index equals to 1, the country is assumed to have a neutral comparative advantage of the commodity. As it is obvious from the foundation, the RCA index has a zero lower bound, whereas it is unbounded from the upper side; this means that the original RCA_{ij} from (1) is asymmetric Thus, RCA indices cannot be viewed as ordinal data with observations that are ranked based on comparison of physical values. That is why Vollrath (1991) brought logarithmic version into action, but it is not a suitable operation when being implemented to exports with zero magnitude. Instead, it is offered to standardize the RCA index such that it takes values between -1 and $+1$, and assumes 0 to be the point of symmetry and neutrality (Yu et al., 2009)

$$SRCA_{ij} = \frac{RCA_{ij} - 1}{RCA_{ij} + 1}.$$

The above equation acts as a normalized index as well such that any index greater than zero translates into a pattern of comparative advantage and vice versa for the index less than zero. Nevertheless, note that any version of RCA indices should be understood as a relative position of a country, not the absolute one (French, 2017). Simultaneously, the RCA index does not take into account tariff and non-tariff barriers, which is of another deficiency of the conception. Even so, its position in modern economics in undeniable. At the same time, there are other numerous representations of trade specialization such as Michaely Index and Net Trade Index that have been later developed. It is worth pointing out that such concepts improved the unbalanced distribution of indices to a significant extent. Therefore, unlike these normalized indices, the original RCA indices are still open to more questions regarding their distribution. That is why this thesis employs the conventional version of RCA index from (1) for empirical purposes in the following chapters. Yet, before moving onward to rather econometric stage of this thesis, below the required mathematical foundation is covered first.

4.2 Mathematics of the Zipf's Law

To recall, the Zipf's Law was proposed by Harvard scholar Kingsley Zipf. Although the law was empirically tested based on experiments in quantitative linguistics, there is a complex mathematical foundation behind it. In statistics, the Zipf's Law is observed through one of power law distributions. Yet, this distribution itself originates from Zeta function that is an integral part of the number theory with voluminous implications. Therefore, this section commences with a definition of the Zeta function, which is then explored through some of its essential and interesting features. Next, the section builds upon its links to statistics, wherein Zeta function is, in fact, described as a distribution. It is further followed by a couple of more distributions which are focal points of this section, contributing to the rationale behind the Zipf's Law.

Around 1730s, the world-famous mathematician Leonhard Euler (1707-1783) starts to develop an infinite series called the ζ function (Greek letter Zeta), which further proved to be a huge development for the analysis of prime numbers and their distribution. In 1749, he precisely formulates the final version of the equation

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + ..., \quad s > 1. \tag{2}$$

Note that the above formulation was initially proposed for real input values s only. Moreover, Euler approximates $\zeta(2)$ and later $\zeta(2n)$ $(n \geq 1,\ n \in \mathbb{N})$ for Zeta function, as shown by Ayoub (1974). Next, he creates a link between primes and Zeta function, following the order of mathematical operations. The process follows the multiplication of the original Zeta function by $1/2^s$, which is then subtracted from the original Zeta function itself. All elements in the series on the right-hand side of (2), which are multiples of 2 (prime), disappear as a result of this operation. If one firmly continues to apply this process with the remaining prime numbers one by one, they will remove all numbers from the right-hand side of the equation. Thus, it is possible to create the following link between Zeta function and primes

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}} = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s > 1.$$

Afterwards, Bernhard Riemann expands Zeta function into the complex plane $\mathbb{C}$ through the above equation. Thus, we can replace the restriction to $s > 1$ with $\Re(s) > 1$, where $\Re(s)$ denotes the real part of a complex number. Further works in the field reveal that the real part can attain negative integers as well. In its expanded functional form, Zeta function can be described with π, and sine and Gamma (Γ) functions. Thus, Zeta function is said to have zeros at $s = -2n \ (n \in \mathbb{N})$, which are called trivial zeros too. The trivial zeros are fairly easy to prove, since $\zeta(s)$ automatically attains zero. What has been more interesting to scholars was the non-trivial roots of Zeta function, just as discussion over them has an intriguing history. Thus, in this context, Riemann later proposes a conjecture regarding Zeta function that is hypothesized to have non-trivial zeros at complex numbers with real part equal to one-half. In other words, the non-trivial zeros are distributed across a line (critical strip) which contains points such that $\zeta\left(\frac{1}{2} + it\right) = 0$, where $t \in \mathbb{R}$ and i is imaginary unit. Notably, a tremendous volume of studies have been conducted since then, targeting to solve the paradigm of critical strip that consists of non-trivial zeros. By the age of computers, already in 2004 scholars (Wang, 2021) were able to compute the first 10^{13} zeros of the Zeta function. Thus, there is almost a unanimous agreement regarding correctness of the Riemann Hypothesis, even though a solid mathematical proof is yet to be discussed.

The mathematics presented above helps to create basic prerequisites for this study. In this regard, it is vital to continue with a follow-up introduction to implications of Zeta function in statistics, since it is certainly those statistical distributions which create proper linkages with economics. Hence, the context will mainly focus on statistical theory from now on with extended implications of Zeta function. In statistics, Zeta function helps to generate Zeta distribution which is a discrete probability distribution. Therefore, there is a special probability mass function f (pmf) for it. In the equation (3) for f, X is a random variable with Zeta distribution for integer $k \geq 1$ (Lin & Hu, 2001)

$$f(k; s) = P(X = k) = \frac{1}{\zeta(s)} k^{-s}, \quad k = 1, 2, \dots\ . \tag{3}$$

In the above equation for Zeta distribution, it is worth pointing out that the requirement of real $s > 1$ holds true as before. In terms of its statistical features, Zeta distribution has a mean of $\zeta(s-1)/\zeta(s)$ (for $s > 2$), whereas the variance is $\left(\zeta(s)\zeta(s-2) - \zeta(s-1)^2\right)/\zeta(s)^2$ (for $s > 3$). The core idea of Kingsley Zipf, as an empirical law, is reflected in statistics via the Zipfian distribution, which is the previously specified power law distribution. In its core, the Zipfian distribution has a mathematical formulation that is quite similar to that of Zeta distribution. The Zipfian distribution is mathematically defined as a probability distribution function as below

$$f(k; s, N) = \frac{1}{H_{N,s}} k^{-s}, \quad k = 1, 2, ..., N. \tag{4}$$

where s is power law exponent and $H_{N,s} = \sum_{n=1}^{N} \frac{1}{n^s}$ is called N-th harmonic number. In comparison to Zeta distribution, $\zeta(s)$ is replaced with $H_{N,s}$ in Zipfian distribution. Therefore, the Zipfian distribution is said to be asymptotically equal to Zeta distribution when $N \longrightarrow \infty$. It is noteworthy that the previous restriction for s changes to $s \geq 0,\ s \in \mathbb{R}$ for Zipfian distribution. Also, note that the verbal formulation of the pmf in (4) incorporates that the formula returns the frequency of the element with rank k out of the population of N elements. Hence, the above formulation is the original realization of the Zipf's Law. Yet, it is not the only such formulation of the Zipf's Law.

At this point, it is worthwhile to mention another distribution called Yule-Simon that indirectly contributes to the logic of Zipf's Law too. Its initial derivation is firstly given by J.C. Willis and G.U. Yule, who discovered that distributions of size of biological species follow a power law; it is further put into a proper mathematical shape by Yule (1925) himself. His empirical work is eventually formed as a probability distribution function by Simon (1955). The following equation (5) represents the pmf of so-called Yule-Simon distribution

$$f(k; \rho) = \rho B(k, \rho + 1), \quad k = 1, 2, \ldots , \tag{5}$$

where $B(k, \rho + 1)$ is Beta function with the integer $k \geq 1$ and real parameter $\rho > 0$. Relatedly, it is also possible to describe the Yule-Simon distribution with the Gamma (Γ) function, which can be derived from Beta (B) function as below

$$B(k, \rho + 1) = \int_0^1 \lambda^{k-1} (1 - \lambda)^{\rho} \, \mathrm{d}\lambda = \frac{\Gamma(k)\Gamma(\rho + 1)}{\Gamma(k + \rho + 1)}, \quad k > 0,\ 0 < \rho < \infty.$$

The above relationship can further be simplified when properties of Gamma function are taken into account. In his lemma Titchmarsh (1939) shows that for a real constant a and $x \longrightarrow \infty$, Gamma function has the following asymptotic behavior, which is used for Yule-Simon distribution as well

$$\frac{\Gamma(x)}{\Gamma(x + a)} \sim x^{-a}.$$

Therefore, the following approximation, which is another realization of Yule-Simon distribution, is consistent as well, being recently studied by Baur and Bertoin (2021) too. It can be inferred that the Yule-Simon distribution is another realization of the Zipfian distribution for $k \longrightarrow \infty$

$$B(k, \rho + 1) \sim \Gamma(\rho + 1) k^{-(\rho+1)}, \quad k \longrightarrow \infty.$$

One can confirm that the above representation of Yule-Simon distribution is similar to a power law distribution. Thus, the Yule-Simon distribution influences formulation of the Zipfian distribution on its own and thus, indirectly establishes reformulation of the Zipf's Law. Therefore, Zipfian and Yule-Simon distributions create a premise for the Zipf's Law. The above theoretical basis obviously suggests that the family of discrete power law probability distributions is fairly complex. Nevertheless, the current content of this subsection suffices to understand the Zipf's Law. Henceforth, it is no longer a challenge to draw a more accurate econometric bridge to economics and, specifically, RCA indices.

