

Zelle, Helge

Doctoral Thesis

The political economy of interstate conflicts and industrial development

PhD Series, No. 227

Provided in Cooperation with:

University of Copenhagen, Department of Economics

Suggested Citation: Zelle, Helge (2022) : The political economy of interstate conflicts and industrial development, PhD Series, No. 227, University of Copenhagen, Department of Economics, Copenhagen

This Version is available at:

<https://hdl.handle.net/10419/266054>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS



PhD Thesis

Helge Zille

The Political Economy of Interstate Conflicts and Industrial Development

Supervisor: John Rand

Handed in: April 28, 2022

Contents

Introduction	v
Acknowledgements	xiii
References to Introduction	xv
English Abstract	xxv
Dansk Abstract	xxvii
1 The Role of Labor Composition and Quality in Determining the Productivity-Wage Gap: Evidence from Industrial Zones in Myanmar	1
2 Brothers in Arms, Brothers in Trade? Measuring the Effect of Violent Conflicts on Trade with Third-Party Countries	55
3 The Ownership of Pioneer Firms: The Role of State-Owned and Multinational Enterprises in Industrial Diversification in Vietnam	111

Introduction

When I first enrolled into studying economics in 2011, barely knowing what I have signed up for, who would have thought I would end up here, writing the introduction to my very own PhD dissertation. Statistics clearly did not predict this path for me. Only one in hundred German children from working class parents achieve a PhD. And that spot was already to be filled up by my older brother. And yet, here I am. This thesis, therefore, starts with expressing how grateful, but also how proud I am, of having been able to achieve this.

This dissertation consists of three self-contained chapters in the fields of international and development economics, and deals with the political economy of violent interstate conflicts and of industrial development in two South-East Asian countries. The first chapter studies the role of industrial zones in Myanmar. The second chapter studies the effect of violent interstate conflicts on trade with third-party countries on a global scale. And the third chapter explores the role of state-owned and multinational enterprises in economic diversification in Vietnam. The three chapters cover a range of topics and areas, and each represents particular interests and stages in my academic career.

The idea for the first chapter arose from my involvement with UNU-WIDER prior to starting my PhD, where I had already studied particular aspects of Myanmar's economic geography. Based on that and together with my co-authors Neda Trifković, John Rand, and Finn Tarp, we decided to further study the role of industrial zones and natural industrial agglomeration in Myanmar. This question falls into the wider discussion in the literature on agglomeration economies and place-based policies. There are numerous potential and widely accepted benefits from industrial agglomeration, such as decreasing production costs, improving access to labor, creation of market linkages, and better transfers of knowledge and technology. These potential benefits unfold due to agglomeration economies, sometimes referred to as Marshallian externalities (Fujita and Thisse, 1996): labor pooling, knowledge spillovers, and input-output linkages (Marshall, 1890; Puga, 2010). A large body of literature has tried to measure the extent to which agglomeration economies unfold, and what the effects of it are. The main focus of this literature is, though, geographically on the

US (and other high-income countries), and conceptually on cities (e.g. Behrens and Robert-Nicoud, 2015; Combes and Gobillon, 2015; Rosenthal and Strange, 2001). Agglomeration is thus often measured as city size or city density, and the outcome of interest is often productivity, measured as wages. Agglomeration is usually found to be associated with higher worker and firm productivity, even though the context- and country-heterogeneity is vast (Combes and Gobillon, 2015; Melo et al., 2009). The debate on agglomeration and, in fact, urbanization, is also directly interlinked with the literature on place-based policies (Kline and Moretti, 2014a). In this literature, the research question is usually not about naturally occurring concentration of economic activity in cities, but rather about policies that encourage and promote economic activity in specific, designated locations. While there is again a strong body of literature studying place-based policy programs in the US, such as enterprise zones or empowerment zones (Busso et al., 2013; Kline and Moretti, 2014a; 2014b), there has been and there is a growing interest in place-based policies in developing countries too. The evidence is, however, still limited and has a strong focus on China, where Special Economic Zones (SEZs) are an integral part of the national development strategy (e.g. Fan and Scott, 2015; Lu and Tao, 2009; Wang, 2013; Zeng, 2015). Numerous studies have shown that Chinese SEZs are quite successful in unfolding agglomeration economies and their benefits, such as attracting foreign investment and generating higher wages. While evidence on other developing countries is limited, there is, nevertheless, a growing research interest in economic zones and agglomeration in general in developing countries (Cohen, 2006; Chhair and Newman, 2014; Deichmann et al., 2008; Farole, 2011; Frick et al., 2018; Newman and Page, 2017; World Bank, 2017). Studies have been conducted on other South and South-East Asian countries, such as India, Indonesia, Vietnam, or Cambodia, as well as on Africa, where also China's strong involvement in SEZs is a subject of interest (Farole, 2011; Zeng, 2015). The findings in these countries are less positive than in China or in developed countries. Especially in Africa, but also in some Asian cases, SEZs and other types of economic zones perform way below their potential, mainly due to a lack of appropriate infrastructure, bad institutions and regulatory frameworks, and an unfavorable investment climate. Nevertheless, different forms of economic zones remain a widely used and promoted policy tool for the stimulation of industrial agglomeration in many developing countries.

In Myanmar, the first zones were established in the 1990s, and remain an integral part of Myanmar's policy basket to promote manufacturing growth, industrial upgrading, and to attract foreign investment. Yet, the extent to which industrial zones in fact contribute to agglomeration and unfold these potential benefits in Myanmar, still remains an open topic in the literature. The first chapter of this thesis aims to contribute to this debate. We specifically study the productivity and wage gains as-

sociated with being located inside industrial zones, taking into account heterogeneity in the labor force composition. We find that firms with an average labor force composition inside an industrial zone have higher value added than firms outside, but their employees are not compensated proportional to these productivity gains. Higher shares of skilled workers do not appear to increase neither productivity nor wages for firms and employees inside the zones. We further find that firms with above-average shares of women in their workforce in industrial zones pay higher wages without additional productivity gains for the firm, suggesting potentially a shortage of female labor in industrial zones. The way in which firms and employees can benefit from operating within industrial zones is, thus, characterized by a large degree of heterogeneity. I believe that these findings can be of high relevance for policy makers in Myanmar, and I hope they can contribute to better, evidence-based industrial policy. I further hope they can contribute to the above-mentioned literature on agglomeration and place-based policies, and make an analytically contribution by disentangling the effects on wages and firm productivity.

Working with Myanmar has been an enriching experience. Being a country that has long been closed to parts of the world, it was a privilege to contribute to some of the first empirical evidence on its economy after opening up. Being involved in the UNU-WIDER project “Towards inclusive development in Myanmar”, I had the chance to meet our partners from Myanmar in several workshops, and to lead a STATA workshop for our collaborators myself. Given these experiences, the news of the military coup in 2021 were extremely shocking to me, and I truly hope the country will be able to return to its path of freedom and democracy in the future.

The second chapter of my dissertation stems from my interest in conflicts that I developed during my undergraduate studies, where I was working for the Heidelberg Institute for International Conflict Research and was attending lectures in international relations. My interest laid in particular in the international political economy (IPE) and the interplay of violent conflicts and international trade. The discussion about the relationship between conflict and trade goes as far back as to thinkers like Smith, Kant, Lenin, or Montesquieu (Reuveny, 2000), and it has become an important field of study in political sciences. In particular, the question whether bilateral trade has a pacifying effect has been the subject of many studies, and influential IPE theories argued against (Marxism, Mercantilism) or in favor of it (Liberalism) (Gilpin, 1987). While the evidence is not unanimous, the empirical literature is in general supportive of the claim that trade can promote cooperation and reduce the likelihood of violent conflicts (e.g. Barbieri, 1996, 2002; Mansfield and Pevehouse, 2000; Martin et al., 2008; Oneal et al., 1996, 2003; Oneal and Russett, 1997; Polacheck, 1980). Other scholars have studied the impact that conflicts have on trade. This discussion falls in a broader trend of estimating the economic costs

and consequences of conflicts. The evidence is rather clear in that violent conflicts decrease trade between the conflict parties (e.g. Gowa, 1994; Keshk et al., 2004; Morrow et al., 1998; Pollins, 1989), and that interstate conflicts and civil wars cause high economic and social costs (e.g. Abadie and Gardeazabal, 2003; Gates et al., 2012; Hoeffler and Reynal-Querol, 2003; Nordhaus, 2002; Stewart et al., 1997).

