Essays in International and Comparative Economics

Inaugural-Dissertation

zur Erlangung des akademischen Grades einer Doktorin der Wirtschaftswissenschaft des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin

vorgelegt von
Maria Polugodina, M.Sc.
aus Moskau, Russland

April 2020

Erstgutachter:	Prof. Theocharis Grigoriadis, C.Sc. Ph.D. Fachbereich Wirtschaftswissenschaft & ZI Osteuropa-Institut Freie Universität Berlin
Zweitgutachter:	Prof. Dr. Nikolaus Wolf Wirtschaftswissenschaftliche Fakultät Humboldt-Universität zu Berlin
Datum der Disputation:	6. August 2020

Folgende Teile der Dissertation wurden mit Koautoren erstellt:

Kapitel:	Koautor:
2 E-commerce and Market Structure	Lennart Jansen (Bundesverband der Deutschen Industrie)
3 East Prussia 2.0	Theocharis Grigoriadis (Freie Universität Berlin)

Folgende Publikationen sind aus der Dissertation entstanden:

Polugodina, Maria (2019). "The Dutch disease paradox: Endowment change and heterogeneous firms". In: *Applied Economics Letters*, pp. 1-5.

Berlin, 16.04.2020

Erklärung gem. § 4 Abs. 2

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

Unterschrift: _____

Berlin, 16.04.2020

Erklärung gem. § 10 Abs. 3

Hiermit erkläre ich, dass ich für die Dissertation folgende Hilfsmittel und Hilfen verwendet habe:

Stata 15.1

QGIS 3.8.3

Microsoft Office 2016

TeXstudio 2.12.16

Mendeley 1.19.4

Auf dieser Grundlage habe ich die Arbeit selbstständig verfasst.

Unterschrift: _____

Contents

List of Tables

List of Figures

Summary

Chapter 1
The Long Shadow of the Short Run. Dynamics of Comparative Advantage with Heterogeneous Firms

This chapter presents a dynamic model of international trade with heterogeneous firms and endowment differences across countries to explore the adjustment of production and trade patterns to exogenous shocks, such as massive endowment changes. On the one hand, it generalizes the "two-by-two" model of Bernard, Redding, and Schott (2007) and shows that most of their propositions hold in a multi-country, multi-sector, and multi-factor case. This gives theoretical support to empirical research on the role of firms in countries' comparative advantage, which cannot be limited to a "two-by-two" case. On the other hand, instead of only comparing equilibria before and after the shock, the chapter focuses on transitory dynamics, or short- to medium-run responses. In this setting, households face a trade-off between current and future consumption, firms need investment to enter the market, and only a share of entrants are productive enough to stay in the market and produce. Thus, the shift to a new equilibrium cannot happen immediately after a shock. Instead, budgetary constraints on demand and investment lead to firm death shortly after the shock, even if industry expansion is expected in the long run. The economy then gradually converges to the new equilibrium, and, if the shock is massive, the transition process can last for several decades. This finding shows the opportunities for further research on the role of macroeconomic policy and targeted government investment in adjustment to such shocks.

Chapter 2
Is Buying on Amazon like Trading with a Digital Atlantis? E-commerce and Market Structure

E-commerce has been growing tremendously in the past decade. Yet, economics research lags behind in modeling and empirically investigating the impacts of ecommerce on intra-sectoral structures (such as market concentration). This chapter closes this gap by developing a model of e-commerce based on the Melitz framework for heterogeneous firms. It shows that some of e-commerce impacts on industrial structures are similar to those of trade liberalization. E-commerce too opens possibilities for firms to engage in new markets, while also increasing competitive pressure. However, contrary to trade liberalization, not necessarily only large, most productive firms profit from e-commerce. If the costs of e-commerce are sufficiently low, small, less productive enterprises can

use e-commerce to survive in the market, or even enter markets, where they would otherwise be uncompetitive. This yields a non-linear relationship between e-commerce costs and market concentration. This relationship is investigated with European data from 2005-2017. The data reveal a hump-shaped relationship between e-commerce costs and market concentration, thus supporting the theoretical findings.

Chapter 3
East Prussia 2.0: Persistent Regions, Rising Nations

This chapter examines the economic and political effects of the breakup of the German province of East Prussia into what is today Poland, Russia and Lithuania. It explores the dissolution of imperial regions into the boundaries of modern states, adding new insights to the research on the imperial legacies. The main hypothesis is that German imperial legacies in the form of advanced economic institutions, and specifically East Prussian legacies of nationalistic and conservative political preferences, persist in the territories of former East Prussia in Poland, Russia and Lithuania compared to neighboring regions in their respective countries. It turns out that no pattern of persistence is found in former East Prussian territories of contemporary Poland, whereas East Prussian persistence appears to be robust in Lithuania. There is strong evidence for the comparative persistence of political preferences in the Kaliningrad region (Russia), whereas no economic spillovers are observed. Evidence from West German electoral data in the aftermath of World War II indicates that the presence of East Prussian refugees is conducive to conservative and nationalist support in the FRG. Hence, the East Prussian legacy relates primarily to the persistence of political preferences and migrating agents.

Kurzfassung

Kapitel 1
Der lange Schatten der Kurzfristigkeit. Dynamik des komparativen Vorteils mit heterogenen Firmen

In diesem Kapitel wird ein dynamisches Modell des internationalen Handels mit heterogenen Firmen und Ressourcenunterschieden zwischen den Ländern vorgestellt, um die Anpassung der Produktions- und Handelsmuster an exogene Schocks wie massive Änderungen in der Ressourcenausstattung zu untersuchen. Einerseits wird das „2x2“-Modell von Bernard, Redding und Schott (2007) verallgemeinert und es wird gezeigt, dass die meisten ihrer Thesen auch in einem Fall mit mehreren Ländern, Sektoren und Faktoren gelten. Dies gibt theoretische Unterstützung für empirische Untersuchungen zur Rolle von Firmen in der Welt des komparativen Vorteils, die nicht auf einen „2x2“-Fall beschränkt werden können. Andererseits, anstatt nur Gleichgewichte vor und nach dem Schock zu vergleichen, konzentriert sich das Kapitel auf die Übergangsdynamik, oder kurz- bis mittelfristige Reaktionen. In diesem Modell stehen die Haushalte vor einem Kompromiss zwischen aktuellem und zukünftigem Verbrauch. Die Firmen benötigen Investitionen, um in den Markt einzutreten, und nur ein Teil der Marktteilnehmer ist produktiv genug, um auf dem Markt zu bleiben und zu produzieren. Daher kann die Anpassung an das neue Gleichgewicht nicht unmittelbar nach einem Schock erfolgen. Stattdessen führen Budgetbeschränkungen, die sowohl Nachfrage als auch Investitionen beeinflussen, kurz nach dem Schock zum Ausscheiden von Firmen, auch wenn langfristig eine Expansion des jeweilgien Sektors erwartet wird. Die Wirtschaft nähert sich dann allmählich dem neuen Gleichgewicht an, und wenn der Schock massiv ist, kann der Übergangsprozess mehrere Jahrzehnte dauern. Diese Erkenntnis zeigt die Möglichkeiten für weitere Untersuchungen zur Rolle der Wirtschaftspolitik und gezielter staatlicher Investitionen bei der Anpassung an solche Schocks.

Kapitel 2
Ist das Kaufen bei Amazon wie das Handeln mit einem digitalen Atlantis? Elektronischer Handel und Marktstruktur

Elektronischer Handel (E-Commerce) ist in den letzten Jahren enorm gewachsen. Die Wirtschaftsforschung bleibt jedoch bei der Modellierung und den empirischen Untersuchungen zu den Auswirkungen des E-Commerce auf sektorinterne Strukturen (z. B. Marktkonzentration) zurück. Dieses Kapitel schließt diese Lücke, indem ein Modell des E-Commerce entwickelt wird, das auf dem Melitz-Modell für heterogene Firmen

basiert. Es zeigt, dass einige der Auswirkungen des elektronischen Handels auf Industriestrukturen denen der Handelsliberalisierung ähnlich sind. Auch der E-Commerce eröffnet Unternehmen die Möglichkeit, neue Märkte zu erschließen, wobei gleichzeitig der Wettbewerbsdruck erhöht wird. Im Gegensatz zur Handelsliberalisierung profitieren jedoch nicht unbedingt nur große, produktivste Firmen vom elektronischen Handel. Wenn die Kosten für E-Commerce ausreichend niedrig sind, können kleine, weniger produktive Unternehmen den elektronischen Handel nutzen, um auf dem Markt zu bleiben oder sogar in Märkte einzutreten, in denen sie sonst nicht wettbewerbsfähig wären. Dies ergibt einen nichtlinearen Zusammenhang zwischen E-Commerce-Kosten und Marktkonzentration. Dieser Zusammenhang wird anhand europäischer Daten von 2005 bis 2017 untersucht. Die Daten zeigen eine buckelförmige Relation zwischen E-Commerce-Kosten und Marktkonzentration und stützen damit die theoretischen Ergebnisse.

Kapitel 3
Ostpreußen 2.0: Persistente Regionen, aufsteigende Nationen

In diesem Kapitel werden die wirtschaftlichen und politischen Auswirkungen des Zerfalls der deutschen Provinz Ostpreußen in das heutige Polen, Russland und Litauen untersucht. Das Kapitel erforscht die Auflösung imperialer Regionen in die Grenzen moderner Staaten und trägt der Erforschung des imperialen Erbes neue Erkenntnisse bei. Die Haupthypothese lautet, dass das deutsche kaiserliche Erbe in Form fortgeschrittener Wirtschaftsinstitutionen und insbesondere das ostpreußische Erbe nationalistischer und konservativer politischer Präferenzen in den Gebieten des ehemaligen Ostpreußens in Polen, Russland und Litauen im Vergleich zu den Nachbarregionen in ihren jeweiligen Ländern fortbesteht. Es erweist sich, dass in den ehemaligen ostpreußischen Gebieten des heutigen Polens kein Persistenzmuster zu finden ist, während die ostpreußische Persistenz in Litauen robust ist. Auch in der Region Kaliningrad (Russland) deuten die Daten auf vergleichsweise dauerhafte politische Präferenzen hin, während keine wirtschaftlichen Auswirkungen festgestellt wurden. Außerdem zeigen die westdeutschen Wahldaten nach dem Zweiten Weltkrieg, dass die Anwesenheit ostpreußischer Flüchtlinge die Unterstützung konservativer und nationalistischer Parteien in der BRD steigerte. Daher bezieht sich das ostpreußische Erbe in erster Linie auf das Fortbestehen politischer Präferenzen und die Migration.

Introduction

The global economy is growing not only in size but also in complexity. New products and financial instruments, new business models appear – also followed by new challenges. Economies across the world are more intertwined than ever. At the same time, the development of a digital world without borders, challenges of sustainable development and political instabilities require considerations beyond individual states. While it is in general accepted that trade and investment drive economic growth (Arkolakis et al. 2012; Eaton and Kortum 2001; Melitz and Redding 2014b), they are also coupled with different types of winners and losers of globalization (compare, for example, the Heckscher-Ohlin model, Krugman 1980 and Melitz 2003). At the same time, the more the countries are connected with each other, the easier the shocks are transmitted within the system. Furthermore, as Dani Rodrik suggests, on a macro-level there is a trade-off between deep economic integration, the nation state and democratic politics: two of them can be combined, but not all three (Rodrik 2011). Hence, international economics is also faced with a need to consider non-economic drivers and effects of economic integration and to be open for multidisciplinary approaches.

For research in economics and social sciences, it is always challenging to cope with the rapid pace of new global developments. There are substantial gaps in the literature regarding e.g. dynamic modeling of trade and research on the digital economy. The goal of this dissertation is to close those gaps by contributing to both theoretical and empirical research on the role of firms in trade and the economics of cultural and institutional persistence.

This dissertation consists of three essays, moving from short- to long-term analysis and from theoretical to empirical research. The overarching theme is the impact of disruptive transformations, be it a shock to countries' resources as in essay 1, a shock of new technology-based business models as in essay 2 or a demographic and institutional shock as in essay 3. Unfolding the different aspects of adjustment to such disruptions, I will go from dynamic (short-run) responses to resource shocks, through the steady-state analysis of the digital economy, to the long-run impacts of changing institutions and forced migration. The first two essays further share the same theoretical perspective, as I will focus on the role of firms in monopolistically competitive markets and show how the respective shocks affect the industrial structures. Who wins and who loses from these disruptions? What are the parallels to the problems economists have solved already? How are the impacts different from the developments we already know well? The third essay abstracts from this agent perspective to explore the insights that economic history can provide with respect to the long-lasting economic impacts of non-economic shocks. This non-economic perspective is especially important to take into account in view of the Rodrik's political trilemma. In today's intertwined world, where a shock in one country is easily transmitted to another, where not only money and goods but also people move

across the globe, the nation-states are often faced with seemingly blurred borders and a fear about the preservation of their own traditions and values. The question is: How resilient are values and attitudes to the shocks from the "outside world"? Can they persist if a country faces, e.g., a major demographic shock? Do they migrate together with the people? If so, what does it mean for the regions that host these people?

These three transformative shocks can also be viewed through the lens of system stability. If a system is pushed out of the equilibrium, will it return to it, or will something new emerge as a result? How far does the new system displace the old? What is the path from the old to the new? Can a system that was formed for years or even centuries be completely dismantled by an external shock, or will its remnants echo through the ages? While, of course, there is much literature devoted to these questions, the following three essays offer new insights, with both state-of-the-art theoretical approaches and empirical investigations of rich new data.

The first essay concentrates on the short-term dynamics after a shock to countries' resource endowments. The Heckscher-Ohlin notion of comparative advantage serves as a conceptual basis for this discussion. According to this concept, the relative endowments of production factors in a country determine its specialization and, hence, trade pattern. Essay 1 asks the following question: How do industries of comparative (dis-)advantage adjust to massive changes in relative endowments? While not receiving much attention in the literature, endowment shocks are, in fact, much more frequent than one might expect. With capital and labor usually considered primary production factors, human-made destruction and resettlement due to armed conflicts, border changes, and even large natural disasters can potentially lead to substantial capital and labor reallocations. In fact, essay 3 will discuss – from a different perspective – an example of such reallocation, namely, the resettlement of German population from the eastern German territories seceded to Poland and the Soviet Union after World War II. Extending the focus beyond two production factors, discoveries of mineral resources add to frequent endowment shocks. Building upon the model of comparative advantage with heterogenous firms by Bernard, Redding, et al. 2007, essay 1 develops a model of dynamic responses to shocks. I show that, while the long-term consequences are in line with the recent literature on comparative advantage, the short-term changes often are *contrary to it or show overreaction to the shock*. Even a positive endowment shock can lead to a short but profound downturn in one or more industries. Moreover, transition can last for decades, depending on the type and scale of the shock, calling for a careful consideration of the term "short run". These findings not only open the way for a discussion on possibilities for government intervention to facilitate transition and offset short-run negative consequences, but also suggest caution when interpreting empirical research on shocks and comparative advantage, which often rely on relatively short time periods.

The elegance of the notion of comparative advantage, originally introduced by Ricardo, manifests itself not only in its power to describe trade among dissimilar countries, but also in its applicability to different agents and sources of relative efficiency. The second essay employs this concept in an investigation of a different type of disruptive shock. Focusing on the emergence of e-commerce, essay 2 again utilizes the framework on heterogenous firms in monopolistically competitive markets to explore the economic impacts of introducing a new production or distribution technology. Extending the framework introduced by Melitz 2003, it models introduction of e-commerce as an emergence of a new variety in the

sector, which is an imperfect substitute for the traditional varieties. On the firm side, it is a second variety that comes with different costs and can be optionally offered together with or instead of the traditional variety. As a result, emergence of e-commerce has many effects similar to those of trade liberalization: it leads to higher competitive pressure, lower sectoral prices and expansion of large firms. Because of its optionality, however, e-commerce can also benefit small and medium-sized firms, *if it comes with a comparative advantage*, i.e. with lower costs relative to the traditional variety. This means, that the introduction of e-commerce has non-linear impacts and allows for different scenarios, depending on the relative cost of e-commerce. At high costs, e-commerce will profit large firms and lead to higher market concentration. At very low costs, e-commerce will lead to a "wave of start-ups" and decrease market concentration. The essay further investigates European data on e-commerce adoption and e-commerce drivers and provides empirical support for the existence of these scenarios.

In the empirical investigation of essay 2, I acknowledge how institutional factors and diversity of consumer preferences can influence the costs of a new technology. More generally, the relevance of institutions and culture for economic behavior has been widely accepted in the literature (Acemoglu et al. 2001; North 2006; Tabellini 2010). Thus, essay 3 fully focuses on this perspective and discusses how persistence of culture in a formerly homogenous region, on which different institutional environments are imposed, depends on the scale of the related demographic shock. As an empirical case, essay 3 investigates the partition of the German province of East Prussia between three states in the aftermath of World War I and World War II. The main hypothesis is that the legacy of East Prussia in terms of political preferences and entrepreneurial activity persists in today's Lithuania, Poland and Russia both in relation to the neighboring territories in these countries and in closer similarities between these countries within former East Prussian territories than outside of it. It turns out, however, that this persistence indeed *depends crucially on the scale of the demographic shock* in East Prussia after World War II. Moreover, the East Prussian political legacy did "migrate" together with the East Prussians, affecting the results of the first parliamentary election in 1949 in the new Federal Republic of Germany.

The three essays create a number of spillovers for the future research. Based on essay 1, I sharpen the focus of the generalized model of comparative advantage and show in Polugodina 2019 how a windfall of resources which can be used for production has a controversial effect on the economy, leading both to a lower overall productivity in the economy (Dutch disease) and a higher productivity in the comparative advantage sectors. Essay 2 provides numerous theoretical insights into the effects of e-commerce, and testing them all empirically would go beyond the scope of one essay. It, thus, suggests several directions for the future theoretical and empirical research. Essay 3 provides a critical perspective on the studies of persistence and the channels that are explicitly or implicitly assumed in these studies.

It should also be noted that new rich datasets were generated for this dissertation. For essay 2, two panel datasets on e-commerce, sector-level and institutional country-level data were assembled from numerous sources, which can be employed well beyond the topic of e-commerce. Essay 3 not only utilized already digitized data on Germany but also compiled detailed data on elections in Lithuania, Poland and Russia, which are geocoded and can readily be used for further regional studies.

1 The Long Shadow of the Short Run: Dynamics of Comparative Advantage with Heterogeneous Firms

1.1 Introduction

Trade can drive economic development, for example, by enhancing specialization and, therefore, productivity growth (e.g. Frankel and Romer 1999; Eaton and Kortum 2001; Dollar and Kraay 2003; López 2005; Melitz and Redding 2014b). Thus, the question of the determinants of specialization patterns (well captured by trade patterns) and their changes is no less important. This is coupled with the problem of analyzing a wide range of winners and losers from trade (compare, for example, the implications of the Heckscher-Ohlin, Krugman 1980 and Melitz 2003 models). Current models of trade typically focus on changes caused by shifts in trade costs, while changes in endowments, which, as will be discussed later, happen much more frequently than one might expect, are typically ignored. With the impacts that an endowment shock can have both in terms of economic reorganization and wealth redistribution, such ignorance seems remarkable.

This chapter builds mainly upon Bernard, Redding, et al. 2007 and Ghironi and Melitz 2005 to develop a formal multi-factor and multi-sector model of dynamic industry responses to macroeconomic and endowment shocks in the presence of trade costs and firm heterogeneity. My contribution is twofold. Firstly, I generalize the model by Bernard, Redding, et al. 2007 beyond the specific case of two countries, industries, or production factors, showing that their propositions continue to hold, albeit in a more general formulation. Because empirical studies typically deal with more than two countries, industries, or production factors, the correctness of their approaches depends crucially on generalizability of the underlying "two-by-two" models. Such generalizations exist for the simpler Heckscher-Ohlin model (see e.g. the overview in Feenstra 2016, Ch. 3), but not for the modern approach. While adjusting the analysis to multiple countries is not necessarily critical, understanding model behavior with multiple sectors and production factors is. The majority of empirical tests, for example of the Rybczinski theorem or of the interplay between Heckscher-Ohlin and Ricardian comparative advantage, rely on differentiating between multiple sectors and usually also at least between skilled and unskilled labor (making, in total, at least three production factors to account for). Recent empirical research is no longer based on the simplest Heckscher-Ohlin model either: the focus has shifted towards investigating the role of firms in adjusting to endowment changes (see e.g. Dustmann and Glitz 2015). For the theoretical validity of such tests, it is vital that modern approaches to comparative advantage can be extended, while still yielding the same, or comparable, results as the "two-by-two" models.

Secondly, besides comparing equilibria before and after the shock, I also calibrate the model to analyze transitory dynamics, or short-run responses, showing that while the long-term consequences are in line with the recent literature on comparative advantage, the short-term changes often are contrary to it or show overreaction to the shock. The important driver of the transition dynamics is the behavior of households, who provide not only access to production factors, but also investment for the firms that wish to enter the market. Thus, there exists a trade-off between current consumption and future consumption, which is secured by current investment. This leads to an interplay of changing demand conditions and competitive toughness, along with levels of investment insufficient to ensure immediate full adjustment to the new equilibrium. Therefore, even a positive endowment shock can lead to a short but profound downturn in one or more industries, followed by a catch-up to the new equilibrium. Since such a downturn is mainly caused by contracting demand and falling price levels, which force less productive firms to leave the market, this opens the way for tools of expansionary fiscal and monetary policy to prevent excessive firm death and mitigate the negative effects of the shock. Also, transition can last for decades, depending on the type and size of the shock, with economies adjusting somewhat faster to a positive shock than to a negative one. Since the constraint on adjustment speed lies in scarce investment, this might also open a possibility for government intervention to facilitate transition and offset short-run negative consequences. Apart from possible policy implications, however, this result suggests caution when interpreting empirical research on shocks and comparative advantage. For example, Blum 2010 analyzes 27 developing and developed countries between 1973 and 1990 and finds that there is little adjustment of production structure to changes in endowments, even in the long-run – after 15 years. He explains this result with innovation bias. My findings suggest that a further possible reason is simply that 15 years might not yet be a "long run". While not undervaluing the importance of technological change, I argue that it is important to follow the development after a shock for much longer periods when testing adjustment dynamics for comparative advantage.

There exists a large body of theoretical literature on comparative advantage that enhances the well-known and very restrictive Heckscher-Ohlin framework with more realistic features such as imperfect competition (Helpman and Krugman 1985), the existence of trade costs, and heterogeneity across firms (Bernard, Redding, et al. 2007; Burstein and Vogel 2011). There is also a growing literature on the short- and mid-term dynamics of firms, trade, and macroeconomic aggregates after a shock (Chen et al. 2009; Costantini and Melitz 2008; Ghironi and Melitz 2005). What is missing, however, is the link between comparative advantage, on the one hand, and firm and trade dynamics, on the other. One can readily analyze differing steady states by using, for example, the model of Bernard, Redding, et al. 2007, but the transition dynamics in an economy where production factors and demand can move across sectors and the speed of such transitions are unclear. To my knowledge, the only two contributions that have so far analyzed dynamics in economies with comparative advantage and heterogeneous firms are Lechthaler and Mileva 2016 and Lechthaler and Mileva 2019. They, however, focus on the effects exerted on the labor market.

Furthermore, while the macroeconomic literature typically focuses on such exogenous shocks as changes in aggregate productivity, production costs (Ghironi and Melitz 2005; Lechthaler and Mileva 2016) or trade openness (Chen et al. 2009; Lechthaler and Mileva

2019), my interest lies primarily in shocks to endowments of production factors, which are not as rare as one might expect. With capital and labor usually considered primary production factors, human-made destruction and resettlement due to armed conflicts, border changes, and even large natural disasters can potentially lead to substantial resource reallocations. Massive in- and outmigration and capital flights can also result in considerable shifts in relative endowments (although such shocks are usually not completely exogenous). Extending the focus beyond two production factors, it is clear that discoveries of mineral resources (e.g. new gas or oil fields, such as the Groningen gas field discovered in 1959 in the Netherlands) add to frequent and massive endowment shocks.

The rest of the chapter is structured as follows. Section 1.2 outlines the formal setup of the model and briefly presents its implications for comparative statics, which are mostly in line with Bernard, Redding, et al. 2007. Section 1.3 discusses the implications for transition dynamics, exploring additional assumptions necessary for calibration in the first two parts and presenting the results of the simulation in the third part. Finally, Section 1.4 concludes.

1.2 Model setup

The basic setup for heterogeneous firms trading in a one-good, one-factor economy was developed by Melitz 2003. Bernard, Redding, et al. 2007 suggested a model with two countries, two factors, and two goods to incorporate firm heterogeneity into the world with comparative advantage. Extending Bernard, Redding, et al. 2007, I will consider a world with S countries (with h indexing the home country and f indexing the foreign country or countries), F production factors (indexed by j, k) and N sectors (indexed by i, l), each featuring a continuum of varieties.[1] For simplicity, it will be assumed that the countries are symmetric up to their factor endowments.[2] In addition to this basic setup, consumers will make an intertemporal choice about their consumption, as in Ghironi and Melitz 2005, so that the mechanism of adjustment to new equilibria can be traced.

1.2.1 Consumption

Consumer preferences are described with a three-level utility function. The upper level characterizes the choice over consumption level in each period, which is made by maximizing intertemporal utility:

[1] The largest part of the model will only be written out for the home country, with all the equations being identical for the foreign country up to the exchange rate reversal. When providing an expression featuring one country's variables, only, the country index will be omitted for notational simplicity. Notions of sectors, industries, and goods will be used interchangeably. While the model could allow for an arbitrary number of factors and industries, in my analysis a restriction of symmetry ($F = N$) will be imposed to ensure that, with known factor prices, factor market clearance is a sufficient determinant for allocation of factors across industries.

[2] Although I will write the model out with the respective country indices for costs, the symmetry assumption will be actively used when discussing implications of the model and running the simulation.

$$U = \left(\sum_{s=t}^{\infty} \beta^{s-t} \frac{C_s^{1-\gamma}}{1-\gamma} \right), \tag{1.1}$$

where C_s is the consumption in period s, $\beta \in (0,1)$ is the discount factor, and $\gamma > 0$ is the inverse of the intertemporal elasticity of substitution.

In each period t, the consumer basket is comprised of goods from N sectors of the economy. Preferences over sectors are given by a Cobb-Douglas utility function, with $\alpha_i \in (0,1)$ denoting the share of income devoted to sector i:

$$C_t = \prod_{i=1}^{N} (Q_{it})^{\alpha_i}, \quad \sum_{i=1}^{N} \alpha_i = 1. \tag{1.2}$$

Finally, the lowest level of the utility function describes consumer preferences across varieties in each sector i. As is usual in the literature on heterogeneous firms in trade, I will assume that the consumption aggregate in each sector is characterized by CES preferences over differentiated varieties $q_{it}(\omega)$, which can, for example, be viewed as different types or brands of some product:

$$Q_{it} = \left(\int_{\omega \in \Omega_{it}} q_{it}(\omega)^{\frac{\theta-1}{\theta}} d\omega \right)^{\frac{\theta}{\theta-1}}, \tag{1.3}$$

where $\theta > 1$ is the elasticity of substitution across varieties. For simplicity, it is assumed that it is the same in all sectors, but this assumption can readily be omitted. Ω_{it} is a period t subset of all possible varieties in sector i ($\Omega_{it} \subset \Omega_i$): in any period t, not all potential varieties may be present in the market. This also means that at any two points of time, the sets of varieties present in the market can differ, which, as will be shown later, reflects the process of firm creation and destruction.

The sectoral price index P_{it} over variety prices $p_{it}(\omega)$ is then defined as

$$P_{it} = \left(\int_{\omega \in \Omega_{it}} p_{it}(\omega)^{1-\theta} d\omega \right)^{\frac{1}{1-\theta}}. \tag{1.4}$$

Finally, the overall price index for the whole economy is the weighted average across the industries:

$$P_t = \prod_{i=1}^{N} \frac{(P_{it})^{\alpha_i}}{\alpha_i}. \tag{1.5}$$

1.2.2 Production, pricing, and profits

There is a continuum of firms operating in a monopolistically competitive market in each of the sectors, with each firm producing a single differentiated variety. There are F factors necessary to produce any variety, and their nominal prices are denoted by W_{jt}. Technologies are identical across countries, but sectors within the countries differ in their factor intensity.[3] Firms differ in their productivity, modeled as a cost shifter φ, with higher productivity lowering the variable production costs. The productivity level of each firm is unknown ex ante and is revealed when the firm actually enters the market; it is drawn randomly from a (known) common distribution $G(\varphi)$. Throughout the firm's lifecycle, its productivity remains unchanged.[4] Domestic production is also subject to a fixed cost of $f_{dit} > 0$ units of a composite production factor.[5] The fixed production cost may refer, for example, to having an administrative unit or to minimal energy supply and other fixed maintenance costs for a plant. Following Bernard, Redding, et al. 2007, I assume that the production function has a Cobb-Douglas form over the usage of production factors, and also assume that fixed costs have the same factor intensity as variable costs in the respective sector. The production cost for variety $q_{it}(\omega)$ can therefore be expressed as

$$TC_{it} = \left(f_{dit} + \frac{q_{it}}{\varphi}\right) \prod_{j=1}^{F} \left(W_{jt}\right)^{\mu_{ij}}, \quad \sum_{j=1}^{F} \mu_{ij} = 1,$$

where $\mu_{ij} \in (0, 1)$ is the cost-share of factor j in sector i. In real terms,

$$tc_{it} = \left(f_{dit} + \frac{q_{it}}{\varphi}\right) \prod_{j=1}^{F} \left(w_{jt}\right)^{\mu_{ij}}, \quad w_{jt} = \frac{W_{jt}}{P_t}, \quad \sum_{j=1}^{F} \mu_{ij} = 1. \tag{1.6}$$

If a firm also exports, for any supply to the foreign market there will be a (bilateral) variable (iceberg) cost $\tau_{it}^{hf} > 1$ (meaning that τ units have to be shipped from the factory in country h so that one unit arrives at the destination in country f) and a fixed export cost of $f_{xit}^{hf} > 0$ units of a composite production factor, which I again assume to have the same factor intensity as variable production costs. The variable trade cost can reflect such ad valorem costs as freight, insurance, and import tariffs. Fixed export costs may refer, for example, to setting up and sustaining the distribution network and service centers. The fixed costs of production are not explicitly included in export activity as it is assumed that any exporting firm also sells domestically and the respective fixed costs are accounted for in domestic production.

With the costs described above and consumer love for variety featuring the same constant elasticity of demand for all industries and countries, the maximization of firm profit will

[3] It is straightforward to add a measure of overall technology level differences or differing factor intensities across sectors and countries to the cost function, but I abstain from that for notational and calculation simplicity. With average productivity in the industry determined endogenously, differences across sectors can, in the presence of endowment differences, produce Ricardian advantage or disadvantage on their own.

[4] This is supported empirically, for instance, by Bernard and Jensen 1999 and Clerides et al. 1998.

[5] I assume fixed costs independent from firm productivity. See Bernard, Redding, et al. 2007 for discussion.

result in the export price of a variety being proportional to the domestic price, which is, in turn, a constant mark-up over the marginal cost:

$$p_{dit}^{h} = \prod_{j=1}^{F} \left(W_{jt}^{h}\right)^{\mu_{ij}} \frac{1}{\varphi} \frac{\theta}{\theta - 1},$$
$$p_{xit}^{hf} = \prod_{j=1}^{F} \left(W_{jt}^{h}\right)^{\mu_{ij}} \frac{\tau_{it}^{hf}}{\varepsilon_{t}^{hf} \varphi} \frac{\theta}{\theta - 1},$$

where ε_t^{hf} is the bilateral nominal exchange rate between the home and the foreign country.

Re-writing prices in real terms as well, we get:

$$\rho_{dit}^{h} = \prod_{j=1}^{F} \left(w_{jt}^{h}\right)^{\mu_{ij}} \frac{1}{\varphi} \frac{\theta}{\theta - 1},$$
$$\rho_{xit}^{hf} = \prod_{j=1}^{F} \left(w_{jt}^{h}\right)^{\mu_{ij}} \frac{\tau_{it}^{hf}}{\mathcal{E}_{t}^{hf} \varphi} \frac{\theta}{\theta - 1}. \tag{1.7}$$

The real exchange rate $\mathcal{E}_t^{hf} = \varepsilon_t^{hf} P_t^f / P_t^h$ is determined endogenously based on purchasing power parity, while the nominal exchange rate is taken as exogenous. For the foreign country, $\mathcal{E}_t^{fh} = 1/\mathcal{E}_t^{hf}$.

I assume that domestic and foreign markets are separated. Thus, the decision whether to export or not to any of the markets only depends on the respective export costs and foreign demand. Therefore, prices, revenues, and profits from the exporting activity can also be separated from those of the domestic activity and other export markets. Equilibrium domestic and foreign revenues of the home country companies can then be expressed as increasing functions of firm productivity:

$$r_{dit}^{h}(\varphi) = \alpha_i C_t^h \left(\frac{\rho_{dit}^{h}(\varphi)}{\mathcal{R}_{it}^{h}} \right)^{1-\theta} = \alpha_i C_t^h \left(\frac{\mathcal{R}_{it}^{h} \varphi}{\prod_{j=1}^{F} \left(w_{jt}^{h}\right)^{\mu_{ij}}} \frac{\theta - 1}{\theta} \right)^{\theta - 1},$$
$$r_{xit}^{hf}(\varphi) = \mathcal{E}_t^{hf} \alpha_i C_t^f \left(\frac{\rho_{xit}^{hf}(\varphi)}{\mathcal{R}_{it}^{f}} \right)^{1-\theta} = \mathcal{E}_t^{hf} \alpha_i C_t^f \left(\frac{\mathcal{E}_t^{hf} \mathcal{R}_{it}^{f} \varphi}{\tau_{it}^{hf} \prod_{j=1}^{F} \left(w_{jt}^{h}\right)^{\mu_{ij}}} \frac{\theta - 1}{\theta} \right)^{\theta - 1}, \tag{1.8}$$

where $\mathcal{R}_{it} = P_{it} \,/\, P_t$ is the real price index for industry i. Revenue decreases with the product's price (or costs) and increases with demand elasticity, industry price level, and aggregate consumption in the recipient market. The exporter will also receive higher revenue if the currency of his home country is depreciated.

If there exists selection to the export market, not all the firms selling domestically also decide to export. Total firm revenue is therefore given by

$$r_{it}^{h}(\varphi) = \begin{cases} r_{dit}^{h} + \sum_{f \in B} r_{xit}^{hf} \text{ , if the firm exports to a subset } B \text{ of countries,} \\ r_{dit}^{h} \text{ , if the firm does not export.} \end{cases} \tag{1.9}$$

Domestic and export profits can be expressed as revenue scaled by the elasticity of substitution less fixed cost and, consequently, are also monotonically increasing in φ:

$$\begin{aligned} \pi_{dit}^{h} &= \frac{r_{dit}^{h}}{\theta} - f_{dit}^{h} \prod_{j=1}^{F} \left(w_{jt}^{h}\right)^{\mu_{ij}}, \\ \pi_{xit}^{hf} &= \frac{r_{xit}^{hf}}{\theta} - f_{xit}^{hf} \prod_{j=1}^{F} \left(w_{jt}^{h}\right)^{\mu_{ij}}. \end{aligned} \tag{1.10}$$

Any firm operating domestically can decide to export to a foreign market at any point of time. There is no additional cost of entering the export market, and so it is only economically reasonable to export if a firm can receive nonnegative profit from the exporting activity: $\pi_{xit}^{hf} > 0$.[6]

Total firm profit is then

$$\pi_{it}^{h} = \pi_{dit}^{h} + \sum_{f \neq h} \max\left(0, \pi_{xit}^{hf}\right). \tag{1.11}$$

To initially start producing in an industry domestically, though, a firm has to bear a fixed entry cost, which is sunk after the firm attempts to produce. As well as other fixed costs, I assume that the entry cost has the same factor intensity as the variable costs. The payment the firm entering industry i in period t has to make is thus:

$$f_{eit} \prod_{j=1}^{F} \left(w_{jt}\right)^{\mu_{ij}} \quad , \quad f_{eit} > 0. \tag{1.12}$$

Only after the entry cost is paid does the firm reveal its productivity φ. If its productivity is so low that it would receive negative profits from domestic production, it exits the market immediately. If its productivity is high enough to earn nonnegative profits, it produces domestically and makes a decision about exporting. Since profit is a monotonic function of productivity, there will exist a unique zero-profit productivity cut-off φ_{dit} such that:

$$\alpha_i C_t \left(\frac{\mathcal{R}_{it} \varphi_{dit}}{\prod_{j=1}^{F} w_{jt}^{\mu_{ij}}} \frac{\theta - 1}{\theta} \right)^{\theta - 1} = f_{dit} \prod_{j=1}^{F} \left(w_{jt}\right)^{\mu_{ij}}. \tag{1.13}$$

[6] Although in reality there are substantial sunk costs of market penetration, for simplicity reasons it is rather usual to model export costs as per-period payments (see Melitz and Redding 2014a for discussion). These can be viewed as comprised of operation costs and installments for repayment of the initial investment.

Similarly, there exists an exporting productivity cutoff φ_{xit}^{hf} as a solution to the zero-profit condition for exports to any market f:

$$\mathcal{E}_t^{hf}\alpha_i C_t^f \left(\frac{\mathcal{E}_t^{hf}\mathcal{R}_{it}^f \varphi_{xit}^{hf}}{\tau_{it}^{hf}\prod_{j=1}^{F}\left(w_{jt}^h\right)^{\mu_{ij}}} \frac{\theta-1}{\theta} \right)^{\theta-1} = f_{xit}^{hf}\prod_{j=1}^{F}\left(w_{jt}^h\right)^{\mu_{ij}} . \tag{1.14}$$

Combining expressions (1.13) and (1.14) yields an equilibrium relationship between the domestic and the export cut-offs:

$$\varphi_{xit}^{hf} = \tau_{it}^{hf}\frac{\mathcal{R}_{it}^h}{\mathcal{E}_t^{hf}\mathcal{R}_{it}^f}\left(\frac{C_t^h}{\mathcal{E}_t^{hf}C_t^f}\frac{f_{xit}^{hf}}{f_{dit}^h}\right)^{\frac{1}{\theta-1}} \varphi_{dit}^h . \tag{1.15}$$

If $\varphi_{xit}^{hf} > \varphi_{dit}^h$, selection exists in the export market, meaning that only more productive firms export, while the less productive ones can only sell domestically. The larger the difference between the domestic and the export cut-offs, the lower the share of exporters. This will be the case, for example, if the variable export cost is significant, thereby raising the price and reducing revenue in the foreign market. The difference will also be large if the fixed cost of exporting is high relative to the fixed cost of domestic production. This would mean that higher revenue is required to cover the fixed cost of exporting and only more productive firms can therefore earn nonnegative profits in the foreign market. Furthermore, the share of exporters will be low if the real exchange rate (based on purchasing power parity) is unfavorable or if the importing country is relatively small, thus representing a smaller and more competitive market.

1.2.3 Industry averages and aggregation

At any time point t, there are M_{dit} domestic firms operating in each industry i. Ex-post distribution of firm productivity is conditional upon successful entry into the domestic market and is, therefore, truncated at the ex-ante probability of successful entry (the zero-profit cut-off φ_{dit}). With the ex-ante productivity distribution $G(\varphi)$, this yields the following ex-post average productivity in the industry:

$$\widetilde{\varphi}_{dit}\left(\varphi_{dit}\right) = \left(\frac{1}{1-G\left(\varphi_{dit}\right)}\int_{\varphi_{dit}}^{\infty}\varphi^{\theta-1}dG\left(\varphi\right)\right)^{\frac{1}{\theta-1}} , \tag{1.16}$$

where $1-G(\varphi_{dit})$ is the ex-ante probability of successful market entry.

Similarly, the mass of exporters and their ex-post productivity distribution are conditional upon receiving nonnegative profit from the exporting activity. Of the firms trying to enter the domestic market, only a share of $1-G(\varphi_{xit}^{hf})$ will not only produce domestically, but also decide to sell on the foreign market h. Ex-post average exporter productivity is then:

$$\widetilde{\varphi}_{xit}^{hf}\left(\varphi_{xit}^{hf}\right)=\left(\frac{1}{1-G\left(\varphi_{xit}^{hf}\right)}\int_{\varphi_{xit}^{hf}}^{\infty}\varphi^{\theta-1}dG\left(\varphi\right)\right)^{\frac{1}{\theta-1}}. \tag{1.17}$$

Average productivity is a monotonic function of the productivity cut-off, and so if selection in the export market exists, it also leads to the average productivity of exporters becoming higher than the average productivity of the domestic firms: $\widetilde{\varphi}_{xit}^{hf}>\widetilde{\varphi}_{dit}^{h}$.

The share of exporters in the industry depends on both the zero-profit cut-off and the export cut-off and can be defined as:

$$\chi_{it}^{hf}=\frac{1-G\left(\varphi_{xit}^{hf}\right)}{1-G\left(\varphi_{dit}^{h}\right)}. \tag{1.18}$$

As noted earlier, this implies that the larger the difference between the zero-profit and the export cut-offs, the lower the share of exporters in the industry.

Average productivity refers to the productivity of a firm earning an average weighted profit:

$$\tilde{\pi}_{it}^{h}=\tilde{\pi}_{dit}^{h}+\sum_{f\neq h}\chi_{it}^{hf}\tilde{\pi}_{xit}^{hf}=\pi_{dit}\left(\widetilde{\varphi}_{dit}^{h}\right)+\sum_{f\neq h}\chi_{it}^{hf}\pi_{xit}\left(\widetilde{\varphi}_{xit}^{hf}\right)$$

In the same way, it refers to average revenue:

$$\tilde{r}_{it}^{h}=\tilde{r}_{dit}^{h}+\sum_{f\neq h}\chi_{it}^{hf}\tilde{r}_{xit}^{hf}=r_{dit}\left(\widetilde{\varphi}_{dit}^{h}\right)+\sum_{f\neq h}\chi_{it}^{hf}r_{xit}\left(\widetilde{\varphi}_{xit}^{hf}\right).$$

Knowing the average revenue per firm and the aggregate industry revenue R_{it}, it is straightforward to express the mass of domestic firms as

$$M_{dit}=\frac{R_{it}}{\tilde{r}_{it}} \tag{1.19}$$

The same holds symmetrically for all of the foreign countries. Thus, in any period t, consumers in the home country are offered M_{dit}^{h} varieties of product i by the domestic firms and $\chi_{it}^{fh}M_{dit}^{f}$ varieties by exporters from each of the foreign countries. Using the equilibrium pricing rule, the industry price index is a weighted average of average variety prices in the home and the foreign countries:

$$\mathcal{R}_{it}^{h}=\left(M_{dit}^{h}\left[\tilde{\rho}_{dit}^{h}\right]^{1-\theta}+\sum_{f\neq h}\chi_{it}^{fh}M_{dit}^{f}\left[\tilde{\rho}_{xit}^{fh}\right]^{1-\theta}\right)^{\frac{1}{1-\theta}}, \tag{1.20}$$

where $\tilde{\rho}_{dit}^{h}=\rho_{dit}(\widetilde{\varphi}_{dit}^{h})$ and $\tilde{\rho}_{xit}^{fh}=\rho_{xit}(\widetilde{\varphi}_{xit}^{fh})$ are prices an average domestic and foreign firm, respectively, charge in the home sector i.

In equilibrium, trade should be balanced. With multilateral clearing, this means that value of total imports of a country should be equal to value of its total exports:

$$\sum_{f\neq h}\sum_{i=1}^{N}\mathcal{E}_t^{hf}\alpha_i C_t^f\left(\frac{\tilde{\rho}_{xit}^{hf}}{\mathcal{R}_{it}^f}\right)^{1-\theta}\chi_{it}^{hf}M_{dit}^h=\sum_{f\neq h}\sum_{i=1}^{N}\alpha_i C_t^h\left(\frac{\tilde{\rho}_{xit}^{fh}}{\mathcal{R}_{it}^h}\right)^{1-\theta}\chi_{it}^{fh}M_{dit}^f. \tag{1.21}$$

Since the real exchange rate is endogenous, it will adjust so that trade is balanced. To ensure market clearing, aggregate production on the industry level should be equal to the sum of domestic and foreign demand:

$$R_i^h=\alpha_i C_t^h M_{dit}^h\left(\frac{\tilde{\rho}_{dit}^h}{\mathcal{R}_{it}^h}\right)^{1-\theta}+\sum_{f\neq h}\mathcal{E}_t^{hf}\alpha_i C_t^f\left(\frac{\tilde{\rho}_{xit}^{hf}}{\mathcal{R}_{it}^f}\right)^{1-\theta}\chi_{it}^{hf}M_{dit}^h. \tag{1.22}$$

1.2.4 Household budget constraint, firm value, and factor markets

What is largely missing in typical models of heterogeneous firms is how firm entry is financed and how profits are used. Although not relevant for a comparative statics analysis, explicit modeling of such financial flows can yield important implications for the transition from one steady state to another. To introduce dynamics to the generalization of Bernard, Redding, et al. 2007, I will follow Ghironi and Melitz 2005 and assume that investments and their returns flow only between households and firms. The representative household owns all factors of production. It can also buy bonds to the amount of B_t, which bring it an income of $(1+r_{t+1})B_t$ in the next period, and a share x_t of a mutual fund of home firms entering in period t. Importantly, there are no inherent growth drivers in this model. The investments go solely into the introduction of new products (firm entry), not into capital accumulation, as is usually assumed in models of endogenous economic growth.

There exists a time-to-build lag, meaning that firms entering the market in period t only start producing and begin receiving profits in period $t+1$. So in each period t, the household buys share x_t of the mutual fund of $M_{ht}=\sum_{i=1}^{N}M_{hit}$ home firms, with $M_{hit}=M_{dit}+M_{eit}$ being the total number of incumbents and firms that entered the market in period t. Of the mass of firms entering industry i in period t, only share $[1-G(\varphi_{dit})]$ will draw high enough productivity to stay in the market and produce in period $t+1$.

There are no cross-border investments: households do not invest in foreign firms. However, factor returns can come from different sectors (e.g. a worker employed in one sector is free to get his capital and land rents from providing them to other sectors as well). In addition, households hold a share in a diversified investment portfolio, which includes firms from all sectors. Portfolio income in period t consists of total firm profits $\sum_{i=1}^{N}M_{dit}\tilde{\pi}_{it}$, which are paid out as dividends in full (there are no retained earnings).

There exists an exogenous negative shock that hits a share $\delta\in(0,1)$ of firms at the end of each period. This shock can hit both incumbents and entrants in any industry. Therefore,

the number of incumbents in period $t+1$ is $M_{di,t+1} = (1-\delta)(M_{dit} + [1 - G(\varphi_{dit})]M_{eit})$. It is unknown which firms will be hit by this "death" shock, and so δ is perceived by firms and households as the probability of the shock. Because of this uncertainty, households finance all incumbents and entrants. In the steady state, where the number of firms in an industry is constant over time, this rule will determine the amount of investment in each period.

Denoting average firm value – and, thus, the price of the respective securities in the portfolio – as $\tilde{v}_{it}$, we can formulate the aggregated budget constraint of the households:

$$C_t + B_t + x_t \sum_{i=1}^{N} M_{hit} \tilde{v}_{it} = \sum_{j=1}^{F} w_{jt} \overline{F}^j + (1 + r_t) B_{t-1} + x_{t-1} \sum_{i=1}^{N} M_{dit} \left(\tilde{v}_{it} + \tilde{\pi}_{it} \right), \quad (1.23)$$

where $\overline{F}^j$ is a country's endowment of factor j.

Under the assumption that asset valuation is done on a cash flow basis, and neither cash flows nor productivity levels in one industry affect firm valuation in other industries, maximization of the utility function (1.1) subject to the budget constraint (1.23) yields the following Euler equations for bonds and shares:

$$\begin{aligned} &Bonds: C_t^{-\gamma} = \beta \left(1 + r_{t+1}\right) C_{t+1}^{-\gamma}, \\ &Shares: \tilde{v}_{it} = (1-\delta)\, \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \left(\tilde{v}_{i,t+1} + \tilde{\pi}_{i,t+1} \right) \left(1 - \frac{G\left(\varphi_{dit}\right) M_{eit}}{M_{dit} + M_{eit}} \right). \end{aligned} \quad (1.24)$$

The last term in the equation for shares can be considered as an inverse measure of industry risk. The larger (and thus, more diversified) and the more productive the industry, the lower the risk. In equilibrium, firm value is determined under assumptions of free market entry and the existence of an unbound edge of potential entrants. There are no barriers to entry except the (sunk) fixed entry cost, which is the same for all the firms in the industry. The number of firms wishing to get investment is unlimited. This means that, in equilibrium, the entry of new firms will keep firm value at a minimal level, just enough to cover the entry cost, as long as the number of entrants is positive:

$$\tilde{v}_{it} = f_{eit} \prod_{j=1}^{F} \left(w_{jt} \right)^{\mu_{ij}}. \quad (1.25)$$

Finally, in equilibrium, factor markets should clear, which means that factor usage in all sectors of the economy should sum up to total endowment of the country:

$$\sum_{i=1}^{N} F_{it}^j = \overline{F}^j, \quad F_{it}^j = F_{it}^{je} + F_{it}^{jp} + F_{it}^{jx}, \; j = 1, \ldots F, \quad (1.26)$$

where indices e, p, x denote the factor employed for market entry, domestic production, and exporting activity, respectively.

1.2.5 Implications for the steady state

Except for the dynamic component, the structure of the model is an immediate generalization of Bernard, Redding, et al. 2007. This feature of the model makes the analysis of comparative statics quite straightforward once we consider simple steady states with zero economic growth. Most of propositions by Bernard, Redding, et al. 2007 will hold, albeit some have to be re-formulated in a more general form. Most of the proposition proofs are similar as well and are, therefore, presented in the appendix. Throughout the rest of the chapter, I will only be interested in internal solutions for the equilibria and assume that no shock can move the economy into a corner solution and make it give up some sectors and fully specialize in others. This assumption is not only simplifying but also highly plausible, unless we look at very disaggregated production patterns. In addition, it affects only few propositions.

The main propositions can be grouped into four categories: a plain comparison of countries with different endowments (under autarky or free trade), moving from autarky to free trade, moving from autarky to costly trade, and a change in endowments, a group not present in Bernard, Redding, et al. 2007.

Equilibria with different endowments

Appendix A.2 shows that in autarky and in free trade equilibria, zero-profit productivity cut-offs are determined independently of endowments or factor prices. Provided that technologies and costs are the same across countries, this implies that zero-profit cut-offs and average productivities will be equal in the same industries across countries, both under autarky and free trade (Proposition A.1 in Appendix A.4). Under free trade, average firm size will also equalize by industry (Proposition A.2 in Appendix A.4). This will not hold under autarky since firm size is co-determined by factor prices, which equalize under free trade but differ under autarky. Moreover, firm mass will vary under free trade. Unlike in the "two-by-two" case, we cannot strictly determine comparative advantage in a multi-sector case. As is shown in Appendix A.4, however, in any country abundant in some factor, there must be an industry with a larger mass of firms than in the same industry of more scarce countries (for proof, see Proposition A.3 in Appendix A.4). Correspondingly, there will also be more entry and exit in such an industry (Proposition A.4 in Appendix A.4).

With their focus on trade costs, Bernard, Redding, et al. 2007 do not consider one of the important elements of the endowment-based Heckscher-Ohlin framework, namely the Rybczinsky theorem. With my focus on endowment shocks, however, it is an important component of comparative statics. As I demonstrate, not only does it continue to hold in the presence of heterogeneous firms, but also in a multi-sector, multi-factor case it is possible to formulate an analogue to a generalized Rybczinsky theorem, as presented in Feenstra 2016, p. 56 for a simpler model not featuring firms.

Theorem. *For an increase in endowment of each factor, there must exist at least one industry that will expand and another industry that will contract.*

Proof. See the proof in Appendix A.3. □

While this formulation provides no identification of precisely which industries will have a larger or lower output, we can imply that at least one industry, other than the industry of largest comparative disadvantage, will expand.

Opening up to free trade

A move from autarky to free trade will keep zero-profit cut-offs and average productivities unchanged. The reason is the same as in an equilibrium comparison: productivities are only determined by model parameters, which are constant. However, relative average firm size will change. A complication of a generalized multi-factor and multi-industry case is that comparative advantage sectors might be difficult to establish. There might be more than one abundant factor, and different abundant factors are used with different intensities across industries. In such cases, it might be unclear whether we can call an industry a comparative advantage industry in the sense of the HO model if it uses one abundant factor more intensively and another less intensively than the scarce factors. Instead, I propose an inverted formulation that approximates the idea of a multi-factor comparative advantage.

Proposition 1.1. *When an economy moves from autarky (a) to free trade (ft), relative average firm size will increase in industries that use the country's scarce factors least intensively:*

$$\frac{\tilde{r}_i^{ft}}{\tilde{r}_l^{ft}} > \frac{\tilde{r}_i^{a}}{\tilde{r}_l^{a}} \quad \text{if } \mu_{ik} < \mu_{lk} \text{ for } k \in \mathcal{J},$$

where $\mathcal{J}$ is a subset of factors that are scarce in the country.

Proof. See Proposition A.6 in Appendix A.4. □

I will define such industries as those of *approximate Heckscher-Ohlin comparative advantage*. The intuition behind this formulation is that under both autarky and free trade, the productivity levels will be the same, and the main determinant of relative firm size is therefore factor cost, which depends on the factor intensity of the industries. If we single out one abundant factor, moving from autarky to free trade will increase its price relative to other factors, and this all the more so, the scarcer the other factors are. Since factor cost is an intensity-weighted average of factor prices, the larger the share of scarce factors, the higher the increase in costs relative to other industries. Thus, the price too will become relatively higher and revenue will become relatively smaller. The smaller the share of scarce factors in the total factor cost (which is the case in industries of approximate Heckscher-Ohlin comparative advantage), the smaller the change in cost and price and the higher the relative revenue.

Opening up to costly trade

Unlike in the free trade case, opening up to costly trade brings about a change in productivity. All industries will become more productive (Proposition A.7 in Appendix A.5), and more so if the country is relatively small, if there is relatively high domestic competition in the respective industry, if the real exchange rate is unfavorable, or if trade costs are relatively low (for proof, see Proposition A.8 in Appendix A.5). In such countries, the share of exporters will also be larger (Proposition A.9 in Appendix A.5). This resembles gravity pull: a relatively small country will face greater foreign demand, the larger the trade partner is, and it will therefore have a higher share of exporters. In contrast, for a large country, foreign demand will be less than domestic demand, and the share of exporters will be lower. As a result, a comparison of countries of similar size will show that they are likely to have a similar average exporter share. If we compare a large country with a small one, however, the economy that is small enough may well have a larger share of exporters in both comparative advantage and comparative disadvantage industries.[7] In addition, competitiveness and trade costs influence exporting activity. All things being equal, the more competitive the domestic market is, the more attractive the foreign market becomes for the domestic companies, with more of them thus deciding to engage in the exporting activity. If, on the other hand, the foreign market is more competitive than the domestic one, only a few, most productive firms will be able to export. Finally, low trade costs allow firms to compete more effectively in the foreign market, and the probability of exporting therefore increases. More importantly, a move from autarky to costly trade will enhance the Ricardian comparative advantage along the lines of the approximate Heckscher-Ohlin comparative advantage:

Proposition 1.2. *The zero-profit productivity cut-offs and average productivities (both domestic and exporter) will increase by more in industries that use the country's scarce factors least intensively:*

$$\begin{aligned} &\Delta\frac{\varphi_{di}^{h}}{\varphi_{dl}^{h}} > \Delta\frac{\varphi_{di}^{f}}{\varphi_{dl}^{f}} \text{ and } \Delta\frac{\widetilde{\varphi}_{di}^{h}}{\widetilde{\varphi}_{dl}^{h}} > \Delta\frac{\widetilde{\varphi}_{di}^{f}}{\widetilde{\varphi}_{dl}^{f}}, \\ &\Delta\frac{\varphi_{xi}^{hf}}{\varphi_{xl}^{hf}} > \Delta\frac{\varphi_{xi}^{fh}}{\varphi_{xl}^{fh}} \text{ and } \Delta\frac{\widetilde{\varphi}_{xi}^{hf}}{\widetilde{\varphi}_{xl}^{hf}} > \Delta\frac{\widetilde{\varphi}_{xi}^{fh}}{\widetilde{\varphi}_{xl}^{fh}}, \end{aligned} \quad \text{if } \mu_{ik} < \mu_{lk} \text{ for } k \in \mathcal{I},$$

where $\mathcal{I}$ is a subset of factors that are scarce in the home country.

Proof. See Proposition A.10 and Proposition A.12 in Appendix A.5. □

The logic of the proposition comes from changing market competitiveness. Opening up to costly trade will render domestic markets more competitive (having lower price levels) in

[7] This might seem to contradict the canonical gravity model of Anderson and Wincoop 2003, which implies that countries of similar size will trade more with each other, and large enough size differences might lead to lower trade volumes. However, the idea here is rather to compare a given country with the size of the trading partner: the larger the partner, the more trade there will be. In addition, the share of exporters translates into trade volume through their size and absolute number, which are typically declining with the size of the economy.

industries of approximate Heckscher-Ohlin comparative advantage, thus leading to their higher productivity. Moreover, opening up to trade will raise the share of exporters by more in the industries of approximate Heckscher-Ohlin comparative advantage, which is again a result of tougher competition in these industries (Proposition A.11 in Appendix A.5).

Finally, in terms of welfare, the Stolper-Samuelson effect holds in the augmented version pointed out by Bernard, Redding, et al. 2007, namely, not only will the abundant factors benefit from increasing real rewards, but possibly also the scarce factors (or, at least, their losses will be dampened). The reason for this is the same as in Bernard, Redding, et al. 2007: if we recall that the real factor price is the ratio of the nominal price and the goods price index, a sufficient decrease in the price index would lead to higher real rewards for all factors. The reason for the decreasing price index is growing productivity in all industries when an economy opens up to costly trade (see Proposition A.7 in Appendix A.5). Opening up to costly trade also allows foreign firms to enter the home country's market, possibly leading to a larger total number of varieties available to consumers, in the style of Krugman 1980, and a further decrease in the price index.

Change in endowments under costly trade

While one of my goals was to show that the analysis of Bernard, Redding, et al. 2007 is valid beyond the empirically unrealistic "two-by-two" case, what is completely missing from their analysis is the comparative statics of an endowment change. Once the implications of opening up to costly trade are clear, it is straightforward to make conclusions about how an endowment shock, or a change to relative endowments of a country, will affect the distribution of productivities across industries. Consider an increase in an endowment of a factor that is abundant in the home country. Under costly trade, the relative price of this factor lies between the autarky ratio as in equation (A.8) and the free trade ratio as in equation (A.9) in Appendix A.1. With an increase in the endowment, both limits of this interval will shift downwards, and so we can safely conclude that the relative price of the affected factor will decline. Note that, unless the factor price ratio is the same as in autarky both before and after the shock, the decline will not be as large as the increase in the endowment: the response will be dampened. This would imply that total factor income and, thus, the total size of the economy, C^h, will increase. Therefore, the producers will face a larger (less competitive) market and generally might afford a lower productivity level (which is an inverse application of Proposition A.8). However, relative productivities will still change along the enhanced approximate Heckscher-Ohlin comparative advantage:

Proposition 1.3. *(a) If the endowment of an abundant factor increases in the home country, relative productivity cut-offs in the home country will increase in industries that use its scarce factors least intensively, as compared to the same industries in the foreign countries.*

$$\Delta\frac{\varphi^h_{di}}{\varphi^h_{dl}} > \Delta\frac{\varphi^f_{di}}{\varphi^f_{dl}} \text{ and } \Delta\frac{\varphi^{hf}_{xi}}{\varphi^{hf}_{xl}} > \Delta\frac{\varphi^{fh}_{xi}}{\varphi^{fh}_{xl}} \quad \text{if } \mu_{ik} < \mu_{lk} \quad \text{for } k \in \mathcal{J}$$

where $\mathcal{I}$ is a subset of factors that are scarce in the home country.

(b) Such a shock implies that relative average productivity in the home country will also increase in industries that use its scarce factors least intensively, as compared to the same industries in the foreign countries.

$$\Delta\frac{\widetilde{\varphi}_{di}^{h}}{\widetilde{\varphi}_{dl}^{h}} > \Delta\frac{\widetilde{\varphi}_{di}^{f}}{\widetilde{\varphi}_{dl}^{f}} \quad \textit{and} \quad \Delta\frac{\widetilde{\varphi}_{xi}^{hf}}{\widetilde{\varphi}_{xl}^{hf}} > \Delta\frac{\widetilde{\varphi}_{xi}^{fh}}{\widetilde{\varphi}_{xl}^{fh}} \qquad \textit{if } \mu_{ik} < \mu_{lk} \textit{ for } k \in \mathcal{I},$$

where $\mathcal{I}$ is a subset of factors that are scarce in the home country.

(c) Such a shock will also increase the probability of exporting in the home country's industries that use its scarce factors least intensively, relative to that in foreign countries.

$$\Delta\frac{\chi_{i}^{hf}}{\chi_{l}^{hf}} > \Delta\frac{\chi_{i}^{fh}}{\chi_{l}^{fh}} \qquad \textit{if } \mu_{ik} < \mu_{lk} \quad \textit{for } k \in \mathcal{I},$$

where $\mathcal{I}$ is a subset of factors that are scarce in the home country.

Proof. Once we have established, that the relative price of the abundant factor decreases with the increase in its endowment, (a) follows directly from Proposition A.10; (b) follows immediately from (a) due to the monotonic relationship between the zero-profit productivity cut-off and average productivity, as also is the case in Proposition A.12; (c) also follows immediately from (a), as in Proposition A.11. □

A decline in the endowment of the abundant factor will lead to exactly the opposite result, namely, the loss of both Heckscher-Ohlin and Ricardian comparative advantage.

This comparison of two steady states does not, however, answer two important questions: how fast does the economy move from one steady state to the other, and how smooth is this movement? In order to find possible paths of transition, I will make a few additional assumptions and find a numerical solution for the model.

1.3 Transition dynamics

1.3.1 Additional assumptions and adjustment mechanism

Since computational complexity increases significantly with each additional industry, factor, or country, I will from now on discuss the simplest combination possible and thus return to the two-good, two-factor, and two-country case. This leaves us with a system of about forty equations to solve for each steady state.

The first additional assumption necessary to find numerical solutions for the different steady states and the transition between them is that concerning the functional form of productivity distribution $G(\varphi)$. The usual approach is to assume Pareto distribution of firm productivity. This will be translated into Pareto distribution of firm size, which is widely supported by the empirical literature (see e.g. Gaffeo et al. 2003). Therefore, it

will be assumed that productivity φ is Pareto distributed with a minimal value $\varphi_{min} = 1$ and a shape parameter $k > \theta - 1$ such that $G(\varphi) = 1 - (1/\varphi)^k$.

The second necessary assumption is related to the speed of adjustment of some variables. I do not explicitly model the behavior of the factor markets but rather keep them in a clearing role. The functions of factor supply and demand are not written down, and factor prices are determined in a general equilibrium. This means, that in two different long-run equilibria, factor prices can differ and depend on the parameters of both home and foreign countries. To also keep the analysis of transition dynamics as simple as possible, it will be assumed that factor markets react immediately to the shock, and nominal factor prices are adjusted instantaneously to the new long-run equilibrium. This leads to a quick adjustment of productivity cut-offs, except when a temporary increase in competition occurs, a case that will be discussed later in my simulation. For computational simplicity, a restriction on changes in consumption is also imposed: it will be assumed that households need time to accommodate market changes, and in the period when the shock happens the amount of consumption C_{t0} remains unchanged, while the consumption expenditure R_{t0} adjusts to the change in price levels.

The adjustment of average productivity and firm mass will depend critically on the long-run effects of the shock. If the shock is positive, it results in increasing total income in the economy in the new long-run equilibrium. It will, therefore, imply higher demand and, thus, a lower level of competition, larger number of firms, and lower domestic productivity cut-offs in at least one of the industries. However, since firm entry is only possible through investment, and investment is limited by income and consumption, the number of firms cannot adjust immediately. Furthermore, because the expansion of an economy needs more investment than would be necessary in any of the two steady states, the initial impact of the shock is likely to be negative: resources will be shifted from current consumption to investment, thus compressing demand. This might force some firms out of the market. As firm productivity is revealed upon entry and does not change after that, average productivity will either increase with the death of some less productive firms, or, if all firms survive the initial impact, stay unchanged immediately after the shock. It will then fall gradually as new firms enter the market, while some incumbents exit due to the regular death shock and are replaced by the entrants, some of which have lower productivity. As the mass of firms converges to the new steady state, average productivity in the industry will decline towards the new equilibrium level, but at a slower pace. This also means that prices will converge to the new steady state much slower than income and number of firms. Recall that the profits generated by firms are accrued to households and can be used for consumption or investment. As the number of firms converges to the new steady state, additionally generated income, together with the gradual downward adjustment of investment to the long-run equilibrium, move consumption closer to the new equilibrium level as well. This mechanism will be discussed within a particular simulation example in section 1.3.3. A negative shock that leads to a decline in total income will induce tougher competition, higher productivity cut-offs, and fewer firms in at least one of the sectors. With the increased productivity cut-offs, some of the firms operating in the market when shock occurs become insufficiently productive to continue their operations and exit the market. Average productivity then rises immediately. New firms successfully entering the market are those whose productivity is also equal to or higher than the new cut-off, and so there is typically no additional adjustment. The

special case of upward adjustment will be discussed later.

The key role in transition dynamics is played by investment behavior of the households. Investment is necessary to finance firm entry into both sectors, yet the model only explicitly defines total investment as a difference between total income and consumption. Thus, the third assumption concerns the weights of the sectors in the investment portfolio in the case of deviations from the steady state. In the long-run equilibrium, the solution is simple: in each period, there is a constant number of firms exiting the market due to the death shock, and the role of investment is virtually to allow exactly as many firms to enter as are necessary to compensate for the death shock. Since the death shock hits the firms in both sectors with the same probability, the number of entering firms will be proportional to the mass of firms in each sector, and so the investment will also be split proportionally to the mass of firms in each sector, weighted by the respective firm values.

The situation changes, however, once the economy is pushed out of the steady state. The mass of firms immediately after the shock will deviate from the new long-run equilibrium. In general, three deviation cases are possible: either the number of firms in both sectors is too large, or it is too small, or there is excess of firms in one sector and a lack in the other. In all of these cases, I will assume that the mutual fund is managed by the social planner, whose aim is to return the economy to its optimal state as soon as possible.

Consider first the simplest case of extreme imbalance immediately after the shock, i.e. when the number of firms in one sector is too small, while in the other sector it is too large. Prompt restoration of the optimal relation between the mass of firms in the two sectors can balance the consumer basket and maximize household utility. The obvious solution is to concentrate the investment exclusively on the "deficit" industry and let the firms in the excess industry naturally "die out". This strategy will inevitably end in a steady state or lead to either the first or the second deviation case. While this case does not directly appear in my simulations, the effect of such a strategy is well illustrated by firm dynamics in the foreign country in Figure B.3 (Appendix B).

Suppose now that the mass of firms in both sectors is too large. In practical terms, this means that total income in the new equilibrium that can be used for consumption is lower than in the old one, and the mass of firms has to be adjusted to lower demand. However, since consumers value their current consumption, a decline in demand will be smoothened. So, on the one hand, a number of firms larger than equilibrium can still operate in the market after a shock, satisfying this transitory demand and paying additional dividends that help to prolong transition. On the other hand, virtually all available income is spent on consumption and no investment is happening. Therefore, the number of firms in both sectors will decrease steadily due to their "natural death". At some point, the economy will reach the state where demand is limited by prices, and not by income. Consumption expenditure will become less than total income in the economy, and households will start investing excess income and smoothen the process of firm death and, thus, their own consumption. There are a number of ways in which this investment can be split between the two sectors, and two will be considered here.

Firstly, one can simply use the steady-state weights. The equilibrium mass of firms is then weighted by firm value and compensates the two industries in the proportion of their steady-state loss. Since the total investment is smaller and the number of incumbents is

larger than in equilibrium, the number of firms entering the market will be too small to offset the firm death. Thus, there will be convergence. However, as I will show later, this approach can prolong transition to optimal firm mass.

The second possibility is to additionally weight the mass of firms by the inverse of the deviation from the steady state. As long as there are imbalances of production composition, such weighting will make sure that sectors with the least excessive mass of firms get the most investment. When the economy returns to the equilibrium balance of firms in both sectors (which does not necessarily mean that their number will reach equilibrium by that time), this weighting scheme will become identical to the first one.

Finally, the third deviation case is when there is too small a number of firms in both sectors. In this case, on the one hand, competition is low and, on the other, demand is limited by prices and, at the same time, investment is necessary to sustain future consumption, with the result that not all available income is spent on consumption, and additional investment takes place. I will consider the same approaches to determining sectoral weights in the investment portfolio as in the case of firm surplus. Firstly, one can apply the steady-state weights, which, again, will ensure convergence. Yet, a more effective approach is to additionally account for the scale of deviation from the new steady state. In the case of firm deficit, however, the weights will be proportional to the deviation from equilibrium such that sectors with the largest deficit get the most investment.

In my simulation, the focus will be on the deficit-weighted portfolio structure, but it will also be shown how dynamics change if the investment is split with steady-state weights.

1.3.2 Calibration

To simulate transition dynamics in the case of an exogenous shock, it is first necessary to determine long-run equilibria before and after the shock. To do this, I start with calibrating the exogenous parameters. As mentioned above, the two countries are assumed to be symmetric up to their endowments. It will also be assumed that exogenous parameters do not change over time, with the exception of the exogenous shocks being simulated. Thus, most exogenous parameters will not be indexed by countries, industries, or time periods.

The periods will be regarded as quarters. Following Ghironi and Melitz 2005 and Costantini and Melitz 2008, the discount factor will be assumed to equal approximately 4-5% per year, or β=0.99. I use the value of intertemporal elasticity of substitution from Ghironi and Melitz 2005 as the usual choice for business cycle models and set γ=2. The value for the elasticity of substitution across varieties is taken from Bernard, Eaton, et al. 2003 and is θ=3.8, which delivers the best fit for the US data. It can be argued that this value implies a very high mark-up over marginal cost of about 35%, but in the presence of fixed entry and production costs, marginal and average costs differ, and the mark-up over average costs remains at a plausible level.[8] In addition, such a mark-up is in line with the Rotemberg and Woodford 1995 estimate for the USA and the Y.-W. Cheung et al. 2001 estimates for the OECD countries. Bernard, Eaton, et al. 2003 also report the average

[8]See Ghironi and Melitz 2005 for discussion.

standard deviation of log US plant sales as 1.67. In the model, the standard deviation of log sales is equal to $1 / [k - (\theta - 1)]$, which together with the choice of θ implies the Pareto shape parameter k=3.4. Note that to ensure that the standard deviation of log firm sales is finite, it is required that $k > \theta - 1$, and the choice of parameters satisfies this condition. The share of expenditure spent on goods from industry 1 is normalized to $\alpha = 0.5$ for direct comparability across industries and countries. The nominal exchange rate is normalized to $\epsilon = 1$ without any loss of generality.

I follow Ghironi and Melitz 2005 to set the ad valorem trade cost to a value of $\tau = 1.3$, based on the overview by Obstfeld and Rogoff 2001. The estimations of a death shock vary. Ghironi and Melitz 2005 follow Bernard, Eaton, et al. 2003 who find that job destruction in the USA is about 10% annually. Bartelsman et al. 2005 suggest approximately 2-6% annual job destruction but 8-11% annual firm exit for the ten OECD countries. Therefore, the probability of a death shock $\delta = 0.025$, or around 2.5%, each quarter is used. The sunk cost of entry is normalized at $f_e = 10$. The relation between domestic and exporting fixed costs is set to match the size difference between exporters and non-exporters. Bernard and Jensen 1999 report a size differential of 70-90%, and for this reason I take the average and set $f_d = f_x / 1.8$. The value of export fixed cost is, in turn, set as a share of the per-period amortized flow value of the entry cost such that the average share of exporters roughly matches the percentage documented by Bernard, Eaton, et al. 2003. This yields $f_x = 0.4 f_e(1 - \beta[1 - \delta]) / \beta[1 - \delta]$, with the share being somewhat higher than in the one-industry, one-factor simulation of Ghironi and Melitz 2005. In the steady state, the probability of successful entry into domestic markets varies from 40% to 65%, which is slightly lower than success rates documented by Bartelsman et al. 2005. On average, 20.6% of all firms decide to export (compared to 21% reported by Bernard, Eaton, et al. 2003), and they are, on average, 60% more productive than firms only producing domestically, which is also in line with the simulation by Ghironi and Melitz 2005. As long as the relationships between the domestic, export, and entry fixed costs hold, their absolute values do not fundamentally affect the transition dynamics. Therefore, the initial normalization of the entry cost can be applied without any loss of generality.

The labor endowments of both countries are normalized to $L^h = 1000$ and $L^f = 1000$. The capital endowment of the capital-abundant (home) country is set to roughly match the capital-to-labor ratio provided by Paulson 2013 for the USA for the last decade, which yields $K^h = 2858$. On the assumption that both countries have more capital than labor, the capital endowment in the relatively labor-abundant (foreign) country is set at 70% of the level of the home country, which corresponds to the capital-to-labor ratio of Great Britain relative to the USA (Janiak and Wasmer 2013). Finally, the first industry is assumed to be capital-intensive, and the cost-share of capital is set to $\mu_1 = 0.6$, which approximately corresponds to agricultural or petroleum production in industrialized countries. For labor-intensive industry, $\mu_2 = 0.4$, which corresponds to textiles or electrical products[9].

Using these parameters, I numerically solve for steady-state values of the endogenous variables, which I will call the baseline case. The same is done for new equilibria after

[9] Based on the Input-Output tables of the OECD STAN Database.

shocks, and relevant model equations are log-linearized around the steady state to determine transition dynamics.

1.3.3 Responses to shocks

The analysis of endowment shocks instead of changes in trade costs is inspired by the fact that these are not such a rare phenomenon, but are rarely analyzed in the trade literature. Human-made destructions occur due to armed conflicts, and border changes can lead to substantial resource reallocations. Both are often related, with border change happening after an interstate or civil war, but there are also examples of rather peaceful changes. Striking war examples are the full redrawing of the European map after World War I and (with somewhat less extreme border modifications, but huge war losses and massive population movement) World War II. A very recent example of a *de facto* loss of endowments without any *de jure* border shifts is the civil war in Ukraine, whose larger share of industrial production was situated in the conflict zone. In contrast, examples of relatively peaceful border changes are the break-up of the Soviet Union and the reunification of Germany.

Obviously, there are also other sources of endowment changes. Natural disasters sometimes exert influence on country's resources, even though such effects can typically only be considered as local. They can, however, be especially relevant for small countries. Consider, for instance, tropical cyclones in the Philippines, where the 2013 Typhoon Haiyan struck shortly after an earthquake and was one of the most destructive disasters in decades, leaving almost 36 000 dead, missing, or injured (which is 0.04% of the total population and does not include those who were displaced) and causing economic loss of about 776 million USD (almost 3% of the country's GDP in 2013) from damage to infrastructure and agriculture (Takagi et al. 2017). More importantly, massive in- and outmigration and capital flights can result in considerable shifts in relative endowments. For example, the so-called European migration crisis may have far-reaching consequences for the economies of both the source and the host countries. However, such phenomena are also often connected to wars and natural disasters, while in other cases their exogeneity may be questionable.

Extending the focus beyond two production factors, changes in available mineral resources add to frequent and massive endowment shocks, for example the discovery of oil in the Gulf countries in the first half of the 20th century, the discovery of the Groningen gas field in the Netherlands, and recent discoveries of new shale gas fields in the USA.

Bearing these examples in mind, I will first concentrate on a case where the home country becomes less capital-abundant, modeling it as a massive capital destruction (or a respective decrease in capital-to-labor ratio) and assuming that all other exogenous parameters (including the rest of the endowments) are unchanged. Since capital is owned by households, it will be assumed that all households are hit by the shock uniformly and capital can be adjusted to the needs of other firms in the industry or even in other industries without substantial additional costs and time losses. Therefore, while some firms die after a shock, the capital they used can be employed immediately by the expanding firms, as non-operative capital implies opportunity costs for the households and they are willing to provide their capital to any other firm that is able to pay the

rent. Several shock intensities, with capital destruction ranging from 5% to 25% of the total endowment, were explored in 5%-steps. Since the results are quite similar, I will discuss the middle case of 15% capital reduction and present the results for the rest of cases in Appendix B.

Figure 1.1 presents the transition dynamics after such a shock for average and relative productivities, number of firms, consumption expenditure, and price indices in the home and the foreign country. The horizontal line at the zero level always represents the baseline steady state, while horizontal dashed lines represent the new steady state. We will commence by comparing the two equilibria.

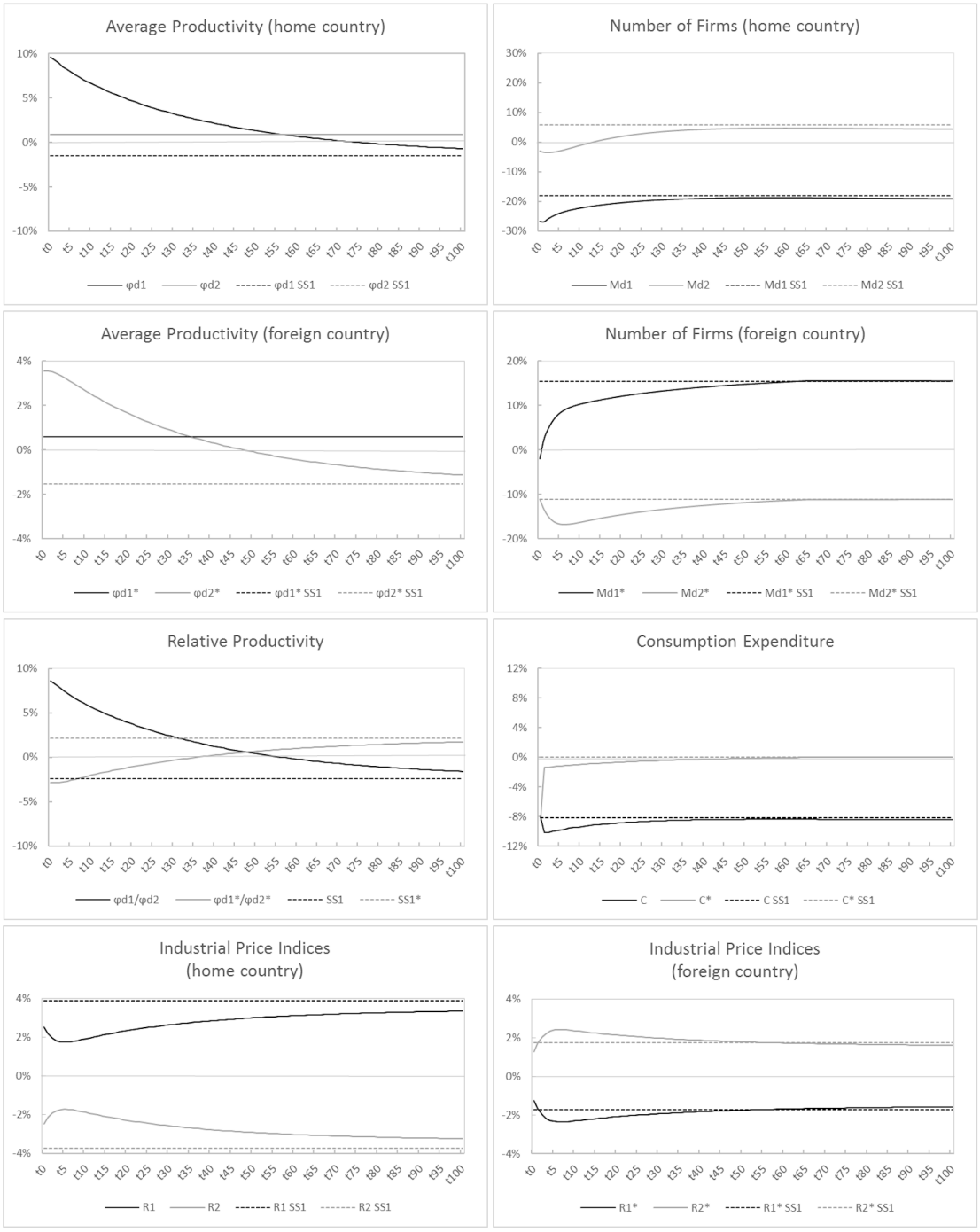
When a shock occurs, an adjustment of nominal factor prices to the new equilibrium leads to an increase in capital rent in the home country, which is coupled with a slight growth of capital rent in the foreign country as well due to a general fall in the world endowment of capital. Since endowments of labor do not change, nominal wages remain virtually unchanged. The effect on relative factor prices is much stronger in the home country, which suffers the loss. In my simulation, the ratio of capital rent to wage increases by 12% in the home country and only by 6% in the foreign country. This leads to capital becoming more expensive in the home country relative both to the foreign country and to the baseline case. As the first sector is capital-intensive, costs increase by more in this sector than in the labor-intensive sector, and by more in the home country than in the same sector in the foreign country. But also in the labor-intensive sector, larger growth in capital cost in the home country leads to a higher increase in costs than in the foreign country.