To elaborate on that bridge, the main aim of this thesis, once again to remind, is to check, if RCA indices are distributed as suggested by the Zipf's Law with a unit power law exponent s. Even though it is an extremely narrow pinpoint in economics, there are certain works focusing on distribution of RCA indices, in general. It is due to the fact that the ordinal features in the analyses of RCA indices have always been of interest in previous research. Despite the fact that this thesis attempts to explain RCA indices via the Zipf's Law, scholars have mostly examined whether RCA indices are normally distributed, which seemed favorable in practice. This is largely because, without normality, RCA indices are not of ordered data with ordinal ranking and, therefore, possess asymmetric shape with an unbalanced mean, as mentioned in the Section 4.1 as well. With this regard, there is also a study conducted by Liu and Gao (2022) who emphasize the issue of normality. Their empirical work is based on one-sample Kolmogorov-Smirnov (K-S) test that is a common non-parametric test used for comparing a sample with respect to a reference distribution. Following the very same logic, they compare a sample of RCA indices to normal distribution, which is indeed the reference distribution. They employ a unique parameter called Deviation from Gaussianity (DfG) which is the difference between the K-S statistic for the specified cumulative distribution function (cdf) and critical value of the K-S test. The following equations summarize the whole process, wherein the equation (6) takes X_i as n independent and identically distributed (i.i.d.) random variables with empirical distribution functions F_n

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[-\infty, x]}(X_i), \tag{6}$$

where $I_{[-\infty,x]}(X_i)$ is called the indicator function which maps to 1 for $X_i \leq x$ and 0 elsewhere. Next, the equation (7) represents the test statistic which is defined as D_n

$$D_n = \sup_x |F_n(x) - F(x)|. \tag{7}$$

The Glivenko-Cantelli theorem proposes that test statistic D_n converges to 0 almost surely for $n \longrightarrow \infty$, when the sample originates from F as well (Tucker, 1959). Lastly, the term DfG, which is introduced by the authors, is determined as the difference between the test statistic D_n and critical value (CV) of the test

$$\text{DfG} = D_n - CV, \tag{8}$$

where the equation translates into a scheme in a way that in case of a negative DfG, one does not reject the null hypothesis about normality of the sample. In contrast, the volume of deviation from normality goes up with respect to an increasing value of a positive DfG and thus, one then reject the null hypothesis about normality. The above study reveals that a country presents a nicer track of economic performance, when the RCA indices are distributed sufficiently closely to normal distribution. It is, of course, a result of regressing derived values of DfG on GDP with a VAR model. In this thesis, the RCA indices are not assumed to be normally distributed, since their distribution is explored through the Zipf's Law. Hence, the main inspirational prior research of this thesis has come forward by Hinloopen and van Marrewijk (2006). They find a strong integration of rank-size rule into trade flows in terms RCA indices. Their findings advocate the fact that RCA indices align with the Zipf's Law most of the time, although there seems to be deviations depending on the choice of sample country. While estimating the power law exponent, they reveal the below fit, which is derived from an OLS model

$$\ln\left(rank - \frac{1}{2}\right) = a - b\ln\left(size\right),$$

where b is the estimated power law exponent, namely the Gabaix-Ibragimov estimate. Apparently, *rank* is shifted by half for a better fit. Since the prior findings by Gabaix and Ibragimov (2011) indicate that bias is in inverse link with sample size, it is necessary to pool a sufficient volume of observations. Also, authors target the above model across 3 different dimensions such as time, country, and sector, while the first dimension is used in this thesis. The findings suggest that the estimates are mostly relevant, being close to 1 with a high goodness-of-fit, and thus in line with the Zipf's Law. Note that the authors find positive values for the exponent term b that are said to be really close to 1; it is due to the fact that b has a negative sign up-front. In contrast, we will be searching for values with negative signs that are close to -1, since we will keep the sign of regression line coefficient as it is positive. Furthermore, the authors also recognize that there is a significant difference within the country dimension. It is initially assumed that it originates from unique characteristic distinctions among countries. As expected, there is an empirical driving power between the power law exponent and specific indicators, including GDP, population, volume of exports. Meanwhile, observing the sector dimension, authors recall the prominent Heckscher-Ohlin theorem since records show that there are differences across export sectors due to factor abundance. It is, in fact, influencing the estimated exponent as well. However, this thesis particularly focuses on time dimension through the time series methodology, as discussed in the following sections.

5 Data and Methodology

This chapter embraces a crucial role by delivering the essence of empirical frameworks used in the thesis. That is to say, the section sheds light to two econometric models that are fed with data set consisting of several variables. Notably, the second model, which is of higher complexity, is explained through previous research in advance.

5.1 Data

The data, that consist of the above mentioned variables, can be recalled from Section 3.2 which clarifies the variables and compares them to trade openness. To be exact, the main goal of those comparisons of economic variables to trade openness was to convey an initial understanding how these indicators respond to trade dynamics. Having the fact that power law exponent acts as a mere measurement of trade in a way too, its implications together with economic variables should be insightful. To recall, the power law exponent is the exponent parameter s of the Zipfian distribution. Thus, the variable Exponent, which will further be derived, is expected to take values around -1 in accordance with the Zipf's Law. Next, when it comes to the remaining variables, GDP has unquestionably been a measure of growth and prosperity in economic sense in previous studies (Borensztein et al., 1998). FDI, meanwhile, is considered a catalyst dynamo of economic development, since it stimulates transfer of technology, investments, and human capital (Makki & Somwaru, 2004). Moreover, Gupta and Singh (2016) use CPI-based (Consumer Price Index) inflation rate that is an essential indicator of economic stability. Lastly, Real Effective Exchange Rate (REER) is defined as a deflated form of its own nominal version with respect to relative prices and thus, the part *"real"* measures the real effect on economic phenomena, or trade balance, for example (Maciejewski, 1983).

It is noteworthy that the explanatory variables have a magnitude of effects on each other; that is why separate time series, representing these variables, are suitable for the time series model that is used in this work. In the initial step, the data of trade statistics are collected from UN Comtrade database through queries from World Bank World Integrated Trade Solution (WITS). The products within the data are classified according to a broad scheme of HS-6 (Harmonized System) which is determined by World Customs Organization (WCO) nomenclature. Furthermore, the data are pre-processed through the calculation phase of RCA indices using the pre-specified formula (1). In the next part, the results from the first model are accompanied in the secondary model by economic data that consist of 4 time series, representing the Czech economy across 1998-2020 on quarterly basis. Essentially, the data set for inflation is collected from OECD and it is then combined with the remaining determinants that are gathered from the IMF database.

5.2 Models

This thesis builds its empirical basis upon two consecutive directions. First of all, it expands the existing literature by inspecting whether RCA indices comply with size-rank rule or not. Hence, the following model is adopted from Hinloopen and van Marrewijk (2006), recalling their models for 3 dimensions which are commodity (i), country (j), and time (t), where the last one is added later

$$\ln\left(rank\left(RCA_{ijt}\right)-\frac{1}{2}\right)=\alpha_{0,t}+\alpha_{1,t}\ln\left(RCA_{ijt}\right)+u_t,\quad t=1998,...,2020. \tag{9}$$

The above OLS model is used in this study to analyze distribution of RCA indices, considering the time dimension t only. It is due to the fact that country dimension j consists of the Czech Republic only. Also, commodity dimension i does not bear a significant importance for this work, just as the trade data are not specifically classified with respect to commodity groups. The model returns per-year estimates for regression coefficient α_1 for a period from 1998 to 2020, and a pooled regression for time dimension in the form of panel (longitudinal) data.

The second objective of this thesis is to probe to what extent a set of economic indicators govern the estimated power law exponent. To perform such a task, the results derived from the first model are inserted into another model together with other economic control variables. This model is previously defined Vector Autoregression (VAR) which is, to say, a time series model introduced by Sims (1980) and based on lagged values that are included to view how variables evolve over time and influence each other simultaneously. In essence, the VAR model is a multivariate generalization of univariate autoregressive (AR) models. Hatemi-J (2004) specifies a typical k-dimensional VAR(ρ) model in the following way

$$y_t=\nu+A_1y_{t-1}+A_2y_{t-2}+...+A_\rho y_{t-\rho}+\varepsilon_t, \tag{10}$$

where ρ denotes the number of lags, y_t and ν are $m\times1$ vectors of variables and constants, respectively, and ε_t is supposedly white noise with $\mathbb{E}(\varepsilon_t)=0$. Note that A_i stands for an $m\times m$ matrix of coefficients corresponding to i ($i=1,...,\rho$) which is lag order. In essence, the equation (10) is designed for a single variable y that is regressed on its lagged values. The model is reflected in this thesis in a similar way yet in a multivariate form with pre-specified explanatory variables such as *GDP*, *FDI*, *REER*, and π (inflation). Since these variables have certain effects on each other as well, VAR model is, once more, a decent framework to properly examine possible existence of endogeneity among variables of interest. Although the VAR model returns five separate regressions for each variable in this case, we are more interested in the one for variable Exponent. Thus, the model with j lags can be standardized in the following condensed way

$$\alpha_t = \beta_0 + \sum_{i=1}^{j} \beta_{1,i}\alpha_{t-i} + \sum_{i=1}^{j} \beta_{2,i}\ln(GDP_{t-i}) + \sum_{i=1}^{j} \beta_{3,i}\ln(FDI_{t-i}) + \sum_{i=1}^{j} \beta_{4,i}REER_{t-i} + \sum_{i=1}^{j} \beta_{5,i}\pi_{t-i} + \varepsilon_t, \tag{11}$$

where β_0 is a constant and the set of parameters $\beta_{n,i}$ $(n = 1, 2, 3, 4)$ constitutes of respective regression coefficients with 4 dimensions representing 4 variables. In parallel, the term α_{t-i} shows lagged power law exponents derived from the first OLS model and π_{t-i} reflects lagged inflation along with other economic variables. Note that t is defined within the same 23-year time frame as in (9). Last but not least, ε_t is a vector of shocks with zero mean and variance-covariance matrix $\sum$. Inside the model, there are $m + m^2 j$ coefficients to be estimated, which often leads to an issue called overparametrization or curse of dimensionality. It is further referred once again while covering the in-depth VAR, suggesting how one can overcome this problem. The rest of the variables represent other economic indicators that are mentioned, when the data are presented previously. It is noteworthy that large values of *GDP* and *FDI* are re-scaled with natural logarithm, which is often used to deal with skewness and improve linearity between variables.

5.3 VAR Model Specifications

Conducting a VAR analysis is a fairly complex process that encapsulates more than just interpreting the regression coefficients and the R^2. Once the estimation is successfully completed, one might attempt at testing model diagnostics (Pfaff, 2008). There are certain assumptions that have to be met in order to derive a model with a stable system. Firstly, a unit root test is required for the model that needs to be fed with a stationary data set. In case the equation for regression model has roots inside the unit circle, one has enough evidence to reject the null hypothesis and claim that the process is stationary. In a likely situation of non-stationarity, due to characteristics of available macroeconomic data, it can be considered to take the first differences of variables, which is performed in this study as well. Next, there are follow-up tests for vital diagnostics such as serial autocorrelation, heteroskedasticity, structural stability, and normality tests, which are also a part of this research for measuring the relative performance of the model. The model that successfully passes through the specified filters is subject to further examination regarding causalities, forecasting, and responses to external dynamics, namely Impulse Response Functions (IRF). Similar to the forecasting tool, the last metric IRF, which demonstrates variables' behavior under shocks, stems from recursive moving averages. Relatedly, the IRF and Granger causality are encapsulated in the course of this thesis.