A less studied and almost neglected question is the effect of conflicts on third-party countries. Some notable exceptions are Murdoch and Sandler (2002a, 2002b, 2004) and De Groot (2010) on civil wars, and, very recently, Feldman and Sadeh (2018) and Korovkin and Makarin (2021) on interstate conflicts and the effect of trade with third-party countries. In the second chapter of my dissertation, I contribute to this scarce literature and explicitly study whether and how violent interstate conflicts affect trade with other countries, and how other political and economic factors influence this relationship. I show that conflicts do have an effect on trade with third-party countries, and that this effect depends on the relationship between the three countries. I find evidence that countries trade more with their allies in times of conflict, and less so with the allies of their enemies. The strongest increase in trade can be observed between countries who have a conflict with same country at the same time. My results indicate a political dimension in the way countries trade with each other, and that trade, to some extent, can be understood as a means of conflict.

This chapter contributes to a literature that has become out of fashion. While wars and conflicts between states have long been the primary focus of conflict research in both political science and economics, the focus has substantially shifted towards interstate conflicts in the last decades. Peace and conflict studies nowadays are centered around civil wars and other intrastate conflicts. This development indeed addresses the fact that the global share of conflicts has shifted towards intrastate conflicts. However, recent events such as the Russian invasion of Ukraine, to name only the most prominent example, show that the time of interstate wars is far from being over, and research needs to take account of that. In that sense, I hope that this chapter can make its contribution to a better understanding of the consequences of international conflicts and on the entanglement of economics and politics.

I further hope to make a contribution to the methodological and conceptual discussion in the field of international political economy. In political science, but also in international economics, the subject of observation has traditionally been individual (countries or firms, in this case) or dyadic (pairs of countries or firms). In recent years, however, more and more scholars stress the insufficiencies of neglecting interdependencies with other observations. International economics have acknowledged and implemented the network structures of products, exporters, or

nations (e.g. Chaney, 2014, 2016; Combes et al., 2005; Hausmann and Hidalgo, 2009; Rauch, 1999, 2001). In the field of international political economy, there is, moreover, now a growing but yet still small literature arguing for alternatives to simple dyadic regressions (e.g. Dorff and Ward, 2013; Hafner-Burton et al., 2009; Ward et al., 2011). Poast (2010, 2016), Fordham and Poast (2016), and Lupu and Poast (2016), for instance, propose the use of k -adic data and apply it to alliance formation. Others suggest network analysis to study alliances (Cranmer et al., 2012a, 2012b; Li et al., 2017; Warren, 2010), the formation of trade agreements (Manger et al., 20012), and the effect of trade on conflict (Cranmer et al., 2015; Kinne, 2014; Lupu and Traag, 2013). My dissertation makes its contribution to this body of literature by proposing a way of how to include a triadic dimension into dyadic data, and to account for the more complex relationships between states.

An integral part of the PhD is the research stay abroad. In January 2020, I headed to New York City to spend a semester at Columbia University, where I was kindly hosted by Eric Verhoogen. I went to Columbia to experience doing research at such a prestigious institution, to take courses from distinguished researchers, and to expand my research network by meeting and collaborating with talented researches from all around the world. I also intended to get a lot of input and inspiration for my future research. To live in New York City and to be at Columbia was indeed an impressive and enriching experience. In the two PhD courses in development economics I attended, taught by Eric Verhoogen and Jack Willis, I learnt a lot, and I am very grateful for that experience.

While I was in New York, however, also started the most difficult period in my PhD. In February 2020, Covid-19 arrived in New York and life was quickly shut down. Instead of interacting with fellow researchers on a daily basis on campus, I spent my days isolated in my small room, cut off from the vivid academic environment I came for. Soon enough, New York City became *the* Covid-19 hotspot. I, therefore, decided to return to Copenhagen around two and a half months earlier than initially planned, only to continue my academic routine not much else than in New York.

This time has not been easy for me. Stripped off from my usual academic and social environment, I had hard times finding any clear thoughts and lacked motivation, inspiration, and concentration. During this time, I also experienced depressive periods for the first time, which would unfortunately accompany me throughout the rest of my PhD. In this constitution, it was, unsurprisingly, extremely difficult to find inspiration for my third PhD chapter, and I lost several valuable months.

It was then in late summer of 2020, when I joined forces with John Rand, Henrik Hansen, and my fellow PhD colleague Bjørn Bo Sørensen, to start researching state-owned enterprises (SOEs) in Vietnam. After I had already collaborated with senior

researches on the Myanmar chapter, and worked on my own on the conflict chapter, I now got the chance to collaborate with a fellow PhD student, which was a valuable and pleasant experience.

Frankly, the exact research question changed direction several times from our first meetings in 2020 until the final chapter. What was clear from the beginning, though, was that we wanted to better understand the role of the different sectors, in particular the state and the foreign sector, in Vietnams economy and its economic development process. State-owned enterprises are often perceived as inefficient monstrosities that should not exist in a functioning economy, while successful economic development and structural transformation are associated with reforms strengthening the private sector and attracting multinationals. In fact, the discussion about what drives development is as old as the economic discipline itself, and especially the economic development in Asia has been studied extensively. The rise of the East Asian Tigers has spurred the question about the state versus the market (Boyer and Drache, 1996; Malesky and London, 2014; Martinussen, 2004; Wade, 1990; White, 1988; World Bank, 1993). More precisely, is it a strong developmental state that drives economic development through strict industrial policies and a high prevalence of SOEs? Or are market-friendly private-sector reforms and the attraction of foreign investments the key to transformation? These questions encompass both the literatures on SOEs as well as on foreign direct investments (FDI) and multinational companies (MNEs).

It is by now a well-established fact that SOEs operate with lower productivity (Megginson and Netter, 2001; Song et al. 2011), erect entry barriers for private firms, and can be sources of resource misallocation (Baccini et al., 2019; Brandt et al., 2020; Hsieh and Song, 2015; Kowalski et al., 2013). Privatization of SOEs is widely found to be associated with both firm- and macro-level productivity growth (Brandt et al., 2008; Brown et al., 2006; Hsieh and Song, 2015; La Porta and Lopez-de-Silanes, 1999; McCaig and Pavcnik, 2017). And yet, SOEs continue to exist and to play a major role in many developing economies, held alive through preferential access to finance and the so-called soft budget constraint (Kornai, 1986; Lin et al., 1998; Lin and Tan, 1999; Song et al., 2011). A growing body of literature acknowledges that SOEs are tasked with a number of policy objectives (Bai et al., 2000, 2006; Christiansen, 2013; Li and Xia, 2008; Lin et al., 1998; Lin and Tan, 1999), which deviate from profit maximization or rent-seeking (Shleifer and Vishny, 1994).

With respect to FDI and MNEs, the debate in the literature centers around the effect of FDI on the host economy. In particular, whether the presence of MNEs unfolds horizontal (intra-industry) and/or vertical (inter-industry) spillovers and linkages, and whether net gains on the host economy are dampened by entry barriers and crowding-out effects. The literature is yet still inconclusive. Some studies find little

to no evidence for the existence of positive spillovers from FDI (e.g. Aitken and Harrison, 1999; Görg and Greenway, 2004; Hanson, 2001), and that foreign firms are associated with crowding-out and entry barriers against domestic firms (De Backer and Sleuwaegen, 2003). Others find FDI to be associated with productivity gains from vertical spillovers and direct linkages along the supply-chain (Ayyagari and Kosová, 2010; Javorcik, 2004; Kugler, 2006; Newman et al., 2015), with quality upgrading and the introduction of more complex or sophisticated products (Bajgar and Javorcik, 2020; Eck and Huber, 2016; Harding and Javorcik, 2012; Javorcik et al., 2018), or to cause economic growth (Hansen and Rand, 2006).