At the same time, there is a change in the total income of households, which consists of factor income and dividends. An increase in factor prices is dampened, as predicted by the model, and cannot fully offset the destruction of endowments in the home country, and so factor income will fall. In the foreign country, in contrast, a slight growth in factor prices is not coupled with any change in endowments, and so factor income will rise. Average profit will increase in the new equilibrium (and more so in the home country), but that will also result in growth of firm value and, therefore, a higher level of investment necessary to sustain the steady state from period to period. Thus, consumption expenditure in the home country (C) will contract, whereas the effect will be quite small for consumption expenditure in the foreign country (C^*).

Together with shifts in firms' costs, this will lead to the realignment of domestic productivity cut-offs in both countries. As predicted in section 1.2.5, the home country will lose its comparative advantage in the capital-intensive sector. These changes are reflected in average productivity levels and relative productivities, but also in the corresponding realignment of exporter shares across sectors. These changes also imply that, driven by the contraction in the first industry, the total number of firms in the home country diminishes, while that in the foreign country grows because of the expansion in the first industry. Finally, the overall (nominal) price index increases in both countries (more so in the home country), as does capital gain in real terms, whereas labor faces a lower real wage.

Many of these changes happen immediately after the shock. For example, a number of firms die immediately, and there is an immediate change in productivity cut-offs and average productivity. An interesting effect, however, is the overshooting of productivity

Figure 1.1: Transition dynamics after a negative endowment shock



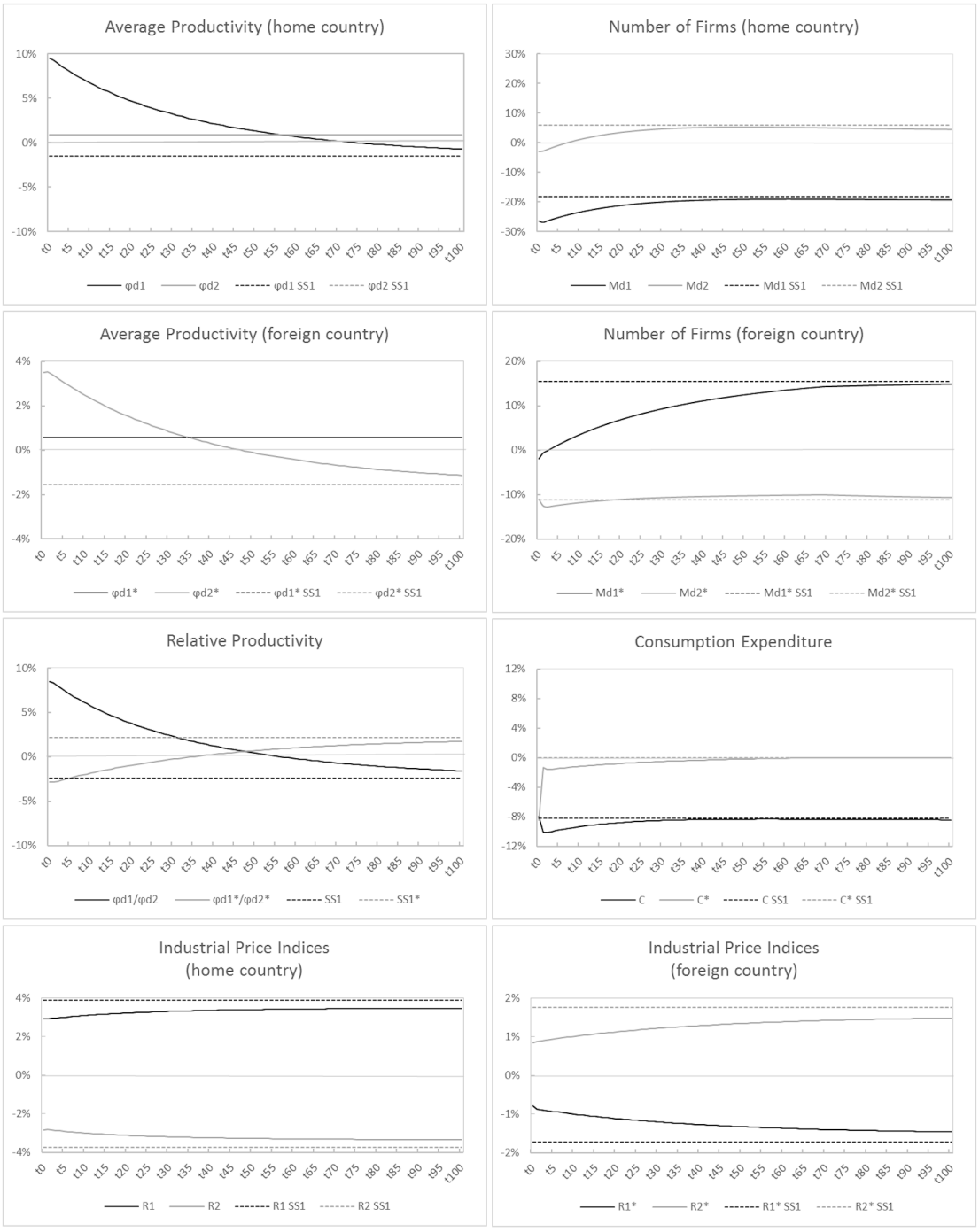
cut-offs in industries where new equilibrium cut-offs fall in period t_0, when the shock happens. With changing factor income and additional investments being necessary to sustain future consumption, world demand contracts, therefore creating additional competition in both markets and initially lowering commodity prices. Thus, the productivity cut-off first increases, causing firm death. This also raises average productivity in the industry. Then consumption expenditure will start increasing, driven by changes in factor income, the profits of the remaining and entering firms, and the adjustment of consumption. The productivity cut-off will then go down to the new equilibrium level. As new firms enter the market, some of which are less productive than the incumbents, the average productivity will steadily decline.

This shock, however, brings about a significant short-term increase in productivity and, thus, a decrease in the overall consumer price index. Since the real price index is, by definition, equal to unity, industry indices mirror each other in their dynamics. If one, however, considers nominal indices, they fall quite sharply and converge extremely slowly to the new equilibrium, despite the much faster convergence of real prices. As one would therefore expect, loss of capital leads to a recession and firm death in the affected country, especially in the first several periods after the shock. Another interesting implication is that the foreign country, which was not directly affected by the shock, also suffers firm death beyond a simple adjustment to the new equilibrium. Since the simulation was carried out with the deviation-based weights in the investment portfolio, the initial investment is mostly in the first industry in both countries, which is further away from the new steady state, in order to return as quickly as possible to the balance of industries in both economies. Thus, in the first few periods, the mass of firms in the second industry continues to decline, and then the trend is reversed and convergence to the new equilibrium starts. The time necessary to adjust to a shock, of course, depends heavily on the properties of the economies and the type of shock. Under the 15%-shock, in the foreign country affected only by contagion, consumption, firm mass, and real price indices adjust in about 15 years, while average productivity gets reasonably close to the new steady state in about 25 years. In the home country, however, which was directly hit by a significant negative shock, most of the catch-up process takes place over the same time, but then it slows down, and full adjustment would take substantially longer. Price indices generally adjust more slowly, as they are affected both by productivities and the number of firms in both countries.

The transition dynamics were shown in Figure 1.1 for deviation-based weights in the investment portfolio. A change in industry weights will mostly prolong convergence of the firm mass by a few years – and, therefore, also the adjustment of the consumer basket, as shown in Figure 1.2. It will, however, have a minor or no effect on other variables: the correction of average productivity occurs slowly in both cases, and a change in the weighting scheme cannot substantially affect it. Total consumption expenditure, driven by both factor income and incumbents' profits, will still converge at approximately the same pace, as the relative deficit of firms and profits in one sector are offset by the relative surplus of those in the other sector.

The transition dynamics for different shock levels, from 5% to 25%, are shown in Appendix B. Several results should be mentioned here. First, as expected, the new steady-state values deviate from the default equilibrium more, the more massive the shock is. Second, the initial impacts of the shocks are almost the same, which is obviously the result of

Figure 1.2: Transition dynamics after a negative endowment shock: steady-state portfolio weights



the assumption about consumption, which only starts to adjust in period t_1. What is interesting, however, is the changes in the dynamics after the initial impact. For productivity levels, more severe shocks imply that consumption shortages influence the economy longer and, thus, the market competition also remains tough longer than after smaller shocks. Therefore, while the transition path follows a parabolic curve after small shocks, it follows an s-shape curve after massive shocks. Investment is reallocated between the industries based on the deviation from the steady state, with a larger share of investment having to go into the expanding second industry in the home country after more massive shocks, and vice versa in the foreign country.

Interestingly, convergence to a new equilibrium firm mass happens faster in the home country as the shock gets more massive, but happens slower in the foreign country. While this result is counterintuitive at first glance, an explanation is actually quite natural. The home economy contracts after a negative shock, doing so all the more, the more severe the shock is. It takes much less investment to reach the new (overall lower) mass of firms, and also less investment to sustain the new equilibrium in the long run. The foreign economy, on the other hand, expands after the shock, and more so relative to the home economy, the more massive the shock is. Therefore, it will take more investment both to reach the new equilibrium and to sustain it later. Thus, not only is the duration of adjustment increased, but also the new equilibrium level of consumption is little affected by the shock intensity although it takes longer to adjust after more severe shocks. A final remark is on price indices, which by their definition are quite interdependent between countries. That is why in one of the countries they seemingly diverge from the equilibrium. The trend is reversed in the longer run as the prices in the other country converge closer to the new steady state.

Having these results for negative shocks in mind, it is tempting to examine whether the development will be exactly the opposite in the case of a positive shock. I therefore repeat the exercise for a 15% increase in capital endowment relative to the baseline case. Abstracting from the question where this additional capital can come from, I will simply use this stylized case and, as previously, assume that all households have 15% more capital from t_0 onwards that they are willing to offer to firms.

One might argue that such a scenario is unrealistic. I will still persevere with this experiment for two reasons. Firstly, if one considers labor and capital to be the factors of interest, it would indeed be more plausible to assume that the increase in an endowment is due to reallocation from one country to the other. For the simulation, however, this would imply a combination of a positive shock in the home country and a negative shock in the foreign country. As shown in the previous simulation, there is a contagion effect between the countries. Thus, any similarities in the transition dynamics to the capital destruction case might actually be driven by the contagion from the negative shock and not by the positive shock per se. Therefore, to keep the experiment clean of side effects, it makes sense to abstract from the reallocation scenario and focus on the less realistic pure positive shock case.

Secondly, labor and capital are not the only factors of production in the real world. If one considers mineral resources as a factor of interest, new discoveries of reserves are indeed a positive shock without any negative consequences for the endowments of other countries. Therefore, in the case of mineral resources, a pure positive shock is not only

an interesting scenario to examine per se (e.g. as a study of the "Dutch disease"), but also a very plausible one. A shock of 15% of endowment would actually represent quite a mild case here: consider oil discoveries in the US, which increased the US proved reserves by about 30% in 1970 and by 8-15% annually in 2008-2014.[10]

The transition dynamics for the positive shock (for deviation-based portfolio weights) are shown in Figure 1.3. As discussed above, the relation of the new steady state to the baseline case is almost exactly the opposite of that in the case of a negative shock. The transition is, however, similar to that of the negative shock in many respects.

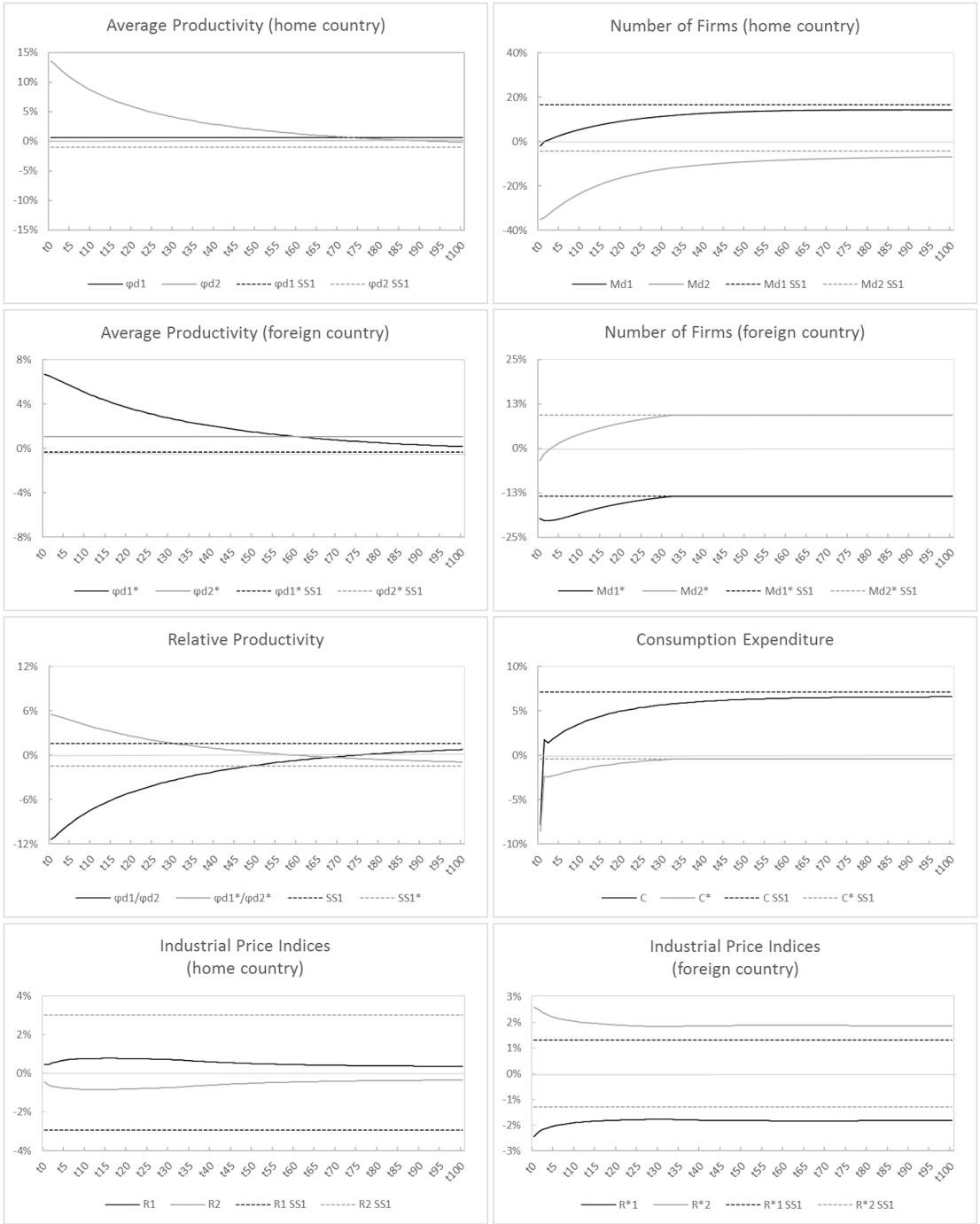
The first important result is that in the new equilibrium, not only the productivity in the first sector relative to the second increases in the home country (as predicted by theory), but, unexpectedly, the productivity cut-off in the first sector of the home country also rises slightly despite the fall in factor costs. A possible reason is the interplay of exporter entry in both countries. On the one hand, because of the expansion of the home economy (mostly driven by the capital-intensive industry), more firms can enter the foreign market, which leads to tougher competition in the foreign country and drives the productivity in the first industry upwards. Higher productivity in the foreign country, in turn, leads to lower prices and increases the attractiveness of the home country as the export market, which creates a boomerang effect. With both countries originally not being too different in size, the competition from the smaller country can be sufficient to influence the market of the home country.

The reason for productivity overshooting in the second industry is that, on the one hand, the consumption level remains unchanged in t_0 by construction and will expand from t_1 onwards. On the other hand, there is a large firm deficit in both industries relative to the new steady state, and massive investment (especially in the first industry) is necessary to sustain consumption in future. Therefore, with the new equilibrium productivity, while all the firms of the second sector would stay in the market, consumption expenditure after the shock is even lower than in the old equilibrium, and tough competition pushes prices downwards and productivity upwards.

The foreign country becomes relatively smaller after the shock, and its firms face more competition from the exporters. Furthermore, since factor prices are, to a small extent, also affected in the foreign country, not only do costs decline but factor income also decreases slightly and suppresses demand. This translates into higher productivity in the comparative advantage sector and only a minor negative response in the comparative disadvantage sector, with some firm deaths in t_0 and gradual adjustment afterwards, as investment allows for new entry. Thus, the number of firms in both industries and in both countries decreases in the short run, leading to short-term job destruction, even when the shock is positive. Interestingly, productivity in the capital-intensive sector of the foreign country barely falls. This is obviously related to the (minor) increase in productivity in this sector in the home country and industry expansion, leading to a considerable impact on the toughness of competition. In order to compete with exporters, domestic companies of the foreign country have to keep their productivity high. The same applies to productivity in the labor-intensive industry in the home country.

[10] See the statistics of the US Energy Information Administration at www.eia.gov.

Figure 1.3: Transition dynamics after a positive endowment shock



Another interesting result it that the transition period is much shorter in the foreign country in the case of a positive shock. It adjusts in just about 6-7 years (but still up to 25 years for average productivity), while the affected home country reaches the new equilibrium in about the same time as with the negative shock.

1.4 Conclusion

This chapter develops a generalized and dynamic model of trade and comparative advantage with heterogeneous firms which explores the short-term adjustment of industries to changes in comparative advantage. On the one hand, I demonstrate that the effects suggested by Bernard, Redding, et al. 2007 hold in a generalized form when more than two factors and two industries are analyzed. On the other hand, I show that while the long-term responses are in line with the comparative statics analysis based on the Heckscher-Ohlin model and its extensions, adjustment immediately after the shock is prone to overreaction and developments opposite to what is expected in the new steady state. The reason for this is the trade-off between consumption and investment that households face, leading to both restrictions on demand early after the shock and prolongation of firm entry due to insufficient investment. Interestingly, this can result in short-term job destruction in some or all industries even in the case of a positive shock, when the economy is expected to expand. The model calibration also shows that the adjustment to massive shocks can be a very long process. In the case of a positive shock, the economy affected by contagion can converge faster, but in the case of a massive negative shock, in particular, the transition may take 15-25 years. A promising challenge in this light is to employ real cases of endowment shocks to empirically estimate how fast the industrial structure and trade patterns change in response to those shocks. The cases of peaceful border changes or discoveries of mineral resources can be used as natural experiments well suited to testing the implications of the model presented here. However, the motivation for this research may also become a complication: with the transition to a new steady state sometimes taking decades, the endowment shocks might happen too frequently to be easily tested for.

Along with the implications, one might be interested in testing the additional restrictive assumptions of the calibration presented here. Does the consumption level adjust in the period of the shock? What are the real sectoral weights in the aggregate investment portfolio? Although these assumptions will not affect the pattern of short-term changes, they can influence the length of the period we call “short term”.

With respect to model limitations, I did not model the option of producing domestically and setting up an additional plant in the foreign country through foreign direct investment, instead of shipping the goods. Although adding FDI may have some interesting implications and the modeling approach would have been very similar (see for instance, Helpman, Melitz, et al. 2004), it would at the same time have substantially increased the complexity of the model. I also abstained from modeling multi-product firms. Explicit modeling of multi-product firms (as, for example, in Bernard, Redding, et al. 2011) would also add substantially to the complexity of the analysis. Therefore, I prefer to remain with the view of varieties as different product lines. Some brands may belong to the same mother company, but there are sunk costs to set up any new brand, and, due

to their inherent characteristics, they may also vary in production productivity. The multi-product firm is then likely to decide upon the production and export of each brand independently of the others[11].

Generally, the trade-off between consumption and investment, which prolongs the transition process, might serve as a basis for a debate on the possible scope of government intervention (e.g. targeted government investments). This could facilitate the adjustment and alleviate negative short-run consequences of a shock. Moreover, tools of fiscal and monetary policy that would keep demand and prices from falling after a shock might prevent some of the firms from leaving the market and, thus, reduce the impact of the shock, if applied promptly. However, it is beyond the scope of this chapter to analyze how such interventions might affect the economy as a whole, or how the government can finance them. Furthermore, since my analysis does not include any steady-state growth path, I do not make any predictions about how a surge in productivity might, for example, influence the technological component of long-run growth. If complementarity exists, it might not be in the long-run interest of a welfare-maximizing government to fully prevent firms' struggle for survival.

[11] This view was suggested by Ghironi and Melitz 2005.

Appendix

A Mathematical proofs

A.1 Steady state

Consumption path

Using the Euler equation for bonds from equation (1.24), the consumption path only depends on the relationship between the discount factor and the risk-free rate: e.g. $r_{t+1} > 1/\beta - 1$ means $C_{t+1} < C_t$. This, in turn, implies that consumption diminishes in time, and so will revenues and firm size. The trend is the opposite for $r_{t+1} < 1/\beta - 1$. Finally, with $r_{t+1} = 1/\beta - 1$, consumption is steady over time and so are revenues, wages and industry size. As there are no inherent drivers of growth in this model, I will consider only the final case for the steady state.

Firm dynamics

For the firm dynamics, steady consumption means that the number of firms in the industry is steady as well: $M_{dit} = M_{di,t+1} = M_{di}$, and the number of entrants should exactly cover the number of dying firms:

$$M_{di} = (1 - \beta) \left(M_{di} + \left[1 - G \left(\varphi_{di} \right) \right] M_{ei} \right).$$

We can then express the number of entrants:

$$M_{ei} = \frac{\delta M_{di}}{\left[1 - G \left(\varphi_{di} \right) \right] \left[1 - \delta \right]}. \tag{A.1}$$

Budget constraint

With (A.1) defining the number of entrants, with constant profits and firm values, and assuming financial autarky (no bonds), the budget constraint (1.23) can be rewritten as:

$$C = \sum_{j=1}^{F} w_j \overline{F}^j + \sum_{i=1}^{N} M_{di} \left[\tilde{\pi}_i - \frac{\delta}{\left[1 - G \left(\varphi_{di} \right) \right] \left[1 - \delta \right]} \tilde{v}_i \right]. \tag{A.2}$$

Firm value

Consequent substitution of future firm values in the Euler equation for shares (1.24) gives a DCF-based firm value:

$$\tilde{v}_i = \sum_{s=t+1}^{\infty} \left[\tilde{\pi}_i \left(1-\delta\right)^{s-t} \prod_{m=t+1}^{s} \frac{1}{1+r_m} \prod_{n=t}^{s-1} \left(1 - \frac{G\left(\varphi_{din}\right) M_{ein}}{M_{din} + M_{ein}}\right) \right].$$

With (A.1) and restriction on the risk-free rate to ensure constant consumption, this simplifies to:

$$\tilde{v}_i = \tilde{\pi}_i \frac{\left(1-\delta\right)\left(1-G\left(\varphi_{di}\right)\right)+\delta}{\left(1-\delta\right)\left(1-G\left(\varphi_{di}\right)\right)\left(1-\beta\right)+\delta}. \tag{A.3}$$

Firm value is not directly dependent on the industry size anymore, but it still depends on the productivity cut-off.

A.2 Factor market clearance and equilibria

Factor employment and factor prices

By definition, the difference between aggregate industry revenue and profit is the total cost, or factor costs in production (both domestic and export goods):

$$R_i - \Pi_i = \sum_{j=1}^{F} w_j \left(F_i^{jp} + F_i^{jx}\right), \ \forall i = 1, \ldots, N.$$

Aggregate industry profit can be split into two elements: coverage of the entry cost of new entrants and pure profit, which is paid out to households as dividends and supports a higher level of consumption. With balanced trade, the value of imports is equal to the value of exports, and consumption is equal to aggregate revenue, so the steady-state budget constraint can be rewritten as:

$$C = \sum_{j=1}^{F} w_j \overline{F}^j + \sum_{i=1}^{N} D_i, \tag{A.4}$$

where total country employment exactly equals the costs of entry and production by the full employment condition (1.26), and the amount of dividends coming from each industry and not used for investment is:

$$D_i = M_{di} \left[\tilde{\pi}_i - \frac{\delta}{\left[1-G\left(\varphi_{di}\right)\right]\left[1-\delta\right]} \tilde{v}_i \right] \tag{A.5}$$

On the industry level, a similar equation holds, describing how the revenue is spent:

$$R_i = M_{di}\tilde{r}_i = \sum_{j=1}^{F} w_j F_i^j + D_i \tag{A.6}$$

Since I assume a Cobb-Douglas production function, usage of each factor is a constant share of total cost:

$$F_i^j = \frac{\mu_{ij}\,(R_i - D_i)}{w_j}. \tag{A.7}$$

Cobb-Douglas preferences over industries imply that expenditure share of each industry is a constant share of aggregate expenditure (so industry revenue and dividends for consumption will be a constant share of their aggregates). Since factors are mobile across industries, factor prices equalize in the whole economy. The full employment condition can be rewritten as:

$$\overline{F}^j = \frac{(C-D)}{w_j}\sum_{i=1}^{N}\mu_{ij}\alpha_i.$$

Then, relative autarky factor prices are uniquely defined by relative endowments of the respective factors and relative weighted factor intensities of the economy:

$$\frac{w_j}{w_k} = \frac{\overline{F}^k}{\overline{F}^j}\frac{\sum_{i=1}^{N}\mu_{ij}\alpha_i}{\sum_{i=1}^{N}\mu_{ik}\alpha_i},\ \forall j,k \neq j. \tag{A.8}$$

Following the same argumentation line, in an FPE, or integrated, world, where factor prices are equal in all countries, relative prices will depend on the relative total world endowments (provided that the technologies and preferences are the same across countries):

$$\frac{w_j^{FPE}}{w_k^{FPE}} = \frac{\sum_{f=1}^{S}\overline{F}^{kf}}{\sum_{f=1}^{S}\overline{F}^{jf}}\frac{\sum_{i=1}^{N}\mu_{ij}\alpha_i}{\sum_{i=1}^{N}\mu_{ik}\alpha_i},\ \forall j,k \neq j. \tag{A.9}$$

From (A.7), usage of factor j relative to factor k in a single industry i is independent of the industry revenue or profits:

$$\frac{F_i^j}{F_i^k} = \frac{\mu_{ij}}{\mu_{ik}}\frac{w_k}{w_j},\ \forall j,k \neq j. \tag{A.10}$$

In the FPE set, the production mix of each country should feasibly employ all its production factors. Since factor usage is determined by (A.10), for any factor j, countries' relative endowment of this factor must lie between the relative factor intensities of the industries using it most and least intensively (weighted by j's relative price):

$$\max\left\{\frac{\mu_{ij}}{\mu_{ik}}\frac{w_k^{FPE}}{w_j^{FPE}}, i=1,\dots,N\right\} > \frac{\overline{F}^{jf}}{\overline{F}^{kf}} > \min\left\{\frac{\mu_{ij}}{\mu_{ik}}\frac{w_k^{FPE}}{w_j^{FPE}}, i=1,\dots,N\right\}, \quad \forall f=1,\ \dots,S, \forall j,k=1,\dots,F. \tag{A.11}$$

This set of inequalities characterizes the FPE set.

Again utilizing the full employment condition $\overline{F}^j = \sum_{i=1}^N F_i^j$, we can rewrite relative country endowment as a function of factor usage in all industries and share λ_i^k of each industry in total employment:

$$\frac{\overline{F}^j}{\overline{F}^k} = \sum_{i=1}^N \frac{F_i^j}{F_i^k}\lambda_i^k, \quad \lambda_i^k = \frac{F_i^k}{\overline{F}^k}, \quad \lambda_i^k \geq 0, \ \sum_{i=1}^N \lambda_i^k = 1. \tag{A.12}$$

Generally, the industry employment share must be nonnegative, but as we are more interested in internal solutions, we will require $\lambda_i^k > 0$. Combining (A.10) and (A.12), the set of following equations for each factor will jointly determine λ_i^k:

$$\sum_{i=1}^N \frac{\mu_{ij}}{\mu_{ik}}\lambda_i^k = \frac{\overline{F}^j}{\overline{F}^k}\frac{w_j}{w_k}, \ \forall j, k \neq j. \tag{A.13}$$

Taking into account restrictions on the sum of employment shares, this yields a system of $2(F-1)$ equations in $2(N-1)$ unknowns. Since I assume $N = F$, there will be a unique solution to the system, provided that the (block) matrix in the following system is non-singular:

$$
\begin{pmatrix}
\frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{N2}}{\mu_{N1}} & \dots & \frac{\mu_{N-1,2}}{\mu_{N-1,1}} - \frac{\mu_{N2}}{\mu_{N1}} & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\frac{\mu_{1F}}{\mu_{11}} - \frac{\mu_{NF}}{\mu_{N1}} & \dots & \frac{\mu_{N-1,F}}{\mu_{N-1,1}} - \frac{\mu_{NF}}{\mu_{N1}} & 0 & \dots & 0 & 0 & \dots & 0 \\
0 & \dots & 0 & \frac{\mu_{11}}{\mu_{1j}} - \frac{\mu_{N1}}{\mu_{Nj}} & \dots & \frac{\mu_{N-1,1}}{\mu_{N-1,j}} - \frac{\mu_{N1}}{\mu_{Nj}} & 0 & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & 0 & \begin{matrix} \frac{\mu_{1,j-1}}{\mu_{1,j}} - \frac{\mu_{N,j-1}}{\mu_{N,j}} \\ \frac{\mu_{1,j+1}}{\mu_{1,j}} - \frac{\mu_{N,j+1}}{\mu_{N,j}} \end{matrix} & \begin{matrix} \dots \\ \dots \end{matrix} & \begin{matrix} \frac{\mu_{N-1,j-1}}{\mu_{N-1,j}} - \frac{\mu_{N,j-1}}{\mu_{Nj}} \\ \frac{\mu_{N-1,j+1}}{\mu_{N-1,j}} - \frac{\mu_{N,j+1}}{\mu_{Nj}} \end{matrix} & 0 & \dots & 0 \\
0 & \dots & 0 & \frac{\mu_{1F}}{\mu_{1j}} - \frac{\mu_{NF}}{\mu_{Nj}} & \dots & \frac{\mu_{N-1,F}}{\mu_{N-1,j}} - \frac{\mu_{NF}}{\mu_{Nj}} & 0 & \dots & 0 \\
0 & \dots & 0 & 0 & \dots & 0 & \frac{\mu_{11}}{\mu_{1F}} - \frac{\mu_{N1}}{\mu_{NF}} & \dots & \frac{\mu_{N-1,1}}{\mu_{N-1,F}} - \frac{\mu_{N1}}{\mu_{NF}} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & 0 & 0 & \dots & 0 & \frac{\mu_{1,F-1}}{\mu_{1F}} - \frac{\mu_{N,F-1}}{\mu_{NF}} & \dots & \frac{\mu_{N-1,F-1}}{\mu_{N-1,F}} - \frac{\mu_{N,F-1}}{\mu_{NF}}
\end{pmatrix} \times
\begin{pmatrix} \lambda_1^1 \\ \vdots \\ \lambda_{N-1}^1 \\ \vdots \\ \lambda_1^F \\ \vdots \\ \lambda_{N-1}^F \end{pmatrix} =
\begin{pmatrix}
\frac{\overline{F}^2}{\overline{F}^1} \frac{w_2}{w_1} - \frac{\mu_{N2}}{\mu_{N1}} \\
\vdots \\
\frac{\overline{F}^F}{\overline{F}^1} \frac{w_F}{w_1} - \frac{\mu_{NF}}{\mu_{N1}} \\
\vdots \\
\frac{\overline{F}^1}{\overline{F}^F} \frac{w_1}{w_F} - \frac{\mu_{N1}}{\mu_{NF}} \\
\vdots \\
\frac{\overline{F}^{F-1}}{\overline{F}^F} \frac{w_{F-1}}{w_F} - \frac{\mu_{NF-1}}{\mu_{NF}}
\end{pmatrix}. \tag{A.14}
$$

The employment share of the last industry N is then just $\lambda_N^j = 1 - \sum_{i=1}^{N} \lambda_i^j$.

Existence of an integrated equilibrium

Let us first look at productivities in an integrated world. Since in such a world any firm can ship costlessly to any location in the world, country- and industry-specific export productivity cutoffs are left out of consideration. Skipping the time indices, the free entry condition (1.25), together with the steady-state firm value (A.3) suggests that:

$$\tilde{\pi}_i \frac{(1-\delta)(1-G(\varphi_{di}))+\delta}{(1-\delta)(1-G(\varphi_{di}))(1-\beta)+\delta} = f_{ei} \prod_{j=1}^{F} (w_j)^{\mu_{ij}}. \tag{A.15}$$

Average profit can be expressed as a linear function of average revenue, which is, in turn, a monotonic function of the zero-profit productivity cut-off, so that:

$$\tilde{\pi}_i = f_{di} \prod_{j=1}^{F} (w_j)^{\mu_{ij}} \left[\frac{1}{1-G(\varphi_{di})} \int_{\varphi_{di}}^{\infty} \left(\frac{\varphi}{\varphi_{di}} \right)^{\theta-1} g(\varphi)\, d\varphi - 1 \right]. \tag{A.16}$$

Substituting this into (A.15) yields an expression that is monotonically decreasing in φ_{di} and determines φ_{di} as a function of model parameters only (independent of factor prices, endowments, or revenues):

$$\frac{(1-\delta)(1-G(\varphi_{di}))+\delta}{(1-\delta)(1-G(\varphi_{di}))(1-\beta)+\delta} \frac{f_{di}}{[1-G(\varphi_{di})]} \int_{\varphi_{di}}^{\infty} \left[\left(\frac{\varphi}{\varphi_{di}} \right)^{\theta-1} - 1 \right] g(\varphi)\, d\varphi = f_{ei}. \tag{A.17}$$

As in (A.16), φ_{di} also uniquely determines average productivity $\widetilde{\varphi}_{di}(\varphi_{di})$. Factor prices w_j are determined from equation (A.8) by substituting world endowment into it. By taking one factor's price as a numeraire, we can uniquely define the prices of all other factors. Factor usage by industry is a solution to the system (A.14).

Goods' prices $\rho_i(\widetilde{\varphi}_{di})$ then follow immediately from (1.7). As mentioned above, revenue and profit are functions of productivity cut-off and factor prices (as in (A.16)). In addition, free entry implies that firm value in the steady-state equilibrium is equal to the entry cost. Average revenue, profit, and firm value, thus, follow directly if the productivity and factor prices are known. With revenue, profit, firm value, and factor payments by industry known, we can calculate the mass of firms M_{di} in each industry combining (A.5) and (A.6), and the steady-state mass of entrants as in (A.1). In the integrated world, industry price index (1.20) is simplified to $\mathcal{R}_i = \tilde{\rho}_{di} [M_{di}]^{1/(1-\theta)}$. The steady-state consumption is the sum of factor income and net dividends (profits minus investment). With that, all unknowns are solved for and equilibrium is fully characterized.

Existence of an FPE-equilibrium

If the countries' endowments lie in the FPE set characterized by (A.11), the world factor prices can be calculated from (A.9). Then industry employment of each factor in each country can be calculated from (A.14). With costless trade, all domestic firms also export to the foreign market, so (A.17) yields the single productivity cut-off for both domestic production and exports, and the share of exporters is 100%. With known productivities and factor prices, aggregate industry prices, average revenues, profits, and firm values follow from the pricing rule (1.7), free entry condition (A.15), slightly amended to incorporate cost-free exports, and firm value (A.3). Next, by combining (A.5) and (A.6), the mass of firms in each industry can be determined. Then, it is straightforward to calculate price indices and aggregate consumption for each country.

As in Bernard, Redding, et al. 2007, as long as the FPE-world matches the integrated world in terms of endowments and costs, the vectors of productivities, factor prices, and product prices will be the same in both worlds. Aggregate variables, summed over all countries, will also match those of the integrated world. However, allocation across countries will not be uniform as long as they differ in their endowments.

Existence of a costly trade equilibrium

Assume first that the vector of factor prices is known.

Industry employment can then be calculated from the system (A.14). Firm values are determined by the free entry condition (1.25). Prices, average firm revenues, and profits can be defined as functions of productivity cut-offs $(\varphi_{di}, \varphi_{xi})$, factor prices (w_j), price indices $(\mathcal{R}_i)$, and aggregate expenditure (C). This allows us to express firm mass (M_{di}) by combining (A.5) and (A.6). Aggregate sectoral revenue can be expressed from the definition of firm mass (1.19), and aggregate expenditure equals the sum of revenues across sectors if trade is balanced. Trade balance (1.21) is secured by adjustment of the real exchange rates $\left(\mathcal{E}_t^{hf}\right)$. In addition, domestic and exporting productivity cut-offs are bound by relation (1.15). Thus, we have $(3N + 1)$ equations for each country, corresponding to $(\varphi_{di}, \varphi_{xi}, \mathcal{R}_i)$ in N industries and aggregate expenditure, plus S equations to determine the same number of exchange rates.

Thus, all necessary variables are known, given the vector of factor prices. Factor prices are pinned down by market clearing condition (1.22), using the relation (A.6). The equilibrium is then fully characterized.

A.3 Analogue of the generalized Rybczinsky theorem

Theorem. *For an increase in endowment of each factor, there must exist at least one industry that will expand, and another industry that will contract.*

Proof. Consider country h abundant in factor j:

$$\frac{\overline{F}^{jh}}{\overline{F}^{kh}} > \frac{\overline{F}^{jf}}{\overline{F}^{kf}}, \ \forall f \neq h, \forall k \neq j.$$

Recall that relative factor employment in industry i is:

$$\frac{F_i^k}{F_i^j} = \frac{\mu_{ik}}{\mu_{ij}} \frac{w_j}{w_k},$$

Denoting this by F_i^{kj}, which is the same for all countries in the FPE equilibrium, let us sort the industries such that $F_1^{kj} > F_2^{kj} > ... > F_N^{kj}$. Using (A.12) and $\lambda_N^j = 1 - \sum_{i=1}^{N} \lambda_i^j$, we get:

$$\begin{aligned}
&\frac{\overline{F}^{kh}}{\overline{F}^{jh}} - \frac{\overline{F}^{kf}}{\overline{F}^{jf}} = \\
&\qquad \sum_{i=1}^{N-1} F_i^{kj} \lambda_i^{jh} + F_N^{kj} \left(1 - \sum_{i=1}^{N-1} \lambda_i^{jh}\right) - \sum_{i=1}^{N-1} F_i^{kj} \lambda_i^{jf} + F_N^{kj} \left(1 - \sum_{i=1}^{N-1} \lambda_i^{jf}\right) = \\
&\qquad \sum_{i=1}^{N-1} \left(F_i^{kj} - F_N^{kj}\right) \left(\lambda_i^{jh} - \lambda_i^{jf}\right) < 0, \qquad \forall f \neq h, \forall k \neq j,
\end{aligned}$$

where the first equation comes from subtracting (A.12) for the home country from that for the foreign country, the second equation comes by rearrangement of terms, and the inequality is implied by factor abundance.

Since the industries are ordered such that $F_i^{kj} > F_N^{kj}$, there must exist some industry i such that $\lambda_i^{jh} < \lambda_i^{jf}$. But, since for each country $\sum_{i=1}^{N} \lambda_i^j = 1$, there also must exist some industry l, such that $\lambda_l^{jh} > \lambda_l^{jf}$. Following the same strategy, it is straightforward to show that, with respective industry reordering, the same will hold for all other (non-abundant) factors:

$$\frac{\overline{F}^{jh}}{\overline{F}^{kh}} - \frac{\overline{F}^{jf}}{\overline{F}^{kf}} = \sum_{i=1}^{N-1} \left(F_i^{jk} - F_N^{jk}\right) \left(\lambda_i^{kh} - \lambda_i^{kf}\right) > 0, \ \forall f \neq h, \forall k \neq j. \tag{A.18}$$

While this does not directly mean that shares of all factors will be larger (or lower) in the same industry, such simultaneity is implied by profit-maximizing behavior of firms, whose production uses all factors in a complementary way. No profit-maximizing firm will hire more of a factor (and bear the respective costs), unless this allows it to increase output and revenue. With a Cobb-Douglas production function, it is only possible if employment of all other factors is increased according to their cost-shares. Thus, if country h has a relatively higher endowment of any factor than country f, there must exist an industry in country h that will have higher output than that in country f, and another industry whose output will be lower. Once treated as two different steady states in one country, instead of two countries, this is a generalized version of the Rybczinsky theorem, as in Feenstra 2016, p. 56. Note that while generally we cannot determine if these will be

comparative advantage industries whose output grows, inequality (A.18) implies that at least some industry other than the most disadvantageous one (with the lowest F_N^{jk}) has to grow. □

A.4 Free trade propositions

Proposition A.1. *ZPC cut-off and average productivities in the same industries across countries are equal both in autarky and under free trade.*

Proof. From (A.17), zero-profit productivity cut-offs in the free trade equilibrium are determined independently of factor prices or endowments. Thus, as long as entry and fixed production costs are equal in the same industries across countries, the zero-profit productivity cut-offs will also be the same. This will hold not only under free trade, but under autarky as well. Since average productivity is a monotonic function of the zero-profit productivity cut-off, as in equation (1.16), average productivities will also equalize for the same industries across countries. □

Proposition A.2. *Under free trade, firm size in the same industries is the same across countries.*

Proof. Given factor price equalization under free trade, average firm size as measured by revenue is uniquely defined by the productivity cut-off. Thus, the result immediately follows from Proposition A.1. □

Proposition A.3. *Under free trade, if country h is abundant in some factor, there must be an industry which it has a larger mass of firms. That is:*

$$\exists\, i: \qquad \frac{M_{di}^h}{M_{dl}^h} > \frac{M_{di}^f}{M_{dl}^f} \qquad \forall l \neq i. \tag{A.19}$$

Proof. From (A.5) and (A.6), firm mass in steady state can be expressed as

$$M_{di} = \frac{\sum_{j=1}^{F} w_j F_i^j}{\tilde{r}_i - \tilde{\pi}_i + \tilde{v}_i \frac{\delta}{(1-\delta)(1-G[\varphi_{di}])}}.$$

In the free trade equilibrium, with equal factor prices and industry productivities across countries, the denominator will be the same across countries. Thus, the relation of denominators across countries would be the same and any differences in relative mass of firms in the home and foreign country are defined by:

$$\frac{M_{di}^h}{M_{dl}^h} > \frac{M_{di}^f}{M_{dl}^f} \quad \Leftrightarrow \quad \frac{\sum_{j=1}^{F} w_j F_i^{jh}}{\sum_{j=1}^{F} w_j F_l^{jh}} > \frac{\sum_{j=1}^{F} w_j F_i^{jf}}{\sum_{j=1}^{F} w_j F_l^{jf}}$$

Separating factor j and dividing both numerator and denominator by the endowment of j, $\overline{F}^{jh}$, the relation can be expressed as follows:

$$\frac{\sum_{j=1}^{F} w_j F_i^{jh}}{\sum_{j=1}^{F} w_j F_l^{jh}} = \frac{\lambda_i^{jh} w_j + \sum_{k \neq j} w_k F_i^{kj} \lambda_i^{jh}}{\lambda_l^{jh} w_j + \sum_{k \neq j} w_k F_l^{kj} \lambda_l^{jh}} = \left(\frac{w_j + \sum_{k \neq j} w_k F_i^{kj}}{w_j + \sum_{k \neq j} w_k F_l^{kj}} \right) \frac{\lambda_i^{jh}}{\lambda_l^{jh}},$$

where $F_i^{kj} = \frac{F_i^k}{F_i^j} = \frac{\mu_{ik}}{\mu_{ij}} \frac{w_j}{w_k}$. A similar equation will hold for the foreign country. Note that F_i^{kj} only depends on factor intensities and factor prices, which are equal across countries. Thus,

$$\frac{M_{di}^h}{M_{dl}^h} > \frac{M_{di}^f}{M_{dl}^f} \quad \Leftrightarrow \quad \frac{\lambda_i^{jh}}{\lambda_l^{jh}} > \frac{\lambda_i^{jf}}{\lambda_l^{jf}}.$$

From the generalized Rybczinsky theorem, there must exist an industry such that $\lambda_i^{jh} > \lambda_i^{jf}$. This means that there must exist an industry such that $\lambda_i^{jh} - \lambda_l^{jh} > \lambda_i^{jf} - \lambda_l^{jf}$. Since industry employment is, by definition, non-negative and we are only interested in internal solutions, the latter inequality is the same as $\lambda_i^{jh}/\lambda_l^{jh} > \lambda_i^{jf}/\lambda_l^{jf}$. Thus, (A.19) must hold. □

Proposition A.4. *In the relatively larger industry from Proposition A.3, there is more entry and exit.*

Proof. From equation (A.1), the relative number of entrants in the steady state is:

$$\frac{M_{ei}^h}{M_{el}^h} = \frac{M_{di}^h}{M_{dl}^h} \frac{1 - G(\varphi_{dl})}{1 - G(\varphi_{di})}.$$

From Proposition A.1,

$$1 - G(\varphi_{dl}) = 1 - G(\varphi_{di}),$$

and so the first part of the proposition follows directly from Proposition A.3.

The number of firms exiting the market each period is a constant share of the mass of firms in the industry: δM_{di}. Thus, the second part of the proposition also follows immediately from Proposition A.3. □

Proposition A.5. *Zero-profit cut-off and average productivities remain unchanged with the move from autarky to free trade.*

Proof. Analogous to Proposition A.1, productivity is determined independently of endowments and factor prices by the parameters of the model. Thus, under both autarky and free trade, it will remain at the same level. □

Proposition A.6. *The move from autarky to free trade increases relative average firm size in industries that use the country's scarce factors least intensively.*

Proof. Consider first the average revenue of a firm. From the zero-profit condition (1.13) and definition of average productivity (1.16), we have

$$\tilde{r}_i = \left(\frac{\widetilde{\varphi}_i}{\varphi_{di}}\right)^{\theta-1} \theta f_{di} \prod_{j=1}^{F} w_j^{\mu_{ij}} = \left(\frac{\widetilde{\varphi}_i}{\varphi_{di}}\right)^{\theta-1} \theta f_{di} w_j \prod_{k\neq j} \left(\frac{w_k}{w_j}\right)^{\mu_{ik}},$$

where the last equality uses the property of the Cobb-Douglas production function that $\sum_{j=1}^{F} \mu_{ij} = 1$.

Considering revenue in industry i relative to some other arbitrary industry $l \neq i$, we get:

$$\frac{\tilde{r}_i}{\tilde{r}_l} = \left(\frac{\widetilde{\varphi}_i/\varphi_{di}}{\widetilde{\varphi}_l/\varphi_{dl}}\right)^{\theta-1} \frac{f_{di}}{f_{dl}} \prod_{k\neq j} \left(\frac{w_j}{w_k}\right)^{\mu_{lk}-\mu_{ik}}.$$

To see how relative revenue changes with the move from autarky to free trade, let us divide the ratio at free trade by that at autarky. Taking into account that productivities will not change, as per Proposition A.5, we get:

$$\frac{\tilde{r}_i^{ft}/\tilde{r}_l^{ft}}{\tilde{r}_i^{a}/\tilde{r}_l^{a}} = \prod_{k\neq j} \left(\frac{w_j^{ft}/w_k^{ft}}{w_j^{a}/w_k^{a}}\right)^{\mu_{lk}-\mu_{ik}}. \tag{A.20}$$

Thus, the change in relative revenue only depends on the change in factor prices and on the two industries' factor intensities. If (A.20) is larger than unity, relative revenue in industry i increases when moving from autarky to free trade. Combining (A.8) and (A.9) and rearranging the terms, it is straightforward to show that the change in relative factor prices only depends on how abundant the home country is:

$$\frac{w_j^{ft}/w_k^{ft}}{w_j^{a}/w_k^{a}} = \frac{1+\sum_{k\neq j} \overline{F}^{kf}/\overline{F}^{kh}}{1+\sum_{k\neq j} \overline{F}^{jf}/\overline{F}^{jh}}.$$

Note also that:

$$\frac{\overline{F}^{jh}}{\overline{F}^{kh}} > \frac{\overline{F}^{jf}}{\overline{F}^{kf}} \Leftrightarrow \frac{\overline{F}^{kf}}{\overline{F}^{kh}} > \frac{\overline{F}^{jf}}{\overline{F}^{jh}}$$

The home country's abundance in factor j implies that relative price of this factor will increase. The more abundant the home country is, the larger the change.

On the other hand, if industry i uses factor j intensively, this means $\mu_{ij} > \mu_{lj}$ and thus $\sum_{k \neq j} \mu_{lk} - \mu_{ik} > 0$. So at least for some factors other than j, the power $(\mu_{lk} - \mu_{ik})$ in equation (A.20) is positive. To ensure that (A.20) is larger than unity, the power should be positive for a sufficient number of factors where the increase in the relative price of factor j is the largest. In other words, most scarce factors (relative to whom the price of j increases most) must be used by industry i less intensively than by the industry in comparison. □

A.5 Costly trade propositions

Proposition A.7. *Moving from autarky to costly trade increases the ZPC cut-off and average productivity in all industries.*

Proof. If there are fixed and variable costs of trading, there will be two different cut-offs for domestic production and exporting. So equation (A.17) will have to incorporate the two types of profit and cut-offs as follows:

$$\frac{(1-\delta)\left(1-G(\varphi_{di}^{h})\right)+\delta}{(1-\delta)\left(1-G(\varphi_{di}^{h})\right)(1-\beta)+\delta}\frac{1}{\left[1-G(\varphi_{di}^{h})\right]}\times \left(f_{di}^{h}\int_{\varphi_{di}^{h}}^{\infty}\left[\left(\frac{\varphi}{\varphi_{di}^{h}}\right)^{\theta-1}-1\right]g(\varphi)d\varphi+ \sum_{f\neq h}f_{xi}^{hf}\int_{\varphi_{xi}^{hf}}^{\infty}\left[\left(\frac{\varphi}{\varphi_{xi}^{hf}}\right)^{\theta-1}-1\right]g(\varphi)d\varphi\right)=f_{ei}^{h}, \tag{A.21}$$

where

$$\varphi_{xi}^{hf}=\tau_{i}^{hf}\frac{\mathcal{R}_{i}^{h}}{\mathcal{E}^{hf}\mathcal{R}_{i}^{f}}\left(\frac{C^{h}}{\mathcal{E}^{hf}C^{f}}\frac{f_{xi}^{hf}}{f_{di}^{h}}\right)^{\frac{1}{\theta-1}}\varphi_{di}^{h}=\Lambda\varphi_{di}^{h} \tag{A.22}$$

is an increasing function of φ_{di}^{h}. Thus, in expression (A.21), f_{ei}^{h} is still a monotonically decreasing function of φ_{di}^{h}. Yet, while f_{ei}^{h} is a parameter that does not change when moving from autarky to trade, the left-hand side is increased by an additive positive term. Thus, to maintain the equality, φ_{di}^{h} has to increase. Note that this and further propositions assume that entry, production, and trade costs are symmetric across countries. □

Proposition A.8. *The domestic productivity cut-off and average productivity increase by more if the country is relatively small, if there is relatively high domestic competition, if the real exchange rate is unfavorable, or if trade costs are relatively low.*

Proof. From equation (A.22), the conditions of Proposition A.8 mean that Λ is smaller and export productivity cut-offs φ_{xi}^{hf} are closer to the domestic productivity cut-off φ_{di}^{h}. Then, the second summand on the left-hand side of (A.21) increases for such a country, and the domestic cut-off has to grow even more to maintain the equality. □

Proposition A.9. *The share of exporters is larger if the country is relatively small, if there is relatively high domestic competition, if the real exchange rate is unfavorable, or if trade costs are relatively low.*

Proof. By definition, the share of exporters is:

$$\chi_i^{hf} = \frac{1 - G\left(\varphi_{xi}^{hf}\right)}{1 - G\left(\varphi_{di}^{h}\right)} = \frac{1 - G\left(\Lambda\varphi_{di}^{h}\right)}{1 - G\left(\varphi_{di}^{h}\right)}.$$

Thus, the share of exporters is monotonically decreasing in Λ. The proposition then follows from the definition of Λ, as in Proposition A.8. □

Proposition A.10. *Moving from autarky to costly trade will raise the zero-profit productivity cut-off by more in industries that use the country's scarce factors least intensively.*

Proof. Let us start by comparing price levels under autarky and under free trade. Under free trade, all domestic firms export, factor prices equalize, and productivity cut-offs in same industries also equalize across countries. Thus, the industry price index (1.20) simplifies to:

$$\mathcal{R}_i = \left(\sum_{f=1}^{S} M_{di}^{f} \left[\tilde{\rho}_{di}^{f}\right]^{1-\theta}\right)^{\frac{1}{1-\theta}}.$$

For the same reason, relative price levels are the same across countries:

$$\frac{\mathcal{R}_i^h}{\mathcal{R}_l^h} = \frac{\mathcal{R}_i^f}{\mathcal{R}_l^f}, \quad \forall i, l = 1, \ldots N; \forall h, f = 1, \ldots S.$$

Under autarky, the industry price index is defined only by average variety price and firm mass in the respective country:

$$\mathcal{R}_{di} = \left(M_{di}\left[\tilde{\rho}_{di}\right]^{1-\theta}\right)^{\frac{1}{1-\theta}} = M_{di}^{\frac{1}{1-\theta}}\tilde{\rho}_{di}.$$

By definition,

$$M_{di} = \frac{R_i}{\tilde{r}_i}.$$

Since I assume Cobb-Douglas preferences over sectors, sector revenue is a fixed share of aggregate expenditure. Thus, relative price level in two sectors only depends on average variety prices and average firm revenues:

$$\frac{\mathcal{R}_i}{\mathcal{R}_l} = \left(\frac{\alpha_i}{\alpha_l}\frac{\tilde{r}_l}{\tilde{r}_i}\right)^{\frac{1}{1-\theta}} \frac{\tilde{\rho}_{di}}{\tilde{\rho}_{dl}}.$$

Substituting for revenues and prices, this equation can be rewritten as:

$$\frac{\mathcal{R}_i}{\mathcal{R}_l} = \left(\frac{\alpha_i}{\alpha_l}\frac{f_{dl}}{f_{di}}\right)^{\frac{1}{1-\theta}} \frac{\varphi_{dl}}{\varphi_d i} \left[\prod_{k\neq j}\left(\frac{w_k}{w_j}\right)^{\mu_{lk}-\mu_{ik}}\right]^{\frac{\theta}{1-\theta}}. \tag{A.23}$$

Provided that preferences, technologies, and costs are the same across countries, domestic productivity cut-offs are pinned down by (A.17) and are equal by sector across countries. Thus, any sector price differences across countries can only stem from differences in factor prices. In autarky, these are determined by country endowments:

$$\frac{\overline{F}^{jh}}{\overline{F}^{kh}} > \frac{\overline{F}^{jf}}{\overline{F}^{kf}} \Leftrightarrow \frac{w_k^h}{w_j^h} > \frac{w_k^f}{w_j^f}, \ \forall k \neq j$$

Analogous to Proposition A.6, equation (A.23) implies that if industry i uses country h's scarce factors less intensively than industry l, the following inequality holds in autarky:

$$\frac{\mathcal{R}_i^h}{\mathcal{R}_l^h} < \frac{\mathcal{R}_i^f}{\mathcal{R}_l^f}.$$

Finally, let us look at the price ratio under costly trade:

$$\frac{\mathcal{R}_i}{\mathcal{R}_l} = \left[\frac{M_{di}^h\left[\tilde{\rho}_{di}^h\right]^{1-\theta} + \sum_{f\neq h}\chi_i^{fh} M_{di}^f\left[\tilde{\rho}_{xi}^{fh}\right]^{1-\theta}}{M_{dl}^h\left[\tilde{\rho}_{dl}^h\right]^{1-\theta} + \sum_{f\neq h}\chi_l^{fh} M_{dl}^f\left[\tilde{\rho}_{xl}^{fh}\right]^{1-\theta}}\right]^{\frac{1}{1-\theta}}.$$

Consider two extreme cases:

1. In both industries, iceberg and fixed trade costs are extremely high: $\tau_i^{fh} \to \infty$, $\tau_l^{fh} \to \infty$, $f_{xi}^{fh} \to \infty$, $f_{xl}^{fh} \to \infty$. In this case, the share of exporters in each sector becomes negligibly low: $\chi_i^{fh} \to 0$ and $\chi_l^{fh} \to 0$. Prices converge to autarky levels.

2. In both industries, trade costs are negligibly low: $\tau_i^{fh} \to 0$, $\tau_l^{fh} \to 0$, $f_{xi}^{fh} \to 0$, $f_{xl}^{fh} \to 0$. Then everybody will export: $\chi_i^{fh} \to 1$ and $\chi_l^{fh} \to 1$. Prices converge to free trade levels.

For intermediate values of trade costs, on which the share of exporters depends monotonically, relative price levels will lie between autarky and free trade values. Thus, industries using country h's scarce factors less intensively will have relatively lower aggregate prices:

$$\frac{\mathcal{R}_i^h}{\mathcal{R}_l^h} < \frac{\mathcal{R}_i^f}{\mathcal{R}_l^f}, \quad \text{or} \quad \frac{\mathcal{R}_i^h}{\mathcal{R}_i^f} < \frac{\mathcal{R}_l^h}{\mathcal{R}_l^f}.$$

If we assume away inter-industry differences in costs, as in earlier propositions, this implies that for such industries:

$$\Lambda_i^h < \Lambda_l^h \text{ and } \Lambda_i^f > \Lambda_l^f.$$

From Proposition A.8, this in turn proves that the domestic productivity cut-off will increase by more in industries using scarce factors least intensively. □

Proposition A.11. *The probability of exporting increases by more in industries using scarce factors least intensively.*

Proof. The same as in Proposition A.9 the probability of exporting

$$\chi_i^{hf} = \frac{1 - G\left(\Lambda \varphi_{di}^h\right)}{1 - G\left(\varphi_{di}^h\right)}$$

decreases monotonically in Λ, and so the result follows immediately from the proof of Proposition A.10. □

Proposition A.12. *When moving from autarky to costly trade, Ricardian comparative advantage is magnified along the lines of the approximate Heckscher-Ohlin comparative advantage: if industry i uses country h's scarce factors least intensively, its average (domestic and exporter) productivity increases by more than in foreign countries:*

$$\Delta \frac{\widetilde{\varphi}_{di}^h}{\widetilde{\varphi}_{dl}^h} > \Delta \frac{\widetilde{\varphi}_{di}^f}{\widetilde{\varphi}_{dl}^f} \quad \text{and} \quad \Delta \frac{\widetilde{\varphi}_{xi}^{hf}}{\widetilde{\varphi}_{xl}^{hf}} > \Delta \frac{\widetilde{\varphi}_{xi}^{fh}}{\widetilde{\varphi}_{xl}^{fh}}.$$

Proof. Average domestic productivity is a monotonic function of the zero-profit productivity cut-off, and so the proposition for domestic productivity follows immediately from Proposition A.10.

Domestic and exporter productivity are related through Λ: $\varphi_{xi}^{hf} = \Lambda \varphi_{di}^h$. From Proposition A.10, both Λ and φ_{di}^h increase by more in the above-mentioned industries. Thus, the exporter productivity cut-off and, consequently, average exporter productivity will also increase by more. □

B Figures

Figure B.1: Average productivities at different shock intensities (negative shock to capital)

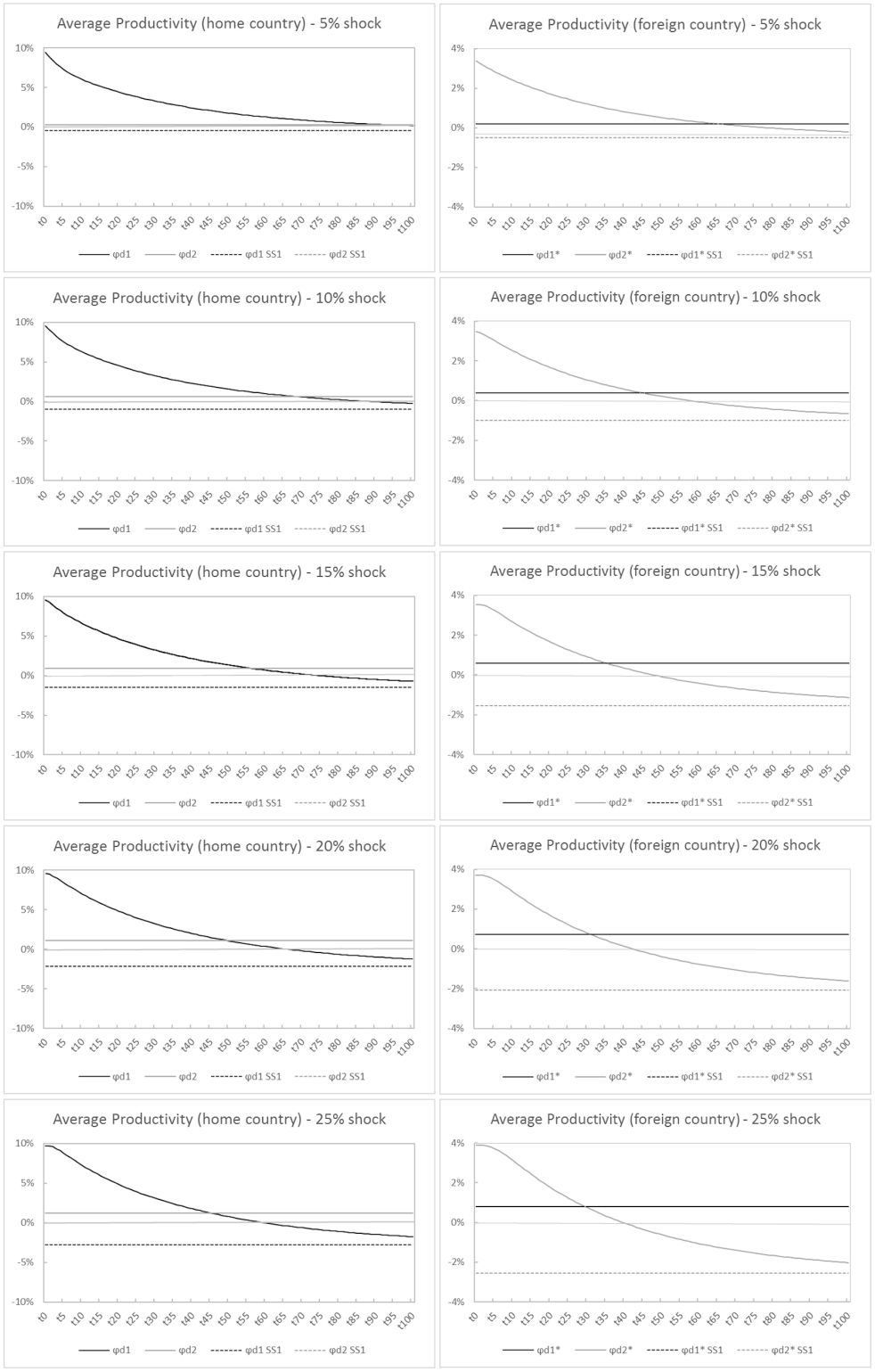


Figure B.2: Relative productivities at different shock intensities (negative shock to capital)

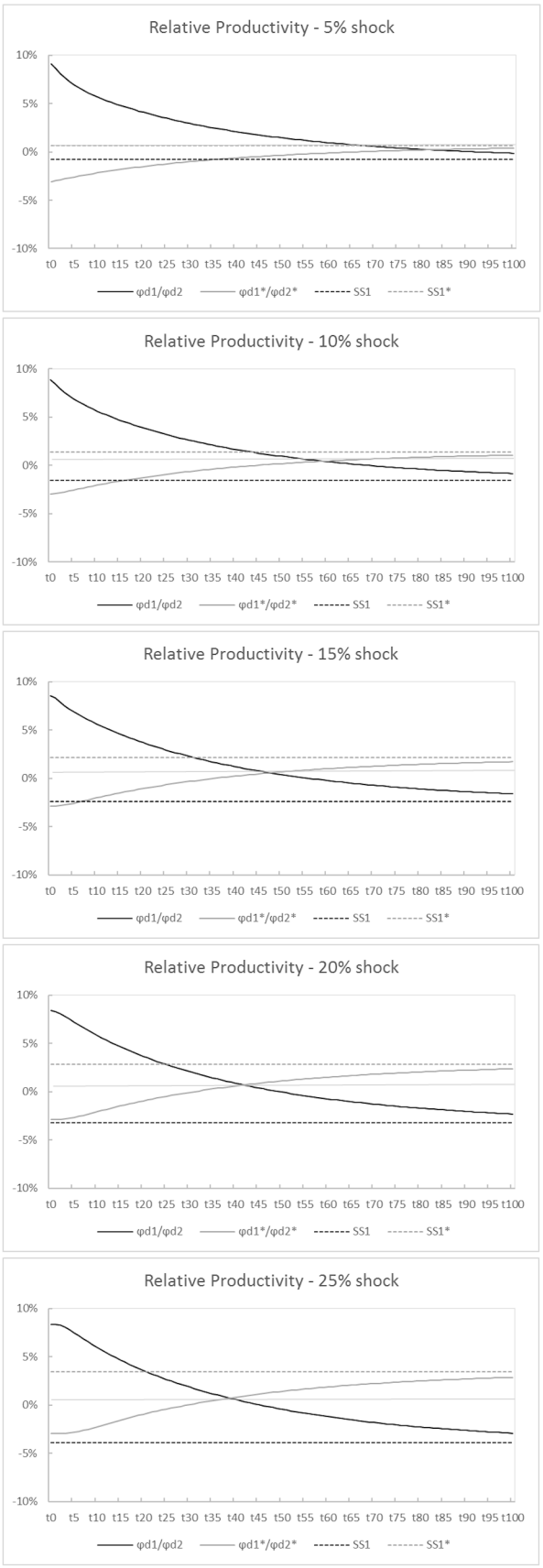


Figure B.3: Firm mass at different shock intensities (negative shock to capital)

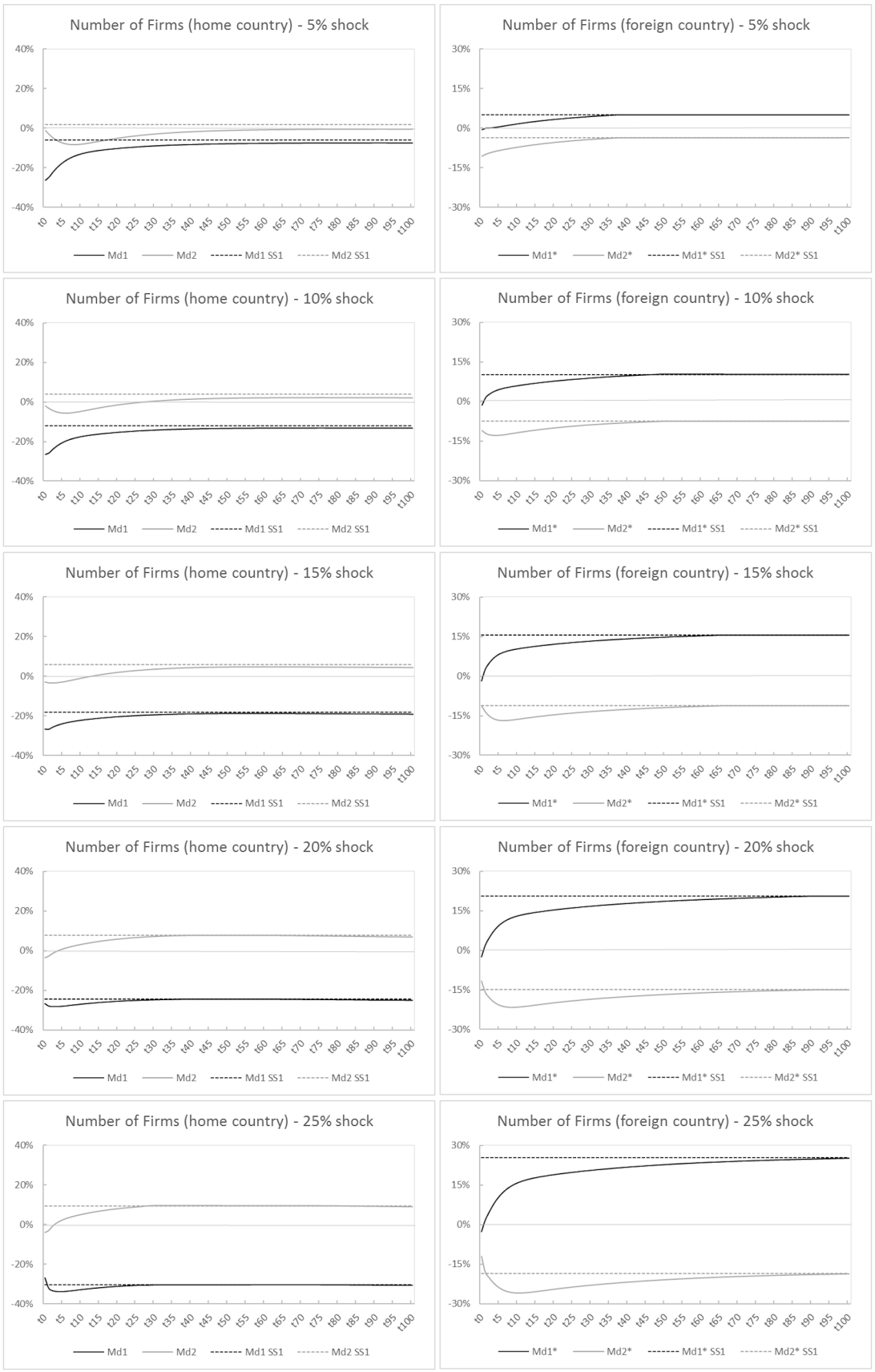
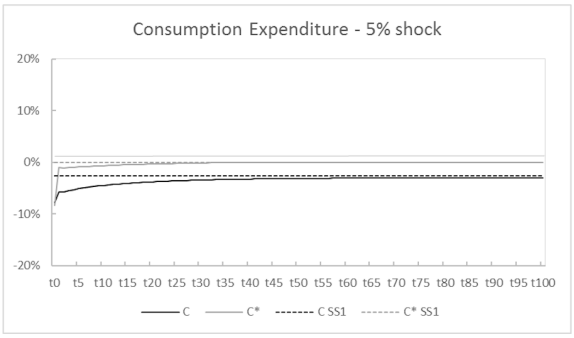
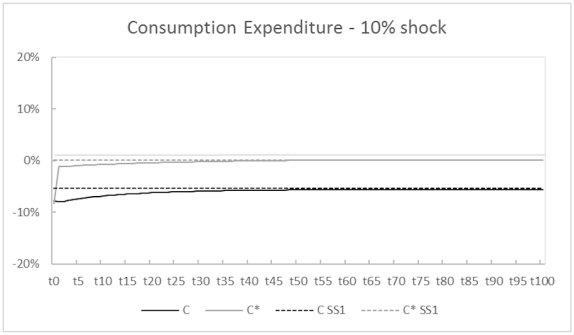
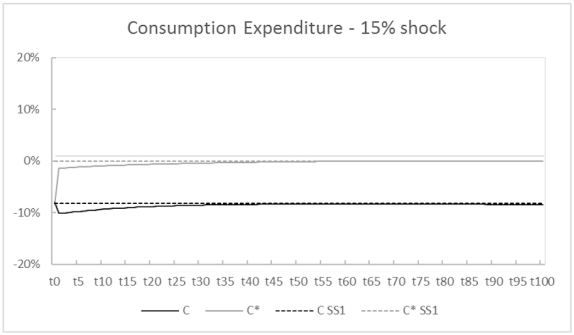
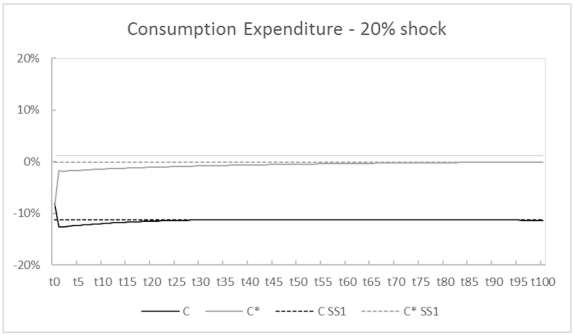


Figure B.4: Consumption expenditure at different shock intensities (negative shock to capital)









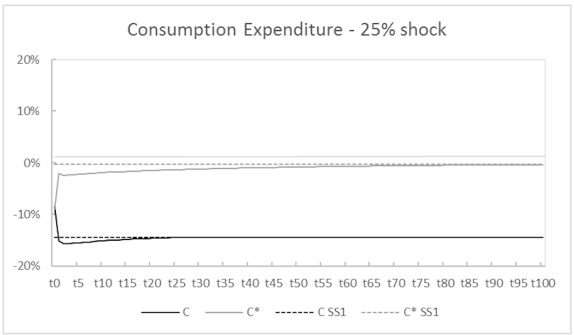
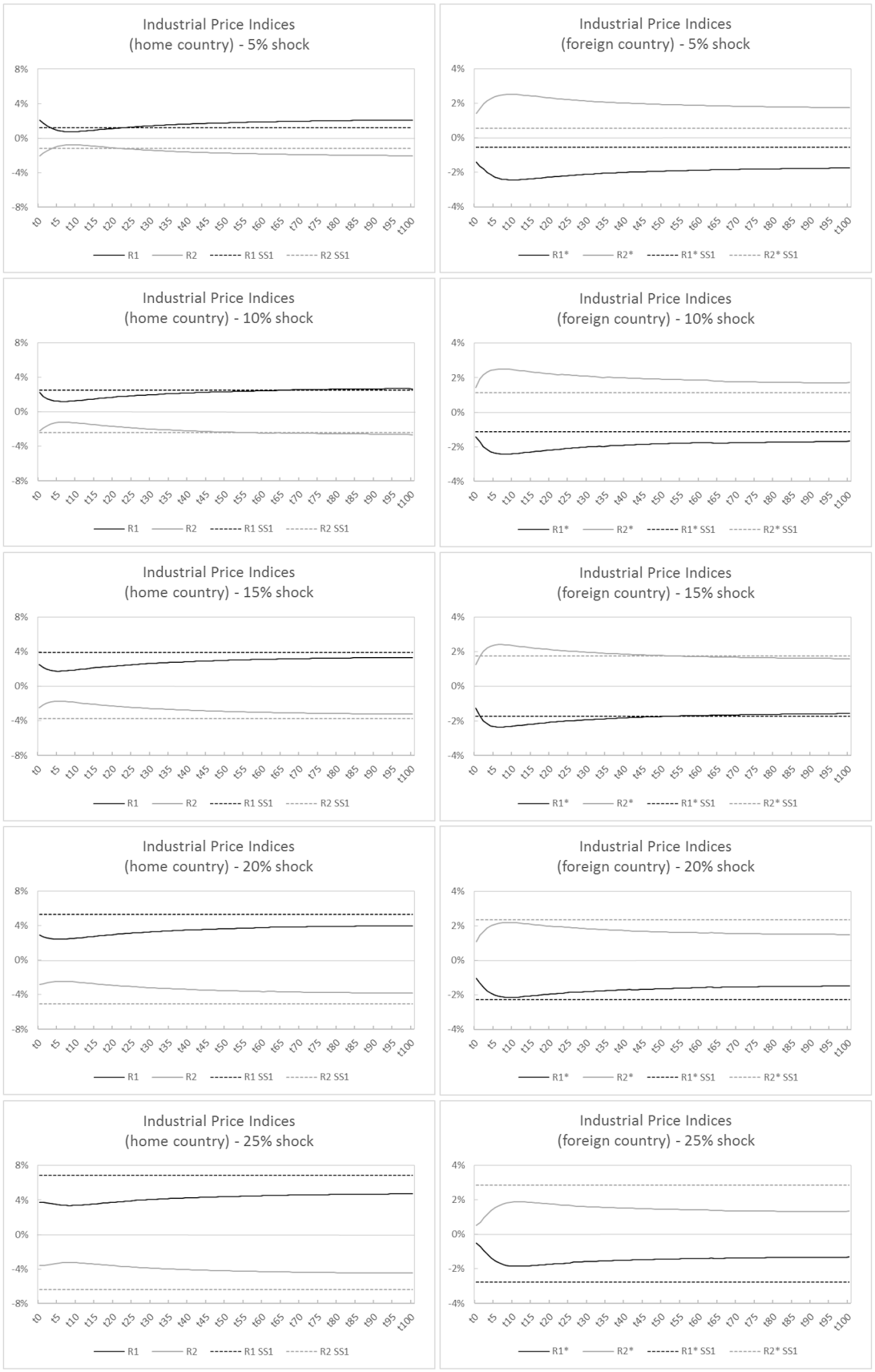


Figure B.5: Industry price indices at different shock intensities (negative shock to capital)



2 Is Buying on Amazon like Trading with a Digital Atlantis? E-commerce and Market Structure[1]

2.1 Introduction

Markets across the world are being reshaped by the online-based business models, such as that of the online retail giant Amazon. At the same time, small firms use online shops to sell their goods and services, and so compete for customers without having to maintain an expensive infrastructure. This leads to the central question of this chapter: How will the introduction of the e-commerce technology affect sectoral structures? And more specifically: Will e-commerce increase or decrease market concentration?

E-commerce is on a continuous growth path along several dimensions. In 2014, 1.3 billion people aged 14 and over were estimated to have bought something online, and this number is predicted to rise to 2.1 billion in 2021 (emarketer 2018). The volume of sales is growing even stronger. Only considering B2C transactions, the sales volume nearly doubled between 2015 and 2017 and reached $2.3 trillion in 2017 (ibid.), which is comparable to the GDP of economies like France or India (each $2.6 trillion in 2017, OECD 2019). Far from showing signs of saturation, the sales volume is predicted to again nearly double and reach $4.9 trillion in 2021 (emarketer 2018). Also the share of e-commmerce sales in total retail sales is predicted to rise from 7.4% in 2015 to 17.5% in 2021 (ibid.). In other words, e-commerce is growing faster that overall retail and eventually at the expense of the traditional retail channels. As B2B transactions tend to be more complex and discreet in nature, estimations vary considerably, some expecting it to reach $6.7 trillion in 2020 (Frost and Sullivan 2014).