Despite its broad implications, the VAR model, in general, has been criticized due to previously stressed overparametrization problem. The most prominent solution to the problem directly originates from one of initial works of Litterman (1979). This approach brings Bayesian modelling together with the VAR models. The Bayesian estimation in this case, that is Bayesian VAR (BVAR), attains a similar functional form as in (11), although the operating mechanism is different in the back end. By definition, a typical BVAR model relies on prior information in order to mitigate the shortages of the original framework. In contrast to a standard VAR model, the state-of-the-art BVAR model can be applied to a non-stationary data set too. Henceforth, the former model is accompanied by the latter as well in this research. To cure the so-called overparametrization issue, the BVAR acts as a shrinkage model which reduces the number of estimated coefficients. It basically pushes parameters of an unrestricted model towards a naïve benchmark and thus, lowers parameter uncertainty (Kuschnig & Vashold, 2021). Choosing informative priors and related hyperparameters is such a critical task, just as it can yield poor estimations. Thus, when it comes to the prior selection, it is occasionally suggested to treat hyperparameters (coefficient parameters of prior distribution) in a hierarchical method that is based on the Bayes' law. The hyperparameters, in turn, are estimated by being assigned to their hyperpriors that are variance parameters. The mathematical formulation of this approach defines an essential parameter Marginal Likelihood (ML) which is a density of the data set as a function of hyperparameters and hyperpriors. The ML, which is a decision criterion for assigning hyperparameters, is as well used to examine the posterior hyperparameters.

This research utilizes a method called Minnesota prior, which is introduced by Litterman (1986) as a baseline approach in the context of BVAR. It is such a flexible method, since it allows users to control crucial parameters such as single-unit-root (SUR) and sum-of-coefficients (SOC). Also, hyperparameters can be adjusted and thus, treated in the hierarchial manner in the Minnesota prior. In its core, Minnesota prior proposes that all variables observe a random walk process. This practice reformulates mean and covariance matrix with further introduced arbitrary parameters such as λ ($0 < \lambda < \infty$), measuring tightness of the prior, and θ, evaluating the degree of shrinkage, and ψ, controlling standard deviation of prior on lags of independent variables. In further empirical works in this thesis, Minnesota prior is used with its default range of parameters. Thereafter, much like the VAR model, the BVAR is briefly investigated through IRF, which knowingly represents reactions of variables to shocks blowing the economy. In spite of other schemes, especially sign restrictions test, that are well applicable to a standard BVAR, they are not covered in this research.

6 Empirical Results and Limitations

In this chapter, results of the empirical work are discussed in detail. It is worth mentioning that the discussion covers implications of both models as well as variables, too.

6.1 Descriptive Statistic Analysis

Before proceeding to the analyses of respective regression models, it is essential to inspect the data set rigorously. Thus, this section covers the descriptive statistics of variables that are utilized within those models. To begin with, as it is known, the first model takes trade data for a 23-year period. The data for each year has a large magnitude such that it covers on average 4852 data points, or, in other words, commodities with HS-6 classification. Although the nature of goods and services is not in the core interest of this study, they are ranging from edibles to raw materials to electronics. Using annual records for these products, RCA indices are calculated in advance and transformed with logarithm as well as their ranks. It is of note that RCA indices that are greater than 1 and show comparative advantage are considered only, since this study is keen on products that the Czech Republic has comparative advantage for. Moreover, the model iterates over the given data for each year and eventually returns results in a usual OLS report form. Each time the coefficients α_1 for RCA indices are considered as the per-year estimates for the Zipf's Law.

It should be stressed that the results for power law exponents are, to recall, in annual structure due to data availability and thus, are not compatible with quarterly economic data. That is why the first dataset requires data-processing based on temporal disaggregation that stems from a methodology called Denton-Cholette (Denton, 1971; Dagum & Cholette, 2006). The process is purely mathematical such that it is based on a predictive regression model, sometimes Generalized Least Squares (GLS), to disaggregate annual data into quarters. The method minimizes sum of squares of deviations among annual and quarterly data points. Thus, the model optimizes the accuracy of the suggested quarterly time series. In our case, the conversed data set, which is not conflicting with dimensions anymore, is then combined with the secondary economic data set. Notably, the main factor for the usage of quarterly data is that the time series models, in general, perform well when they are fed with the sufficient amount of data. Ultimately, the ready-to-use data for the second model have 92 observations for 5 variables, excluding *Time* column. While exploring the FDI data set, it is obvious that it is a net determinant such that it has negative figures where FDI inflows exceed outflows. Firstly, this study is more interested in the effect of FDI outflows on these power law exponents that are computed from exports, which are in turn outflows in a way too. Secondly, the negative values are problematic since the VAR model from equation (11) re-scales FDI with logarithm, which is possible for positive values only. Therefore, the positive values are considered only as the data for net FDI outflows, accounting for 79 observations in total. However, not to neglect negative data points for potentially important information, another VAR model is employed that takes the whole 92 observations with GDP and FDI data as smaller input values re-scaled in billions and millions, respectively. Table 1 below recapitulates the summary statistics for variables used in the first VAR model. Figures show that the Czech Republic actually experienced deflation of almost -0.4% in the first quarter of 2003. Moreover, the maximum rate of REER positioned around 107 in late 2008, when the Czech koruna appreciated the most. Meanwhile, GDP and FDI followed a rather stable yet an increasing evolvement over time such that they improved to the highest volumes until the pre-pandemic period. In terms of descriptive statistics, it is often suggested to deal with outliers that can distort the actual results most of the time. In this case, almost no outliers are found in the data set, although the z-score method reveals that there are two outliers in the inflation data. However, these data points are kept in the sample, largely because of the fact that these two points might capture important information.

Table 1: Summary statistics of variables

Name	Observations	Mean	Std. Dev.	Minimum	Maximum
Inflation	79	2.746	2.508	−0.387	13.251
REER	79	87.578	12.661	63.075	106.968
ln(GDP)	79	27.564	0.186	27.259	27.854
ln(FDI)	79	22.928	1.785	18.05	25.129
Exponent	79	−1.06	0.057	−1.177	−0.949

Furthermore, working with time series models, an extremely vital aspect to consider and check in advance is definitely stationarity of the data, which is a property based on a stable mean, variance, and autocorrelation with no trend and seasonality. In this regard, the Augmented Dickey-Fueller test (ADF) is used to test the null hypothesis whether there is a unit root and thus, non-stationarity in the data set (Dickey & Fueller, 1979). The ADF test for our sample of 79 observations, in the initial step, shows non-stationarity which probably asserts an unstable mean since the variance should be balanced with log-transformation. Hence, differencing process is applied by taking differences of consecutive values of each variable. Now the ADF test assures a stationary data set at 5% significance level for almost all variables, except for GDP and REER that have p-values very slightly above 0.05, which should be tolerable in this case. Accordingly, one has enough evidence to reject the null hypothesis and claim that variables are stationary. The full data set, including negative FDI values, is as well stationary with all variables. Another step is to check the persistence of the data set, which can be explored through the Autocorrelation Function (ACF). As depicted in the Appendix A, the process returns a 5×5 matrix of plots depicting the autocorrelation of each variable with its lagged values subject to the confidence bands. Notably, there is only a tiny volume of persistence in each variable, yet the model does not seem to be affected much. Relatedly, in Figure 8 one can find plots for individual variables in the diagonal, while their combinations are demonstrated in triangles above and below the diagonal.

6.2 Regression Analyses and Results

First Model

As debated earlier, the empirical part of this thesis comprises of two models and the very first one of them is a particular OLS model from equation (9). In other words, the model regresses RCA indices on their respective ranks, which are priorly shifted by half. The indices, that cover the period between 1998 and 2020, are calculated using the equation (1) in advance before being used in the econometric model. In this part, the model returns separate results as well as a pooled output for α_1 coefficient which also stands for the power law exponent. Thus, Table 2 summarizes the findings in each case with corresponding coefficients and values of R-Squared (R^2). Clearly, the first parameter measures how far the estimated exponent is from -1, as indicated by the Zipf's Law. Next, the R^2 interprets how good the model is by demonstrating to what extent variance of the dependent variable (Rank) is explained by the independent variable (RCA index). From the table below, it can be inferred that coefficients are strikingly close to the desired value of -1 along with being statistically significant at all times. Notably, the pooled regression model yields a coefficient around -1.054, which is supported with an R^2 of about 0.988 or 98.8%.