Vietnam represents an ideal case for the above-mentioned discussions. Its development in the past decades has been a huge success story, being one of the most growing countries in the world, despite having a large and influential state-sector. At the same time, Vietnam also passed a wide range of substantial economic reforms that strengthened the private domestic sector and attracted a vast amount foreign investment.

We focus in this chapter on industrial pioneers as a basic unit of economic diversification and development (Jara-Figueroa et al., 2018). Numerous studies have established that industrial diversification and innovation are crucial for economic growth and development (Akçomak and ter Weel, 2009; Brown et al., 2009; Dodgson and Rothwell, 1995; Grossman and Helpman, 1989, 1993; Hausmann and Hidalgo, 2009; Jovanovic and Nyarko, 1996). Especially the role of unrelated diversification is stressed, thus diversification into industries that require knowledge and capabilities different from the ones a country currently possesses. Reaching unrelated diversification is difficult, though. Countries mostly diversify into closely related industries and products, following a path-dependence (Hidalgo et al., 2007; Neffke et al., 2011, 2018). Due to high discovery costs and the presence of externalities, countries might moreover lack the optimal amount of private pioneering (Hausmann et al., 2007; Hausmann and Klinger, 2006, 2007; Hausmann and Rodrik, 2003; Hoff, 1997; Wei et al., 2021). Foreign firms can help to overcome these “idea gaps” (Romer, 1992, 1993) through learning and knowledge spillovers (as discussed above), but the state might also directly foster innovation and diversification through SOEs. This discussion is addressed in this chapter.

We find that the vast majority of new industries on the provincial level is established by private domestic firms (PDEs), whereas SOEs and MNEs have a higher propensity to do so. We further show that there are qualitative and spatial differences, in the sense that MNEs tend to pioneer in more complex industries and around the major economic centers where they are predominantly located, whereas SOEs tend to pioneer in less complex industries in Vietnam’s less developed provinces.

When we look at the dynamics in these pioneered industries, we find hetero-

geneity between the different ownership types too. MNE-pioneered industries employ substantially more workers than PDE-pioneered industries, even though most of this employment is situated in the pioneer itself or other emulating MNEs. Moreover, MNEs are responsible for almost 40 percent of all employment in infant industries established between 2004 and 2017. For SOE-pioneered industries, on the contrary, we find no such employment premium, and emulating MNEs and PDEs in fact employ fewer workers in these industries.

These findings provide interesting insights into the above discussions about regional diversification and the role of the state and foreign firms in development. Our results suggest that regional pioneering by SOEs, especially in less developed areas, is, at least to some extent, attributable to a policy mandate. Our results are, moreover, evident of the important role that foreign investment plays in regional diversification and employment generation. These findings are not only relevant for Vietnam, but potentially also for understanding the role of the state vis-à-vis the foreign and private sectors in other developing countries.

With this final chapter, both my dissertation and my time as a PhD student have come to an end. I have certainly learnt a lot from it, both on an academic and on a personal level, as well as about the countries that I have studied. I hope that the present dissertation will give the reader some new insights too, and that it will eventually spark some fruitful thoughts for both the academic as well as for the political discussion.

Acknowledgements

Even though the work as a researcher can often be lonesome, feeling as if it was only you and your data against the world, a PhD does not happen in a vacuum. There are numerous people who contributed to my PhD and supported me in different ways, and to whom I owe great thanks.

First and foremost, I want to thank my supervisor John Rand for his excellent guidance and support throughout my PhD. His office door was at all times open for me, regardless of the question or problem, and his advice was always valuable. It has been great to work with him, both on a professional and on a personal level. Similarly, I want to thank Henrik Hansen, who provided me a great amount of supervision during the second half of my PhD, and who has been a great mentor too.

I would like to thank Neda Trifković and Finn Tarp for the great collaboration on the first chapter of my dissertation. I also want to thank Agnete Vienberg Hansen for helping me with the Danish abstract of this thesis. A deep and special thanks to my final co-author, Bjørn Bo Sørensen, for spending countless long evenings with me in the office and for supporting me especially through the final period of my PhD. Your support really meant a lot to me, and I am grateful for our collaboration.

I want to thank the Department of Economics and the Development Economics Research Group (DERG) at the University of Copenhagen for giving me the opportunity to work on this thesis. I want to further thank Finn Tarp and John Rand for putting their trust in me in the first place when I became engaged with DERG and UNU-WIDER and started my academic career.

One of the biggest pleasures of doing my PhD has been my many great colleagues at the Department of Economics and in particular at DERG. I have enjoyed being part of DERG since joining it in 2016, and have at all times felt welcomed and respected. Even when work was sometimes hard, being around my colleagues has always been a great pleasure, and I am grateful to call many of them my good friends now.

Finally, I want to express my gratitude for the support I got from outside of the university, from my great friends and from my wonderful family. My mum, my dad, Khanh, and especially my brother Heiner have always supported me with their full hearts and never stopped believing in me.

Last, but most importantly, I want to thank my wonderful girlfriend Lisa. Who has always been there for me, who always believed in me, and who helped me through all the most difficult times of my PhD. Without you, I could not have done it. Thank you.

Helge Zille
Copenhagen, April 2022

References to Introduction

Bibliography

- Abadie, A. & Gardeazabal J. (2003).** The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1), 113–132.
- Aitken, B. J. & Harrison, A. E. (1999).** Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. *American Economic Review*, 89(3), 605–618.
- Akçomak, İ. S. & ter Weel, B. (2009).** Social Capital, Innovation and Growth: Evidence from Europe. *European Economic Review*, 53(5), 544–567.
- Ayyagari, M. & Kosová, R. (2010).** Does FDI Facilitate Domestic Entry? Evidence from the Czech Republic. *Review of International Economics*, 18(1), 14–29.
- Baccini, L., Impullitti, G. & Malesky, E. J. (2019).** Globalization and State Capitalism: Assessing Vietnam’s Accession to the WTO. *Journal of International Economics*, 119, 75–92.
- Bai, C.-E., Li, D. D., Tao, Z. & Wang, Y. (2000).** A Multitask Theory of State Enterprise Reform. *Journal of Comparative Economics*, 28(4), 716–738.
- Bai, C.-E., Lu, J. & Tao, Z. (2006).** The Multitask Theory of State Enterprise Reform: Empirical Evidence from China. *American Economic Review*, 96(2), 353–357.
- Bajgar, M. & Javorcik, B. (2020).** Climbing the Rungs of the Quality Ladder: FDI and Domestic Exporters in Romania. *Economic Journal*, 130(628), 937-955.
- Barbieri, K. (1996).** Economic Interdependence: A Path to Peace or a Source of Interstate Conflict? *Journal of Peace Research*, 33(1), 29-49.
- Barbieri, K. (2002).** *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor: University of Michigan Press.

-
- Behrens, K. & Robert-Nicoud, F. (2015).** Agglomeration Theory with Heterogeneous Agents. In G. Duranton, J. V. Henderson & W. C. Strange (eds.) *Handbook of Regional and Urban Economics* (pp. 171–245). Elsevier.
- Boyer, R. & Drache, D. (eds.) (1996).** *States Against Markets: The Limits of Globalization*. Psychology Press.
- Brandt, L., Hsieh, C.-t. & Zhu, X. (2008).** Growth and Structural Transformation in China. In Brandt, L. & Rawski, T. G. (eds.) *China's Great Economic Transformation*, (pp. 683-728). Cambridge University Press.
- Brandt, L., Kambourov, G. & Storesletten, K. (2020).** Barriers to Entry and Regional Economic Growth in China. CEPR Discussion Paper No. DP14965.
- Brown, J. D., Earle, J. S. & Telegdy, Á. (2006).** The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia, and Ukraine. *Journal of Political Economy*, 114(1), 61-99.
- Brown, J. R., Fazzari, S. M. & Petersen, B. C. (2009).** Financing Innovation and Growth: Cash Flow, External Equity, and the 1990s R&D Boom. *Journal of Finance*, 64(1), 151–185.
- Busso, M., Gregory, J. & Kline, P. (2013).** Assessing the Incidence and Efficiency of a Prominent Place Based Policy. *American Economic Review*, 103(2), 897–947.
- Chaney, T. (2014).** The Network Structure of International Trade. *American Economic Review*, 104(11), 3600–3634.
- Chaney, T. (2016).** Networks in International Trade. In Y. Bramoullé, A. Galeotti & B. W. Rogers (eds.) *The Oxford Handbook of the Economics of Networks* (pp. 753–775). Oxford University Press.
- Chhair, S., & Newman, C. (2014).** Clustering, Competition, and Spillover Effects: Evidence from Cambodia. WIDER Working Paper Series No. 065. World Institute for Development Economic Research (UNU-WIDER).
- Christiansen, H. (2013).** Balancing Commercial and Non-Commercial Priorities of State-Owned Enterprises. OECD Corporate Governance Working Papers 6.
- Cohen, B. (2006).** Urbanization in Developing Countries: Current Trends, Future Projections, and Key Challenges for Sustainability. *Technology in Society*, 28(1–2), 63–80.
- Combes, P.-P. & Gobillon, L. (2015).** The Empirics of Agglomeration Economies. In G. Duranton, J. V. Henderson & W. C. Strange (eds.) *Handbook of Regional and Urban Economics* (pp. 247–348). Elsevier.