At this scale of growth, traditional firms, wage earners, consumers and policy-makers are often overwhelmed by the impacts and the disruption caused by e-commerce. And, while there is much literature in business administration and managerial science about firm- and consumer-level impacts of e-commerce, economics research still lags behind in providing a solid theoretical basis for the macroeconomic effects we observe. Nor are the macroeconomic effects of e-commerce well-researched empirically.

This chapter works towards closing this gap. First, we formulate a theoretical model of e-commerce within the Melitz framework from the area of international trade theory. Second, we investigate empirically the implications of our model with a rich dataset on the countries of the European Union and the EU accession candidates.

[1] This chapter is a result of joint work with Lennart Jansen (Federation of German Industries). To honor his contribution, "we" will be used throughout this chapter.

As an intuitive basis for our theoretical approach, we first present several new stylized facts about e-commerce. We show that there is much heterogeneity across firms in the adoption of e-commerce, but also across sectors in consumers' preference to buy online. At the same time, while e-commerce has shown tremendous growth in the last decades, large online-based firms also increasingly engage in offline sales.

To account for these facts, we extend the well-established Melitz model by introducing products marketed via an e-commerce technology as a new variety within a sector, which comes with (potentially) different costs for producers. With this extension, we show how the emergence of online-markets has a similar effect on industrial structures as trade liberalization. Just as trade liberalization, e-commerce brings new opportunities, going hand in hand with increased competitive pressure on the established business structures. Similar to export opportunities, e-commerce allows the most productive firms to compete for more market shares by selling via additional channels. At the same time, like import competition, e-commerce tends to push less productive business models out of markets. Just as imports do not fully replace domestic production, e-commerce co-exists with traditional forms of distribution.

Contrary to the effects of trade liberalization, however, we also show that, if the cost of e-commerce becomes sufficiently low, small firms can enter the market by focusing on an e-commerce business model (or e-commerce "channel"). There are, in fact, several possible scenarios, depending on the relative costs of serving a market segment with the e-commerce technology, instead of the "traditional" technology. Each scenario suggests different types of firms profiting and losing from e-commerce. While trade liberalization tends to increase market concentration, the relationship between market concentration and e-commerce is non-linear. If in a sector e-commerce is relatively costly, only the large firms profit from this second channel, and so concentration in the sector grows. However, if the costs of e-commerce are very low, less productive enterprises will adopt e-commerce too. Eventually, even very small, low-productivity firms will be able to use e-commerce to enter markets, where they would otherwise be uncompetitive. This can in turn decrease market concentration.

Using European data between 2005 and 2017 to investigate the link between relative e-commerce costs and market concentration, we provide empirical support for the implications of our theoretical model. We show that, indeed, the relation between e-commerce costs and e-commerce adoption on the one hand and market concentration on the other tends to be hump-shaped. Especially the high-cost scenario is well supported by the data. In high-cost sectors, a slight decrease in e-commerce costs benefits more firms at the upper tail of the productivity distribution and leads to higher market concentration.

This chapter relates to a well-developed branch of literature on firm heterogeneity in open economies and a growing literature on the economic effects of e-commerce. The role of firm heterogeneity in trade was introduced in the seminal work of Melitz 2003. In the last two decades, the Melitz model was actively extended to incorporate a lot of features, such as interactions between the Heckscher-Ohlin and the Ricardian comparative advantage (Bernard, Redding, et al. 2007; Lechthaler and Mileva 2019; Polugodina 2019), short-term dynamics of trade liberalization (Chen et al. 2009; Costantini and Melitz 2008; Lechthaler and Mileva 2019), multi-product firms (Bernard, Redding, et al. 2011), foreign direct investment (Helpman, Melitz, et al. 2004) or interactions with the labor market (Helpman

and Redding 2010). Interestingly, Helpman and Redding 2010 show theoretically and Helpman, Itskhoki, et al. 2017 empirically, how wage inequality is non-linearly related to trade liberalization. A move from autarky to an open economy increases wage inequality, while at high openness levels further liberalization decreases it. With regard to market structure, however, the Melitz 2003 model implies that trade liberalization leads to higher market concentration, which has so far been supported by the empirical research (Baccini et al. 2017). The Melitz model is perfectly suited to incorporate e-commerce as well. We will show in section 2.3, how an elegant extension within the original model explains the firm-side stylized facts and delivers important insights into the economic effects of e-commerce. Similarly to Helpman and Redding 2010 and Helpman, Itskhoki, et al. 2017 for wage inequality and liberalization, we find that introduction of e-commerce will first increase and then lower inequality between firms.

The phenomenon of e-commerce and its influence on an open economy from a macro-perspective have been barely discussed in the literature so far. The closest contribution that both theoretically models and empirically investigates the effects of e-commerce is that by Freund and Weinhold 2004. The authors, however, focus on buyer-seller matching and a resulting reduction in fixed costs, empirically investigating the effect on trade volume. Our model is structured more broadly, allowing for interactions between different types of costs, considering the interactions between the traditional and e-commerce markets and offering insights into the effects of e-commerce on the market structure. Related to that is the model by Goldmanis et al. 2009, yet also this model focuses on buyer-seller matching, while empirically only investigating a few industries. Other literature focuses on the empirical investigation of trade volumes (Terzi 2011, Visser 2019) or size and distribution of benefits from e-commerce in terms of welfare or GDP (Anvari and Norouzi 2016; Couture et al. 2018; Dolfen et al. 2019). Two noteworthy empirical studies in the context of our research are Duch-Brown et al. 2017 and M. Falk and Hagsten 2015. Duch-Brown et al. 2017 show, for the sector of consumer electronics, that e-commerce and traditional channels co-exist, yet there is also substantial diversion from the traditional to the online sales. Moreover, the additional distribution channel of e-commerce increases competition. M. Falk and Hagsten 2015 show that e-commerce is positively related to labor productivity and that small firms gain more from e-commerce. Our model will provide a comprehensive theoretical basis for these findings.

The rest of the chapter is structured as follows. Section 2.2 provides several stylized facts about e-commerce, which serve as a basis for the intuition of our model. Section 2.3 lays out our theoretical model, while section 2.4 discusses its implications and formulates testable hypotheses. Section 2.5 describes our data and the empirical approach, and section 2.6 discusses the empirical findings. Finally, section 2.7 concludes.

2.2 Stylized facts about online markets

Taking a closer look at the definition of e-commerce and our focus in that respect, the EU categorizes "e-commerce" as *"the trading of goods or services over computer networks such as the internet"* (Eurostat 2019b). A similar definition is offered by OECD 2019. In contrast, business transactions not facilitated through the internet are often labeled "traditional" or "offline". Just as traditional markets, online markets can be sub-divided

along dimensions of product types and affiliated parties. Affiliated parties categories are business-to-consumer (B2C), business-to-business (B2B) and consumer-to-consumer (C2C). Product types are divided into sectoral classifications with standard goods and services categories.[2]

Three points are of importance here. First, the defining difference between e-commerce and the traditional business is the technology used to serve costumers. While there is a demand side aspect to it (not all consumers do consume via the internet), this means the strategic decisions involved are mostly supply side-driven. Second, for most product types, online and traditional markets co-exist, even though to different degrees, depending on sector specifics. Third, there are some specific business models closely related with e-commerce, most prominently the giant online platforms such as Amazon, Ebay and Alibaba for goods and Google, Facebook and Netflix for services. These platforms have their own economic dynamics, such as the "race for monopoly" and network effects, resulting in strategic interactions.[3] The investigation of this chapter will, however, focus on e-commerce as a broad phenomenon, with bigger and smaller players, and abstract from platform particularities. So the reference to Amazon as an illustrative example should be understood under the assumptions of monopolistic competition, i.e. Amazon being the biggest and most successful but not strategically dominant firm.

Before we come to the theoretical modeling of e-commerce, this section outlines a few stylized facts about e-commerce.

2.2.1 There is strong firm-level heterogeneity

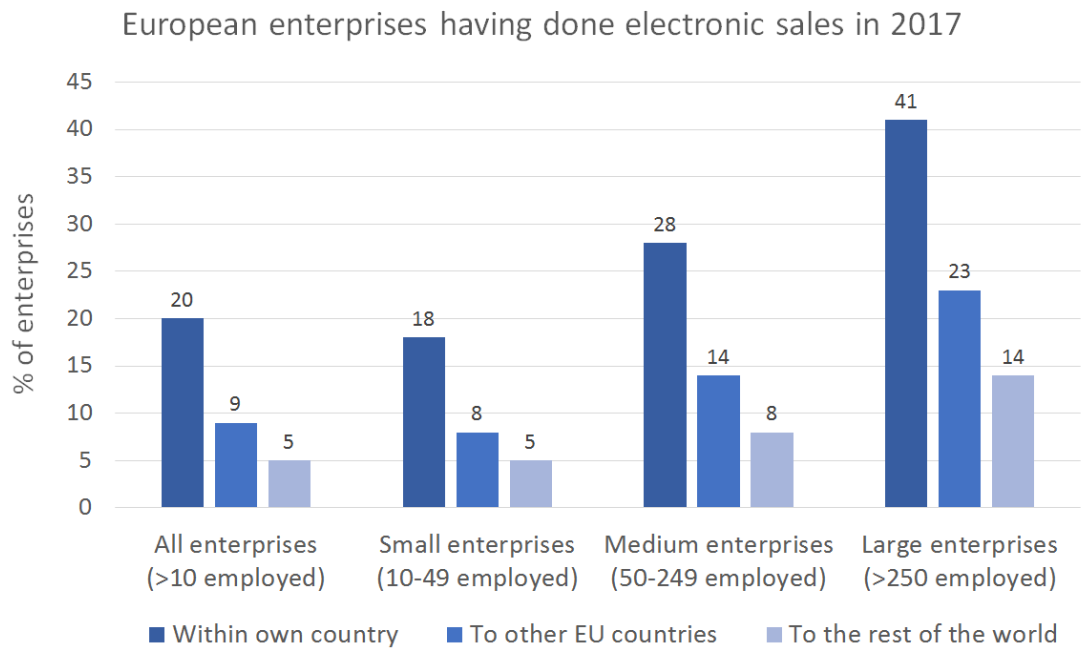
In the European Union, data on e-commerce are collected on a firm level. The data reveal two interesting aspects, as illustrated in Figure 2.1. First, bigger firms use e-commerce more often than smaller firms. In 2017, 41% of large firms but only 18% of small firms used e-commerce channels to sell to the domestic costumers. Second, only a fraction of firms serving e-commerce channels do so to sell in foreign markets. For all firm sizes, the share of firms with e-sales to the other EU countries is much lower than those with e-sales in home countries. And even less firms use e-commerce technology to sell to countries outside the EU. However, the pattern that bigger firms use e-commerce channels more often holds across all destinations.

There are strong similarities to patterns found in international economics, which are that a) only a fraction of firms use the opportunity to export and b) exporting firms tend to be bigger than non-exporters (Bernard and Jensen 1999). These patterns are strongly related to the development of the Melitz 2003 model, therefore it is promising to apply an extended version of this model for an investigation of similar patterns in e-commerce.

[2] Certain digital product types, e.g. video streaming services, are sometimes considered as e-commerce by nature, even if in some cases they are sold in an "offline" shop.

[3] On these issues, the German Federal Cartel Office has published a very comprehensive working paper, covering the scientific frontier on the relation of market power and (online) platforms (Bundeskartellamt 2016).

Figure 2.1: Firm level heterogeneity in e-commerce application



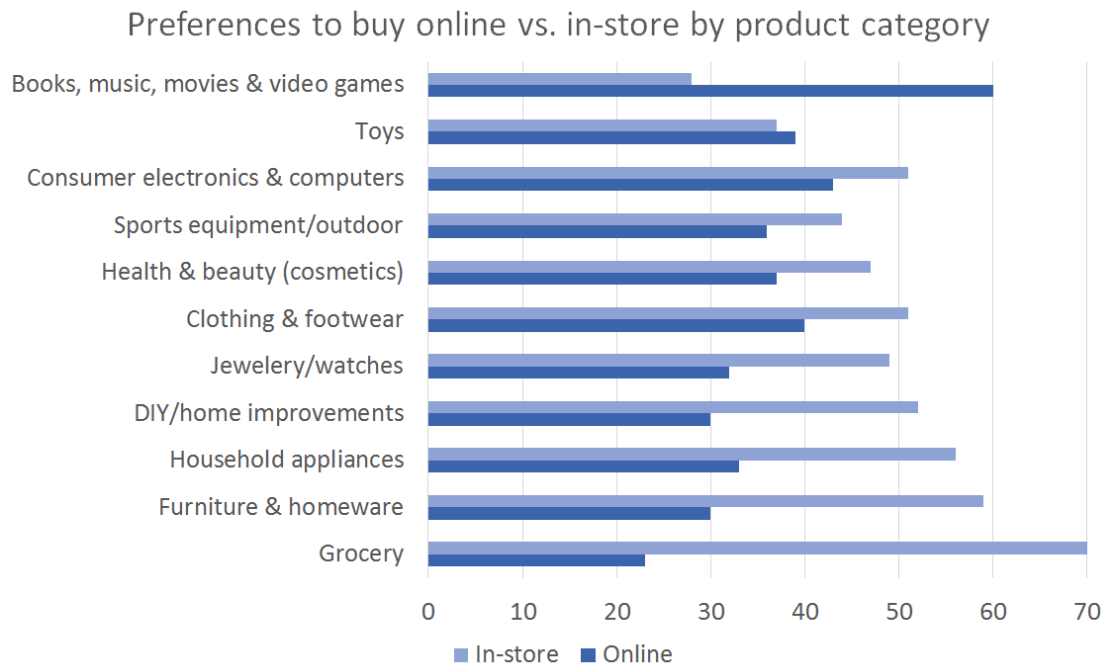
Data: Eurostat 2019a

2.2.2 Consumer preferences for e-commerce versus traditional consumption differ across sectors

Having stated that e-commerce business is defined by the technology applied by firms, it is vital to stress that there is a demand-side aspect to it. Consumers do have preferences about e-commerce consumption versus traditional consumption, and these vary across sectors. Figure 2.2 depicts data from a survey by PwC 2017, in which the respondents were asked the following question: "*Which method do you most prefer for buying your purchases in the following product categories?*". Consumers tend to choose the online channel for toys, books, music, movies and video games. On the other hand, they strongly prefer to buy groceries in stores.

A possible intuition behind this variation is that standardized goods are more conveniently ordered online, whereas for goods, whose quality cannot be examined audio-visually via a screen, people prefer to go to physical shops. As will be discussed later, there are factors other than product characteristics that drive the distribution of these preferences across markets, e.g. demography, infrastructure, or contract reliability. For our theoretical model, it is sufficient to recognize that there are some online versus traditional preferences. In our empirical investigation, we will return to the impact of these patterns.

Figure 2.2: Consumer preferences to buy online versus in-store



Data: PwC 2017

2.2.3 Multichannel-marketing: big e-commerce players heavily invest in traditional infrastructure

While e-commerce is growing at the expense of the traditional retail, the phenomenon that giant e-commerce firms invest in traditional shops is making headlines. Most prominently, in 2017 Amazon bought Whole Foods, a premium groceries chain, for $13.4 billion and in 2018 also announced plans to further invest billions into inner-city brick and mortar shops under the label AmazonGo (Levy 2017). In this context, the New York Times quoted John E. Lopatka, a professor for anti-trust law, as follows: "One question would be, does an online seller of groceries compete with a brick-and-mortar grocery store, and I think the answer is 'yes, at some level, but that overlap is probably not terribly great'" (Wingfield and Merced 2018).

This relation of traditional retail and e-commerce is discussed under the business administration term "multi-channel marketing" (see e.g. Weinberg et al. 2007). Originally coined as multiple channels for advertisement, it now specifically aims at the complementarity of online and offline channels to serve costumers. Businesses are aware that consumers value choice and flexibility, both individually and on a macro-level. A business can, thus, become more attractive for (additional) costumers, if it invests in multiple channels to serve them. From the economic theory perspective, this translates into the online and offline channels being imperfect substitutes, with potential cost-side complementarities. This, in turn, speaks in favor of an investigation under the assumption of *monopolistic competition*.

2.3 E-commerce in the Melitz framework: "Twin varieties"

Having laid out the intuition in section 2.2, this section will discuss how e-commerce can be incorporated into the Melitz framework. The central technical novelty towards Melitz 2003 is an additional e-commerce "channel" through which firms can optionally serve markets. The approach will be labeled as the "Twin Varieties Option" (TVO). It will technically introduce e-commerce as a co-emergence of a new variety that firms can offer at different costs and a slight adaption in consumer preferences to this novelty.

The methodological path is to deduct a new key set of sectoral equilibrium equations, which will allow us to derive e-commerce effects on the sectoral key parameters (most important: market entry productivity cut-offs φ and demand per variety B). Based on these, the effects of e-commerce can then be traced within the already well-researched Melitz framework, which, in its basic version, we can also use as a benchmark. We will methodologically trace the effects of e-commerce as ceteris paribus differences between the benchmark Melitz framework's costly trade equilibrium and the equilibrium of our e-commerce TVO-extension. The superscript ec will denote the parameter aligned with the e-commerce (ec-)version of any product. Parameters with a tr or without a superscript will refer to the respective traditional (tr-)version. This is because the latter is per definition equivalent to the benchmark Melitz model. If not explicitly noted otherwise, all equations, variable definitions and connections between variables and parameters in this extension are exactly as in the consolidated version of the Melitz model laid out in Melitz and Redding 2014a.

2.3.1 The dual nature of e-commerce

How does e-commerce differ from traditional forms of supplying goods and services to markets? The surprising answer is: not very much. From an economic perspective, all agents on e-commerce markets behave along the same principles as they do on traditional markets. Firms invest, hire and produce if they expect to receive a (marginal) profit. Consumers spend their limited income on a basket of goods that best suits their preferences, with some internal trade-offs between quantity, quality and variety.

One aspect where e-commerce seems to differ is efficiency, as digitization allows firms to do certain aspects of value-creation at lower costs. It is, for example, often cheaper to distribute goods via one online-shop with one warehouse, compared to the vast network of physical shops necessary to reach a similar amount of costumers. So, is e-commerce just a new, potentially less costly, technology to serve customers? It is tempting to say: Yes, because the vast majority of e-commerce business consists of goods that are also available in traditional shops. On a library shelf, a volume of Harry Potter bought online is indistinguishable from the same book bought in a physical book shop. They are only provided to the market with different infrastructure and logistics. The questions to answer are, however: If one technology is simply less costly than the other, why would e-commerce differ from any ordinary increase of productivity parameters? Why would any firm not simply use the comparatively cheaper technology to serve its market? Why would a firm with a cheaper technology to supply the identical product not capture the whole market? Why, if Amazon has a superior technology, do book shops still exist?

The answer is: some people like book shops. They are willing to pay a premium to buy the exact same book there rather than ordering it on Amazon. Thus, as people's preferences to buy online or traditionally differ both across sectors and consumers themselves – as the stylized fact 2.2.2 suggests – the identical product sold online or offline will be perceived as somewhat different varieties by the representative consumer. And even if e-commerce is the cheaper technology and offers more competitive prices, firms will eventually still have an incentive to invest in dual channels and infrastructures. As in the case of export decisions, they will do so if they can expect sufficient *additional* revenue to make additional investments pay off.

One notable conclusion from this argumentation is that, at this point, we can abstain from the restrictive assumption that e-commerce is a cheaper technology to serve any particular market. It is sufficient to state that establishing an e-commerce channel to supply a certain good would imply potentially *different* fixed and marginal costs than establishing a traditional channel to supply the (ex-post) identical good, both domestically and in foreign markets. If there is a strong preference among customers to buy either online or via traditional channels, we should expect a market structure with some sales share being facilitated as e-commerce and the other share via traditional means. These thoughts are the intuitive core of our approach to modeling e-commerce.

2.3.2 E-commerce properties in the Melitz framework

In order to investigate the effects of e-commerce, we need to introduce e-commerce market segments, consisting of a demand side and a supply side. Note that these will be additional to the traditional market segments, which are, by assumption, those modeled in Melitz and Redding 2014a.[4]

Supply side

One assumption already mentioned above is that the Harry Potter book sold online is indeed "ex-post identical" to the one in the bookshop. This is rather intuitive with any book, which is written and published by the same people, no matter how it is distributed afterwards. Now, we might be tempted to ask where the traditional and e-commerce varieties start to deviate from each other. The Melitz framework offers a very hands-on approach to this demarcation. When a firm decides whether to sell its product via e-commerce and/or traditional channels, it already knows the product. In other words, the e-commerce version and the traditional version of any product are *identical in the sense that both are the result of the same R&D process.* Therefore, both will share the same productivity parameter φ.

The two varieties then differ in the technology of production and distribution. Similarly to firms' self-selection into different export markets in the Melitz model, in this TVO-extension the firm, after entering the market, can decide to serve the demand for its traditional *and/or* e-commerce version in its home and eventually a set of export markets.

[4] To avoid repetitiveness, only the important concepts, novelties and differences of the extension will be laid out. Whenever some element of the Melitz framework is not explicitly dealt with in this section, it is so because it works just as described in the benchmark model.

In other words, every firm has two channels to potentially generate revenue within every country, one with the traditional technology and one with the e-commerce technology. Consider a firm in sector j in country i selling via e-commerce to any country n. In line with section 2.3.1, we assume it to bear different fixed costs of production than those of traditional varieties: $f^{ec}_{nij} \neq f_{nij}$[5]. Differently to the benchmark model, the firm will also bear different marginal costs if it sells online. This is captured by an *additional* e-commerce iceberg factor τ^{ec}_{nij}. In case of exports, the iceberg trade cost, τ_{nij}, still applies to both versions of the good. So, the second, e-commerce, technology is characterized by the following total cost l^{ec}_{nij} of producing q^{ec}_{nij} units of a good:[6]

$$l^{ec}_{nij} = f^{ec}_{nij} + \tau^{ec}_{nij} * \tau_{nij} * \frac{q^{ec}_{nij}}{\varphi}. \tag{2.1}$$

Note, that while $\tau_{nij} \geq 1$ ($\tau_{iij} = 1$), the e-commerce iceberg factor τ^{ec}_{nij} can be smaller, equal or larger than 1, where $\tau^{ec}_{nij} = 1$ implies no difference to the traditional product version. Also note that the e-commerce iceberg factor is multiplicative to the benchmark model's trade cost iceberg factor. Thus, the relation $\frac{\tau^{ec}_{nij}}{\tau_{nij}}$ will in most cases differ between (home and export) markets.

One important aspect regarding τ^{ec}_{nij} is that it captures differences in marginal costs but simultaneously differences in qualitative evaluation of the ec-version against the tr-version by the representative consumer. The reason is that quantities in the Melitz framework are, by definition, quality-adjusted, and potential taste differences between consumers in different countries are, by construction, captured by the trade cost iceberg factor. The same is now the case for e-commerce, so we have to understand τ^{ec}_{nij} as a marginal cost parameter per unit of relative, quality-adjusted quantity, where quality adjustment can differ across markets. As international taste differences between the actual products are already covered by the trade iceberg factor, the ec-iceberg factor will capture a) the difference in marginal costs to serve some markets' ec-demand relative to serving the traditional demand, and b) the potential differences in the relative preference of each market's representative consumer to consume via e-commerce versus via traditional means. This implicit normalization has far-reaching consequences. If, for example, consumers tend to find it more comfortable to order products online, the ec-iceberg factor can be below unity, even if marginal costs are not lower in the ec-sector.[7]

Based on the condition that there are – as will be discussed below – separate residual demands for both product versions, profit maximization results in standard monopolistic competition equations for price, revenue and profit aligned with the ec-version:

[5] For domestic sales, $n = i$.

[6] As in Melitz and Redding 2014a, we focus on a one-factor economy with the normalized factor price $w = 1$. This factor, however, can also be easily interpreted as a composite of several production factors.

[7] Note that, technically, it is also possible to separate the qualitative and cost aspects into two distinct iceberg factors. This will, however, have no qualitative effect on the results, therefore we abstain from that for notational simplicity.

$$
\begin{aligned}
p_{nij}^{ec}(\varphi) &= \frac{\sigma}{\sigma-1}\tau_{nij}^{ec}\frac{\tau_{nij}}{\varphi}, \\
r_{nij}^{ec}(\varphi) &= A_{nj}^{ec}p_{nij}^{ec}(\varphi)^{1-\sigma}, & A_{nj}^{ec} &= X_{nj}(P_{nj}^{ec})^{\sigma-1}, \quad X_{nj} = \beta_{nj}Y, \\
\pi_{nij}^{ec}(\varphi) &= B_{nj}^{ec}\left(\tau_{nij}^{ec}\frac{\tau_{nij}}{\varphi}\right)^{1-\sigma} - f_{nij}^{ec}, & B_{nj}^{ec} &= \frac{(\sigma-1)^{\sigma-1}}{\sigma^{\sigma}}A_{nj}^{ec},
\end{aligned}
\tag{2.2}
$$

where σ is the elasticity of substitution between varieties, Y is the aggregate demand, β_{nj} is the demand share of sector j and P_{nj}^{ec} is the sectoral price index.[8] Here, it is important to note that the sectoral price index is a weighted average of all varieties in the sector. In the TVO, this includes both the ec-varieties and the traditional varieties. Therefore, there is only one price index ($P_{nj}^{ec} = P_{nj}$) per sector and, as a result, demand parameters are also equal: $A_{nj}^{ec} = A_{nj}$ and $B_{nj}^{ec} = B_{nj}$. In the following, we will drop the *ec* superscript for these three parameters.

We will later discuss the restrictions on cost parameter constellations resulting in particular scenarios within the model outcome. At this point, it should be stressed that the TVO extension allows for *any* pattern of cost parameter constellations. Both, fixed and marginal costs can individually be higher, lower or the same in e-commerce, as compared to the traditional product. The patterns can – and realistically will – differ across markets.

One might also question whether indeed every product will have twin-varieties or some products are "by nature" only distributable as an e-commerce product or a traditional product. The assumption in this framework is that there is no such limitation, as there may be a prohibitive cost for either channel, so that some products will ex-post only be sold via one channel, while ex-ante both channels are possible for all products.

Demand side and the twin varieties option (TVO)

Helpman, Melitz, et al. 2004 investigate how, in a Melitz framework, an option for firms to choose between two technologies to serve any (export) market affects the general equilibrium. Their focus is, however, on a straightforward proximity-concentration trade-off, i.e. between higher fixed versus marginal costs. Our e-commerce extension will instead investigate how the option for firms to serve *two separate demands* with two different technologies will affect the general equilibrium. Thus, based on the stylized fact 2.2.2, the most important assumption here is that *customers differ in their preference to buy online or via traditional means.* In particular, in any sector there will be a distribution of preferences, with some costumers having a higher and others a lower willingness to buy a product online, compared to buying the same product via traditional channels. The empirical nature of this distribution is not important for this model, but calls for a conceptualization of this assumption in the representative consumer's utility function. The representative consumer will have to perceive the Harry Potter book bought online as something different to the same book bought in a bookshop, despite the fact that they

[8] Recall that the Melitz model utilizes a Cobb-Douglas utility over sectors and a CES utility over varieties within a sector.

come from the very same line of production. Eventually, the representative consumer will see it as a qualitative difference on the one hand and as improved choice/variety on the other, if a product is offered via both channels. In other words, the ec-version of any product is perceived as good a substitute to the traditional version of that same product as it is to any other product in the sector.

Intuitively, it may seem unreasonable to not assume a Harry Potter book offered via Amazon was a closer substitute to the Harry Potter book sold in a book store, than it is to any other book sold in the book store. The argument in support of this simplifying assumption is that there is no reason either to assume that, within any sector, the technology of distribution would define a variety's uniqueness any less than the uniqueness properties defined in the R&D process. While some customers want a specific book and will compare the options of buying it online or in a traditional store, others will have a preference to go to some specific store (online or traditional) "to buy a present for their mother" and then rather randomly choose between any book presented in that store. Taking this example further, for the same film that a consumer can watch in a cinema or on Netflix, the distribution channel (online or physically in the cinema) might in fact matter more than other characteristics of the product. From the perspective of economic theory, there is no reason to assume one of the approaches dominant. At the same, a departure from the assumption of demand-side neutrality in substitution elasticities will require to use some utility function other than the CES utility, which will complicate the analysis.[9] Furthermore, the qualitative differences of ec-varieties versus tr-varieties are already captured by τ_{ni}^{ec} within every market. Thus, treating the ec-version of a product as another *symmetrical* variety is not as unreasonable as it might seem at first glance. In other words, this approach is based on the assumption that, for the representative consumer, e-commerce facilitation is a uniqueness property for any product, just as any other uniqueness property that distinguishes one variety from another variety within a sector.

The technical implication of this demand-side assumption is that there will be separate residual demands for a traditional version and an e-commerce version of the same good. The underlying assumption of our extension is that the ec-version of any good is a new, or "twin", variety *within the same sector*, which is an imperfect substitute both to the traditional version of the same product and to the other varieties in the sector. This means that every firm's self-selection will take place as independent decisions to offer none, one or both varieties on the home and/or foreign markets. This is a significant structural deviation from Melitz 2003. However, it can still be incorporated within the concept of monopolistic competition and free market entry. In this concept, the mass of varieties is "large" by definition, and natural oligopolies are ruled out as long as there is a meaningful limitation on the number of varieties any single firm can offer within

[9] In fact, an alternative "Twin Sector Option" (TSO) was also developed for the theoretical part of this chapter. Although following slightly different intuitions, both approaches yield qualitatively equivalent results. The TSO-version follows the idea that consumers have inherent priorities on consuming via e-commerce versus traditionally, just as they have inherent priorities to spend across sectors. Technically this is facilitated via a new Cobb-Douglas / CES utility function including a split of each sectoral demand weight β_j into further two sub-sectors. While it is technically more complicated in the calculation of the general equilibrium, the advantage lies in distinct and independent ec- and tr-price indexes and demand parameters B_{nj}. For our theoretical investigation, the TSO-Option became a victim of Ockham's razor.

a market. This, however, does not have to be one variety per firm. Any firm with two varieties will still hold competitive monopolistic positions for both of its product varieties, and its second variety will only be one symmetrical substitute of many to its first variety. In other words, provided that the (large) number of varieties in the market is determined in a general equilibrium, the equilibrium residual demand per variety cannot be influenced by any single firm. Given that we are interested in the duality of e-commerce and traditional markets, it makes sense to technically allow for two (but not more) well defined "twin"-varieties.[10]

The TVO approach, therefore, allows us to work with the utility function from the original Melitz model. The only difference here is that some of the varieties available to the consumer ($\omega \in \Omega_{nj}$) will be ec-versions and others – tr-versions of products. The actual composition of this subset of available varieties within any market will be determined by firms' self-selection in general equilibrium, technically equivalent to the benchmark framework.[11]

Therefore, on the demand side, the TVO only redefines the elements of the utility function, without tampering with its mathematical construction. As a result, the residual demands will be technically equivalent to the benchmark model, where the residual demand for any variety in a market is symmetrical. However, any firm will now face two such – symmetrical yet distinct, i.e. twin – residual demands for a sector j in any country n.

$$q_{nj}^{tr}(\omega) = A_{nj} p_{nj}^{tr}(\omega)^{-\sigma_{nj}} \quad and \quad q_{nj}^{ec}(\omega) = A_{nj} p_{nj}^{ec}(\omega)^{-\sigma_{nj}}. \tag{2.3}$$

Note that, while demands are symmetrical, prices of the two varieties and, thus, quantities will differ as long as the e-commerce costs are different from the traditional ones. And, while the sectoral price index is calculated in the same manner as in the benchmark framework, the new general equilibrium set of available varieties, quantities and prices will result in a new value of the price index.

New set of equilibrium equations

From now on, we will be analyzing one sectoral equilibrium, as in Melitz and Redding 2014a, so in the following we will drop the sector subscript j. The core role of the sectoral equilibrium equations does not change. This "engine" still reflects firms' self-selection, both on the market level and on the aggregated sectoral level. The difference in comparison to the benchmark framework lies in the fact that now there are additional zero-profit conditions for the e-commerce varieties in every market. Also, the free-entry equation is modified to capture the fact that firms, ex-ante to the R&D lottery, can expect profits from ec- and tr-varieties of their product:

[10] In fact, there are frameworks with an extended Melitz structure that allow for more than one variety per firm. However, they still use some modeling tools to limit the number of varieties per firm in order to avoid outcomes with one monopolist dominating every market (see e.g. Bernard, Redding, et al. 2011).

[11] An advantageous feature of the TVO is that, for any given general equilibrium, there will be an endogenous market share of e-commerce varieties within any market, just as there is an endogenous market share of home versus import products.

$$\begin{aligned} \pi_{ni}(\varphi^*_{ni}) = 0 \quad &\Longleftrightarrow \quad B_n(\tau_{ni})^{1-\sigma}(\varphi^*_{ni})^{\sigma-1} = f_{ni}, \\ \pi^{ec}_{ni}(\varphi^{*ec}_{ni}) = 0 \quad &\Longleftrightarrow \quad B_n(\tau^{ec}_{ni})^{1-\sigma}(\tau_{ni})^{1-\sigma}(\varphi^{*ec}_{ni})^{\sigma-1} = f^{ec}_{ni}, \end{aligned} \tag{2.4}$$

$$\begin{aligned} \int_0^\infty [\pi_i(\varphi) + \pi^{ec}_i(\varphi)] dG_i(\varphi) = f_{Ei} \Longleftrightarrow \\ \sum_n \int_{\varphi^*_{ni}}^\infty \left[B_n \tau_{ni}^{1-\sigma} \varphi^{\sigma-1} - f_{ni}\right] dG_i(\varphi) + \\ \sum_n \int_{\varphi^{*ec}_{ni}}^\infty \left[B_n (\tau^{ec}_{ni} * \tau_{ni})^{1-\sigma} (\varphi^{*ec}_{ni})^{\sigma-1} - f^{ec}_{ni}\right] dG_i(\varphi) = f_{Ei}. \end{aligned} \tag{2.5}$$

This set of equilibrium equations implicitly determines the equilibrium vector of N demand parameters B_n (one for each country), NxN zero-profit productivity cut-offs for the traditional varieties φ_{ni} and NxN zero-profit productivity cut-offs for the e-commerce varieties φ^{ec}_{ni}. The difference towards the benchmark framework is that, for every market, there are one home and $N-1$ export productivity cut-offs separately for both traditional *and* e-commerce varieties. Apart from the differences in the self-selection equilibrium and the introduced differences at the demand and supply side, all elements of the new equilibrium are calculated in exactly the same manner as in the model by Melitz and Redding 2014a. In other words, from this point on, the TVO extension works similarly to the benchmark framework. Thus, we can deduct all differences between the benchmark framework's equilibrium and the TVO-equilibrium from the differences in the set of equilibrium parameters only.

2.3.3 Theoretical implications of the TVO extension

In order to meaningfully compare the TVO-extension's equilibrium to the costly trade equilibrium in Melitz and Redding 2014a, we have to make ceteris paribus assumptions on all parameters and productivity distributions, so that the only change is the introduction of the ec-technology.

The best way to understand the of effect e-commerce is to compare the zero-profit productivity cut-offs of the benchmark framework's varieties φ^{*b}_{ni} to the extension's traditional varieties φ^{*tr}_{ni}. From that, we can also draw conclusions about the sectoral price level (P_n) and the market demand per variety (B_n). The rest of outcomes are straightforward to deduct from changes in productivity and the price index, which is directly proportional to B_n.

Proposition 2.1. *For every market n, i, $\varphi^{*tr}_{ni} > \varphi^{*b}_{ni}$, i.e. the traditional cut-off will increase with the introduction of e-commerce, even if traditional cost and demand parameters are unchanged. This result is independent of any particular (positive) values of e-commerce cost parameters, provided that at least one e-commerce market is served.*

Proof. If we combine any country's zero-profit equations as in (2.4) into its free-entry equation (2.5), we get

$$\sum_n \left[f_{ni} J_i(\varphi_{ni}^{*tr})\right] + \sum_n \left[f_{ni}^{ec} J_i^{ec}(\varphi_{ni}^{*ec})\right] = f_{Ei}, \quad \text{where} \tag{2.6}$$

$$J_i(\varphi_{ni}^{*tr}) = \int_{\varphi_{ni}^{*tr}}^{\infty} \left[\frac{\varphi}{\varphi_{ni}^{*tr}}^{\sigma-1} - 1\right] dG(\varphi) \quad \text{and} \quad J_i(\varphi_{ni}^{*ec}) = \int_{\varphi_{ni}^{*ec}}^{\infty} \left[\frac{\varphi}{\varphi_{ni}^{*ec}}^{\sigma-1} - 1\right] dG(\varphi).$$

The same can be done for the benchmark framework's costly trade equilibrium equations, so that there is a similar sum of equations $J(\varphi_{ni}^{*b})$ equal to the market entry cost f_{Ei}. As the entry cost is unchanged, we can relate the two sets of equations:

$$\sum_n \left[f_{ni} J_i(\varphi_{ni}^{*tr})\right] + \sum_n \left[f_{ni}^{ec} J_i^{ec}(\varphi_{ni}^{*ec})\right] = \sum_n \left[f_{ni} J_i(\varphi_{ni}^{*b})\right]. \tag{2.7}$$

Combining the zero-profit equations as in (2.4), one can also select one productivity cut-off for any particular equilibrium and express all other productivity cut-offs in relation to it and the cost parameters. For example, for the domestic ($n = i$) e-commerce cut-off as a function of the domestic traditional cut-off, we get:

$$\varphi_{ii}^{*ec} = \varphi_{ii}^{*tr} \tau_{ii}^{ec} \left(\frac{f_{ii}^{ec}}{f_{ii}^{tr}}\right)^{\frac{1}{\sigma-1}}. \tag{2.8}$$

For the comparison of the two equilibria, we shall choose φ_{ii}^{*tr} as the reference cut-off for the e-commerce extension and φ_{ii}^{*b} for the benchmark framework. Then we can express equation (2.7) as a function of these two cut-offs only: φ_{ii}^{*tr} on the left hand side (LHS) and φ_{ii}^{*b} on the right hand side (RHS). Other than these two cut-offs, the RHS and LHS will only differ because the LHS has the second sum of e-commerce channel-related functions J(.). This sum, in turn, is strictly positive in all elements, as long as at least one e-commerce (domestic or foreign) market is served.[12] The only way the equation can re-balance this additional sum of e-commerce channels is if the functions $J(\varphi_{ii}^{*tr})$ on the LHS have smaller values than the respective functions $J(\varphi_{ii}^{*b})$ on the RHS. As J(.) is downward sloping in the respective φ^*, the equation indeed can only hold if $\varphi_{ii}^{*tr} > \varphi_{ii}^{*b}$. The same logic applies for every export market ni. Therefore, all traditional cut-offs will be above their respective cut-off in the benchmark framework. □

Proposition 2.2. *Introduction of e-commerce will decrease the price index P_n.*

Proof. By combining the formula for the equilibrium residual demand parameter B_n as in equation (2.2) and the respective zero profit condition (2.4), the price index can be expressed as a function of any cut-off such as φ_{ni}^{*tr} or φ_{ni}^{*b}:

[12] This will always be the case if productivity distributions have no or sufficient upper limits. In a trivial corner case, where e-commerce technology is uncompetitive in all sub-markets ni and, thus, not applied at all, the model is equivalent to the benchmark framework.

$$P_n^b = \frac{\sigma}{\sigma-1}\left(\frac{f_{ni}\sigma}{\beta\bar{L}_n}\right)^{1/\sigma-1}\frac{1}{\varphi_{ni}^{*b}}, \quad vs. \quad P_n^{TVO} = \frac{\sigma}{\sigma-1}\left(\frac{f_{ni}\sigma}{\beta\bar{L}_n}\right)^{1/\sigma-1}\frac{1}{\varphi_{ni}^{*tr}}. \tag{2.9}$$

Given that $\varphi_{ni}^{*tr} > \varphi_{ni}^{*b}$ as per Proposition 2.1 and all other elements of the price index are constant, within the TVO extension e-commerce will decrease the price index in all markets. □

Note that this important result holds even if e-commerce is characterized by higher cost parameters and prices. In our framework of monopolistic competition, this decrease in the price index can be interpreted as increased competition within the sector. The two economic reasons for increased competition will be discussed below.

In the Melitz framework, lower sectoral prices technically imply an increased welfare. Since aggregated profits are zero, weighted price indexes inversely determine welfare because they imply higher real wages:

$$U_n = \prod_{j=0}^{J} P_{nj}^{-\beta_j} \tag{2.10}$$

From Proposition 2.2, we can state that e-commerce increases welfare, because of higher real wages. This result, however, builds on the Melitz framework's technical assumptions and should be interpreted with caution. Nevertheless, within this framework, the TVO extension suggests that *e-commerce will benefit consumers with some combination of lower prices and/or increased variety.*

To put these technical results in an economic context, let us recall the basic results of the benchmark Melitz model. Opening up to costly trade increases the zero-profit productivity cut-off. As a result, the least productive firms have to leave the market, the most productive firms serve both domestic and export markets and increase their market shares, and there is profit redistribution in the middle, with some firms only serving domestic markets and potentially loosing their market shares. The mechanics guaranteeing this hierarchy do not change in the e-commerce extension. What does change is a) the number of potential varieties per firm per market and b) the option to either serve the residual demand for a product's traditional variety with a traditional technology, or serve the residual demand for a product's e-commerce variety with an e-commerce technology, or do both.

The e-commerce extension has two mechanisms contributing to increased competition within each sector. The first mechanism is access to additional channels allowing to compete for additional market share ("*multi-channel marketing*").[13] The second mechanism is that the firms get an additional technological option, which potentially allows them to switch to a cheaper technology and, thus, decrease prices in order to gain market share ("*comparative advantage*"). In both cases, increased competition – mirrored

[13] Channel is defined as a revenue stream coming from selling one variety in one market. In the move from autarky to costly trade, the number of channels increases through access to foreign markets. E-commerce allows for a second potential variety per (domestic and foreign) market.

in a lower sectoral price index – ceteris paribus decreases residual demand per variety for all firms active in the sector.

Under the first mechanism, the most productive firms will find it worthwhile to serve residual demands for both their product's e-commerce and traditional variety in each market (domestic and foreign). Through multi-channel marketing, they increase their market share by offering separate product varieties to those consumers preferring e-commerce and to those who prefer traditional consumption. It will only be the most productive firms who find it worthwhile to invest in multiple channels, bearing double fixed costs (dual infrastructures). However, as R&D investments (entry costs) are the same, independent of how many channels are served, there is a cost-side complementarity to serving more channels. Still, given that the firms engaging in multi-channel marketing have, by definition, the most competitive products, this will increase competitive pressure on all varieties sold in the sector. This is reflected in a decreased price index and a lower demand per variety.

In other words: As e-commerce gives the most productive firms an additional channel to compete in each market, which can be understood as diversifying their portfolio of varieties, competitive pressure for all firms will increase. As in the case of trade liberalization and import competition, decreased demand per variety will hurt all firms, including firms selling via two channels. However, some high-productivity firms will potentially be able to overcompensate decreased profits per channel by an increased number of channels contributing to aggregated profits. Note that it is the application of e-commerce technology to serve demand for e-commerce "twin"-varieties, that drives both, the increased competitive pressure and the increased market share of firms applying the ec-technology.

In this sense, we can argue that *the emergence of online markets has similar effects on industrial structures as trade liberalization.* Just as trade liberalization, introduction of e-commerce will, in equilibrium, put pressure on all firms ("import competition"), while giving the biggest firms additional channels ("export opportunities") to increase market shares and eventually profits. Just as imports do not fully replace domestic production, traditional channels will have their place side by side with e-commerce channels. Supporting stylized facts outlined in section 2.2, big firms will engage in multi-channel marketing, investing in dual infrastructures to serve e-commerce *and* traditional consumer preferences with tailored varieties. And, just as trade liberalization, e-commerce will cause increased competition reflected in lower prices and/or more varieties available for consumers.

The second mechanism, however, can lead to cases where the effects of e-commerce are different from those of trade liberalization. Trade liberalization tends to push out the least productive firms, which are not productive enough to find exporting worthwhile and, thus, have no additional channels to compensate for the decreased demand in their home market. E-commerce will make formerly profitable *traditional* business models unprofitable for low-productive firms, as shown in Proposition 2.1. In contrast to trade liberalization, however, e-commerce gives small and medium productive firms an option to eventually *switch towards the new technology, if e-commerce is characterized by lower costs.* In some scenarios, small firms can avoid dropping out by switching the technology.

The reason is that e-commerce is the *option* of a second technology that can be applied *flexibly* to serve some residual demand. Firms at medium levels of productivity, who in equilibrium do not find it worthwhile to invest in dual infrastructures, will serve one, but not the second residual demand for their product in some market. Firms will switch from traditional to e-commerce technology, if the latter allows for higher profits, or, mathematically, if for a market n in country i the following inequality holds:

$$\frac{B_n(\tau_{ni})^{1-\sigma}\varphi^{\sigma-1} - f_{ni}}{B_n(\tau_{ni}^{ec}\tau_{ni})^{1-\sigma}\varphi^{\sigma-1} - f_{ni}^{ec}} \leq 1. \tag{2.11}$$

In order to provide comparatively higher profits, the ec-technology should offer relatively lower marginal and/or fixed costs. On the one hand, lower marginal costs will allow firms to set lower prices and get higher sales. On the other hand, lower fixed costs will allow some less productive firms to receive non-negative profits. They will then be able to compete in markets, where market entry via the traditional channel would be unprofitable. As a result, the number of firms in the market increases. Both effects decrease the sectoral price index. Note that this will also be the case if fixed *or* marginal costs are individually higher than those of traditional varieties, as long as equation (2.11) holds.

Intentionally introducing Ricardo's term here, we can state that *the e-commerce technology will be applied by small and medium firms if and only if it provides them with a "comparative advantage" vis á vis the benchmark traditional technology.* The competitive pressure will also be increased through this mechanism, through more competitors and/or lower prices. However, those firms that do cause this increased competition by switching to e-commerce technology can potentially overcompensate due to lower costs and higher market shares.

Note that the multi-channel mechanism of the most productive firms applying ec-technology additionally to their traditional technology and, thus, increasing competitive pressure is inevitable. In absence of an upper limit on productivity, even in the markets where e-commerce is a comparatively more expensive technology some firms will use e-commerce as a second channel. The mechanism of comparative advantage, however, is only present if and to the degree that the e-commerce technology is in equilibrium comparatively more competitive than the benchmark traditional technology in some markets. Such a case is given whenever the inequality (2.11) holds. The interesting aspect is, however, that some medium productive firms can potentially benefit from e-commerce, despite the fact that the most productive firms will apply it in a multi-channel strategy, inevitably decreasing demand per variety.

Given that the ec-technology is an option, it is, however, clear that for those markets, in which firms do not apply it and stick to the traditional technology, firms will suffer from additional competition and decreased demand for their product. This is the intuition why productivity thresholds for traditional sub-markets rise, as was shown in Proposition 2.1. Whether or not a firm in total gains or loses profits, depends on the interplay of two factors *across all markets*: a) how much additional profit can the firm realize from additional e-commerce channels (opportunities from multi-channel marketing and technology switching) versus b) how much profit does a firm loose through decreased

sales in traditional channels, including markets it must eventually stop serving with the traditional technology (increased competition).

The model allows for an infinite amount of different scenarios, depending on parameter constellations, not only differing between the e-commerce and traditional technology but also across countries and sectors. This broad variety of scenarios is a strength of the TVO, as it allows for a realistic (empirical) investigation of different sector realities. This broad variety, however, also complicates intuitive graphical representation, and discussing all possible scenarios in detail is beyond the scope of this chapter. We will focus on the three basic scenarios, which we will then use to formulate our empirical hypotheses.

2.4 E-commerce scenarios and empirical hypotheses

Depending on parameter constellations, different scenarios will apply that result in specific hierarchies of zero-profit productivity cut-offs. These scenarios will differ for sub-markets.[14] For every sub-market nij, there are three distinct zero-profit productivity cut-offs: (1) benchmark framework (φ^{*b}_{nij}), (2) traditional in TVO (φ^{*tr}_{nij}) and (3) e-commerce in TVO (φ^{*ec}_{nij}). From Proposition 2.1, $\varphi^{*tr}_{nij} > \varphi^{*b}_{nij}$, and so there are three possible hierarchies: (I) $\varphi^{*ec}_{nij} > \varphi^{*tr}_{nij} > \varphi^{*b}_{nij}$, (II) $\varphi^{*tr}_{nij} > \varphi^{*ec}_{nij} > \varphi^{*b}_{nij}$ and (III) $\varphi^{*tr}_{nij} > \varphi^{*b}_{nij} > \varphi^{*ec}_{nij}$. For example, the e-commerce cut-off will be above the traditional cut-off whenever the following inequality holds:

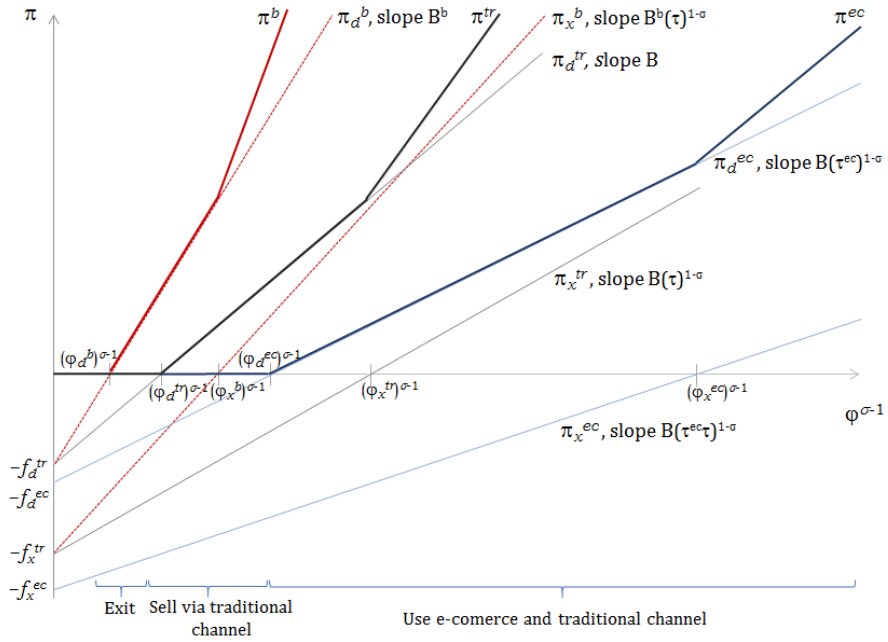
$$\tau^{ec}_{nij} > \left(\frac{f^{tr}_{nij}}{f^{ec}_{nij}} \right)^{\frac{1}{\sigma_j - 1}} . \tag{2.12}$$

It is important to stress that the framework is open for any hierarchy of productivity cut-offs. The famous result by Melitz 2003 that all active firms serve their home market and only the most productive ones serve export markets is a result of parametrical restrictions within the framework. *Such restrictions, however, can be applied to ensure any hierarchy within the model that is considered reasonable.* Melitz's restriction was based on the broad empirical evidence for exporting firms being significantly larger and more productive than domestic-only firms (as shown e.g. by Bernard and Jensen 1999). Such scientifically rigid empirical findings do not yet exist for e-commerce. We will, therefore, briefly discuss all three potential hierarchies outlined above and their implications for the changes in the market structure induced by e-commerce.[15] Based on these contrasting scenarios, we will formulate several hypotheses for our empirical investigation.

[14] A sub-market nij is understood as a sector j in country n which is served form country i, where $n = i$ means firms are serving their domestic market. Any market, or sector, j will is a weighted aggregate of these sub-markets.

[15] Depending on parameters, each of the three scenarios can also deliver several hierarchies of the domestic versus export productivity cut-offs for traditional and e-commerce varieries. We will abstract from those considerations and focus solely on domestic cut-offs. The figures in this section will feature the same hierarchies among export cut-offs as among the domestic ones for illustrative purposes only.

Figure 2.3: Profits and productivity in an economy with e-commerce (tr, ec) vs. the benchmark model (b), scenario 1



Notes: $B^b > B$ as per section 2.3. Subscripts d and x refer to the domestic and the export markets respectively.

2.4.1 Scenario I

E-commerce is only attractive as a second channel for highly productive firms

The first scenario will apply to sub-markets where e-commerce is, in equilibrium, the more "expensive" technology and, thus, comes with a comparative disadvantage. This scenario is tentatively illustrated in figure 2.3.[16] Note, that this can be either because of actually higher fixed and/or marginal costs, or because consumers in a market nj tend to comparatively dislike buying via e-commerce from destination i, which also results in a high τ_{nij}^{ec}. This can, for example, be the case when consumers have a strong preference against buying the sector's products online, which seems to be the case for groceries, as the stylized fact 2.2.2 suggests. However, it can also be that consumers in a country n do not trust online shops of country i because e.g. postal shipping from that destination is considered unreliable or language barriers make online shopping unattractive.

In any such case, however, e-commerce will still be applied by the most productive firms from country i, as, with enough (cost- or quality-related) productivity, they will attract sufficient demand to return a profit. Yet, this will also mean that, in this scenario, e-commerce will *only* be applied as a second channel in competition for consumers who

[16] For simplicity, we assume for this illustration that both $\tau_{nij}^{ec} > 1$ and $f_{nij}^{ec} > f_{nij}^{tr}$, which, however, does not have to be the case simultaneously. For illustrative purposes, respective assumptions are made in the other two scenarios too.

have a "love for variety". An example might again be groceries, where e-commerce is comparatively unattractive vis-a-vis traditional channels, but some of the big retail chains still invest in it as part of a multi-channel strategy within a market. Only the most productive firms will benefit. Some highly productive firms will increase their market share but suffer reduced profitability because of dual fixed costs. And all firms not productive enough to serve a second channel will keep traditional business strategies and suffer from increased competitive pressure from the biggest firms' additional channels. The least productive firms will have to leave the (sub-)market.

On a macro level, this scenario can result in increased average sector productivity because of the reallocation of market shares and input factors from low- to high-productivity firms. This scenario would, thus, further stress the similarities between e-commerce and trade liberalization. In terms of the market structure, in this scenario, one would expect a larger market share taken by the large, productive firms, while small firms loose their market share. Therefore, we formulate the first empirical hypothesis as follows:

Hypothesis 2.1. *Markets with very high e-commerce costs are characterized by higher concentration and very large size of companies trading online.*

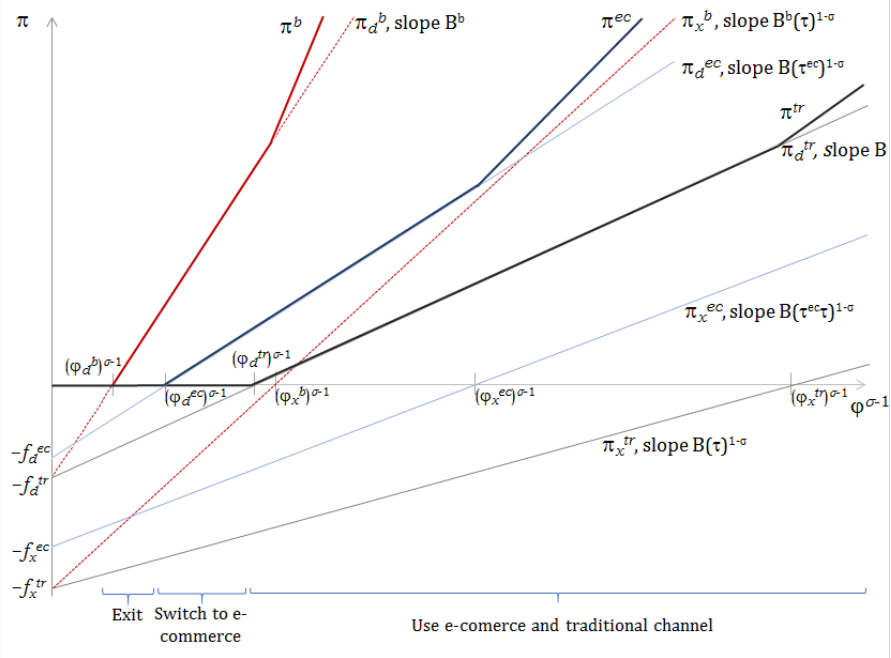
2.4.2 Scenario II

Switching to e-commerce will allow medium productive companies to stay in business, despite increased competitive pressure

The second scenario describes sub-markets nij, where e-commerce is comparatively more attractive than traditional business strategies but not attractive enough to decrease the sub-markets' entry cut-off below the benchmark scenario's cut-off. This scenario is illustrated in figure 2.4. In this case, the biggest and most productive firms will apply a multi-channel strategy, just as described for scenario I. So they will e.g. have a chain of physical shops as well as a complementary e-commerce infrastructure. The defining characteristic of scenario II is that, at the same time, some medium productive firms will find it comparatively more profitable to *switch* from a traditional business strategy to an e-commerce strategy, because this allows them to decrease costs and/or even increase their market share. For some of these firms, it will be a pure question of survival, as switching is the only thing they can do to handle the increased competition. Given that both types of e-commerce appliers will increase competitive pressure, small firms will leave the market if they are not productive enough to find switching to e-commerce profitable. This means that the new smallest firms are, in fact, bigger than the old smallest firms. This is not counter-intuitive when we think about some specialized bookshop chain now serving a national market via the internet, which is small compared to Amazon, but still bigger than the small traditional bookshops, which have had their local market-niche before the emergence of e-commerce competition.

In this scenario, where e-commerce technology comes with a small "comparative advantage", one would expect a strong disruption, reflected in reallocation of market shares and input factors from less to more productive firms and from traditional to e-commerce business concepts and infrastructures. The smallest firms will exit, and some of the medium firms who, despite switching, cannot fully compensate the loss in the market

Figure 2.4: Profits and productivity in an economy with e-commerce (tr, ec) vs. the benchmark model (b), scenario 2



Notes: $B^b > B$ as per section 2.3. Subscripts d and x refer to the domestic and the export markets respectively.

share due to higher competition will lose market shares. The biggest firms with dual channels and some more productive firms who can eventually overcompensate by switching the technology will expand. In a new equilibrium dominated by scenario II, the sector will be characterized by a higher average productivity. As in scenario I, it is uncertain in how far variety is affected, because, on the one hand, costumers will have the additional possibility to consume online varieties, but on the other hand, small unproductive firms and their products will have left the market.

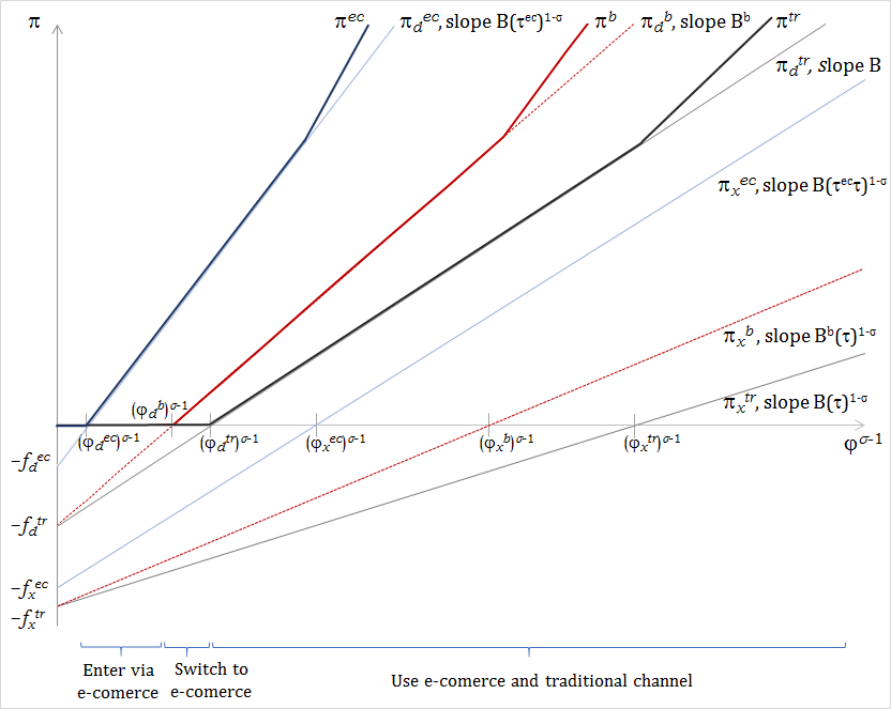
This scenario, by illustrating the switching process, is characterized by two important features: a) the e-commerce costs are not extremely high or low, and b) middle-sized firms increasingly turn to the channel of e-commerce to survive in the market. Any small decrease in the e-commerce costs will allow a slightly less productive firm to survive by switching to e-commerce. Therefore, regarding the dynamics of e-commerce costs we formulate the second hypothesis as follows:

Hypothesis 2.2. *Decreasing e-commerce costs lead to lower average productivity in e-commerce, thus, to smaller average firm size.*

2.4.3 Scenario III

E-Commerce will allow small, low-productivity companies to enter business

Figure 2.5: Profits and productivity in an economy with e-commerce (tr, ec) vs. the benchmark model (b), scenario 3



Notes: $B^b > B$ as per section 2.3. Subscripts d and x refer to the domestic and the export markets respectively.

The third scenario is an accelerated version of scenario (II), with some reversed effects. In this case, the comparative advantage of e-commerce is so strong that the new e-commerce market entry cut-off is even below the benchmark scenario's cut-off: $(\varphi^{*b}_{nij} > \varphi^{*ec}_{nij})$. It will lead to something we could label "the wave of start-ups": low costs and the attractiveness of e-commerce will allow for new market entrants at the lower bound of the productivity distribution, as illustrated in figure 2.5.

As an illustration, imagine a sector formerly dominated by big, highly productive firms, because very high fixed costs made market entry unprofitable for less productive firms. If now e-commerce is characterized by massively lower fixed costs, it will trigger a heavy disruption towards an industrial structure with many additional medium and small competitors. Most big companies will loose despite applying multi-channel strategies, as the loss in the traditional channels' market shares is so strong that the additional e-commerce market shares will not compensate. And most switching companies will also loose, as they end up with a small e-commerce market share in a now highly competitive market instead of a larger traditional market share in a less competitive basic scenario. However, the "wave" of small entrants gain because e-commerce will allow them to get a small but profitable share of a market, in which their product had no chance to participate without e-commerce.

This scenario may attract a Robin Hood sort of sympathy, as there are no firms leaving the market, and instead small firms enter it, capturing market shares of bigger firms and

decreasing market concentration. Because of this, we should also expect higher variety available to consumers. However, the Melitz model (and with it the TVO extension) predict that this could result in a sector characterized by decreased average productivity and higher prices. In contrast to the trade liberalization effect, this scenario suggests that it could well be the other way around. While a possible decrease in productivity already features hypothesis 2.2, this case of extremely low costs of e-commerce is a distinct opposite to hypothesis 2.1, which captures the difference between the impacts of e-commerce and those of trade liberalization:

Hypothesis 2.3. *Markets with very low e-commerce costs are characterized by a large number of online-only firms and overall lower market concentration.*

2.5 Data and empirical approach

For the empirical investigation of our theoretical predictions, we use country- and sector-level data on 35 EU countries and (potential) EU accession candidates between 2005 and 2017[17]. There are several limitations connected to this empirical exercise.

Firstly, in an ideal case, we would want to compare the situation in the same sector with and without e-commerce, or at least before and after e-commerce emerged. The first is, however, not possible, as it is also the case in the studies of trade liberalization versus autarky. The second is, unfortunately, not possible either – in this case, due to the data limitations. It is not until e-commerce became a widespread phenomenon that statistical bodies started gathering information on it. Even when the data are collected, the scope of information is quite limited. Therefore, instead of the "e-commerce vs. no e-commerce" comparison, we rely on a panel dataset and use the sector, country and time variation in e-commerce costs, e-commerce adoption and market structure to investigate the relationships between those three. The unit of observation is, thus, constituted by a sector j in a country i at period t.[18] We will primarily focus on hypotheses 2.1 and 2.3, which simultaneously yield a prediction of an *hump-shaped* relationship between e-commerce costs and market concentration. Namely, starting from high levels of e-commerce costs, a decrease in costs will first increase market concentration (hypothesis 2.1), but at some point the relationship will be reversed, and low and falling e-commerce costs will lead to lower concentration (hypothesis 2.3).

The inclusion of e-commerce adoption in this investigation closely relates to its second limitation. Just like in empirical research on trade, there is no direct measure of e-commerce costs, f^{ec}_{nij} and τ^{ec}_{nij}. These may include observable and unobservable costs of production and distribution, including consumer preferences. Thus, before we can relate e-commerce costs to market concentration, we have to identify proxies for those costs. We, therefore, perform our analysis in three steps.

The first step of the analysis is to identify the best proxies for e-commerce costs among the different potential factors. The underlying assumption is that lower e-commerce costs

[17] The full list of countries is provided in table C.2 in appendix C

[18] In some cases, we have to look at sector groups instead of individual sectors. These cases will be discussed below in more detail.

lead to increasing e-commerce adoption. Our model makes no prediction about the shape of this relationship, and so we concentrate on the simple linear specification:

$$ec_{ijt} = \beta X_{ijt} + \alpha_{ij} + u_{ijt}, \tag{2.13}$$

where ec_{ijt} is a measure of e-commerce adoption in country i and sector j at time t (share of enterprises with digital, web-, or electronic sales or share of turnover from electronic sales), X_{ijt} is a set of potential proxies for e-commerce costs, α_{ij} are country-sector fixed effects, and u_{ijt} is the error term. As we are looking at the panel data differentiating between countries and sectors, we focus on the fixed effects regressions throughout the analysis. We work with this step as an exploratory analysis, so we do not specify any restricted set of the cost proxies at this stage. Instead, any of the potential proxies, which we discuss below, can serve as variables here.

The second step is an intermediary step before relating e-commerce costs to market concentration. In this step, we investigate the relationship between e-commerce adoption and concentration. E-commerce adoption is the channel connecting e-commerce costs to market structure. High e-commerce costs lead to higher market concentration because of low levels of e-commerce adoption, whereby only large enterprises use it and profit from it. Very low e-commerce costs lead to lower market concentration because many small enterprises can use e-commerce to enter the market. Therefore, also for e-commerce adoption, simultaneous consideration of hypotheses 2.1 and 2.3 yields an hump-shaped relationship to market concentration:

$$mc_{ijt} = \beta_1 ec_{ijt} + \beta_2 {ec^2}_{ijt} + \gamma Z_{ijt} + \alpha_{ij} + u_{ijt}, \tag{2.14}$$

where mc_{ijt} is a measure of market concentration in country i and sector j at time t, ec_{ijt} is a measure of e-commerce adoption, Z_{ijt} is a set of potential concentration drivers other than e-commerce (e.g. R&D intensity), α_{ij} are country-sector fixed effects, and u_{ijt} is the error term. If hypotheses 2.1 and 2.3 both hold, we expect $\beta_1 > 0$ and $\beta_2 < 0$.

We estimate this equation with two approaches. First, we use the robust fixed-effects regression. Second, to both handle the potential endogeneity issue and to build a bridge to e-commerce costs, we also estimate the equation with an IV approach, where e-commerce adoption is considered endogenous and e-commerce costs are used as instruments. Since some of e-commerce costs may also be simultaneous to e-commerce use and/or market structure, we use lagged variables to deal with that endogeneity.

In the final step, we relate e-commerce costs directly to market concentration. Here we also make use of a much more detailed dataset, as will be discussed below. According to hypotheses 2.1 and 2.3, we expect that this relationship is non-linear:

$$mc_{ijt} = \beta_1 X_{ijt} + \beta_2 {X^2}_{ijt} + \gamma Z_{ijt} + \alpha_{ij} + u_{ijt}, \tag{2.15}$$

where mc_{ijt} is a measure of market concentration, X_{ijt} is a set of proxies for e-commerce costs, Z_{ijt} is a set of potential concentration drivers other than e-commerce, α_{ij} are country-sector fixed effects, and u_{ijt} is the error term.

The third limitation to our empirical analysis is data availability. We focus on the EU statistics, as the EU provides a combination of both some of the most comprehensive statistics in terms of the variables collected and a time and geographic span that allows for a sufficient number of observations, while offering harmonized data throughout the whole dataset. The time span between 2005 and 2017 is also restricted by data availability. Even in this case, not all variables are available for the whole period, especially with regard to e-commerce usage and consumers' online shopping behavior. The Eurostat data are further complemented by the data from the European Central Bank, the World Bank and the Universal Postal Union. The full list of variables is summarized in table C.5 in appendix C, here we limit ourselves to briefly describing the main groups of variables:

Basic enterprise characteristics. The basic data about the enterprises include their number, turnover, employment, and value added and are very detailed in terms of sectors represented. Based on those data, we derive our measure of market concentration, which is the gini coefficient, calculated from the firm turnover and firm number within five enterprise size classes.[19]

E-commerce in enterprises. The data on ICT usage and e-commerce adoption on the firm side represent the share of enterprises using e-commerce or share of the turnover devoted to it and include different notions of e-commerce, e.g. digital sales as the broadest concept, electronic sales, and web sales, which only include sales via websites or own apps. However, the data only feature broad sector groups instead of individual sectors (e.g., one of the groups would include food, wood and textile processing, clothes and leather products). As a result of this, we have to split our data into two distinct samples. One, which includes data on e-commerce adoption, will be used in the first two steps of the analysis. The second, featuring individual sectors but no information about e-commerce adoption, will be used in the third step.

Consumers' ICT usage and online shopping behavior. This group of variables concerns the consumer side of potential e-commerce costs. Beside the differences in preferences highlighted in the stylized fact 2.2.2, a number of empirical studies highlight the role of ICT infrastructure and consumers' awareness about and trust to online shopping (Alyoubi 2015; Choshin and Ghaffari 2017; Gefen 2000; Martinsons 2008; Iglesias-Pradas et al. 2013; Oliveira et al. 2017). We, therefore, use the data on access to the internet, computer usage, frequency and purchase amounts of online shopping, concerns about buying online, and online shopping by products to proxy for preferences and infrastructure access on the consumer side. The variables are measured as percentage of population and, thus,

[19] In fact, the most-used concentration measures are the concentration ratio and the Herfindahl-Hirschman index (HHI) (Bikker and Haaf 2002), however, both require firm-level information. Our data only contain information grouped by employment size classes – a data type the gini coefficient can handle very well. We experimented with calculating the HHI by assuming uniform distribution of firm turnover within one size class – being aware, however, that this would significantly understate the HHI. Indeed, the mean value of the HHI was unreasonably low at 0.03 (with standard deviation of 0.067), and it barely correlated with the gini coefficient (the correlation coefficient was -0.124).

are available on the country level, however, with time variation. In addition, we use the product-related variables to construct a country-sector measure of online shopping.[20]

Logistics and postal services. Unlike much of the traditional business, e-commerce relies much on delivery of their products by post or other logistics companies (see e.g. Gomez-Herrera et al. 2013). Therefore, we utilize the data by Universal Postal Union, which provide a comprehensive overview on post services, including the population covered by post services, the number and characteristics of post offices, collections and deliveries of post, dispatch and receipt of parcels as well as financial data of the postal companies in respective countries. The data are complemented by the logistics performance index of the World Bank.

Other sectoral data. To control for further potential drivers of both e-commerce adoption and market concentration, we also collect a set of sectoral data. We use two product classifications, which can potentially reflect the durability or complexity of goods. However, both are time-invariant and, due to the use of fixed effects, can only be utilized in interaction with other variables. The first classification is the Main Industrial Grouping (MIG), which classifies manufacturing products as consumer products (food, non-durable and durable goods), intermediate goods and investment (capital) goods. The second classification is based on the "Search-Experience-Credence" (SEC) framework and describes the goods in terms of how easy information about them can be obtained. Using this classification is based on the idea from the stylized fact 2.2.2 that consumers might have a preference against buying products they cannot credibly examine audio-visually. An overview of our SEC classification is presented in table C.3 in appendix C. To use the two classifications in our regression analysis, we also convert them to a score of "durability" or "complexity". A further potentially important product characteristic is product bulkiness. As e-commerce largely depends on individually shipping the orders by post or logistic services, bulky goods can substantially increase the shipment costs.

Factors that potentially drive concentration are R&D expenditure (Gayle 2001; Dolata 2017), intangible assets (Crouzet and Eberly 2019), capital intensity (Curry and George 1983; Crouzet and Eberly 2019), trade openness (H. Egger and P. Egger 2003), and advertising (Greer 1971; Sá 2015). The relation of the latter to concentration, however, is not quite clear (ibid.), and the respective data are not available, so we acknowledge that this factor might be missing in our analysis, but are convinced that it will not significantly influence our results. For R&D expenditure, intangible assets and capital intensity, based on data availability, we use the respective investment data.

Country aggregates and institutional data. Finally, we also collect a set of country-level economic and institutional data, for factors that might influence market concentration and/or adoption of e-commerce. Both economic prosperity and institutional environment have been found to play a role in e-commerce adoption (Alyoubi 2015; Gomez-Herrera et al. 2013; Ho et al. 2007; Martinsons 2008). Thus, we collect basic economic data

[20] Clearly, not all sectors can be covered by such measure, but assuming that, e.g., online purchase of food can theoretically relate to both grocery retailers and food producers, this measure covers a reasonable amount of manufacturing and services sectors.

(such as GDP per capita, individual and household consumption, country-level wages and productivity, transport infrastructure, population and employment), data on payments infrastructure (e.g., the number and value of card transactions) and institutional measures (such as credit rights and contract enforcement, the ease of starting business or customs burden).

The data availability issues leave us with several distinct datasets to use in further analysis. Their main characteristics are summarized in table C.1 in appendix C. The first dataset is based on the sector group data, where information on e-commerce adoption by enterprises is available. This is the dataset we work with when estimating equations (2.13) and (2.14). In presence of data on enterprise number and turnover, we could also extrapolate the dataset to several more (smaller) groups. We use the extrapolated dataset, which we call *ec-extra* sample for brevity, for our analysis. Being aware of potential data quality issues related to such extrapolation, we also repeat the main steps with the original (smaller) dataset (*ec* sample) to check the robustness of our results. Overall, without considering gaps in individual variables, the *ec-extra* sample includes 35 countries, 22 sector groups and 13 years, making a total of about 10 thousand observations. The number of observations used in any given model specification, however, is much smaller and depends on the number and overlaps of data gaps. The *ec* sample only has 18 sector groups, totaling a maximum of 8,190 possible observations.

The second dataset, which we focus on when estimating equation (2.15), consists of all non-agricultural sectors of European economies and combines basic enterprise statistics with investment data and country-level consumer ICT data, economic aggregates and institutional measures. Depending on the variable, the sectoral detalization goes up to the 4th NACE digit (with a total of 649 sectors), though most of the data are only available at the 2-digit classification (with a total of 95 sectors). The maximum possible number of observations in this sample is 295,295. However, as will be discussed in section 2.6, the panel is very unbalanced at the variable level, and in any particular model specification, only a small fraction of this data can be used. We will further refer to this dataset as the *full* sample.

2.6 Results and discussion

Overall, our dataset counts about 240 possible independent variables as proxies of e-commerce costs and controls for market concentration. It is impossible to cover all results in the scope of this chapter, so we only show those most statistically significant and best performing in terms of the model explanatory power. Of those, only the major results are presented in this section, the rest can be viewed in appendix D. Another important point is that there are a lot of gaps in the data, which often do not overlap across variables. Different countries, sectors and years are missing for different variables, and data on e-commerce adoption and online shopping behavior are only available for a few years. As a result, there is massive variation in the number of observations, depending on the specification, which is especially the case for the full sample. This can be observed in the descriptive statistics of the selected variables in table C.6 in appendix C.

Before turning to the empirical evidence on hypotheses 2.1 and 2.3, we will briefly discuss our exploratory study of e-commerce costs in the ec-extra sample, which generated the cost proxies for further analysis.

2.6.1 Exploratory analysis of e-commerce costs

The results in this section are based on estimating equation (2.13), substituting X_{ijt} with different combinations of potential explanatory variables. This step serves to limit the high number of variables available in our dataset to a few that can well explain the adoption of e-commerce. The main results are presented in table 2.1.

Overall, the explanatory power of the specifications for the enterprises doing digital sales and electronic sales is quite low. However, there is a difference between manufacturing and services. As shown in columns (2), (3) and (7) of table 2.1, the e-commerce cost proxies explain the share of enterprises selling digitally or electronically quite well in manufacturing, with R-squared reaching 20-30%. For services, however, the explanatory power is very low (mostly below 5%). For *turnover* from electronic sales, on the contrary, selected variables perform very well both in the whole sample and separately in manufacturing and services, with highly significant coefficients and R-squared reaching close to 50%. It is, in fact, very reasonable, as digital and electronic sales measure the number of enterprises adopting e-commerce, independent of their size or importance of e-commerce in their sales, while the turnover variables measure the *scale* of e-commerce adoption in the sector. Turning to specific cost proxies, there are several distinct groups of factors that can well explain e-commerce adoption.

The first group is online shopping behavior and buyers' concerns about buying online. We consider these as indications of preferences towards or against buying online. The percentage of individuals buying online is positively related to e-commerce adoption (also in lagged form). As shown table D.1 in appendix D.1, not buying online, vice versa, is negatively related to e-commerce adoption. High number of buyers concerned about different (potential) problems of online shopping is, generally, also negatively related to e-commerce adoption (see full list of results in table D.3 of appendix D.1). However, only concerns about seller fraud consistently significantly affect the share of e-commerce. This result is as intuitive as the overall preference towards online shopping: lack of trust towards online sellers induces them to put additional effort in their image and customer protection policies, thus increasing the cost of selling via e-commerce.

The second group of important factors is indeed the availability of postal services. Lack of postal coverage consistently negatively impacts e-commerce adoption, while the number of post offices and frequency of urban deliveries are positively related to it.

The third group is that of aggregate economic and institutional data. Here, the evidence is somewhat mixed. Overall income level (lagged GDP per capita) seems to be positively yet inconsistently related to e-commerce, and this is about the only economic characteristic having any impact. In terms of institutions, the development of payment systems (lagged number of card transactions per capita) seems positively related to e-commerce, but only if looking at turnover. With regard to the business environment, the time to register property or build a warehouse tends to be negatively related to e-commerce, as expected, but the cost of starting business actually shows an opposite relation (though

Table 2.1: Estimation results for e-commerce adoption, ec-extra sample

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Digital sales	Digital sales manufacturing	Digital sales services	Digital sales	Digital sales	E-sales	E-sales manufacturing	E-sales	E-sales turnover	E-sales turnover
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP p.c. -CLV EUR$_{t-1}$	0.001**									
	(0.000)									
E-shoppers$_{t-1}$	0.266***	0.231***	0.347***	0.294***	0.171***	0.256***	0.204***	0.101***	0.092***	0.092
	(0.025)	(0.030)	(0.036)	(0.028)	(0.020)	(0.024)	(0.027)	(0.013)	(0.019)	(0.068)
E-shopping problem: fraud	−0.904***	−1.141***	−0.818***	−0.771***	−0.544**	−0.874***	−1.136***		−0.505***	2.044***
	(0.216)	(0.275)	(0.272)	(0.197)	(0.251)	(0.180)	(0.259)		(0.156)	(0.352)
E-shopping problem: high price			0.931***							
			(0.243)							
Post offices$_{t-1}$		0.246***		0.211***	0.090	0.169***	0.209***	−0.276*	0.166***	−1.997***
		(0.042)		(0.067)	(0.403)	(0.061)	(0.037)	(0.140)	(0.024)	(0.548)
Deliveries urban$_{t-1}$	10.562***	10.441***	6.933***	8.501***	9.060***	7.453***	8.235***	5.474***		
	(2.210)	(1.350)	(2.559)	(2.612)	(1.116)	(1.935)	(1.304)	(1.810)		
No post	−10.745***	−138.897***		−100.549**	−130.134***	−126.680***	−156.827***		−50.035***	12.752
	(40.439)	(22.401)		(40.777)	(22.012)	(41.517)	(24.374)		(9.841)	(16.946)
Post financial revenue$_{t-1}$			0.237**							
			(0.097)							
Production$_{t-1}$									0.402***	0.710***
									(0.037)	(0.106)
Property reg. time				0.031		−0.013*	−0.013***			
				(0.028)		(0.008)	(0.003)			
Warehouse time				−0.009*					−0.006*	−0.006*
				(0.005)					(0.003)	(0.003)
Card payments -Nr p.c.$_{t-1}$									0.096***	0.118***
									(0.010)	(0.036)
Business start cost									0.139**	
									(0.064)	
Customs burden									−1.194**	−1.434
									(0.540)	(0.938)
R&D rate current$_{t-1}$					30.930***			−2.003***		−12.082
					(5.635)			(0.596)		(20.088)
R&D rate capital$_{t-1}$					−175.498***					−49.211
					(37.535)					(45.938)
Firm share world					−51.681***			−31.970		−38.837***
					(13.168)			(10.976)		(7.415)

Table 2.1 – continued from previous page

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Digital sales	Digital sales manufacturing	Digital sales services	Digital sales	Digital sales	E-sales	E-sales manufacturing	E-sales	E-sales turnover	E-sales turnover
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
N.obs.	1267	482	711	1161	325	1160	462	1064	1200	347
N.groups	471	191	261	446	195	449	182	238	441	162
R-squared	0.039	0.318	0.054	0.091	0.156	0.064	0.284	0.130	0.458	0.302
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. In specificaton (1), following (statisticaly insignificant) variables are not shown: Post offices person coverage, Price index '15_{t-1}

not always statistically significant). The latter result is very interesting and calls for caution when interpreting this set of variables. These variables, in fact, might rather measure sunk entry costs (f_{Eij}) than the fixed costs of production (f^{ec}_{nij}) and, thus, affect both traditional and e-commerce sectors the same way[21]. Within our theoretical model, this would mean that, if e-commerce comes with lower costs and, thus, allows more enterprises to break-even on the entry cost, this channel will be preferred to the traditional one. Therefore, in combination with other e-commerce cost proxies, one might see a positive relation between the entry cost and e-commerce adoption.