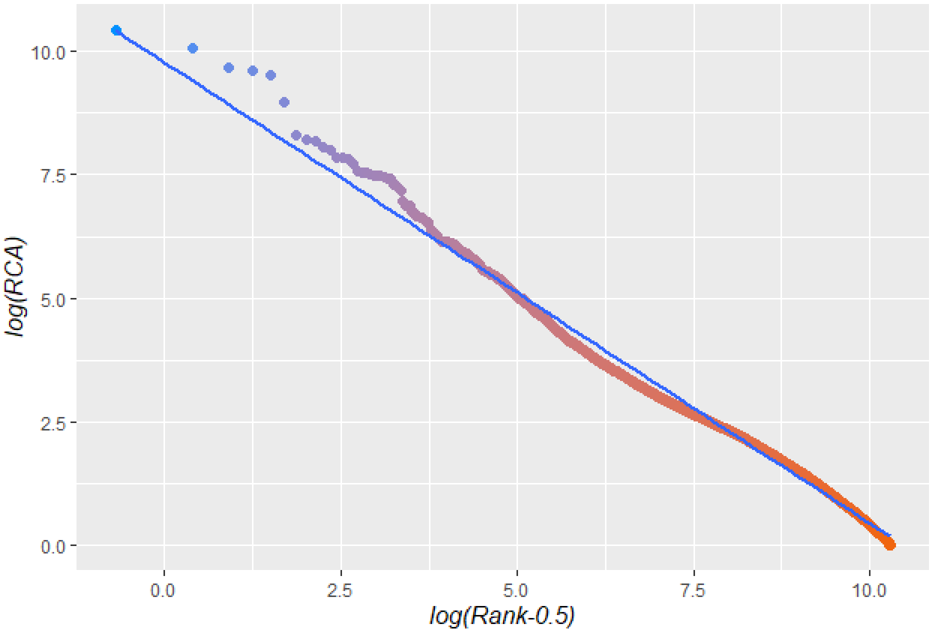
Table 2: Per-year and pooled estimates for power law exponent

1998	1999	2000	2001	2002
-0.9781^{***}	-1.0413^{***}	-0.9895^{***}	-0.9633^{***}	-1.0548^{***}
(0.0036)	(0.0039)	(0.0033)	(0.0031)	(0.004)
2003	2004	2005	2006	2007
-1.0244^{***}	-1.0505^{***}	-1.0248^{***}	-1.0136^{***}	-1.0104^{***}
(0.004)	(0.0036)	(0.0032)	(0.0025)	(0.0039)
2008	2009	2010	2011	2012
-1.0201^{***}	-1.0815^{***}	-1.0033^{***}	-1.0601^{***}	-1.0612^{***}
(0.0041)	(0.0032)	(0.0031)	(0.0035)	(0.0032)
2013	2014	2015	2016	2017
-1.0843^{***}	-1.1468^{***}	-1.1438^{***}	-1.1464^{***}	-1.0717^{***}
(0.0033)	(0.0035)	(0.0041)	(0.0037)	(0.0033)
2018	2019	2020		
-1.1125^{***}	-1.1171^{***}	-1.1646^{***}		
(0.0034)	(0.0033)	(0.0041)		
Pooled		-1.054^{***} (0.0006)		
R^2		0.9885		
Observations		32,554		

Note: *p<0.1; **p<0.05; ***p<0.01

From the above outcomes, it is obvious that patterns in export structure of the Czech Republic nicely complies with the Zipf's Law to a certain extent, although there are occasional deviations, specifically, in the upper part. However, as reported, both per-year and pooled power law exponents are statistically significant at 1% significance level. Indeed, the outcomes can also be checked through the following Figure 5, which is provided to draw the inverse relationship between RCA indices and their ranks. As anticipated, the figure portrays a negative log-linear association accompanied by a fit with a slope close to -1. Note that the concentration of indices increases from left to right.

Figure 5: Distribution of RCA indices



Second Model

The above outputs, as debated earlier, constitute of 23 power law exponents that are extremely close to that of Zipf's Law. Using the previous Denton-Cholette method, these exponents are converted into quarterly data which is then compatible to be merged with the economic data for the second model. As summarized in Table 3, there are three models, which have the quite similar mathematical formulations but certain technical differences. The models, to say, consist of a BVAR with 4 lags and two VARs with 3 and 5 corresponding lags. The former VAR(3) takes log-transformed data with 79 observations, while the latter VAR(5) is assigned the complete data including negative FDI values as well. In between the BVAR(4) likewise utilizes rendered data in logarithmic scale. Note that the models are mainly compared through IRFs, since the regular regression outputs for the Bayesian model are slightly different from those for a standard VAR.

Table 3: Regression results with different lags

	Dependent variable: Exponent		
	VAR(3)	**BVAR(4)**	**VAR(5)**
Exponent-lag1	**0.920*****	1.002	**0.650*****
	(0.141)		**(0.109)**
GDP-lag1	0.106	0.0003	−0.0001
	(0.158)		(0.00004)
FDI-lag1	−0.0001	0.00026	−0.00004
	(0.002)		(0.00002)
REER-lag1	0.0001	0.00013	0.0002
	(0.001)		(0.0003)
Inflation-lag1	0.002	−0.001	−0.001
	(0.002)		(0.001)
Exponent-lag2	−0.289	−0.001	0.109
	(0.181)		(0.090)
GDP-lag2	−0.081	−0.00023	0.00004
	(0.184)		(0.0001)
FDI-lag2	0.001	−0.00012	−0.0001
	(0.002)		(0.00004)
REER-lag2	0.0002	0.000022	**−0.001****
	(0.001)		**(0.0004)**
Inflation-lag2	−0.002	−0.0023	−0.001
	(0.002)		(0.001)
Exponent-lag3	**−0.231***	−0.00075	−0.036
	(0.133)		(0.090)
GDP-lag3	−0.026	−0.00007	−0.0001
	(0.185)		(0.0001)
FDI-lag3	0.001	−0.00006	−0.0001
	(0.002)		(0.00005)
REER-lag3	0.0001	−0.00008	**−0.001****
	(0.001)		**(0.0003)**
Inflation-lag3	−0.0004	−0.00074	0.00000
	(0.002)		(0.001)
Exponent-lag4		−0.0003	**−0.806*****
			(0.088)
GDP-lag4		0.0004	−0.0001
			(0.0001)
FDI-lag4		−0.00024	−0.0001
			(0.00004)
REER-lag4		−0.00015	**−0.001****
			(0.0003)
Inflation-lag4		0.00026	0.001
			(0.001)
Exponent-lag5			**0.472*****
			(0.102)
GDP-lag5			0.0001
			0.0001
FDI-lag5			−0.00003
			(0.00003)
REER-lag5			**−0.001*****
			(0.0003)
Inflation-lag5			−0.001
			(0.001)
Constant	−0.001	0.019	−0.0001
	(0.002)		(0.001)
R^2	0.701	–	0.853
Observations	79	79	92

Note: *p<0.1; **p<0.05; ***p<0.01

The regression coefficients are summoned above for three different models with three consecutive lags. The number of lags are determined in accordance with different information criteria. In particular, the lag selection compares techniques ranging from Akaike Information Criterion (AIC) to Schwarz Information Criterion (SIC), which quantify the volume of information lost by the model. Note that lags with the given statistical significance are highlighted in bold alongside coefficients and standard errors.

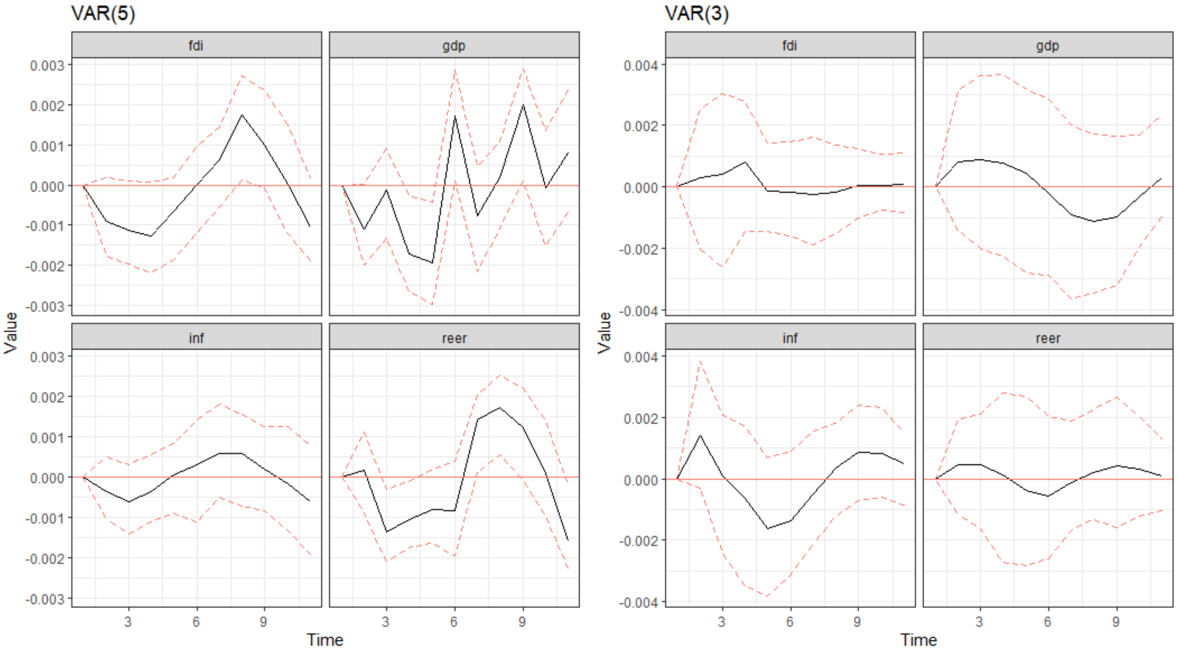
To begin with, in VAR(3) there are two statistically significant independent variables which are the first and third lagged values of Exponent itself at corresponding 1% and 10% significance levels. Similarly, the second VAR with 5 lags has statistically significant coefficients of Exponent for almost all lags, except for second and third lags, at 1% significance levels. At the same time, all lagged values of REER, except the first one, are statistically significant indicators. It is worth pointing out the fifth lag of REER is statistically significant even at 1% significance level. In both cases, the remaining variables, namely GDP, FDI, and inflation have seemingly little to no impact on Exponent with very small coefficients. The first model has an R^2 of 0.701, which recalls the fact that approximately 70% of variance in dependent variable is explained by independent lagged variables; this metric is defined at 0.853 or 85.3% for the second VAR, which is improved by a lot with additional lagged values. Note that the statistical significance of coefficients and standard errors are not stressed for the BVAR(4) in R software by default. Similarly, the R^2, as a measurement of model performance, is not pointed out for BVAR models. Although Gelman et al. (2019) propose the rstanarm package for calculation of the R^2 for certain Bayesian models, it is as well not compatible with Bayesian time series.