-
- Combes, P.-P., Lafourcade, M. & Mayer, T. (2005).** The Trade-Creating Effects of Business and Social Networks: Evidence from France. *Journal of International Economics*, 66(1), 1–29.
- Cranmer, S. J., Desmarais, B. A. & Kirkland, J. H. (2012a).** Toward a Network Theory of Alliance Formation. *International Interactions*, 38(3), 295–324.
- Cranmer, S. J., Desmarais, B. A. & Menninga, E. J. (2012b).** Complex Dependencies in the Alliance Network. *Conflict Management and Peace Science*, 29(3), 279–313.
- Cranmer, S. J., Menninga, E. J., & Mucha, P. J. (2015).** Kantian Fractionalization Predicts the Conflict Propensity of the International System. *Proceedings of the National Academy of Sciences*, 112(38), 11812–11816.
- De Backer, K. & Sleuwaegen, L. (2003).** Does Foreign Direct Investment Crowd Out Domestic Entrepreneurship? *Review of Industrial Organization*, 22(1), 67–84.
- Deichmann, U., Lall, S. V., Redding, S. J., & Venables, A. J. (2008).** Industrial Location in Developing Countries. *World Bank Research Observer*, 23(2), 219–246.
- Dodgson, M. & Rothwell, R. (eds.) (1995).** *The Handbook of Industrial Innovation*. Edward Elgar Publishing.
- Dorff, C. & Ward, M. D. (2013).** Networks, Dyads, and the Social Relations Model. *Political Science Research and Methods*, 1(2), 159–178.
- Eck, K. & Huber, S. (2016).** Product Sophistication and Spillovers from Foreign Direct Investment. *Canadian Journal of Economics/Revue Canadienne d'Économique*, 49(4), 1658–1684.
- Fan, C. C., & Scott, A. J. (2009).** Industrial Agglomeration and Development: A Survey of Spatial Economic Issues in East Asia and a Statistical Analysis of Chinese Regions. *Economic Geography*, 79(3), 295–319.
- Farole, T. (2011).** Special Economic Zones in Africa. Comparing Performance and Learning from Global Experience. World Bank.
- Feldman, N. & Sadeh, T. (2018).** War and Third-Party Trade. *Journal of Conflict Resolution*, 62(1), 119–142.
- Fordham, B. & Poast, P. (2016).** All Alliances are Multilateral: Rethinking Alliance Formation. *Journal of Conflict Resolution*, 60(5), 840–865.

-
- Frick, S. A., Rodríguez-Pose, A. & Wong, M. D. (2018).** Toward Economically Dynamic Special Economic Zones in Emerging Countries. *Economic Geography*, 95(1), 30–64.
- Fujita, M. & Thisse, J.-F. (1996).** Economics of Agglomeration. *Journal of the Japanese and International Economies*, 10(4), 339–378.
- Gates, S., Hegre, H., Nygård, H. M. & Strand, H. (2012).** Development Consequences of Armed Conflict. *World Development*, 40(9), 1713–1722.
- Gilpin, R. (1987)** *The Political Economy of International Relations*. Princeton: Princeton University Press.
- Görg, H. & Greenaway, D. (2004).** Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment? *World Bank Research Observer*, 19(2), 171–197.
- Gowa, J. (1994).** *Allies, Adversaries, and International Trade*. Princeton: Princeton University Press.
- De Groot, O. J. (2010).** The Spillover Effects of Conflict on Economic Growth in Neighbouring Countries in Africa. *Defence and Peace Economics*, 21(2), 149–164.
- Grossman, G. M. & Helpman, E. (1989).** Product Development and International Trade. *Journal of Political Economy*, 97(6), 1261–1283.
- Grossman, G. M. & Helpman, E. (1993).** *Innovation and Growth in the Global Economy*. MIT Press.
- Hafner-Burton, E. M., Kahler, M. & Montgomery, A. H. (2009).** Network Analysis for International Relations. *International Organization*, 63(3), 559–592.
- Hansen, H. & Rand, J. (2006).** On the Causal Links Between FDI and Growth in Developing Countries. *World Economy*, 29(1), 21–41.
- Hanson, G. H. (2001).** Should Countries Promote Foreign Direct Investment? G-24 Discussion Paper Series. UNCTAD.
- Harding, T. & Javorcik, B. S. (2012).** Foreign Direct Investment and Export Upgrading. *Review of Economics and Statistics*, 94(4), 964–980.
- Hausmann, R. & Hidalgo, C. A. (2009).** The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570–10575.
- Hausmann, R., Hwang, J. & Rodrik, D. (2007).** What you Export Matters. *Journal of Economic Growth*, 12(1), 1–25.

-
- Hausmann, R. & Klinger, B. (2006).** Structural Transformation and Patterns of Comparative Advantage in the Product Space. KSG Working Paper No. RWP06-041, CID Working Paper No. 128.
- Hausmann, R. & Klinger, B. (2007).** The Structure of the Product Space and the Evolution of Comparative Advantage. CID Working Paper No. 146.
- Hausmann, R. & Rodrik, D. (2003).** Economic Development as Self-Discovery. *Journal of Development Economics*, 72, 603–633.
- Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R. (2007).** The Product Space Conditions the Development of Nations. *Science*, 317(5837), 482–487.
- Hoeffler, A. & Reynal-Querol, M. (2003).** Measuring the Costs of Conflict. World Bank.
- Hoff, K. (1997).** Bayesian Learning in an Infant Industry Model. *Journal of International Economics*, 43(3-4), 409–436.
- Hsieh, C.-T. & Song, Z. (2015).** Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China. NBER Working Paper Series No. 21006.
- Jara-Figueroa, C., Jun, B., Glaeser, E. L. & Hidalgo, C. A. (2018).** The Role of Industry-Specific, Occupation-Specific, and Location-Specific Knowledge in the Growth and Survival of New Firms. *Proceedings of the National Academy of Sciences*, 115(5), 12646–12653.
- Javorcik, B. S. (2004).** Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages. *American Economic Review*, 94(3), 605-627.
- Javorcik, B. S., Lo Turco, A. & Maggioni, D. (2018).** New and Improved: Does FDI Boost Production Complexity in Host Countries? *Economic Journal*, 128(614), 2507–2537.
- Jovanovic, B. & Nyarko, Y. (1996).** Learning by Doing and the Choice of Technology. *Econometrica*, 64(6), 1299-1310.
- Keshk, O. M. G., Pollins, B. M., & Reuveny, R. (2010).** Trade and Conflict: Proximity, Country Size, and Measures. *Conflict Management and Peace Science*, 27(1), 3-27.
- Kinne, B. J. (2014).** Does Third-Party Trade Reduce Conflict? Credible Signaling Versus Opportunity Costs. *Conflict Management and Peace Science*, 31(1), 28-48.