The fourth, and final, group of factors is that of the sector characteristics. The variables relating specifically to product attributes, such as MIG score or bulkiness, neither show any significant impact on e-commerce by themselves, nor when interacted with the measure of countries' infrastructure. On the producer's side, interesting and very indicative, the share of foreign ownership tends to have a negative impact on e-commerce. It is fully in line with the literature highlighting the importance of location and market potential for the FDI: many companies set up enterprises or branches in foreign countries to better reach consumers through physical presence. But for e-commerce, not based on physical shops, investing in these structures is redundant.

There are also factors, whose investigation is not reported here, as they did not deliver any significant results. These are, in the first place, the measures of the level and type of internet access and computer usage. One group worth mentioning is also the perception of the companies themselves about the difficulties of selling online. These data are only available in this sample, so we do not relate this analysis to market concentration and only report the summary of the results in the appendix (table D.2). Supporting the importance of postal services, the fitness of the products and logistical difficulties are the only two problem groups consistently affecting adoption of e-commerce.

2.6.2 E-commerce and market structure

Having chosen the best-performing proxies of e-commerce costs, in the second step, we look at the relationship between market concentration (gini) and adoption of e-commerce. In other words, in this section we estimate equation (2.14), substituting ec_{ijt} with either enterprises with digital sales, or e-sales, or with turnover share from e-sales. The results are presented in table 2.2, some further specifications can also be viewed in appendix D.2. In addition, figure 2.6 illustrates the relationship between e-commerce adoption and market concentration.

For digital sales (column (1)), the coefficients speak for an hump-shaped relationship, however, the statistical significance of this result is not consistent throughout the model specifications (compare to table D.9 in appendix D.2). Yet, this relationship becomes inverted, if digital sales are instrumented with e-commerce cost proxies. At the same time, the explanatory power of the IV specification is extremely low. The results are similar for electronic sales and are, therefore, not reported.

[21] This is not necessarily the case for property and warehouses, but highly likely for the overall business setup process

Table 2.2: Estimation results for the gini coefficient and e-commerce adoption, ec-extra sample

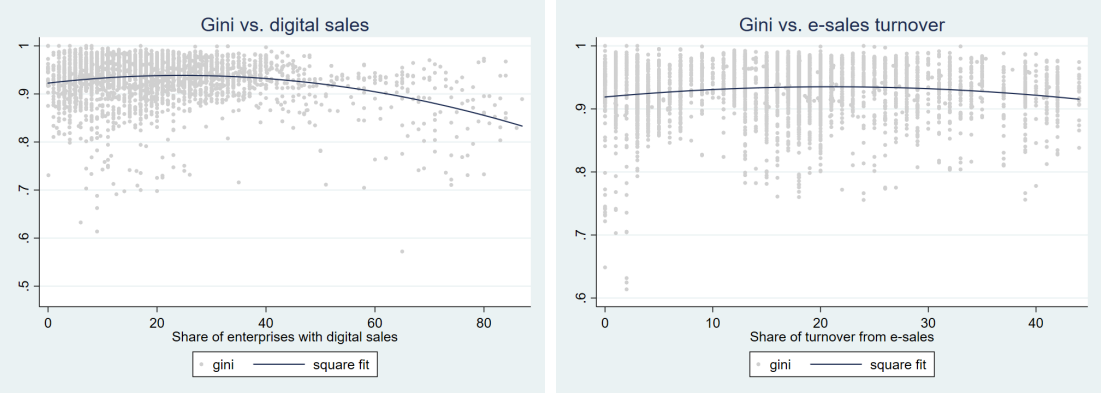
Method	Robust FE	IV, Robust FE	Robust FE	Robust FE	IV, Robust FE	IV, Robust FE	IV, Robust FE
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Digital sales	0.063*	−0.484					
	(0.035)	(0.358)					
Digital sales^2	−0.131	1.965*					
	(0.087)	(1.138)					
E-sales turnover			0.117***	0.449***	0.706***	−0.052	1.275***
			(0.040)	(0.104)	(0.257)	(0.188)	(0.184)
E-sales turnover^2			−0.272***	−1.063***	−1.255***	−0.046	−2.440***
			(0.084)	(0.258)	(0.412)	(0.400)	(0.382)
GDP p.c.		2.992**	−1.057***	−4.829***	−6.433**	0.244	−6.728***
-CLV EUR$_{t-1}$		(1.887)	(0.368)	(1.309)	(2.673)	(0.646)	(2.379)
Price index '15$_{t-1}$			0.000***		0.000***	0.000	
			(0.000)		(0.000)	(0.000)	
Wage per hour PPS			0.002***	0.003**	0.001	0.001***	−0.004**
			(0.000)	(0.001)	(0.001)	(0.000)	(0.002)
Labor productivity per hour			0.241**	0.104	0.111	0.240	
			(0.099)	(0.298)	(0.660)	(0.224)	
R&D rate current$_{t-1}$				−0.090			
				(0.142)			
R&D rate capital$_{t-1}$	−0.086***			−0.019			
	(0.021)			(0.019)			
Investment in buildings	−11.192***			−18.919***			−17.721
	(4.129)			(7.169)			(18.943)
Firm share world	−0.857***	−0.869***					
	(0.084)	(0.153)					
VA share world	0.026*						
	(0.015)						
N.obs.	585	324	3688	821	828	1715	390
N.groups	131	186	550	167	375	507	198
R-squared	0.419	0.029	0.037	0.119	0.122	0.028	0.235
Prob > F/ Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2.2 – continued from previous page

Method Dep. variable: gini	Robust FE (1)	IV, Robust FE (2)	Robust FE (3)	Robust FE (4)	IV, Robust FE (5)	IV, Robust FE (6)	IV, Robust FE (7)
Instruments		E-shoppers$_{t-1}$, E-shopping problem: fraud, Deliveries urban$_{t-1}$, R&D rate current$_{t-1}$, R&D rate capital$_{t-1}$			E-shoppers$_{t-1}$, E-shopping problem: fraud, Post offices$_{t-1}$, Production$_{t-1}$, Card payments -Nr p.c.$_{t-1}$, Business start cost, No post, Warehouse time, Customs burden	E-shoppers$_{t-1}$, Warehouse time, LPI shipment time, LPI customs, LPI shipment price	E-shoppers$_{t-1}$, E-shopping problem: fraud, Post offices$_{t-1}$, No post, Production$_{t-1}$, Card payments -Nr p.c.$_{t-1}$, Warehouse time, Customs burden, Firm share world
AR, Prob > chi2		0.000			0.000	0.000	0.000
Wald, Prob > chi2		0.074			0.007	0.313	0.000
Omitted controls	Turnover share world, Investment share world, Employment share world		E-access	Investment in construction, Investment in machinery			

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. "Omitted controls" lists variables in the model specification, whose coefficients were not statistically significant and which are not shown in this table for the sake of brevity.

Figure 2.6: Market concentration and e-commerce adoption



For electronic sales as turnover share, however, the hump-shape is present and statistically significant. This finding is robust to adding control variables (income level, price level, hourly compensation of employees). The expansion to more control variables and use of IV improves the model fit substantially. An exception, however, is the use of logistics performance indices as instruments. While doing quite well in explaining e-commerce turnover in the first analysis step (compare to table D.8 in appendix D.1), these variables perform very poorly as instruments, as shown by the Wald test and very low explanatory power of the specification in column (6). Overall, however, the data tend to lend support to hypotheses 2.1 and 2.3 so far.

In the third step, we relate market concentration directly to the proxies of e-commerce costs, which is a more direct application of hypotheses 2.1 and 2.3. We perform this exercise both for the ec-extra and the full sample. Based on the previous results, we start by adding all variables that were able to well explain e-commerce adoption and test for a presence of a hump-shaped relationship for each of them. Equation (2.15) then translates into the following basic version:

$$\begin{aligned} gini_{ijt} = \; & \beta_1 E\_Shopping_{it} + \beta_2 E\_Shopping^2{}_{it} + \beta_3 Postal\_Services_{it} + \beta_4 Postal\_Services^2{}_{it} + \\ & \beta_5 Institutions_{it} + \beta_6 Institutions^2{}_{it} + \beta_7 Sector_{ijt} + \beta_8 Sector^2{}_{ijt} + \\ & \gamma_1 Aggregates_{it} + \gamma_2 Aggregates^2{}_{it} + \alpha_{ij} + u_{ijt}, \end{aligned}$$

where *E_Shopping* is the lagged share of population buying online and share of buyers concerned about fraud, *Postal_Services* are number of post offices with delivery staff and share of population not covered by post, *Institutions* are institutional variables (number of card payments per capita, cost of starting business, time to build a warehouse, customs burden), *Sector* are sector-related variables (share of foreign enterprises, current and capital R&D rate, capital investment in buildings, price index) and *Aggregates* are country-level economic variables (GDP per capita, hourly wage, hourly labor productivity). Indices i, j, t refer to countries, sectors and years respectively.

The main results are reported in table 2.3. The evidence is somewhat mixed here. The model fit depends crucially on the presence of sectoral variables, such as R&D intensity and foreign ownership, which can be considered both as e-commerce cost drivers and direct drivers of market concentration. In their absence (column (1)), the explanatory power of the model is very low and, for many of the e-commerce cost proxies, the relationship to market concentration is rather linear. That being said, the direction of the impact is, in most cases as expected: higher e-commerce costs (or lower preference for e-commerce) lead to higher concentration for linear relationships, or reveal an hump-shape. Two exceptions with highly significant coefficients are, in the first two specifications, concerns about fraud and the number of post offices. For the latter, however, we also tested statistically for the u-shape and found no significant result: the extreme point of the square function was very close to the upper bound of the variable values (for specification (1), 26017.62 out of 27600), so one would rather speak of a diminishing negative effect on market concentration, which is also in line with the expectations.

We also experiment with alternative dependent variables, which might potentially reflect market structure: share of turnover and share of enterprises for small and large enterprises (by employment size class). We are aware, however, that the number of enterprises does not well represent the market structure, as the financial aspect is missing, while the turnover share might not be a good measure in this particular case. The reason for the latter is that the data are based on the employment size classification, which depends a lot on the country and sector under investigation. For example, an enterprise providing call center services (highly labor-based activity) with over 300 employees might still have lower turnover than a furniture producer with automated plants and mostly administrative personnel. Continuing this example, a less productive furniture producer, with lower turnover, might actually have less possibilities to automate the process and need to hire more factory workers. While the first issue can be handled by the country-sector fixed effects, the second reverts the connection between employment and financial size. It is, therefore, by no means granted that, for employment or turnover shares, we find the same relationship with e-commerce as for the gini coefficient.

Indeed, in most cases no significant relationship is found (the summary of results can be viewed in table D.10 in appendix D.3). There are two notable exceptions, however. First, for e-sales turnover, for large enterprises (over 250 employees) the relationship actually seems to be a u-shaped one (not hump-shaped, as would be expected) and, mirroring this, we find an hump-shaped relationship for very small enterprises (less than 10 employees). The same holds for the share of large or, respectively, very small enterprises in the total enterprise number. The second – very intriguing – exception is the u-shaped relationship with e-commerce adoption for small-medium enterprises (10-19 employees), when measured by turnover share. In other words, at low levels of e-commerce adoption, small-medium enterprises lose their market shares, while regaining it at higher e-commerce levels and again facing tougher competition as e-commerce becomes widespread. This is, in fact, exactly what is predicted by hypotheses 2.1 and 2.3.

To check the robustness of the results we obtained so far to the extrapolation of the original e-commerce dataset with the data on absolute number and turnover of enterprises, we also re-run the main specifications on a smaller, non-extrapolated sample. The results are presented in appendix D.4 and are very similar to those of the larger sample.

Table 2.3: Estimation results for the gini coefficient and e-commerce costs

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Sample	ec-extra	ec-extra	full	full	full	full
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)
E-shoppers$_{t-1}$	0.043	0.131*	0.167***	0.116***	0.114***	
	(0.048)	(0.078)	(0.057)	(0.022)	(0.022)	
E-shoppers$_{t-1}^2$	0.176**	−0.216	−0.203***	−0.104***	−0.104***	
	(0.081)	(0.140)	(0.075)	(0.032)	(0.032)	
E-shoppers$_{t-1}$* lowcost sf					0.037	
					(0.027)	
E-shoppers$_{t-1}$* highcost sf					0.036*	
					(0.021)	
E-shopping problem: fraud	−1.299***					
	(0.387)					
E-shopping problem: fraud^2	20.575***					
	(7.903)					
Post office$_{t-1}$	−0.013***	0.033**	0.011*	0.001	0.001	
	(0.004)	(0.016)	(0.006)	(0.000)	(0.000)	
Post office$_{t-1}$* lowcost sf					−0.003***	
					(0.001)	
Post office$_{t-1}$* highcost sf					0.001	
					(0.001)	
Post office$_{t-1}^2$	0.000**	−0.004*	−0.003**			
	(0.000)	(0.003)	(0.001)			
No post	−0.062	0.000	−0.317*			
	(0.226)	(0.058)	(0.163)			
No post^2	10.374**	1.894	11.209**			
	(5.228)	(1.734)	(4.583)			
EC cost						−2.633**
						(1.179)
EC cost^2						3.258**
						(1.473)
EC cost*lowcost						3.270**
						(1.269)
EC cost^2*lowcost						−5.558**
						(2.270)
EC cost*highcost						−2.511***
						(0.931)
Production$_{t-1}$	1.418	−1.234	−0.053	0.227*	0.231*	
	(0.891)	(1.039)	(0.135)	(0.121)	(0.121)	
Production$_{t-1}^2$	−2.012	7.145	−0.035	−0.150	−0.153	
	(7.446)	(7.082)	(0.143)	(0.116)	(0.116)	
Business start cost	−0.003*					
	(0.002)					
Warehouse time$_{t-1}$	0.185***	−0.442	−0.140***	−0.043*	−0.044*	0.196
	(0.070)	(0.275)	(0.050)	(0.026)	(0.026)	(0.250)
Warehouse time$_{t-1}^2$	−0.275**	0.653*	0.181**	0.095***	0.096***	−0.738
	(0.121)	(0.382)	(0.079)	(0.037)	(0.037)	(0.519)
Customs burden		−0.009	−0.201**	0.054***	0.054	0.071**
		(0.050)	(0.092)	(0.015)	(0.014)	(0.034)
Customs burden^2	−0.001*	0.001	0.020**	−0.007***	−0.007***	−0.008*
	(0.001)	(0.005)	(0.010)	(0.002)	(0.002)	(0.004)
Firm share world		−1.017***	−1.171***	−0.637***	−0.636***	
		(0.140)	(0.225)	(0.142)	(0.143)	
Firm share world^2		0.036	1.443**	0.444	0.441	
		(0.682)	(0.567)	(0.341)	(0.342)	

Table 2.3 – continued from previous page

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Sample	ec-extra	ec-extra	full	full	full	full
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)
R&D rate current$_{t-1}$		−0.679	−0.074	−0.037***	−0.037***	
		(0.610)	(0.099)	(0.012)	(0.012)	
R&D rate current$_{t-1}^2$		1.979	0.012	0.007***	0.007***	
		(1.599)	(0.019)	(0.002)	(0.002)	
GDP p.c.	0.693	−2.137	−0.780	−2.117***	−2.080**	4.590**
-CLV EUR$_{t-1}$	(2.278)	(1.360)	(1.674)	(0.811)	−(0.814)	(2.020)
Price index '15$_{t-1}$	−0.001***					
	(0.000)					
Productivity per hour	1.090*					
	(0.598)					
N.obs.	1010	581	1057	3232	3232	1200
N.groups	430	136	286	864	864	207
R-squared	0.048	0.181	0.142	0.183	0.181	0.004
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000
Omitted controls	Card payments -Nr p.c.$_{t-1}$, Card payments -Nr p.c.$_{t-1}^2$, Business start cost^2, Wage per hour PPS	R&D rate capital$_{t-1}$, R&D rate capital$_{t-1}^2$, Inv. in buildings	R&D rate capital$_{t-1}$, R&D rate capital$_{t-1}^2$, Inv. in buildings	Inv. rate$_{t-1}$	Inv. rate$_{t-1}$	Inv. rate$_{t-1}$

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. "Omitted controls" lists variables in the model specification, whose coefficients were not statistically significant and which are not shown in this table.

Continuing our investigation of the relationship between e-commerce costs and market structure, we turn to our full sample, detailed by NACE Rev 2. sectors. We start with the specification we ended the ec-extra sample with. Columns (3) and (4) of table 2.3 show the basic version with all variables and a version, where highly insignificant variables were dropped. As in the previous sample, for the most factors, which, from earlier steps, can directly be attributed to e-commerce (e.g., online shopping and postal services), there is evidence for an hump-shaped relationship, as predicted by hypotheses 2.1 and 2.3. This is not the case for some institutional and sectoral variables, such as customs burden or the foreign ownership. At the same time, as discussed above, these variables might also have a direct effect on concentration, and, therefore, it cannot be differentiated whether this result is the "wrong" effect of e-commerce or the direct effect of these variables on market structure.

We also experimented with classifying industries into those with high, mid and low e-commerce costs. We used two approaches for such classification, summarized in table C.4 in appendix C, and both yielded very similar results. First, we relied on the stylized fact 2.2.2 to single out a few sectors that can be plausibly referred to as high-cost or low-cost ones for e-commerce. The top two high-cost sectors are then food products, including subcategories, and furniture, while the low-cost sectors are media products (books, video, games, etc.) and toys. Second, we created an sector-based index of e-commerce cost, based on product perishability (food in MIG classification) and complexity (SEC classsification), bulkiness and percentage of online-shoppers buying certain products. The cost index shows two distinct breaks (see also figure C.1 in appendix C), which were then used to identify high- and low-cost sectors. We, however, consider this classification less plausible in terms of low-cost sectors, as the sectors related to recorded media (books, video, etc.), typically considered as good examples of e-commerce, are classified as mid-cost sectors, while sectors such as computers and consumer electronics, for which there is much less consensus, are instead put in the low-cost group.[22] At the same time, the classification into high-cost sectors (food, beverages and tobacco products) seems plausible.

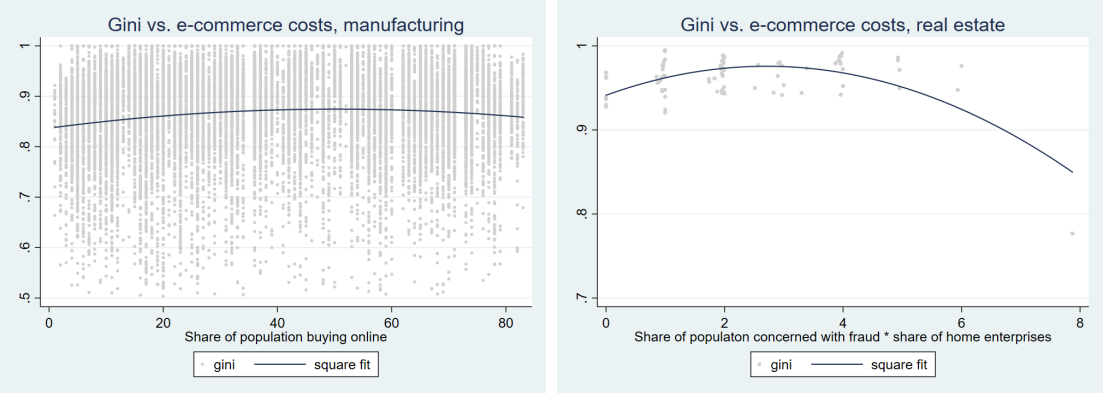
In both classifications, we created indicator variables for the low- and high-cost sectors (1 for low-/high-cost sector respectively, 0 otherwise). We then included the interactions of the main e-commerce cost proxies (online shopping behavior and postal services) with these indicator variables. The results (columns (5) and (6) of table 2.3) indicate a negative relationship between e-commerce costs and market concentration in high-cost industries, which is in line with hypothesis 2.1.[23] For low-cost sectors, the evidence is less clear. For the first cost classification, the significant coefficients point in the direction of a positive relationship between e-commerce costs and market concentration (as postulated by hypothesis 2.3). For the second classification, however, the relationship seems to be an hump-shaped one, while this shape is reversed for the e-commerce cost term (indicating mid-cost sectors). This inconclusive result might also be caused by the issues of the second classification with regard to low- and mid-cost sectors.

The large size of the sample and inclusion of very disaggregated sectors also allow us to analyze the relationship between e-commerce and market structure for different sector groups. Some interesting patterns are revealed here. For the sake of brevity, we present

[22] Consider stylized fact 2.2.2 and see Mityko 2012 on differences in valuation of electronic products.

[23] It is important to remember here that share of population buying online and number of post offices are assumed to be negatively related to e-commerce costs.

Figure 2.7: Market concentration and e-commerce costs: manufacturing and real estate



the technical results in appendix D.5 and only highlight a few points here. An example for one cost driver is also illustrated for manufacturing and real estate in figure 2.7.

Firstly, also here the model performs substantially better for the manufacturing subsample than for trade or services. At the same time, for the non-consumer sector of mining and quarrying, e-commerce-related variables have no statistical significance.

Secondly, for trade (retail and wholesale), we could not find any specification, which would deliver a reasonable explanatory power. Most of e-commerce cost proxies remain statistically insignificant. While surprising at a first glance, one has to consider that, for the most part, retail and wholesale companies are those buying from producers and reselling their products. On the one hand, they can also use e-commerce as a further sales channel. On the other hand, with growth of e-commerce, they stand in competition with producers' direct online sales and are more than ever just another sales channel for the producers. As such, trade firms might just keep an online shop "because everybody does it today" and further put most their efforts into the physical shops, as the major traditional sales channel for producers, independent of how the costs of e-commerce develop.

The third observation of the sectoral analysis is that there is also much variation in the explanatory power among services, especially if specifications are adjusted on a sectoral basis to best fit the data. For example, for the real estate activities, more than a third of gini variation can be explained by online shoppers' concerns about fraud and high prices, interacted with the share of home enterprises and complemented only with the number of post offices and inflation. This finding is very indicative of the nature of the service.

We also use the full sample to briefly look into hypothesis 2.2, namely the negative relationship between e-commerce costs and firm productivity and size. In terms of firm size, we look at turnover per firm, and we proxy firm productivity with apparent labor productivity. The model fit is relatively low both for enterprise turnover and (even more so) for labor productivity (see table D.19 in appendix D.5). At the same time, we find some interesting effects in case of labor productivity. For small-medium enterprises (10-19 employees), we find an hump-shaped relationship for a number of explanatory variables, which supports the finding of the ec-extra sample. The notable exception, lack

of post coverage, shows a u-shape with an extreme point very close to the lower bound (0.015 against the values interval 0 to 5). In other words, for the majority of the value range, better post coverage allows for lower productivity, and only at very high coverage levels do small enterprises suffer an inverse impact – potentially, due to costs being so low that even more e-commerce adopters increase competition. Comparing this result to the turnover of small enterprises, there is little correlation, and for turnover, most variables remain insignificant. Another interesting finding with regard to productivity is that, for medium enterprises, there seems to be a u-shaped relation to e-commerce costs. This would support the idea of section 2.4.2 that medium firms only profit from a certain decrease in e-commerce costs, which potentially allows them to replace the (unprofitable) traditional channel. After costs fall enough, many smaller enterprises can enter the market, which raises competition and forces the less productive of the medium firms to lose market share (and maybe even shrink in size to eventually fall in the smaller size category). It would indeed be interesting to investigate, for these enterprise classes, how much of the firm's turnover is covered by online and traditional channels, but the available data, unfortunately, are not detailed enough to allow for such an investigation.

2.7 Conclusion

This chapter started with the presentation of several stylized facts about e-commerce and illustrated the macroeconomic relevance of e-commerce due to pure volume, but also because of the similarities in the patterns of variation in e-commerce usage by firms to those known from international trade. It also showed how e-commerce markets are characterized by love for variety and multi-channel marketing. Building upon these concepts, the chapter formulates a theoretical extension to the well-known Melitz framework that accommodates the effects of e-commerce. To our knowledge, this is the first work to depart from the narrow focus on buyer-seller matching and to consider e-commerce as both a substitute and a complement to the traditional market channels.

We show that the emergence of e-commerce is in many aspects similar to trade liberalization. E-commerce will increase competition, as it will allow some firms to compete for more market shares via (additional) e-commerce channels. Traditional business models, or channels, will, as a result, inevitably lose market shares and profits. Therefore, only some firms will benefit from the new opportunity of e-commerce. This pattern is quite similar to the one caused by trade liberalization, where export opportunities for some firms go hand in hand with import competition for all firms. This pattern is well established in the theoretical and empirical literature on trade liberalization. Our data do not allow for testing this implication for e-commerce, as no differentiation between traditional and e-commerce firms is possible on the sectoral level. An empirical investigation based on firm-level data would be a logical next step in research on e-commerce.

We further show that e-commerce also has non-linear effects, which are different from those of trade liberalization. The twin-varieties extension proposed here provides a more complex insight into the effects of e-commerce, as it reflects that firms can *optionally* apply e-commerce, *either as a additional channel or an alternative technology.* This optionality means that, depending on parameter constellations, there are up to three

types of firms applying e-commerce. Whenever e-commerce is comparatively unattractive for costumers and/or associated with higher costs, only the biggest firms will establish e-commerce as additional sales channels (multi-channel marketing). Whenever e-commerce is comparatively more attractive and/or less costly, there will be medium-sized firms switching from traditional business concepts to e-commerce business concepts, additionally to the large multi-channel firms. If the comparative advantage of the e-commerce technology is large, additional small e-commerce firms will enter the market on top of the multi-channel firms and the switching firms. Thus, the effect of e-commerce on average sector productivity can be positive, as in the case of trade liberalization (in the first two scenarios), but it can also be negative, if low costs of e-commerce lead to a fragmentation of industrial structures. The same holds for market concentration.

The data on the European countries lend support to the existence of the high-cost and low-cost scenarios. Relating the first scenario to the expansion of large firms and rising market concentration and the third scenario to the market entry of small firms and decreasing market concentration, we indeed find that the relationship between market concentration and e-commerce adoption and costs tends to follow an hump-shape. The empirical results are not as conclusive with regard to labor productivity. Here, we have to acknowledge, however, that the data at hand are not ideally suited for investigating the hypotheses on productivity or size of firms using e-commerce versus those using traditional channels.

Therefore, the theoretical framework proposed in this chapter provides a fruitful basis for further research. From the theoretic perspective, the Melitz framework and the TVO extension abstract from any market failure. Investigation of, e.g., macroeconomic externalities or information asymmetries in the context of e-commerce are interesting fields for further research. As there are extensions to the Melitz model that introduce Heckscher-Ohlin comparative advantage in the context of heterogeneous firms (Bernard, Redding, et al. 2007), it is also promising to further extend it to investigate the effect of the three scenarios outlined in this chapter on different factor owners. On a more fundamental level, a further important question is that of market power within the "Melitz branch" of literature. In relation to e-commerce, especially questions surrounding strategically dominating online-platforms, such as Amazon, might require a different economic analysis and potentially a different technical approach than CES utility and open monopolistic competition.

Last but not least, e-commerce requires much more macroeconomic empirical research than has already been done. The bridge between e-commerce and trade theory literature – the Melitz model and TVO extension in particular – can further help empirical researchers as inspiration or practical starting point to structure their investigations. Several steps have already been taken, and in particular, in this chapter, but even more detailed, rich data need to be collected to test the implications of our model, and this is a task for future research.

Appendix

C Data description

Table C.1: Description of the samples used

Sample	List of sectors / sector groups & data aggregation		Basic period
ec	Sectors:	C10 to C18, C10 to C33, C19 to C22, C19 to C23, C23 to C25, C24 to C25, C26 to C33, D35 to E39, F41 to F43, G45 to G47, G47, H49 to H53, I55, I56, J58 to J63, L68, M69 to M74, N77 to N82	2005-2017 (e-commerce: 2009-2017)
	Data aggregation:	Scores & indices: weighted by the sector size (production). Absolute values: simple sum. Rates derived from absolute values: recalculation from aggregated variables.	
ec-extra	Sectors:	C10 to C18, C10 to C33, C19 to C22, C19 to C23, C19 to C25, C19 to C33, C23, C23 to C25, C24 to C25, C26 to C33, D35 to E39, F41 to F43, G45 to G46, G45 to G47, G47, H49 to H53, I55, I56, J58 to J63, L68, M69 to M74, N77 to N82 (Extrapolation of the ec-sample)	2005-2017 (e-commerce: 2009-2017)
	Data aggregation:	Scores & indices: weighted by the sector size (production). Absolute values: simple sum. Rates derived from absolute values: recalculation from aggregated variables.	
full	Sectors:	All NACE Rev. 2 sectors from B to S up to the 4th digit (e.g. B: B05, B051, B052, B06, B061, B062, B07, B071, B072, B08, B081, B089, B0891, B0892, B0893, B0899, B09, B091, B099)	2005-2017
	Data aggregation:	None	

Table C.2: List of the countries in the samples

EU Code	Country	EU Code	Country	EU Code	Country
AT	Austria	FR	France	NL	Netherlands
BE	Belgium	HR	Croatia	NO	Norway
BG	Bulgaria	HU	Hungary	PL	Poland
CH	Switzerland	IE	Ireland	PT	Portugal
CY	Cyprus	IS	Iceland	RO	Romania
CZ	Czech Republic	IT	Italy	RS	Serbia
DE	Germany	LI	Liechtenstein	SE	Sweden
DK	Denmark	LT	Lithuania	SI	Slovenia
EE	Estonia	LU	Luxembourg	SK	Slovakia
EL	Greece	LV	Latvia	TR	Turkey
ES	Spain	MK	North Macedonia	UK	United Kingdom
FI	Finland	MT	Malta		

Table C.3: Goods in the SEC classification

Product	**SEC classification**	**Source**
Clothing	Experience-1	Girard, Korgaonkar, et al. 2003
Furniture	Search	Siegel and Vitaliano 2007
Footwear	Experience-1	Mityko 2012
Carpets	Search	Siegel and Vitaliano 2007
Mattresses	Experience-1	Girard and Dion 2010
Perfumes	Experience-1	Girard, Korgaonkar, et al. 2003, Girard and Dion 2010
Music/video	Search	Girard and Dion 2010, Figueiredo 2000
Health/beauty	Experience-1	Girard and Dion 2010, Mityko 2012
Cigarettes	Experience-1	Siegel and Vitaliano 2007
Food	Experience-1	Mityko 2012
Cleaners/detergents	Experience-1	Siegel and Vitaliano 2007
Newspapers	Search	Siegel and Vitaliano 2007
Office supplies	Search	Girard and Dion 2010, Kiang et al. 2011
Housing	Experience-2	Girard and Dion 2010
Automobiles	Experience-1	Girard and Dion 2010
Appliances	Experience-2	Girard, Korgaonkar, et al. 2003, Girard and Dion 2010, Figueiredo 2000
Hardware	Search	Kiang et al. 2011
Drugs	Credence	Girard, Korgaonkar, et al. 2003, Kiang et al. 2011, Mityko 2012
Glasses	Experience-2	Siegel and Vitaliano 2007

Table C.3 – continued from previous page

Product	SEC classification	Source
Software	Experience-2	Siegel and Vitaliano 2007
Books	Search	Girard, Korgaonkar, et al. 2003, Mityko 2012
Sporting goods	Search	Kiang et al. 2011
Toys	Search	Figueiredo 2000
Advertising	Experience-1	Siegel and Vitaliano 2007
Transportation	Search	Girard and Dion 2010, Mityko 2012
Vacations	Experience-1	Siegel and Vitaliano 2007
Education	Credence	Siegel and Vitaliano 2007
Training	Experience-1	Siegel and Vitaliano 2007
Tours	Experience-1	Siegel and Vitaliano 2007
Banking	Experience-2	Siegel and Vitaliano 2007
Car rentals	Experience-1	Siegel and Vitaliano 2007
Entertainment	Experience-1	Siegel and Vitaliano 2007
Real estate	Experience-1	Siegel and Vitaliano 2007
Cargo	Experience-1	Siegel and Vitaliano 2007
Job placement	Experience-1	Siegel and Vitaliano 2007
Nursing homes	Experience-1	Siegel and Vitaliano 2007
Sports clubs	Experience-1	Siegel and Vitaliano 2007
Hotels	Experience-1	Siegel and Vitaliano 2007
Waste collection	Experience-1	Siegel and Vitaliano 2007
Landscaping	Experience-1	Siegel and Vitaliano 2007
Investments	Credence	Siegel and Vitaliano 2007
Trusts	Credence	Siegel and Vitaliano 2007
Portfolio Management	Credence	Siegel and Vitaliano 2007
Mutual funds	Credence	Siegel and Vitaliano 2007
Insurance	Credence	Siegel and Vitaliano 2007
Health care	Credence	Siegel and Vitaliano 2007
Weight control	Credence	Siegel and Vitaliano 2007
Car repairs	Credence	Siegel and Vitaliano 2007

Notes: Distinction of experience goods is as follows: Experience-1: quality cannot be known until sampling/use; Experience-2: quality is harder/costlier to estimate than sampling/use

Table C.4: Classifications of NACE Rev. 2 sectors to low- and high-cost sectors

Classification	Low-cost sectors	High-cost sectors	Selection criteria
Consumer preference (stylized fact 2.2.2)	C18 printing & reproduction of recorded media C182 reproduction of recorded media C324 games & toys G476 retail sale of cultural & recreation goods J581 publishing of books, periodicals J582 software publishing J59 motion picture, video, television, sound, music J591 motion picture, video, television J592 sound & music recording	C10 food products C101 meat C102 fish C103 fruit & vegetables C105 dairy G472 retail sale of food C31 furniture	Low-cost [*Lowcost SF*]: preference towards online purchase as in PwC 2017 High-cost [*Highcost SF*]: preference towards in-store purchase as in PwC 2017; delivery considerations: food as perishable, furniture as bulky product
E-commerce cost index *ECcost*	C262 computers & peripheral equipment C264 consumer electronics C323 sports goods	C10 food products C11 beverages C12 tobacco products	Low-cost [*Lowcost*]: cost index below 0.28 High-cost [*Highcost*]: cost index above 0.6

Table C.5: Description of variables

Variable name	Meaning	Detalization	Based on database variable	Source
Main dependent variables (market structure) and their supporting variables				
Gini	Gini coefficient (based on firm turnover)	Country, sector		Calculated from number of enterprises, enterprise turnover
Enterprise number	Number of enterprises (per sector / size class)	Country, sector, size class	V11110	Business statistics (Eurostat)
Turnover	Total turnover in sector (by size class), mio EUR	Country, sector, size class	V12110	ibid.
Labor productivity	Apparent labour productivity (value added per person employed) in sector (by size class), thd EUR	Country, sector, size class	V91110	ibid.
Employees per enterprise	Number of persons employed per enterprise in sector (by size class)	Country, sector, size class	V92100	ibid.
Enterprise turnover	Turnover per enterprise in sector (by size class), mio EUR	Country, sector, size class	V11110, V12110	ibid.
Turnover group share	Share of the size class in sector turnover	Country, sector, size class	V12110	ibid.
Turnover per employee	Turnover per person employed in sector (by size class), thd EUR	Country, sector, size class	V91100	ibid.
Value added	Total value added in sector (by size class), mio EUR	Country, sector, size class	V12150	ibid.
Employees	Number of persons employed in sector (by size class)	Country, sector, size class	V16110	ibid.
E-commerce in enterprises				
Digital sales	Enterprises having received orders via computer mediated networks, % of enterprises	Country, sector	E_AESELL	Digital Economy and Society (Eurostat)
E-sales	Enterprises selling online (at least 1% of turnover), % of enterprises	Country, sector	E_ESELL	ibid.
Web-sales	Enterprises having received orders via a website or apps (web sales), % of enterprises	Country, sector	E_AWSELL	ibid.
Web-sales b2bg	Enterprises which sold via a website or apps - to other enterprises or the government (B2B & B2G), % of enterprises	Country, sector	E_AWS_B2BG	ibid.
Web sales b2c	Enterprises which sold via a website or apps to consumers (B2C), % of enterprises	Country, sector	E_AWS_B2C	ibid.
Web sales b2c 10+	Enterprises where B2C web sales are 10% or more of the web sales, % of enterprises	Country, sector	E_AWSVAL_ B2C_GE10WS	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Web sales b2c 1+	Enterprises where B2C web sales are more than 1% of the web sales, % of enterprises	Country, sector	E_AWS_ B2C_GT1WS	ibid.
Web sales 1+/ 10+ws	Enterprises where web sales are more than 1% of total turnover and B2C web sales more than 10% of the web sales, % of enterprises	Country, sector	E_AWS_GT1_ B2C_GT10WS	ibid.
Web sales home	Enterprises with web sales to the own country, % of enterprises	Country, sector	E_AWSHM	ibid.
Web sales EU-foreign	Enterprises with web sales to other EU countries, % of enterprises	Country, sector	E_AWSEU	ibid.
Web sales RoW	Enterprises with web sales to the rest of the world, % of enterprises	Country, sector	E_AWSWW	ibid.
E-sales home	Enterprises having done electronic sales to the own country, % of enterprises	Country, sector	E_AESHM	ibid.
E-sales EU-foreign	Enterprises having done electronic sales to other EU countries, % of enterprises	Country, sector	E_AESEU	ibid.
E-sales RoW	Enterprises having done electronic sales to the rest of the world, % of enterprises	Country, sector	E_AESWW	ibid.
E-sales EU-RoW	Enterprises having done electronic sales to other EU countries and the rest of the world, % of enterprises	Country, sector	E_AE-SEUWW	ibid.
E-sales Webpay	Enterprises accepting online payment for sales via website, % of enterprises	Country, sector	E_AES-PAYON	ibid.
E-sales b2c-marketplace	Enterprises which sold via a website or apps - B2C and via an e-commerce marketplace, % of enterprises	Country, sector	E_AWS_ B2C_CMP	ibid.
E-sales b2c 10+ marketplace	Enterprises where B2C web sales are 10% or more of the total web sales and which sold via an e-commerce marketplace, % of enterprises	Country, sector	E_AWSVAL_ B2C_ GE10WS_CMP	ibid.
Web sales ownapp	Enterprises which sold via a website or apps - via their own website or apps, % of enterprises	Country, sector	E_AWS_ COWN	ibid.
Web sales marketplace	Enterprises which sold via a website or apps - via an e-commerce marketplace, % of enterprises	Country, sector	E_AWS_CMP	ibid.
Web sales 20+ marketplace	Enterprises which sold via a website or apps - via an e-commerce marketplace for at least 20% of the web sales, % of enterprises	Country, sector	E_AWS_ CMP_GE20	ibid.
E-sales turnover home	Enterprises' turnover from electronic sales to own country, % of total turnover	Country, sector	E_AESVHM	ibid.
E-sales turnover EU	Enterprises' turnover from electronic sales to other EU countries, % of total turnover	Country, sector	E_AESVEU	ibid.
E-sales turnover RoW	Enterprises' turnover from electronic sales to the rest of the world, % of total turnover	Country, sector	E_AESVWW	ibid.
E-sales turnover	Enterprises' total turnover from e-commerce, % of total turnover	Country, sector	E_ETURN	ibid.
Web sales turnover ownapp	Enterprises' turnover from web sales via own websites or apps, % of total turnover	Country, sector	E_AWSVAL_ COWN	ibid.
Web sales turnover marketplace	Enterprises' turnover from web sales via e-commerce marketplaces, % of total turnover	Country, sector	E_AWSVAL_ CMP	ibid.
Web sales turnover	Enterprises' turnover from web sales, % of total turnover / turnover from e-commerce (ec)	Country, sector	E_AWSVAL	ibid.
Web sales turnover b2c	Enterprises' turnover from web sales - B2C, % of total turnover / turnover from e-commerce (ec) / web sales (ws)	Country, sector	E_AWSVAL_ B2C	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Web sales turnover b2bg	Enterprises' turnover from web sales - B2B and B2G, % of total turnover / turnover from e-commerce (ec) / web sales (ws)	Country, sector	E_AWSVAL_B2BG	ibid.
Web sales turnover 1+/ b2c10+	Web sales of the enterprises where these are more than 1% of total turnover and B2C web sales more than 10% of the web sales, % of total turnover / turnover from e-commerce (ec) / web sales (ws)	Country, sector	E_AWSVAL_GT1_B2C_GT10WS	ibid.
Web sales problem: fit	Obstacles to selling online: The enterprise's goods or services are not suitable - enterprises selling via website, % of enterprises / enterprises doing web sales (ws)	Country, sector	E_AWS_OSUIT	ibid.
Web sales problem: logistic	Obstacles to selling online: Problems related to logistics (shipping of goods or delivery of services) - enterprises selling via website, % of enterprises / enterprises doing web sales (ws)	Country, sector	E_AWS_OLOG	ibid.
Web sales problem: pay	Obstacles to selling online: Problems related to payments - enterprises selling via website, % of enterprises / enterprises doing web sales (ws)	Country, sector	E_AWS_OPAY	ibid.
Web sales problem: security	Obstacles to selling online: Problems related to ICT security or data protection - enterprises selling via website, % of enterprises / enterprises doing web sales (ws)	Country, sector	E_AWS_OSEC	ibid.
Web sales problem: legal	Obstacles to selling online: Problems related to the legal framework - enterprises selling via website, % of enterprises / enterprises doing web sales (ws)	Country, sector	E_AWS_OLF	ibid.
Web sales problem: cost	Obstacles to selling online: The costs of introducing web sales too high compared to the benefits - enterprises selling via website, % of enterprises / enterprises doing web sales (ws)	Country, sector	E_AWS_OCOST	ibid.
No web sales problem: fit	Obstacles to selling online: The enterprise's goods or services are not suitable - enterprises not selling via website, % of enterprises / enterprises not doing web sales (nws)	Country, sector	E_AWSX_OSUIT	ibid.
No web sales problem: logistic	Obstacles to selling online: Problems related to logistics (shipping of goods or delivery of services) - enterprises not selling via website, % of enterprises / enterprises not doing web sales (nws)	Country, sector	E_AWSX_OLOG	ibid.
No web sales problem: pay	Obstacles to selling online: Problems related to payments - enterprises not selling via website, % of enterprises / enterprises not doing web sales (nws)	Country, sector	E_AWSX_OPAY	ibid.
No web sales problem: security	Obstacles to selling online: Problems related to ICT security or data protection - enterprises not selling via website, % of enterprises / enterprises not doing web sales (nws)	Country, sector	E_AWSX_OSEC	ibid.
No web sales problem: legal	Obstacles to selling online: Problems related to the legal framework - enterprises not selling via website, % of enterprises / enterprises not doing web sales (nws)	Country, sector	E_AWSX_OLF	ibid.
No web sales problem: cost	Obstacles to selling online: The costs of introducing web sales too high compared to the benefits - enterprises not selling via website, % of enterprises / enterprises not doing web sales (nws)	Country, sector	E_AWSX_OCOST	ibid.
	Consumers' ICT usage and online shopping behavior			
Internet broad	Household internet connection type: broadband, % of households (by household type)	Country	H_BROAD	ICT in households (Eurostat)
Internet fbroad	Household internet connection type: fixed broadband, % of households (by household type)	Country	H_BBFIX	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Internet mbroad	Household internet connection type: mobile broadband, % of households (by household type)	Country	H_BBMOB	ibid.
Internet dsl	Household internet connection type: DSL, % of households (by household type)	Country	H_DSL	ibid.
Internet modem	Household internet connection type: modem or ISDN, % of households (by household type)	Country	H_DIALUP	ibid.
Internet dialup	Household internet connection type: dial-up access via normal telephone line or ISDN, % of households (by household type)	Country	H_DIALUP1	ibid.
Internet mobile	Household internet connection type: mobile phone over narrowband, % of households (by household type)	Country	H_MPHNAR	ibid.
E-shopping hotel	Online purchases: holiday accommodation, % of individuals	Country	I_BHOLAC	ibid.
E-shopping other travel	Online purchases: other travel arrangements (transport tickets, car hire, etc.), % of individuals	Country	I_BOTA	ibid.
E-shoppers 3mon	Last online purchase: in the last 3 months, % of individuals	Country	I_BUY3	ibid.
E-shopping 3-12mon	Last online purchase: between 3 and 12 months ago, % of individuals	Country	I_B3_12	ibid.
E-shoppers	Last online purchase: in the 12 months, % of individuals	Country	I_BLT12	ibid.
E-shoppers over 12mon	Last online purchase: more than a year ago, % of individuals	Country	I_BUMT12	ibid.
No e-shopping/ over 12mon	Individuals who ordered goods or services, over the Internet, for private use, more than a year ago or have never ordered, % of individuals	Country	I_BUMT12X	ibid.
No e-shopping	Individuals who never ordered goods or services, over the Internet, for private use, % of individuals	Country	I_BUX	ibid.
E-shopping food	Online purchases: food/groceries, % of individuals	Country	I_BFOOD	ibid.
E-shopping household	Online purchases: household goods, % of individuals	Country	I_BFURN	ibid.
E-shopping film/ music	Online purchases: films/music, % of individuals	Country	I_BFILM	ibid.
E-shopping e-learning	Online purchases: e-learning material, % of individuals	Country	I_BELRN	ibid.
E-shopping book/ magazine	Online purchases: books/ magazines/ e-learning material, % of individuals	Country	I_BBOOK	ibid.
E-shopping book-news	Online purchases: books/magazines/newspapers, % of individuals	Country	I_BBOOKNL	ibid.
E-shopping clothes	Online purchases: clothes, sports goods, % of individuals	Country	I_BCLOT	ibid.
E-shopping soft	Online purchases: computer software, % of individuals	Country	I_BSOFT	ibid.
E-shopping hardware	Online purchases: computer hardware, % of individuals	Country	I_BHARD	ibid.
E-shopping electro	Online purchases: electronic equipment, % of individuals	Country	I_BEEQU	ibid.
E-shopping finance	Online purchases: shares/insurance/other financial services, % of individuals	Country	I_BFIN	ibid.
E-shopping event	Online purchases: tickets for events, % of individuals	Country	I_BTICK	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
E-shopping travel	Online purchases: travel and holiday accommodation, % of individuals	Country	I_BHOLS	ibid.
E-shopping med	Online purchases: medecine, % of individuals	Country	I_BMED	ibid.
E-shopping telecom	Online purchases: telecom services, % of individuals	Country	I_BTS	ibid.
E-shopping e-film/music	Online purchases: films/music, delivered or upgraded online, % of individuals	Country	I_BFILMO	ibid.
E-shopping e-book	Online purchases, downloaded or accessed from websites or apps: e-books, % of individuals	Country	I_BE-BOOKO	ibid.
E-shopping e-book/news	Online purchases, downloaded or accessed from websites or apps: e-magazines, e-newspapers, % of individuals	Country	I_BMGNWO	ibid.
E-shopping game/soft	Online purchases: video games software and upgrades, % of individuals	Country	I_BGSOFT	ibid.
E-shopping soft-nogame	Online purchases: computer software other than video games and upgrades, % of individuals	Country	I_BOSOFT	ibid.
E-shopping product	Online purchases by product, in product categories listed for variables above, % of individuals	Country	n.a.	Compiled from e-shopping variables
E-shopping domestic	Online purchases: from national sellers, % of individuals	Country	I_BFDOM	ICT in households (Eurostat)
E-shopping EU-foreign	Online purchases: from sellers from other EU countries, % of individuals	Country	I_BFEU	ibid.
E-shopping foreign	Online purchases: from sellers abroad (other EU or non EU countries), % of individuals	Country	I_BFFOR	ibid.
E-shopping RoW	Online purchases: from sellers from the rest of the world (non-EU), % of individuals	Country	I_BFWRLD	ibid.
E-shopping no origin	Online purchases: from sellers with unknown country of origin, % of individuals	Country	I_BFUNK	ibid.
E-shopping foreign physical	Online purchases from sellers abroad: physical goods (e.g. electronics, clothes, toys, food, groceries, books, CDs/DVDs), % of individuals	Country	I_BFFOR_PGD	ibid.
E-shopping foreign e-product	Online purchases from sellers abroad: products downloaded or accessed from websites or apps (e.g. films, music, e-books, e-newspapers, games), % of individuals	Country	I_BFFOR_DWL	ibid.
E-shopping foreign travel	Online purchases from sellers abroad: travel, accommodation or holiday arrangements (e.g. tickets and documents by mail or printed by oneself), % of individuals	Country	I_BFFOR_TRH	ibid.
E-shopping foreign other	Online purchases from sellers abroad: other services (e.g. tickets for events received by mail, telecom subscriptions), % of individuals	Country	I_BFFOR_OSV	ibid.
E-shopping rare	Frequency of online purchases in the last 3 months: 1 or 2 times, % of individuals	Country	I_BF_1_2	ibid.
E-shopping low-mid frequent	Frequency of online purchases in the last 3 months: 3 to 5 times, % of individuals	Country	I_BF_3-5	ibid.
E-shopping high-mid frequent	Frequency of online purchases in the last 3 months: 6 to 10 times, % of individuals	Country	I_BF_6-10	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
E-shopping often10	Frequency of online purchases in the last 3 months: more than 10 times, % of individuals	Country	I_BF_GT10	ibid.
E-shopping often6	Frequency of online purchases in the last 3 months: 6 times or more, % of individuals	Country	I_BF_HI	ibid.
E-shopping lowcost	Online purchases in the last 3 months for less than 50 euro, % of individuals	Country	I_IBV_LT50	ibid.
E-shopping cost 50-99eur	Online purchases in the last 3 months for between 50 and 99 euro, % of individuals	Country	I_IBV_ 50-99	ibid.
E-shopping high-cost 100	Online purchases in the last 3 months for 100 euro or more, % of individuals	Country	I_IBV_HI	ibid.
E-shopping cost 100-499eur	Online purchases in the last 3 months for between 100 and 499 euro, % of individuals	Country	I_IBV_ 100-499	ibid.
E-shopping cost 500-999eur	Online purchases in the last 3 months for between 500 and 999 euro, % of individuals	Country	I_IBV_ 500-999	ibid.
E-shopping highcost 1000	Online purchases in the last 3 months for 1000 euro or more, % of individuals	Country	I_IBV_GE1000	ibid.
Computer use 12mon	Last computer use: within last 12 months, % of individuals	Country	I_CLT12	ibid.
Computer use 3mon	Last computer use: within last 3 months, % of individuals	Country	I_C3	ibid.
Computer use 3-12mon	Last computer use: between 3 and 12 months ago, % of individuals	Country	I_C3_12	ibid.
Computer use over12	Individuals who used a computer more than a year ago, % of individuals	Country	I_CUMT12	ibid.
Computer use never	Computer use: never, % of individuals	Country	I_CUX	ibid.
Computer use	Individuals who have ever used a computer, % of individuals	Country	I_CEVR	ibid.
Internet access	Households with access to internet, % of households (by household type)	Country		ibid.
E-shopping problem: long delivery	Individuals who encountered the following problem when making purchases over the Internet: Speed of delivery longer than indicated, % of individuals / % of individuals having ordered online in the past year	Country	I_BSPD	ibid.
E-shopping problem: high price	Individuals who encountered the following problem when making purchases over the Internet: Delivery costs or final price higher than indicated, % of individuals / % of individuals having ordered online in the past year	Country	I_BCPR	ibid.
E-shopping problem: wrong good	Individuals who encountered the following problem when buying/ordering over the Internet: wrong or damaged good/services delivered, % of individuals / % of individuals having ordered online in the past year	Country	I_BWDN	ibid.
E-shopping problem: security	Individuals who encountered the following problem when making purchases over the Internet: Lack of security of payments, % of individuals / % of individuals having ordered online in the past year	Country	I_BSEC	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
E-shopping problem: fraud	Individuals who encountered the following problem when buying/ordering over the Internet: problems with fraud, % of individuals / % of individuals having ordered online in the past year	Country	I_BFRA	ibid.
E-shopping problem: complaint	Individuals who encountered the following problem when making purchases over the Internet: Complaints and redress were difficult or no satisfactory, % of individuals / % of individuals having ordered online in the past year	Country	I_BCR	ibid.
E-shopping problem: other	Individuals who encountered the following problem when making purchases over the Internet: Other, % of individuals / % of individuals having ordered online in the past year	Country	I_BOTH	ibid.
E-shopping problem: technical	Individuals who encountered the following problem when buying/ordering over the Internet: technical failure, % of individuals / % of individuals having ordered online in the past year	Country	I_BTFW	ibid.
E-shopping problem: info	Individuals who encountered the following problem when buying/ordering over the Internet: difficulties finding information concerning guarantees, etc., % of individuals / % of individuals having ordered online in the past year	Country	I_BDGL	ibid.
E-shopping problems	Individuals who encountered problems when buying/ordering goods or services over the internet for private use, % of individuals / % of individuals having ordered online in the past year	Country	I_BARR1Y	ibid.
E-shopping no problems	Individuals who did not encounter problems when buying/ordering goods or services over the internet for private use, % of individuals / % of individuals having ordered online in the past year	Country	I_BARR1X	ibid.
E-shopping problem: coverage	Individuals who encountered the following problem when buying/ordering over the internet: foreign retailer did not sell in my country, % of individuals / % of individuals having ordered online in the past year	Country	I_BDNS	ibid.
	Logistics and postal services			
Post collections urban	Average number of collections from boxes per working day in urban areas	Country	n.a.	Universal Postal Union
Post collections rural	Average number of collections from boxes per week in rural areas	Country	n.a.	ibid.
Letter boxes	Number of letter-boxes	Country	n.a.	ibid.
Deliveries urban	Average number of deliveries per working day in urban areas	Country	n.a.	ibid.
Deliveries rural	Average number of deliveries per week in rural areas	Country	n.a.	ibid.
Post office boxes	Number of post office boxes	Country	n.a.	ibid.
Delivery to boxes	Percentage of items delivered through post office boxes	Country	n.a.	ibid.
Home delivery	Percentage of the population having mail delivered at home	Country	n.a.	ibid.
Delivery to post	Percentage of the population having to collect mail from a postal establishment	Country	n.a.	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
No post	Percentage of the population without postal services	Country	n.a.	ibid.
Area	Country area, km^2	Country	n.a.	ibid.
Exchange rate	Rate of exchange, national currency to special drawing rights	Country	n.a.	ibid.
Express domestic	Number of express items, domestic service	Country	n.a.	ibid.
Express int'l dispatch	Number of express items, international service - dispatch	Country	n.a.	ibid.
Express int'l receipt	Number of express items, international service - receipt	Country	n.a.	ibid.
Parcels domestic	Number of parcels, domestic service	Country	n.a.	ibid.
Parcels int'l dispatch	Number of parcels, international service - dispatch	Country	n.a.	ibid.
Parcels int'l receipt	Number of parcels, international service - receipt	Country	n.a.	ibid.
Post revenue	Operating revenue, special drawing rights	Country	n.a.	ibid.
Post expenditure	Operating expenditure, special drawing rights	Country	n.a.	ibid.
Post result	Operating result, special drawing rights	Country	n.a.	ibid.
Post net result	Net result, special drawing rights	Country	n.a.	ibid.
Post letter revenue	Percentage of income linked to letter post	Country	n.a.	ibid.
Post parcel revenue	Percentage of income linked to parcels and logistics services	Country	n.a.	ibid.
Post financial revenue	Percentage of income linked to postal financial services	Country	n.a.	ibid.
Post other revenue	Percentage of income linked to other products	Country	n.a.	ibid.
Post phil. revenue	Income from philately as a percentage of total income	Country	n.a.	ibid.
Post permanent office	Total number of permanent post offices	Country	n.a.	ibid.
Post office administrative	Number of permanent offices staffed by administration officials	Country	n.a.	ibid.
Post office	Number of permanent offices staffed by people from outside the administration	Country	n.a.	ibid.
Post office area coverage	Average area covered by a permanent office, km^2	Country	n.a.	ibid.
Post office person coverage	Average number of inhabitants served by a permanent office	Country	n.a.	ibid.
Post mobile office	Number of mobile post offices (including rural delivery staff)	Country	n.a.	ibid.
Post financial office	Number of post offices (permanent and mobile) accepting financial transactions	Country	n.a.	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Post sorting centre	Number of sorting centres	Country	n.a.	ibid.
Post internet point	Number of post offices providing public Internet access points	Country	n.a.	ibid.
Post office network	Number of permanent post offices connected to an electronic network	Country	n.a.	ibid.
Post office auto	Number of permanent post offices using counter automation systems	Country	n.a.	ibid.
Post staff total	Total number of staff	Country	n.a.	ibid.
Post staff fulltime	Number of full-time staff	Country	n.a.	ibid.
Post staff parttime	Number of part-time staff	Country	n.a.	ibid.
Post positions	Number of posts (in full-time equivalent)	Country	n.a.	ibid.
Post female staff	Female employees as a percentage of total staff	Country	n.a.	ibid.
Post delivery staff	Percentage of delivery staff	Country	n.a.	ibid.
	Other sectoral data			
EC cost	Index of e-commerce cost: average of $SEC\ score$ (excluding Credence goods), $Perischable$, $Bulkiness$ & $E-shopping\ product$, each re-scaled linearly to the maximum of 1	Country, sector	n.a.	Own calculation
Production	Production value, mio EUR	Country, sector	V12120	Business statistics (Eurostat)
Labor productivity adj	Wage adjusted labour productivity (Apparent labour productivity by average personnel costs) - percentage	Country, sector	V91120	ibid.
Price index '10	Total output price index, national currency, index 2010 = 100	Country, sector	PRON	Country statistics (Eurostat)
Price index '15	Total output price index, national currency, index 2015 = 100	Country, sector	PRON	ibid.
Bulkiness	Goods bulkiness: ratio of exports value to exports volume (calculated as in Lee and Pak 2018 for internationally traded goods only)	Country, sector	n.a.	Foreign trade statistics (Eurostat)
MIG	MIG (Main Industrial Groupings) classification (food; consumer durable/non-durable; intermediate; investment goods)	Sector	n.a.	Eurostat
MIG score	MIG durability score, 0 to 4 (0 = food, 4 = investment goods)	Sector		Converted from MIG
SEC	Product classification in search-experience-credence (SEC) framework	Sector	n.a.	Various sources
SEC score	SEC classification score, 0 to 3 (0 = search products; 3 = credence products)	Sector		Converted from SEC
Perishable	Perishable goods (food): 1 if MIG score = 0, 0 otherwise	Sector	n.a.	Converted from MIG
R&D rate	R&D investment rate (total R&D investment to value added at factor cost)	Country, sector	n.a.	Calculated from Total R&D & total value added

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
R&D rate capital	R&D capital investment rate (capital R&D investment to value added at factor cost)	Country, sector	n.a.	Calculated from R&D capital & total value added
R&D rate current	R&D capital investment rate (capital R&D investment to value added at factor cost)	Country, sector	n.a.	Calculated from R&D current & total value added
R&D total	Total investment in research and development, mio EUR	Country, sector	n.a.	Business statistics (Eurostat)
R&D current	R&D investment - current investment, mio EUR	Country, sector	n.a.	ibid.
R&D labor	R&D investment - current labor cost, mio EUR	Country, sector	n.a.	ibid.
R&D current other	R&D investment - other current investment, mio EUR	Country, sector	n.a.	ibid.
R&D capital	R&D investment - capital investment, mio EUR	Country, sector	n.a.	ibid.
Concessions	Investment in intangibles: Gross investment in concessions, mio EUR	Country, sector	V15420	ibid.
Software	Investment in intangibles: Investment in purchased software, mio EUR	Country, sector	V15441	ibid.
Investment rate	Investment rate (gross investment in tangible goods to value added at factor cost)	Country, sector		Calculated from investment tangible & total value added
Investment tangible	Gross investment in tangible goods, mio EUR	Country, sector	V15110	Business statistics (Eurostat)
Investment in land	Gross investment in land, mio EUR	Country, sector	V15120	ibid.
Investment in buildings	Gross investment in existing buildings and structures, mio EUR	Country, sector	V15130	ibid.
Investment in construction	Gross investment in construction and alteration of buildings, mio EUR	Country, sector	V15140	ibid.
Investment in machinery	Gross investment in machinery and equipment, mio EUR	Country, sector	V15150	ibid.
Net investment tangible	Net investment in tangible goods, mio EUR	Country, sector	V15250	ibid.
Foreign firms	Number of enterprises under foreign control, total & offshore	Country, sector	V11110	ibid.
Turnover of foreign firms	Turnover or gross premiums written by enterprises under foreign control, total & offshore, mio EUR	Country, sector	V12110	ibid.
Production of foreign firms	Production value by enterprises under foreign control, total & offshore, mio EUR	Country, sector	V12120	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Value added of foreign firms	Value added at factor cost by enterprises under foreign control, total & offshore, mio EUR	Country, sector	V12150	ibid.
Investment by foreign firms	Gross investment in tangible goods by enterprises under foreign control, total & offshore, mio EUR	Country, sector	V15110	ibid.
Employment in foreign firms	Persons employed by enterprises under foreign control, total & offshore, number	Country, sector	V16110	ibid.
Firm share x	Share of enterprises under foreign control (total: x=world & offshore: x=off)	Country, sector		Calculated from Enterprise number & Foreign firms
Turnover share x	Share of turnover from enterprises under foreign control (total: x=world & offshore: x=off)	Country, sector		Calculated from Turnover & Tutnover of foreign firms
VA share x	Share of value added from enterprises under foreign control (total: x=world & offshore: x=off)	Country, sector		Calculated from Vallue added total & Value added of foreign firms
Investment share x	Share of gross investment in tangible goods enterprises under foreign control (total: x=world & offshore: x=off)	Country, sector		Calculated from Investment tabgible & Investment by foreign firms
Employment share x	Share of employment in enterprises under foreign control (total: x=world & offshore: x=off)	Country, sector		Calculated from Employees & Employment in foreign firms
	Country aggregates and institutional data			
Infrastructure length	Total length of waterways, roads and railways, km	Country	CNL, MWAY, RD_OTH, RIV, RL, RL_ELC, RL_TGE2	Country statistics (Eurostat)
Infrastructure density	Total length of waterways roads and railways per km2 of country area	Country		Calculated from Infrastructure length and Area
Wage EUR	Compensation per employee, EUR	Country	D1_SAL_PER	Country statistics (Eurostat)
Wage PPS	Compensation per employee, purchasing power standard	Country	D1_SAL_PER	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Wage per hour EUR	Compensation of employees per hour worked, EUR	Country	D1_SAL_HW	ibid.
Wage per hour PPS	Compensation of employees per hour worked, purchasing power standard	Country	D1_SAL_HW	ibid.
ULC per person	Index of nominal unit labour cost based on persons, 2010 = 100	Country	NULC_PER	ibid.
ULC per hour	Index of nominal unit labour cost based on hours worked, 2010 = 100	Country	NULC_HW	ibid.
Productivity per person	Index of real labour productivity per person, 2010 = 100	Country	RLPR_PER	ibid.
Productivity per hour	Index of real labour productivity per hour worked, 2010 = 100	Country	RLPR_HW	ibid.
GDP p.c.-CLV	GDP per capita, Chain linked volumes, index 2010 = 100	Country	B1GQ	ibid.
GDP p.c.-CLV EUR	GDP per capita, Chain linked volumes (2010), EUR	Country	B1GQ	ibid.
GDP p.c.-EUR	GDP per capita in current prices, EUR	Country	B1GQ	ibid.
GDP p.c.-PPS	GDP per capita in current prices, purchasing power standard	Country	B1GQ	ibid.
Consumption, HH -CLV	Household consumption expenditure per capita, Chain linked volumes, index 2010 = 100	Country	P31_S14	ibid.
Consumption, HH -CLV EUR	Household consumption expenditure per capita, Chain linked volumes (2010), EUR	Country	P31_S14	ibid.
Consumption, HH -EUR	Household consumption expenditure per capita in current prices, EUR	Country	P31_S14	ibid.
Consumption, HH -PPS	Household consumption expenditure per capita in current prices, purchasing power standard	Country	P31_S14	ibid.
Consumption, ind. -CLV	Actual individual consumption per capita, Chain linked volumes, index 2010 = 100	Country	P41	ibid.
Consumption, ind. -CLV EUR	Actual individual consumption per capita, Chain linked volumes (2010), EUR	Country	P41	ibid.
Consumption, ind. -EUR	Actual individual consumption per capita in current prices, EUR	Country	P41	ibid.
Consumption, ind. -PPS	Actual individual consumption per capita in current prices, purchasing power standard	Country	P41	ibid.
Employment NC	Total employment national concept, thd persons	Country	EMP_NC	ibid.
Employment DC	Total employment domestic concept, thd persons	Country	EMP_DC	ibid.
Population	Total population national concept, thd persons	Country	POP_NC	ibid.
Employees DC	Employees domestic concept, thd persons	Country	SAL_DC	ibid.
Employees NC	Employees national concept, thd persons	Country	SAL_NC	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Self-employed DC	Self-employed domestic concept, thd persons	Country	SELF_DC	ibid.
Self-employed NC	Self-employed national concept, thd persons	Country	SELF_NC	ibid.
Cash withdrawals	Number of cash withdrawals via customer terminals, mio transactions	Country	n.a.	European Central Bank
Card payments -Nr p.c.	Number of card payments per million inhabitants, mio transactions	Country	n.a.	ibid.
Card payments -value p.c.	Value of card payments per million inhabitants, mio EUR	Country	n.a.	ibid.
Credit rights	Strength of legal rights index measures the degree to which collateral and bankruptcy laws protect the rights of borrowers and lenders and thus facilitate lending. The index ranges from 0 to 12, with higher scores indicating that these laws are better designed to expand access to credit.	Country	IC.LGL. CRED.XQ	World Bank Development Indicators
Enforcement time	Time required to enforce a contract is the number of calendar days from the filing of the lawsuit in court until the final determination and, in appropriate cases, payment.	Country	IC.LGL.DURS	ibid.
Property reg. time	Time required to register property is the number of calendar days needed for businesses to secure rights to property.	Country	IC.PRP.DURS	ibid.
License time	Time required to obtain operating license is the average wait to obtain an operating license from the day the establishment applied for it to the day it was granted.	Country	IC.FRM.DURS	ibid.
Business start time	Time required to start a business is the number of calendar days needed to complete the procedures to legally operate a business. If a procedure can be speeded up at additional cost, the fastest procedure, independent of cost, is chosen.	Country	IC.REG.DURS	ibid.
Warehouse time	Time required to build a warehouse is the number of calendar days needed to complete the required procedures for building a warehouse. If a procedure can be speeded up at additional cost, the fastest procedure, independent of cost, is chosen.	Country	IC.WRH.DURS	ibid.
E-connection wait	The average wait, in days, experienced to obtain an electrical connection from the day an establishment applies for it to the day it receives the service.	Country	IC.ELC.DURS	ibid.
Warehouse procedures	Number of procedures to build a warehouse is the number of interactions of a company's employees or managers with external parties, including government agency staff, public inspectors, notaries, land registry and cadastre staff, and technical experts apart from architects and engineers.	Country	IC.WRH.PROC	ibid.
Property reg. procedures	Number of procedures to register property is the number of procedures required for a businesses to secure rights to property.	Country	IC.PRP.PROC	ibid.
Theft loss -sales	Average losses as a result of theft, robbery, vandalism or arson that occurred on the establishment?s premises, % of annual sales. The value represents the average losses for all firms which reported losses.	Country	IC.FRM. CRIM.ZS	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
LPI infrastructure	Logistics performance index: Quality of trade and transport-related infrastructure (1=low to 5=high)	Country	LP.LPI.INFR.XQ	ibid.
LPI overall	Logistics performance index: Overall (1=low to 5=high). Logistics Performance Index overall score reflects perceptions of a country's logistics based on efficiency of customs clearance process, quality of trade- and transport-related infrastructure, ease of arranging competitively priced shipments, quality of logistics services, ability to track and trace consignments, and frequency with which shipments reach the consignee within the scheduled time.	Country	LP.LPI.OVRL.XQ	ibid.
LPI shipment time	Logistics performance index: Frequency with which shipments reach consignee within scheduled or expected time (1=low to 5=high)	Country	LP.LPI.TIME.XQ	ibid.
LPI customs	Logistics performance index: Efficiency of customs clearance process (1=low to 5=high)	Country	LP.LPI.CUST.XQ	ibid.
LPI shipment price	Logistics performance index: Ease of arranging competitively priced shipments (1=low to 5=high)	Country	LP.LPI.ITRN.XQ	ibid.
LPI services	Logistics performance index: Competence and quality of logistics services (1=low to 5=high)	Country	LP.LPI.LOGS.XQ	ibid.
LPI tracking	Logistics performance index: Ability to track and trace consignments (1=low to 5=high)	Country	LP.LPI.TRAC.XQ	ibid.
Profit tax	Profit tax is the amount of taxes on profits paid by the business.	Country	IC.TAX.PRFT.CP.ZS	ibid.
Theft loss -firms	Percent of firms experiencing losses due to theft, robbery, vandalism or arson that occurred on the establishment's premises.	Country	IC.FRM.THEV.ZS	ibid.
Blackout -firms	Percent of firms experiencing electrical outages during the previous fiscal year.	Country	IC.ELC.OUTG.ZS	ibid.
E-access	Access to electricity, % of population. Electrification data are collected from industry, national surveys and international sources.	Country	EG.ELC.ACCS.ZS	ibid.
E-access rural	Access to electricity, rural, % rural population.	Country	EG.ELC.ACCS.RU.ZS	ibid.
E-access urban	Access to electricity, urban, % of urban population.	Country	EG.ELC.ACCS.UR.ZS	ibid.
Remittance cost origin	Average transaction cost of sending remittances from a specific country, %	Country	SI.RMT.COST.OB.ZS	ibid.
Remittance cost destination	Average transaction cost of sending remittances to a specific country, %	Country	SI.RMT.COST.IB.ZS	ibid.
Customs burden	Burden of Customs Procedure measures business executives' perceptions of their country's efficiency of customs procedures. The rating ranges from 1 to 7, with a higher score indicating greater efficiency. (1=extremely inefficient to 7=extremely efficient)	Country	IQ.WEF.CUST.XQ	ibid.
Business start cost	Cost of business start-up procedures, % of GNI per capita	Country	IC.REG.COST.PC.ZS	ibid.
Customs duties	Customs and other import duties, % of tax revenue	Country	GC.TAX.IMPT.ZS	ibid.
Education BA	Educational attainment, at least Bachelor's or equivalent, population 25+, total (%) (cumulative)	Country	SE.TER.CUAT.BA.ZS	ibid.
Education post-sec.	Educational attainment, at least completed post-secondary, % of total population 25+ (cumulative)	Country	SE.SEC.CUAT.PO.ZS	ibid.

Table C.5 – continued from previous page

Variable name	Meaning	Detalization	Based on database variable	Source
Education upper-sec.	Educational attainment, at least completed upper secondary, % of total population 25+ (cumulative)	Country	SE.SEC.CUAT.UP.ZS	ibid.
Education MA	Educational attainment, at least Master's or equivalent, % of total population 25+ (cumulative)	Country	SE.TER.CUAT.MS.ZS	ibid.
Inflation, consumer price	Inflation, consumer prices, annual %	Country	FP.CPI.TOTL.ZG	ibid.
Inflation, deflator	Inflation, GDP deflator, annual %	Country	NY.GDP.DEFL.KD.ZG	ibid.
Inflation, deflator LS	Inflation, GDP deflator - linked series, annual %	Country	NY.GDP.DEFL.KD.ZG.AD	ibid.
Blackout frequency	Power outages in firms in a typical month, number	Country	IC.ELC.OUTG	ibid.
Business start procedures	Start-up procedures to register a business, number	Country	IC.REG.PROC	ibid.