The discussed previously models require a range of diagnostic tests to check the overall performance. Firstly, the outputs from VAR(3) model suggest that unit roots, which can be interpreted as eigenvalues too, are inside the unit circle and thus, the system is stable at all. Next, the Portmanteau test is applied in order to control the serial autocorrelation of residuals. Having the fact that p-value of this test is 0.33, one does not have enough evidence to reject the null hypothesis at 5% significance level and thus, it can be claimed that there is not serial autocorrelation. Furthermore, the Lagrange Multiplier test is considered for Autoregressive Conditional Heteroskedastic (ARCH) effect in residuals. Likewise, a large p-value in this test determines that there is not enough evidence to reject the null hypothesis and concludes that heteroskedasticity has not been detected in residuals. Lastly, CUSUM test (Cumulative Sum Control Chart) is employed to scrutinize structural stability in residuals. It is a technique to monitor deviations in a model subject to a threshold value. Thus, the Figure 9 in Appendix B suggests that the requirement of stability is met, since there are not structural breaks exceeding the confidence bands in red. In case all these previous tests are duplicated for the VAR(5) model, the results are almost identical, although the Portmanteau test detects serial autocorrelation. Lastly, as a result of Jarque-Bera test, a small p-value manifests that residuals are not normally distributed in both VAR models.

Beyond these, the BVAR(4) model differs from the previous ones in terms of standard diagnostic analysis of outputs. In essence, along with a brief overview of log-likelihood, that is about 14.51, or covariance matrix (Σ), that is a 5×5 positive-definite matrix, a typical BVAR recommends the analysis of model hyperparameters. In the context of this study, Appendix C provides a set of plots that explains structure of hyperparameters, including ML, λ, SUR, and SOC. Plots from Figure 10 propose convergence for hyperparameters. One can assess the convergence behavior in detail by using the interface for Markov Chain Monte Carlo (MCMC) simulations. Note that the MCMC is a method to randomly sample from the high-dimensional probability distribution and then approximate the previously mentioned hyperparameters. Trace plots advocate the fact that there are not any strong outliers and that is why the chain is said to explore the posterior effectively.

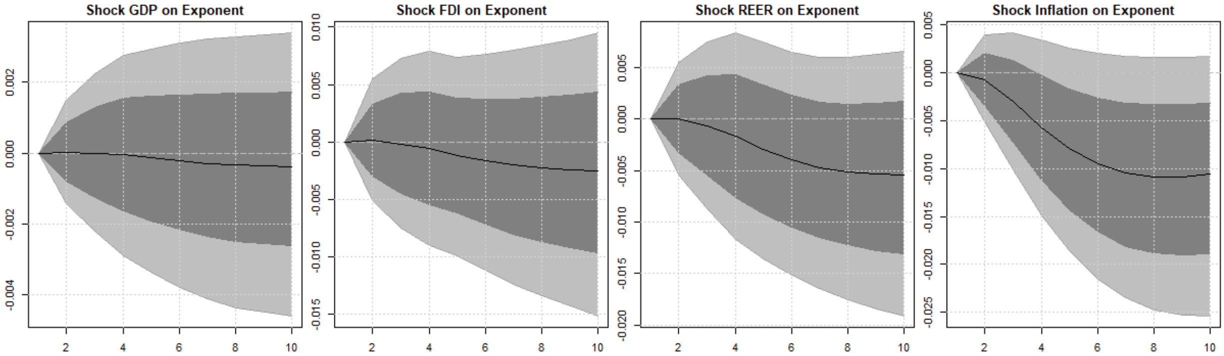
Once the diagnostic requirements are fulfilled, it is suggested to proceed with further implications of outputs from regression models. Therefore, investigation of IRFs attains an integral role in this context. As portrayed in Figure 6, both VAR models are assembled through the comparison of responses of power law exponent to economic shocks in a 10-year period.

Figure 6: Impulse Responses from Exponent in VAR



The VAR(5) is demonstrated through above IRFs with relatively narrow confidence intervals for the impulse responses. Meanwhile, the impulse responses from the VAR(3) define a relatively less accurate situation with broader bands in red, which suggest potential absence of statistical significance in responses of Exponent. In comparison, The differences made by initial data transformation for both models are too obvious to be recognized. In Figure 6, the economic shocks bear an initial negative impact on Exponent in almost all cases in VAR(5) models. However, these effects later shift to be a positive one in the remaining long-term period. In contrast, VAR(3) suggests IRFs with positive initial effects, which might turn to be negative in the long-run. It is worth pointing out that both positive and negative effects have nearly the same amplitudes in each cycle. More importantly, in almost all economic shocks, the VAR(5) performs well such that the diagrams occasionally include significant impulse responses from Exponent. It is due to the fact that at some points, the zero baseline is positioned outside the bands in red. Noticeably, red dashed lines represent 95% confidence intervals in each graph. Next, the Figure 7 below summarizes for the BVAR(4) the impulse responses of exponent subject to four different shocks during the same time frame.

Figure 7: Impulse Responses from Exponent in BVAR



Accordingly, the IRFs have a less complicated shape with no sign of volatile ups and downs, which are particularly visible in the previous Figure 6 for VAR models. In contrast, the dark gray areas stand for 68% confidence bands that are then followed by 90% confidence intervals, which are set by default. Comparison of the IRFs for standard VAR and BVAR models yields some characteristic differences. Apparently, VAR models represent responses with both negative and positive effects; in contrast, the BVAR demonstrates a consistent negative shock only. This might actually stem from the data selection. In other words, this is largely because the former VAR models are fed with differenced stationary data. Specifically, the first difference is taken for VAR(3), while VAR(5) model even required the second differences for stationarity. Nevertheless, the BVAR(4) is assigned with unprocessed non-stationary data.

In Figure 7, the shocks have a negative impact on Exponent in all cases. Interestingly, the severity of shocks increases from GDP (left) to inflation (right). A potential inference from this might be drawn with a basic geometric explanation. In plain terms, it is obvious that the variable Exponent attains a more negative value as a result of a negative shock. We also know that the variable Exponent is identical with the slope of the model fit, which portrays the log-log relation between the rank and RCA index. Thus, when the slope gets a more negative value, the fit becomes steeper. Therefore, when comparing the steeper fit with the original one, the same rank from horizontal axis returns a smaller RCA index from vertical axis. Thus, it might be interpreted that an economic shock might reduce the comparative advantage of a country over a commodity due to the negative effect.

Afterwards, one of the next implications of regression outputs is testing for Granger causality that is often involved in time series analyses. The procedure, being a statistical hypothesis test, simply asks for the fact whether prior values of a time series object can predict future values of another time series. Thus, the null hypothesis basically asserts that a variable does not Granger-cause the another one. Tests that are conducted for VAR(3) propose that only GDP can Granger-cause the remaining variables, just as the null hypothesis is rejected with a p-value of 0.001. While on the other hand, VAR(5) model has three variables, namely Exponent, GDP, and REER, that Granger-cause the other time series in the model. It is supposedly drawn from the fact that the null hypothesis is rejected at 5% significance level in all three cases with p-values slightly above 0.01.

6.3 Limitations

The main external limitation of this study is data availability. Despite the fact that Denton-Cholette conversion method is utilized to overcome this problem, it would have been unquestionably more appropriate to work with raw quarterly trade data that are suggested by official sources. Nevertheless, the available data set is provided on annual basis at its best. Next, another shortcoming is the corresponding FDI data that consist of net values only, although this thesis attempted at analyzing the dynamic between exports and FDI outflows. Also, some negative net FDI values create computational problems while being applied to the logarithm operation in the regression model. That is why, the first VAR model considers positive values only as net outflows. Not to neglect the negative FDI figures that might capture essential information, the second VAR model is employed through the full data. Beyond these, the scope of this work is certainly open to future improvements in the field. Firstly, this context has potential for an expansion into commodity-wise studies. In other words, upcoming research can analyze how individual commodity groups respond to the Zipf's Law in terms of specialization. Moreover, one can consider conducting studies on a broad regional scale, taking more than one sample country into account. Hence, it is thereby more straightforward to reveal differences with regard to the Zipf's Law in the country level. Then, it would be practicable to quantify multidimensional association of power law exponents with economic indicators.

7 Conclusion

In this concluding chapter, results from the previous empirical analyses are summarized. This thesis noteworthily focused on investigation of trade, specifically export, structure of the Czech Republic during the period 1998-2020, using a unique empirical tool called the Zipf's Law. The first goal of this work has been to verify whether the well-known Revealed Comparative Advantage (RCA) indices follow the size-rank rule and are correspondingly distributed across a straight line. Moreover, the regression coefficients, namely power law exponents, from the first model have been subject to time series examinations with respect to a set of economic indicators including GDP, FDI, REER, and inflation.

It is worth to recall the contributions of this thesis, since it is of the pioneer works in the field that analyzes implications of power law probability distributions in trade theory. More importantly, this study is not only one of few works elaborating on the distribution of RCA indices, but the first ever exploration of a country-specific analysis of RCA indices alongside the Zipf's Law. The outcomes of the first model suggested that RCA indices are nearly distributed throughout a line in log-log scale with a slope close to -1, which is in line with the Zipf's Law. This fact is supported by both per-year estimates and a large pooled model. Thus, it is inferred that the export structure of the Czech Republic aligns with the expectations of the Zipf's Law to a large extent.

Next, the meaning of power law exponents are expanded into a more economic sphere using some of the prominent economic determinants. Thus, the respective exponents, after a certain pre-processing phase, are inspected with previously mentioned indicators through three Vector Autoregression (VAR) models that are sophisticated time series models. The first two models are typical VAR models with 3 and 5 lags, whereas the last one, that is BVAR with 4 lags, has particular differences based on Bayesian configurations. According to the former model, Exponent is significantly affected by some of its own lags. Into the bargain, with an increasing size of lags, it is revealed that REER is a statistically significant indicator, along with Exponent itself, nearly in all lags. In the majority of cases, however, the remaining variables have little to no effect on the Exponent. The remaining BVAR model, by default, has relatively more intricating interpretations rather than a mere analysis of statistical significance of coefficients.

The diagnostic tests, which measured performance of models, are followed by deeper econometric frameworks called Impulse Response Function (IRF) and Granger causality. The former tool metricizes reactions of variables to external shocks; meanwhile the latter test whether a time series helps to predict another one. Initial outcomes from IRFs of

VAR models depicted a situation in which the Exponent is subject to both negative and positive shocks. In the VAR model with 5 lags, the shocks mostly bear an initial negative effect, which eventually turns to be a positive one in the very long-run. While on the other hand, economic shocks yield an initial positive impact which changes its sign from time to time. Lastly, the Bayesian model presents a clearer insight such that economic shocks ensure a negative impact on Exponent, as demonstrated by the respective IRFs. This effect persistently lasts throughout the whole period. It is suggested that the characteristic differences in impulse responses might originate from data-processing, since the data intake differs across models. Afterwards, tests that are conducted for Granger causality reveal that GDP, in fact, Granger-causes the other variables in the first VAR model with 3 lags. In contrast, variables GDP and REER alongside Exponent are said to Granger-cause the remaining indicators in VAR(5).