-
- Kline, P. & Moretti, E. (2014a).** Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*, 129(1), 275–331.
- Kline, P. & Moretti, E. (2014b)** People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs. *Annual Review of Economics*, 6(1), 629–662.
- Kornai, J. (1986).** The Soft Budget Constraint. *Kyklos*, 39(1), 3-30.
- Korovkin, V. & Makarin, A. (2021).** Conflict and Inter-Group Trade: Evidence from the 2014 Russia-Ukraine Crisis. SSRN Working Paper.
- Kowalski, P., Büge, M., Sztajerowska, M. & Egeland, M. (2013)** State-Owned Enterprises: Trade Effects and Policy Implications. OECD Trade Policy Papers 147.
- Kugler, M. (2006).** Spillovers from Foreign Direct Investment: Within or Between Industries? *Journal of Development Economics*, 80(2), 444–477.
- La Porta, R. & Lopez-de-Silanes, F. (1999).** The Benefits of Privatization: Evidence from Mexico. *Quarterly Journal of Economics*, 114(4), 1193–1242.
- Li, S. & Xia, J. (2008).** The Roles and Performance of State Firms and Non-State Firms in China’s Economic Transition. *World Development*, 36(1), 39–54.
- Li, W., Bradshaw, A. E., Clary, C. B., & Cranmer, S. J. (2017).** A Three-Degree Horizon of Peace in the Military Alliance Network. *Science Advances*, 3(3), e1601895.
- Lin, J. Y., Cai, F. & Li, Z. (1998).** Competition, Policy Burdens, and State-Owned Enterprise Reform. *American Economic Review*, 88(2), 422–427.
- Lin, J. Y. & Tan, G. (1999).** Policy Burdens, Accountability, and the Soft Budget Constraint. *American Economic Review*, 89(2), 426–431.
- Lu, J. & Tao, Z. (2009).** Trends and Determinants of China’s Industrial Agglomeration. *Journal of Urban Economics*, 65(2), 167–180.
- Lupu, Y. & Poast, P. (2016).** Team of Former Rivals: A Multilateral Theory of Non-Aggression Pacts. *Journal of Peace Research*, 53(3), 344–358.
- Lupu, Y. & Traag, V. A. (2013).** Trading Communities, the Networked Structure of International Relations, and the Kantian Peace. *Journal of Conflict Resolution*, 57(6), 1011-1042.

-
- Malesky, E. & London, J. (2014).** The Political Economy of Development in China and Vietnam. *Annual Review of Political Science*, 17(1), 395–419.
- Manger, M. S., Pickup, M. A. & Snijders, T. A. B. (2012).** A Hierarchy of Preferences: A Longitudinal Network Analysis Approach to PTA Formation. *Journal of Conflict Resolution*, 56(5), 853–878.
- Mansfield, E. D. & Pevehouse, J. C. (2000).** Trade Blocs, Trade Flows, and International Conflict. *International Organization*, 54(4), 775-808.
- Marshall, A. (1890).** *Principles of Economics*. MacMillan.
- Martin, P., Mayer, T., & Thoenig, M. (2008).** Make Trade Not War? *Review of Economic Studies*, 75(3), 865-900.
- Martinussen, J. (2004).** *Society, State and Market: A Guide to Competing Theories of Development*. HSRC Publishers.
- McCaig, B. & Pavcnik, N. (2017)** Moving out of Agriculture: Structural Change in Viet Nam. In McMillan, M. S., Rodrik, D. & Sepulveda, C. P. (eds.) *Structural Change, Fundamentals, and Growth: A Framework and Case Studies* (pp. 81-124), International Food Policy Research Institute.
- Meggison, W. L. & Netter, J. M. (2001).** From State to Market: A Survey of Empirical Studies on Privatization. *Journal of Economic Literature*, XXXIX, 321–389.
- Melo, P. C., Graham, D. J. & Noland, R. B. (2009).** A Meta-Analysis of Estimates of Urban Agglomeration Economies. *Regional Science and Urban Economics*, 39(3), 332–342.
- Morrow, J. D., Siverson, R. M., & Tabares, T. E. (1998).** The Political Determinants of International Trade: The Major Powers, 1907–1990. *American Political Science Review*, 92(3), 649-661.
- Murdoch, J. C. & Sandler, T. (2002a).** Economic Growth, Civil Wars, and Spatial Spillovers. *Journal of Conflict Resolution*, 46(1), 91-110.
- Murdoch, J. C. & Sandler, T. (2002b).** Civil Wars and Economic Growth: A Regional Comparison. *Defence and Peace Economics*, 13(6), 451-464.
- Murdoch, J. C. & Sandler, T. (2004).** Civil Wars and Economic Growth: Spatial Dispersion. *American Journal of Political Science*, 48(1), 138-151.
- Neffke, F., Hartog, M., Boschma, R. & Henning, M. (2018).** Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification. *Economic Geography*, 94(1), 23–48.

-
- Neffke, F., Henning, M. & Boschma, R. (2011).** How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, 87(3), 237–265.
- Newman, C., Rand, J., Talbot, T. & Tarp, F. (2015).** Technology Transfers, Foreign Investment and Productivity Spillovers. *European Economic Review*, 76, 168–187.
- Newman, C., & Page, J. (2017).** Industrial Clusters: The Case for Special Economic Zones in Africa. WIDER Working Paper Series No. 015. World Institute for Development Economic Research (UNU-WIDER).
- Nordhaus, W. (2002).** The Economic Consequences of a War in Iraq. NBER Working Paper 9361.
- Oneal, J. R., Oneal, F. H., Maoz, Z., & Russett, B. (1996).** The Liberal Peace: Interdependence, Democracy, and International Conflict, 1950-85. *Journal of Peace Research*, 33(1), 11-28.
- Oneal, J. R. & Russett, B. (1997).** The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985. *International Studies Quarterly*, 41(2), 267-294.
- Oneal, J. R., Russett, B., & Berbaum, M. L. (2003).** Causes of Peace: Democracy, Interdependence, and International Organizations, 1885–1992. *International Studies Quarterly*, 47(3), 371-393.
- Poast, P. (2010).** (Mis)Using Dyadic Data to Analyze Multilateral Events. *Political Analysis*, 18(4), 403-425.
- Poast, P. (2016).** Dyads are Dead, Long Live Dyads! The Limits of Dyadic Designs in International Relations Research. *International Studies Quarterly*, 60(2), 369-374.
- Polachek, S. W. (1980).** Conflict and Trade. *Journal of Conflict Resolution*, 24(1), 55-78.
- Pollins, B. M. (1989).** Conflict, Cooperation, and Commerce: The Effect of International Political Interactions on Bilateral Trade Flows. *American Journal of Political Science*, 33(3), 737-761.
- Puga, D. (2010).** The Magnitudes and Causes of Agglomeration Economies. *Journal of Regional Science*, 50(1), 203–219.
- Rauch, J. E. (1999).** Networks Versus Markets in International Trade. *Journal of International Economics*, 48, 7-35.

-
- Rauch, J. E. (2001).** Business and Social Networks in International Trade. *Journal of Economic Literature*, 39(4), 1177–1203.
- Reuveny, R. (2000).** The Trade and Conflict Debate: A Survey of Theory, Evidence and Future Research. *Peace Economics, Peace Science and Public Policy*, 6(1), 23–49.
- Romer, P. M. (1992).** Two Strategies for Economic Development: Using Ideas and Producing Ideas. *World Bank Economic Review*, 63–91.
- Romer, P. M. (1993).** Idea Gaps and Object Gaps in Economic Development. *Journal of Monetary Economics*, 32(3), 543–573.
- Rosenthal, S. S., & Strange, W. C. (2004).** Evidence on the Nature and Sources of Agglomeration Economies. In J. V. Henderson & J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2119–2171). Elsevier.
- Shleifer, A. & Vishny, R. W. (1994).** Politicians and Firms. *Quarterly Journal of Economics*, 109(4), 995–1025.
- Song, Z., Storesletten, K. & Zilibotti, F. (2011).** Growing Like China. *American Economic Review*, 101(1), 196–233.
- Stewart, F., Humphreys, F. P. & Lea, N. (1997).** Civil Conflict in Developing Countries Over the Last Quarter of a Century: An Empirical Overview of Economic and Social Consequences. *Oxford Development Studies*, 25(1), 11–41.
- Wade, R. (1990).** *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton University Press.
- Wang, J. (2013).** The Economic Impact of Special Economic Zones: Evidence from Chinese municipalities. *Journal of Development Economics*, 101, 133–147.
- Ward, M. D., Stovel, K. & Sacks, A. (2011).** Network Analysis and Political Science. *Annual Review of Political Science*, 14(1), 245–264.
- Warren, T. C. (2010).** The Geometry of Security: Modeling Interstate Alliances as Evolving Networks. *Journal of Peace Research*, 47(6), 697–709.
- Wei, S.-J., Wei, Z. & Xu, J. (2021).** On the Market Failure of “Missing Pioneers”. *Journal of Development Economics*, 152, 102705.
- White, G. (ed.) (1988).** *Developmental States in East Asia*. Palgrave Macmillan.
- World Bank (1993).** *The East Asian Miracle: Economic Growth and Public Policy*. Oxford University Press.