Table C.6: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Major dependent variables					
ec-extra					
Gini	6,366	0.921	0.054	0.475	1
Digital sales	4,463	19.082	14.708	0	87
E-sales	4,404	16.942	14.028	0	87
Web-sales	3,563	15.835	14.379	0	87
E-sales turnover	4,872	17.206	10.767	0	44
ec					
Gini	5,260	0.923	0.053	0.475	1
Digital sales	3,787	19.030	15.189	0	87
E-sales	3,737	16.763	14.481	0	87
Web-sales	2,984	15.956	15.123	0	87
E-sales turnover	4,032	17.027	10.809	0	44
full					
Gini	51,750	0.888	0.105	0.00018	1
Labor productivity 0-9	59,388	246.062	15135.03	-11623	1943718
Labor productivity 10-19	46,827	46.323	112.910	-1057.8	12385.2
Labor productivity 20-49	44,911	51.455	113.573	-1590.72	11734.4
Labor productivity 50-249	40,749	58.712	99.309	-439.74	9013.9
Labor productivity 250+	27,318	66.049	75.078	-196.6	2791.5
Turnover per employee 0-9	59,778	527.483	20091.68	0.2	2582238
Turnover per employee 10-19	47,423	178.698	461.244	0.1	34236.3
Turnover per employee 20-49	45,651	204.375	479.189	0.7	21034.8
Turnover per employee 50-249	41,733	239.981	530.752	1.46	23905.3
Turnover per employee 250+	28,437	262.725	546.266	-11.3	17448.8
Major independent variables					
ec-extra					
E-shoppers	8,998	36.56	23.08	1	83

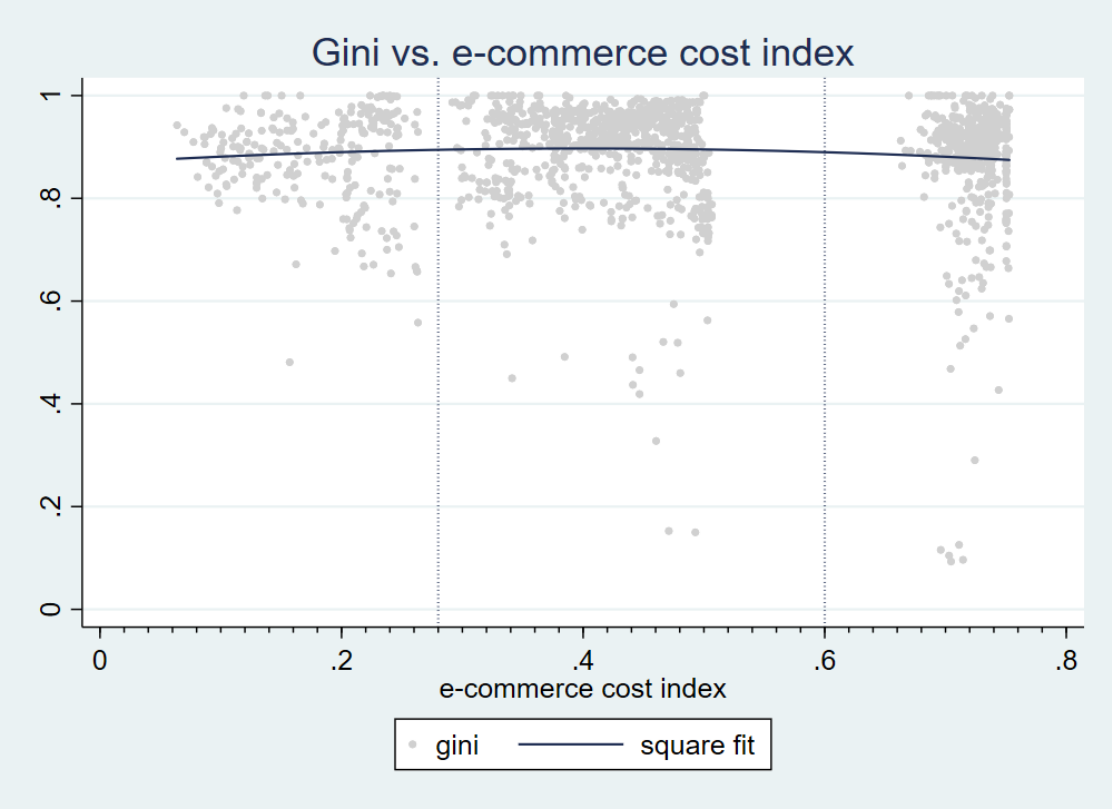
Table C.6 – continued from previous page

Variable	Obs	Mean	Std. Dev.	Min	Max
No e-shopping	8,998	30.80	10.76	10	61
E-shopping problem: high price	3,608	2.93	2.96	0	21
E-shopping problem: fraud	2,662	2.21	1.57	0	8
E-shopping problem: wrong good	3,828	5.17	3.97	0	20
E-shopping problems	2,772	20.63	16.36	2	68
Deliveries urban	7,348	1.02	0.12	1	2
Post office	6,820	1830.6	3905.0	0	27600
No post	6,666	0.02	0.29	0	5
Production_{t-1}	8,117	11784.0	36810.9	22.5	529882.9
GDP p.c. -CLV EUR_{t-1}	9,701	25572.6	18063.8	2900	84400
Price index $'15_{t-1}$	8,843	96.04	17.39	43.5	185.8
Property reg. time	6,688	28.69	33.90	1	391
Warehouse time	6,688	182.75	105.77	64	677
Customs burden	6,710	4.75	0.62	3	6.3
$\text{R\&D rate}_{t-1}$	3,446	0.04	0.08	5.87E-09	2.637
$\text{R\&D rate current}_{t-1}$	2,334	0.03	0.07	2.27E-07	2.525
$\text{R\&D rate capital}_{t-1}$	2,269	0.00	0.02	5.87E-09	0.420
Firm share world	5,453	0.04	0.05	6.55E-05	0.5
		ec			
E-shoppers	7,362	36.562	23.078	1	83
No e-shopping	7,362	30.804	10.759	10	61
E-shopping problem: high price	2,952	2.933	2.956	0	21
E-shopping problem: fraud	2,178	2.207	1.569	0	8
E-shopping problem: wrong good	3,132	5.172	3.971	0	20
E-shopping problems	2,268	20.627	16.362	2	68
Deliveries urban	6,012	1.022	0.117	1	2
Post office	5,580	1830.6	3905.1	0	27600
No post	5,454	0.018	0.287	0	5
Production_{t-1}	6,641	11783.4	36811.7	22.5	529882.9
GDP p.c. -CLV EUR_{t-1}	7,937	25572.4	18064.1	2900	84400
Price index $'15_{t-1}$	2,436	97.083	10.512	42.537	167.9
Property reg. time	5,474	29.114	34.901	1	391
Warehouse time	5,474	184.477	108.752	64	677
Customs burden	5,491	4.750	0.619	3	6.3
$\text{R\&D rate}_{t-1}$	2,688	0.037	0.082	5.87E-09	2.637
$\text{R\&D rate current}_{t-1}$	1,804	0.030	0.081	2.27E-07	2.525
$\text{R\&D rate capital}_{t-1}$	1,749	0.005	0.020	5.87E-09	0.420
Firm share world	4,417	0.038	0.055	0.00007	0.5
		full			
E-shoppers	265,441	36.562	23.076	1	83
No e-shopping	265,441	30.804	10.758	10	61
E-shopping problem: high price	106,436	2.933	2.955	0	21
E-shopping problem: fraud	78,529	2.207	1.569	0	8
E-shopping problem: wrong good	112,926	5.172	3.970	0	20
E-shopping problems	81,774	20.627	16.358	2	68
Deliveries urban	216,766	1.022	0.117	1	2
Post office	201,190	1830.6	3904.7	0	27600
No post	196,647	0.018	0.287	0	5
Production_{t-1}	140,493	5294.1	70513.0	-54386	20300000

Table C.6 – continued from previous page

Variable	Obs	Mean	Std. Dev.	Min	Max
GDP p.c. -CLV EUR$_{t-1}$	286,208	25573.2	18062.1	2900	84400
Price index '15$_{t-1}$	30,818	96.313	12.563	31.7	426
Property reg. time	197,296	28.688	33.895	1	391
Warehouse time	197,296	182.753	105.759	64	677
Customs burden	197,945	4.749	0.615	3	6.3
R&D rate$_{t-1}$	11,593	0.066	0.398	-32.690	7.490
R&D rate current$_{t-1}$	7,328	0.050	0.156	-0.128	4.803
R&D rate capital$_{t-1}$	7,368	0.006	0.026	-0.012	1.036
Firm share world	27,322	0.050	0.123	0	14

Figure C.1: Market concentration and e-commerce cost index



D Estimation results

D.1 Proxies of e-commerce costs, ec-extra sample

Table D.1: Online shopper's behavior

Method	Robust FE	Robust FE	Robust FE
Dep. variable	Digital sales	Digital sales	Digital sales
	(1)	(2)	(3)
GDP p.c. -CLV EUR	0.001***		0.001***
	(0.000)		(0.000)
GDP p.c. -CLV		0.205*	
		(0.117)	
Price index '15	0.002		
	(0.007)		
Population		−0.008	
		(0.005)	
MIG*Infrastructure	0.064	0.219*	0.392**
	(0.168)	(0.122)	(0.193)
SEC score	0.802		4.874***
	(1.161)		(1.157)
No e-shopping	−0.080**		−0.078
	(0.032)		(0.061)
Internet mobile		0.400	
		(0.308)	
E-shopping domestic		−0.060	
		(0.104)	
E-shopping EU-foreign		−0.232	
		(0.259)	
E-shopping RoW		0.442*	
		(0.260)	
E-shopping lowcost		−0.046	
		(0.350)	
E-shopping highcost 1000		−0.189	
		(0.208)	
E-shopping no problems			−0.047**
			(0.020)
N.obs.	1564	224	703
N.groups	239	120	235
R-squared	0.224	0.211	0.286
Prob > F	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.2: Problems when selling online

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Digital sales	E-sales	Web-sales	Digital sales	E-sales	Web-sales
	(1)	(2)	(3)	(4)	(5)	(6)
Web-sales problem: fit	0.530***	0.295	0.503***			
	(0.159)	(0.185)	(0.140)			
Web-sales problem: logistics	0.441***	0.550***	0.519***			
	(0.121)	(0.151)	(0.125)			
Web-sales problem: payment	−0.035	0.047	0.090			
	(0.147)	(0.178)	(0.146)			
Web-sales problem: security	0.144	0.069	0.119			
	(0.131)	(0.193)	(0.143)			
Web-sales problem: legal	0.245	0.313	0.151			
	(0.195)	(0.229)	(0.192)			
Web-sales problem: cost	0.253	0.119	0.232			
	(0.232)	(0.305)	(0.226)			
No web-sales problem: fit				−0.154***	−0.142***	−0.158***
				(0.036)	(0.035)	(0.037)
No web-sales problem: logistics				0.123**	0.168***	0.105**
				(0.050)	(0.045)	(0.048)
No web-sales problem: payment				−0.093	−0.096*	−0.109*
				(0.058)	(0.056)	(0.059)
No web-sales problem: security				−0.045	−0.041	−0.039
				(0.066)	(0.086)	(0.065)
No web-sales problem: legal				−0.025	−0.061	−0.001
				(0.082)	(0.087)	(0.084)
No web-sales problem: cost				0.046	0.041	0.031
				(0.050)	(0.052)	(0.049)
N.obs.	915	885	920	1033	987	1027
N.groups	463	461	467	515	510	513
R-squared	0.464	0.396	0.461	0.387	0.302	0.411
Prob > F	0.000	0.000	0.000	0.002	0.000	0.001

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.3: Problems when buying online

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable:										
Digital sales	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP p.c.	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
-CLV EUR_{t-1}	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MIG*	0.382*	0.373*	0.416**	0.331	0.352*	0.361*	0.386*	0.353*	0.266**	0.374*
Infrastructure	(0.217)	(0.194)	(0.194)	(0.209)	(0.209)	(0.201)	(0.197)	(0.198)	(0.105)	(0.197)
SEC score	4.477***	3.820***	3.677**	4.454***	3.610**	4.121**	3.584***	4.202***	0.822	3.775***
	(1.242)	(1.456)	(1.469)	(1.522)	(1.459)	(1.466)	(1.462)	(1.527)	(3.120)	(1.489)
E-shoppers	0.114***	0.117***	0.098***	0.132***	0.150***	0.104***	0.102***	0.099***	−0.085*	0.099***
	(0.036)	(0.035)	(0.033)	(0.033)	(0.031)	(0.033)	(0.034)	(0.032)	(0.046)	(0.032)
E-shopping problems	0.010									
	(0.034)									
E-shopping problem: long delivery		−0.024								
		(0.044)								
E-shopping problem: high price			0.181							
			(0.172)							
E-shopping problem: wrong good				−0.241**						
				(0.094)						
E-shopping problem: fraud					−1.141***					
					(0.194)					
E-shopping problem: complaint						−0.100				
						(0.168)				
E-shopping problem: technical							0.040			
							(0.046)			
E-shopping problem: info								−0.116		
								(0.082)		
E-shopping problem: coverage									−0.289***	
									(0.107)	
E-shopping problem: other										0.047
										(0.023)
N.obs.	509	694	674	688	667	674	688	674	542	681
N.groups	232	233	228	232	228	228	232	228	191	230
R-squared	0.283	0.301	0.285	0.307	0.282	0.289	0.302	0.299	0.271	0.287
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.4: Post and logistics - deliveries and post coverage

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable: Digital sales	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP p.c. -CLV EUR	0.000	0.001***	0.001**	0.001**	0.001***	0.000**	0.001**	0.000	0.001***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MIG* Infrastructure	−0.192	0.210	−0.051	−0.165	−0.079	−0.060	0.073	0.128	−0.191	
	(0.224)	(0.316)	(0.239)	(0.222)	(0.236)	(0.271)	(0.271)	(0.255)	(0.249)	
MIG score										−2.763*
										(1.463)
SEC score	2.350	6.173***	3.619*	3.205	4.011**	2.901*	3.610**	15.816***	3.056*	
	(1.746)	(2.342)	(2.096)	(2.174)	(1.856)	(1.633)	(1.652)	(5.595)	(1.678)	
E-shoppers	0.181***	0.153***	0.184***	0.174***	0.176***	0.198***	0.152***	0.290***	0.146***	0.192***
	(0.041)	(0.035)	(0.037)	(0.045)	(0.035)	(0.039)	(0.040)	(0.062)	(0.050)	(0.039)
E-shopping problem: fraud	−1.398***	−1.152***	−1.103***	−1.648***	−1.250***	−1.048***	−1.108***	−1.375***	−1.703***	−1.343***
	(0.266)	(0.240)	(0.206)	(0.316)	(0.203)	(0.222)	(0.219)	(0.290)	(0.361)	(0.257)
No post	−131.861***									−129.030***
	(25.928)									(25.558)
Home delivery		−0.130***								
		(0.037)								
Post office person coverage			−0.001***						0.003**	−0.001**
			(0.000)						(0.001)	(0.000)
Delivery to post				0.119***					0.062	
				(0.039)					(0.079)	
Deliveries urban					7.679***				13.605***	7.630***
					(2.290)				(3.770)	(2.285)
Post offices						0.000***			0.000	
						(0.000)				
Post delivery staff							0.062***		0.007	
							(0.026)		(0.031)	
Post office auto								0.001***		
								(0.000)		
N.obs.	509	545	565	453	572	539	489	379	391	513
N.groups	203	206	207	189	207	203	190	156	172	204
R-squared	0.274	0.237	0.250	0.248	0.272	0.342	0.228	0.093	0.241	0.224
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.5: Post and logistics - parcels and revenue

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable: Digital sales	(1)	(2)	(3)	(4)	(5)	(6)
GDP p.c. -CLV EUR	0.000	0.000	0.000	0.000	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MIG*Infrastructure	0.037	0.056	−0.247	0.028	0.099	0.065
	(0.304)	(0.271)	(0.227)	(0.281)	(0.248)	(0.276)
SEC score	4.178*	1.937	2.039	1.700	6.134**	5.001*
	(2.318)	(1.930)	(1.991)	(1.719)	(2.382)	(2.568)
E-shoppers	0.177***	0.155***	0.207***	0.230***	0.156***	0.176***
	(0.043)	(0.044)	(0.037)	(0.042)	(0.034)	(0.043)
E-shopping problem: fraud	−1.932***	−1.595***	−1.681***	−1.666***	−1.306***	−1.365***
	(0.337)	(0.301)	(0.240)	(0.280)	(0.212)	(0.266)
Express int'l dispatch	3.240**					
	(1.440)					
Express int'l receipt		−8.380***				
		(2.930)				
Parcels domestic			0.046**			
			(0.020)			
Parcels int'l dispatch				−5.920*		
				(3.270)		
Post revenue					0.000*	
					(0.000)	
Post parcel revenue						0.139*
						(0.072)
N.obs.	304	321	366	321	562	436
N.groups	115	121	147	121	201	181
R-squared	0.079	0.104	0.164	0.177	0.274	0.307
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.6: Institutional factors in e-commerce sales

Dep. variable:	E-sales	E-sales manufacturing	E-sales	E-sales	E-sales manufacturing	E-sales	E-sales	E-sales manufacturing	E-sales	Digital sales
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
E-shoppers$_{t-1}$	0.251***	0.204***	0.262***	0.261***	0.202***	0.239***	0.261***	0.214***	0.265***	0.220***
	(0.024)	(0.025)	(0.024)	(0.024)	(0.026)	(0.028)	(0.023)	(0.026)	(0.025)	(0.039)
E-shopping problem: fraud	−0.867***	−1.133***	−1.287**	−0.884***	−1.209***	−0.829***	−0.886***	−1.223***	−0.854***	−0.758***
	(0.180)	(0.253)	(0.533)	(0.182)	(0.264)	(0.182)	(0.175)	(0.260)	(0.181)	(0.229)
Post offices$_{t-1}$	0.163***	0.211***	0.165***	0.178***	0.244***	0.201***	0.176***	0.221***	0.157**	0.248***
	(0.061)	(0.038)	(0.061)	(0.060)	(0.037)	(0.063)	(0.059)	(0.037)	(0.067)	(0.078)
Deliveries urban$_{t-1}$	7.615***	9.256***	7.798***	7.706***	7.760***	8.419***	8.624***	10.052***	8.873***	8.110**
	(1.931)	(1.353)	(1.929)	(1.946)	(1.509)	(1.995)	(1.760)	(1.416)	(1.821)	(3.287)
No post	−121.27***	−148.84***	−128.39***	−127.99***	−159.15***	−141.12***	−127.77***	−153.70***	−120.92***	−153.31***
	(41.803)	(24.186)	(41.606)	(41.599)	(24.477)	(42.735)	(41.544)	(24.063)	(40.622)	(42.410)
Enforcement time	0.005	0.004*								
	(0.003)	(0.002)								
Fraud* Enforcement			0.001							
			(0.001)							
Warehouse time				−0.003	−0.008**					
				(0.004)	(0.003)					
Business start cost						−0.153				
						(0.106)				
E-access							0.705	3.835***		
							(0.773)	(0.821)		
Customs burden									−0.874	
									(0.887)	
Education upper-sec.										0.482*
										(0.253)
N.obs.	1160	462	160	1160	462	1160	1211	481	1170	710
N.groups	449	182	449	449	182	449	469	193	451	370
R-squared	0.057	0.233	0.064	0.067	0.308	0.057	0.060	0.267	0.061	0.015
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.7: Institutional factors in e-commerce turnover

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable: E-sales turnover	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
E-shoppers$_{t-1}$	0.069***	0.103***	0.088***	0.080***	0.087***	0.095***	0.086***	0.080***	−0.272***	0.088***
	(0.020)	(0.022)	(0.020)	(0.020)	(0.019)	(0.019)	(0.018)	(0.019)	(0.044)	(0.020)
E-shopping problem: fraud	−0.528***	−2.281***		−0.451***	−0.450***	−0.506***	−0.490***	−0.503***	−0.913***	−0.512***
	(0.146)	(0.452)		(0.146)	(0.147)	(0.153)	(0.136)	(0.142)	(0.187)	(0.155)
Post offices$_{t-1}$	0.173***	0.156***	0.200***	0.199***	0.203***	0.185***	0.200***	0.173***	0.226***	0.169***
	(0.023)	(0.026)	(0.021)	(0.020)	(0.019)	(0.019)	(0.017)	(0.022)	(0.039)	(0.025)
No post	−58.501***	−76.926***	−65.596***	−71.529***	−70.974***	−58.033***	−71.304***	−58.979***	−88.660***	−52.280***
	(7.422)	(5.743)	(6.967)	(6.557)	(6.691)	(7.993)	(5.613)	(8.040)	(10.384)	(9.857)
Production$_{t-1}$	0.346***	0.353***	0.382***	0.373***	0.376***	0.382***	0.370***	0.389***	0.398***	0.400***
	(0.034)	(0.034)	(0.034)	(0.034)	(0.035)	(0.035)	(0.033)	(0.035)	(0.148)	(0.037)
Card payments -Nr p.c.$_{t-1}$	0.082***	0.078***	0.083***	0.086***	0.085***	0.090***	0.085***	0.095***	0.215***	0.095***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.014)	(0.010)
Enforcement time	0.011***									
	(0.003)									
Fraud* Enforcement		0.003***	0.000							
		(0.001)	(0.000)							
Property reg. time				−0.009**						−0.006
				(0.005)						(0.005)
Warehouse time					−0.005*					−0.005
					(0.003)					(0.003)
Business start cost						0.161**				0.146**
						(0.068)				(0.067)
E-access							1.831			
							(2.783)			
Customs burden								−1.455***		−0.993*
								(0.508)		(0.559)
Education upper-sec.									1.076***	
									(0.153)	
N.obs.	1203	1203	1203	1203	1203	1203	1263	1216	723	1200
N.groups	441	441	441	441	441	441	462	441	387	441
R-squared	0.419	0.457	0.463	0.461	0.471	0.469	0.470	0.450	0.044	0.457
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

Table D.8: Logistics performance index

Method	Robust FE	Robust FE	Robust FE
Dep. variable:			
E-sales turnover	(1)	(2)	(3)
E-shoppers$_{t-1}$	0.190***	0.130***	0.143***
	(0.021)	(0.015)	(0.013)
E-shopping problem: fraud	−0.242		
	(0.274)		
Card payments -Nr p.c.$_{t-1}$	−0.009	−0.005	
	(0.008)	(0.007)	
Business start cost	0.040	−0.067**	
	(0.061)	(0.031)	
Warehouse time	0.028***	0.013***	0.007**
	(0.004)	(0.003)	(0.003)
LPI shipment time	−1.944***	−1.521***	−1.068**
	(0.698)	(0.381)	(0.435)
LPI customs	6.411***	2.625***	1.260**
	(1.388)	(0.590)	(0.548)
LPI shipment price			1.469***
			(0.422)
N.obs.	869	1950	2096
N.groups	503	546	586
R-squared	0.188	0.399	0.561
Prob > F	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

D.2 E-commerce adoption and market structure, ec-extra sample

Table D.9: Estimation results for the gini coefficient, ec-extra sample

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Sample	all	all	all	all	manu-facturing	manu-facturing
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)
Digital sales	0.050**			0.194	0.144**	
	(0.020)			(0.144)	(0.068)	
Digital sales^2	−0.050			−0.355	−0.284**	
	(0.033)			(0.329)	(0.135)	
E-sales		0.035*				0.159**
		(0.019)				(0.075)
E-sales^2		−0.026				−0.357**
		(0.033)				(0.156)
E-sales turnover			0.199***			
			(0.037)			
E-sales turnover^2			−0.391***			
			(0.078)			
R&D rate capital$_{t-1}$				−0.075**		
				(0.032)		
Investment in buildings				−13.504***		
				(4.874)		
GDP p.c. -CLV EUR$_{t-1}$					−2.280***	−2.130***
					(0.726)	(0.749)
Price index '15$_{5-1}$					0.000**	0.000**
					(0.000)	(0.000)
Infrastructure density					−0.011	−0.011
					(0.013)	(0.013)
Wage per hour PPS					0.003***	0.004***
					(0.001)	(0.001)
Labor productivity per hour					0.216	0.124
					(0.225)	(0.218)
N.obs.	3275	3237	3741	641	1251	1225
N.groups	542	544	556	145	234	237
R-squared	0.000	0.010	0.013	0.061	0.157	0.092
Prob > F	0.008	0.040	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

D.3 E-commerce costs and market structure, ec-extra sample

Table D.10: Alternative measures of market structure

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable:	Turnover group share 0-9	Turnover group share 250+	Firm number group share 0-9	Firm number group share 10-19	Firm number group share 250+
	(1)	(2)	(3)	(4)	(5)
E-sales turnover	0.271^{***}	-0.152^{***}	0.528^{*}	-0.200^{***}	-0.032^{***}
	(0.077)	(0.088)	(0.171)	(0.047)	(0.009)
E-sales turnover^2	-0.685^{***}	0.345^{***}	-1.114^{*}	0.432^{***}	0.065^{***}
	(0.162)	(0.204)	(0.362)	(0.094)	(0.017)
GDP p.c.	−1.437	−3.664	0.373	0.579^{***}	0.014
-CLV EUR	(1.172)	(2.812)	(1.017)	(0.176)	(0.033)
Price index '15	0.000^{*}	0.000	0.000	0.000^{***}	0.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Wage per hour PPS	0.000^{***}	0.003	0.006^{***}	-0.003^{***}	0.000^{***}
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
N.obs.	3618	3085	4122	4102	3860
N.groups	562	489	587	588	551
R-squared	0.004	0.000	0.005	0.004	0.007
Prob > F	0.000	0.009	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

D.4 Robustness check, ec sample

Table D.11: Estimation results for e-commerce adoption, ec sample

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Digital sales	Digital sales manufacturing	Digital sales services	Digital sales	E-sales	E-sales manufacturing	E-sales	E-sales turnover	E-sales turnover	Web-sales turnover
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
GDP p.c. -CLV EUR$_{t-1}$	−0.001									
	(0.000)									
E-shoppers$_{t-1}$	0.264***	0.269***	0.333***	0.170***	0.261***	0.232***	0.115***	0.082***	0.089	−0.050***
	(0.049)	(0.041)	(0.037)	(0.025)	(0.028)	(0.038)	(0.015)	(0.021)	(0.078)	(0.008)
E-shopping problem: fraud	−1.439***	−1.129***	−0.831***	−0.475*	−0.834***	−1.174***		−0.527***	1.837***	0.136***
	(0.383)	(0.343)	(0.284)	(0.280)	(0.200)	(0.325)		(0.168)	(0.430)	(0.034)
E-shopping problem: high price			0.953***							
			(0.250)							
Post offices$_{t-1}$		0.219***		0.247	0.169**	0.189***	−0.290*	0.182***	−1.730***	0.053***
		(0.061)		(0.534)	(0.072)	(0.052)	(0.169)	(0.027)	(0.659)	(0.012)
Deliveries urban$_{t-1}$	8.850***	10.128***	5.850**	8.368***	7.521***	9.010***	5.276**			
	(1.815)	(1.551)	(2.661)	(1.424)	(2.235)	(1.969)	(2.410)			
No post	−130.744***	−137.047***		130.463***	−120.882**	−166.356***		−51.679***	8.148	−76.739***
	(38.164)	(27.764)		(29.818)	(49.241)	(28.836)		(10.940)	(20.563)	(1.939)
Post financial revenue$_{t-1}$			0.229**							
			(0.102)							
Production$_{t-1}$								0.410***	0.691***	−0.071***
								(0.041)	(0.126)	(0.015)
Property reg. time					−0.015*	−0.012***				
					(0.009)	(0.003)				
Warehouse time								−0.007**	−0.006	
								(0.003)	(0.004)	
Card payments -Nr p.c.$_{t-1}$								0.099***	0.117***	0.014***
								(0.011)	(0.042)	(0.002)
Busiess start cost								0.225***		
								(0.056)		
Customs burden								−0.888	−1.456	
								(0.597)	(1.040)	
R&D rate current$_{t-1}$				20.877***			−2.290***		−16.928	
				(5.194)			(0.369)		(20.157)	

Table D.11 – continued from previous page

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Digital sales	Digital sales manufacturing	Digital sales services	Digital sales	E-sales	E-sales manufacturing	E-sales	E-sales turnover	E-sales turnover	Web-sales turnover
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
R&D rate capital$_{t-1}$				−162.787***					−33.358	
				(39.364)					(50.430)	
Firm share world				−81.007***			−26.518**		−41.018***	
				(10.019)			(16.172)		(8.075)	
N.obs.	330	349	664	247	985	335	840	990	263	846
N.groups	133	143	240	129	382	137	198	361	125	360
R-squared	0.029	0.333	0.052	0.095	0.057	0.297	0.124	0.465	0.330	0.019
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. In specificaton (1), following (statisticaly insignificant) variables are not shown: Post offices person coverage, Price index '15$_{t-1}$

Table D.12: Estimation results for the gini coefficient and e-commerce adoption, ec sample

Method	Robust FE	IV, Robust FE	Robust FE	Robust FE	IV, Robust FE	IV, Robust FE	IV, Robust FE
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Digital sales	0.045	−0.558					
	(0.041)	(0.495)					
Digital sales^2	−0.100	2.380					
	(0.100)	(1.651)					
E-sales turnover			−0.105***	0.412***	−0.204**	−0.202	1.265***
			(0.037)	(0.109)	(0.093)	(0.151)	(0.224)
E-sales turnover^2			0.189**	−0.990***	0.668***	0.298	−2.452***
			(0.088)	(0.284)	(0.212)	(0.382)	(0.459)
GDP p.c.		4.200	−1.910***	−3.950***	−4.200**	0.944	−5.823**
-CLV EUR$_{t-1}$		(2.600)	(0.692)	(1.440)	(1.790)	(0.846)	(2.401)
Price index '15$_{t-1}$			0.000		0.000	0.000**	
			(0.000)		(0.000)	(0.000)	
Wage per hour PPS			0.003***	0.002	0.004***	0.001*	−0.005**
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Labor productivity			−0.032	0.191	−0.070	−0.109	
per hour			(0.166)	(0.346)	(0.280)	(0.174)	
R&D rate current$_{t-1}$				0.051			
				(0.054)			
R&D rate capital$_{t-1}$	−0.073***			−0.014			
	(0.022)			(0.016)			
Investment in	−8.700*			−14.300**			−22.279
buildings	(4.730)			(7.290)			(24.188)
Firm share world	−0.838***	−0.780***					
	(0.117)	(0.209)					
VA share world	0.024						
	(0.018)						
N.obs.	440	263	1090	616	208	525	283
N.groups	102	154	154	123	104	146	143
R-squared	0.440	0.019	0.001	0.100	0.107	0.076	0.210
Prob > F / Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table D.12 – continued from previous page

Method	Robust FE	IV, Robust FE	Robust FE	Robust FE	IV, Robust FE	IV, Robust FE	IV, Robust FE
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Instruments		E-shoppers$_{t-1}$, E-shopping problem: fraud, Deliveries urban$_{t-1}$, R&D rate current$_{t-1}$, R&D rate capital$_{t-1}$			E-shoppers$_{t-1}$, E-shopping problem: fraud, Post offices$_{t-1}$, Production$_{t-1}$, Card payments -Nr p.c.$_{t-1}$, Business start cost, No post, Warehouse time, Customs burden	E-shoppers$_{t-1}$, Warehouse time, LPI shipment time, LPI customs, LPI shipment price	E-shoppers$_{t-1}$, E-shopping problem: fraud, Post offices$_{t-1}$, No post, Production$_{t-1}$, Card payments -Nr p.c.$_{t-1}$, Warehouse time, Customs burden, Firm share world
AR, Prob > chi2		0.000			0.000	0.083	0.000
Wald, Prob > chi2		0.118			0.000	0.057	0.000
Omitted controls	Turnover share world, Investment share world, Employment share world		E-access	Investment in construction, Investment in machinery			

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. "Omitted controls" lists variables in the model specification, whose coefficients were not statistically significant and which are not shown in this table.

Table D.13: Estimation results for the gini coefficient and e-comemrce costs, ec sample

Method	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable: gini	(1)	(2)	(3)	(4)
GDP p.c. -CLV EUR_{t-1}	−0.684		−2.623	−1.842
	(6.865)		(1.759)	(1.528)
Price index '15	0.000	0.000		
	(0.000)	(0.000)		
Wage per hour PPS	−0.004	0.005***		
	(0.003)	(0.001)		
Labor productivity per hour	1.358**			
	(0.574)			
Investment in buildings			−22.066	−22.486
			(16.123)	(16.363)
E-shoppers$_{t-1}$	−0.620***	−0.026*	0.128	0.096
	(0.234)	(0.014)	(0.087)	(0.070)
E-shoppers$_{t-1}^2$	0.892**		−0.226	−0.187
	(0.347)		(0.163)	(0.144)
E-shopping problem: fraud	−1.112**	0.061		
	(0.475)	(0.403)		
E-shopping problem: fraud^2	26.727**	1.241		
	(11.530)	(10.152)		
Post offices$_{t-1}$	0.014**	0.001	0.040**	0.034**
	(0.006)	(0.004)	(0.019)	(0.016)
Post offices$_{t-1}^2$	−0.001**	0.000	−0.005***	−0.004*
	(0.000)	(0.000)	(0.003)	(0.002)
No post	0.938**	−0.237***	0.024**	0.142
	(0.429)	(0.071)	(0.061)	(0.054)
No post^2	−51.779***		2.357	
	(15.262)		(2.078)	
Production$_{t-1}$	74.357***	1.170	−1.747	−2.199
	(26.029)	(3.112)	(1.501)	(1.811)
Production$_{t-1}^2$	−3248.300***		10.848	15.831
	(1036.910)		(10.441)	(13.261)
Card payments -Nr p.c.$_{t-1}$	0.058	−0.001		
	(0.324)	(0.067)		
Card payments -Nr p.c.$_{t-1}^2$	−1.220			
	(0.866)			
Business start cost	0.000	0.001**		
	(0.003)	(0.000)		
Business start cost^2	0.000*			
	(0.000)			
Warehouse time	2.782**	−0.330	−0.650*	−0.545*
	(1.114)	(0.269)	(0.369)	(0.321)
Warehouse time^2	−4.348**	0.699	0.958*	0.810*
	(1.928)	(0.481)	(0.521)	(0.459)
Customs burden		0.021**	−0.048	
		(0.010)	(0.060)	
Customs burden^2	−0.005**		0.005	
	(0.002)		(0.006)	
Firm share world			−1.058***	−1.018***
			(0.125)	(0.064)
Firm share world^2			0.100	
			(0.588)	
R&D rate current$_{t-1}$			−0.742	0.077**
			(0.657)	(0.034)
R&D rate current$_{t-1}^2$			2.134	

Table D.13 – continued from previous page

Method	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable: gini	(1)	(2)	(3)	(4)
			(1.678)	
R&D rate capital$_{t-1}$			−0.113	
			(0.178)	
R&D rate capital$_{t-1}^2$			0.014	
			(0.331)	
N.obs.	243	243	438	441
N.groups	118	118	98	99
R-squared	0.011	0.159	0.155	0.168
Prob > F	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

D.5 E-commerce costs and market structure, full sample

Table D.14: Best-performing full sample specification applied to sector groups

Method	FE	FE	FE	FE	FE	FE	FE	FE	FE
Sector group	B	C	G	H	I	J	L	M	N
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GDP p.c. -CLV EUR$_{t-1}$	20.354	−0.811	0.106	−2.748	−10.382*	−3.037***	−28.317***	−1.033	1.987
	(23.244)	(0.922)	(1.524)	(2.904)	(5.944)	(1.167)	(8.522)	(1.065)	(3.690)
Investment rate$_{t-1}$	0.137	−0.001	−0.065*	−0.009	−0.059	−0.014	0.031	−0.006	−0.070**
	(0.113)	(0.001)	(0.036)	(0.010)	(0.055)	(0.018)	(0.034)	(0.004)	(0.028)
E-shoppers$_{t-1}$	0.677*	0.132***	−0.049	0.066	0.132	0.112***	0.348	0.050	0.047
	(0.416)	(0.028)	(0.057)	(0.076)	(0.182)	(0.031)	(0.266)	(0.035)	(0.097)
E-shoppers$_{t-1}{}^2$	−0.936	−0.128***	0.087	−0.144	−0.328	−0.071*	−0.284	−0.029	0.032
	(0.513)	(0.037)	(0.074)	(0.109)	(0.252)	(0.040)	(0.368)	(0.047)	(0.135)
Post offices$_{t-1}$	−0.001	0.001	0.000	0.003	0.036	−0.001	−0.002	0.003**	−0.002
	(0.010)	(0.001)	(0.003)	(0.002)	(0.028)	(0.001)	(0.016)	(0.001)	(0.004)
Production$_{t-1}$	10.613	−0.208	0.005	0.538	2.968	0.613	8.780	0.217	−0.370
	(13.344)	(0.251)	(0.345)	(1.261)	(6.363)	(0.665)	(9.766)	(0.443)	(1.843)
Production$_{t-1}{}^2$	−100.139	0.168	0.003	0.642	−15.042	−0.796	−40.388	−0.138	2.075
	(112.205)	(0.258)	(0.782)	(3.554)	(38.883)	(1.499)	(104.809)	(0.771)	(4.428)
Warehouse time	−2.677	−0.017	−0.004	−0.003	−0.089	−0.002	1.770	0.016	−0.114
	(3.285)	(0.063)	(0.093)	(0.147)	(0.311)	(0.067)	(3.060)	(0.085)	(0.223)
Warehouse time^2	5.129	0.009	0.045	0.047	0.259	0.083	−3.367	0.012	0.186
	(6.926)	(0.103)	(0.125)	(0.170)	(0.311)	(0.101)	(7.690)	(0.092)	(0.273)
Customs burden	−0.389	0.020	0.019	−0.009	0.047	0.099***	0.098	0.073***	0.211**
	(0.360)	(0.022)	(0.044)	(0.058)	(0.138)	(0.026)	(0.296)	(0.023)	(0.092)
Customs burden^2	0.046	−0.003	−0.003	0.001	−0.006	−0.012***	−0.017	−0.009***	−0.026**
	(0.043)	(0.002)	(0.005)	(0.007)	(0.016)	(0.003)	(0.032)	(0.003)	(0.010)
Firm share world	3.892	−0.881***	−0.635***	0.467	−5.691*	0.050	−1.896*	−1.209***	−1.441**
	(2.877)	(0.081)	(0.237)	(0.325)	(3.105)	(0.194)	(1.047)	(0.391)	(0.694)
Firm share world^2	−15.429	0.906***	3.231***	−2.654***	134.825	−1.857*	5.001	16.155***	3.155
	(10.421)	(0.211)	(0.893)	(0.873)	(93.546)	(1.064)	(4.728)	(3.678)	(4.934)
R&D rate current$_{t-1}$	−11.885**	−0.048	−0.052	−3.928	−33.955	−0.079	13.816	−0.033	−0.428
	(5.279)	(0.035)	(0.386)	(6.143)	(61.463)	(0.074)	(45.315)	(0.021)	(3.257)

Table D.14 – continued from previous page

Method	FE	FE	FE	FE	FE	FE	FE	FE	FE
Sector group	B	C	G	H	I	J	L	M	N
Dep. variable: gini	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
R&D rate current$_{t-1}{}^2$	204.490	0.009	0.815	795.563	76361.6	−0.508	-4960.26	0.008	4.900
	(140.975)	(0.006)	(2.888)	(1105.36)	(74905.6)	(0.371)	(11463.4)	(0.012)	(223.808)
N.obs.	56	1606	155	252	67	518	56	198	324
N.groups	17	388	34	75	19	156	18	54	103
R-squared	0.088	0.184	0.001	0.074	0.002	0.008	0.001	0.083	0.024
Prob > F	0.385	0.000	0.010	0.000	0.075	0.000	0.063	0.000	0.017

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. Sectors D, E, F, K, P, Q, R, S omitted due to insufficient observations. The sectors are as follows: B=Mining and quarrying; C=Manufacturing; G=Wholesale and retail trade; repair of motor vehicles and motorcycles; H=Transportation and storage; I=Accommodation and food service activities; J=Information and communication; L=Real estate activities; M=Professional, scientific and technical activities; N=Administrative and support service activities.

Table D.15: Best-performing sector-specific models, sectors B-D, full sample

Method	Robust FE		Robust FE		Robust FE
Sector group	B		C		D
Dep. variable: gini	(1)		(2)		(3)
GDP p.c. -CLV EUR$_{t-1}$	4.214	GDP p.c. -CLV EUR$_{t-1}$	−0.710	E-shoppers$_{t-1}$	0.097***
	(9.681)		(1.099)		(0.029)
Warehouse time	−2.320***	E-shoppers$_{t-1}$	0.129***	Production$_{t-1}$	1.035**
	(0.442)		(0.024)		(0.450)
Warehouse time^2	4.645***	E-shoppers$_{t-1}{}^2$	−0.088***	Production$_{t-1}{}^2$	−5.468**
	(0.867)		(0.030)		(2.522)
R&D rate capital$_{t-1}$	−6.140	No post	0.299***	Firm share world	−0.706***
	(5.086)		(0.080)		(0.085)
		Firm share world	−1.118***	Firm share world^2	1.988***
			(0.202)		(0.162)
		Firm share world^2	1.477***		
			(0.517)		
		R&D rate current$_{t-1}$	−0.001		
			(0.002)		
N.obs.	77		1685		392
N.groups	18		382		54
R-squared	0.140		0.225		0.122
Prob > F	0.000		0.000		

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. The sectors are as follows: B=Mining and quarrying; C=Manufacturing; D=Electricity, gas, steam and air conditioning supply

Table D.16: Best-performing sector-specific models, sectors E-H, full sample

Method	Robust FE		Robust FE		Robust FE
Sector group	E		G		H
Dep. variable: gini	(1)		(2)		(3)
GDP p.c. -CLV EUR$_{t-1}$	−1.516*	Inflation, deflator LS	0.001	GDP p.c. -CLV EUR$_{t-1}$	−4.293*
	(1.768)		(0.001)		(2.462)
Investment rate$_{t-1}$	0.011	Profit tax	0.001**	Investment rate$_{t-1}$	−0.019*
	(0.006)		(0.001)		(0.010)
E-shoppers$_{t-1}$	0.191	E-shoppers$_{t-1}$	−0.001	E-shoppers$_{t-1}$	0.078
	(0.085)		(0.050)		(0.054)
E-shoppers$_{t-1}^2$	−0.155	E-shoppers$_{t-1}^2$	−0.040	E-shoppers$_{t-1}^2$	−0.100
	(0.116)		(0.036)		(0.066)
Customs burden	−0.049	E-shopping product	0.189	Post offices$_{t-1}$	1.381*
	(0.054)		(0.128)		(0.727)
Customs burden^2	0.005	E-shopping product^2	−0.179	Production$_{t-1}$	0.458**
	(0.007)		(0.169)		(0.174)
Firm share world	−0.293	Post offices$_{t-1}$	−0.002**	Firm share world	−0.550***
	(0.304)		(0.001)		(0.045)
				R&D rate current$_{t-1}$	−3.669
					(2.311)
				R&D rate current$_{t-1}^2$	841.154***
					(318.589)
N.obs.	544		466		270
N.groups	107		78		75
R-squared	0.019		0.031		0.183
Prob > F	0.011		0.060		0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. The sectors are as follows: E=Water supply; sewerage, waste management and remediation activities; G=Wholesale and retail trade; repair of motor vehicles and motorcycles; H=Transportation and storage

Table D.17: Best-performing sector-specific models, sectors I-L, full sample

Method	Robust FE		Robust FE		Robust FE
Sector group	I		J		L
Dep. variable:					
gini	(1)		(2)		(3)
Investment rate$_{t-1}$	−0.017	Inflation, deflator LS	0.001***	Inflation, deflator LS	0.001**
	(0.011)		(0.000)		(0.000)
Cash withdrawals$_{t-1}$	0.000*	E-shoppers$_{t-1}$	0.174***	E-shopping problem: fraud* Home enterp.	1.841***
	(0.000)		(0.027)		(0.315)
Education upper-sec.	2.750	E-shoppers$_{t-1}{}^2$	−0.171***	E-shopping problem: fraud* Home enterp.^{2}	−30.699***
	(1.493)		(0.033)		(7.432)
Education upper-sec.^{2}	−1.651	Production$_{t-1}$	0.245**	E-shopping problem: high price* Home enterp.	1.422***
	(1.087)		(0.125)		(0.271)
E-shoppers$_{t-1}$	0.040	Firm share world	−0.419***	E-shopping problem: high price* Home enterp.^{2}	−25.405***
	(0.128)		(0.102)		(4.381)
E-shoppers$_{t-1}{}^2$	−0.294**			Post offices$_{t-1}$	−0.974***
	(0.132)				(0.090)
E-shopping problem: long delivery	0.851*				
	(0.501)				
E-shoping problem: wrong good	0.984				
	(0.957)				
E-shopping problem: wrong good^2	−15.188**				
	(8.170)				
Warehouse time	0.001				
	(0.000)				
Warehouse time^2	0.000**				
	(0.000)				
Business start cost	0.006				
	(0.004)				
E-shopping problem: info	−0.341				
	(0.581)				
Internet access rural	−0.253				
	(0.181)				
Internet access rural^2	0.182				
	(0.198)				
N.obs.	464		1929		148
N.groups	213		283		54
R-squared	0.046		0.074		0.381
Prob > F	0.000		0.000		0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. The sectors are as follows: I=Accommodation and food service activities; J=Information and communication; L=Real estate activities

Table D.18: Best-performing sector-specific models, sectors M-S, full sample

Method	Robust FE		Robust FE		Robust FE
Sector group	M		N		S
Dep. variable: gini	(1)		(2)		(3)
Post offices$_{t-1}$	3.961*	Cash withdrawals$_{t-1}$	−0.001***	Labor productivity per hour$_{t-1}$	0.065***
	(2.252)		(0.000)		(0.021)
Customs burden	0.056***	E-shoppers$_{t-1}$	0.476*	Education MA	−0.098**
	(0.020)		(0.278)		(0.037)
Customs burden^2	−0.007***	E-shoppers$_{t-1}{}^2$	0.537	E-shoppers$_{t-1}$	0.777**
	(0.002)		(0.384)		(0.330)
Firm share world	−1.024**	Post offices$_{t-1}$	-443.09**	E-shoping problem: fraud	−1.361*
	(0.457)		(213.62)		(0.798)
Firm share world^2	14.111***	License time	0.002***	E-shoping problem: complaint	−6.352***
	(3.593)		(0.001)		(2.073)
R&D rate current$_{t-1}$	−0.018**	Business start procedures	0.028		
	(0.008)		(0.021)		
Internet access rural	0.028**				
	(0.011)				
N.obs.	193		329		87
N.groups	54		217		52
R-squared	0.077		0.017		0.163
Prob > F	0.000		0.000		0.084

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. The sectors are as follows: M=Professional, scientific and technical activities; N=Administrative and support service activities; S=Other service activities

Table D.19: Estimation results for the full sample, turnover and productivity

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Enterprise turnover total	Labor productivity 10-19	Labor productivity 20-49	Enterprise turnover 0-9	Enterprise turnover 10-19	Labor productivity 0-9	Labor productivity 10-19
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP p.c. -CLV EUR_{t-1}		−0.908	−0.852	−0.030	0.209	−2.437	0.117
		(0.875)	(0.605)	(0.020)	(0.177)	(1.737)	(0.804)
Investment in buildings		5.036		−0.036		−0.061	−26.609
		(3.517)		(0.072)		(8.634)	(27.749)
Investment in concessions	0.022						
	(0.017)						
Price index '15	0.092						
	(0.058)						
E-$shoppers_{t-1}$	8.467	65.838	−15.662	−1.185*		−59.720	23.092***
	(3.716)	(37.331)	(15.543)	(0.712)		(41.190)	(7.934)
E-$shoppers_{t-1}^2$		−61.441*	64.688	1.933*		98.541**	
		(59.955)	(26.464)	(1.074)		(49.794)	
No e-shopping					0.007		
					(0.035)		
E-shopping problem: high price					−0.214**		
					(0.104)		
E-shopping problems					0.036*		
					(0.020)		
Post $offices_{t-1}$	2.716	6.809	0.279	−0.119		0.446	
	(1.870)	(7.141)	(1.966)	(0.107)		(4.063)	
Post $offices_{t-1}^2$		−0.771*		0.016		−0.346	
		(1.747)		(0.018)		(0.825)	
No post		−90.477***	38.156	−0.869	−6.178	−66.314	49.973*
		(49.114)	(26.021)	(0.756)	(7.289)	(72.397)	(29.058)
No $post^2$		2927.2		23.882		2822.4	
		(870.06)		(20.933)		(1855.7)	
$Production_{t-1}$					39.970	−125.410	
					(28.182)	(219.153)	
$Production_{t-1}^2$					−18.723	155.046	
					(13.465)	(211.149)	
Warehouse time		−16.685	−15.929*		8.385*	8.373	
		(32.353)	(13.180)		(4.302)	(37.759)	
Warehouse $time^2$		41.776	29.161		−11.349**	2.261	
		(41.861)	(15.565)		(5.116)	(42.122)	
Customs burden	−0.763	−55.099		2.552		163.632	−55.717**
	(1.441)	(58.465)		(1.674)		(131.446)	(26.963)
Customs $burden^2$		5.841		−0.262		−15.924	6.460**
		(6.573)		(0.171)		(12.836)	(3.156)
Firm share world	136.971	−43.979		−0.414	14.444***	−214.403	−50.185*
	(51.503)	(30.667)		(2.733)	(4.481)	(238.638)	(29.057)
Firm $share\ world^2$				1.798		384.013	47.757
				(4.609)		(403.667)	(19.643)
R&D $rate\ current_{t-1}$		51.073	−13.397	−1.770		−91.930	−11.897
		(19.729)	(1.092)	(1.229)		(105.045)	(3.378)
R&D $rate\ current_{t-1}^2$		−12.408		0.249		20.423	
		(3.422)		(0.245)		(20.956)	

Table D.19 – continued from previous page

Method	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE	Robust FE
Dep. variable	Enterprise turnover total	Labor productivity 10-19	Labor productivity 20-49	Enterprise turnover 0-9	Enterprise turnover 10-19	Labor productivity 0-9	Labor productivity 10-19
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
R&D rate capital$_{t-1}$,%	−0.335			0.138		7.179	
	(0.661)			(0.085)		(7.398)	
R&D rate capital$_{t-1}$,^{2}%	0.072			−0.019		−1.200	
	(0.064)			(0.014)		(1.285)	
N.obs.	402	1070	3061	1088	3272	1078	1241
N.groups	233	287	866	297	1868	295	323
R-squared	0.231	0.112	0.123	0.095	0.066	0.181	0.160
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

3 East Prussia 2.0: Persistent Regions, Rising Nations[1]

3.1 Introduction

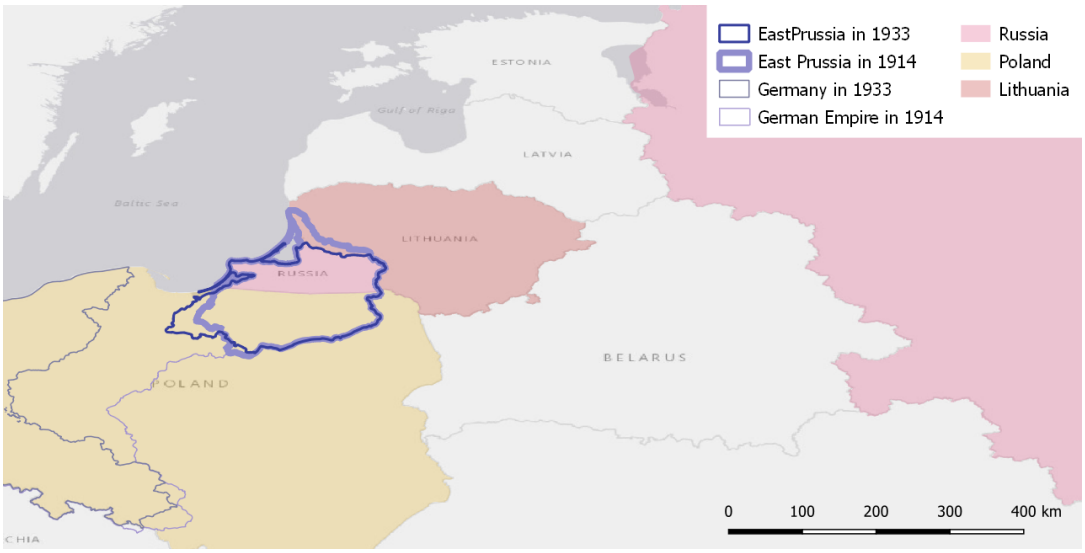
It is widely accepted that cultural and institutional factors have a massive impact on economic behavior and economic development. Economic research on persistence of cultural traits has gained much momentum in the last few decades, and an important role is played by the literature on legacies of states long vanished. It includes studies of the European empires that broke down in World War I, and of the long-lasting impact of socialism. The case of Poland, in particular, a single state reborn from the ashes of three empires, is popular in the literature as it shows how cultural traits of the different empires have persisted for a century, even in a unified political and institutional space.

Our goal is to investigate the question of cultural persistence from a different perspective, namely, through a lens of state dissolution. If one installs different institutional environments in a homogenous region, will the similarities across this region persist? Our regional focus lies in the former German province of East Prussia, which, by the end of World War II, was partitioned between Lithuania, Poland and Russia (at that time as a member of the Soviet Union). The region's location within the modern states is illustrated in Figure 3.1. Like borders that vanish *de jure* but are *de facto* visible in socioeconomic data for decades, regions that vanish through such dissolution might remain visible for quite long. Especially in terms of geographically small regions like East Prussia, one can ask whether there is a tradeoff between nation-building, which culminated in the development of nation-states during the twentieth century, and the preservation of regional ties, both economic and cultural, which have existed for centuries. In economics, for example, geographic proximity and common history or culture are an important factor in economic integration (see, e.g., Anderson and Wincoop 2003, and Wolf et al. 2011). More broadly, there is still the debate in economic geography whether it is the region or its population that determines the distribution of economic activity.

In line with research on cultural transmission in families and through inter-family spillovers (see, e.g., Bisin and Verdier 2001), we study cultural persistence in former East Prussia through the lens of the demographic shock and pre-shock diversity in that region. Our empirical approach to capture "culture" is to compare political preferences, as revealed by voting outcomes, and entrepreneurial activity in and around former East Prussia. Using detailed regional data on modern Lithuania, Poland and Russia, we first investigate whether the regions of those countries located in former East Prussia are different from those located outside it. For Lithuania and Poland, we do so using

[1] This chapter is a result of joint work with Theocharis Grigoriadis (Freie Universität Berlin). To honor his contribution, "we" will be used throughout this chapter.

Figure 3.1: East Prussia before World War II and the modern states



Source: Authors' work. Base map: GADM, HGIS Germany & ESRI Gray

the regression discontinuity design. In the case of Russia, as the region of Kaliningrad (formerly Königsberg) is geographically detached from the "mainland", we employ what is a novel method in economics: coarsened exact matching, an automated algorithm that stratifies covariates and offers the degree of post-matching covariate imbalance. Second, we also test for the similarities between the regions of the three countries in former East Prussia. Finally, as the region experienced a massive population outflow in the aftermath of World War II, we also investigate how the migrating East Prussians affected the political preferences in the regions they moved to.

We find differential patterns of East Prussian persistence across the Polish, Lithuanian and Russian territories of former East Prussia. In Poland, we observe no pattern of East Prussian persistence, with the former East Prussian territories of Poland exhibiting no significant difference in entrepreneurship in services that are more reliant on economic institutions, compared to the areas of Poland on the other side of the border. The same observation holds for political preferences: nationalism and political conservatism are lower on the East Prussian side of the internal Polish border. In Lithuania, in contrast, we observe strong patterns of persistence when we evaluate both political and economic outcomes. While the political legacy of East Prussia is significantly stronger than the economic one, we also find that economic institutions, as exhibited through entrepreneurship types, are stronger on the East Prussian side of the internal Lithuanian border. For Russia, too, we show that the East Prussian political legacies of nationalism and conservatism persist in Kaliningrad, whereas there is no evidence of persistence in relation to economic activity.

Moreover, we find that the massive population movement from East Prussia to other German regions after its partition between Poland and the Soviet Union in 1945 affected the voting patterns in the expellees' host regions in West Germany. In line with the historical patterns in East Prussia, the regions with higher shares of East Prussian

expellees in total population tend to vote more conservatively and nationalistically.

These findings perfectly illustrate the persistence channel usually postulated, namely inter-generational transmission of values (such as political preferences) and transmission of skills and networks (conducive to economic development). We find that the regions where persistence is "broken" are those that were most ethnically homogenous and where an almost complete population exchange took place after World War II. In contrast, we find evidence for persistence in Lithuania, where the Memel region was annexed as early as shortly after World War I and the German share of the population gradually decreased in the next 25 years. Thus the assimilation and gradual departure of the German population may have secured a better transmission. The absence of persistence in Poland is also in line with the notion of interrupted transmission and with the literature on migration, which states that migrating people are different from those staying in home regions (younger, more entrepreneurial, more liberal-minded). In fact, the voting patterns we observe today may well be the result of the selection bias of migration. The findings on West Germany also suggest that East Prussia "moved out", with its population taking their preferences with them and voting accordingly in their new homes. Overall, our findings not only highlight the importance of inter-generational transmission ("the people") in socioeconomic development, but also question the extent to which current studies on persistence implicitly rely on this channel.

The rest of the chapter is structured as follows. In Section 3.2, we survey the related empirical literature on the role of culture in economics, border persistence, and the role of refugees on political polarization and economic development. In Section 3.3, we provide an overview of the historical development of East Prussia. In Section 3.4, we discuss our data and empirical strategy. Section 3.5 reports persistence results from the constituent territories of East Prussia. In Section 3.6, we discuss the effects of East Prussian migration on West German political outcomes. Section 3.7 concludes.

3.2 Related literature

The role of non-economic forces in economic behavior is gaining growing attention in the scholarly literature. This link is often set up through institutions and culture that can influence the determinants of long-term economic growth and hence have long-lasting impacts. The literature receiving the greatest attention is on domestic institutions as has been introduced by Acemoglu et al. 2001, Engerman and Sokoloff 1997 or La Porta, Lopez-de-Silanes, Shleifer, and Vishny 1997 and La Porta, Lopez-de-Silanes, and Shleifer 2008, who investigate mechanisms through which colonial rule affects current development and domestic political and economic institutions. In the context of our current investigation, relevant research refers primarily to the significant persistent impact of the institutional legacy of the former pre-WWI empires on economic, political and social development. Schulze and Wolf 2009 find that the political borders that separated the Habsburg Empire's successor states after World War I became visible in the economy as early as from the mid-1880s onwards. They explain this effect of a "border before a border" by the rise of nationalism along ethno-linguistic lines, controlling for the role of physical geography, changes in infrastructure, and patterns of integration with neighboring regions outside of the Habsburg customs and monetary union.

Yet, despite this socioeconomic disintegration in empires before the war, the legacy of inter-imperial differences looks substantial. In Eastern and Southern Europe, Dimitrova-Grajzl 2007 shows that the Habsburg successor states in South-Eastern and Central Europe have more efficient institutions accommodating their modern market economies than the successor states of the Ottoman Empire. Grosjean 2011a; Grosjean 2011b has found negative and persistent effects of Ottoman rule in South-Eastern Europe on financial development and social norms of trust. Becker et al. 2016 identify a positive legacy of Habsburg rule on the lack of corruption and on the levels of trust in state institutions in Eastern European countries. Peisakhin 2015 explores how the division of a homogenous Ukrainian population between the Austrian and Russian Empires for almost 150 years has affected political attitudes and behavior in the regions that were subject to divergent imperial treatments. He surveys individuals in settlements within 15 miles (25 kilometers) of the historical Austrian-Russian border and shows that the two survey clusters differ in attitudes toward Russia and Europe. He refers to it as a "cultural legacy of historical institutions". Lechevalier and Wielgohs 2013 discuss the effects of the borders between Poland and other countries, but additionally provide a broader view on border effects using evidence on Abkhazia and Israel. Furthermore, research on Ukraine (Löwis 2015), Czech Republic (Šimon 2015), Romania (Rammelt 2015) and Serbia (Tomić 2016) provides evidence for differences in social and political attitudes and for the reappearance of historical borders. Šimon 2015 also points out the importance of demographic discontinuity, with its negative effect on civic engagement. His findings on electoral turnout in the Czech Republic are in line with research of Urbatsch 2017 on effects of ethnic cleansing in Poland. We will show, however, that demographic discontinuity has more broad effects too.

In the case of Poland, research on the differences between three regions – former partitions by neighboring empires – is numerous. Such differences manifest themselves in physical infrastructure and technology (Grosfeld and Zhuravskaya 2015; Hryniewicz 2003; Zukowski 2004), political behavior and voting patterns (Bartkowski 2003; Grosfeld and Zhuravskaya 2015; Wysokinska 2015), economic development (Bartkowski 2003; Wysokinska 2015; Zukowski 2004), interpersonal and institutional trust (Hryniewicz 2003), cultural capital (Lewicka 2005; Zarycki 2015), and social capital (Bukowski 2015; Lewicka 2005; Zukowski 2004). There is also a lot of research produced on the population composition of Western and Northern Poland. It deals with the local and regional identity of populations (e.g. Kozłowski 2003; Eberhardt 2010), social and cultural adaptation processes of new inhabitants (see the review in Michalak et al. 2011), collective memory and identity of displaced persons (Giedrojć 2005; Wylgała 2014), integration within Poland (Sakson 2006; Wolf 2005), and identity after EU accession (Makowski 2008). Methodologically, recent studies of imperial legacies often rely on spatial regression discontinuity analysis to estimate discontinuous jumps in social, economic and political characteristics at the internal border of the former empires (Russia, Germany, Austria-Hungary) in contemporary states (see, e.g., Bukowski 2015; Grosfeld and Zhuravskaya 2015). We use the regression discontinuity design for some of our hypotheses, but also apply additional methods to handle geographic discontinuities and differences across the three countries in our data.

The majority of the literature discussed above focuses on the persistence of borders that no longer formally exist. The notable exceptions are Becker et al. 2016 and Grosjean

2011a; Grosjean 2011b, who focus on legacies of the Habsburg and Ottoman Empires, respectively, in several modern states in Central and South-Eastern Europe. A further important example is Beestermöller and Rauch 2014 on trade within the former Austro-Hungarian Empire. The authors show that the successor states of the Habsburg Empire still trade more with each other than can be predicted by traditional models and explain this effect with the persistence of trading capital through cultural memory unrelated to personal contacts or networks. They identify this channel instead by excluding other possible transmission channels. Our goal is to focus on the persistence of certain attitudes in a region dissected by new borders and characterized by quite different demographic developments after that partition.

Most of the research explicitly or implicitly explains the persistence of border effects through the persistence of cultural traits. Culture is generally regarded as a fundamental determinant of economic development. North 2006, p. 4 defines "culture" as "the accumulated beliefs and inherited institutions from the past that provide the framework within which we begin thinking about [research] problems". Tabellini 2010 proposes that culture is shaped by contemporaneous social interactions as well as the cultural traditions inherited from earlier generations and finds that the proxies for culture are quantitatively significant determinants of per capita GDP levels and growth rates across European regions. An important theoretical contribution in this regard is that of Bisin and Verdier 2001, who show how cultural traits of children are formed by family socialization and the social environment of their neighborhood. They argue that the socialization efforts of parents inside the family are higher in more diverse societies, thus leading to stable equilibria with high cultural heterogeneity. In more homogenous societies, much of socialization is transferred from the family to the society. An implicit conclusion one might draw is that, in more diverse societies, cultural persistence can be more resilient to large-scale demographic shocks as it is less dependent on the social ties outside the family. Another relevant contribution on preference persistence is that of Alesina and Fuchs-Schündeln 2007, who study German division and reunification and argue that it will take one to two generations for East Germans' preferences to converge toward those of West Germans completely. This implies that dismantling the differences in preferences takes about as much time as it does to develop them in the first place.

Specifically when discussing the persistence or transmission of political preferences, we have to take into account the massive flow of refugees from the eastern regions of Germany into what was to become the FRG and the GDR (as discussed in the next section). Here, it is also important to take into consideration the literature on the formation of political attitudes and the impact of migration. Especially relevant here is the large body of literature on migration and right-wing party preferences. Increased migration has been shown to be correlated with higher levels of right-wing and extremist right support, especially among the poorer and less educated population (Corneo 2010; Decker et al. 2014; A. Falk et al. 2011; Mayda 2006; O'Rourke and Sinnott 2006; Shayo 2009). The argument also generalizes to the overall perceived threat of migrants to the natives, e.g. also because they bring other attitudes and cultural traits with them (Berning and Schlueter 2016; Decker et al. 2014; Konitzer and Grujić 2009). Interestingly, however, people tend to react with less hostility to refugees (O'Rourke and Sinnott 2006), and the evidence on the relation between the influx of refugees and right-wing party preferences is mixed (Konitzer and Grujić 2009; Sekeris and Vasilakis 2016; Steinmayr 2016). This

relation depends on, among other things, the integration of the refugees (Böhm et al. 2018; Konitzer and Grujić 2009; Steinmayr 2016), which can be facilitated by refugees' own skills, a favorable economic situation in the host regions or cultural similarities (Braun and Dwenger 2018; S. Y. Cheung and Phillimore 2014; Gericke et al. 2018). At the same time, the way the governments handle the influx of refugees echoes in right-wing party preferences (Hälbig and Lorenz 2019; Steinmayr 2016). That being said, right-wing party preferences and anti-migrant attitudes are not always fueled by the (actual) migration. There is evidence for intergenerational transmission of these preferences – the channel of preference persistence also discussed above (Avdeenko and Siedler 2017). Related to this, Ochsner and Roesel 2016 show how refugees themselves having more nationalistic attitudes influences the voting patterns of their host regions in the long term. This finding connects directly to our hypothesis on the influence of the expelled East Prussians in their host regions, as to be discussed below.

3.3 Historical background

The German presence in Eastern Europe dates back to the beginning of the thirteenth century, when the Teutonic Order started its crusades against the pagan Baltic tribes (Jasinski 1993). Prussian borders started taking more permanent shape in the fifteenth century, with the eastern border determined in 1422 and remaining unchanged until the end of World War I (Forstreuter 1955). In the west of Prussia, the fifteenth century brought the Teutonic Order losses to the Polish Kingdom, and the second treaty of Thorn in 1466 defined the Prussian border in the west and south for three centuries to come, until the first partition of Poland in 1772. The centuries-long warfare led to depopulation of the region by native Baltic peoples, especially in the eastern borderlands. The population losses were mostly recovered through migration of the German population into the region, thus leading to large-scale Germanization of Prussia (Forstreuter 1955).

In 1525, the Teutonic Order's Grand Master Albert of Brandenburg converted to Lutheranism and secularized the Prussian territories, converting Prussia into a hereditary duchy, which was inherited by the elector of Brandenburg in 1618. In 1701, Frederic III of the House of Brandenburg was crowned "King in Prussia" Frederic I in Königsberg, uniting Brandenburg and Prussia into one Kingdom of Prussia. What was originally the Duchy of Prussia would later become the province of East Prussia (Solsten 1996). The borders of Prussia changed significantly around the turn of the nineteenth century. Through the partition of Poland, it regained the territories lost through the treaty of Thorn, while also receiving substantial lands in the east and south. Some of these were, however, lost to the Russian Empire during the Napoleonic wars. The borders set in the Congress of Vienna would remain unchanged for a century, until the end of World War I.

Ever since the secularization of Prussia, the duchy, and later the province of East Prussia, was the stronghold of German agricultural nobility (Berdahl 2014). While industrialization transformed western regions of Germany throughout the nineteenth century, East Prussia was almost untouched by this process. The main economic sector was still agriculture, and most industry was mainly supplementary to it or was based on raw materials produced locally or, especially towards the end of the century, imported

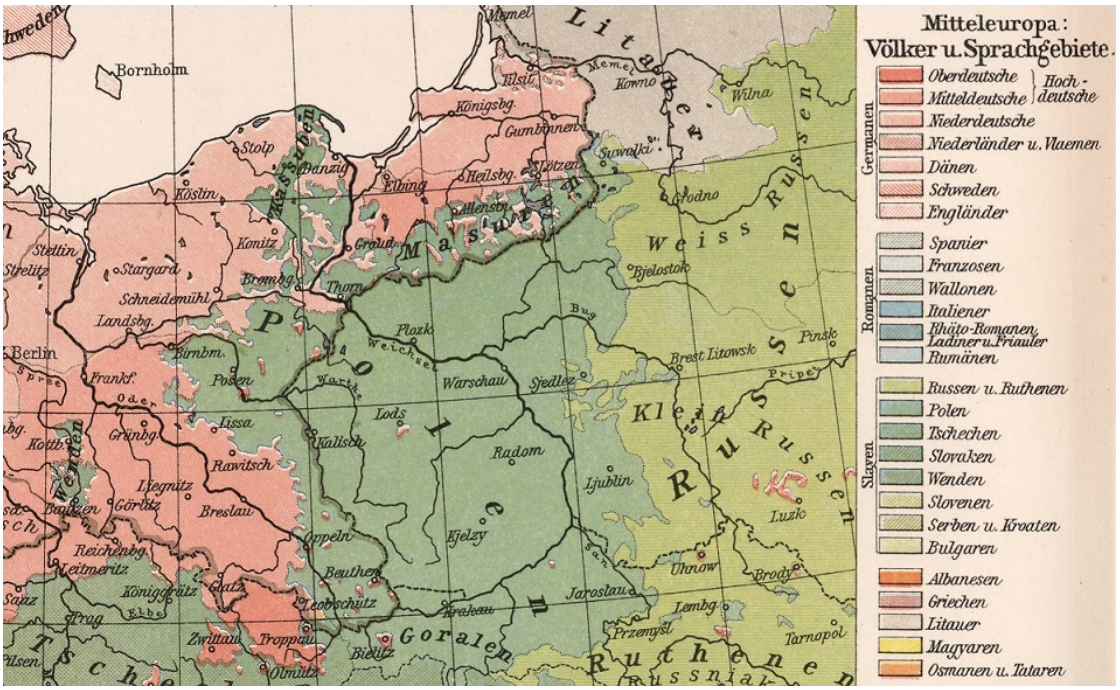
from the Russian Empire (Ambrassat 1912). Being located along the shore of the Baltic Sea, East Prussia also had several important seaports, Königsberg being the main one. The export goods were primarily grain and wood, again not only produced locally, but also increasingly shipped from the Russian Empire by internal waterways (Ambrassat 1912). Thus, it is not surprising that, politically, East Prussia was also a stronghold of conservatism and growing nationalism in the German Empire. It was one of the largest supporters of conservative Lutheranism and, towards the second half of the nineteenth century, with Russia rising as a direct competitor in agricultural production, also of protectionist trade policy (Beck 1997).

In terms of population, the long history of Prussian state formation led to very high demographic diversity in the east of the Kingdom of Prussia (and later of the German Empire). Figure 3.2 shows an extract of a 1906 map of ethnicities in Central-Eastern Europe. The border of the East Prussian province followed the course of the Vistula river (German: Weichsel), yet shifted a few kilometers to the east, so that the city of Elblag (German: Elbing) was at this border on the West Prussian side. Within the nineteenth century borders, especially in the provinces of West Prussia, Posen and Silesia, there was a substantial Polish population, leaving the Germans a minority in many areas (especially in the countryside). East Prussia was much more homogenous, with an almost purely German population in the north, but Polish majorities in some southern areas and a slight Lithuanian majority in the Memel region (Eberhardt 2002). The complex demographic structure in the south, however, differed from the Polish areas in other German provinces. A large share of the East Prussian Poles were Evangelical Protestant Masurians, who mostly considered themselves neither Polish nor German, even though many adopted either Polish or German identity (Eberhardt 2002).

Thus, the national conflict rising in the German Empire at the end of the nineteenth century touched East Prussia to a lesser extent. The suppression of the Catholic Church through Bismarck's Kulturkampf was less critical for the Protestant Poles, and the predominance of Poles in areas where they were in the majority was usually not as substantial as in other provinces, although the imperial policy of Germanization naturally contributed to tensions in East Prussia too (Eberhardt 2002; Tilse 2011). Co-habitation of Germans and Poles resulted in extensive cultural exchange, and much of the population (especially among the Poles) was bilingual. Also, mixed marriages were quite common in the Prussian East. Many Germans and Poles could adopt either Polish or German identity, depending on what was more advantageous in specific situations (Eberhardt 2002). Thus, for example, Tilse 2011 speaks of the process of cultural transnationalism, instead of exchange or assimilation.

The situation was similarly less intense in the Memel region (in Figure 3.2, the area with the Lithuanian majority in the north of East Prussia, between the Neman river (German: Memel) and the state border). Despite the Lithuanian national identity generally having developed relatively late (Staliunas 2016), most of the conflict on the Russian side increased, while the population in Lithuania Minor (Memel region) did not actively identify themselves with a Lithuanian nation. Even the naming of the "two" nations was different in the Lithuanian language (Vareikis 2002). On top of that, Lithuanian nationalist discontent was directed against the Polish population, predominant in the south of Lithuania and in Vilnius, rather than against the Russians or Germans (Staliunas 2016; Vareikis 2002). As a result, the Lithuanians in the Memel region were much less

Figure 3.2: Ethnic groups in Central Europe in 1906



Notes: The color red denotes prevalence of German ethnic groups. Green and light green denote prevalence of Polish and Russian/Belarussian populations, respectively. Gray denotes prevalence of Lithuanian population. The bold gray line delineates the border of the German Empire in 1906. *Source:* Scobel 1906, p. 31.

eager to separate from the German Empire and unite with Lithuania Major – the part of Lithuania under Russian control (Vareikis 2002).

Border re-drawing in the 1919 Paris Conference in the aftermath of World War I was primarily based on the ethnicity principle, which resulted in huge territorial losses in West Prussia, Posen and Silesia and, later, in silent acceptance of the separation of the Memel region from East Prussia (compare also the borders of 1914 and 1933 in Figure 3.1). The very mixed demographic structure in the south of East Prussia, however, made use of ethnicity principle there virtually impossible, so the peacemakers had to take the will of the majority of the population into account (Eberhardt 2002). While the results of the plebiscite raised discontent in the reborn Polish state (Wrzesinski 1985), the outcome was an almost unchanged southern and western border of East Prussia for a further thirty years to come.[2] During the interwar time, the German share of the population in the Memel region gradually decreased, but, with generally good German-Lithuanian relations, no massive outflows occurred (Eberhardt 2002; Nikzentaitis 2002).

An important outcome of the Versailles treaty, however, was the major discontent in the German provinces about the new borders, especially those in the east, with this remaining a public issue well after the treaty came into effect (Harvey 2000). Nationalistic

[2] In fact, the western border of East Prussia was moved further west to accommodate a small part left in the east of the former West Prussia into the province.

Table 3.1: The number of refugees from East Prussia, 1944-1955

	East Prussia total	RB Königsberg	RB Gumbinnen	RB Allenstein
1944	185701	25118	134326	24886
Jan 1945	551734	305129	33585	117951
May 1945	74725	33620	13954	19084
Aug 1945	32817	11546	2250	14419
1946	23834	7997	1681	5358
1947	19924	12828	908	4154
1948	11105	9020	958	934
1949	843	408	131	217
1950	526	118	119	209
1951	111	59	29	10
1952	121	49	18	42
1953	391	116	81	151
1954	278	109	27	122
1955	210	43	29	118

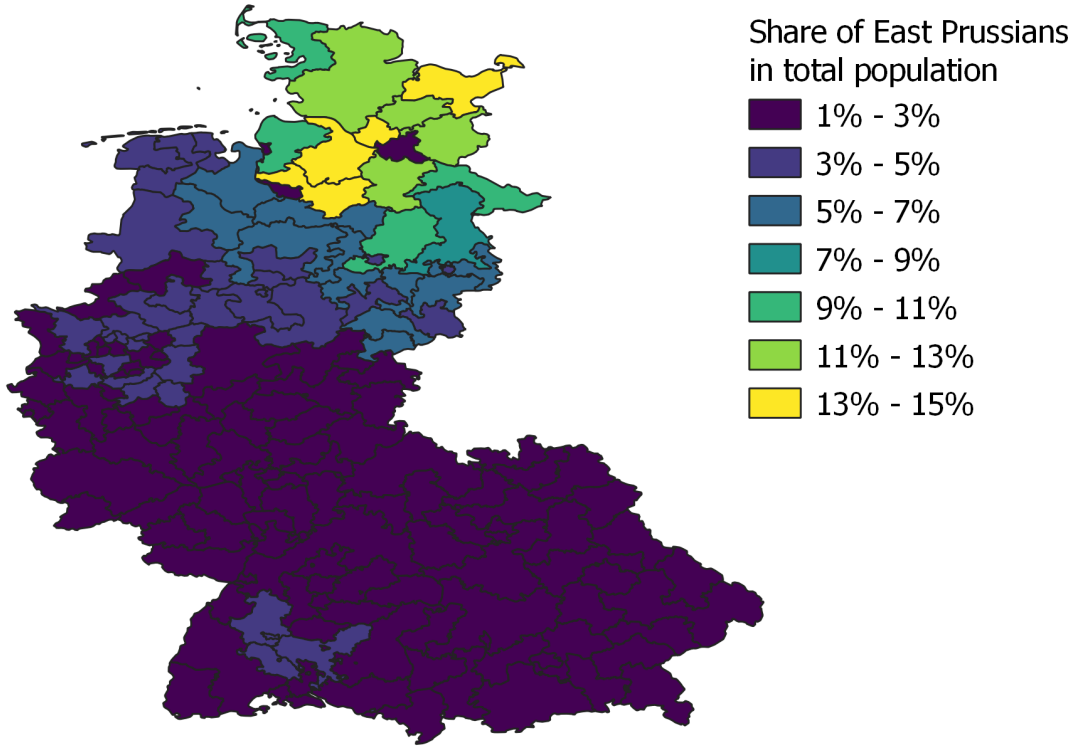
Source: Besser 2007

organizations and political parties actively used this issue in their rhetoric, especially focusing on the preservation of the "Germandom" and military and cultural vulnerabilities created by the new borders (Harvey 2000). In addition, the eastern provinces were largely agricultural, making them more prone to nationalism due to both the overall lower income level of the population (Friedrich 1937, see also Section 3.2) and the farmers' high dependence on their place of birth (Friedrich 1937). With Germany already being in a difficult economic situation as a result of after-war hyperinflation and unsettled reparations issues, the Great Depression aggravated the economic distress. East Prussia, now also in its disadvantageous position as an exclave, suffered severely from the crisis (Harvey 2000). By 1933, East Prussia became one of the major supporters of the nationalists (see also Figure G.3 in the Appendix).

The German territorial losses in the East were drastic after World War II. The rest of East Prussia first became a Soviet occupation zone and then was divided between Poland (constituting the Warmińsko-Mazurskie voivodship) and the Soviet Union (with the city of Königsberg, renamed Kaliningrad). As the Soviet army advanced into East Prussia in early 1945, the inhabitants massively fled from their homes. Those who had remained in the occupied areas east of the Oder-Neisse line until the end of the war were expelled in the next few years. The majority of Germans left East Prussia by the end of 1945, as shown in Table 3.1 for East Prussia as a whole and each of its districts (Regierungsbezirke). Especially before their expulsion by the Polish and the Soviet governments, the easiest way for people to leave East Prussia was by way of the Baltic Sea. Thus the entry points for the East Prussian expellees in the West were mostly in the north of Germany. As a result, in West Germany most of the East Prussian expellees landed in Bremen, Hamburg, and the states of Schleswig-Holstein and Lower Saxony.

The massive inflow of the population from the former eastern provinces – but also from the Soviet occupation zone – required that the West German authorities be actively involved

Figure 3.3: Share of East Prussian expellees in FRG population, 1950



Source: Authors' work. Base map: MPIDR and CGG 2011a, population data: see Table E.1

in managing this migration. They determined the places of settlement for the expellees and refugees, with local economic conditions playing a minor role – the choice was mostly driven by the availability of housing (Braun and Dwenger 2018). The expellees were then prohibited from changing their residence until 1947, after which they could only move with the permission of the authorities. The relocation ban was completely lifted in May 1949, with the result that the geographical distribution of expellees was almost the same in 1950 as in 1946 (Braun and Dwenger 2018). East Germans (including East Prussians) were quite similar to the native West German population in most socio-demographic characteristics (such as age and education) (Braun and Dwenger 2018). However, they were distinguishable, for example, by their eastern dialect and often treated as foreigners in their host settlements (Glück and Sauer 1997). The distribution of the East Prussian expellees was highly uneven across the states of West Germany, reaching up to almost 15% of the total population in some districts of Schleswig-Holstein and Lower Saxony but barely exceeding 2% in the southern states (see Figure 3.3).

Back in East Prussia, of some 1.2 million of its former German inhabitants, a total of a few thousand remained in the three new states throughout the Cold War period (Eberhardt 2002; Zyromski 1985). The demographic shock was somewhat less severe in Lithuania, since as early as 1925 only 43.5% of the population in the Memel region was German, and this share decreased even further during the interwar period (Eberhardt 2002, p. 40). Also, unlike Latvia and Estonia, Lithuania was little affected by Soviet

internal migration. Some Russian population came to the republic throughout the Soviet period, mainly as employees in the bureaucracy, military staff and technical staff working on industrialization plans. These migrants, however, remained quite dispersed across towns and in the countryside, and their total share was kept very low by the fast natural growth of the Lithuanian population (Eberhardt 2002).

In Poland and the Kaliningrad region, on the other hand, the aftermath of the war meant an almost complete exchange of population. In the new Polish Warmińsko-Mazurskie voivodship, no more than 25% of the population were pre-war residents. The huge loss was recovered mainly through in-migration from the Warsaw region and former eastern Poland (which was ceded to the Soviet Belarus after the Second World War) and largely comprised a younger population (Zyromski 1985). In the Kaliningrad region, the effect was even more devastating, as the north of East Prussia was a predominantly German region before the war. Population replacement there was complete. While loyalty to the regime of course played a role in the choice of the settlers, the choice of source regions for resettlement tended to be driven by convenience: the Russian-speaking population mostly came in from the regions of Pskov and Smolensk, and in rural areas also from the "black earth" region in Central Russia and Ukraine, which combined proximity to Kaliningrad and a relatively high population density (Diener and Hagen 2011; Eberhardt 2002).

After the fall of the Iron Curtain, migration was quite mild, relative to what happened during and after World War II. For example, in Poland, the total yearly number of newcomers in internal migration between voivodships rarely reached 2% of the incumbent population.[3] Several thousand ethnic Germans appeared in the first post-Soviet census, but, it seems, these tended to be bilingual Germans, who proclaimed themselves Polish after the war to avoid expulsion (Eberhardt 2002). Moreover, their share in the total (Polish) population was still negligible. Similarly, a few thousand ethnic Germans moved to Kaliningrad from other regions of the former Soviet Union, but here too they were too few to constitute any meaningful minority (Eberhardt 2002). Thus, one can safely assume that the current population in former East Prussia is mostly comprised of the descendants of after-war migrants. In Lithuania, however, there is still a substantial share of descendants of pre-WWI inhabitants.

3.4 Data and methodology

3.4.1 Hypotheses

While our case focuses on the legacy of one region, instead of the long-term effects of borders, the underlying question is quite similar. Our expectation is that culture is persistent. It is transmitted directly by people in their families, but also indirectly through neighborhood socialization and the choice of the socialization environment. A solid theoretical foundation for both direct and indirect channels was suggested by Bisin and Verdier 2001. Moreover, as the literature discussed in Section 3.2 suggests, not only does it take decades to level out the differences created by varying institutional settings,

[3] Authors' calculation based on official Polish statistics.

but it also takes a similarly long time to create such differences by installing varying institutional settings in the first place. More specifically, we might expect that the legacies of East Prussia are persistent and still traceable in all three states under consideration. One can apply the argument of imperial legacies, with long-lasting institutional impact, not only to differences within one modern country, but also to our opposite case. Some features of East Prussia apply more to this particular province and less to the German Empire as a whole (and even less so to the diverse regions of the Russian Empire). Thus, we may expect that:

Hypothesis 3.1. *In modern Lithuania, Poland and Russia, the regions located in former East Prussia and those located outside former East Prussia show differences in culture as captured by political preferences and entrepreneurial activity.*

At the same time, given the relative homogeneity within East Prussia, we can also expect that the regions formerly located in East Prussia are even more similar across the modern state borders if intergenerational transmission takes place:

Hypothesis 3.2. *In modern Lithuania, Poland and Russia, the regions located in former East Prussia are less different across the modern borders than regions located outside former East Prussia.*

The idea behind the persistence argument is that, even though the population structure was mixed in some areas of East Prussia, cultural assimilation (or cultural transnationalism) between the Germans and the Poles or Lithuanians, respectively, provided for a certain level of homogeneity in values. By simple historical predominance of the German population in East Prussia and through the effects of German schooling, the relations developed more in direction of Germanization of the Polish and Lithuanian population than vice versa. Thus, even if the German population moved out of the region, the remaining Germanized Poles and Lithuanians would transmit the attitudes to their descendants, and also possibly to the migrants coming into the region. In addition, the political and ideological systems were quite similar in all three states between 1945 and 1989, as Lithuania was directly a part of the Soviet Union and Poland was largely under Soviet control. Thus, for more than a half of the partitioning period, the possibilities for the three states to drift apart were limited. The divergence was more likely to unfold during the transition period.

The major argument against any persistence is, of course, the scale of the after-war demographic shock. With most of the population decimated in the regions ceded to Poland and the Soviet Union, remaining inhabitants were likely too few to transfer any values to the migrants coming to fill the demographic vacuum. If anything, they might have been more likely to assimilate with the migrants if these had any unifying value sets.[4] Given the scale of the demographic shock in the Polish and Soviet parts of East

[4] Another argument might be that the development during the Soviet era was shaped by military interests, especially in Kaliningrad as the main Soviet naval base in the Baltic Sea. While this role of Kaliningrad definitely had an impact on economic development in the region and the city of Kaliningrad was essentially a closed military area until the 1980s (Diener and Hagen 2011), the hinterland resembled the countryside in other Soviet regions, and the focus on military manufacturing can well be considered a mirror to the general disequilibrium in the Soviet economy, with excessive attention to heavy industry. Structural and regional imbalances were characteristic for all of the Soviet

Prussia, it is likely that the patterns of persistence postulated in Hypotheses 3.1 and 3.2 will not be present to the same degree in the three countries under investigation.