In conclusion, this thesis has attempted to scrutinize applicability of the Zipf's Law in the case of trade structure of the Czech Republic. Essentially, the outcomes from the trade analysis have a nice alignment with the Zipf's Law and that is why RCA indices, in logarithmic level, are distributed accordingly. Moreover, results from the following empirical models disclose the fact that power law exponents are, indeed, affected by particular economic variables in a way too. The first couple of standard VAR models suggest that the economic shocks might have both negative and positive impact on Exponent, whereas these effects change their signs in the very long-run. More interestingly, the Bayesian VAR model reveals that shocks to the economic indicators have a persistent negative effect on Exponent, which, in turn, might reduce the comparative advantage of an economy in production of a product.

References

Aquino, A. (1981). Changes over time in the pattern of comparative advantage in manufactured goods. An empirical analysis for the period 1962-1974. *European Economic Review*, 15(1), 41–62. https://doi.org/10.1016/0014-2921(81)90066-0

Ayoub, R. (1974). Euler and the Zeta Function. *The American Mathematical Monthly*, 81(10), 1067–1086. https://doi.org/10.2307/2319041

Balassa, B. (1975). Reforming the system of incentives in developing countries. *World Development*, 3(6), 365-382. https://doi.org/10.1016/0305-750X(75)90022-4

Balassa, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage. *The Manchester School*, 33(2), 99-123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x

Balassa, B., & Noland, M. (1989). "Revealed" Comparative Advantage in Japan and the United States. *Journal of International Economic Integration*, 4(2), 8-22. https://doi.org/10.11130/jei.1989.4.2.8

Batra, A., & Khan, Z. (2005). Revealed comparative advantage: An analysis for India and China, *Working Paper, No. 168*, Indian Council for Research on International Economic Relations (ICRIER), New Delhi. http://www.icrier.org/pdf/wp168.pdf

Baur, E., & Bertoin, J. (2021). On a Two-Parameter Yule-Simon Distribution. *A Lifetime of Excursions Through Random Walks and Lévy Processes. Progress in Probability*, 78, 59-82. https://doi.org/10.1007/978-3-030-83309-1_4

Bebek, U. (2017). RCA: Choosing the Right Measure. *Department of Economics, University of Birmingham.* https://www.researchgate.net/publication/325905987

Binici, M., Cheung, Y. W., & Lai, K. (2012). Trade Openness, Market Competition, and Inflation: Some Sectoral Evidence from OECD Countries. *International Journal of Finance and Economics*, 17(4), 321-336. https://doi.org/10.1002/ijfe.1451

Borensztein, E., De Gregorio, J., & Lee, J. W. (1998). How does foreign direct investment affect economic growth? *Journal of International Economics*, 45(1), 115-135, https://doi.org/10.1016/S0022-1996(97)00033-0

Bowdler, C., & Nunziata, L. (2006). Trade Openness and Inflation Episodes in the OECD. *Journal of Money, Credit and Banking*, 38(2), 553-563. http://www.jstor.org/stable/3839134

Bowen, H. P., Leamer, E. E., & Sveikauskas, L. (1987). Multicountry, Multifactor Tests of the Factor Abundance Theory. *The American Economic Review*, 77(5), 791-809. http://www.jstor.org/stable/1810209

Brun, J., & Gnangnon, S.K. (2017). Does Trade Openness contribute to driving Financing Flows for Development? *WTO Staff Working Papers* ERSD-2017-06, World Trade Organization (WTO), Economic Research and Statistics Division. https://doi.org/10.30875/1f74d977-en

CNB. (2011). *Germany – the Czech Republic's main trading partner.* https://www.cnb.cz/en/monetary-policy/inflation-reports/boxes-and-annexes-contained-in-inflation-reports/Germany-the-Czech-Republics-main-trading-partner

CZSO. (2021). *International trade in goods (change of ownership) – December 2021.* czso.cz/csu/czso/ari/international-trade-in-goods-change-of-ownership-december-2021

Dagum, E. B., & Cholette, P. A. (2006). *Benchmarking, Temporal Distribution, and Reconciliation Methods for Time Series.* Lecture Notes in Statistics. Springer-Verlag, New York.

Dandashly, A., & Verdun, A. (2015). Boarding the Euro Plane: Euro Adoption in the Czech Republic and Slovakia. *Review of European and Russian Affairs*, 9(2), 1-26, https://doi.org/10.22215/cjers.v9i2.2492

Davis, D. R. (1995). Intra-industry trade: A Heckscher-Ohlin-Ricardo approach. *Journal of International Economics*, 39(3-4), 201-226. https://doi.org/10.1016/0022-1996(95)01383-3.

De Castro, T., Vlčková, J., & Hnát, P. (2017). Trade and investment relations between the Czech Republic and China: The Czech Republic as a gateway to the EU? *Society and Economy*, 39(4), 481-499. https://doi.org/10.1556/204.2017.39.4.2

Deb, K., & Hauk, W.R. (2017). RCA indices, multinational production and the Ricardian trade model. *International Economics and Economic Policy*, 14(1), 1-25. https://doi.org/10.1007/s10368-015-0317-z

Dědek, O. (2004). Adopting the Euro: Brake on or Engine for True Convergence? *Eastern European Economics*, 42(2), 45-62. https://doi.org/10.1080/00128775.2004.1104

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, 66(333), 99-102. https://doi.org/10.1080/01621459.1971.10482227

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366), 427-431. https://doi.org/10.2307/2286348

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3), 297-308. http://www.jstor.org/stable/1831401

Dornbusch, R. (1974). Tariffs and nontraded goods. *Journal of International Economics*, 4(2), 177-185. https://doi.org/10.1016/0022-1996(74)90060-9

European Commission. (2020). Convergence Report. *Institutional Paper 129, Brussels.* https://doi.org/10.2765/508213

Fatima, S., Chen, B., Ramzan, M., & Abbas, Q. (2020). The Nexus Between Trade Openness and GDP Growth: Analyzing the Role of Human Capital Accumulation. *SAGE Open*, 10(4), 1-18. https://doi.org/10.1177/2158244020967377

Fetahi-Vehapi, M., Sadiku, L., & Petkovski, M. (2015). Empirical Analysis of the Effects of Trade Openness on Economic Growth: An Evidence for South East European Countries. *Procedia Economics and Finance*, 19(15), 17-26. https://doi.org/10.1016/S2212-5671(15)00004-0.

Fitzová, H., & Zidek, L. (2015). Impact of trade on economic growth in the Czech and Slovak Republics. *Economics and Sociology*, 8(2), 36-50. https://doi.org/10.14254/2071-789X.2015/8-2/4.

Frankel, J. A., & Romer, D. H. (1999). Does Trade Cause Growth? *The American Economic Review*, 89(3), 379-399. https://doi.org/10.1257/aer.89.3.379

French S. (2017). Revealed comparative advantage: What is it good for? *Journal of International Economics*, 106, 83-103. https://doi.org/10.1016/j.jinteco.2017.02.002.

Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M., & Souma, W. (2004). Do Pareto-Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A: Statistical Mechanics and Its Applications*, 335(1-2), 197-216. https://doi.org/10.1016/j.physa.2003.12.015

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics*, 1(1), 255-294. https://doi.org/10.1146/annurev.economics.050708.142940

Gabaix, X. (1999). Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics*, 114(3), 739-767. http://www.jstor.org/stable/2586883

Gabaix, X., & Ibragimov, R. (2011). Rank − 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business and Economic Statistics*, 29(1), 24-39. http://www.jstor.org/stable/25800776

Gantman, E.R., & Dabós, M.P. (2018). Does trade openness influence the real effective exchange rate? New evidence from panel time-series. *SERIEs*, 9, 91-113. https://doi.org/10.1007/s13209-017-0168-7

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian Regression Models. *The American Statistician*, 73(3), 307-309. https://doi.org/10.1080/00031305.2018.1549100

Greenaway, D., Morgan, W., & Wright, P. (2002). Trade liberalisation and growth in developing countries, *Journal of Development Economics*, 67(1) 229-244, https://doi.org/10.1016/S0304-3878(01)00185-7

Grubel, H. G., & Lloyd, P. J. (1971). The Empirical Measurement of Intra- Industry Trade. *Economic Record*, 47(4), 494-517. https://doi.org/10.1111/j.1475-4932.1971.tb00772.x

Gupta, P., & Singh, A. (2016). Determinants of Foreign Direct Investment Inflows in BRICS Nations: A Panel Data Analysis. *Emerging Economy Studies*, 2(2), 181–198. https://doi.org/10.1177/2394901516661095

Hatemi-J, A. (2004). Multivariate tests for autocorrelation in the stable and unstable VAR models. *Economic Modelling*, 21(4), 661-683. https://doi.org/10.1016/j.econmod.2003.09.005

Hau, H. (2002). Real Exchange Rate Volatility and Economic Openness: Theory and Evidence. *Journal of Money, Credit, and Banking*, 34(3), 611-630. https://doi.org/10.1353/mcb.2002.0002

Hinloopen, J., & van Marrewijk, C. (2008). Empirical relevance of the Hillman condition for revealed comparative advantage: 10 stylized facts. *Applied Economics*, 40(18), 2313-2328. https://doi.org/10.1080/00036840600949488

Hinloopen, J., & van Marrewijk, C. (2006). Comparative Advantage, the Rank-Size Rule, and Zipf's Law. *Tinbergen Institute Discussion Papers* 06-100/1. https://doi.org/10.2139/ssrn.943370

International Trade Administration. (2022). *Czech Republic – Country Commercial Guide. Market Overview.* https://www.trade.gov/country-commercial-guides/czech-republic-market-overview

Jakab, Z. M., Kovács, M. A., & Oszlay, A. (2001). How far has trade integration advanced?: An analysis of the actual and potential trade of three central and Eastern European countries. *Journal of Comparative Economics*, 29(2), 276-292. https://doi.org/10.1006/jcec.2001.1709