World Bank (2017). Special Economic Zones. An Operational Review of Their Impacts. World Bank.

Zeng, D. Z. (2021). The Past, Present, and Future of Special Economic Zones and Their Impact. *Journal of International Economic Law*, 24(2), 259–275.

English Abstract

This PhD thesis consists of three self-contained chapters within the fields of international and development economics.

Chapter 1: The Role of Labor Composition and Quality in Determining the Productivity-Wage Gap: Evidence from Industrial Zones in Myanmar – with John Rand, Finn Tarp, and Neda Trifković

The first chapter uses a large nationally representative survey of private manufacturing enterprises to estimate productivity and wage gains associated with industrial zones in Myanmar. Moreover, we investigate the contribution of the industrial zones in facilitating agglomeration and human capital returns to firms and workers. Our results show that being located in an industrial zone is associated with higher value added per worker, but the extent to which this is reflected in worker wages depends on the labor force composition. Productivity gains related to agglomeration and hiring of trained workers are larger in industrial zones than elsewhere. While increasing the share of women workers can lead to short-term productivity losses, our estimates show that worker wages improve in firms with a higher share of women workers. Our results highlight the role of labor force quality and composition in affecting the firm performance, but also illustrate that the benefits from industrial zones extend beyond agglomeration, bringing out the advantages of planned industrial zones as opposed to relying on innate industrial clustering.

Chapter 2: Brothers in Arms, Brothers in Trade? Measuring the Effect of Violent Conflicts on Trade with Third-Party Countries

The second chapter contributes to the old and ongoing discussion about the relationship between violent conflicts and international trade. Empirical research in the 1990s and early 2000s has established that violent interstate conflicts harm international trade. While most of this literature dates back at least 10 to 20 years, the effect of interstate conflicts on trade with third-party countries has been neglected for most of the time in the literature. In this chapter, I attempt to fill this gap. A period of 46 years is covered in the analysis, using more than 500 thousand dyad-

year observations. The third-party country dimension is derived from a triadic data set, which covers all possible country-triad combinations for the studied period. I find that violent interstate conflicts reduce trade with third-party countries, and that they cause a shift in trade towards allied countries and away from the enemy's allies. Countries increase imports from members of the same security alliance by between 1 and 4 percent, and trade more with countries that have the same enemies by between 5 and 7 percent. They reduce trade with the formal allies of their enemies by between 9 and 14 percent. This negative trade shifting effect is further amplified by the size of the respective conflict country. This chapter contributes to the literature on conflict and trade in two ways: First, by adding to the scarce literature introducing a third-country dimension into standard gravity models and into the literature on conflict and trade. And second, by showing the importance of a spatially dynamic perspective on interstate conflicts.

Chapter 3: The Ownership of Pioneer Firms: The Role of State-Owned and Multinational Enterprises in Industrial Diversification in Vietnam – with Bjørn Bo Sørensen, Henrik Hansen, and John Rand

In the third chapter, we study the ownership of pioneer firms that establish market activities in provinces where such activities did not previously exist. Using the Vietnamese Enterprise Survey (2001-2017), which allows us to track close to one million formal firms over time, we explore whether Vietnam's remarkable industrial diversification during the past two decades was driven by state-owned, multinational, or domestic and privately owned enterprises (SOEs, MNEs, or PDEs). We document significantly higher ownership-specific pioneering frequencies in SOEs and MNEs compared to PDEs, also after controlling for a broad spectrum of observable characteristics. Using a simple conceptual framework, we attribute these differences to variations in access to capital, knowledge of firm- and industry productivity, and firm-specific strategies and policy objectives. Second, we investigate the dynamic employment effects in industries pioneered by different firms. We find a large employment premium in MNE-pioneered industries. The premium is driven entirely by employment in pioneering MNEs and other MNEs following the pioneer in subsequent years. In contrast, there is no employment premium in industries pioneered by SOEs relative to PDEs. Pioneering SOEs are systematically larger than pioneering PDEs, but this direct employment effect is counterbalanced by less employment in PDEs and MNEs in SOE-pioneered industries in the long run.

Dansk Abstract

Denne PhD afhandling indeholder tre selvstændige kapitler inden for international- og udviklingsøkonomi.

Kapitel 1: Arbejdster Komposition og Kvalitets rolle i Determinering af Produktivitetstilvækst: Evidens fra Industrielle Zoner i Myanmar – med John Rand, Finn Tarp, and Neda Trifković

Det første kapitel bruger en stor nationalt repræsentativ undersøgelse af private fremstillingsvirksomheder til at estimere produktivitet og lønstigninger forbundet med industrizoner i Myanmar. Desuden undersøger vi industrizonernes bidrag til facilitering af agglomeration og afkast af human kapital til virksomheder og arbejdere. Vores resultater viser, at det at være placeret i en industrizone er forbundet med højere værditilvækst pr. arbejder, men afspejles ikke helt i arbejdernes lønninger. Produktivitetstilvækst relateret til agglomeration og ansættelse af uddannet arbejdskraft er større i industrizoner end andre steder. Selvom en stigning i andelen af kvindelige arbejdere kan føre til kortsigtede produktivitetstab, viser vores estimater, at arbejdernes lønninger forbedres i virksomheder med en højere andel af kvindelige arbejdere. Vores resultater fremhæver den rolle, som arbejdsstyrkens kvalitet og sammensætning spiller for at påvirke virksomhedens præstationer, men illustrerer også, at fordelene ved industrizoner strækker sig ud over agglomeration, og fremhæver fordelene ved planlagte industrizoner i modsætning til at stole på naturlig dannelse af industrieklynger.

Kapitel 2: Brødre i Krig, Brødre i Handel? Effektmåling af Voldelige Konflikter på Handel med Tredjepartslande

Andet kapitel bidrager til den gamle og igangværende diskussion om forholdet mellem voldelig konflikt og international handel. Empirisk forskning i 1990'erne og begyndelsen af 2000'erne har fastslået, at voldelige mellemstatslige konflikter skader international handel. Mens det meste af denne litteratur går mindst 10 til 20 år tilbage, er effekten af mellemstatslige konflikter på handel med tredjepartslande blevet forsømt det meste af tiden i litteraturen. I dette kapitel forsøger jeg at udfylde

det hul. Analysen inkluderer 46 år, med mere end 500 tusinde dyad-års observationer. Tredjeparts lantedimensionen er udledt af et triadisk datasæt, som dækker alle mulige lande-triadekombinationer for den undersøgte periode. Jeg finder, at volde-lige mellemstatslige konflikter reducerer handelen med tredjepartslande, og at de forårsager et skift i handelen hen imod allierede lande og væk fra fjendens allierede. Lande øger importen fra medlemmer af den samme sikkerhedsalliance med mellem 1 og 4 procent og handler mere med lande, der har de samme fjender, med mellem 5 og 7 procent. De reducerer handelen med deres fjenders formelle allierede med mellem 9 og 14 procent. Den negative handelsskiftende effekt forstærkes yderligere af størrelsen af det respektive konfliktland. Dette kapitel bidrager til litteraturen om konflikt og handel på to måder: For det første ved at tilføje til den sparsomme litteratur ved at indføre en tredjelandsdimension i standardgravitationmodeller og i litteraturen om konflikt og handel. Og for det andet ved at vise vigtigheden af et rumligt dynamisk perspektiv på mellemstatslige konflikter.