Furthermore, in view of the transmission channel discussed above, the follow-up question we pose is: Did East Prussian migrants affect the voting patterns in West Germany, either through their own preferences or through the reaction of the native population to the migration inflow? The inter-generational transmission of values postulated as the main channel in most studies of persistence is based on individuals' values and preferences. While being forced to leave home for a new region is certainly, among other effects, a massive psychological shock, there is little reason to believe that the values a person developed during his/her entire life would be completely reversed by this shock. Therefore, we hypothesize that the political preferences were affected in the regions where the East Prussian refugees found their new homes immediately after the war:

Hypothesis 3.3. *The regions in the FRG that hosted more East Prussian refugees were characterized by more conservative political preferences and higher support for nationalistic parties shortly after World War II.*

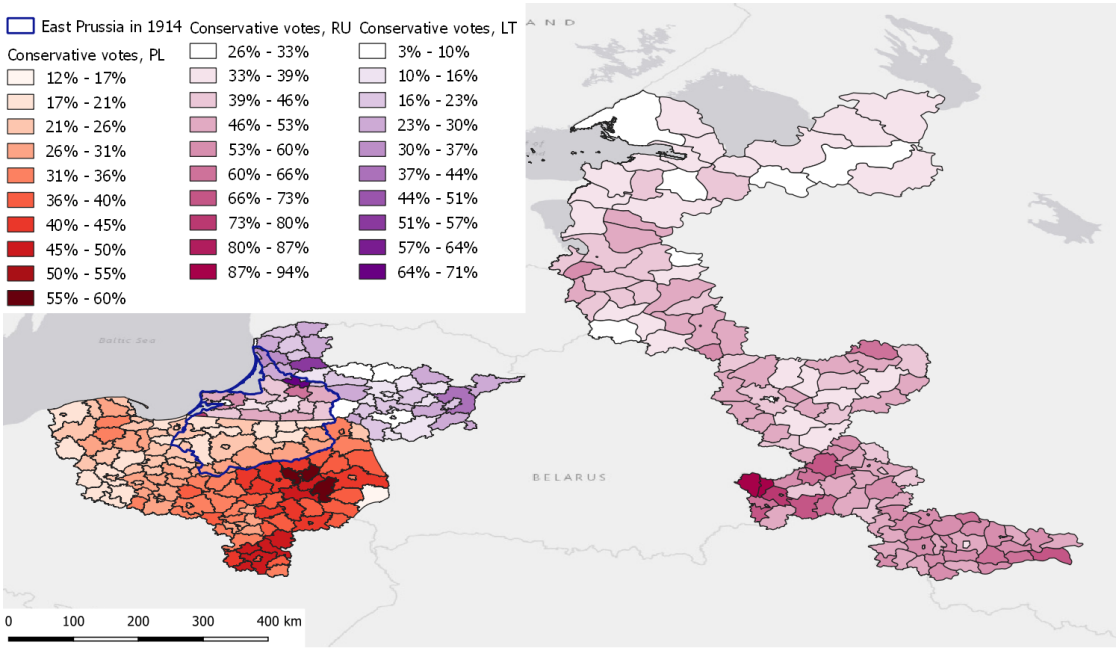
3.4.2 Data

We have collected statistical data on political and socioeconomic outcomes in Lithuania, Poland and Russia. The political data are compiled from official statistics and are treated as cross-section datasets, even though the elections do not coincide in all three countries. We, however, chose elections whose timing was closest to each other, and focused on the period before 2014, so that the outcomes are not influenced by the deteriorating East-West relations or entry of openly populist parties and candidates into governmental bodies. We consider parliamentary elections of 2011 (Poland, Russia) and 2012 (Lithuania) as a cross-sectional dataset. The data include the turnout and the number of votes for each of the parties, which we coded along the political spectrum (left-right position) and ideology (liberal, conservative, nationalist, etc.). We can, thus, calculate the share of votes that conservative or nationalistic parties received in respective elections. While we admit that survey data on political preferences would reflect the attitudes better than the political outcomes, we are limited by the geographical representation of such surveys, which is critical for an analysis of the attitude differences in this relatively small region. The election data on very low levels of administrative division are, however, readily available.

The dataset for Russia is compiled at the county (*raion*) level, which is the second level of administrative division. In addition, due to the country size, we only look at the counties within former East Prussia and in the neighboring regions of the western mainland of Russia (regions of Leningrad, Smolensk, Bryansk, Pskov and Kursk). For Lithuania and Poland, the data are available at the third level of administrative division (*gmina*) in Poland and for polling districts in Lithuania. As a result, we analyze a total of 158 counties in Russia, 2480 *gminas* in Poland and 2000 polling districts in Lithuania. Figure 3.4 offers a visualization of our political dataset for Lithuania, Poland and Russia with respect to conservative political preferences and in relation to the former East Prussian

Union (Escoe 1995). Thus, while the military importance of Kaliningrad might have contributed to the creation of new identities and values in the region, it is unlikely to be the most important factor.

Figure 3.4: Share of votes for conservative parties in parliament elections in Lithuania (2012), Poland, and Russia (2011): East Prussia and neighboring regions



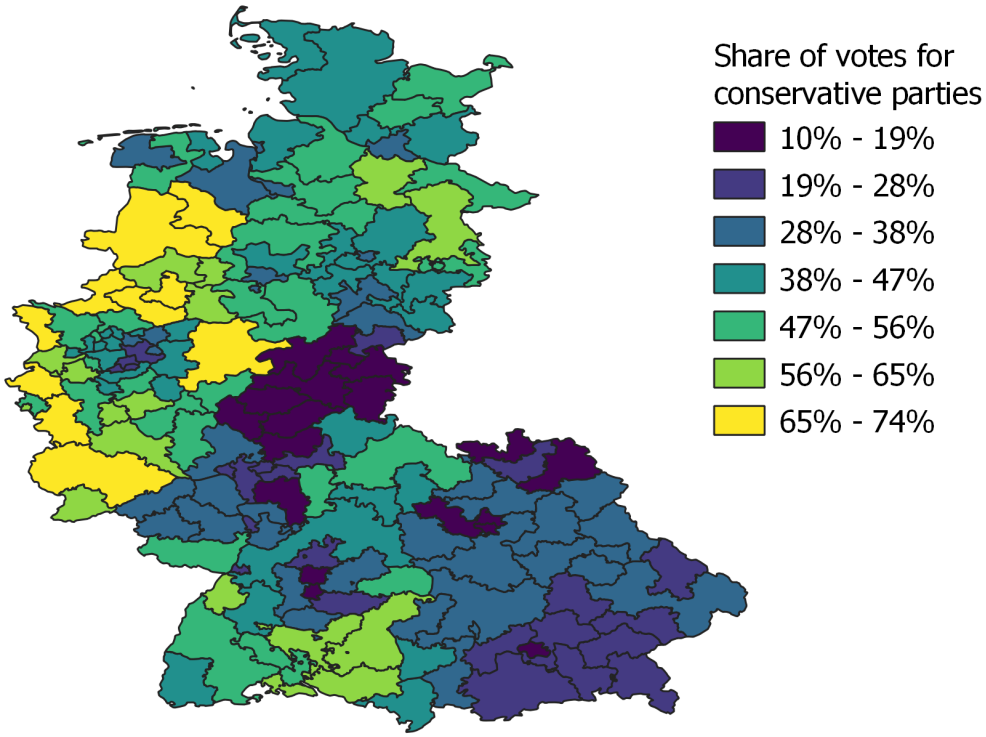
Source: Authors' work. Base map: GADM & ESRI Gray, election data: see Table E.1

region (for electoral turnout and nationalist political preferences see the respective Figures G.1 and G.2 in the Appendix).

With respect to economic outcomes, we collect data on the number of enterprises in different sectors based on the first level of the NACE Rev. 2 classification. In the case of Lithuania, the data on sectoral employment are additionally available. It is also the most geographically detailed dataset, based on a grid map of Lithuania, with the size of grid cell equal to one kilometer. With four years of data from 2015 to 2018, this generates a total of 29,963 observations. In the case of Poland, the data are based on the administrative divisions and are collected for *gminas* for the years 2015-2017, with a total of 7,434 observations. Finally, in the case of Russia, the data are collected for districts (*raions*), as is the case for electoral data too. Some districts are missing, however, and the classification of sectors was only harmonized with the NACE classification of 2017. Thus, with the two years of data, we are left with 274 observations.

In order to explore Hypothesis 3.3, we collect the data on the 1949 Bundestag election results by electoral districts, which we code along the ideologies of participating parties so as to calculate the share of votes received by conservative and nationalistic parties. To calculate the density of the East Prussian refugees, we use the number of refugees from each former eastern province in each of the West German districts (*Kreise*) reported in the population census of 1950 (Braun and Dwenger 2018). We then calculate the share of each province's migrants in the total population. To control for other socioeconomic variables and political preferences before the war, we also utilize the electoral and socioeconomic data of 1920-1933 compiled by Falter and Hänisch 1990. The *Kreise* are then aggregated

Figure 3.5: Share of votes for conservative parties in parliament elections in the FRG, 1949



Source: Authors' work. Base map: MPIDR and CGG 2011a, election data: see Table E.1

to 194 "mega-districts" to handle the geographical mismatch between the electoral and the administrative districts. Figure 3.5 offers a visualization of our political dataset for West Germany with respect to conservative political preferences (for nationalist political preferences see Figure G.4 in the Appendix).

3.4.3 Methodology

Our approach is mainly based on a regression discontinuity design. The underlying assumption is that the border of former East Prussia is exogenous. Throughout the formation of the Prussian state, the borders were determined by warfare and negotiations with Poland, Lithuania and later the new neighbor, Russia, and often cut through historical ethnic areas. After World War II, the border between Poland and the Soviet Union did not follow any ethnic or economic criteria either, but rather was determined by the balance of power and strategic military considerations. The only case where a border was more or less determined by ethnic composition of the area's population, was the separation of the Memel region through quiet acceptance of Lithuanian annexation. However, as we showed in the previous section, the German population in the region was quite substantial, and the Lithuanian population was Germanized. The argument of inherent national unity with Lithuania Major was thus questionable. Lithuanian influence

in the Memel region was indeed weaker than the German in the interwar period (Vareikis 2002). In addition, the new border quite conveniently followed the course of the Neman River, although there were also some predominantly Lithuanian areas south of the river, which remained in East Prussia (see also Figure 3.2). Therefore, we can safely assume that the borders of both East Prussia and the successor states were drawn exogenously.

Robust RDD in Poland & Lithuania

The absence of territorial continuity between the Kaliningrad region and Russia does not allow us to perform the robust regression discontinuity design as introduced by Calonico, Cattaneo, Farrell, et al. 2017. We use Calonico, Cattaneo, and Titiunik 2014; Calonico, Cattaneo, and Titiunik 2015 only for the Polish and Lithuanian data. Thus, our baseline regression is:

$$Y_i = \alpha_i EastPrussia_i + \beta_1 Distance + \beta_2 EastPrussia_i Distance_i + \beta_3 X_i + \epsilon_i, \quad (3.1)$$

where Y_i is the electoral or economic outcome: turnout, share of votes for conservatives, share of votes for nationalists and number of economic entities in an economic sector. X_i is the set of additional controls such as city dummy, altitude, latitude and longitude. *Distance* denotes the distance from the centroid of the territorial unit to the East Prussian border, which is the forcing variable in our model. The interaction term $EastPrussia^* Distance$ shows that the distance of each territorial unit to the East Prussian border varies with its historical attachment to East Prussia.

CEM in Russia

We correct for the territorial discontinuity between the Kaliningrad region and the rest of Russia by introducing Coarsened Exact Matching (CEM) to compare economic and political outcomes between Kaliningrad, on the one hand, and neighboring Russian regions such as Pskov, Leningrad, Smolensk, Briansk and Kursk, on the other (Datta 2015; Iacus et al. 2009). Thus, the proposed baseline model is the following:

$$L_1(f, g) = \frac{1}{2} \sum_{l_1,\ldots,l_k} |f_{l_1,\ldots,l_k} - g_{l_1,\ldots,l_k}|$$

where $L_1 \in [0, 1]$ is the measure of multivariate imbalance, $f_{(l_1,\ldots,l_k)}$ denotes the relative multivariate frequency distributions of treatment units and $g_{(l_1,\ldots,l_k)}$ denotes the relative multivariate frequency distributions of control units in k-dimensional space. Furthermore, if $TE_i = Y_i(T_i = 1) - Y_i(T_i = 0)|X_i$, where TE_i is the treatment effect, Y_i the outcome variable, T_i the treatment variable and X_i the set of pre-treatment covariates, then we compute the local sample average treatment effect such that

$$LSATT = \frac{1}{m_T} \sum_{i \in T^m} TE_i \quad (3.2)$$

where m_T is the number of matched treated units and T^m the subset of matched treated units (ibid.). The main advantage of the CEM method is that it does not require the common pre-treatment trends for both treatment and control observations, and there is no data extrapolation (Datta 2015).

Similarities in East Prussia

To investigate Hypothesis 3.2, we use an approach similar to the regression discontinuity design for a pooled dataset on Lithuania, Poland and Russia. Because of this pooling, however, we have to account for the differences between the countries. We do so by introducing the country dummies and differentiating between the effects of East Prussia in the three countries:

$$\begin{aligned} Y_i = \ & \alpha + \beta_1 LT_i + \beta_2 PL_i + \beta_3 LTEP_i + \beta_4 PLEP_i + \beta_5 RUEP_i + \\ & \gamma_1 Distance + \gamma_2 EP_i^* Distance_i + \delta X_i + \epsilon_i, \end{aligned} \tag{3.3}$$

where Y_i is the electoral outcome: turnout, share of votes for conservatives and share of votes for nationalist parties. LT_i and PL_i are 1 for Lithuania and Poland, respectively, and 0 otherwise (Russia serves as a base), $LTEP_i$, $PLEP_i$ and $RUEP_i$ are 1 for regions in Lithuania, Poland and Russia, respectively, located in former East Prussia and 0 otherwise. Note that they are additive to the overall country effects, meaning the cumulative effect of being, for example, in Lithuania in former East Prussia is $\alpha + \beta_1 + \beta_3$. Similarly, one can also calculate the effects for the other two countries. X_i is the set of additional controls, such as city dummy, latitude and longitude. *Distance* denotes the distance from the centroid of the territorial unit to the East Prussian border, and the interaction term $EP^*Distance$ shows that the distance of each territorial unit to the East Prussian border varies with its historical attachment to East Prussia. We only perform this exercise for the political outcomes, as the economic outcomes are not directly comparable across countries due to the differences in collecting the statistics.

In terms of Hypothesis 3.1, such a regression should deliver results similar to the RDD and CEM approaches discussed above. Its advantage, however, is that such pooling and the use of interaction terms allows us to test statistically if the effects we observe for particular groups (e.g. regions in former East Prussia) are different.

Political preferences in the FRG

We use the geographic variation in the number of East Prussian refugees in the FRG to identify the effect of the East Prussian migration on the voting results in the first parliament elections after the end of World War II. We consider this variation exogenous for, as discussed above, the settlement pattern was determined by geography and the settlement decisions of authorities, mostly dependent on housing availability (hence, indirectly on war damage). We run robust OLS regressions of the vote shares of conservative and nationalistic parties on the share of East Prussian refugees in the total population, controlling for other refugees, large cities, age and gender structure, religion, prewar electoral patterns, and economic structure of the respective districts. Based on data

availability for the control variables, some further geographic aggregation was necessary, and so in most specifications we end up with 162 regions to analyze. The basic regression is:

$$Y_i = \alpha + \beta_1 ShareEastPrussia_i + \beta_2 X_i + \epsilon_i, \tag{3.4}$$

where Y_i is the electoral outcome in 1949, $ShareEastPrussia_i$ is the share of East Prussian refugees in the host region i, and X_i is the set of control variables outlined above.

With possible concerns that, despite the way most of the expellees left their home province and how they were settled by the authorities, their distribution in West Germany might still be endogenous, we also check for the robustness of our results with an IV regression, using latitude as an instrument for the share of East Prussians in the total population. The latitude reflects the evacuation routes over the Baltic Sea and the respective settlements primarily in the North of Germany very well and, based on the Montiel-Pflueger test, appears to be a strong instrument for the geographical distribution of the East Prussian expellees.

Finally, since spatial data are being analyzed, there is a possible concern that any results we find are not driven by the independent variables but by the spatial autocorrelation between the regions. As Kelly 2019 suggests, spatial autocorrelation can exaggerate the t-statistics and the Conley procedure does not fully correct this, mostly because either only one neighboring region is given non-zero weight in the adjustment or the cutoff radius for the zero weight is set very low (and so still too few regions get non-zero weights). Taking this into account, we test for spatial dependence in our data and check the robustness of our OLS results with a spatial autoregressive model (SAR), using the inverse distance to weight the covariance matrix and not setting any cut-off on the weights. Thus, we assume decreasing but non-zero mutual influence of the regions that are further apart.

3.5 Persistence in former East Prussia

We first concentrate on economic and political outcomes at the historical borders of East Prussia within modern-day Poland and Lithuania. It is obvious that the internal Polish-East Prussian border reveals no statistical significance when it comes to several areas of entrepreneurial activity (see Table F.3 and Figure G.7 in the Appendix). Nevertheless, political outcomes in what used to be East Prussia within the boundaries of contemporary Poland differ significantly from respective political outcomes on the Polish side of the East Prussian–Polish border. Table 3.2 and Figure 3.6 show the estimation results for conservative, liberal-conservative and nationalistic parties (the results for all parties can be viewed in Table F.1 and Figure G.5 in the Appendix). As these results indicate, in the territories of pre-war Poland, the conservative party *Prawo i Sprawiedliwośc* (Law and Justice) vote share is higher by a mean difference of 4.6 percentage points, statistically significant at the 5% level with a bias-corrected confidence interval (CI) and at the 10% level with a robust CI. Similarly, in the territories of pre-war Poland, the

Table 3.2: RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia

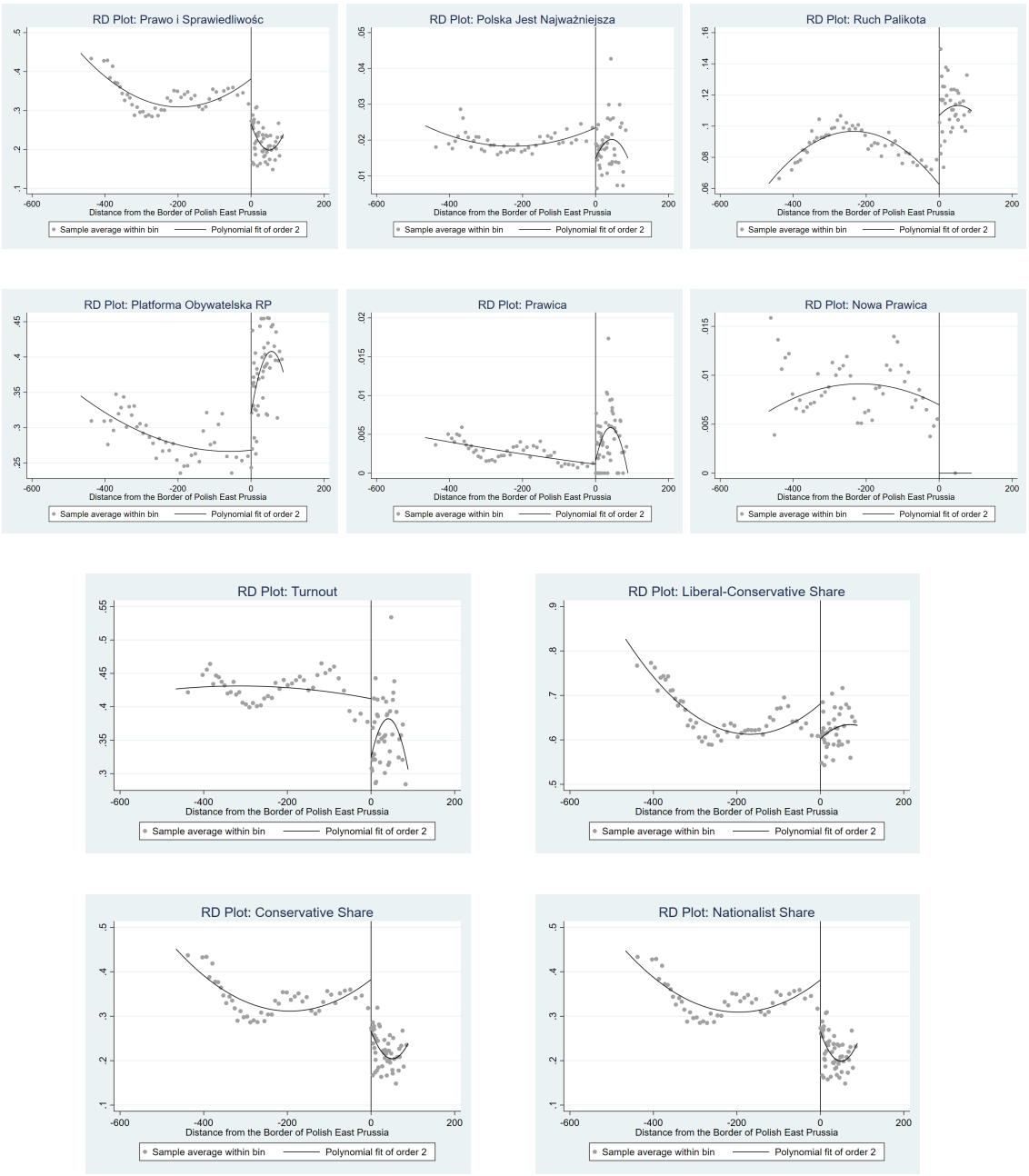
Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Prawo i Sprawiedliwośc	Conventional	−0.061***	0.018	−3.42	−0.096	−0.026	2479
	Bias-Corrected	0.046**	0.018	−2.59	−0.081	−0.011	2479
	Robust	−0.046*	0.028	−1.68	−0.100	0.008	2479
Polska Jest Najważniejsza	Conventional	−0.009*	0.005	−1.71	−0.018	0.001	2479
	Bias-Corrected	−0.014***	0.005	−2.72	−0.023	−0.004	2479
	Robust	−0.014	0.009	−1.46	−0.032	0.005	2479
Platforma Obywatelska RP	Conventional	0.068***	0.026	2.60	0.017	0.119	2479
	Bias-Corrected	0.058**	0.026	2.22	0.007	0.109	2479
	Robust	0.058	0.040	1.47	−0.019	0.136	2479
Prawica	Conventional	0.001	0.001	1.38	0.000	0.002	2479
	Bias-Corrected	0.001	0.001	1.19	−0.001	0.002	2479
	Robust	0.001	0.001	0.72	−0.002	0.003	2479
Nowa Prawica	Conventional	−0.004***	0.001	−3.43	−0.007	−0.002	2479
	Bias-Corrected	−0.006***	0.001	−4.66	−0.008	−0.003	2479
	Robust	−0.006***	0.002	−3.04	−0.010	−0.002	2479
Turnout	Conventional	−0.038**	0.016	−2.37	−0.069	−0.006	2479
	Bias-Corrected	−0.035**	0.016	−2.22	−0.066	−0.004	2479
	Robust	−0.035	0.022	−1.62	−0.078	0.007	2479
Conservative Share	Conventional	−0.060***	0.018	−3.32	−0.096	−0.025	2479
	Bias-Corrected	−0.045**	0.018	−2.51	−0.081	−0.010	2479
	Robust	−0.045	0.028	−1.63	−0.100	−0.009	2479
Liberal-Conservative Share	Conventional	−0.005	0.020	−0.25	−0.043	0.033	2479
	Bias-Corrected	−0.007	0.020	−0.34	−0.045	0.032	2479
	Robust	−0.007	0.029	−0.23	−0.063	0.050	2479
Nationalist Share	Conventional	−0.061***	0.018	−3.42	−0.096	−0.026	2479
	Bias-Corrected	−0.046**	0.018	−2.59	−0.081	−0.011	2479
	Robust	−0.046*	0.028	−1.68	−0.100	0.008	2479

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km.

liberal-conservative party *Polska Jest Najważniejsza* (Poland Comes First) vote share is higher by a mean difference of 1.4 percentage points, statistically significant at the 1% level with a bias-corrected CI, that of the labor party *Polska Partia Pracy - Sierpień 80* higher by a mean difference of 0.4 percentage points, statistically significant at the 5% level with a bias-corrected CI, and that of the socialist party *Nasz Dom Polska* (Our Home Poland) higher by a mean difference of 0.2 percentage points, statistically significant at the 5% level with a bias-corrected CI.

Furthermore, aggregated conservative and nationalist vote shares in the territories of pre-war Poland are also higher than those in former East Prussia by an average difference of 4.5 and 4.6 percentage points, respectively, which is statistically significant at the 5% level with a bias-corrected CI. The vote share of the right-wing party *Nowa Prawica* (Congress of the New Right) is also higher in the pre-war territories of Poland and the same observation holds for electoral turnout, an indication of a politically mobilized society. In contrast, the vote share of the party *Sojusz Lewicy Demokratycznej* (Democratic Left

Figure 3.6: RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia



Alliance) is higher in former East Prussia by an average difference of 2.6 percentage points, which is statistically significant at the 1% level with both bias-corrected and robust CIs. Similarly, the vote share in former East Prussia is higher for the liberal party *Ruch Palikota* by a mean difference of 1.9 percentage points, statistically significant at the 1% level with a bias-corrected CI and at the 10% level with a robust CI. The vote share of the liberal-conservative *Platforma Obywatelska RP* is higher in "Polish East

Prussia" by a mean difference of 5.8 percentage points, statistically significant at the 5% level with a bias-corrected CI. Several robustness checks including covariates such as city dummy, latitude and longitude or an increase of the border bandwidth from 60 km to 100 km do not change the baseline findings reported for both economic and political outcomes (see Tables F.11-F.13 in the Appendix).

Overall, the East Prussian lands of modern Poland reveal lower levels of political conservatism and support for nationalistic parties compared to the pre-war Polish territories on the other side of the border. "Polish East Prussia" appears to be more progressive rather than conservative in terms of electoral results. This, in fact, is the opposite of Hypothesis 3.1. The former East Prussian territories do not show a higher level of economic institutions either, as indicated by entrepreneurship in the sectors of information and communication, real estate activities, professional and scientific activities, financial and insurance activities, and other services.

When we evaluate the internal East Prussian border in Lithuania, we find that, in terms of economic outcomes, only enterprises offering professional and scientific services appear to be more developed on the East Prussian side of the border, with a mean difference of 2.289, statistically significant at the 1% level both with bias-corrected and robust CIs (see Table F.4 in the Appendix). Financial and insurance enterprises do not reveal any significant difference in terms of their frequency across the historical border, whereas real estate enterprises, information & communication as well as other services are significant in the opposite direction, i.e. reveal a discontinuity in favor of the pre-war Lithuanian territories (see also Table F.4 in the Appendix). With respect to political outcomes, the results are presented in Table 3.3 and Figure 3.7 for conservative, liberal-conservative and nationalistic parties and in Table F.2 and Figure G.5 in Appendix for all parties in the sample. We find that, in the pre-war Lithuanian territories, the *Social Democratic Party* vote share is higher by an average difference of 6.2 percentage points, statistically significant at the 5% level both with bias-corrected and robust CIs. A similar observation holds for the *Homeland Union* and the *Liberals Movement*; the mean difference is 3.9 and 3.4 percentage points and is statistically significant at the 1% and 5% levels, respectively.

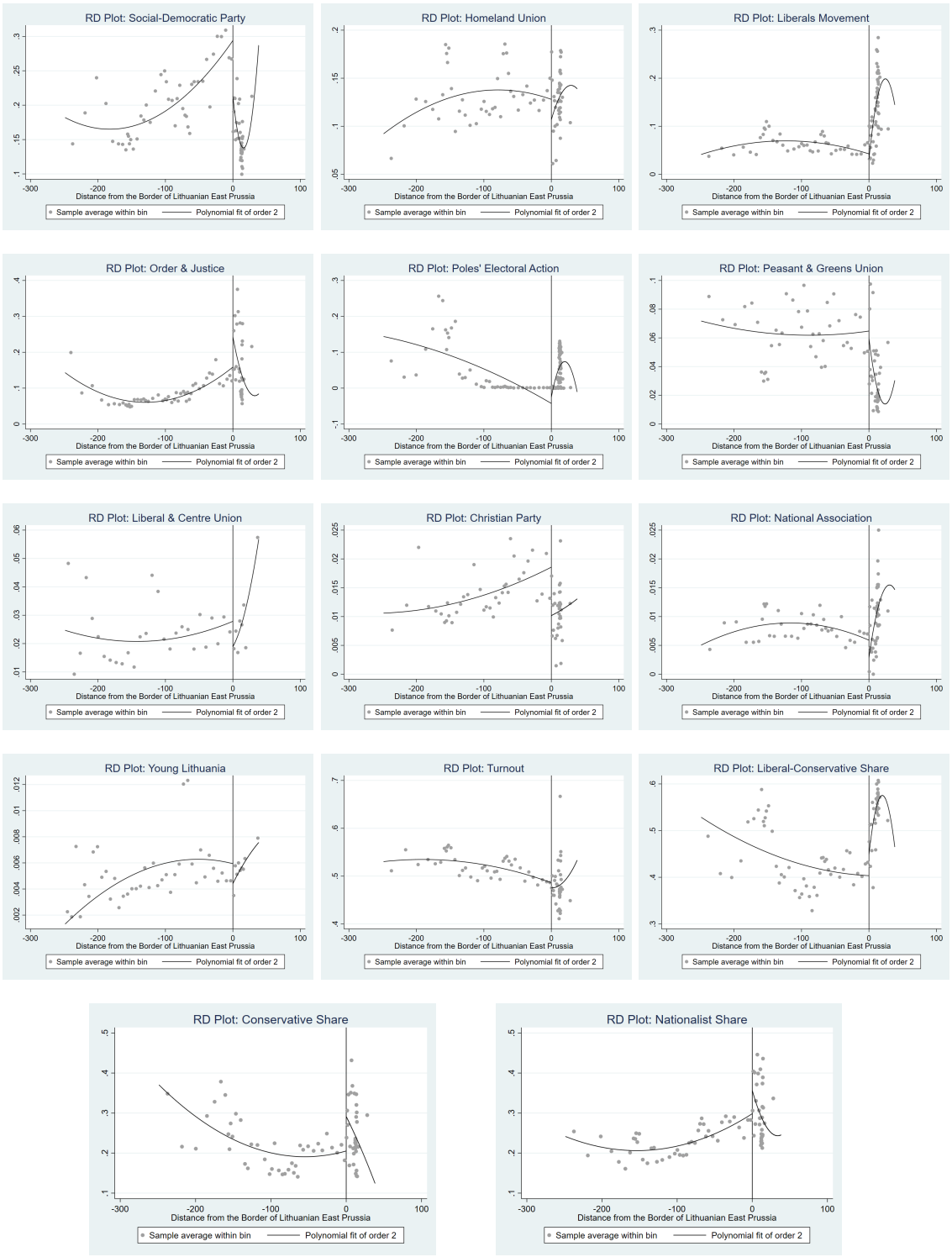
Moreover, vote shares of parties such as *The Way of Courage*, *Order and Justice*, *Poles' Electoral Action*, *Socialist People's Front*, *National Association* and the *Emigrants' Party* are significantly higher on the pre-war Lithuanian side of the border. However, the *Labor Party* exhibits an average difference of 6.1 percentage points in favor of East Prussian territories in Lithuania, which is statistically significant at the 5% level with a bias-corrected CI and at the 10% level with a robust CI. Similarly, the conservative vote share is higher in the former East Prussian territories of Lithuania by a mean difference of 10.6 percentage points, statistically significant with a bias-corrected CI at the 1% level and with a robust CI at the 5% level. The nationalist vote share is also higher in the same direction by an average difference of 8.4 percentage points, statistically significant at the 1% level with a bias-corrected CI and at the 5% level with a robust CI. We introduce several robustness checks here as well by changing the border bandwidth from 60 km to 100 km and by introducing covariates in the robust RDD such as city dummy, city distance, latitude and longitude (see Tables F.17-F.19 in the Appendix). Our initial results (Tables 3.3 and F.2 as well as Figures 3.7 and G.6) are reinforced.

Table 3.3: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Homeland Union	Conventional	-0.028^{***}	0.011	−2.65	−0.049	−0.007	2000
	Bias-Corrected	-0.039^{***}	0.011	−3.68	−0.060	−0.018	2000
	Robust	-0.039^{***}	0.013	−2.91	−0.067	−0.013	2000
Liberals Movement	Conventional	−0.0004	0.013	−0.04	−0.027	0.026	2000
	Bias-Corrected	-0.034^{**}	0.013	−2.57	−0.060	−0.008	2000
	Robust	-0.034^{**}	0.017	−2.03	−0.067	−0.001	2000
Order & Justice	Conventional	-0.104^{***}	0.037	2.79	0.031	0.177	2000
	Bias-Corrected	-0.128^{***}	0.037	3.42	0.055	0.201	2000
	Robust	-0.128^{***}	0.044	2.87	0.040	0.214	2000
Poles' Electoral Action	Conventional	-0.004^{***}	0.002	−2.54	−0.007	−0.001	2000
	Bias-Corrected	-0.027^{***}	0.002	−17.45	−0.030	−0.024	2000
	Robust	-0.027^{***}	0.002	−11.70	−0.032	−0.023	2000
Peasant & Greens Union	Conventional	−0.001	0.012	−0.10	−0.025	0.023	2000
	Bias-Corrected	0.011	0.012	0.91	−0.013	0.035	2000
	Robust	0.011	0.016	0.70	−0.020	0.042	2000
Liberal & Centre Union	Conventional	−0.008	0.005	−1.59	−0.017	0.002	2000
	Bias-Corrected	−0.007	0.005	−1.37	−0.016	0.003	2000
	Robust	−0.007	0.006	−1.15	−0.018	0.005	2000
Christian Party	Conventional	-0.006^{**}	0.003	−2.06	−0.011	0.000	2000
	Bias-Corrected	-0.005^{*}	0.003	−1.86	−0.010	0.000	2000
	Robust	−0.005	0.003	−1.47	−0.012	0.002	2000
National Association	Conventional	−0.003	0.002	−1.52	−0.006	0.001	2000
	Bias-Corrected	-0.004^{**}	0.002	−2.35	−0.008	−0.001	2000
	Robust	-0.004^{*}	0.002	−1.78	−0.009	0.000	2000
Young Lithuania	Conventional	-7.0×10^{-6}	0.001	−0.01	−0.002	0.002	2000
	Bias-Corrected	4.3×10^{-5}	0.001	0.04	−0.002	0.002	2000
	Robust	4.3×10^{-5}	0.001	0.03	−0.002	0.003	2000
Turnout	Conventional	−0.011	0.019	−0.59	−0.049	0.026	2000
	Bias-Corrected	−0.013	0.019	−0.67	−0.051	0.025	2000
	Robust	−0.013	0.024	−0.54	−0.060	0.034	2000
Conservative Share	Conventional	0.093^{**}	0.036	2.58	0.023	0.164	2000
	Bias-Corrected	0.106^{***}	0.036	2.94	0.036	0.177	2000
	Robust	0.106^{**}	0.042	2.53	0.024	0.189	2000
Liberal-Conservative Share	Conventional	0.057^{*}	0.030	1.88	−0.002	−0.660	2000
	Bias-Corrected	0.026	0.030	0.87	−0.033	0.085	2000
	Robust	0.026	0.037	0.72	−0.046	0.098	2000
Nationalist Share	Conventional	0.073^{**}	0.034	2.13	0.006	0.140	2000
	Bias-Corrected	0.084^{***}	0.034	2.45	0.017	0.151	2000
	Robust	0.084^{**}	0.042	2.01	0.002	0.166	2000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km.

Figure 3.7: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia



The Lithuanian border confirms our first hypothesis in terms of the dynamics of East Prussian persistence. Political, and to a lesser extent economic, outcomes suggest the presence of an East Prussian legacy that favors more advanced economic institutions and higher levels of political conservatism and support for nationalistic parties. Rather than offering a linear narrative of post-imperial persistence, what we find is that the extent of the demographic shock may be a powerful predictor of long-run persistence. While populations usually follow the path of a defeated army and evacuate territories that are conceded to the rival military adversary as a result of an international truce or treaty, what is crucial is the prior existence of ethnic and linguistic diversity in the province or territory conceded, its prior sectoral and resource structure, and the degree of violence of the population transfer *per se*. This is why Poland deviates much more significantly from the hypothesis of East Prussian persistence *in situ* than Lithuania.

We now turn to the case of Russia. As already mentioned above, the matching algorithm CEM allows us to coarsen the values of the covariates with the purpose of equalizing the number of treated and control units within each stratum of the covariates (Datta 2015; Iacus et al. 2009). As shown in Tables 3.4-3.5 as well as in Tables F.5-F.6, we run two different matching exercises, one with city dummy and altitude as the set of covariates (Match I) and another one with an augmented set of covariates including the distance to the Russian border (Match II). As Figure 3.8 indicates, coarsening includes many more strata in the second rather than in the first matching model. The same observation holds for economic outcomes (see Figure G.9). When it comes to political outcomes compared between Kaliningrad and the neighboring – and territorially discontinuous – Russian regions, Table 3.4 reports 17 treated units matched to 40 control units with a post-matching multivariate imbalance of $1.874^{*}10^{-16}$. Similarly, Table 3.5 reports 13 treated units matched to 19 control units with a post-matching imbalance of 0.25. When we evaluate comparative economic outcomes with the CEM algorithm, we find that the degree of matching efficiency between the treated and control units is lower. Table F.5 reports 38 treated units matched to 64 control units with a post-matching multivariate imbalance of 0.232, while Table F.6 shows 28 treatment units matched to 20 control units with a post-matching multivariate imbalance of 0.679.

Figure 3.8: CEM weights vs. strata (political outcomes)

(a) Match I (b) Match II

Table 3.4: Political matching results for Russian East Prussia (Kaliningrad) coarsening, Match I

		Treated	Control
Number of strata:	15		
Number of matched strata:	5		
All		22	136
Matched		17	40
Unmatched		5	96
Multivariate imbalance measure:	L1 = 1.874×10^{-16}		

Univariate imbalance measures:

Variable	L1	Mean	Min	25%	50%	75%	Max
City	1.00×10^{-17}	-2.10×10^{-17}	0	0	0	0	0
Altitude	1.80×10^{-16}	-2.244	16	-6	0	-6	-17

Table 3.5: Political matching results for Russian East Prussia (Kaliningrad) coarsening, Match II

		Treated	Control
Number of strata:	53		
Number of matched strata:	7		
All		22	136
Matched		13	16
Unmatched		9	120
Multivariate imbalance measure:	L1 = 0.25		

Univariate imbalance measures:

Variable	L1	Mean	Min	25%	50%	75%	Max
City	0	0	0	0	0	0	0
Altitude	0.026	-4.276	-2	-11	-5	-11	-17
Distance to the Russian border (rescaled)	0.083	-0.080	0.15	-0.067	-0.1	0.067	-0.567

Table 3.6 summarizes the political effect of East Prussia for two different matching models. Assuming constant treatment across strata, we estimate the local sample average treatment effect on the treated units (LSATT) both for Match I and Match II. We find that, for United Russia, there is an increase of 7.7 percentage points in Match I, statistically significant at the 1% level, while there is an increase of 5.4 percentage points in Match II, statistically significant at the 10% level. In other words, the share of votes given to the United Russia party is higher in former East Prussia (the Kaliningrad region). For the party Patriots of Russia, there is a statistically significant increase of 0.5 percentage points at the 1% level. The conservative vote share is also significantly higher

Table 3.6: Political estimates of East Prussian impact in Russia (Kaliningrad)

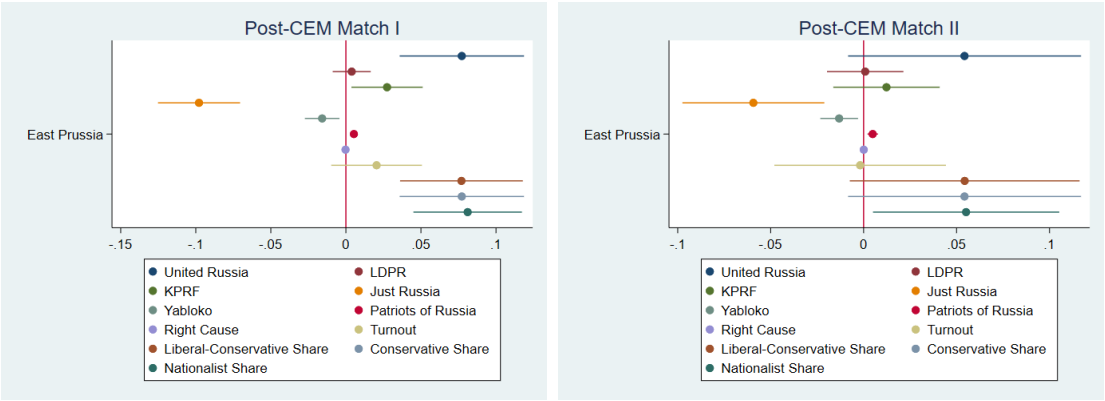
Outcome	Matching Model	Coefficient	Std. Err.	t	N	R-squared
United Russia	Match I	0.077***	0.021	3.74	57	0.202
	Match II	0.054*	0.031	1.78	29	0.105
LDPR	Match I	0.004	0.006	0.62	57	0.007
	Match II	0.001	0.010	0.09	29	0.000
CPRF	Match I	0.027**	0.012	2.32	57	0.089
	Match II	0.012	0.014	0.88	29	0.028
Just Russia	Match I	−0.098***	0.014	−7.17	57	0.483
	Match II	−0.059***	0.019	−3.19	29	0.274
Yabloko	Match I	−0.016***	0.006	−2.75	57	0.121
	Match II	−0.013**	0.005	−2.65	29	0.207
Patriots of Russia	Match I	0.005***	0.001	5.49	57	0.354
	Match II	0.005***	0.001	3.60	29	0.324
Right Cause	Match I	−0.0002	0.000	−0.54	57	0.005
	Match II	0.0001	0.001	0.17	29	0.001
Turnout	Match I	0.020	0.015	1.36	57	0.032
	Match II	−0.002*	0.022	−0.08	29	0.000
Liberal-Conservative Share	Match I	0.077***	0.020	3.77	57	0.206
	Match II	0.054*	0.030	1.81	29	0.108
Nationalist Share	Match I	0.081***	0.018	4.50	57	0.269
	Match II	0.055**	0.024	2.26	29	0.159
Conservative Share	Match I	0.077***	0.021	3.74	57	0.202
	Match II	0.054*	0.031	1.78	29	0.105

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively.

in the Kaliningrad region than in the neighboring regions of Russia in our sample: in Match I there is an increase of 7.7 percentage points, statistically significant at the 1% level, and in Match II an increase of 5.4 percentage points, statistically significant at the 10% level. The results of the conservative vote appear to be driven by the United Russia vote share. The nationalist vote share is higher in Kaliningrad as well, producing an increase of 8.1 percentage points in Match I, statistically significant at the 1% level, and an increase of 5.5 percentage points in Match II, statistically significant at the 5% level. The liberal-conservative vote share shows the same results as the conservative share.

Similarly, more progressive and left-wing parties receive higher vote shares in western Russian regions of our sample other than in Kaliningrad. The vote shares of Just Russia and Yabloko exhibit a decrease of 9.8 percentage points and 1.6 percentage points, respectively, in Match I, statistically significant at the 1% level for both parties. In Match II, we observe a decrease of 5.9 percentage points for Just Russia and of 1.3 percentage points for Yabloko, statistically significant at the 1% and 5% levels, respectively. Figure 3.9 visualizes Table 3.6 both for Match I and Match II. Table F.7 and Figure G.10 present matching results and LSATT estimates for economic outcomes in Kaliningrad and its control western Russian regions. Our results show no pattern of persistence in terms of institutions-intensive entrepreneurial activity in Kaliningrad. The respective LSATT estimates of information and communication, real estate activities, professional and scientific activities, financial and insurance activities, and other services are either statistically insignificant or point in the opposite direction (real estate activities). Hence,

Figure 3.9: Post-CEM regressions (political outcomes)



(a) Match I (b) Match II

we observe that the political legacy of East Prussia persists in Kaliningrad in the form of conservative politics and a significant role for nationalist politics, while there is no persistence of German economic institutions in entrepreneurial activity.

Turning to our analysis of the pooled sample (as in specification (3.3)), the results are presented in Table 3.7 for the full sample and a sample in which Russia is excluded (denoted LTPL). We experimented with excluding Russia due to a possible objection that the Russian ("hybrid-democratic") political system is not comparable to that of Lithuania and Poland. Exclusion of Russia, however, has no significant effect on the other results. Similar to the RDD and the CEM results, regions in East Prussia tend to vote less conservatively and nationalistically in Poland and more conservatively and nationalistically in Lithuania and Russia. All these effects are highly significant, except for nationalistic voting in Russia. Interestingly, Lithuania and Poland in general tend to vote more conservatively than Russia. Lithuanian regions also tend to lend less support to nationalistic parties than Russia and less support to both conservative and nationalistic parties than Poland.

At the bottom of Table 3.7, we also measure pairwise if Lithuania, Russia and Poland are statistically distinguishable inside East Prussia and, for the full sample, if Lithuania and Poland are distinguishable outside East Prussia. The latter measure is also repeated through the Lithuanian country effect in the smaller sample, and the coefficients are quite close. As could be expected, all three countries are significantly different from each other both inside and outside former East Prussia. An interesting result, however, is that the difference between Lithuania and Poland in East Prussia seems to be less in absolute terms than that outside East Prussia. The last three lines of Table 3.7, therefore, also report pairwise the difference in absolute disparities between the modern countries within versus outside East Prussia. It is important to note that we only compare the *magnitude* of the disparity in this case, even if East Prussian and non-East Prussian effects go in different directions. The negative coefficient means that the country difference within East Prussia is smaller than outside it, whereas the positive coefficient indicates a larger difference within East Prussia. Indeed, we find Lithuania and Poland are more similar within East Prussia than outside it with respect to conservative and nationalistic voting.

Table 3.7: Estimation results for East Prussia, pooled sample

Sample	Full	Full	Full	LTPL	LTPL	LTPL
Dependent variable	turnout	cons.	national.	turnout	cons.	national.
	(1)	(2)	(3)	(4)	(5)	(6)
LT	0.004	0.052**	−0.283***	0.070**	−0.253***	−0.171***
	(0.014)	(0.027)	(0.019)	(0.006)	(0.010)	(0.010)
PL	−0.077***	0.266***	−0.134			
	(0.017)	(0.034)	(0.025)			
LTEP	−0.053***	0.110***	0.099***	−0.052***	0.125***	0.103***
	(0.009)	(0.015)	(0.013)	(0.009)	(0.015)	(0.013)
PLEP	−0.054***	−0.070***	−0.088***	−0.061***	−0.083***	−0.101***
	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
RUEP	0.052*	0.443***	0.180			
	(0.031)	(0.044)	(0.032)			
City	0.053***	−0.021***	−0.002	0.057***	−0.020***	0.001
	(0.004)	(0.008)	(0.004)	(0.004)	(0.008)	(0.005)
Dist. to EP border (tkm)	0.071***	0.285***	0.065***	0.083***	0.348***	0.082***
	(0.012)	(0.024)	(0.020)	(0.013)	(0.024)	(0.021)
Dist. to EP border (EP) (tkm)	−0.032	−1.493***	−1.187***	0.106	−1.358***	−1.004***
	(0.092)	(0.290)	(0.248)	(0.311)	(0.256)	(0.240)
Location	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4637	4637	4637	4479	4479	4479
R-squared	0.366	0.244	0.341	0.354	0.221	0.200
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000
LT vs RU in EP	−0.100***	−0.280***	−0.364***			
PL vs. RU in EP	−0.182***	−0.247***	−0.401***			
LT vs. PL in EP	0.082***	−0.033**	0.037***	0.079***	−0.045***	0.033***
LT vs. PL outside EP	0.081***	−0.214***	−0.150***			
Δ LT vs. RU in EP / outside EP	0.096***	0.227***	0.081***			
Δ PL vs. RU in EP / outside EP	0.105***	−0.020	0.268***			
Δ LT vs. PL in EP / outside EP	0.001	−0.180***	−0.112***	0.009	−0.208***	−0.138***

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. LT = Lithuania, PL = Poland, RU = Russia, EP = East Prussia, LTEP = Lithuania in East Prussia, PLEP = Poland in East Prussia, RUEP = Russia in East Prussia.

They are, however, indistinguishable in this regard with respect to turnout. At the same time, both Lithuania and Poland seem to be more different from Russia within East Prussia than they are outside it. Thus, our results both support and contradict Hypothesis 3.2. Given the history of the region, this might also reflect how the scale of the demographic shock, together with pre-shock diversity, affects cultural persistence.

For Lithuania and Poland, we also repeated the exercise limiting the sample to a bandwidth of 60 km and 100 km from the East Prussian border, which brings the specifications even closer to the RD design. The results are reported in Table F.20 in the Appendix and are fully in line with the results reported above . Moreover, within these smaller bands, Lithuania and Poland become statistically indistinguishable from each other within East Prussia with respect to conservative voting and, for the bandwidth of 100 km, also with regard to nationalistic voting. This lends more support for Hypothesis 3.2 in the region with more geographic proximity, more initial diversity and a (somewhat) lower scale of the demographic shock.

3.6 Migrating East Prussians

The results for the voting behavior in the FRG after World War II are presented in Table 3.8 for conservative voting and Table 3.9 for nationalist voting. For the sake of brevity, we only present the specifications with relevant control variables. For conservative voting, we find a significant positive effect of the share of East Prussians in the district population on the share of votes the conservative parties receive. The coefficient is statistically significant and highly stable, and this result is robust across a variety of specifications, including IV-estimations. We thus conclude that the presence of the East Prussian expellees in West Germany indeed resulted in more conservative voting, which is consistent with the hypothesized conservative political preferences of the East Prussian population. The effect appears even stronger when we control for the presence of refugees from the other Eastern provinces (compare specifications (1) and (2) in Table 3.8).

The results for nationalistic voting also support our hypotheses. Namely, the share of the East Prussian expellees in the total district population is positively related to the share of votes the nationalistic parties received in 1949. This result is also highly robust, with only one exception: controlling simultaneously for the share of workers as well as shares of agriculture and of trade and financial sector in the economy renders the effect of the East Prussian expellees insignificant (but still positive) in the IV-specification. Given the high correlation between these three controls, however, and the fact that this is the only combination that influences the significance of the result, it is unclear whether it is due to weak robustness or the specification error.

In view of the literature discussed in Section 3.2, the explanation for this positive relation can be twofold. On the one hand, this may be the reaction of the native West German population to the semi-foreigners from the Eastern provinces. This is supported by the prevalence of positive relations to refugees from other provinces (except Berlin and Pomerania, partially statistically significant effects) and is in line with the research on immigration and right-wing voting (see Section 3.2). On the other hand, as the share of nationalistic voting in the 1920s and 1930s predicts the respective patterns after the

war in the West German states, so may the high prewar support of nationalistic parties in East Prussia (see, e.g., Figure G.3 in the Appendix) convert to higher support of nationalistic parties by the East Prussian expellees in 1949.

As a "placebo" test, we also regress the share of East Prussian expellees on the share of votes for nationalistic and conservative parties in 1933, as reported in Table F.23 in the Appendix, and find that the share of East Prussian expellees is negatively related to the conservative vote, although the relationship becomes positive if further controls are added. In contrast, the share of East Prussians is positively related to the nationalistic vote, both with and without controls. Adding the controls takes away about half of the effect. Thus, there must be some unobservable variables affecting both the (long-term) voting patterns and the settlement of the refugees. However, if we use the IV approach (with latitude as an instrument), as in the main regression, the explanatory power of the "placebo" regression drops substantially, and the coefficient on the share of East Prussians approaches zero. This is not the case with the main regression and lends support to our IV approach.

There are several possible objections to the results presented. The first is the issue of self-selection, which may both influence the choice of place of residence and reflect a selection bias in preferences. In these circumstances, however, we consider this bias very small, if any. As discussed in Section 3.3, almost all the surviving population of East Prussia had to leave the area, and so the possibility to self-select into migration did not exist. The expellees could not choose their place of residence strategically either, due to settlement by the military authorities and the subsequent relocation ban, as discussed earlier in Section 3.3 too. Furthermore, our IV approach tackles this problem well.

Another possible – and very valid – objection is that the question of the eastern provinces was not quite settled by the time of the first Bundestag election. It might be the case that the East Prussian expellees voted for the parties who favored the support and integration of the refugees in West Germany and who refused to accept the Oder-Neisse line. The former is barely a concern, as essentially all parties expressed the necessity of proper integration in their programs. Only some very regionally focused parties with a relatively small electoral base, like the Bavarian Party (Bayerische Partei), expressed a demand not to receive the expellees and/or to send them back. Still, even they typically acknowledged that those who had already been settled should have the possibilities to integrate in the social and economic life in their host regions Mintzel 1986. The acceptance of the new borders was not that universal, however. Some of the parties explicitly claimed in their electoral programs that the eastern territories (both, the Soviet occupation zone and the former eastern provinces) should belong to Germany, some did not take any side on this question, while the communist party explicitly stated the Oder-Neisse line should not be revised. We thus used the party programs and/or their closed session resolutions and singled out the parties that were explicitly against the Oder-Neisse line. We then checked the robustness of our result with respect to the conservative vote by splitting the sample into parties rejecting the Oder-Neisse line and those not doing so. This exercise was not possible with respect to the nationalistic parties as there were only two of them, one being negligibly small. We did, however, test whether the presence of the East Prussian expellees is positively related to voting for parties against the Oder-Neisse line independently of their ideology.

Table 3.8: Estimation results for conservative voting in West Germany

Method	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	IV	IV
Dep. variable: Conservative votes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share of EP expellees	1.178***	2.139***	2.804***	2.457***	2.427***	2.417***	2.906***	2.759***	2.702***	3.923***	2.941***
	(0.263)	(0.678)	(0.293)	(0.311)	(0.309)	(0.308)	(0.296)	(0.284)	(0.304)	(0.503)	(0.562)
City	−0.091***	−0.026	0.043**	0.041**	0.035*	0.029	0.030	0.020	0.055*	0.035*	0.021
	(0.024)	(0.028)	(0.019)	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.029)	(0.020)	(0.026)
Conservative votes 1933			0.868***	1.071***	1.087***	1.091***	0.890***	0.834***	0.832***	0.884***	0.888***
			(0.055)	(0.078)	(0.080)	(0.082)	(0.057)	(0.056)	(0.084)	(0.050)	(0.099)
Male			−2.575**	−2.410**	−2.436**	−2.475**	−2.897***	−2.405**	−0.794		
			(1.027)	(1.021)	(1.026)	(1.047)	(1.062)	(1.020)	(1.173)		
Protestants				0.146***		−0.169			0.012		0.039
				(0.051)		(0.350)			(0.056)		(0.069)
Catholics					−0.152***	−0.318					
					(0.050)	(0.342)					
Entrepreneurs											−0.532
											(0.416)
Industry							0.141**		0.152		
							(0.070)		(0.110)		
Public sector								0.632***	0.535**		−0.068
								(0.211)	(0.218)		(0.178)
Wage ratio (EP to host)	−0.637***	−0.642***							−0.430***		−0.519
	(0.104)	(0.094)							(0.116)		(0.093)
Agriculture ratio (EP to host)									−0.001		
									(0.001)		
Entrepreneurship ratio (EP to host)									−0.093***		−0.088**
									(0.035)		(0.037)
Other eastern regions	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
N	194	194	162	162	162	162	162	162	162	163	162
R-squared	0.197	0.323	0.690	0.704	0.705	0.706	0.699	0.706	0.772	0.609	0.727
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. EP = East Prussia

Table 3.9: Estimation results for nationalist voting in West Germany

Method	Robust	Robust	Robust	Robust	Robust	Robust	Robust	IV	IV	IV	IV
Dep. variable:	OLS	OLS	OLS	OLS	OLS	OLS	OLS				
Nationalistic share	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share of EP	2.512***	2.719***	2.675***	2.637***	2.671***	2.696***	2.820***	1.797***	1.502***	1.853***	0.541
expellees	(0.502)	(0.464)	(0.461)	(0.470)	(0.450)	(0.501)	(0.432)	(0.284)	(0.400)	(0.360)	(0.468)
City	0.022**	0.037***	0.047***	0.046***	0.036***	0.008	0.018	0.020*	0.045***	0.027*	0.022*
	(0.010)	(0.011)	(0.013)	(0.013)	(0.011)	(0.013)	(0.014)	(0.012)	(0.013)	(0.015)	(0.013)
Nationalistic		0.078**	0.080**	0.081**	0.083**	0.098***	0.069**	0.146***	0.107***	0.098***	0.147***
share 1933		(0.035)	(0.035)	(0.036)	(0.035)	(0.035)	(0.033)	(0.035)	(0.037)	(0.034)	(0.040)
Male		0.501	0.600	0.031	0.336	−0.120	1.241*				
		(0.538)	(0.555)	(0.463)	(0.487)	(0.464)	(0.726)				
Workers		−0.150***							0.111	−0.097	0.600***
		(0.047)							(0.103)	(0.103)	(0.184)
Entrepreneurs			0.391***				0.048		−1.189***	−0.813**	−0.943***
			(0.117)				(0.326)		(0.364)	(0.384)	(0.354)
Agriculture				0.070***			0.016		0.158**		0.571***
				(0.025)			(0.067)		(0.065)		(0.123)
Industry					−0.095***				−0.107***		
					(0.024)				(0.041)		
Trade & finance						0.178*	0.377***			−0.026	0.911***
						(0.106)	(0.132)			(0.131)	(0.215)
Wage ratio	0.087***						0.018				
(EP to host)	(0.031)						(0.040)				
Agriculture ratio							0.000				
(EP to host)							(0.001)				
Entrepreneurship							−0.055**		−0.073***	−0.074***	−0.086***
ratio (EP to host)							(0.022)		(0.019)	(0.022)	(0.022)
Other eastern regions	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
N	194	162	162	162	162	162	162	162	162	162	162
R-squared	0.730	0.774	0.773	0.767	0.776	0.763	0.806	0.649	0.706	0.683	0.665
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Montiel-Pflueger F								68.213	40.903	50.501	32.549

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. EP = East Prussia. Column (7) does not show the following (statistically insignificant) variables: Industry ratio (EP to host), Workers ratio (EP to host)

We present the test results in Table F.21 in the Appendix. Indeed, voting for parties against the Oder-Neisse line is positively related to the share of the East Prussian expellees in the total population, even after controlling for other expellee groups and socioeconomic characteristics of the districts. This result holds both for conservative parties and in general for all parties that expressed a position against the new borders. At the same time, the effect of the East Prussian expellees on voting for conservative parties who did not declare a position with regard to the Oder-Neisse line is insignificant. This might seem to overturn our finding of support for conservative parties in favor of revision of the new borders. However, a closer look at the composition of parties rejecting the Oder-Neisse line shows that the positive results are completely driven by the German Party (Deutsche Partei), which was both conservative, nationalistic and against the new borders. The effect of the East Prussian presence on the group of parties against the new borders does not withstand the exclusion of the German Party: the coefficient becomes insignificant when controlling for the other expellees and even changes the sign when not controlling for the other expellees (see specifications (7) and (8) in Table F.21). At the same time, voting for the German party is consistently positively related to the share of East Prussians, both when taking the full sample and when restricting it only to the regions where the party was actually present: Bremen, Hamburg, Schleswig-Holstein and Lower Saxony, which are also the regions with high shares of East Prussians (see Table F.22 in the Appendix).

Another argument against voting for border revision only is that there were several parties that expressed their position against the Oder-Neisse line but were not conservative or nationalistic, with the liberal FDP (Free Democratic Party/Freie Demokratische Partei) being most explicit among all parties in its demand to return the eastern provinces to Germany while supporting the Eastern expellees in their host regions (FDP 1948; FDP 1949). At the same time, as noted above, the communist party (KPD) explicitly stated that the revision of the borders would endanger peace and should not be attempted (Parteivorstand der KPD 1949). Had the East Prussian expellees voted primarily in the hope of restoring the old borders, we would see a positive relationship to voting for FDP and a negative relationship to voting for KPD. This is, however, not the case. In fact, the share of East Prussians is negatively related to the FDP share, as could be expected given the hypothesized East Prussian conservative preferences (see Table F.22 in the Appendix). This result is robust across several specifications. At the same time, there is no consistent relationship between the share of East Prussians and voting for KPD: the negative effect vanishes once the shares of other expellees are controlled for.

We, therefore, conclude that the positive relationship between conservative voting and the share of East Prussians in the district population is not completely driven by the parties' position with regard to the new German borders. Neither is nationalistic voting likely to be driven only by the dissent of the native population, as the German Party, for which there is the strongest positive relation with the share of East Prussians, demanded proper integration and support for the refugees (Mommsen 1960). As such, the German Party had a strong standing among the expellees, including the East Prussians. It expressed rather conservative and nation-centered views, demanded good living conditions for the refugees, declared Germany to be both to the west and to the east of the Oder-Neisse line, and above all, drew a lot of members from the associations of the Prussian-conservative DKP/DRP and provided "political shelter" to the expellees, who were not yet allowed to

form their own parties (Mommsen 1960; Schmollinger 1986).

The robustness check with the spatial autoregression largely confirms our findings. The Moran's I-test indicates both for the conservative and nationalist shares that there is spatial dependence.[5] The overview of the SAR results can be found in Table F.24 in the Appendix. For the conservative share of votes, the results are robust to individual and any combination of the spatial lags of the dependent and independent variable and of the error term. The direct effect of the share of East Prussian expellees is comparable to that in the OLS estimations, while the total effect is mostly even larger. For the share of votes for nationalistic parties, the results are robust to spatial lags of the share of East Prussians and of the error term, but not to the lag of the dependent variable. Inclusion of the latter inflates the standard errors, especially of the indirect effect, and thus no conclusive evidence is obtained. At the same time, this result depends on the estimation method: using robust maximum likelihood instead of the generalized spatial two-stage least squares (GS2SLS) generates more moderate standard errors. The direct effect of East Prussian expellees is also close to that in the OLS estimations throughout all specification. Still, the outlier result of the spatial lag of the dependent variable calls for some caution when interpreting the results for nationalistic voting.

3.7 Conclusion

While most economic research on cultural persistence investigates the long-term impacts of different cultural or institutional environments within one modern state or across otherwise similar states, our focus is on the opposite case. In this chapter, we investigate whether a relatively homogenous region can persist as such after being exposed to different political settings. While not questioning that the new regions diverge, we ask whether this divergence is slowed down by common history.

We find that legacies of former East Prussia only partially persist in the region in the case of both political and economic outcomes. This persistence greatly depends on how much of the original East Prussian population was left in the area and who filled the demographic vacuum created by World War II and the partitioning of the province. We find most evidence for persistence in Lithuania, where the departure of the German population was much more gradual, stretching over 25 years. There is little evidence for economic persistence in Poland or Russia. In fact, in Poland, also in the case of political preferences, our hypothesis of persistence is rejected. Moreover, we find that the flight and expulsion of the East Prussian population in the aftermath of the Second World War changed the political outcomes in the first parliamentary election in the West German regions where the expellees were settled. This finding implicitly supports the idea of intergenerational transmission as the main persistence channel, but also advocates caution in the interpretation of persistence in Eastern Europe. With the massive population movements in the mid-twentieth century, it might be tempting to

[5] The Moran test for the conservative share returns $\chi^2 = 19.14$ (the independent variables were the share of East Prussians, city, conservative share in 1920 and wage in East Prussia relative to host region). The Moran test for the nationalistic share returns $\chi^2 = 8.83$ (with the share of East Prussians, city, nationalistic share in 1920 and wage in East Prussia relative to host region).

describe culture as persistent, where it is actually more likely to be determined by a selection bias of migration.

Appendix

E Data description

Table E.1: Data profile and sources

Variable	Unit	Period	Data source	Notes
Political data in former East Prussia				
Turnout	%	2011 (Poland, Russia), 2012 (Lithuania)	The Central Electoral Commission of the Republic of Lithuania; Polish National Electoral Commission; Central Election Commission of the Russian Federation	Share of valid ballot papers in total number of voters
Votes for conservative parties (Lithuanian parliamentary election)	%	2012	The Central Electoral Commission of the Republic of Lithuania	Share of people voting for: Order and Justice; Lithuanian Poles' Electoral Action; Lithuanian Peasant and Greens Union; Christian Party; 'Young Lithuania'
Votes for conservative parties (Polish parliamentary election)	%	2011	National Electoral Commission	Share of people voting for: Prawo i Sprawiedliwość (Law and Justice); Prawica (Right Wing of the Republic)
Votes for conservative parties (Russian parliamentary election)	%	2011	Central Election Commission of the Russian Federation	Share of people voting for: United Russia
Votes for liberal-conservative parties (Lithuanian parliamentary election)	%	2012	The Central Electoral Commission of the Republic of Lithuania	Share of people voting for: Order and Justice; Lithuanian Poles' Electoral Action; Lithuanian Peasant and Greens Union; Christian Party; 'Young Lithuania'; Homeland Union - Lithuanian Christian Democrats; Liberals Movement of the Republic of Lithuania; Liberal and Centre Union; National Association 'For Lithuania in Lithuania'

Table E.1 – continued from previous page

Variable	Unit	Period	Data source	Notes
Votes for liberal-conservative parties (Polish parliamentary election)	%	2011	National Electoral Commission	Share of people voting for: Prawo i Sprawiedliwość (Law and Justice); Prawica (Right Wing of the Republic); Polska Jest Najważniejsza (Poland Comes First); Platforma Obywatelska RP (Civic Platform); Nowa Prawica - Janusza Korwin-Mikke (Congress of the New Right)
Votes for liberal-conservative parties (Russian parliamentary election)	%	2011	Central Election Commission of the Russian Federation	Share of people voting for: United Russia; Civic Platform
Votes for nationalistic parties (Lithuanian parliamentary election)	%	2012	The Central Electoral Commission of the Republic of Lithuania	Share of people voting for: Homeland Union - Lithuanian Christian Democrats; Order and Justice; National Association 'For Lithuania in Lithuania'; 'Young Lithuania'
Votes for nationalistic parties (Polish parliamentary election)	%	2011	National Electoral Commission	Share of people voting for: Prawo i Sprawiedliwość (Law and Justice)
Votes for nationalistic parties (Russian parliamentary election)	%	2011	Central Election Commission of the Russian Federation	Share of people voting for: United Russia, LDPR
Economic data in former East Prussia				
Overall	Number of enterprises	2015-2017 (Poland), 2015-2018 (Lithuania), 2017-2018 (Russia)	Geospatial data on economic entities in operation by grid, Statistics Lithuania; Local Data Bank, Statistics Poland; Database of Indicators of Municipalities, Federal State Statistic Service (Russia)	Number of economic entities in operation (Lithuania, Poland) Number of reporting enterprises (Russia)
Agriculture				Number of economic entities by economic activity: agriculture, forestry and fishing
Manufacturing				Number of economic entities by economic activity: manufacturing, mining and quarrying, and other industry
Construction				Number of economic entities by economic activity: construction
Wholesale & Retail Trade				Number of economic entities by economic activity: wholesale and retail trade, transportation and storage, accommodation and food service activities
Information & Communication				Number of economic entities by economic activity: information and communication
Financial & Insurance Activities				Number of economic entities by economic activity: financial and insurance activities

Table E.1 – continued from previous page

Variable	Unit	Period	Data source	Notes
Real Estate Activities				Number of economic entities by economic activity: real estate activities
Professional & Scientific Activities				Number of economic entities by economic activity: professional, scientific and technical activities, administrative and support service activities
Public Administration				Number of economic entities by economic activity: public administration and defense, education, human health and social work
Other Services				Number of economic entities by economic activity: other services
	Political and socioeconomic data in West Germany			
Conservative votes in FRG	%	1949	Der Bundeswahlleiter 2016	Share of people voting for: Christian Democratic Union (CDU); Christian Social Union (CSU); German Party (DP); Centre Party (Zentrum); German Conservative Party / German Right Party (DKP/DRP)
Nationalistic votes in FRG	%	1949	Der Bundeswahlleiter 2016	Share of people voting for: German Party (DP); European People's Movement in Germany (EVD)
Share of expellees in FRG districts	%	1950	Braun and Dwenger 2018	Share of expellees from East Prussia, Berlin, East Brandenburg, Silesia, or Pomerania in the total district population
German elections 1920-1933	%	1920-1933	Falter and Hänisch 1990	*Conservative parties*: share of people voting for Centre / Bavarian People's Party (separate parties until 1928); German National People's Party; Christian Social People's Service (from 1930) *Nationalistic parties*: share of people voting for NSDAP (from 1928); National People's Party; German People's Party
Share of industry, agriculture, or entrepreneurship in East Prussia vs. FRG districts	%	1939	Falter and Hänisch 1990	Share of employees in industry, agriculture, or entrepreneurship in total employment – ratio of the share in East Prussia to that in the FRG district
Demographics of FRG districts	%	1950	Schmitt et al. 1994	Share of males; young people (aged 15-20); old people (aged over 65); Protestants; or Catholics in total population
Employment structure in FRG districts	%	1950	Schmitt et al. 1994	Share of workers; entrepreneurs; employees in agriculture, industry, trade & finance, or public sector in total employment

Table E.2: Descriptive statistics: Main variables

Variable	Lithuania Obs	Mean	Std. Dev.	Min	Max	Poland Obs	Mean	Std. Dev.	Min	Max	Russia Obs	Mean	Std. Dev.	Min	Max
Distance to East Prussian border	29,963	96.82	60.96	5.043	252.2	7,434	223.5	120.1	343	466.2	274	602	288	13	1060
East Prussia	29,963	0.066	0.248	0	1	7,434	0.038	0.192	0	1	274	0.161	0.368	0	1
City	29,963	0.158	0.364	0	1	7,434	0.122	0.328	0	1	274	0.131	0.338	0	1
Agriculture	29,963	0.311	0.862	0	24	7,434	30.2	41.20	0	1301	274	0.653	2.006	0	12
Manufacturing	29,963	1.240	4.218	0	87	7,434	164	745.24	2	28510	274	5.047	11.22	0	63
Construction	29,963	1.004	4.461	0	114	7,434	201.15	814.19	4	31612	274	0.723	2.968	0	23
Wholesale & Retail Trade	29,963	4.834	20.57	0	460	7,434	579.6	3071.6	13	125483	274	2.5	7.555	0	46
Information & Communication	29,963	0.443	3.756	0	130	7,434	57.83	761.87	0	37717	274	0.135	0.766	0	6
Financial & Insurance Activities	29,963	0.108	1.174	0	61	7,434	50.51	378.52	0	16171	274	0.015	0.242	0	4
Real Estate Activities	29,963	0.618	4.601	0	225	7,434	97.17	651.94	0	25645	274	0.595	2.339	0	20
Professional & Scientific Activities	29,963	1.812	13.36	0	462	7,434	221.46	2140.4	1	100236	274	0.314	1.759	0	16
Public Administration	29,963	1.036	4.995	0	170	7,434	165.09	879.11	6	33669	274	0.631	2.793	0	17
Other Services	29,963	1.963	9.091	0	243	7,434	143.27	803.54	4	34250	274	0.42	2.023	0	14
Number of employees	11,569	410.1	1451	0	47307	na	na	na	na	na	na	na	na	na	na
Number of economic entities	29,963	13.37	62.40	1	1796	7,434	1712.4	10186	68	434676	274	17.25	31.71	0	196
Turnout	2,000	0.518	0.075	0.299	0.962	2,479	0.425	0.069	0.231	0.726	158	0.565	0.086	0.401	0.865
Conservative Share	2,000	0.220	0.149	0.025	0.950	2,479	0.335	0.126	0.040	0.831	158	0.475	0.116	0.257	0.936
Liberal-Conservative Share	2,000	0.441	0.134	0.077	0.957	2,479	0.653	0.111	0.184	0.924	158	0.479	0.115	0.266	0.936
Nationalist Share	2,000	0.234	0.093	0.007	0.704	2,479	0.332	0.124	0.040	0.829	158	0.606	0.093	0.40	0.952

Table E.3: Descriptive statistics: Individual parties

	Obs	Mean	Std. Dev.	Min	Max
Lithuania					
Labor Party	2,000	0.248	0.110	0.029	0.760
Social Democratic Party	2,000	0.201	0.101	0.000	0.779
Homeland Union	2,000	0.131	0.069	0.000	0.487
Liberals Movement	2,000	0.066	0.051	0.000	0.343
The Way of Courage	2,000	0.067	0.046	0.000	0.294
Order & Justice	2,000	0.089	0.075	0.000	0.635
Poles' Electoral Action	2,000	0.050	0.146	0.000	0.950
Peasant & Greens Union	2,000	0.062	0.069	0.000	0.811
Liberal & Centre Union	2,000	0.023	0.031	0.000	0.383
Union YES	2,000	0.012	0.015	0.000	0.157
Socialist People's Front	2,000	0.011	0.013	0.000	0.153
Christian Party	2,000	0.014	0.014	0.000	0.171
National Association	2,000	0.008	0.009	0.000	0.150
Young Lithuania	2,000	0.005	0.006	0.000	0.053
Democratic Labor & Unity Party	2,000	0.004	0.005	0.000	0.063
Emigrants' Party	2,000	0.002	0.003	0.000	0.030
Republican Party	2,000	0.003	0.007	0.000	0.156
People's Party	2,000	0.003	0.004	0.000	0.064
Poland					
Prawo i Sprawiedliwość	2,479	0.332	0.124	0.040	0.829
Polska Jest Najważniejsza	2,479	0.020	0.012	0.003	0.199
Sojusz Lewicy Demokratycznej	2,479	0.082	0.049	0.003	0.529
Ruch Palikota	2,479	0.089	0.031	0.014	0.226
Polskie Stronnictwo Ludowe	2,479	0.163	0.109	0.011	0.766
Polska Partia Pracy - Sierpien 80	2,479	0.006	0.005	0.000	0.077
Platforma Obywatelska RP	2,479	0.290	0.135	0.000	0.703
Nasz Dom Polska	2,479	0.001	0.003	0.000	0.042
Nowa Prawica	2,479	0.008	0.009	0.000	0.086
Prawica	2,479	0.003	0.004	0.000	0.048
Russia					
United Russia	158	0.475	0.116	0.257	0.936
LDPR	158	0.131	0.033	0.015	0.207
CPRF	158	0.204	0.048	0.035	0.318
Just Rusia	158	0.150	0.056	0.011	0.315
Yabloko	158	0.025	0.019	0.001	0.118
Patriots of Russia	158	0.011	0.004	0.000	0.032
Right Cause	158	0.004	0.002	0.000	0.010

F Estimation results

F.1 Persistence in East Prussia

Table F.1: RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia (all parties)

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Prawo i Sprawiedliwośc	Conventional	−0.061***	0.018	−3.42	−0.096	−0.026	2479
	Bias-Corrected	0.046**	0.018	−2.59	−0.081	−0.011	2479
	Robust	−0.046*	0.028	−1.68	−0.100	0.008	2479
Polska Jest Najważniejsza	Conventional	−0.009*	0.005	−1.71	−0.018	0.001	2479
	Bias-Corrected	−0.014***	0.005	−2.72	−0.023	−0.004	2479
	Robust	−0.014	0.009	−1.46	−0.032	0.005	2479
Sojusz Lewicy Demokratycznej	Conventional	0.021***	0.007	3.06	0.008	0.035	2479
	Bias-Corrected	0.026***	0.007	3.82	0.013	0.040	2479
	Robust	0.026***	0.009	3.04	0.009	0.044	2479
Ruch Palikota	Conventional	0.029***	0.007	4.23	0.016	0.043	2479
	Bias-Corrected	0.019***	0.007	2.66	0.005	0.032	2479
	Robust	0.019*	0.010	1.92	0.000	0.037	2479
Polskie Stronnictwo Ludowe	Conventional	−0.039	0.025	−1.59	−0.088	0.009	2479
	Bias-Corrected	−0.032	0.025	−1.31	−0.081	0.016	2479
	Robust	−0.032	0.036	−0.91	−0.103	0.038	2479
Polska Partia Pracy – Sierpień 80	Conventional	−0.004*	0.002	−1.90	−0.007	0.000	2479
	Bias-Corrected	−0.004**	0.002	−2.06	−0.008	0.000	2479
	Robust	−0.004	0.003	−1.55	−0.009	0.001	2479
Platforma Obywatelska RP	Conventional	0.068***	0.026	2.60	0.017	0.119	2479
	Bias-Corrected	0.058**	0.026	2.22	0.007	0.109	2479
	Robust	0.058	0.040	1.47	−0.019	0.136	2479
Nasz Dom Polska	Conventional	−0.003***	0.001	−2.79	−0.005	−0.001	2479
	Bias-Corrected	−0.002**	0.001	−2.03	−0.004	0.000	2479
	Robust	−0.002	0.001	−1.39	−0.005	0.001	2479
Prawica	Conventional	0.001	0.001	1.38	0.000	0.002	2479
	Bias-Corrected	0.001	0.001	1.19	−0.001	0.002	2479
	Robust	0.001	0.001	0.72	−0.002	0.003	2479
Nowa Prawica	Conventional	−0.004***	0.001	−3.43	−0.007	−0.002	2479
	Bias-Corrected	−0.006***	0.001	−4.66	−0.008	−0.003	2479
	Robust	−0.006***	0.002	−3.04	−0.010	−0.002	2479
Turnout	Conventional	−0.038**	0.016	−2.37	−0.069	−0.006	2479
	Bias-Corrected	−0.035**	0.016	−2.22	−0.066	−0.004	2479
	Robust	−0.035	0.022	−1.62	−0.078	0.007	2479
Conservative Share	Conventional	−0.060***	0.018	−3.32	−0.096	−0.025	2479
	Bias-Corrected	−0.045**	0.018	−2.51	−0.081	−0.010	2479
	Robust	−0.045	0.028	−1.63	−0.100	−0.009	2479
Liberal-Conservative Share	Conventional	−0.005	0.020	−0.25	−0.043	0.033	2479
	Bias-Corrected	−0.007	0.020	−0.34	−0.045	0.032	2479
	Robust	−0.007	0.029	−0.23	−0.063	0.050	2479

Table F.1 – continued

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Nationalist Share	Conventional	-0.061^{***}	0.018	-3.42	-0.096	-0.026	2479
	Bias-Corrected	-0.046^{**}	0.018	-2.59	-0.081	-0.011	2479
	Robust	-0.046^{*}	0.028	-1.68	-0.100	0.008	2479

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km.