Janda, K., & Münich, D. (2004). The Intra-Industry Trade of the Czech Republic in the Economic Transition. *Emerging Markets Finance and Trade*, 40(2), 27–50. https://doi.org/10.1080/1540496x.2004.11052565

Josic, M., & Nikic, M. (2013). Gravity Model and Zipf's Law: An In-Depth Study into the Nature of International Trade. *Academic Journal of Interdisciplinary Studies*, 2(9), 583-588. https://doi.org/10.5901/ajis.2013.v2n9p583

Kaitila, V. (1999). Trade and Revealed Comparative Advantage: Hungary, the Czech Republic, and the European Union. *BOFIT Discussion Paper* No. 8/1999. https://doi.org/10.2139/ssrn.1016686

Krugman, P. (2009). The Increasing Returns Revolution in Trade and Geography. *The American Economic Review*, 99(3), 561-571. http://www.jstor.org/stable/25592474

Krugman, P. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics*, 9(4), 469-479. https://doi.org/10.1016/0022-1996(79)90017-5

Kuschnig, N., & Vashold, L. (2021). BVAR: Bayesian Vector Autoregressions with Hierarchical Prior Selection in R. *Journal of Statistical Software*, 100(14), 1-27. https://doi.org/10.18637/jss.v100.i14

Kuzmenko, E., Rumankova, L., Benesova, I., & Smutka, L. (2022). Czech Comparative Advantage in Agricultural Trade with Regard to EU-27: Main Developmental Trends and Peculiarities. *Agriculture*, 12(2). https://doi.org/10.3390/agriculture12020217

Lafay, G. (1992). *The Measurement of Revealed Comparative Advantages.* In International Trade Modelling; Springer: New York, NY, USA.

Lancaster, K. (1957). The Heckscher-Ohlin Trade Model: A Geometric Treatment. *Economica*, 24(93), 19-39. https://doi.org/10.2307/2551625

Laursen, K. (2015). Revealed comparative advantage and the alternatives as measures of international specialization. *Eurasian Business Review*, 5(1), 99-115. https://doi.org/10.1007/s40821-015-0017-1

Leamer, E. E. (1995). Princeton, NJ : International Finance Section, Department of Economic, Princeton University.

Leamer, E. E. (1980). The Leontief Paradox, Reconsidered. *Journal of Political Economy*, 88(3), 495-503. http://www.jstor.org/stable/1831928

Leontief, W. (1953). Domestic Production and Foreign Trade; The American Capital Position Re-Examined. *Proceedings of the American Philosophical Society*, 97(4), 332-349. http://www.jstor.org/stable/3149288

Leromain, E., & Orefice, G. (2014). New revealed comparative advantage index: Dataset and empirical distribution. *International Economics*, 139, 48-70. https://doi.org/10.1016/j.inteco.2014.03.003

Liargovas, P. G., & Skandalis, K. S. (2012). Foreign Direct Investment and Trade Openness: The Case of Developing Economies. *Social Indicators Research*, 106(2), 323-331. https://doi.org/10.1007/s11205-011-9806-9

Lien, N. T. K. (2021). The Effect of Trade Openness on Foreign Direct Investment in Vietnam. The *Journal of Asian Finance, Economics and Business*, 8(3), 111-118. https://doi.org/10.13106/jafeb.2021.vol8.no3.0111

Lin, G. D., & Hu, C.-Y. (2001). The Riemann Zeta Distribution. *Bernoulli*, 7(5), 817-828. https://doi.org/10.2307/3318543

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business and Economic Statistics*, 4(1), 25-38. https://doi.org/10.2307/1391384

Litterman, R. B. (1979). "Techniques of forecasting using vector autoregressions." *Federal Reserve Bank of Minneapolis Working Paper*, 115.

Liu, B., & Gao, J. (2022). Normality in the Distribution of Revealed Comparative Advantage Index for International Trade and Economic Complexity. *Applied Sciences*, 12(3), 1125. https://doi.org/10.3390/app12031125

Ma, A. S. (2013). Revealed Comparative Advantage Measure : ASEAN-China Trade Flows. *Journal of Economics and Sustainable Development*, 4(7), 136-145.

Maciejewski, E. B. (1983). "Real" effective exchange rate indices: a reexamination of the major conceptual and methodological issues. *IMF Staff Papers*, 30(3), 491-541. https://doi.org/10.5089/9781451972740.024

Makki, S. S., & Somwaru, A. (2004). Impact of foreign direct investment and trade on economic growth: Evidence from developing countries. *American Journal of Agricultural Economics*, 86(3), 795-801. https://doi.org/10.1111/j.0002-9092.2004.00627.x

Mundell, R. A. (1961). A Theory of Optimum Currency Areas. *The American Economic Review*, 51(4), 657-665. http://www.jstor.org/stable/1812792

Munir, S., & Kiani, A. K. (2011). Relationship between trade openness and inflation: Empirical evidences from Pakistan (1976-2010). *Pakistan Development Review*, 50(4), 853-876. https://doi.org/10.30541/v50i4iipp.853-876

Neary, J. P. (2009). Putting the "New" into New Trade Theory: Paul Krugman's Nobel Memorial Prize in Economics. *Scandinavian Journal of Economics*, 111(2), 217-250. https://doi.org/10.1111/j.1467-9442.2009.01562.x

Nowak, A., Rozanska-Boczula, M., & Krukowski, A. (2020). Competitiveness of Agriculture in New Member States of the European Union. *European Research Studies Journal*, 23(1), 160-175. https://doi.org/10.35808/ersj/1752

Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27(4), 1-32. https://doi.org/10.18637/jss.v027.i04

Reed, W. J. (2001). The Pareto, Zipf and other power laws. *Economics Letters*, 74(1), 15-19. https://doi.org/10.1016/S0165-1765(01)00524-9

Ricardo, D. (1817). *On the principles of political economy and taxation (3rd ed)*. John Murray.

Romer, D. (1993). Openness and inflation: theory and evidence. *Quarterly Journal of Economics*, 108(4), 869-903. https://doi.org/10.2307/2118453

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5), 71-102. http://www.jstor.org/stable/2937632

Rovna, L. A. (2001). The enlargement of the European Union: The case of the Czech Republic. *Perspectives on European Politics and Society*, 2(1), 105-126. https://doi.org/10.1080/1570585018458754

Rybczynski, T. M. (1955). Factor Endowment and Relative Commodity Prices. *Economica*, 22(88), 336-341. https://doi.org/10.2307/2551188

Sachs, J. D., Warner, A., Åslund, A., & Fischer, S. (1995). Economic Reform and the Process of Global Integration. *Brookings Papers on Economic Activity*, 1995(1), 1-118. https://doi.org/10.2307/2534573

Samuels, J. M. (1965). Size and the Growth of Firms. *The Review of Economic Studies*, 32(2), 105-112. https://doi.org/10.2307/2296055

Šimáková, J. (2014). The Impact of Exchange Rate Development on Czech Trade Flows. *Procedia Economics and Finance*, 15, 129-136. https://doi.org/10.1016/s2212-5671(14)00460-2

Simon, H. A. (1955). On a Class of Skew Distribution Functions. *Biometrika*, 42(3/4), 425-440. https://doi.org/10.2307/2333389

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48. https://doi.org/10.2307/1912017

Soete, L. G., & Wyatt, S. M. E. (1983). The use of foreign patenting as an internationally comparable science and technology output indicator. *Scientometrics*, 5(1), 31-54. https://doi.org/10.1007/BF02097176

Solow, R. M. (1957). Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, 39(3), 312-320. https://doi.org/10.2307/1926047

Stolper, W. F., & Samuelson, P. A. (1941). Protection and real wages. *Review of Economic Studies*, 9(1), 58-73. https://doi.org/10.2307/2967638

The World Bank. (2019). *Czech Republic trade statistics: Exports, Imports, Products, Tariffs, GDP and related Development Indicator.* https://wits.worldbank.org/CountryProfile/en/CZE

Tichý, F. (2007). Impact of Accession to EMU on International Trade - Case of the Czech Republic. *Prague Economic Papers*, 16(4), 336-346. https://doi.org/10.18267/j.pep.312

Titchmarsh, E.C. (1939). *The Theory of Functions, 2nd ed.* Oxford University Press.

Tucker, H. G. (1959). A Generalization of the Glivenko-Cantelli Theorem. *The Annals of Mathematical Statistics*, 30(3), 828-830. https://doi.org/10.1214/aoms/1177706212

Vollrath, T.L. (1991). A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage. *Weltwirtschaftliches Archiv*, 127(2), 265-280. https://doi.org/10.1007/BF02707986

Wang, J. (2021). The Proof of Riemann Hypothesis, the Key to the Door Is the Periodicity. *Applied Mathematics*, 12(8), 723-765. https://doi.org/10.4236/am.2021.128050

Yu, R., Cai, J., & Leung, P. S. (2009). The normalized revealed comparative advantage index. Annals of Regional Science, 43(1), 267-282. https://doi.org/10.1007/s00168-008-0213-3

Yule, G. U. (1925). A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London.* Series B, Containing Papers of a Biological Character, 213, 21-87. https://doi.org/10.1098/rstb.1925.0002

Zaman, Q. U., Donghui, Z., Yasin, G., Zaman, S., & Imran, M. (2018). Trade Openness and FDI Inflows: A Comparative Study of Asian Countries. *European Online Journal of Natural and Social Sciences*, 7(2), 386-396. http://www.european-science.com386