Kapitel 3: Pionervirksomheders Ejerskab: Statsejede og Multinationale Virksomheders Rolle i Industriel Diversificering i Vietnam – med *Bjørn Bo Sørensen, Henrik Hansen, and John Rand*

I det tredje kapitel studerer vi ejerskabet af pionervirksomheder, der etablerer markedsaktiviteter i provinser, hvor sådanne aktiviteter ikke tidligere fandtes. Ved at bruge den vietnamesiske virksomhedsundersøgelse (2001-2017), som giver os mulighed for at spore næsten en million formelle virksomheder over tid, undersøger vi, om Vietnams bemærkelsesværdige industrielle diversificering i løbet af de sidste to årtier var drevet af statsejede, multinationale eller indenlandske og private ejede virksomheder (hhv. SOE'er, MNE'er eller PDE'er). Vi dokumenterer betydeligt højere ejerskabsspecifikke banebrydende frekvenser i SOE'er og MNE'er sammenlignet med PDE'er, også efter at have kontrolleret for et bredt spektrum af observerbare karakteristika. Ved hjælp af en simpel konceptuel ramme tilskriver vi disse forskelle variationer i adgang til kapital, viden om virksomheds- og industriproduktivitet og virksomhedsspecifikke strategier og politiske mål. For det andet undersøger vi de dynamiske beskæftigelseseffekter i brancher, pioneret af forskellige virksomheder. Vi finder en stor beskæftigelsespræmie i MNE-pionerindustrier. Præmien er udelukkende drevet af ansættelse i banebrydende multinationale selskaber og andre multinationale selskaber, der følger pioneren i de efterfølgende år. I modsætning hertil er der ingen beskæftigelsespræmie i industrier, der er pioneret af SOE'er i forhold til PDE'er. Pionerende SOE'er er systematisk større end pionerende PDE'er, men den direkte beskæftigelseseffekt opvejes af mindre beskæftigelse i PDE'er og MNE'er i SOE-pionerindustrier på lang sigt.

Chapter 1

The Role of Labor Composition and Quality in Determining the Productivity-Wage Gap: Evidence from Industrial Zones in Myanmar

The Role of Labor Composition and Quality in Determining the Productivity-Wage Gap: Evidence from Industrial Zones in Myanmar*

John Rand, Finn Tarp, Neda Trifković, and Helge Zille

Department of Economics, University of Copenhagen

This paper uses a large nationally representative survey of private manufacturing enterprises to estimate productivity and wage gains associated with industrial zones in Myanmar. Moreover, we investigate the contribution of the industrial zones in facilitating agglomeration and human capital returns to firms and workers. Our results show that being located in an industrial zone is associated with higher value added per worker, but the extent to which this is reflected in worker wages depends on the labor force composition. Productivity gains related to agglomeration and hiring of trained workers are larger in industrial zones than elsewhere. While increasing the share of women workers can lead to short-term productivity losses, our estimates show that worker wages improve in firms with a higher share of women workers. Our results highlight the role of labor force quality and composition in affecting the firm performance, but also illustrate that the benefits from industrial zones extend beyond agglomeration, bringing out the advantages of planned industrial zones as opposed to relying on innate industrial clustering. (JEL D24, L25, R12, J31, J81)

*A previous version of this paper is published in the WIDER working paper series under the title *Industrial Agglomeration in Myanmar* (Rand et al., 2019). This research formed part of the project Towards Inclusive Development in Myanmar financed by Danida and implemented by UNU-WIDER, the Central Statistical Organization (CSO) of the Ministry of Planning and Finance of Myanmar, and the University of Copenhagen Development Economics Research Group (UCPH-DERG). The authors are grateful to the partners involved for enabling data access. We extend our thanks to colleagues and collaborators, including Dr. Tomohiro Machikita, Prof. Toshihiro Kudo,

Prof. Tetsushi Sonobe, Dr. Mari Tanaka, Dr. Yutaka Arimoto, Dr. Kensuke Teshima, as well as seminar participants at IDE-JETRO, GRIPS, Hitotsubashi University, Columbia University, and the Nordic Conference on Development Economics for many helpful comments and suggestions.

1. INTRODUCTION

Promoting industrial or special economic zones (SEZs) has been an integral tool of industrialization policies in many low- and middle- income countries.¹ Industrial hubs may provide several benefits, such as decreasing production costs, creating market linkages, or improving access to labor and knowledge. Industrial zones are usually created to increase manufacturing exports, industrial upgrading, and employment (Cirera and Lakshman, 2017). By providing infrastructure and an enabling business environment, industrial zones can also play an important role in attracting foreign investments. Clusters are an innate result of industrial growth and transformation, while industrial zones are the result of deliberate planning and government intervention. In many cases, the development of industrial zones and the formation of industrial clusters overlap, as industrial zones can provide favorable conditions for clustering of economic activity (Tsuji et al., 2007).

The benefits of industrial zones arise from access to better infrastructure, skilled labor, and agglomeration (geographic clustering of industrial activity). In a number of studies, industrial clustering has been shown to increase firm productivity. For instance, Greenstone et al. (2010) find a 12 percent higher total factor productivity among firms located in industrial clusters. Hu et al. (2015) attribute 14 percent of China's productivity growth to industrial agglomeration. Particularly strong evidence comes from the Chinese textile industry (Lin et al., 2011). Similarly, SEZs have been found to increase productivity, foreign direct investment, and wages (Wang, 2013). Alongside productivity, agglomeration externalities can also include wage effects (Fafchamps and Hamine, 2017), transfer of technology and skills (Howard et al., 2016), creation of input-output linkages (Ellison et al., 2010), and better access to market centers (Lall et al., 2004).

However, the benefits of SEZs are by no means guaranteed or distributed equally. This depends, for example, on whether the zone is in a strategic location and how the industrial zone strategy is integrated in the overall development strategy (Zeng, 2021). To illustrate, zones located in more remote areas seem to lag behind those in larger cities (Combes et al., 2012; Frick and Rodríguez-Pose, 2019). A lack of trust may decrease interaction between business owners in the same sectors and prevent knowledge spillovers (Robertson and Taung, 2015). Additional risks and challenges of agglomeration include congestion and pollution, as well as fierce competition and growth of

¹ See, for example, Newman et al. (2016), Newman and Page (2017), and Page and Tarp (2017).

within-country inequality (Deichmann et al., 2008; Hu et al., 2015). Besides, the continuing rise of the importance of information and telecommunications technology may threaten to reduce the value of physical proximity, and hence the value of agglomeration economies (Giuliano et al., 2019).

The existence and exact magnitude of benefits associated with industrial zones remain open questions, despite their theoretical and practical potential. This paper estimates the magnitude of benefits to firms and workers associated with industrial zones. First, we compare the relative marginal products and relative wages of workers employed in private manufacturing enterprises operating in and outside industrial zones in Myanmar. Under perfect competition, relative marginal products and wages equal out and firms are indifferent between locating in or outside an industrial zone. When this equality is not obtained, and the relative marginal product is higher than relative wages, profit-maximizing or cost-minimizing firms will prefer to be located in an industrial zone. Such a result is a sign of some deviation from competitive spot markets, such as a lack of complementarity between labor and technology. Second, we investigate to what extent different agglomeration profiles of industrial zones (measured by the Ellison-Glaser Index) can account for their differential performance. Third, we estimate the contribution of industrial zones in facilitating human capital returns to firms and workers, measured as interactions between the industrial zone indicator and the share of skilled workers, the share of trained workers, and the share of women workers. This focus is important in coming to grips with the economic mechanisms behind place-based industrialization policies.