Table F.2: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia (all parties)

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Labor Party	Conventional	0.059**	0.026	2.27	0.008	0.110	2000
	Bias-Corrected	0.061**	0.026	2.36	0.010	0.112	2000
	Robust	0.061*	0.033	1.84	−0.004	0.127	2000
Social Democratic Party	Conventional	−0.105***	0.024	−4.29	−0.153	−0.057	2000
	Bias-Corrected	−0.062**	0.024	−2.55	−0.110	−0.014	2000
	Robust	−0.062**	0.032	−1.97	−0.125	0.000	2000
Homeland Union	Conventional	−0.028***	0.011	−2.65	−0.049	−0.007	2000
	Bias-Corrected	−0.039***	0.011	−3.68	−0.060	−0.018	2000
	Robust	−0.039***	0.013	−2.91	−0.067	−0.013	2000
Liberals Movement	Conventional	−0.0004	0.013	−0.04	−0.027	0.026	2000
	Bias-Corrected	−0.034**	0.013	−2.57	−0.060	−0.008	2000
	Robust	−0.034**	0.017	−2.03	−0.067	−0.001	2000
The Way of Courage	Conventional	−0.010	0.007	−1.30	−0.024	0.005	2000
	Bias-Corrected	−0.023***	0.007	−3.13	−0.038	−0.009	2000
	Robust	−0.023**	0.010	−2.39	−0.042	−0.004	2000
Order & Justice	Conventional	−0.104***	0.037	2.79	0.031	0.177	2000
	Bias-Corrected	−0.128***	0.037	3.42	0.055	0.201	2000
	Robust	−0.128***	0.044	2.87	0.040	0.214	2000
Poles' Electoral Action	Conventional	−0.004***	0.002	−2.54	−0.007	−0.001	2000
	Bias-Corrected	−0.027***	0.002	−17.45	−0.030	−0.024	2000
	Robust	−0.027***	0.002	−11.70	−0.032	−0.023	2000
Peasant & Greens Union	Conventional	−0.001	0.012	−0.10	−0.025	0.023	2000
	Bias-Corrected	0.011	0.012	0.91	−0.013	0.035	2000
	Robust	0.011	0.016	0.70	−0.020	0.042	2000
Liberal & Centre Union	Conventional	−0.008	0.005	−1.59	−0.017	0.002	2000
	Bias-Corrected	−0.007	0.005	−1.37	−0.016	0.003	2000
	Robust	−0.007	0.006	−1.15	−0.018	0.005	2000
Union YES	Conventional	0.007	0.005	1.45	−0.002	0.016	2000
	Bias-Corrected	0.010**	0.005	2.02	0.000	0.019	2000
	Robust	0.010	0.006	1.49	−0.003	0.022	2000
Socialist People's Front	Conventional	−0.004***	0.001	−3.57	−0.007	−0.002	2000
	Bias-Corrected	−0.007***	0.001	−5.63	−0.009	−0.004	2000
	Robust	−0.007***	0.002	−4.25	−0.010	−0.004	2000
Christian Party	Conventional	−0.006**	0.003	−2.06	−0.011	0.000	2000
	Bias-Corrected	−0.005*	0.003	−1.86	−0.010	0.000	2000
	Robust	−0.005	0.003	−1.47	−0.012	0.002	2000
National Association	Conventional	−0.003	0.002	−1.52	−0.006	0.001	2000
	Bias-Corrected	−0.004**	0.002	−2.35	−0.008	−0.001	2000
	Robust	−0.004*	0.002	−1.78	−0.009	0.000	2000
Young Lithuania	Conventional	-7.0×10^{-6}	0.001	−0.01	−0.002	0.002	2000
	Bias-Corrected	4.3×10^{-5}	0.001	0.04	−0.002	0.002	2000
	Robust	4.3×10^{-5}	0.001	0.03	−0.002	0.003	2000
Democratic Labor & Unity Party	Conventional	−0.001	0.001	−0.73	−0.003	0.001	2000
	Bias-Corrected	0.0003	0.001	0.32	−0.002	0.002	2000
	Robust	0.0003	0.001	0.28	−0.002	0.003	2000

Table F.2 – continued

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Emigrants' Party	Conventional	−0.0003	0.001	−0.53	−0.001	0.001	2000
	Bias-Corrected	−0.001*	0.001	−1.74	−0.002	0.000	2000
	Robust	−0.001	0.001	−1.49	−0.002	0.000	2000
Republican Party	Conventional	−0.0002	0.001	−0.39	−0.001	0.001	2000
	Bias-Corrected	−0.0002	0.001	−0.37	−0.001	0.001	2000
	Robust	−0.0002	0.001	−0.29	−0.002	0.001	2000
People's Party	Conventional	0.0003	0.001	0.39	−0.001	0.002	2000
	Bias-Corrected	0.0005	0.001	0.68	−0.001	0.002	2000
	Robust	0.0005	0.001	0.56	−0.001	0.002	2000
Turnout	Conventional	−0.011	0.019	−0.59	−0.049	0.026	2000
	Bias-Corrected	−0.013	0.019	−0.67	−0.051	0.025	2000
	Robust	−0.013	0.024	−0.54	−0.060	0.034	2000
Conservative Share	Conventional	0.093**	0.036	2.58	0.023	0.164	2000
	Bias-Corrected	0.106***	0.036	2.94	0.036	0.177	2000
	Robust	0.106**	0.042	2.53	0.024	0.189	2000
Liberal-Conservative Share	Conventional	0.057*	0.030	1.88	−0.002	−0.660	2000
	Bias-Corrected	0.026	0.030	0.87	−0.033	0.085	2000
	Robust	0.026	0.037	0.72	−0.046	0.098	2000
Nationalist Share	Conventional	0.073**	0.034	2.13	0.006	0.140	2000
	Bias-Corrected	0.084***	0.034	2.45	0.017	0.151	2000
	Robust	0.084**	0.042	2.01	0.002	0.166	2000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km.

Table F.3: RD results with robust bias-corrected CIs: Socio-economic outcomes in Poland vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	4.616**	2.198	2.10	0.307	8.924	7434
	Bias-Corrected	−0.406	2.198	−0.18	−4.714	3.902	7434
	Robust	−0.406	2.857	−0.18	−6.006	5.195	7434
Manufacturing	Conventional	−11.542	12.442	−0.93	−35.927	12.843	7434
	Bias-Corrected	4.904	12.442	0.39	−19.481	28.289	7434
	Robust	4.904	13.629	0.36	−21.807	31.616	7434
Construction	Conventional	−0.172	15.274	−0.01	−30.108	29.764	7434
	Bias-Corrected	13.461	15.274	0.88	−16.475	43.397	7434
	Robust	13.461	16.94	0.79	−19.740	46.663	7434
Wholesale & Retail Trade	Conventional	−63.149	47.570	−1.33	−156.384	30.087	7434
	Bias-Corrected	8.756	47.570	0.18	−84.480	101.991	7434
	Robust	8.756	50.425	0.17	−90.075	107.586	7434
Information & Communication	Conventional	−5.253	3.666	−1.43	−12.439	1.933	7434
	Bias-Corrected	2.683	3.666	0.73	−4.503	9.869	7434
	Robust	2.683	3.717	0.73	−4.602	9.969	7434
Financial & Insurance Activities	Conventional	−7.251	5.694	−1.27	−18.410	3.908	7434
	Bias-Corrected	4.166	5.694	0.73	−6.994	15.325	7434
	Robust	4.166	5.706	0.73	−7.019	15.350	7434
Real Estate Activities	Conventional	−8.160	19.374	−0.42	−46.132	29.812	7434
	Bias-Corrected	8.298	19.374	0.43	−29.674	46.270	7434
	Robust	8.298	19.954	0.42	−30.812	47.408	7434
Professional & Scientific Activities	Conventional	−32.301	19.339	−1.67	−70.205	5.602	7434
	Bias-Corrected	15.761	19.339	0.82	−22.143	53.665	7434
	Robust	15.761	19.274	0.82	−22.015	53.537	7434
Public Administration	Conventional	−16.464	20.089	−0.82	−55.839	22.910	7434
	Bias-Corrected	20.355	20.089	1.01	−19.019	59.729	7434
	Robust	20.355	21.307	0.96	−21.405	62.115	7434
Other Services	Conventional	−13.358	15.953	−0.84	−44.627	17.910	7434
	Bias-Corrected	16.439	15.953	1.03	−14.830	47.707	7434
	Robust	16.439	16.385	1.00	−15.675	48.552	7434

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. Bandwidth is 60 km.

Table F.4: RD results with robust bias-corrected CIs: Socio-economic outcomes in Lithuania vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	-0.154^{***}	0.014	−10.90	−0.182	−0.127	29963
	Bias-Corrected	-0.161^{***}	0.014	−11.37	−0.189	−0.133	29963
	Robust	-0.161^{***}	0.020	−8.14	−0.200	−0.122	29963
Manufacturing	Conventional	-0.326^{***}	0.035	−9.42	−0.393	−0.258	29963
	Bias-Corrected	-1.272^{***}	0.035	−36.79	−1.339	−1.204	29963
	Robust	-1.272^{***}	0.053	−23.96	−1.376	−1.168	29963
Construction	Conventional	-0.429^{***}	0.031	−13.74	−0.490	−0.368	29963
	Bias-Corrected	-1.406^{***}	0.031	−45.02	−1.467	−1.344	29963
	Robust	-1.406^{***}	0.052	−27.06	−1.507	−1.304	29963
Wholesale & Retail Trade	Conventional	-3.197^{***}	0.088	−36.20	−3.370	−3.024	29963
	Bias-Corrected	-6.974^{***}	0.088	−78.95	−7.147	−6.800	29963
	Robust	-6.974^{***}	0.151	−46.24	−7.269	−6.678	29963
Information & Communication	Conventional	-0.118^{***}	0.008	−14.11	−0.134	−0.101	29963
	Bias-Corrected	-0.297^{***}	0.008	−35.73	−0.314	−0.281	29963
	Robust	-0.297^{***}	0.014	−21.39	−0.325	−0.270	29963
Financial & Insurance Activities	Conventional	−0.047	0.004	−11.88	−0.054	−0.039	29963
	Bias-Corrected	−0.137	0.004	−34.84	−0.145	−0.129	29963
	Robust	−0.137	0.006	−22.48	−0.149	−0.125	29963
Real Estate Activities	Conventional	-0.539^{***}	0.020	−27.62	−0.577	−0.501	29963
	Bias-Corrected	-1.226^{***}	0.020	−62.83	−1.264	−1.187	29963
	Robust	-1.226^{***}	0.031	−39.27	−1.287	−1.164	29963
Professional & Scientific Activities	Conventional	-0.893^{***}	0.027	−32.69	−0.947	−0.840	29963
	Bias-Corrected	2.289^{***}	0.027	−83.75	−2.343	−2.235	29963
	Robust	2.289^{***}	0.046	−49.57	−2.38	−2.199	29963
Public Administration	Conventional	-0.770^{***}	0.024	−31.73	−0.818	−0.723	29963
	Bias-Corrected	-1.475^{***}	0.024	−60.76	−1.522	−1.428	29963
	Robust	-1.475^{***}	0.038	−38.67	−1.550	−1.400	29963
Other Services	Conventional	-1.067^{***}	0.044	−24.32	−1.153	−0.981	29963
	Bias-Corrected	-2.472^{***}	0.044	−56.35	−2.558	−2.386	29963
	Robust	-2.472^{***}	0.071	−34.75	−2.611	−2.332	29963
Number of Employees	Conventional	-103.13^{***}	9.426	−10.94	−121.59	−84.651	11569
	Bias-Corrected	-289.67^{***}	9.426	−30.73	−308.14	−271.19	11569
	Robust	-289.67^{***}	13.433	−21.56	−315.99	−263.34	11569
Income of Economic Entities	Conventional	0.074	0.202	0.36	−0.322	0.469	5457
	Bias-Corrected	-0.397^{**}	0.202	−1.97	−0.792	−0.001	5457
	Robust	−0.397	0.269	−1.48	−0.923	0.130	5457

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km.

Table F.5: Economic matching results for Russian East Prussia (Kaliningrad) coarsening, Match I

			Treated				Control
Number of strata:	17						
Number of matched strata:	5						
All			44				230
Matched			38				64
Unmatched			6				166
Multivariate imbalance measure:	L1 = 0.232						
Univariate imbalance measures:							
Variable	L1	Mean	Min	25%	50%	75%	Max
City	1.2×10^{-16}	1.4×10^{-16}	0	0	0	0	0
Altitude	0.147	-3.305	16	-6	-8	1	11

Table F.6: Economic matching results for Russian East Prussia (Kaliningrad) coarsening, Match II

			Treated				Control
Number of strata:	52						
Number of matched strata:	6						
All			44				230
Matched			28				20
Unmatched			16				210
Multivariate imbalance measure:	L1 = 0.679						
Univariate Imbalance Measures:							
Variable	L1	Mean	Min	25%	50%	57%	Max
City	1.9×10^{-16}	-5.6×10^{-17}	0	0	0	0	0
Altitude	0.071	-3.196	-7	0	-2	1	11
Distance to the Russian border (rescaled)	0.518	-0.012	0.002	-0.017	-0.015	0	0.002

Table F.7: Economic estimates of East Prussian impact in Russia (Kaliningrad)

Outcome	Matching Model	Coefficient	Std. Err.	t	N	R-squared
Agriculture	Match I	-2.189^{***}	0.568	−3.86	102	0.130
	Match II	-1.938^{**}	0.861	−2.25	48	0.099
Manufacturing	Match I	-10.779^{***}	3.032	−3.56	102	0.112
	Match II	-9.071^{**}	4.289	−2.11	48	0.087
Construction	Match I	−1.095	0.838	−1.31	102	0.017
	Match II	0.071	0.944	0.08	48	0.000
Wholesale & Retail Trade	Match I	-7.211^{***}	2.304	−3.13	102	0.089
	Match II	-7.429^{**}	3.467	−2.14	48	0.091
Information & Communication	Match I	−0.074	0.239	−0.31	102	0.001
	Match II	0.214	0.383	0.56	48	0.007
Financial & Insurance Activities	Match I	0.105	0.081	1.30	102	0.017
	Match II	0.143	0.170	0.84	48	0.015
Real Estate Activities	Match I	-1.300^{**}	0.648	−2.01	102	0.039
	Match II	-1.393^{*}	0.736	−1.89	48	0.072
Professional & Scientific Activities	Match I	0.100	0.517	0.19	102	0.000
	Match II	0.571	0.979	0.58	48	0.007
Public Administration	Match I	−0.321	0.627	−0.51	102	0.003
	Match II	0.071	0.871	−0.33	48	0.002
Other Services	Match I	−0.374	0.520	−0.72	102	0.269
	Match II	0.071	0.871	0.08	48	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively.

Table F.8: RD results with robust bias-corrected CIs: Socio-economic outcomes in Poland vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	8.946***	1.919	4.66	5.186	12.706	7434
	Bias-Corrected	1.549	1.919	0.81	−2.211	5.309	7434
	Robust	1.549	2.422	0.64	−3.197	6.295	7434
Manufacturing	Conventional	26.726**	10.813	2.47	5.533	47.920	7434
	Bias-Corrected	−9.528	10.813	−0.88	−30.722	11.665	7434
	Robust	−9.528	16.513	−0.58	−41.893	22.837	7434
Construction	Conventional	42.131***	12.816	3.29	17.012	67.251	7434
	Bias-Corrected	−4.393	12.816	−0.34	−29.513	20.726	7434
	Robust	−4.393	19.305	−0.23	−42.230	33.443	7434
Wholesale & Retail Trade	Conventional	64.921	39.925	1.63	−13.331	143.172	7434
	Bias-Corrected	−70.617	39.925	−1.77	−148.869	7.634	7434
	Robust	−70.617	59.724	−1.18	−187.673	46.439	7434
Information & Communication	Conventional	11.305***	3.284	3.44	4.869	17.740	7434
	Bias-Corrected	−1.949	3.284	−0.59	−8.384	4.487	7434
	Robust	−1.949	6.166	−0.32	−14.034	10.137	7434
Financial & Insurance Activities	Conventional	6.424	4.645	1.38	−2.680	15.289	7434
	Bias-Corrected	−6.135	4.645	−1.32	−15.240	2.969	7434
	Robust	−6.135	7.308	−0.84	−20.458	8.188	7434
Real Estate Activities	Conventional	23.947	15.680	1.53	−6.784	54.679	7434
	Bias-Corrected	−2.229	15.680	−0.14	−32.961	28.502	7434
	Robust	−2.229	23.128	−0.10	−47.559	43.101	7434
Professional & Scientific Activities	Conventional	37.988**	15.252	2.49	8.095	67.880	7434
	Bias-Corrected	−30.859**	15.252	−2.02	−60.752	−0.967	7434
	Robust	−30.859	27.440	−1.12	−84.640	22.921	7434
Public Administration	Conventional	25.106	15.578	1.61	−5.427	55.639	7434
	Bias-Corrected	−16.484	15.578	−1.06	−47.017	14.049	7434
	Robust	−16.484	25.195	−0.65	−65.866	32.898	7434
Other Services	Conventional	24.447**	12.355	1.98	0.232	48.662	7434
	Bias-Corrected	−19.572	12.355	−1.58	−14.830	4.643	7434
	Robust	−19.572	19.824	−0.99	−15.675	19.281	7434

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km.

Table F.9: Covariate-adjusted RD results with robust bias-corrected CIs: Socio-economic outcomes in Poland vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	4.220*	2.228	1.89	−0.148	8.588	7434
	Bias-Corrected	−1.646	2.228	−0.74	−6.014	2.721	7434
	Robust	−1.646	2.898	−0.57	−7.326	4.034	7434
Manufacturing	Conventional	−7.503	11.291	−0.66	−29.633	14.628	7434
	Bias-Corrected	7.994	11.291	0.71	−14.137	30.124	7434
	Robust	7.994	12.770	0.63	−17.035	33.022	7434
Construction	Conventional	5.784	13.565	0.43	−20.803	32.371	7434
	Bias-Corrected	15.783	13.565	1.16	−10.803	42.370	7434
	Robust	15.783	15.300	1.03	−14.205	45.771	7434
Wholesale & Retail Trade	Conventional	−38.641	40.035	−0.97	−117.108	39.826	7434
	Bias-Corrected	19.970	40.035	0.50	−58.498	98.437	7434
	Robust	19.970	44.691	0.45	−67.624	107.563	7434
Information & Communication	Conventional	−3.578	3.093	−1.16	−9.640	2.485	7434
	Bias-Corrected	3.377	3.093	1.09	−2.685	9.440	7434
	Robust	3.377	3.081	1.10	−2.661	9.415	7434
Financial & Insurance Activities	Conventional	−4.703	4.811	−0.98	−14.132	4.726	7434
	Bias-Corrected	5.515	4.811	1.15	−3.914	14.944	7434
	Robust	5.515	5.176	1.07	−4.629	15.659	7434
Real Estate Activities	Conventional	−3.355	16.026	−0.21	−34.765	28.055	7434
	Bias-Corrected	13.196	16.026	0.82	−18.214	44.606	7434
	Robust	13.196	17.390	0.76	−20.887	47.279	7434
Professional & Scientific Activities	Conventional	−24.008	17.177	−1.40	−57.673	9.657	7434
	Bias-Corrected	18.902	17.177	1.10	−14.763	52.568	7434
	Robust	18.902	17.788	1.06	−15.962	53.767	7434
Public Administration	Conventional	−7.881	16.716	−0.47	−40.644	24.882	7434
	Bias-Corrected	23.612	16.716	1.41	−9.151	56.375	7434
	Robust	23.612	17.721	1.33	−11.120	58.344	7434
Other Services	Conventional	−6.252	13.283	−0.47	−32.287	19.784	7434
	Bias-Corrected	19.801	13.283	1.49	−6.235	45.836	7434
	Robust	19.801	14.250	1.39	−8.129	47.730	7434

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km. Covariates include latitude, longitude and city dummy.

Table F.10: Covariate-adjusted RD results with robust bias-corrected CIs: Socio-economic outcomes in Poland vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	8.973***	1.998	4.49	5.057	12.890	7434
	Bias-Corrected	1.270	1.998	0.64	−2.647	5.186	7434
	Robust	1.270	2.520	0.50	−3.669	6.208	7434
Manufacturing	Conventional	36.294***	13.345	2.72	10.138	62.450	7434
	Bias-Corrected	2.532	13.345	0.19	−23.625	28.688	7434
	Robust	2.532	19.124	0.13	−34.951	40.014	7434
Construction	Conventional	55.194***	14.294	3.86	27.178	83.209	7434
	Bias-Corrected	8.034	14.294	0.56	−19.982	36.049	7434
	Robust	8.034	20.850	0.39	−32.831	48.898	7434
Wholesale & Retail Trade	Conventional	113.620***	43.082	2.64	29.186	198.063	7434
	Bias-Corrected	−28.895	43.082	−0.67	−113.334	55.544	7434
	Robust	−28.895	63.043	−0.46	−152.457	94.667	7434
Information & Communication	Conventional	15.603***	4.660	3.35	6.470	24.736	7434
	Bias-Corrected	3.282	4.660	0.70	−5.851	12.414	7434
	Robust	3.282	7.295	0.45	−11.017	17.580	7434
Financial & Insurance Activities	Conventional	11.587	4.851	2.39	2.078	21.095	7434
	Bias-Corrected	−1.375	4.851	−0.28	−10.884	8.133	7434
	Robust	−1.375	7.575	−0.18	−16.222	13.472	7434
Real Estate Activities	Conventional	33.687**	13.331	2.53	7.559	54.679	7434
	Bias-Corrected	8.391	13.331	0.63	−17.737	28.502	7434
	Robust	8.391	21.037	0.40	−32.841	49.623	7434
Professional & Scientific Activities	Conventional	56.933***	19.381	2.94	18.946	94.919	7434
	Bias-Corrected	−11.611	19.381	−0.60	−49.598	26.375	7434
	Robust	−11.611	30.803	−0.38	−71.985	48.762	7434
Public Administration	Conventional	42.960***	15.332	2.80	12.909	73.011	7434
	Bias-Corrected	−2.987	15.332	−0.19	−33.038	27.064	7434
	Robust	−2.987	24.437	−0.12	−50.883	44.908	7434
Other Services	Conventional	38.256***	11.754	3.25	15.218	61.294	7434
	Bias-Corrected	−8.894	11.754	−0.76	−31.932	14.144	7434
	Robust	−8.894	19.050	−0.47	−46.231	28.443	7434

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km. Covariates include latitude, longitude and city dummy.

Table F.11: RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Prawo i Sprawiedliwośc	Conventional	−0.078***	0.015	−5.18	−0.107	−0.048	2479
	Bias-Corrected	−0.047***	0.015	−3.11	−0.076	−0.017	2479
	Robust	−0.047**	0.021	−2.25	−0.087	−0.006	2479
Polska Jest Najważniejsza	Conventional	−0.007**	0.003	−2.09	−0.013	0.000	2479
	Bias-Corrected	−0.010***	0.003	−3.14	−0.016	−0.004	2479
	Robust	−0.010*	0.009	−1.68	−0.021	0.002	2479
Sojusz Lewicy Demokratycznej	Conventional	0.017***	0.006	2.86	0.005	0.029	2479
	Bias-Corrected	0.022***	0.006	3.69	0.011	0.034	2479
	Robust	0.022***	0.008	2.91	0.007	0.038	2479
Ruch Palikota	Conventional	0.033***	0.006	5.72	0.022	0.045	2479
	Bias-Corrected	0.026***	0.006	4.49	0.015	0.038	2479
	Robust	0.026***	0.008	3.33	0.011	0.042	2479
Polskie Stronnictwo Ludowe	Conventional	−0.049**	0.020	−2.40	−0.088	−0.009	2479
	Bias-Corrected	−0.032	0.020	−1.60	−0.072	0.007	2479
	Robust	−0.032	0.029	−1.13	−0.088	0.024	2479
Polska Partia Pracy – Sierpień 80	Conventional	−0.004**	0.001	−2.53	−0.006	0.001	2479
	Bias-Corrected	−0.003**	0.001	−2.12	−0.006	0.000	2479
	Robust	−0.003	0.002	−1.41	−0.007	0.001	2479
Platforma Obywatelska RP	Conventional	0.091***	0.021	4.33	0.050	0.132	2479
	Bias-Corrected	0.050**	0.021	2.38	0.009	0.091	2479
	Robust	0.050	0.030	1.64	−0.010	0.110	2479
Nasz Dom Polska	Conventional	−0.003***	0.001	−3.38	−0.005	−0.001	2479
	Bias-Corrected	−0.003***	0.001	−3.09	−0.004	−0.001	2479
	Robust	−0.003**	0.001	−2.26	−0.005	0.000	2479
Prawica	Conventional	0.002***	0.001	3.07	0.001	0.003	2479
	Bias-Corrected	0.0002	0.001	0.44	−0.001	0.002	2479
	Robust	0.0002	0.001	0.31	−0.002	0.002	2479
Nowa Prawica	Conventional	−0.004***	0.001	−4.97	−0.006	−0.003	2479
	Bias-Corrected	−0.004***	0.001	−4.98	−0.006	−0.003	2479
	Robust	−0.004***	0.001	−3.15	−0.007	−0.002	2479
Turnout	Conventional	−0.020	0.014	−1.45	−0.047	0.007	2479
	Bias-Corrected	−0.051***	0.014	−3.71	−0.077	−0.024	2479
	Robust	−0.051***	0.019	−2.73	−0.087	−0.014	2479
Conservative Share	Conventional	−0.076***	0.015	−4.97	−0.105	−0.046	2479
	Bias-Corrected	−0.045***	0.015	−3.05	−0.076	−0.017	2479
	Robust	−0.045**	0.021	−2.21	−0.087	−0.005	2479
Liberal-Conservative Share	Conventional	0.004	0.016	0.26	−0.028	0.036	2479
	Bias-Corrected	−0.011	0.016	−0.66	−0.043	0.021	2479
	Robust	−0.011	0.023	−0.47	−0.056	0.034	2479
Nationalist Share	Conventional	−0.078***	0.015	−5.18	−0.107	−0.048	2479
	Bias-Corrected	−0.047***	0.015	−3.11	−0.076	−0.017	2479
	Robust	−0.047**	0.021	−2.25	−0.087	−0.006	2479

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km.

Table F.12: Covariate-adjusted RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Prawo i Sprawiedliwośc	Conventional	−0.055***	0.014	−3.99	−0.083	−0.028	2479
	Bias-Corrected	−0.052***	0.014	−3.76	−0.080	−0.025	2479
	Robust	−0.052**	0.023	−2.29	−0.097	−0.008	2479
Polska Jest Najważniejsza	Conventional	−0.008	0.005	−1.61	−0.017	0.002	2479
	Bias-Corrected	−0.013***	0.005	−2.64	−0.022	−0.003	2479
	Robust	−0.013	0.009	−1.42	−0.030	0.005	2479
Sojusz Lewicy Demokratycznej	Conventional	0.018***	0.006	2.88	0.006	0.030	2479
	Bias-Corrected	0.024***	0.006	3.82	0.012	0.036	2479
	Robust	0.024***	0.008	3.16	0.009	0.039	2479
Ruch Palikota	Conventional	0.030***	0.005	5.45	0.019	0.041	2479
	Bias-Corrected	0.022***	0.005	4.07	0.012	0.033	2479
	Robust	0.022***	0.008	2.75	0.006	0.038	2479
Polskie Stronnictwo Ludowe	Conventional	−0.039*	0.023	−1.66	−0.085	0.007	2479
	Bias-Corrected	−0.038	0.023	−1.61	−0.084	0.008	2479
	Robust	−0.038	0.035	−1.09	−0.106	0.030	2479
Polska Partia Pracy – Sierpień 80	Conventional	−0.004**	0.002	−1.99	−0.007	0.000	2479
	Bias-Corrected	−0.004**	0.002	−2.07	−0.008	0.000	2479
	Robust	−0.004	0.003	−1.55	−0.009	0.001	2479
Platforma Obywatelska RP	Conventional	0.063***	0.022	2.87	0.020	0.106	2479
	Bias-Corrected	0.066***	0.022	3.02	0.023	0.109	2479
	Robust	0.066*	0.036	1.86	−0.003	0.136	2479
Nasz Dom Polska	Conventional	−0.003***	0.001	−2.77	−0.005	−0.001	2479
	Bias-Corrected	−0.001	0.001	−1.50	−0.003	0.001	2479
	Robust	−0.001	0.002	−0.94	−0.005	0.002	2479
Prawica	Conventional	0.001**	0.001	2.15	0.000	0.003	2479
	Bias-Corrected	0.001*	0.001	1.77	0.000	0.002	2479
	Robust	0.001	0.001	1.06	−0.001	0.003	2479
Nowa Prawica	Conventional	−0.004***	0.001	−3.10	−0.006	−0.001	2479
	Bias-Corrected	−0.005***	0.001	−4.57	−0.008	−0.003	2479
	Robust	−0.005***	0.002	−3.01	−0.009	−0.002	2479
Turnout	Conventional	−0.034**	0.016	−2.19	−0.065	−0.004	2479
	Bias-Corrected	−0.034**	0.016	−2.19	−0.065	−0.004	2479
	Robust	−0.034	0.022	−1.59	−0.077	0.008	2479
Conservative Share	Conventional	−0.054***	0.014	−3.87	−0.082	−0.027	2479
	Bias-Corrected	−0.051***	0.014	−3.66	−0.079	−0.024	2479
	Robust	−0.051**	0.023	−2.22	−0.097	−0.006	2479
Liberal-Conservative Share	Conventional	−0.002	0.020	−0.12	−0.041	0.036	2479
	Bias-Corrected	−0.003	0.020	−0.15	−0.041	0.035	2479
	Robust	−0.003	0.029	−0.10	−0.060	0.054	2479
Nationalist Share	Conventional	−0.056***	0.014	−3.99	−0.083	−0.028	2479
	Bias-Corrected	−0.052***	0.014	−3.76	−0.080	−0.025	2479
	Robust	−0.052**	0.023	−2.29	−0.097	−0.008	2479

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km. Covariates include latitude, longitude and city dummy.

Table F.13: Covariate-adjusted RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Prawo i Sprawiedliwośc	Conventional	-0.068^{***}	0.011	−6.01	−0.090	−0.046	2479
	Bias-Corrected	-0.044^{***}	0.011	−3.90	−0.066	−0.022	2479
	Robust	-0.044^{***}	0.017	−2.68	−0.077	−0.012	2479
Polska Jest Najważniejsza	Conventional	-0.006^{**}	0.003	−2.02	−0.012	0.000	2479
	Bias-Corrected	-0.009^{***}	0.003	−3.01	−0.015	−0.003	2479
	Robust	−0.009	0.006	−1.61	−0.020	0.002	2479
Sojusz Lewicy Demokratycznej	Conventional	0.015^{***}	0.006	2.67	0.004	0.026	2479
	Bias-Corrected	0.019^{***}	0.006	3.47	0.008	0.030	2479
	Robust	0.019^{***}	0.007	2.74	0.005	0.033	2479
Ruch Palikota	Conventional	0.032^{***}	0.005	6.57	0.022	0.042	2479
	Bias-Corrected	0.027^{***}	0.005	5.55	0.018	0.037	2479
	Robust	0.027^{***}	0.007	4.13	0.014	0.040	2479
Polskie Stronnictwo Ludowe	Conventional	-0.046^{**}	0.019	−2.37	−0.084	−0.008	2479
	Bias-Corrected	-0.037^{*}	0.019	−1.92	−0.076	0.001	2479
	Robust	−0.037	0.027	−1.36	−0.091	0.016	2479
Polska Partia Pracy – Sierpień 80	Conventional	-0.004^{***}	0.001	−2.65	−0.007	0.001	2479
	Bias-Corrected	-0.003^{**}	0.001	−2.24	−0.006	0.000	2479
	Robust	−0.003	0.002	−1.49	−0.007	0.001	2479
Platforma Obywatelska RP	Conventional	0.082^{***}	0.018	4.65	0.047	0.116	2479
	Bias-Corrected	0.053^{***}	0.018	3.03	0.019	0.088	2479
	Robust	0.053^{**}	0.026	2.04	0.002	0.105	2479
Nasz Dom Polska	Conventional	-0.003^{***}	0.001	−3.46	−0.004	−0.001	2479
	Bias-Corrected	-0.002^{***}	0.001	−2.83	−0.004	0.000	2479
	Robust	-0.002^{*}	0.001	−1.96	−0.005	0.000	2479
Prawica	Conventional	0.002	0.001	4.05	0.001	0.003	2479
	Bias-Corrected	0.001	0.001	1.00	−0.001	0.002	2479
	Robust	0.001	0.001	0.68	−0.001	0.002	2479
Nowa Prawica	Conventional	-0.004^{***}	0.001	−4.73	−0.006	−0.002	2479
	Bias-Corrected	-0.004^{***}	0.001	−4.67	−0.006	−0.002	2479
	Robust	-0.004^{***}	0.002	−2.98	−0.006	−0.001	2479
Turnout	Conventional	−0.017	0.014	−1.21	−0.043	0.010	2479
	Bias-Corrected	-0.045^{***}	0.014	−3.27	−0.072	−0.018	2479
	Robust	-0.045^{**}	0.019	−2.39	−0.082	−0.008	2479
Conservative Share	Conventional	-0.066^{***}	0.011	−5.79	−0.088	−0.044	2479
	Bias-Corrected	-0.044^{***}	0.011	−3.84	−0.066	−0.021	2479
	Robust	-0.044^{***}	0.017	−2.63	−0.076	−0.011	2479
Liberal-Conservative Share	Conventional	−0.006	0.017	0.34	−0.028	0.039	2479
	Bias-Corrected	−0.003	0.017	−0.20	−0.037	0.030	2479
	Robust	−0.003	0.024	−0.14	−0.050	0.043	2479
Nationalist Share	Conventional	-0.068^{***}	0.011	−6.01	−0.090	−0.046	2479
	Bias-Corrected	-0.044^{***}	0.011	−3.90	−0.066	−0.022	2479
	Robust	-0.044^{***}	0.016	−2.68	−0.077	0.012	2479

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km. Covariates include latitude, longitude and city dummy.

Table F.14: RD results with robust bias-corrected CIs: Socio-economic outcomes in Lithuania vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	−0.133***	0.013	−10.13	−0.159	−0.107	29963
	Bias-Corrected	−0.159***	0.013	−12.09	−0.185	−0.133	29963
	Robust	−0.159***	0.019	−8.46	−0.196	−0.122	29963
Manufacturing	Conventional	0.198***	0.031	6.40	0.137	0.258	29963
	Bias-Corrected	−0.984***	0.031	−31.88	−1.045	−0.924	29963
	Robust	−0.984***	0.049	−19.94	−1.081	−0.888	29963
Construction	Conventional	0.043	0.026	1.62	−0.009	0.094	29963
	Bias-Corrected	−1.086***	0.026	−41.34	−1.137	−1.034	29963
	Robust	−1.086***	0.045	−23.91	−1.175	−0.997	29963
Wholesale & Retail Trade	Conventional	−1.389***	0.069	−20.07	−1.524	−1.253	29963
	Bias-Corrected	−5.860***	0.069	−84.70	−5.996	−5.725	29963
	Robust	−5.860***	0.125	−46.82	−6.106	−5.615	29963
Information & Communication	Conventional	0.009***	0.007	1.20	−0.006	0.024	29963
	Bias-Corrected	−0.226***	0.007	−30.33	−0.241	−0.211	29963
	Robust	−0.226***	0.013	−17.20	−0.252	−0.200	29963
Financial & Insurance Activities	Conventional	−0.008***	0.004	−1.98	−0.015	0.000	29963
	Bias-Corrected	−0.121***	0.004	−31.51	−0.129	−0.114	29963
	Robust	−0.137***	0.006	−21.13	−0.132	−0.110	29963
Real Estate Activities	Conventional	−0.225***	0.017	−12.96	−0.259	−0.191	29963
	Bias-Corrected	−1.045***	0.017	−60.26	−1.078	−1.011	29963
	Robust	−1.045***	0.026	−39.84	−1.096	−0.993	29963
Professional & Scientific Activities	Conventional	−0.235***	0.025	−9.27	−0.284	−0.185	29963
	Bias-Corrected	−1.920***	0.025	−75.92	−1.970	−1.871	29963
	Robust	−1.920***	0.042	−45.79	−2.002	−1.838	29963
Public Administration	Conventional	−0.457***	0.021	−21.78	−0.498	−0.416	29963
	Bias-Corrected	−1.257***	0.021	−59.91	−1.298	−1.216	29963
	Robust	−1.257***	0.035	−35.40	−1.326	−1.187	29963
Other Services	Conventional	−0.332***	0.039	−8.57	−0.408	−0.256	29963
	Bias-Corrected	−2.142***	0.039	−55.32	−2.217	−2.066	29963
	Robust	−2.142***	0.065	−32.79	−2.270	−2.014	29963
Number of Employees	Conventional	11.186	8.443	1.32	−5.363	27.734	11569
	Bias-Corrected	−245.83***	8.443	−29.12	−262.38	−229.28	11569
	Robust	−245.83***	12.489	−19.68	−270.31	−221.35	11569
Income of Economic Entities	Conventional	0.398**	0.181	2.20	0.044	0.752	5457
	Bias-Corrected	−0.222	0.181	−1.23	−0.575	−0.132	5457
	Robust	−0.222	0.251	−0.88	−0.713	−0.270	5457

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km.

Table F.15: Covariate-adjusted RD results with robust bias-corrected CIs: Socio-economic outcomes in Lithuania vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	-0.143^{***}	0.014	−10.07	−0.170	−0.115	29963
	Bias-Corrected	-0.155^{***}	0.014	−10.96	−0.183	−0.127	29963
	Robust	-0.155^{***}	0.020	−7.85	−0.194	−0.116	29963
Manufacturing	Conventional	-1.099^{***}	0.037	−29.70	−1.172	−1.027	29963
	Bias-Corrected	-1.674^{***}	0.037	−45.24	−1.747	−1.602	29963
	Robust	-1.674^{***}	0.055	−30.31	−1.783	−1.566	29963
Construction	Conventional	-1.251^{***}	0.035	−35.95	−1.319	−1.183	29963
	Bias-Corrected	-1.843^{***}	0.035	−52.95	−1.911	−1.775	29963
	Robust	-1.843^{***}	0.055	−33.33	−1.951	−1.734	29963
Wholesale & Retail Trade	Conventional	-6.478^{***}	0.104	−62.26	−6.682	−6.274	29963
	Bias-Corrected	-8.723^{***}	0.104	−83.83	−8.926	−8.519	29963
	Robust	-8.723^{***}	0.165	−52.99	−9.045	−8.400	29963
Information & Communication	Conventional	-0.279^{***}	0.009	−31.65	−0.296	−0.262	29963
	Bias-Corrected	-0.379^{***}	0.009	−42.97	−0.396	−0.361	29963
	Robust	-0.379^{***}	0.014	−26.46	−0.407	−0.350	29963
Financial & Insurance Activities	Conventional	-0.115^{***}	0.004	−27.90	−0.123	−0.107	29963
	Bias-Corrected	-0.172^{***}	0.004	−41.73	−0.180	−0.164	29963
	Robust	-0.172^{***}	0.006	−27.46	−0.184	−0.160	29963
Real Estate Activities	Conventional	-1.142^{***}	0.022	−51.88	−1.185	−1.099	29963
	Bias-Corrected	-1.550^{***}	0.022	−70.40	−1.593	−1.506	29963
	Robust	-1.550^{***}	0.033	−46.49	−1.615	−1.484	29963
Professional & Scientific Activities	Conventional	-2.144^{***}	0.034	−62.57	−2.211	−2.077	29963
	Bias-Corrected	-2.946^{***}	0.034	−85.96	−3.013	−2.879	29963
	Robust	-2.946^{***}	0.053	−56.10	−3.049	−2.843	29963
Public Administration	Conventional	-1.230^{***}	0.026	−47.65	−1.280	−1.179	29963
	Bias-Corrected	-1.698^{***}	0.026	−65.80	−1.749	−1.648	29963
	Robust	-1.698^{***}	0.040	−42.91	−1.776	−1.621	29963
Other Services	Conventional	-2.127^{***}	0.048	−44.43	−2.221	−2.033	29963
	Bias-Corrected	-3.010^{***}	0.048	−62.87	−3.104	−2.916	29963
	Robust	-3.010^{***}	0.075	−40.27	−3.156	−2.863	29963
Number of Employees	Conventional	-240.29^{***}	10.099	−23.79	−260.08	−220.49	11569
	Bias-Corrected	-355.20^{***}	10.099	−35.17	−374.99	−335.41	11569
	Robust	-355.20^{***}	14.182	−25.05	−333.00	−327.41	11569
Income of Economic Entities	Conventional	-0.472^{**}	0.211	−2.24	−0.885	−0.059	5457
	Bias-Corrected	-0.669^{***}	0.211	−3.18	−1.082	−0.256	5457
	Robust	-0.669^{**}	0.284	−2.36	−1.225	−0.113	5457

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km. Covariates include latitude, longitude, city distance and city dummy.

Table F.16: Covariate-adjusted RD results with robust bias-corrected CIs: Socio-economic outcomes in Lithuania vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Agriculture	Conventional	−0.137***	0.013	−10.42	−0.163	−0.111	29963
	Bias-Corrected	−0.158***	0.013	−11.99	−0.184	−0.132	29963
	Robust	−0.158***	0.019	−8.39	−0.195	−0.121	29963
Manufacturing	Conventional	−1.022***	0.036	−28.67	−1.092	−0.952	29963
	Bias-Corrected	−1.802***	0.036	−50.53	−1.872	−1.732	29963
	Robust	−1.802***	0.054	−33.30	−1.908	−1.696	29963
Construction	Conventional	−1.120***	0.032	−35.10	−1.182	−1.057	29963
	Bias-Corrected	−1.879***	0.032	−58.91	−1.941	−1.816	29963
	Robust	−1.879***	0.051	−36.80	−1.979	−1.779	29963
Wholesale & Retail Trade	Conventional	−6.206***	0.097	−63.91	−6.397	−6.016	29963
	Bias-Corrected	−9.108***	0.097	−93.79	−9.298	−8.918	29963
	Robust	−9.108***	0.152	−59.91	−9.406	−8.810	29963
Information & Communication	Conventional	−0.355***	0.009	−37.90	−0.374	−0.337	29963
	Bias-Corrected	−0.471***	0.009	−50.26	−0.490	−0.453	29963
	Robust	−0.471***	0.015	−31.39	−0.501	−0.442	29963
Financial & Insurance Activities	Conventional	−0.110***	0.004	−26.47	−0.118	−0.102	29963
	Bias-Corrected	−0.189***	0.004	−45.58	−0.197	−0.181	29963
	Robust	−0.189***	0.006	−31.25	−0.201	−0.177	29963
Real Estate Activities	Conventional	−1.039***	0.021	−49.77	−1.080	−0.998	29963
	Bias-Corrected	−1.599***	0.021	−76.58	−1.640	−1.558	29963
	Robust	−1.599***	0.030	−53.71	−1.657	−1.541	29963
Professional & Scientific Activities	Conventional	−2.070***	0.036	−57.43	−2.140	−1.999	29963
	Bias-Corrected	−3.160***	0.036	−87.67	−3.230	−3.089	29963
	Robust	−3.160***	0.053	−59.52	−3.264	−3.055	29963
Public Administration	Conventional	−1.221***	0.024	−50.25	−1.269	−1.173	29963
	Bias-Corrected	−1.768***	0.024	−72.75	−1.815	−1.720	29963
	Robust	−1.768***	0.039	−45.59	−1.843	−1.691	29963
Other Services	Conventional	−1.976***	0.046	−42.97	−2.066	−1.886	29963
	Bias-Corrected	−3.262***	0.046	−70.93	−3.352	−3.172	29963
	Robust	−3.262***	0.072	−45.03	−3.404	−3.120	29963
Number of Employees	Conventional	−223.70***	9.817	−22.79	−242.94	−204.46	11569
	Bias-Corrected	−396.55***	9.817	−40.40	−415.79	−377.31	11569
	Robust	−396.55***	14.119	−28.09	−424.22	−368.88	11569
Income of Economic Entities	Conventional	−0.402**	0.190	−2.12	−0.773	−0.030	5457
	Bias-Corrected	−0.703***	0.190	−3.71	−1.074	−0.331	5457
	Robust	−0.703***	0.267	−2.64	−1.225	−0.180	5457

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km. Covariates include latitude, longitude, city distance and city dummy.

Table F.17: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Labor Party	Conventional	0.057**	0.025	2.33	0.009	0.106	2000
	Bias-Corrected	0.057**	0.025	2.31	0.009	0.105	2000
	Robust	0.057*	0.033	1.74	−0.007	0.121	2000
Social Democratic Party	Conventional	−0.118***	0.023	−5.21	−0.162	−0.074	2000
	Bias-Corrected	−0.081***	0.023	−3.57	−0.125	−0.036	2000
	Robust	−0.081***	0.029	−2.76	−0.138	−0.023	2000
Homeland Union	Conventional	−0.020**	0.010	−2.02	−0.039	−0.001	2000
	Bias-Corrected	−0.029***	0.010	−2.90	−0.048	−0.009	2000
	Robust	−0.029**	0.013	−2.25	−0.054	−0.004	2000
Liberals Movement	Conventional	0.013	0.013	0.95	−0.013	0.038	2000
	Bias-Corrected	−0.026**	0.013	−1.98	−0.052	0.000	2000
	Robust	−0.026	0.016	−1.58	−0.058	0.006	2000
The Way of Courage	Conventional	0.000	0.007	−0.03	−0.014	0.013	2000
	Bias-Corrected	−0.012***	0.007	−1.69	−0.025	0.002	2000
	Robust	−0.012	0.009	−1.25	−0.030	0.007	2000
Order & Justice	Conventional	0.087**	0.037	2.38	0.015	0.159	2000
	Bias-Corrected	0.120***	0.037	3.28	0.048	0.192	2000
	Robust	0.120***	0.044	2.70	0.033	0.207	2000
Poles' Electoral Action	Conventional	0.002*	0.001	1.74	−0.001	0.005	2000
	Bias-Corrected	−0.028***	0.001	−19.43	−0.031	−0.024	2000
	Robust	−0.028***	0.003	−10.94	−0.033	−0.023	2000
Peasant & Greens Union	Conventional	−0.008	0.012	−0.71	−0.031	0.014	2000
	Bias-Corrected	0.008	0.012	0.71	−0.014	0.031	2000
	Robust	0.008	0.015	0.53	−0.022	0.038	2000
Liberal & Centre Union	Conventional	−0.008*	0.005	−1.71	−0.017	0.001	2000
	Bias-Corrected	−0.006	0.005	−1.37	−0.015	0.003	2000
	Robust	−0.006	0.006	−1.13	−0.017	0.005	2000
Union YES	Conventional	0.008	0.005	1.74	−0.001	0.017	2000
	Bias-Corrected	0.011***	0.005	2.35	0.002	0.019	2000
	Robust	0.011**	0.006	1.71	−0.002	0.023	2000
Socialist People's Front	Conventional	−0.003***	0.001	−3.04	−0.005	−0.001	2000
	Bias-Corrected	−0.007***	0.001	−6.32	−0.009	−0.005	2000
	Robust	−0.007***	0.002	−4.48	−0.010	−0.004	2000
Christian Party	Conventional	−0.007***	0.003	−2.81	−0.012	−0.002	2000
	Bias-Corrected	−0.004*	0.003	−1.74	−0.009	0.001	2000
	Robust	−0.004	0.003	−1.30	−0.011	0.002	2000
National Association	Conventional	−0.002	0.002	−1.07	−0.005	0.002	2000
	Bias-Corrected	−0.004**	0.002	−2.38	−0.007	−0.001	2000
	Robust	−0.004*	0.002	−1.71	−0.009	0.001	2000
Young Lithuania	Conventional	0.0003	0.001	0.38	−0.002	0.002	2000
	Bias-Corrected	0.0004	0.001	0.45	−0.001	0.002	2000
	Robust	0.0004	0.001	0.36	−0.002	0.003	2000
Democratic Labor & Unity Party	Conventional	−0.001	0.001	−1.38	−0.003	0.001	2000
	Bias-Corrected	0.0002	0.001	−0.16	−0.002	0.002	2000
	Robust	0.0002	0.001	−0.14	−0.003	0.002	2000

Table F.17 – continued

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Emigrants' Party	Conventional	0.000	0.001	−0.17	−0.001	0.001	2000
	Bias-Corrected	−0.001	0.001	−1.34	−0.002	0.000	2000
	Robust	−0.001	0.001	−1.07	−0.002	0.001	2000
Republican Party	Conventional	0.000	0.001	−0.07	−0.001	0.001	2000
	Bias-Corrected	0.000	0.001	−0.05	−0.001	0.001	2000
	Robust	0.000	0.001	−0.04	−0.002	0.001	2000
People's Party	Conventional	0.000	0.001	0.24	−0.001	0.002	2000
	Bias-Corrected	0.001	0.001	1.22	−0.001	0.002	2000
	Robust	0.001	0.001	1.12	−0.001	0.002	2000
Turnout	Conventional	−0.013	0.018	−0.71	−0.049	0.023	2000
	Bias-Corrected	−0.003	0.018	−0.15	−0.039	0.033	2000
	Robust	−0.003	0.024	−0.11	−0.049	0.043	2000
Conservative Share	Conventional	0.074**	0.036	2.05	0.003	0.146	2000
	Bias-Corrected	0.097***	0.036	2.66	0.025	0.168	2000
	Robust	0.097**	0.042	2.29	0.014	0.179	2000
Liberal-Conservative Share	Conventional	0.059*	0.028	2.10	0.004	0.114	2000
	Bias-Corrected	0.036	0.028	1.26	−0.020	0.091	2000
	Robust	0.036	0.035	1.00	−0.034	0.105	2000
Nationalist Share	Conventional	0.065**	0.033	1.99	0.001	0.130	2000
	Bias-Corrected	0.089***	0.033	2.67	0.023	0.152	2000
	Robust	0.089**	0.041	2.13	0.007	0.169	2000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km.

Table F.18: Covariate-adjusted RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Labor Party	Conventional	0.068**	0.026	2.59	0.017	0.119	2000
	Bias-Corrected	0.068**	0.026	2.59	0.016	0.119	2000
	Robust	0.068**	0.034	2.00	0.001	0.134	2000
Social Democratic Party	Conventional	−0.085***	0.021	−3.99	−0.126	−0.043	2000
	Bias-Corrected	−0.069***	0.021	−3.25	−0.111	−0.027	2000
	Robust	−0.069***	0.026	−2.66	−0.120	−0.018	2000
Homeland Union	Conventional	−0.039***	0.010	−3.70	−0.059	−0.018	2000
	Bias-Corrected	−0.034***	0.010	−3.23	−0.054	−0.013	2000
	Robust	−0.034**	0.014	−2.48	−0.060	−0.007	2000
Liberals Movement	Conventional	−0.016	0.011	−1.39	−0.037	0.006	2000
	Bias-Corrected	−0.014	0.011	−1.30	−0.036	0.007	2000
	Robust	−0.014	0.014	−1.00	−0.043	0.014	2000
The Way of Courage	Conventional	−0.014**	0.007	−2.01	−0.028	0.000	2000
	Bias-Corrected	−0.018**	0.007	−2.58	−0.031	−0.004	2000
	Robust	−0.018**	0.009	−1.98	−0.036	0.000	2000
Order & Justice	Conventional	−0.099***	0.037	2.64	0.025	0.173	2000
	Bias-Corrected	−0.083**	0.037	2.20	0.009	0.157	2000
	Robust	−0.083*	0.044	1.87	−0.004	0.170	2000
Poles' Electoral Action	Conventional	−0.007**	0.003	−2.20	−0.014	−0.001	2000
	Bias-Corrected	−0.016***	0.003	−4.65	−0.022	−0.009	2000
	Robust	−0.016***	0.004	−3.73	−0.024	−0.024	2000
Peasant & Greens Union	Conventional	0.008	0.012	0.66	−0.016	0.032	2000
	Bias-Corrected	0.011	0.012	0.90	−0.013	0.035	2000
	Robust	0.011	0.016	0.69	−0.020	0.042	2000
Liberal & Centre Union	Conventional	−0.007	0.005	−1.34	−0.017	0.003	2000
	Bias-Corrected	−0.006	0.005	−1.14	−0.016	0.004	2000
	Robust	−0.006	0.006	−0.96	−0.018	0.006	2000
Union YES	Conventional	0.007	0.005	1.46	−0.002	0.016	2000
	Bias-Corrected	0.008*	0.005	1.75	−0.001	0.017	2000
	Robust	0.008	0.006	1.29	−0.004	0.021	2000
Socialist People's Front	Conventional	−0.004***	0.001	−3.74	−0.007	−0.002	2000
	Bias-Corrected	−0.005***	0.001	−4.58	−0.008	−0.003	2000
	Robust	−0.005***	0.002	−3.48	−0.009	−0.002	2000
Christian Party	Conventional	−0.008***	0.003	−2.91	−0.013	−0.002	2000
	Bias-Corrected	−0.006**	0.003	−2.22	−0.011	−0.001	2000
	Robust	−0.006*	0.003	−1.74	−0.012	0.001	2000
National Association	Conventional	−0.003	0.002	−1.47	−0.006	0.001	2000
	Bias-Corrected	−0.003	0.002	−1.54	−0.006	0.001	2000
	Robust	−0.003	0.002	−1.17	−0.007	0.002	2000
Young Lithuania	Conventional	0.0001	0.001	0.06	−0.002	0.002	2000
	Bias-Corrected	0.001	0.001	0.50	−0.001	0.003	2000
	Robust	0.001	0.001	0.40	−0.002	0.003	2000
Democratic Labor & Unity Party	Conventional	−0.0004	0.001	−0.41	−0.003	0.002	2000
	Bias-Corrected	0.0003	0.001	0.31	−0.002	0.002	2000
	Robust	0.0003	0.001	0.28	−0.002	0.003	2000

Table F.18 – continued

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Emigrants' Party	Conventional	0.000	0.001	−0.69	−0.001	0.001	2000
	Bias-Corrected	−0.001	0.001	−1.42	−0.002	0.000	2000
	Robust	−0.001	0.001	−1.22	−0.002	0.001	2000
Republican Party	Conventional	−0.0001	0.001	−0.20	−0.001	0.001	2000
	Bias-Corrected	−0.0001	0.001	−0.19	−0.001	0.001	2000
	Robust	−0.0001	0.001	−0.15	−0.002	0.001	2000
People's Party	Conventional	0.0002	0.001	0.39	−0.001	0.002	2000
	Bias-Corrected	0.001	0.001	1.11	−0.001	0.002	2000
	Robust	0.001	0.001	0.93	−0.001	0.002	2000
Turnout	Conventional	−0.006	0.020	−0.33	−0.045	0.032	2000
	Bias-Corrected	−0.012	0.020	−0.60	−0.050	0.027	2000
	Robust	−0.012	0.025	−0.48	−0.060	0.036	2000
Conservative Share	Conventional	0.093**	0.036	2.55	0.021	0.164	2000
	Bias-Corrected	0.073**	0.036	2.01	0.002	0.144	2000
	Robust	0.073*	0.042	1.74	−0.009	0.155	2000
Liberal-Conservative Share	Conventional	0.032	0.029	1.10	−0.025	0.088	2000
	Bias-Corrected	0.019	0.029	0.66	−0.038	0.076	2000
	Robust	0.019	0.034	0.57	−0.047	0.085	2000
Nationalist Share	Conventional	0.058*	0.035	1.65	−0.011	0.128	2000
	Bias-Corrected	0.047	0.035	1.33	−0.022	0.116	2000
	Robust	0.047	0.042	1.11	−0.036	0.130	2000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 60 km. Covariates include latitude, longitude, city distance and city dummy.

Table F.19: Covariate-adjusted RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia – Bandwidth of 100 km

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Labor Party	Conventional	0.079**	0.025	3.20	0.031	0.127	2000
	Bias-Corrected	0.053**	0.025	2.14	0.004	0.101	2000
	Robust	0.053	0.033	1.59	−0.012	0.117	2000
Social Democratic Party	Conventional	−0.078***	0.021	−3.67	−0.120	−0.036	2000
	Bias-Corrected	−0.076***	0.021	−3.56	−0.118	−0.034	2000
	Robust	−0.076***	0.025	−2.99	−0.126	−0.026	2000
Homeland Union	Conventional	−0.047***	0.010	−4.60	−0.068	−0.027	2000
	Bias-Corrected	−0.030***	0.010	−2.92	−0.050	−0.010	2000
	Robust	−0.030**	0.014	−2.11	−0.058	−0.002	2000
Liberals Movement	Conventional	−0.014	0.011	−1.24	−0.035	0.008	2000
	Bias-Corrected	−0.023**	0.011	−2.09	−0.044	−0.001	2000
	Robust	−0.023	0.014	−1.61	−0.051	0.005	2000
The Way of Courage	Conventional	−0.017**	0.007	−2.34	−0.030	−0.003	2000
	Bias-Corrected	−0.005	0.007	−0.77	−0.019	0.008	2000
	Robust	−0.005	0.009	−0.58	−0.024	0.013	2000
Order & Justice	Conventional	−0.091**	0.037	2.48	0.019	0.163	2000
	Bias-Corrected	−0.105***	0.037	2.85	0.033	0.177	2000
	Robust	−0.105**	0.044	2.38	0.019	0.191	2000
Poles' Electoral Action	Conventional	−0.004**	0.002	−1.99	−0.007	0.000	2000
	Bias-Corrected	−0.025***	0.002	−14.26	−0.029	−0.022	2000
	Robust	−0.025***	0.003	−8.90	−0.031	−0.020	2000
Peasant & Greens Union	Conventional	0.008	0.012	0.68	−0.015	0.030	2000
	Bias-Corrected	0.011	0.012	0.93	−0.012	0.033	2000
	Robust	0.011	0.016	0.68	−0.020	0.042	2000
Liberal & Centre Union	Conventional	−0.007	0.005	−1.49	−0.016	0.002	2000
	Bias-Corrected	−0.006	0.005	−1.21	−0.015	0.003	2000
	Robust	−0.006	0.006	−0.99	−0.017	0.006	2000
Union YES	Conventional	0.007	0.005	1.42	−0.002	0.015	2000
	Bias-Corrected	0.011**	0.005	2.46	0.002	0.020	2000
	Robust	0.011*	0.006	1.80	−0.001	0.023	2000
Socialist People's Front	Conventional	−0.004***	0.001	−3.61	−0.006	−0.002	2000
	Bias-Corrected	−0.007***	0.001	−6.18	−0.009	−0.005	2000
	Robust	−0.007***	0.002	−4.40	−0.010	−0.004	2000
Christian Party	Conventional	−0.010***	0.002	−4.05	−0.015	−0.005	2000
	Bias-Corrected	−0.006**	0.002	−2.31	−0.010	−0.001	2000
	Robust	−0.006*	0.003	−1.74	−0.012	0.001	2000
National Association	Conventional	−0.002	0.002	−1.33	−0.005	0.001	2000
	Bias-Corrected	−0.004*	0.002	−2.15	−0.007	0.000	2000
	Robust	−0.004	0.002	−1.55	−0.008	0.001	2000
Young Lithuania	Conventional	−0.001	0.001	−0.65	−0.003	0.001	2000
	Bias-Corrected	0.001	0.001	1.31	−0.001	0.003	2000
	Robust	0.001	0.001	1.03	−0.001	0.004	2000
Democratic Labor & Unity Party	Conventional	−0.001	0.001	−0.89	−0.003	0.001	2000
	Bias-Corrected	−0.0002	0.001	−0.18	−0.002	0.002	2000
	Robust	−0.0002	0.001	−0.15	−0.003	0.002	2000

Table F.19 – continued

Outcome Variable	Method	Coefficient	Std. Err.	z	95% Confidence Interval		N
Emigrants' Party	Conventional	−0.0004	0.001	−0.91	−0.001	0.001	2000
	Bias-Corrected	−0.001	0.001	−1.06	−0.002	0.001	2000
	Robust	−0.001	0.001	−0.85	−0.002	0.000	2000
Republican Party	Conventional	0.0000	0.001	−0.06	−0.001	0.001	2000
	Bias-Corrected	0.0001	0.001	0.11	−0.001	0.001	2000
	Robust	0.0001	0.001	0.08	−0.001	0.002	2000
People's Party	Conventional	0.000	0.001	−0.03	−0.001	0.001	2000
	Bias-Corrected	0.001*	0.001	1.67	0.000	0.003	2000
	Robust	0.001	0.001	1.56	0.000	0.003	2000
Turnout	Conventional	−0.010	0.019	−0.53	−0.047	0.027	2000
	Bias-Corrected	0.003	0.019	0.13	−0.035	0.040	2000
	Robust	0.003	0.024	0.11	−0.045	0.050	2000
Conservative Share	Conventional	0.085**	0.036	2.33	0.014	0.156	2000
	Bias-Corrected	0.086**	0.036	2.36	0.014	0.157	2000
	Robust	0.086**	0.042	2.05	0.004	0.168	2000
Liberal-Conservative Share	Conventional	0.017	0.028	0.60	−0.039	0.073	2000
	Bias-Corrected	0.027	0.028	0.96	−0.028	0.083	2000
	Robust	0.027	0.034	0.81	−0.039	0.093	2000
Nationalist Share	Conventional	0.041	0.035	1.18	−0.027	0.109	2000
	Bias-Corrected	0.072**	0.035	2.09	0.005	0.140	2000
	Robust	0.072*	0.042	1.72	−0.010	0.155	2000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Bandwidth is 100 km. Covariates include latitude, longitude, city distance and city dummy.

Table F.20: Estimation results for the similarities with East Prussia, Lithuania and Poland – Bandwidth of 60 km & 100 km

Bandwidth Dep. variable	60 km turnout (1)	60 km cons. (2)	60 km national. (3)	100 km turnout (4)	100 km cons. (5)	100 km national. (6)
LT	0.089^{***}	-0.209^{***}	-0.136^{***}	0.088^{***}	-0.260^{***}	-0.141^{***}
	(0.010)	(0.021)	(0.020)	(0.009)	(0.019)	(0.018)
LTEP	-0.017	0.128^{***}	0.091^{***}	-0.024^{**}	0.111^{***}	0.033^{**}
	(0.012)	(0.023)	(0.021)	(0.010)	(0.016)	(0.015)
PLEP	-0.035^{***}	-0.089^{***}	-0.099^{***}	-0.034^{***}	-0.143^{***}	-0.135^{***}
	(0.013)	(0.016)	(0.015)	(0.012)	(0.014)	(0.013)
City	0.011	-0.094^{***}	-0.062^{***}	0.038^{***}	-0.086^{***}	0.009
	(0.012)	(0.019)	(0.017)	(0.005)	(0.006)	(0.006)
Dist. to EP border (tkm)	0.604^{***}	0.396^{**}	-0.022	0.475^{***}	-0.390^{*}	-0.555^{***}
	(0.143)	(0.193)	(0.204)	(0.061)	(0.104)	(0.097)
Dist. to EP border (EP) (tkm)	0.458	-1.673^{***}	-1.388^{***}	-0.219	-0.503^{***}	-0.386
	(0.420)	(0.438)	(0.397)	(0.317)	(0.277)	(0.260)
Location	Yes	Yes	Yes	Yes	Yes	Yes
Observations	894	894	894	1611	1611	1611
R-squared	0.406	0.280	0.150	0.369	0.358	0.180
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000
Lithuania vs. Poland in EP	0.106^{***}	0.008	0.054^{**}	0.099^{***}	-0.006	0.027
Δ LT vs. PL in EP / outside EP	0.018	-0.201^{***}	-0.082^{***}	0.011	-0.254^{***}	-0.114^{***}

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

F.2 East Prussian refugees in West Germany

Table F.21: Estimation results for parties against the Oder-Neisse line or those not taking position

Method	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	IV	Robust OLS	Robust OLS
Dependent variable	cons_ noaligned	cons_ noaligned	cons_noline	all_noline	all_noline	all_noline	no_wt_dp	no_wt_dp
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of EP expellees	0.411	−0.047	2.858***	2.357***	1.868***	2.221***	−0.849***	−0.298
	(0.374)	(0.844)	(0.616)	(0.735)	(0.707)	(0.646)	(0.243)	(0.443)
City	0.002	0.005	0.037***	0.048***	−0.018	0.055***	0.038**	0.023
	(0.016)	(0.018)	(0.013)	(0.015)	(0.020)	(0.018)	(0.017)	(0.016)
Nationalistic votes '33	0.945***	0.950***	−0.082**	−0.712***	−0.379***	−0.769***	−0.763***	−0.654***
	(0.047)	(0.048)	(0.034)	(0.048)	(0.072)	(0.044)	(0.042)	(0.049)
Male	−0.104**	−2.633		0.654	2.845***	1.285	2.881***	0.813
	(0.870)	(1.026)		(1.013)	(0.970)	(0.975)	(0.729)	(0.903)
Old (65+)					2.181**			
					(1.051)			
Protestants					0.240***			
					(0.046)			
Workers					−0.554***			
					(0.160)			
Entrepreneurs					−1.030**			
					(0.496)			
Trade & finance					0.417**			
					(0.209)			
Wage ratio (EP to host)					−0.044	−0.271***		
					(0.092)	(0.095)		
Other eastern regions	No	Yes	Yes	Yes	Yes	No	No	Yes
Observations	162	162	162	162	162	162	162	162

Table F.21 – continued

Method	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	IV	Robust OLS	Robust OLS
Dependent variable	cons_ noaligned	cons_ noaligned	cons_noline	all_noline	all_noline	all_noline	no_wt_dp	no_wt_dp
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R-squared	0.690	0.774	0.757	0.803	0.886	0.726	0.649	0.709
Prob > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. Dependent variables are as follows:

cons_noaligned	conservative parties aligned with the Neisse-Oder line
cons_noline	conservative parties against the Neisse-Oder line
all_noline	all parties against the Neisse-Oder line
no_wt_dp	all parties against the Neisse-Oder line, except Deutsche Partei (German Party)

Table F.22: Estimation results for selected parties in West Germany

Method	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	Robust OLS	IV	Robust OLS	Robust OLS
Dep. variable	dkp_share	dp_share	dp_share	dp_share	fdp_share	fdp_share	fdp_share	fdp_share	kpd_share	kpd_share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share of EP expellees	0.203	2.655***	1.210**	2.032***	−1.112***	−1.705***	−1.566***	−1.128***	−0.484***	0.094
	(0.274)	(0.563)	(0.580)	(0.694)	(0.329)	(0.349)	(0.371)	(0.292)	(0.071)	(0.195)
City	0.012**	0.025**	0.031		−0.055***	−0.041***	−0.044***	−0.024	0.032***	0.030***
	(0.005)	(0.011)	(0.028)		(0.013)	(0.013)	(0.013)	(0.016)	(0.007)	(0.008)
Nationalistic votes '33	−0.024	−0.058**	−0.169**	−0.261***	−0.335***		−0.043	−0.351***	−0.098***	−0.110***
	(0.019)	(0.025)	(0.063)	(0.062)	(0.045)		(0.057)	(0.044)	(0.016)	(0.017)
Male	0.234	−0.159		7.722***	−1.629**		0.201	−1.361**	1.841***	1.152***
	(0.254)	(0.554)		(2.376)	(0.707)		(0.918)	(0.654)	(0.350)	(0.383)
Old (65+)				5.462***		2.189***	1.299*			
				(1.854)		(0.837)	(0.747)			
Protestants						0.207***	0.197***			
						(0.026)	(0.036)			
Entrepreneurs						−0.342	0.410**			
						(0.462)	(0.180)			
Agriculture						0.175*				
						(0.105)				
Trade & finance				−0.492**						
				(0.238)						
Other eastern regions	Yes	Yes	No	Yes[a]	Yes	Yes	Yes	No	No	Yes
Observations	162	162	34	34	162	163	162	162	162	162
R-squared	0.459	0.758	0.303	0.742	0.559	0.671	0.666	0.426	0.420	0.553
Prob > F	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.

[a] One significant region only

Table F.23: "Placebo" regression for voting outcomes in 1933

Method	Robust OLS	IV	Robust OLS	IV	Robust OLS	IV	Robust OLS	IV
Dep. variable	cons. share '33	cons. share '33	cons. share '33	cons. share '33	nat. share '33	nat. share '33	nat. share '33	nat. share '33
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of EP expellees	-1.853^{***}	-0.000005^{***}	0.492^{**}	0.000007^{***}	1.945^{***}	0.000006^{***}	0.901^{***}	-0.000003^{*}
	(0.231)	(0.000002)	(0.191)	(0.000002)	(0.244)	(0.000001)	(0.291)	(0.000002)
City	-0.081^{***}	-0.063^{*}	0.026^{***}	0.033	-0.025	-0.045	-0.041^{***}	-0.062^{***}
	(0.024)	(0.033)	(0.014)	(0.024)	(0.016)	(0.028)	(0.013)	(0.020)
Conservative votes '20			0.392^{*}	0.418^{***}				
			(0.050)	(0.060)				
Nationalistic votes '20							0.419^{***}	0.177^{*}
							(0.098)	(0.107)
Protestants			-0.303^{***}	-0.382^{***}			0.107^{**}	0.311^{***}
			(0.033)	(0.046)			(0.051)	(0.072)
Observations	162	162	160	160	162	162	160	160
R-squared	0.145	0.060	0.844	0.614	0.222	0.039	0.660	0.533
Prob > F	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. IV instrument: latitude (East Prussian share as endogenous variable).