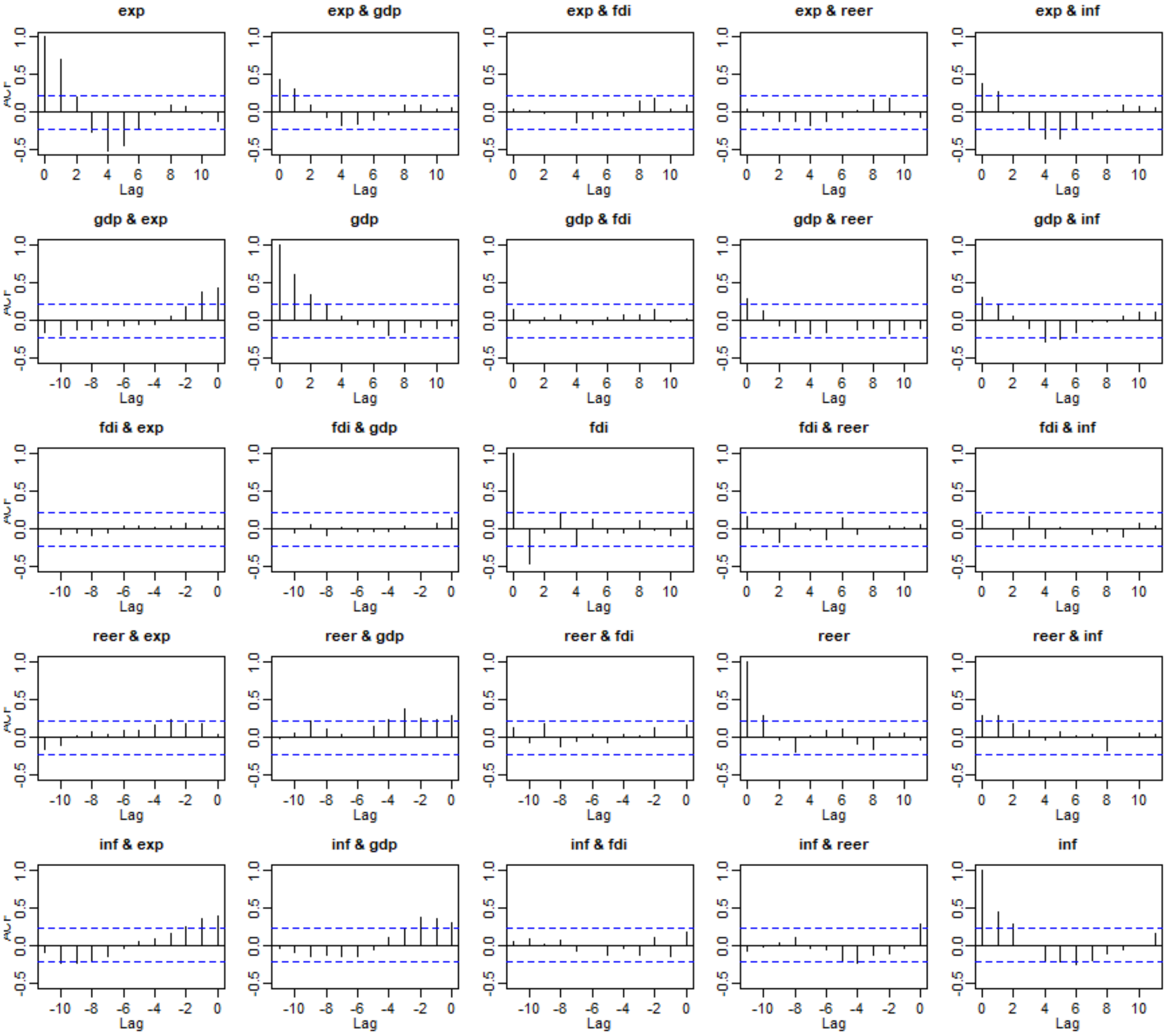
Zipf, G. (1949). *Human Behavior and the Principle of Last Effort.* Cambridge, MA: Addison-Wesley.

Appendices

Appendix A: Matrix of ACF Plots

Below Figure 8 displays a 5 × 5 matrix of ACF plots that portray the degree of persistence for each individual variable as well as their combinations.

Figure 8: ACF plots for degree of persistence

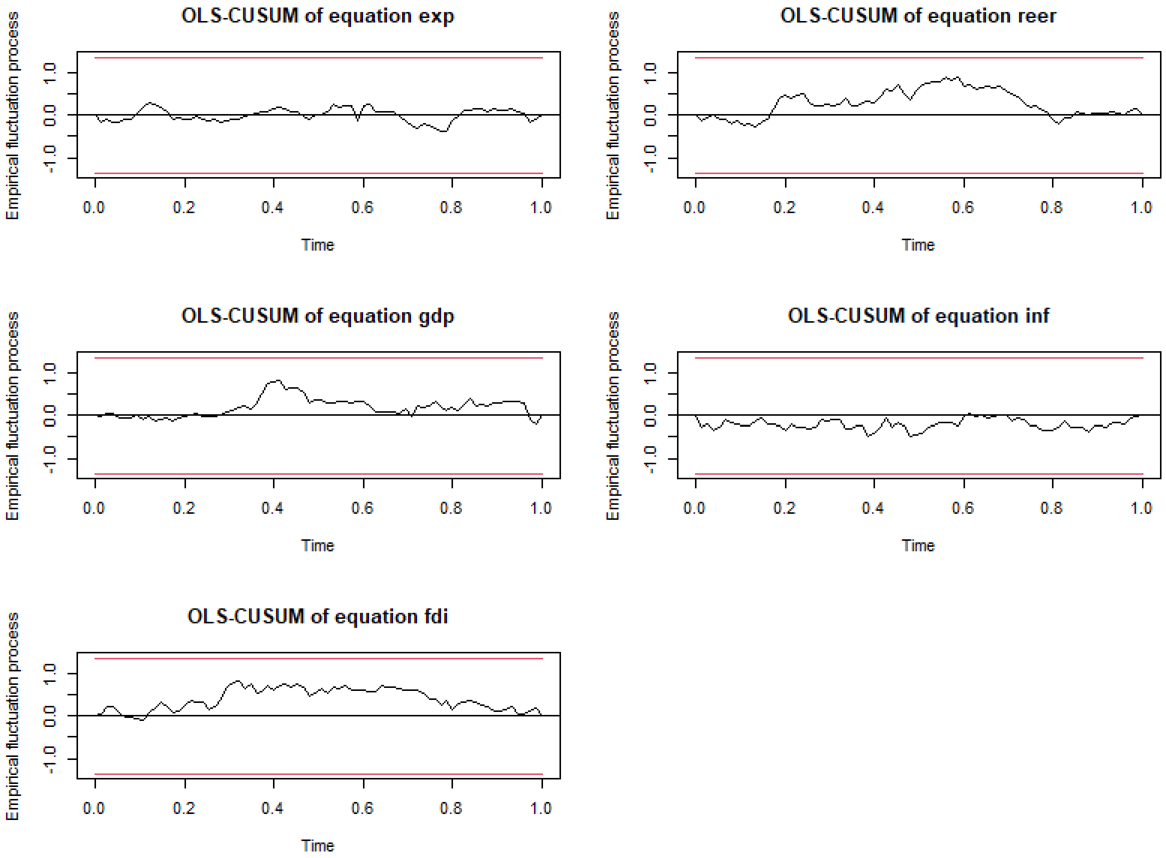


Source: Author's own visualizations from R

Appendix B: Plots of CUSUM Test

The following Figure 9 demonstrates five diagrams for CUSUM test to check structural stability.

Figure 9: CUSUM test for structural stability

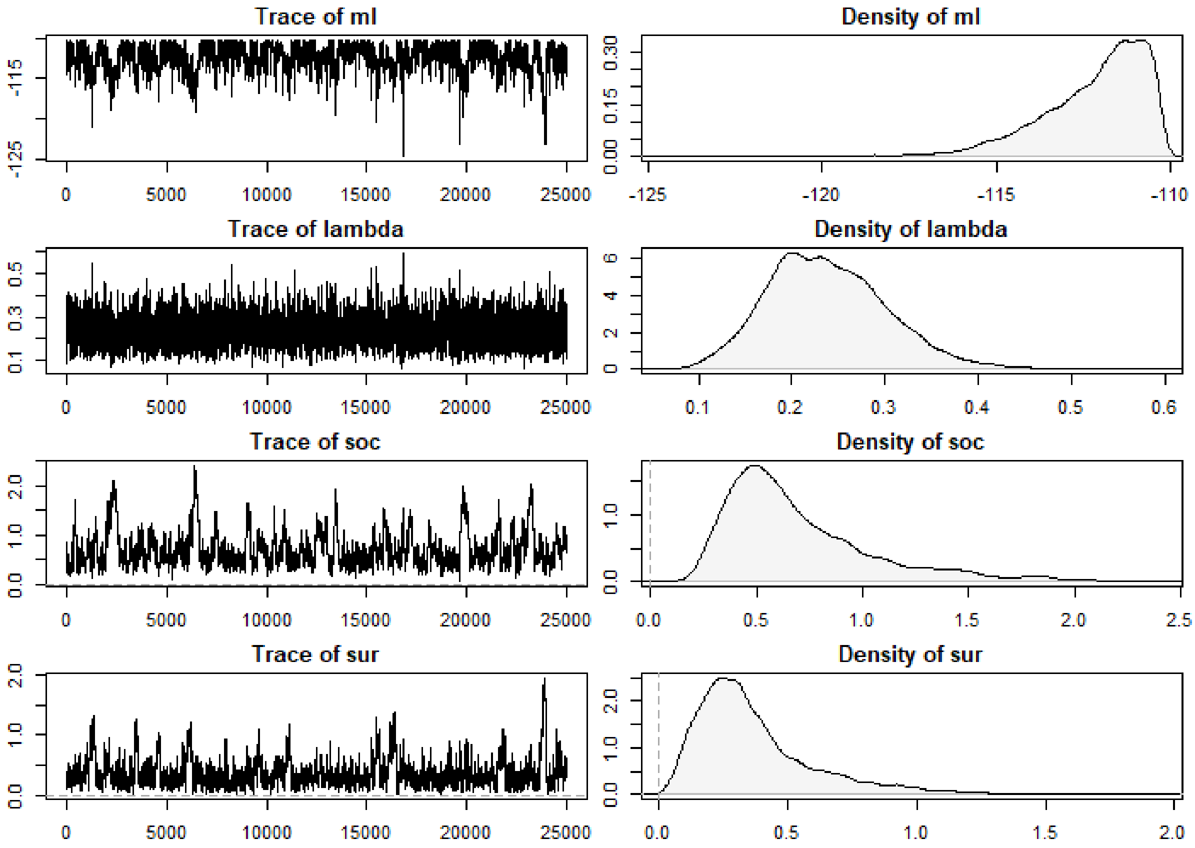


Source: Author's own visualizations from R

Appendix C: Plots of Hyperparameters

The below diagrams depict trace and density of BVAR hyperparameters.

Figure 10: Trace and density plots of hierarchial hyperparameters of BVAR



Source: Author's own visualizations from R