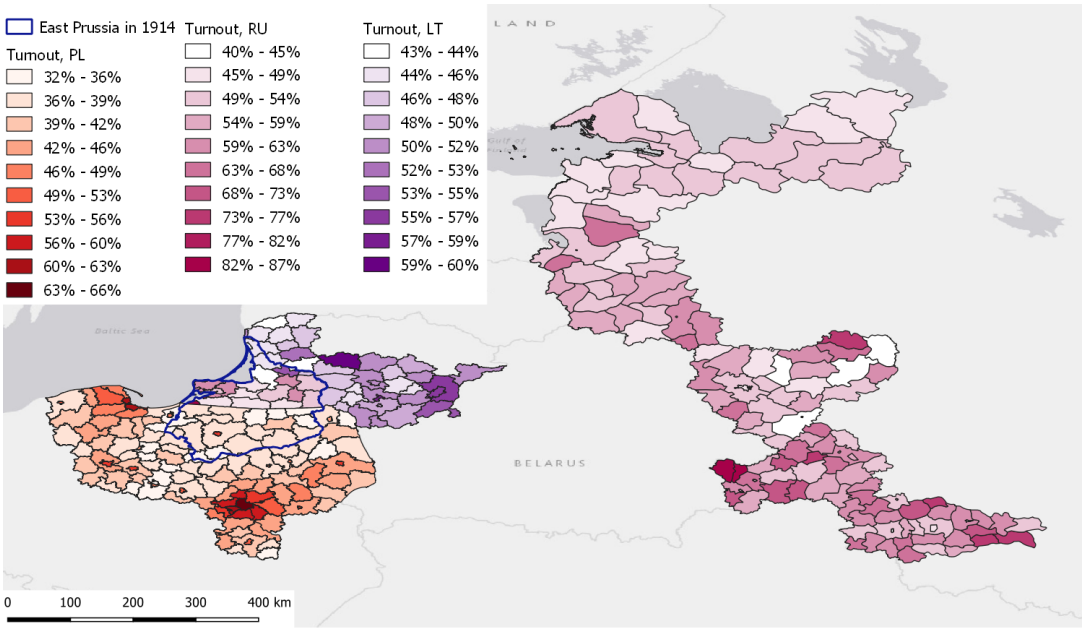
Table F.24: Spatial autoregression for conservative and nationalistic votes

Method	gs2sls	gs2sls	gs2sls	gs2sls	gs2sls	gs2sls	gs2sls	gs2sls	gs2sls	ml	ml
Dep. variable	cons.	cons.	cons.	cons.	nat.	nat.	nat.	nat.	nat.	nat.	nat.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Share of EP expellees	2.402***	2.267***	1.865***	1.825***	2.017***	2.272***	2.187***	1.953***	2.164***	2.095***	2.069***
	(0.343)	(0.344)	(0.448)	(0.502)	(0.355)	(0.333)	(0.369)	(0.349)	(0.374)	(0.141)	(0.164)
City	0.032	0.032	0.028*	0.022	0.022**	0.026**	0.027***	0.026***	0.026**	0.023**	0.024**
	(0.021)	(0.021)	(0.015)	(0.015)	(0.010)	(0.012)	(0.010)	(0.009)	(0.010)	(0.011)	(0.011)
Wage ratio	−0.511***	−0.501***	−0.294***	−0.190*	0.140***	0.107**	0.028	0.022	0.033	0.135***	
(EP to host)	(0.086)	(0.087)	(0.094)	(0.113)	(0.031)	(0.051)	(0.054)	(0.035)	(0.065)	(0.048)	
Conservative votes '20	0.560***	0.569***	0.541***	0.535***							
	(0.053)	(0.052)	(0.046)	(0.042)							
Nationalistic votes '20					−0.016	0.011	−0.042	−0.027	−0.050	−0.007	−0.037
					(0.028)	(0.023)	(0.028)	(0.026)	(0.033)	(0.028)	(0.029)
Spatial lag											
Dep. variable	0.072			0.091	0.959*			2.031***		0.628***	0.449
	(0.069)			(0.172)	(0.512)			(0.481)		(0.184)	(0.674)
Share of EP expellees		1.352		3.143		−0.357		−2.608***	0.494		0.561
		(1.140)		(2.574)		(0.741)		(0.481)	(1.404)		(2.158)
Error term			3.052***	6.502			3.015***	0.736	2.822***		0.898***
			(0.697)	(4.114)			(0.556)	(2.238)	(0.473)		(0.152)
Average impact of East Prussia											
Direct	2.402***	2.267***	1.865***	1.828***	2.314	2.272***	2.187***	2.296	2.164***	2.115***	2.081***
	(0.343)	(0.344)	(0.448)	(0.503)	(3.587)	(0.333)	(0.369)	(6.697)	(0.374)	(0.138)	(0.165)
Indirect	0.163	1.183		3.165*	37.504	−0.312		6.952	0.432	2.932	2.283
	(0.173)	(0.998)		(1.670)	(482.14)	(0.649)		(159.37)	(1.229)	(2.195)	(2.464)
Total	2.564***	3.450***	1.865***	4.994***	39.818	1.959***	2.187***	9.249	2.596**	5.047**	4.364*
	(0.438)	(0.970)	(0.448)	(1.465)	(485.72)	(0.494)	(0.369)	(166.06)	(1.155)	(2.184)	(2.465)
Observations	161	161	161	161	161	161	161	161	161	161	161
Pseudo R-sq.	0.576	0.580	0.548	0.551	0.313	0.643	0.627	0.331	0.617	0.628	0.513
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: ***, **, * - significance at 1%, 5% and 10%, respectively. Standard errors in parentheses. gs2sls = heteroskedastic generalized spatial two-stage least squares; ml = robust maximum likelihood

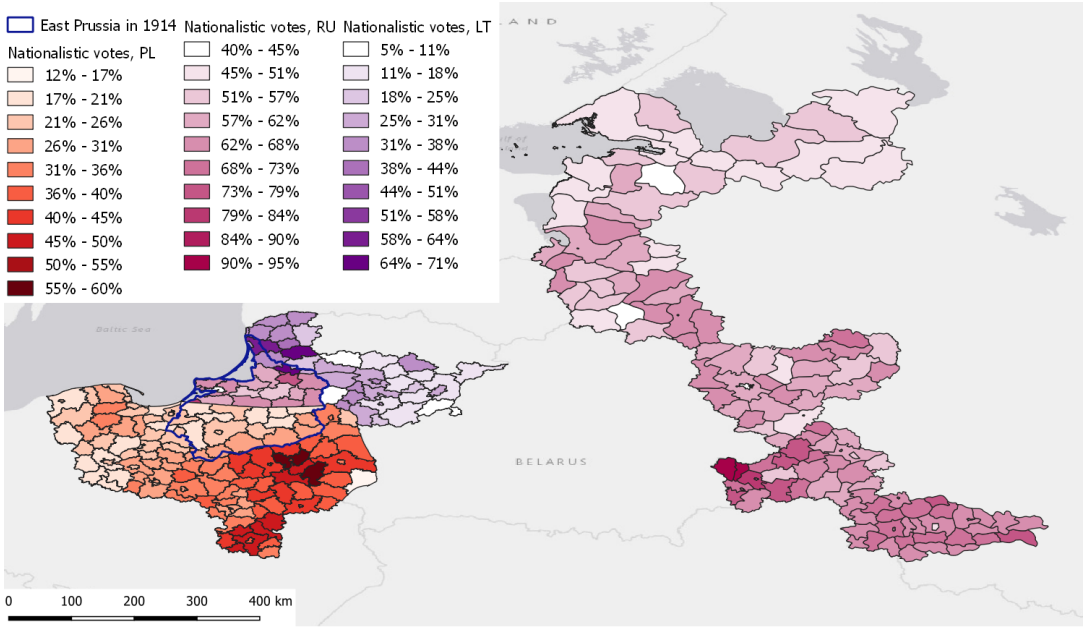
G Figures

Figure G.1: Turnout in parliament elections in Lithuania (2012), Poland and Russia (2011): East Prussia and neighboring regions



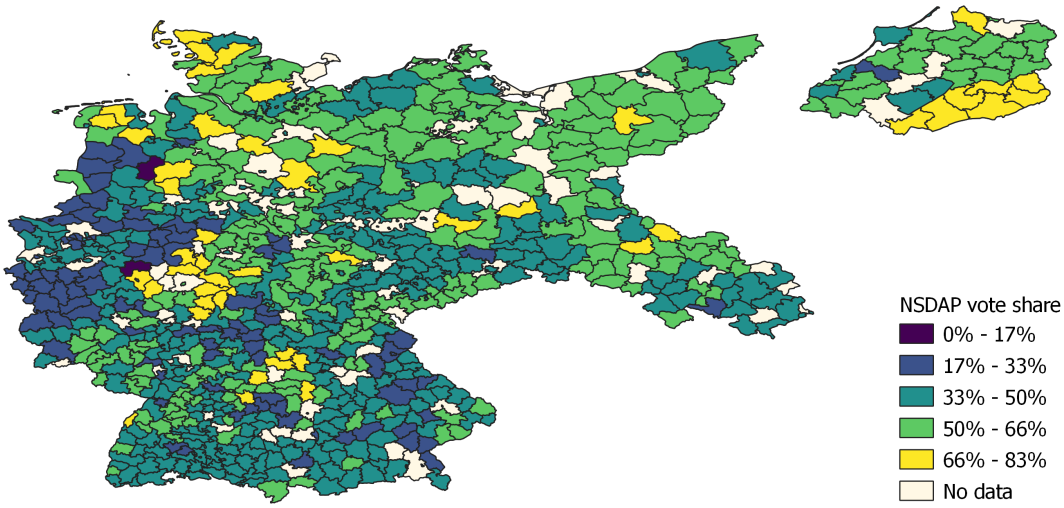
Source: Authors' work. Base map: GADM & ESRI Gray, election data: see Table E.1

Figure G.2: Share of votes for nationalistic parties in parliament elections in Lithuania (2012), Poland and Russia (2011): East Prussia and neighboring regions



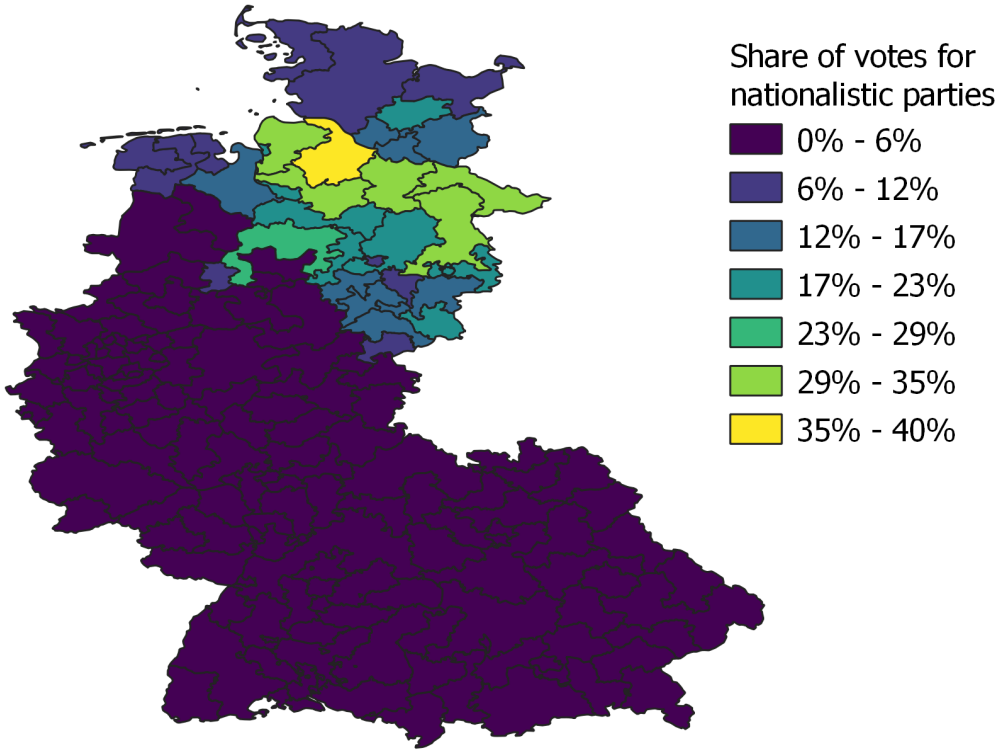
Source: Authors' work. Base map: GADM & ESRI Gray, election data: see Table E.1

Figure G.3: Share of votes for NSDAP in Germany, 1933



Source: Authors' work. Base map: MPIDR and CGG 2011b, election data: see Table E.1

Figure G.4: Share of votes for nationalistic parties in parliament elections in the FRG, 1949



Source: Authors' work. Base map: MPIDR and CGG 2011a, election data: see Table E.1

Figure G.5: RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia (all parties)

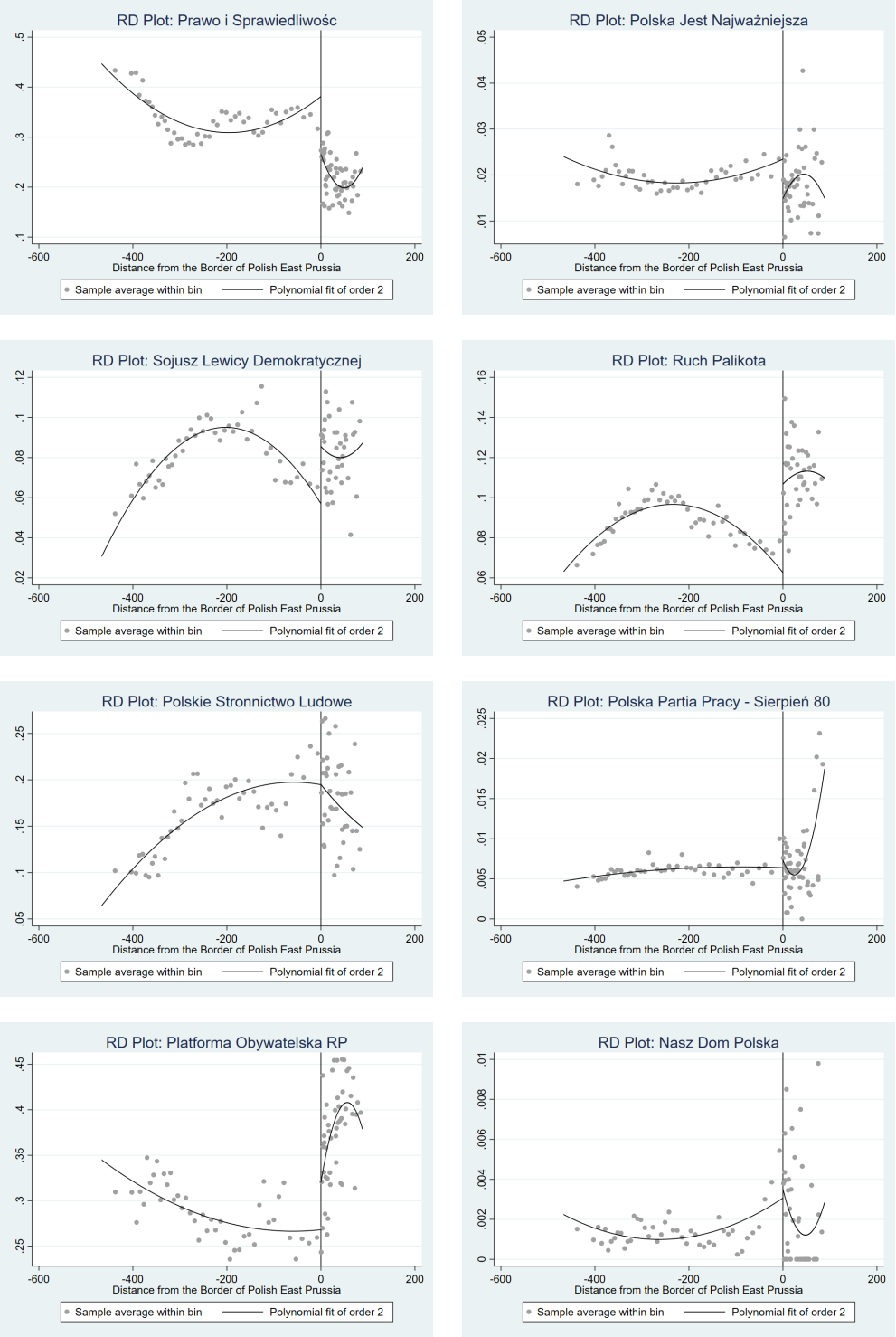


Figure G.5: RD results with robust bias-corrected CIs: Political outcomes in Poland vs. East Prussia (all parties) – continued

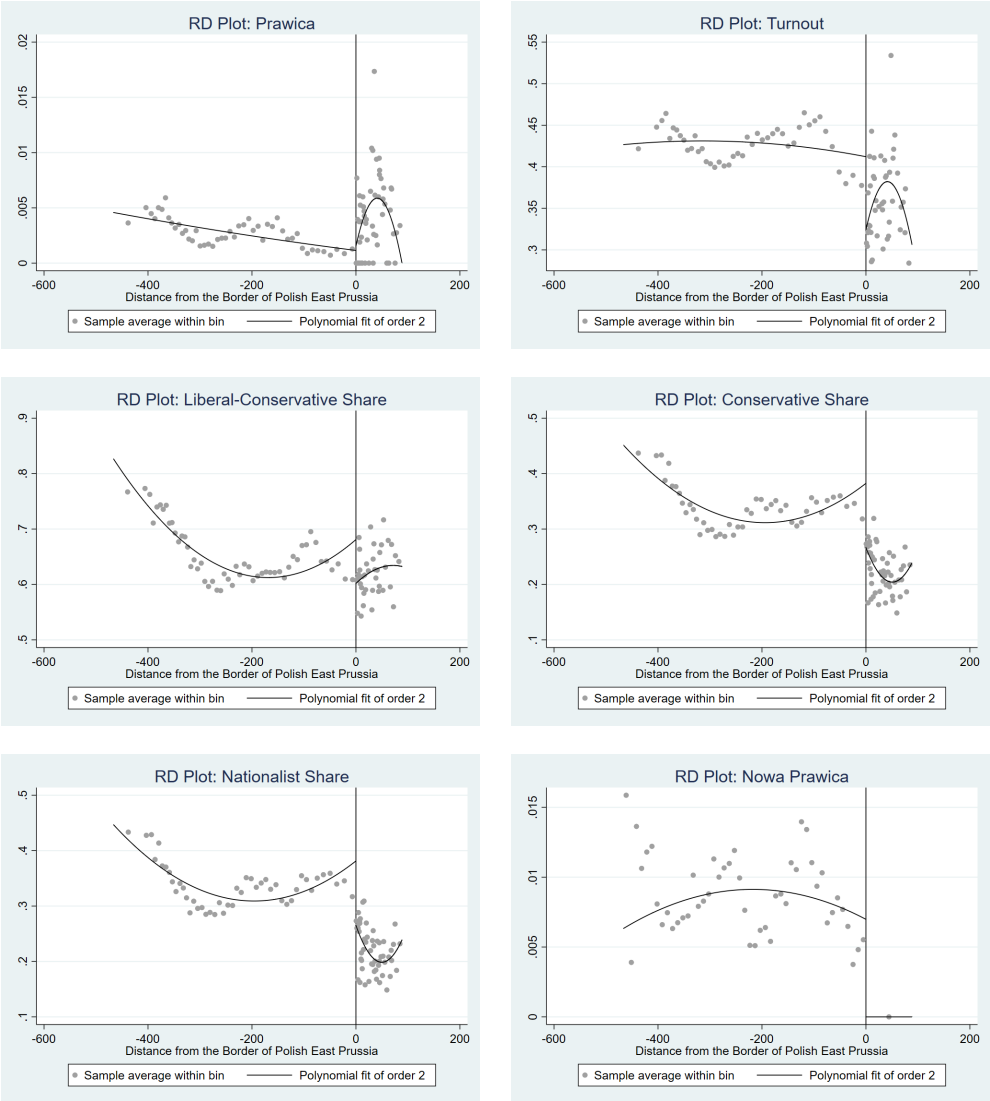


Figure G.6: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia (all parties)

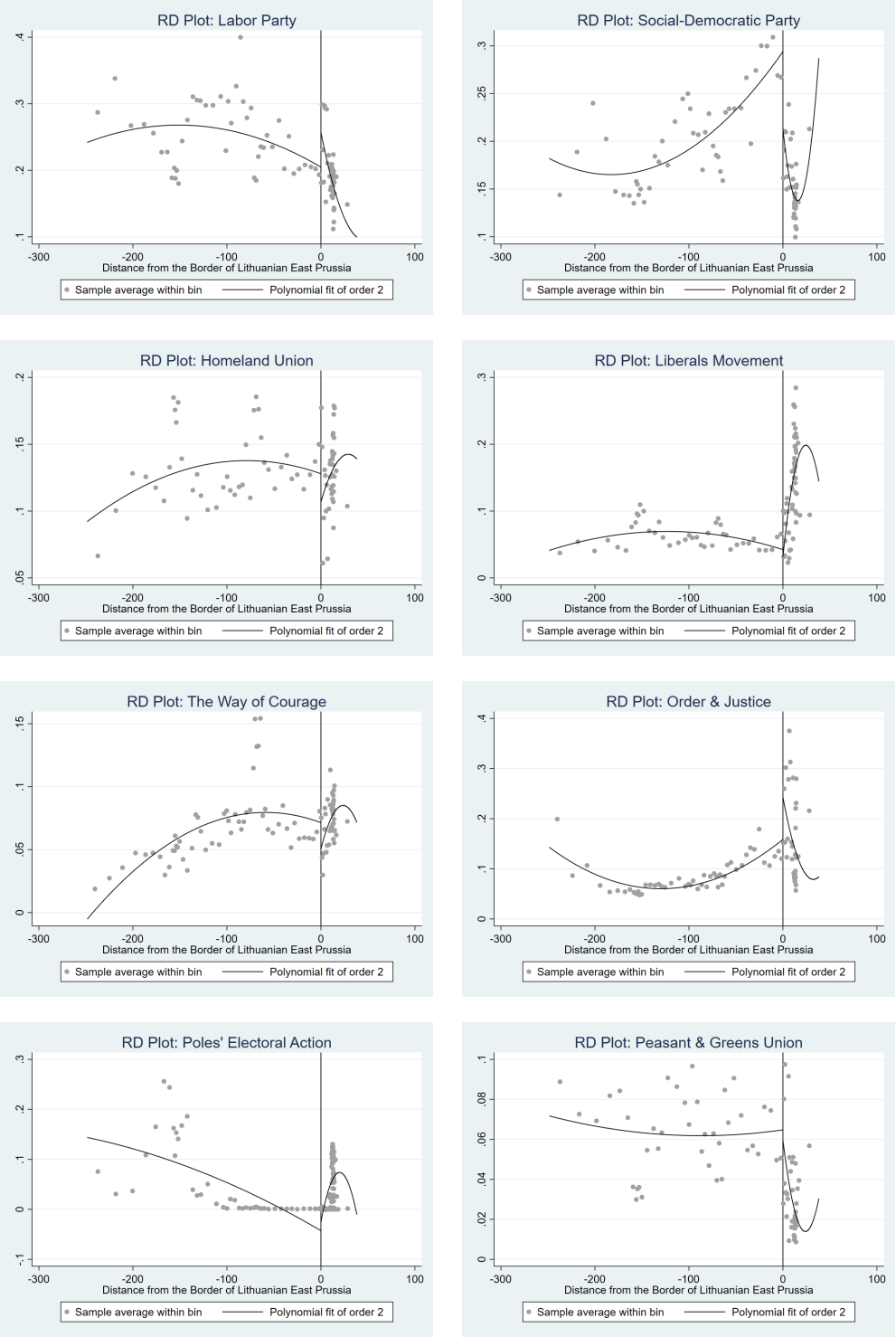


Figure G.6: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia (all parties) – continued

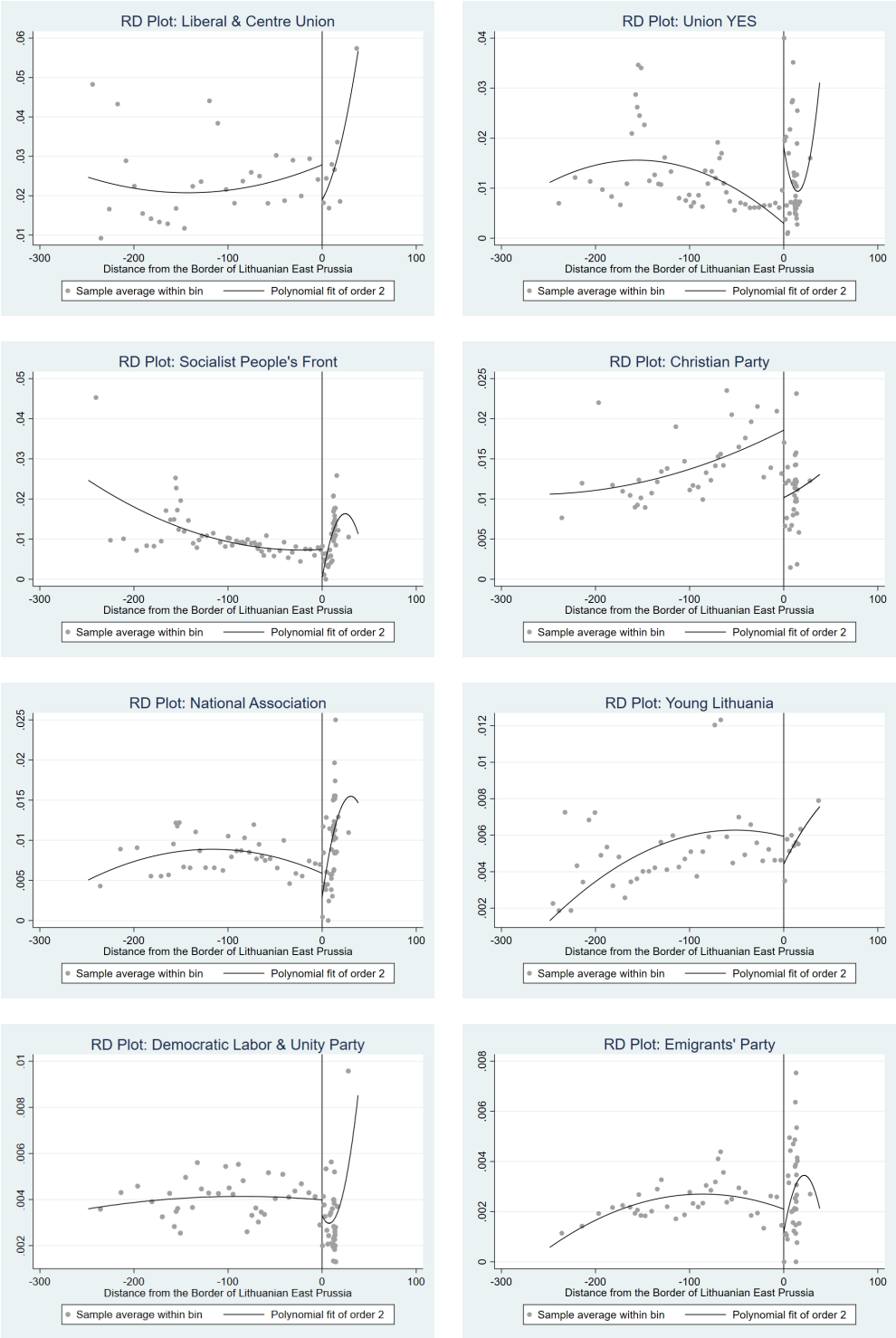


Figure G.6: RD results with robust bias-corrected CIs: Political outcomes in Lithuania vs. East Prussia (all parties) – continued

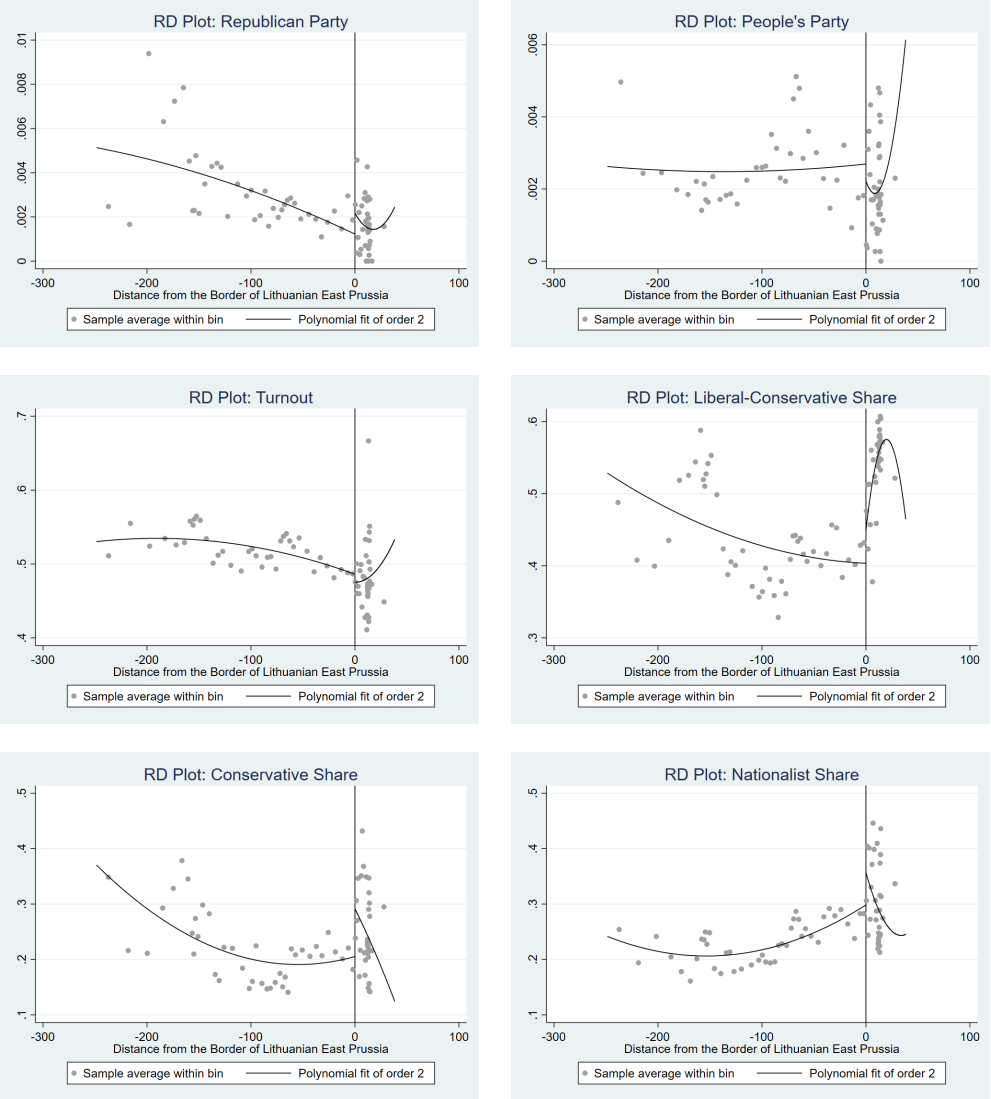


Figure G.7: RD results with robust bias-corrected CIs: Socio-economic outcomes in Poland vs. East Prussia

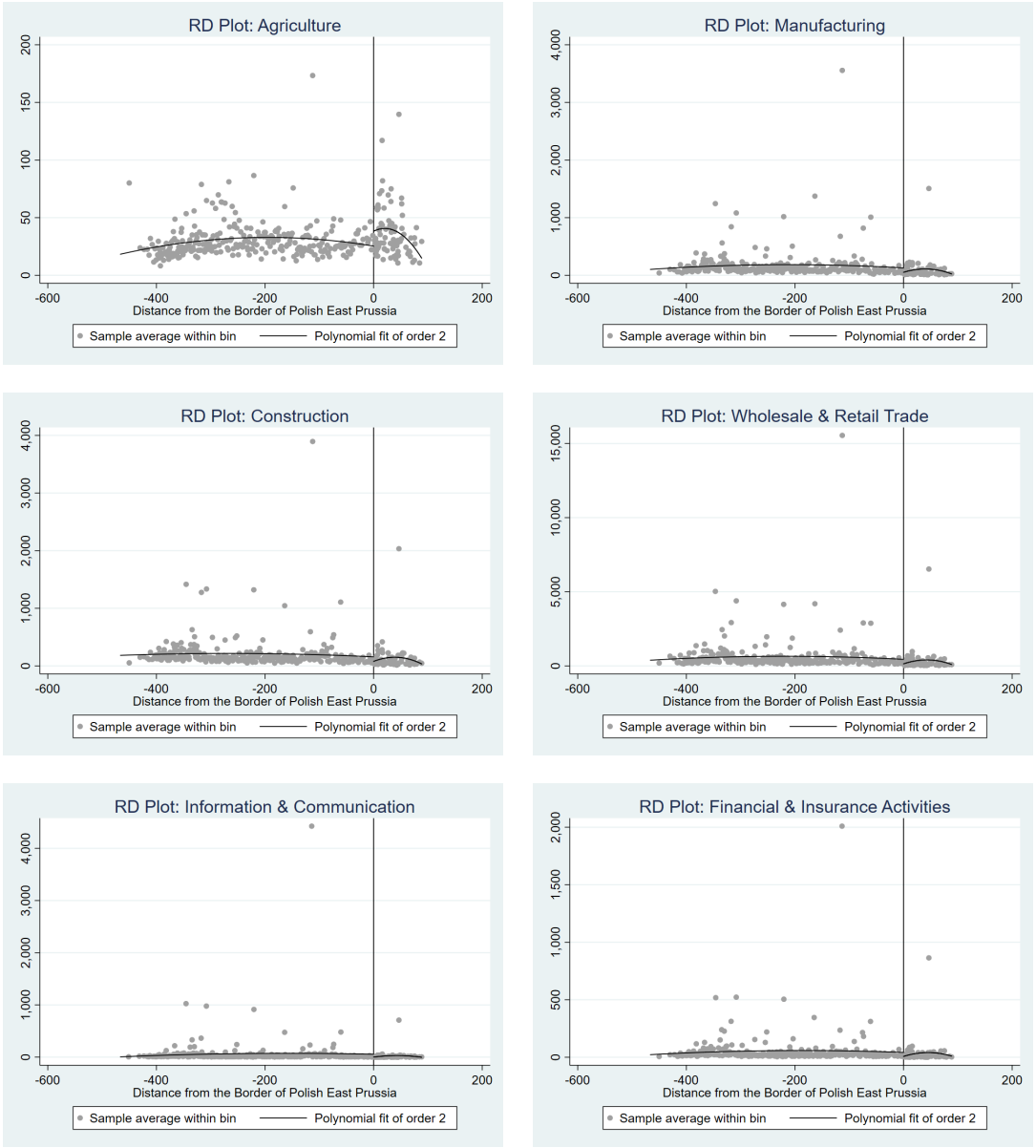


Figure G.7: RD results with robust bias-corrected CIs: Socio-economic outcomes in Poland vs. East Prussia – continued

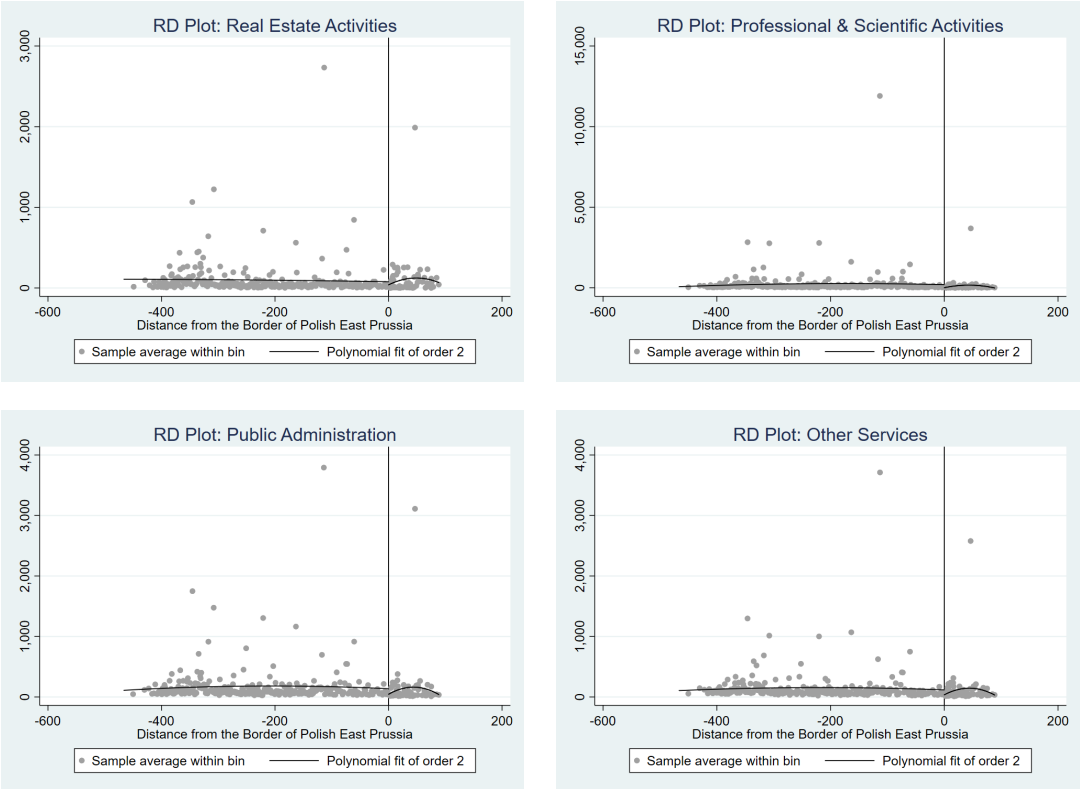


Figure G.8: RD results with robust bias-corrected CIs: Socio-economic outcomes in Lithuania vs. East Prussia

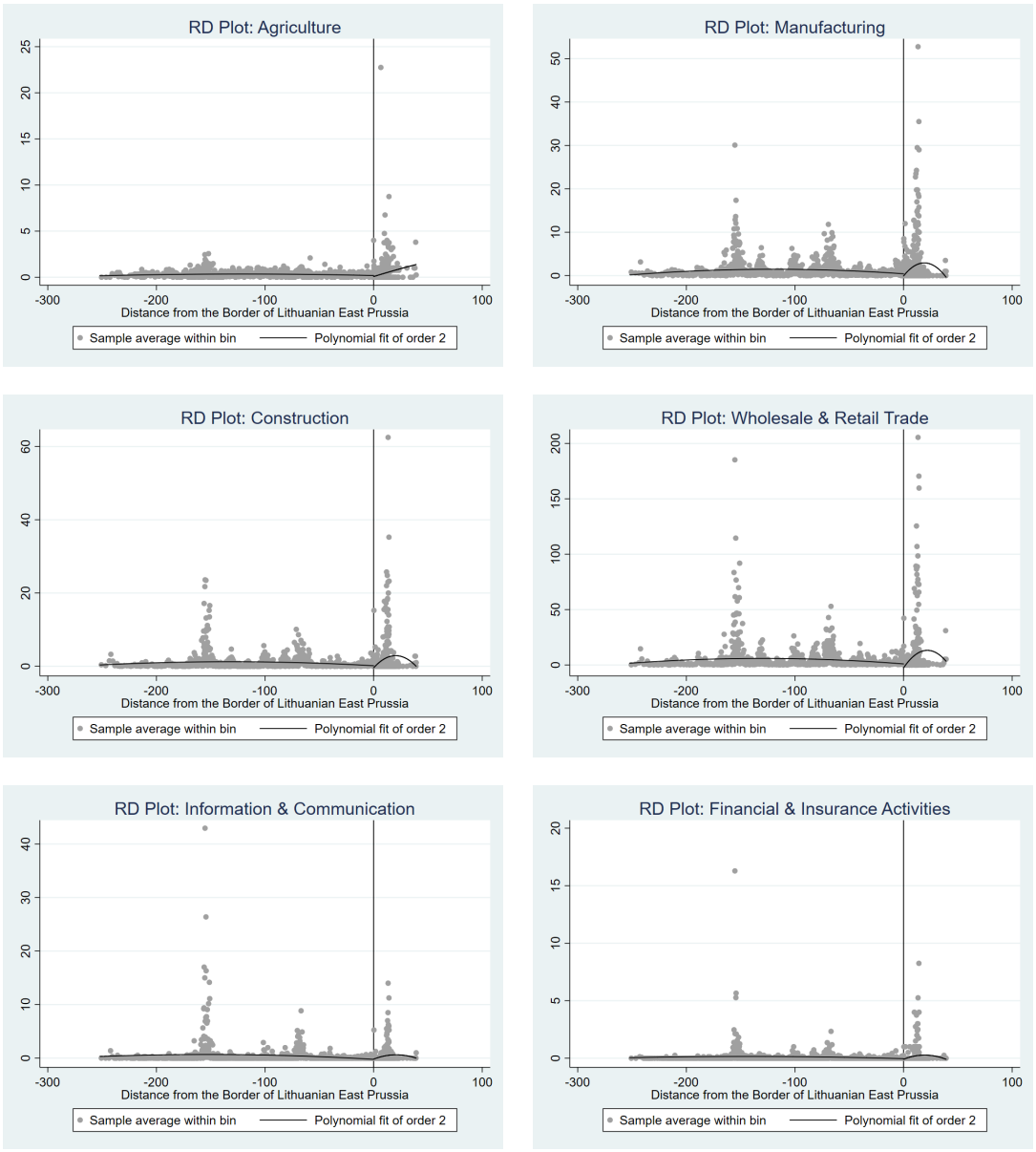


Figure G.8: RD results with robust bias-corrected CIs: Socio-economic outcomes in Lithuania vs. East Prussia – continued

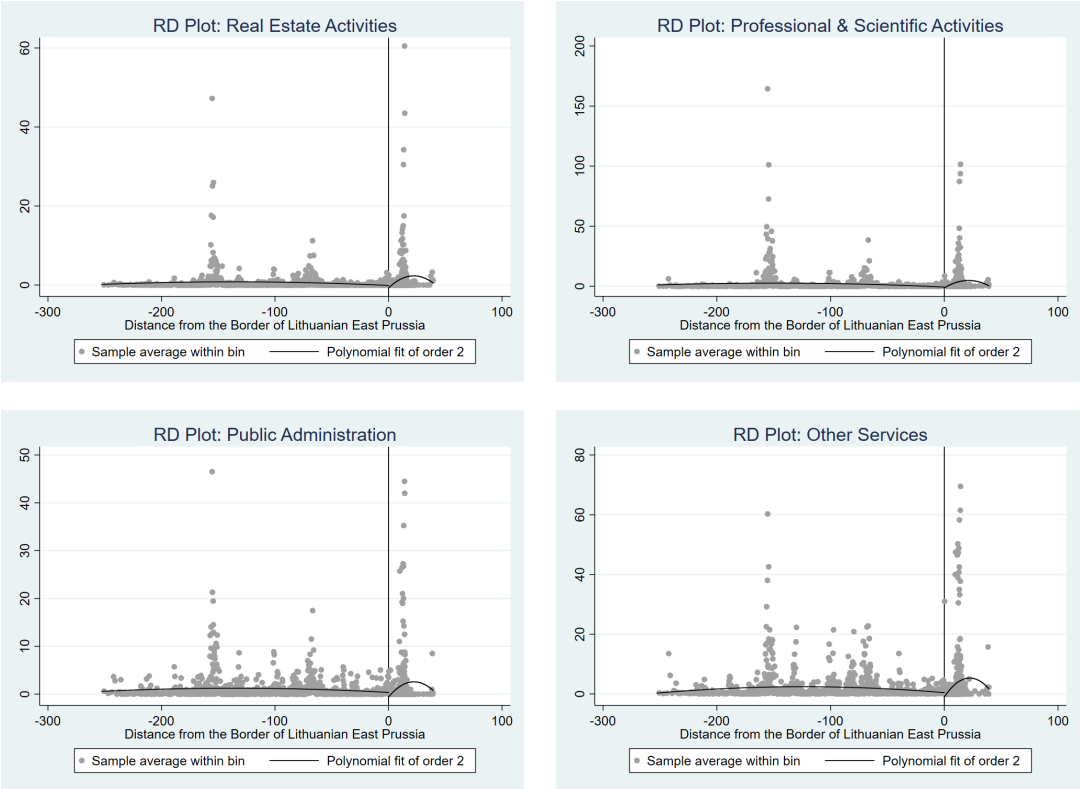
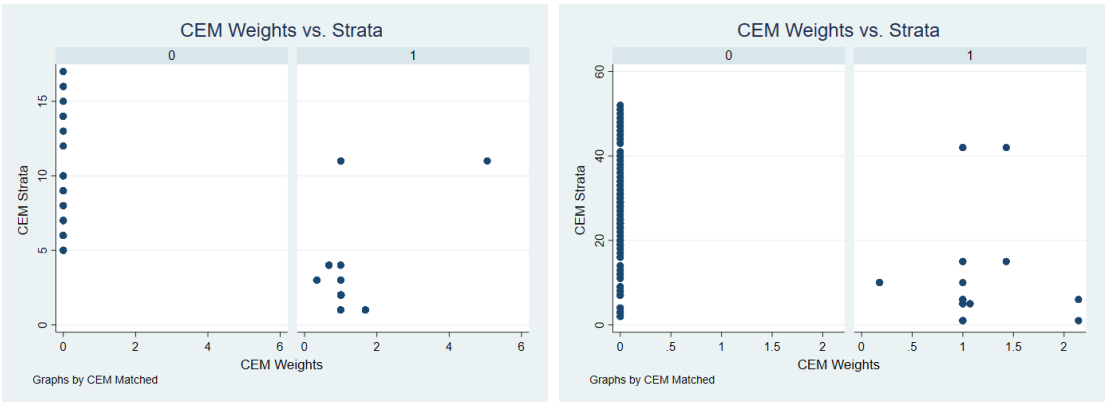
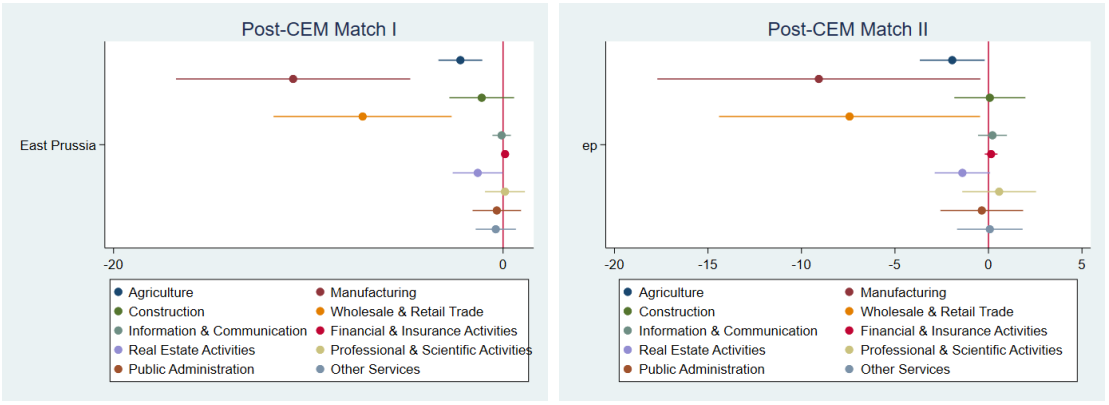


Figure G.9: CEM weights vs. strata (economic outcomes)



(a) Match I (b) Match II

Figure G.10: Post-CEM regressions (economic outcomes)



(a) Match I (b) Match II

Concluding remarks

The goal of this dissertation was to explore different aspects of the impacts of disruptive transformations, or shocks. Looking at various shocks and shock types it started with short-run dynamics and moved towards long-run impacts of shocks, while also moving from predominantly theoretical to predominantly empirical approaches. Thematically, it started with a firm perspective and a discussion of the shocks that directly impact firm behavior. Emphasizing the importance of non-economic factors, it then turned to a multidisciplinary investigation of long-lasting impacts of culture and institutions.

The first essay, therefore, focused on theoretical modeling and simulation of short-run dynamics after a shock. I chose a change in resource endowments of a country as a shock type, which was discussed within the concept of comparative advantage. The essay generalized the "two-by-two" model of comparative advantage and showed how, also in this generalized setting, the Heckscher-Ohlin comparative advantage reinforces the Ricardian comparative advantage. The essay further showed how the short-run impacts are different from those of the long-run equilibrium. Even a windfall of resources, which benefits the country in the long run, can lead to a recession and massive firm death in the first years after the shock. Moreover, economies may need decades to adjust to a new equilibrium. The time to adjust depends on the type and scale of the shock, with typically faster adjustment to positive shocks than to negative ones.

The second essay formulated a theoretical model explaining the economic impacts of e-commerce. It showed, on the one hand, that the emergence of e-commerce has similar effects on industrial structures to those of trade liberalization. Just as trade liberalization, e-commerce creates additional markets to compete for and, thus, benefits some firms, while also increasing competitive pressure. On the other hand, the model revealed more complex impacts of e-commerce. Unlike in the case of trade liberalization, these are not only large, most profitable firms, which can benefit from e-commerce. If e-commerce comes with much lower costs than the traditional channels, it can allow small and medium-sized firms to survive in the market by switching to e-commerce, or even to use e-commerce to enter markets where they would otherwise generate negative profits. In this case, the introduction of e-commerce would, unlike trade liberalization, lead to lower market concentration. This yields a non-linear, hump-shaped relationship between e-commerce costs (and adoption) and market concentration, which is supported by the European data.

The third essay focused on cultural and institutional perspectives of economic development and explored how persistent the legacy of a homogenous region is after its exposure to divergent institutional settings and different scales of a demographic shock. The essay emphasized the importance of interpersonal relationships and intergenerational transmission as the mechanisms of cultural persistence by showing how East Prussian

culture was dismantled in the regions with massive population exchange after World War II.

In terms of system stability, as discussed in the introduction, the three essays offer several interesting insights. I acknowledge that they focused on very different types of transformations and they are, therefore, not directly comparable to each other. Yet, exactly because of this diversity, it is striking how similar some of the results are in terms of equilibrium displacement and durability of the old system's legacy. It is important to note that all shocks considered here were permanent in the sense that the endowments, technology or institutions never return to their pre-shock state. As a result, it is obvious that in all three cases we observe new equilibria. At the same time, in essay 1, the new equilibrium rescales the economy and industry sizes but remains structurally the same, unless the endowment shock is so huge that it reverts the comparative advantage. In essay 2, the new e-commerce technology leads to relocation of market shares but does not push all traditional firms out of the market. The old and new technologies co-exist in the new equilibrium. Essay 3 provides evidence that in the regions, where the demographic shock did not completely interrupt the transmission mechanisms, the legacy of East Prussia persisted through decades. Thus, also here the new system did not completely replace the old one. As in essay 1, the degree of displacement was related to the scale of shock. Furthermore, both essays 1 and 3 show how the old equilibria echo through ages. In essay 1, this is observed as a decades-long transition to the new steady state. In essay 3, the lasting impact of the old – East Prussian – equilibrium manifests itself in cultural persistence in the regions with lower magnitudes of the demographic shock.

Overall, this dissertation not only contributes to international economics and economic history on several dimensions, but also opens the way for further fascinating research. Firstly, any new theoretical insights call for empirical investigations, and exploring empirically all theoretical implications presented in this dissertation would go far beyond the scope of this work. In terms of transition dynamics after endowment shocks, the logical next step is to select a real-world case, e.g. a discovery of new mineral resources, on which the model implications can be tested. The exciting challenge here might just be the same as the motivation for the first essay: since the transition to the new equilibrium may take decades, the endowment shocks may happen too frequently to be easily tested for.

In terms of e-commerce impacts, collection of firm-level data would deliver new insights into the interplay of the traditional and e-commerce channels within firms, which cannot be investigated with sector-level data. Moreover, there are numerous directions in which the model of e-commerce could be extended. For example, one could explore the interplay of e-commerce and the Heckscher-Ohlin comparative advantage or the price and market structure effects of market power and strategically dominating firms.

When it comes to cultural persistence, this dissertation concentrated on the channel of interpersonal and intergenerational transmission. It is promising to explore through what other channels culture can persist, especially in relation to disruptive shocks such as population exchange. Furthermore, as currently the data on values and attitudes are quite limited for such regional studies across borders, a challenging research project would be to enhance the coverage of existing value surveys or even conduct a detailed survey on a selected region with low-level geographic aggregation. So far, such surveys

have only been conducted in specific countries (e.g., Poland); therefore, an international geographically detailed dataset could be invaluable.

Bibliography

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). "The colonial origins of comparative development: An empirical investigation". In: *American Economic Review* 91.5, pp. 1369–1401. DOI: 10.1257/aer.91.5.1369.

Alesina, Alberto and Nicola Fuchs-Schündeln (2007). "Good bye Lenin (or not?): The effect of Communism on people's preferences". In: *American Economic Review* 97.4, pp. 1507–1528. ISSN: 0002-8282.

Alyoubi, Adel A. (2015). "E-commerce in developing countries and how to develop them during the introduction of modern systems". In: *Procedia Computer Science* 65, pp. 479–483. ISSN: 18770509. DOI: 10.1016/j.procs.2015.09.127.

Ambrassat, August (1912). *Die Provinz Ostpreußen, ein Handbuch der Heimatkunde*. Königsberg: J.H. Bon.

Anderson, James E. and Eric van Wincoop (2003). "Gravity with gravitas: A solution to the border puzzle". In: *The American Economic Review* 93.1, pp. 170–192. ISSN: 0002-8282. DOI: 10.1257/000282803321455214.

Anvari, Rana Deljavan and Davoud Norouzi (2016). "The impact of e-commerce and R&D on economic development in some selected countries". In: *Procedia - Social and Behavioral Sciences* 229, pp. 354–362.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare (Feb. 2012). "New trade models, same old gains?" In: *American Economic Review* 102.1, pp. 94–130. DOI: 10.1257/aer.102.1.94.

Avdeenko, Alexandra and Thomas Siedler (2017). "Intergenerational correlations of extreme right-wing party preferences and attitudes toward immigration". In: *Scandinavian Journal of Economics* 119.3, pp. 768–800. ISSN: 14679442. DOI: 10.1111/sjoe.12190.

Baccini, Leonardo, Pablo M. Pinto, and Stephen Weymouth (2017). "The distributional consequences of preferential trade liberalization: Firm-level evidence". In: *International Organization* 71.2, pp. 373–395. ISSN: 15315088. DOI: 10.1017/S002081831700011X.

Bartelsman, Eric, Stefano Scarpetta, and Fabiano Schivardi (2005). "Comparative analysis of firm demographics and survival: Evidence from micro-level sources in OECD countries". In: *Industrial and Corporate Change* 14.3, pp. 365–391. ISSN: 09606491. DOI: 10.1093/icc/dth057.

Bartkowski, Jerzy (2003). *Tradycja i polityka: wpływ tradycji kulturowych polskich regionów na współczesne zachowania społeczne i polityczne*. Warszawa: Zak.

Beck, Hermann (1997). *The origins of the authoritarian welfare state in Prussia: Conservatives, bureaucracy, and the social question, 1815-70*. Social History, Popular Cultur. University of Michigan Press. ISBN: 9780472084289.

Becker, Sascha O, Katrin Boeckh, Christa Hainz, and Ludger Woessmann (2016). "The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy". In: *Economic Journal* 126.590, pp. 40–74. DOI: 10.1111/ecoj.12220.

Beestermöller, Matthias and Ferdinand Rauch (2014). *A dissection of trading capital: Cultural persistence of trade in the aftermath of the fall of the iron curtain*. Munich Discussion Paper.

Berdahl, Robert M (2014). *The politics of the Prussian nobility: The development of a conservative ideology, 1770-1848*. Princeton Legacy Library. Princeton University Press. ISBN: 9781400859788.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum (2003). "Plants and productivity in international trade". In: *The American Economic Review* 93.4, pp. 1268–1290. ISSN: 00028282. DOI: 10.2307/3132288. arXiv: arXiv:1011.1669v3.

Bernard, Andrew B. and Bradford J. Jensen (1999). "Exceptional exporter performance: Cause, effect, or both?" In: *Journal of International Economics* 47.1, pp. 1–25. ISSN: 00221996. DOI: 10.1016/S0022-1996(98)00027-0.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott (2007). "Comparative advantage and heterogeneous firms". In: *The Review of Economic Studies* 74.1, pp. 31–66. ISSN: 0034-6527. DOI: 10.1111/j.1467-937X.2007.00413.x.

– (2011). "Multiproduct firms and trade liberalization". In: *Quarterly Journal of Economics* 126.3, pp. 1271–1318. ISSN: 00335533. DOI: 10.1093/qje/qjr021.

Berning, Carl C. and Elmar Schlueter (2016). "The dynamics of radical right-wing populist party preferences and perceived group threat: A comparative panel analysis of three competing hypotheses in the Netherlands and Germany". In: *Social Science Research* 55, pp. 83–93. DOI: 10.1016/j.ssresearch.2015.09.003.

Besser, Christoph (2007). *Die deutschen Vertriebenen in der Bundesrepublik Deutschland von 1945 bis 1969*. Cologne.

Bikker, Jacob and Katharina Haaf (2002). "Measures of competition and concentration in the banking industry: A review of the literature". In: *Economic & Financial Modelling* 9.2, pp. 53–98.

Bisin, Alberto and Thierry Verdier (2001). "The economics of cultural transmission and the dynamics of preferences". In: *Journal of Economic Theory* 97, pp. 298–319. DOI: 10.1006?jeth.2000.2678.

Blum, Bernardo S (2010). "Endowments, output, and the bias of directed innovation". In: *Review of Economic Studies* 77.2, pp. 534–559. DOI: 10.1111/j.1467-937X.2009.00579.x.

Böhm, Robert, Maik M.P. Theelen, Hannes Rusch, and Paul A.M. Van Lange (2018). "Costs, needs, and integration efforts shape helping behavior toward refugees". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.28, pp. 7284–7289. ISSN: 10916490. DOI: 10.1073/pnas.1805601115.

Braun, Sebastian and Nadja Dwenger (2018). *The local environment shapes refugee integration: Evidence from post-war Germany*. Tech. rep.

Bukowski, Pawel (2015). *How history matters for student performance. Lessons from the partitions of Poland*. CenEA Working Paper.

Bundeskartellamt (June 2016). *Market power of platforms and networks*. Working Paper B6-113/15. URL: https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Berichte/Think-Tank-Bericht-Langfassung.html?nn=3591568.

Burstein, Ariel Tomás and Jonathan Vogel (2011). *Factor prices and international trade: A unifying perspective*. NBER Working Paper 16904. URL: http://www.nber.org/papers/w16904.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik (2017). "rdrobust: Software for regression-discontinuity designs". In: *Stata Journal* 17.2, pp. 372–404. ISSN: 1536867X. DOI: 10.1177/1536867x1701700208.

Calonico, Sebastian, Matias D. Cattaneo, and Rocío Titiunik (2014). "Robust data-driven inference in the regression-discontinuity design". In: *Stata Journal* 14.4, pp. 909–946. ISSN: 1536867X. DOI: 10.1177/1536867x1401400413.

– (2015). "Optimal data-driven regression discontinuity plots". In: *Journal of the American Statistical Association* 110.512, pp. 1753–1769. ISSN: 1537274X. DOI: 10.1080/01621459.2015.1017578.

Chen, Natalie, Jean Imbs, and Andrew Scott (2009). "The dynamics of trade and competition". In: *Journal of International Economics* 77.1, pp. 50–62. ISSN: 00221996. DOI: 10.1016/j.jinteco.2008.10.003.

Cheung, Sin Yi and Jenny Phillimore (2014). "Refugees, social capital, and labour market integration in the UK". In: *Sociology* 48.3, pp. 518–536. ISSN: 00380385. DOI: 10.1177/0038038513491467.

Cheung, Yin-Wong, Menzie Chinn, and Eiji Fujii (2001). "Market structure and the persistence of sectoral real exchange rates". In: *International Journal of Finance and Economics* 6, pp. 95–114. DOI: 10.1002/ijfe.151.

Choshin, Mahdi and Ali Ghaffari (2017). "An investigation of the impact of effective factors on the success of e-commerce in small- and medium-sized companies". In: *Computers in Human Behavior* 66, pp. 67–74. ISSN: 07475632. DOI: 10.1016/j.chb.2016.09.026.

Clerides, Sofronis K., Saul Lach, and James R. Tybout (1998). "Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco". In: *Quarterly Journal of Economics* 113.3, pp. 903–947. DOI: 10.1162/003355398555784.

Corneo, Giacomo (2010). "Nationalism, cognitive ability, and interpersonal relations". In: *International Review of Economics* 57.2, pp. 119–141. ISSN: 18651704. DOI: 10.1007/s12232-010-0097-2.

Costantini, James A. and Marc J. Melitz (2008). "The dynamics of firm-level adjustment to trade liberalization". In: *The Organization of Firms in a Global Economy* 4, pp. 107–141.

Couture, Victor, Benjamin Faber, Yizhen Gu, and Lizhi Liu (2018). *Connecting the countryside via e-commerce: Evidence from China*. NBER Working Paper 24384.

Crouzet, Nicolas and Janice C Eberly (2019). *Understanding weak capital investment: The role of market concentration and intangibles*. NBER Working Paper 25869.

Curry, B and K.D. George (1983). "Industrial concentration: A survey". In: *Journal of Industrial Economics* 31.3, pp. 203–255.

Datta, Nirupam (2015). "Evaluating impacts of watershed development program on agricultural productivity, income, and livelihood in Bhalki watershed of Bardhaman District, West Bengal". In: *World Development* 66, pp. 443–456. ISSN: 0305750X. DOI: 10.1016/j.worlddev.2014.08.024.

Decker, Oliver, Johannes Kiess, and Elmar Brähler (2014). *Die stabilisierte Mitte - Rechtsextreme Einstellung in Deutschland 2014*. "Mitte"-Studien der Universität Leipzig. Leipzig: Universität Leipzig, Kompetenzzentrum für Rechtsextremismus- und Demokratieforschung der Universität Leipzig, p. 74. URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-389999.

Der Bundeswahlleiter (2016). *Wahl zum 1. Deutschen Bundestag (14. August 1949). Endgültige Ergebnisse der Stimmen nach Wahlkreisen*. Wiesbaden. URL: https://www.bundeswahlleiter.de/bundestagswahlen/1949.html.

Diener, Alexander and Joshua Hagen (2011). "Geopolitics of the Kaliningrad exclave and enclave: Russian and EU perspectives". In: *Eurasian Geography and Economics* 52.4, pp. 567–592. ISSN: 1538-7216. DOI: 10.2747/1539-7216.52.4.567.

Dimitrova-Grajzl, Valentina (2007). "The reat divide revisited: Ottoman and Habsburg legacies on transition". In: *Kyklos* 60.4, pp. 539–558. ISSN: 00235962.

Dolata, Ulrich (2017). *Apple, Amazon, Google, Facebook, Microsoft: Market concentration - competition - innovation strategies*. SOI Discussion Paper.

Dolfen, Paul, Liran Einav, Peter J. Klenow, Benjamin Klopack, Jonathan D. Levin, Laurence Levin, and Wayne Best (2019). *Assessing the gains from e-commerce*. NBER Working Paper 25610.

Dollar, David and Aart Kraay (2003). "Institutions, trade, and growth". In: *Journal of Monetary Economics* 50, pp. 133–162. DOI: 10.1016/S0304-3932(02)00206-4.

Duch-Brown, Néstor, Lukasz Grzybowski, André Romahn, and Frank Verboven (2017). "The impact of online sales on consumers and firms. Evidence from consumer electronics". In: *International Journal of Industrial Organization* 52, pp. 30–62. ISSN: 01677187. DOI: 10.1016/j.ijindorg.2017.01.009.

Dustmann, Christian and Albrecht Glitz (2015). "How do industries and firms respond to changes in local labor supply?" In: *Journal of Labor Economics* 33.3, pp. 711–750. ISSN: 0734-306X. DOI: 10.1086/679684.

Eaton, Jonathan and Samuel Kortum (2001). "Technology, trade, and growth: A unified framework". In: *European Economic Review* 45, pp. 742–755. DOI: 10.1016/S0014-2921(01)00129-5.

Eberhardt, Piotr (2002). *Ethnic groups and population changes in twentieth-century Central-Eastern Europe: History. Data. Analysis*. London & New York: Taylor & Francis Group. ISBN: 0-7656-0665-8.

– (2010). *Migracje polityczne na ziemiach polskich (1939-1950)*. Poznań: Instytut Zachodni.

Egger, Hartmut and Peter Egger (2003). "On market concentration and international outsourcing". In: *Applied Economics Quarterly* 49.1, pp. 49–64. ISSN: 08954356. DOI: 10.1016/s0895-4356(05)00103-4.

emarketer (Jan. 2018). *Worldwide retail and e-commerce sales: eMarketer's updated forecast and new e-commerce estimates for 2016-2021*. Report. URL: https://www.emarketer.com/Report/Worldwide-Retail-Ecommerce-Sales-eMarketers-Updated-Forecast-New-Mcommerce-Estimates-20162021/2002182.

Engerman, Stanley L and Kenneth L Sokoloff (1997). "Factor endowments, institutions, and differential paths of growth among New World economies". In: *How Latin America Fell Behind. Essays on the Economic Histories of Brazil and Mexico, 1800-1914*. Ed. by Stephen Haber. Stanford University Press. Chap. 10, pp. 260–296.

Escoe, Gisela Meyer (1995). "The demise of Soviet industry: A regional perspective". In: *Journal of Comparative Economics* 21.3, pp. 336–352. ISSN: 10957227. DOI: 10.1006/jcec.1995.0005.

Eurostat (2019a). *E-commerce sales (isoc_ec_eseln2)*. URL: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc%5C_ec%5C_eseln2%5C%7B%5C&%7Dlang=en.

– (2019b). *E-commerce statistics*. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/E-commerce%7B%5C_%7Dstatistics.
Falk, Armin, Andreas Kuhn, and Josef Zweimüller (2011). "Unemployment and right-wing extremist crime". In: *Scandinavian Journal of Economics* 113.2, pp. 260–285. ISSN: 03470520. DOI: 10.1111/j.1467-9442.2011.01648.x.
Falk, Martin and Eva Hagsten (2015). "E-commerce trends and impacts across Europe". In: *International Journal of Production Economics* 170, pp. 357–369. ISSN: 09255273. DOI: 10.1016/j.ijpe.2015.10.003.
Falter, Jürgen W. and Dirk Hänisch (1990). *Election and social data of the districts and municipalities of the German Empire from 1920 to 1933*. Cologne. DOI: http://dx.doi.org/10.4232/1.8013.
FDP (1948). *Heppenheimer Proklamation der Freien Demokratischen Partei*. URL: https://www.freiheit.org/grundsatzprogramme-und-dokumente.
– (1949). *Die Bremer Plattform der Freien Demokratischen Partei*. URL: https://www.freiheit.org/grundsatzprogramme-und-dokumente.
Feenstra, Robert C. (2016). *Advanced international trade: theory and evidence*. 2nd ed. Princeton and Oxford: Princeton University Press.
Figueiredo, John M. de (2000). "Finding sustainable profitability in electronic commerce". In: *Sloan Management Review* Summer 200, pp. 41–52.
Forstreuter, Kurt (1955). *Preussen und Russland. Von den Anfängen des Deutschen Ordens bis zu Peter dem Grossen*. Göttinger Bausteine zur Geschichtswissenschaft: Bd. 23. Göttingen, Musterschmidt [1955].
Frankel, Jeffrey A. and David Romer (1999). "Does trade cause growth?" In: *The American Economic Review* 89.3, pp. 379–399.
Freund, Caroline L. and Diana Weinhold (2004). "The effect of the Internet on international trade". In: *Journal of International Economics* 62, pp. 171–189.
Friedrich, Carl Joachim (1937). "The agricultural basis of emotional nationalism". In: *Public Opinion Quarterly* 1, pp. 50–61. ISSN: 0033362X. DOI: 10.1086/265061.
Frost and Sullivan (Dec. 2014). *Future of B2B online retailing*. Report MA4E-01-00-00-00. URL: https://store.frost.com/future-of-b2b-online-retailing.html.
Gaffeo, Edoardo, Mauro Gallegati, and Antonio Palestrini (2003). "On the size distribution of firms: Additional evidence from the G7 countries". In: *Physica A* 324.1-2, pp. 117–123. ISSN: 03784371. DOI: 10.1016/S0378-4371(02)01890-3.
Gayle, Philip G. (2001). *Market concentration and innovation: New empirical evidence on the schumpeterian hypothesis*. Working Paper. Kansas State University.
Gefen, David (2000). "E-commerce: The role of familiarity and trust". In: *Omega* 28, pp. 725–737. ISSN: 03050483. DOI: 10.1016/S0305-0483(00)00021-9.
Gericke, Dina, Anne Burmeister, Jil Löwe, Jürgen Deller, and Leena Pundt (2018). "How do refugees use their social capital for successful labor market integration? An exploratory analysis in Germany". In: *Journal of Vocational Behavior* 105, pp. 46–61. DOI: 10.1016/j.jvb.2017.12.002.
Ghironi, Fabio and Marc J. Melitz (2005). "International trade and macroeconomic dynamics with heterogeneous firms". In: *Quarterly Journal of Economics* 120.3, pp. 865–915.
Giedrojć, Marzenna (2005). *Kształtowanie tożsamości kulturowej mieszkańców Pomorza Zachodniego w drugiej połowie XX wieku*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Girard, Tulay and Paul Dion (2010). "Validating the search, experience, and credence product classification framework". In: *Journal of Business Research* 63.9-10, pp. 1079–1087. DOI: 10.1016/j.jbusres.2008.12.011.

Girard, Tulay, Pradeep Korgaonkar, and Ronnie Silverblatt (2003). "Relationship of type of product, shopping orientations, and demographics with preference for shopping on the internet". In: *Journal of Business and Psychology* 18.1, pp. 101–120.

Glück, Helmut and Wolfgang Werner Sauer (1997). "Demographie und Sprachwandel: Flüchtlinge, Besatzer, Gastarbeiter, Touristen, Aussiedler". In: *Gegenwartsdeutsch*. Stuttgart: J.B. Metzler, pp. 87–97. ISBN: 978-3-476-04110-4. DOI: 10.1007/978-3-476-04110-4_7.

Goldmanis, Maris, Ali Hortacsu, Chad Syverson, and Onsel Emre (2009). "E-commerce and the market structure of retail industries". In: *The Economic Journal* 120, pp. 651–682.

Gomez-Herrera, Estrella, Bertin Martens, and Geomina Turlea (2013). *The drivers and impediments for cross-border e-commerce in the EU*. Digital Economy Working Paper. Institute for Prospective Technological Studies.

Greer, Douglas F. (1971). "Advertising and market concentration". In: *Southern Economic Journal* 38.1, pp. 19–32.

Grosfeld, Irena and Ekaterina Zhuravskaya (2015). "Cultural vs. economic legacies of Empires: Evidence from the partition of Poland". In: *Journal of Comparative Economics* 43.1, pp. 55–75. ISSN: 10957227.

Grosjean, Pauline (2011a). "The institutional legacy of the Ottoman Empire: Islamic rule and financial development in South Eastern Europe". In: *Journal of Comparative Economics* 39.1, pp. 1–16. DOI: 10.1016/j.jce.2010.05.003.

– (2011b). "The weight of history on European cultural integration: A gravity approach". In: *The American Economic Review* 101.3, pp. 504–508.

Hälbig, Mirja C. and Jürgen R. Lorenz (2019). "How Stalin and Roosevelt influenced the federal elections in 2017 East Germany". In: *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2019: 30 Jahre Mauerfall - Session: Economic History II, No. D07-V1*. Kiel, Hamburg: ZBW - Leibniz-Informationszentrum Wirtschaft.

Harvey, Elizabeth (2000). "Pilgrimages to the 'bleeding border': Gender and rituals of nationalist protest in Germany, 1919-39". In: *Women's History Review* 9.2, pp. 201–229. ISSN: 09612025. DOI: 10.1080/09612020000200246.

Helpman, Elhanan, Oleg Itskhoki, Marc Andreas Muendler, and Stephen J. Redding (2017). "Trade and inequality: From theory to estimation". In: *Review of Economic Studies* 84.1, pp. 357–405. ISSN: 1467937X. DOI: 10.1093/restud/rdw025.

Helpman, Elhanan and Paul R. Krugman (1985). *Market structure and foreign trade: Increasing returns, imperfect competition and the international economy*. London: The MIT Press. ISBN: 9780262580878. DOI: 10.1016/0022-1996(86)90014-0.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple (2004). "Export versus FDI". In: *American Economic Review* 94.1, pp. 300–316. DOI: 10.1257/000282804322970814.

Helpman, Elhanan and Stephen J. Redding (2010). "Inequality and unemployment in a global economy". In: *Econometrica* 78.4, pp. 1239–1283. ISSN: 0012-9682. DOI: 10.3982/ecta8640.

Ho, Shu Chun, Robert J. Kauffman, and Ting Peng Liang (2007). "A growth theory perspective on B2C e-commerce growth in Europe: An exploratory study". In: *Electronic*

Commerce Research and Applications 6, pp. 237–259. ISSN: 15674223. DOI: 10.1016/j.elerap.2006.06.003.

Hryniewicz, Janusz T. (2003). "Polityczne i gospodarcze nastpstwa zaboru pruskiego". In: *Studia Regionalne i Lokalne* 4.13, pp. 61–78.

Iacus, Stefano M., Gary King, and Giuseppe Porro (2009). "cem : Software for Coarsened Exact Matching". In: *Journal of Statistical Software* 30.9, pp. 1–27. ISSN: 1548-7660. DOI: 10.18637/jss.v030.i09. URL: http://www.jstatsoft.org/v30/i09/.

Iglesias-Pradas, Santiago, Félix Pascual-Miguel, Ángel Hernández-García, and Julián Chaparro-Peláez (2013). "Barriers and drivers for non-shoppers in B2C e-commerce: A latent class exploratory analysis". In: *Computers in Human Behavior* 29.2, pp. 314–322. ISSN: 07475632. DOI: 10.1016/j.chb.2012.01.024.

Janiak, Alexandre and Etienne Wasmer (2013). *EPL and capital-labor ratios*. Working Paper. University of Chile.

Jasinski, Tomasz (1993). "Die Rolle des Deutschen Ordens bei der Städtegründung in Preußen im 13. Jahrhundert". In: *Stadt und Orden. Das Verhältnis des Deutschen Ordens zu den Städten in Livland, Preußen und im Deutschen Reich*. Ed. by Udo Arnold. Marburg: N.G. Elwert Verlag, pp. 94–111.

Kelly, Morgan (2019). *The standard errors of persistence*. Working Paper. DOI: 10.13140/RG.2.2.26903.21922. URL: https://www.researchgate.net/publication/333571096.

Kiang, Melody Y., Qiang Ye, Yuanyuan Hao, Minder Chen, and Yijun Li (2011). "A service-oriented analysis of online product classification methods". In: *Decision Support Systems* 52, pp. 28–39. ISSN: 0167-9236. DOI: 10.1016/j.dss.2011.05.001.

Konitzer, Andrew and Jelena Grujić (2009). "An electorate adrift: Refugees and elections in post-Milošević Serbia". In: *Europe - Asia Studies* 61.5, pp. 857–874. ISSN: 09668136. DOI: 10.1080/09668130902905057.

Kozłowski, Kazimierz (2003). *Osadnictwo polskie na Pomorzu Zachodnim. Mity i rzeczywistość*. Szczecin.

Krugman, Paul R. (1980). "Scale economies, product differentiation, and the pattern of trade". In: *American Economic Review* 70.5, pp. 950–959. URL: https://www.jstor.org/stable/1805774.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer (2008). "The Economic Consequences of Legal Origins". In: *Journal of Economic Literature* 46.2, pp. 285–332.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny (1997). "Legal determinants of external finance". In: *The Journal of Finance* LII.3, pp. 1131–1151. ISSN: 0022-1082.

Lechevalier, Arnaud and Jan Wielgohs, eds. (2013). *Borders and border regions in Europe: Changes, challenges and chances*. Bielefeld: transcript Verlag, pp. 1–2. ISBN: 9783837624427.

Lechthaler, Wolfgang and Mariya Mileva (2016). *Inter-industry trade and business cycle dynamics*. Working Paper 2041. Kiel Institute for the World Economy.

– (2019). "Trade liberalization and wage inequality: New insights from a dynamic trade model with heterogeneous firms and comparative advantage". In: *Review of World Economics* 155.3, pp. 407–457. DOI: 10.1007/s10290-019-00347-0.

Lee, Raphael S.-H. and Mathilde Pak (2018). *Pro-competitive effects of globalisation on prices, productivity and markups: Evidence in the Euro Area*. INSEE Working Paper. Institut National de la Statistique et des Études Économiques.

Levy, Adam (2017). *Why would Amazon want to open 3000 cashier-less stores?* URL: https://www.fool.com/investing/2018/09/24/why-would-amazon-want-to-open-3000-cashierless-sto.aspx.

Lewicka, Maria (2005). "Ways to make people active: The role of place attachment, cultural capital, and neighborhood ties". In: *Journal of Environmental Psychology* 25.4, pp. 381–395. ISSN: 02724944.

López, Ricardo A. (2005). "Trade and growth: Reconciling the macroeconomic and microeconomic evidence". In: *Journal of Economic Surveys* 19.4, pp. 623–648. ISSN: 09500804. DOI: 10.1111/j.0950-0804.2005.00264.x.

Löwis, Sabine von (2015). "Phantom borders in the political geography of East Central Europe: An introdution". In: *Erdkunde* 69.2, pp. 99–106. ISSN: 00140015.

Makowski, Adam, ed. (2008). *Polacy na ziemiach odzyskanych: Regiony przygraniczne w poszukiwaniu tożsamości: Przed i po akcesji*. Szczecin: Instytut Historii i Stosunków Mdzynarodowych.

Martinsons, Maris G. (2008). "Relationship-based e-commerce: Theory and evidence from China". In: *Information Systems Journal* 18.4, pp. 331–356. ISSN: 13501917. DOI: 10.1111/j.1365-2575.2008.00302.x.

Mayda, Anna Maria (2006). "Who is against immigration? A cross-country investigation of individual attitudes toward immigrants". In: *Review of Economics and Statistics* 88.3, pp. 510–530. ISSN: 00346535. DOI: 10.1162/rest.88.3.510.

Melitz, Marc J. (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity". In: *Econometrica* 71.6, pp. 1695–1725. DOI: 10.1111/1468-0262.00467.

Melitz, Marc J. and Stephen J. Redding (2014a). "Heterogeneous firms and trade". In: *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. Vol. 4. Elsevier. Chap. 1, pp. 1–54. DOI: 10.1016/B978-0-444-54314-1.00001-X.

– (2014b). "Missing gains from trade?" In: *American Economic Review* 104.5, pp. 317–321. ISSN: 00028282. DOI: 10.1257/aer.104.5317.

Michalak, Andrzej, Andrzej Sakson, and Ż Stasieniuk, eds. (2011). *Polskie Ziemie Zachodnie – Studia socjologiczne*. Poznań: Instytut Zachodni.

Mintzel, Alf (1986). "Die Bayernpartei". In: *Parteien-Handbuch: Die Parteien der Bundesrepublik Deutschland 1945-1980. Sonderausgabe Band 1: AUD bis CDU*. Ed. by Richard Stöss. Wiesbaden: Springer Fachmedien, pp. 395–489. DOI: 10.1007/978-3-663-14349-9_6.

Mityko, Delia Sorana Varvara (2012). "The search experience credence product classification paradigm in the eyes of the electronic consumer". In: *Management & Marketing: Challenges for the Knowledge Society* 7.3, pp. 449–464.

Mommsen, Wilhelm (1960). "Deutsche Partei". In: *Deutsches Handbuch der Politik, Band 1: Deutsche Parteiprogramme*. München: Olzog Verlag, pp. 714–715.

MPIDR and CGG (2011a). *MPIDR population history GIS collection (partly based on Bundesamt für Kartographie und Geodäsie 2011: VG 2500 Verwaltungsgebiete (Ebenen) 1:2.500.000. Stand 01.01.2009 – Frankfurt am Main)*. Rostock.

– (2011b). *MPIDR population history GIS collection (partly based on Hubatsch and Klein (eds.) 1975 ff.: Grundriß der deutschen Verwaltungsgeschichte – Marburg)*. Rostock.

Nikzentaitis, Alvydas (2002). "Ostpreußen in der litauischen Geschichtsschreibung". In: *Baltisch-deutsche Sprachen- und Kulturenkontakte in Nord-Ostpreussen: Methoden ihrer Erforschung*. Ed. by Jochen D Range. Essen: Die Blaue Eule, pp. 143–164.

North, Douglass C. (2006). "What is missing from political economy". In: *The Oxford Handbook of Political Economy*. Ed. by Donald A. Wittman and Barry R. Weingast. Oxford University Press. ISBN: 9780199548477. URL: www.oxfordhandbooks.com/10.1093/oxfordhb/9780199548477.001.0001/oxfordhb-9780199548477-e-057.

O'Rourke, Kevin H. and Richard Sinnott (2006). "The determinants of individual attitudes towards immigration". In: *European Journal of Political Economy* 22.4, pp. 838–861. ISSN: 01762680. DOI: 10.1016/j.ejpoleco.2005.10.005.

Obstfeld, Maurice and Kenneth Rogoff (2001). "The six major puzzles in international macroeconomics: Is there a common cause?" In: *NBER Macroeconomics Annual 2000*. Ed. by Ben S. Bernanke and Kenneth Rogoff. Vol. 15. MIT Press, pp. 339–412. ISBN: 0262025035. DOI: 10.1162/08893360052390428.

Ochsner, Christian and Felix Roesel (2016). *Migrating extremists*. CESifo Working Paper. Center for Economic Studies & ifo Institute.

OECD (2019). *OECD Economic Outlook, Interim Report March 2019*. DOI: https://doi.org/https://doi.org/10.1787/69e962c6-en. URL: https://www.oecd-ilibrary.org/content/publication/69e962c6-en.

Oliveira, Tiago, Matilde Alhinho, Paulo Rita, and Gurpreet Dhillon (2017). "Modelling and testing consumer trust dimensions in e-commerce". In: *Computers in Human Behavior* 71, pp. 153–164. ISSN: 07475632. DOI: 10.1016/j.chb.2017.01.050.

Parteivorstand der KPD (1949). *Einheit Frieden Wohlstand. Das Wahlprogramm der KPD*. Frankfurt/M.

Paulson, James W. (2013). *Where Wall Street meets Main Street... Corner of capital & labor*. Tech. rep. Wells Capital Management, pp. 1–8.

Peisakhin, Leonid (2015). "Cultural legacies: Persistence and transmission". In: *The Political Economy of Governance: Institutions, Political Performance and Elections*. Ed. by Norman Schofield and Gonzalo Caballero. Cham: Springer International Publishing, pp. 21–39. ISBN: 978-3-319-15551-7.

Polugodina, Maria (2019). "The Dutch disease paradox: Endowment change and heterogeneous firms". In: *Applied Economics Letters*, pp. 1–5.

PwC (2017). *Total retail: 10 retailer investments for an uncertain future*. Reportr. URL: https://www.pwc.com/gx/en/industries/assets/total-retail-2017.pdf.

Rammelt, Henry (2015). "Shadows of the past: Common effects of communism or different pre-communist legacies? An analysis of discrepancies in social mobilization throughout Romanian Regions". In: *Erdkunde* 69.2, pp. 151–160.

Rodrik, Dani (2011). "Political Trilemma". In: *Globalization Paradox: Democracy and the Future of the World Economy*. Norton & Company. Chap. 9, pp. 184–206.

Rotemberg, Julio J. and Michael Woodford (1995). "Dynamic general equilibrium models with imperfectly competitive output markets". In: *Frontiers of Business Cycle Research*. Ed. by T. E. Cooley. Princeton, NJ: Princeton University Press.

Sá, Nelson (2015). "Market concentration and persuasive advertising: a theoretical approach". In: *Journal of Economics* 114.2, pp. 127–151. ISSN: 09318658. DOI: 10.1007/s00712-013-0387-8.

Sakson, Andrzej, ed. (2006). *Ziemie Odzyskane/Ziemie Zachodnie i Północne 1945-2005 - 60 lat w granicach państwa polskiego, Seria Ziemie Zachodnie. Studia i Materiały nr 23*. Poznań: Instytut Zachodni.

Schmitt, Karl, Hans Rattinger, and Dieter Oberndörfer (1994). *District data (Censuses 1950-1987)*. Cologne. DOI: http://dx.doi.org/10.4232/1.2472.

Schmollinger, Horst (1986). "Die Deutsche Partei". In: *Parteien-Handbuch: Die Parteien der Bundesrepublik Deutschland 1945-1980. Sonderausgabe Band 2: CSU bis DSU*. Ed. by Richard Stöss. Opladen: Westdeutscher Verlag GmbH, pp. 1025–1111.

Schulze, Max-Stephan and Nikolaus Wolf (2009). "On the origins of border effects: Insights from the Habsburg Empire". In: *Journal of Economic Geography* 9, pp. 117–136. DOI: 10.1093/jeg/lbn040.

Scobel, A., ed. (1906). *Andrees allgemeiner Handatlas in 139 Haupt- und 161 Nebenkarten nebst vollstandigem alphabetischem Namenverzeichnis. Fünfte, vollig neubearbeitete und vermehrte Auflage. Jubiläumsausgabe*. Bielefeld & Leipzig: Velhagen & Klasing. URL: https://searchworks.stanford.edu/view/10448358.

Sekeris, Petros and Chrysovalantis Vasilakis (2016). *The Mediterranean refugees crisis and extreme right parties: Evidence from Greece*. MPRA Paper 72222. MPRA. DOI: 10.1227/01.NEU.0000349921.14519.2A.

Shayo, Moses (2009). "A model of social identity with an application to political economy: Nation, class, and redistribution". In: *American Political Science Review* 103.2, pp. 147–174. ISSN: 00030554. DOI: 10.1017/S0003055409090194.

Siegel, Donald S. and Donald F. Vitaliano (2007). "An empirical analysis of the strategic use of corporate social responsibility". In: *Journal of Economics & Management Strategy* 16.3, pp. 773–792. DOI: 10.1111/j.1530-9134.2007.00157.x.

Šimon, Martin (2015). "Measuring phantom borders: The case of Czech/Czechoslovakian electoral geography". In: *Erdkunde* 69.2, pp. 139–150.

Solsten, Eric, ed. (1996). *Germany: A country study*. Washington, DC: Federal Research Division, Library of Congress. ISBN: 0844408530.

Staliunas, Darius, ed. (2016). *Spatial concepts of Lithuania in the long nineteenth century*. Boston: Academic Studies Press.

Steinmayr, Andreas (2016). *Exposure to refugees and voting for the far-right: (Unexpected) results from Austria*. IZA Discussion Papers.

Tabellini, Guido (2010). "Culture and institutions: Economic development in the regions of Europe". In: *Journal of the European Economic Association* 8.4, pp. 677–716.

Takagi, H., M. Esteban, T. Shibayama, T. Mikami, R. Matsumaru, M. De Leon, N. D. Thao, T. Oyama, and R. Nakamura (2017). "Track analysis, simulation, and field survey of the 2013 Typhoon Haiyan storm surge". In: *Journal of Flood Risk Management* 10.1, pp. 42–52. ISSN: 1753318X. DOI: 10.1111/jfr3.12136.

Terzi, Nuray (2011). "The impact of e-commerce on international trade and employment". In: *Procedia - Social and Behavioral Sciences* 24, pp. 745–753.

Tilse, Mark (2011). *Transnationalism in the Prussian East. From national conflict to synthesis, 1871-1914*. Palgrave Macmillan.

Tomić, Đorđe (2016). *Phantomgrenzen und regionale Autonomie im postsozialistischen Südosteuropa. Die Vojvodina und das Banat im Vergleich*. Göttingen: Wallstein-Verlag.

Urbatsch, R. (2017). "Long-term effects of ethnic cleansing in the former Polish-German borderlands". In: *Political Geography* 58, pp. 56–66. ISSN: 09626298. DOI: 10.1016/j.polgeo.2017.01.004. URL: http://dx.doi.org/10.1016/j.polgeo.2017.01.004.

Vareikis, Vygantas (2002). “Germano-Lithuanian relations in Lithuania Minor and certain aspects of the ”Memelland Identity” in the nineteenth and twenteeth centuries”. In: *Baltisch-deutsche Sprachen- und Kulturenkontakte in Nord-Ostpreussen: Methoden ihrer Erforschung*. Ed. by Jochen D Range. Essen: Die Blaue Eule, pp. 81–98.

Visser, Robin (2019). “The effect of the Internet on the margins of trade”. In: *Information Economics and Policy* 46, pp. 41–54.

Weinberg, Bruce D., Salvatore Parise, and Patricia J. Guinan (2007). “Multichannel marketing: Mindset and program development”. In: *Business Horizons* 50.5, pp. 385–394. ISSN: 0007-6813. DOI: https://doi.org/10.1016/j.bushor.2007.04.002.

Wingfield, Nick and Micheal J. de la Merced (2018). *Amazon to buy Whole Foods for 13.4 billion*. URL: https://www.nytimes.com/2017/06/16/business/dealbook/amazon-whole-foods.html.

Wolf, Nikolaus (2005). “Path dependent border effects: The case of Poland’s reunification (1918 – 1939)”. In: *Explorations in Economic History* 42, pp. 414–438. DOI: 10.1016/j.eeh.2004.10.002.

Wolf, Nikolaus, Max-Stephan Schulze, and Hans-Christian Heinemeyer (2011). “On the economic consequences of the peace: Trade and borders after Versailles”. In: *Journal of Economic History* 71.4, p. 2011. DOI: 10.1017/S0022050711002191.

Wrzesinski, Wojciech (1985). “Disputes over East Prussia’s history between 1920 and 1939”. In: *Polish Western Affairs* 26.2, pp. 219–239.

Wylgała, Anna (2014). *Przesiedlenia a pamić: studium (nie) pamici społecznej na przykładzie ukraińskiej Galicji i polskich” Ziem Odzyskanych”*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika.

Wysokinska, Agnieszka (2015). *Invisible wall: How history matters for economic development*. Mimeo. Faculty of Management, University of Warsaw.

Zarycki, Tomasz (2015). “The electoral geography of Poland: Between stable spatial structures and their changing interpretations”. In: *Erdkunde* 69.2, pp. 107–124. ISSN: 00140015.

Zukowski, Ryszard (2004). “Historical path dependence, institutional persistence, and transition to market economy: The case of Poland”. In: *International Journal of Social Economics* 31.10, pp. 955–973. ISSN: 03068293.

Zyromski, Stanislaw (1985). “Population processes in Warmia and Mazury in 1945-1949”. In: *Polish Western Affairs* 26.2, pp. 291–302.