We focus on Myanmar for several reasons. First, it presents a unique setting for disentangling the effects of planned as compared to natural industrial agglomeration. The first industrial zones were established in Myanmar in the mid-1990s. Unlike in other South-East Asian countries, firms in Myanmar were initially forced to locate in the industrial zones by policy prescription, rather than by financial incentives such as tax cuts (Robertson and Taung, 2015; Tsuji et al., 2007). Tax exemptions were introduced in the 2010s, mainly to attract foreign investment. Second, prior to the coup d'état in 2021, profound economic and political reforms were underway, including a transition from a controlled into a market-based economy. The intensity of industrial activity increased, and the number of jobs in manufacturing grew. Between 1990 and 2015, the industrial sector grew from 10 to 34 percent of gross domestic product (GDP), while the contribution of the agricultural sector fell from 60 to 27 percent of GDP (CSO, 2017). Third, structural transformation extended

employment opportunities to women, in particular in manufacturing, following the expansion of industrial zones (Myint and Rasiah, 2012). Early accounts of the expansion of manufacturing jobs in the country show that women earned higher wages than men (Francisco, 2007). Later accounts, however, indicate that there may be household welfare losses associated with women migrating to find work in manufacturing (Mueller et al., 2020) and that significant wage differences exist between female and male manufacturing workers, with women, *ceteris paribus*, earning 13 percent less than men (Hansen et al., 2020). Finally, industrial zones in Myanmar have been criticized for a lack of infrastructure, investment, regulation, and efficiency (Min and Kudo, 2013; Robertson and Taung, 2015), which makes it relevant to assess the benefits of industrial zones amidst the difference in the quality of services they provide.

This paper contributes to a growing body of literature on agglomeration economies in developing countries. The evidence on industrial, export processing, or special economic zones in developing countries so far *inter alia* comes from other more developed Asian countries, including China, India, and Vietnam (see, e.g., Alkon, 2018; Ge, 2009; Howard et al., 2016; Hu et al., 2015; Huang et al., 2008; Ito et al., 2015; Lall et al., 2004; Li et al., 2012; Wen, 2004). Potentially the most comparable economic setting is Cambodia, where Chhair and Newman (2014) investigate productivity effects of industrial agglomeration, but they do not focus on the returns to capital and labor. In addition, we contribute to the literature by investigating whether the productivity gains of industrial agglomeration benefit workers in any way. This extends earlier literature on the link between exports and working conditions in manufacturing firms in Myanmar (see, e.g., Tanaka, 2019) and the research on the returns to women's employment in manufacturing enterprises (Mueller et al., 2020). Fafchamps and Hamine (2017) and Wang (2013) estimate productivity and wage effects of agglomeration and SEZs, but they neither account for different agglomeration profiles of industrial zones and innate industrial clusters, nor do they consider the differential contribution from the labor force composition to enterprise and wage gains.

We use data from the first two nationally representative surveys of manufacturing enterprises in 2017 and 2019 (carried out prior to the coup d'état in 2021). Quantitative results are supplemented with findings from qualitative interviews with enterprise owners, workers, and key institutional actors. This enables a better understanding of contextual factors relevant for business performance. The panel structure of the data enables us to account for the changing composition of the workforce

over time. We use the control function approach to estimate the production function, which allows accounting explicitly for endogeneity of human capital variables. The production function estimates, in turn, provide us with a measure of unobserved worker characteristics, which we include in the wage equation to obtain consistent estimates of the impact of human capital variables on production worker wages. We extend the approach in Konings and Vanormelingen (2015) by distinguishing between observed and unobserved firm factors and by allowing for slope effects in returns to human capital investments and for level effects in the observed firm factors.

We find that operating in an industrial zone associates with higher value added and wages for manufacturing enterprises with average composition of skills, training, and women workers under average agglomeration levels. Productivity gains vary with agglomeration and the share of trained workers. We do not find statistically significant wage gains from operating in industrial zones for enterprises with above-average labor quality and above-average level of agglomeration. An exception is found for enterprises in industrial zones with an above-average share of women workers, for which gains in worker wages are offset by losses in value added. We also discover that increasing the share of skilled workers above average is associated with lower value added and wages. Our findings illustrate the importance of labor input quality and composition for achieving productivity gains. The rents captured by employers may be undermined by labor shortage (in particular of women workers) or frequent employment turn-over among workers. This calls for consideration of the benefits of a sound legal and regulatory framework targeted at improving working conditions in the manufacturing sector.

We proceed as follows. Section 2 presents a review of the literature on industrial agglomeration and its associated benefits, such as firm productivity. It also describes the establishment of industrial zones, while Section 3 describes the data. Section 4 introduces our estimation strategy, Section 5 presents the results, before Section 6 includes robustness checks and a discussion. Section 7 concludes, highlighting the policy implications of our findings.

2. LITERATURE REVIEW

Marshall (1890) argued that spatial agglomeration affects firm productivity through knowledge spillovers, input sharing, and labor market pooling. Numerous studies support Marshall's theories (e.g., Ellison et al., 2010; Greenstone et al., 2010), categorizing the agglomeration effects into localization and urbanization economies. Localization economies attribute the improved efficiency to firm clustering in the same industry (Rosenthal and Strange, 2004), while urbanization economies suggest that benefits arise from concentration of different kinds of activities in a given area and need not be industry-specific (Jacobs, 1969).

The agglomeration of industrial activities has significant impacts on firm performance, including productivity (Ciccone and Hall, 1996), organization of production processes (Holmes, 1999), and innovation (Carlino et al., 2007; Feldman and Audretsch, 1999). The benefits of agglomeration arise through several channels.

First, close geographic proximity promotes the potential for sharing of knowledge through formal and informal interactions, among either firms or individuals (Marshall, 1890). New technology, patent citations, and innovations are found to be more likely to originate from the same geographic area (Audretsch and Feldman, 1996; Jaffe et al., 1993; Kelley and Helper, 1999). The same is argued for management skills and business knowledge (Rosenthal and Strange, 2004; Sorenson and Audia, 2000). Entry into export markets is facilitated when a larger number of exporting firms is concentrated in the same area, as the sharing of knowledge about foreign markets can reduce the entry costs for new exporters (Greenaway and Kneller, 2008; Ito et al., 2015). If agglomeration of economic activity leads to knowledge spillovers, it is also likely to lead to higher productivity and we should observe that firms located in denser areas are more productive than comparable firms located in less dense areas (Greenstone et al., 2010).

Second, industrial agglomeration enables the production of specialized intermediate inputs to reach the level that enables benefiting from scale economies (Marshall, 1890). This, in turn, enables firms to outsource a higher share of their intermediate inputs and specialize in the most profitable activities (Holmes, 1999).

Third, geographical concentration of firms and workers allows more effective matching between employers and workers when many firms offer jobs and many workers look for jobs in the same

area. Increased labor market density reduces the risk of unemployment for workers and the risk that a firm cannot fill its vacancies (Krugman, 1993). The impact of agglomeration on productivity could be two-sided. If thickness in the labor market only improves worker-firm matches, firms located in denser areas would be more productive than comparable firms located in less dense areas. If instead thickness in the labor market leads to lower unemployment risk, productivity would not change, but wages may vary, which may affect the relative use of labor and capital (Greenstone et al., 2010).

Finally, market size is also relevant as it enables benefiting from economies of scale and lower transportation costs. Proximity to a larger market may attract industries if transport costs are neither too low nor too high (Krugman, 1991). Firms located in denser areas are likely to enjoy cheaper and faster delivery of local services and local intermediate goods (Greenstone et al., 2010). The implications are lower production costs in denser areas.

Not all firms will benefit from agglomeration externalities in the same way. This depends on both the cluster and the firm characteristics. Positive external effects in the form of technology transfers are more likely for firms located within more productive clusters that produce similar products (Chhair and Newman, 2014). Consider, for instance, a case of labor movement between firms where employees transfer their skills and knowledge from one employer to the next. Consider also the evidence that the diffusion of new technology is faster in geographical areas with more knowledge about such technologies and that it depends positively on the proximity of early users (Baptista, 2001). Greenstone et al. (2010) find larger productivity spillovers among plants sharing similar labor and technology pools.

In terms of firm characteristics, Henderson (2003) finds that the net benefits of industrial clustering mainly go to high-tech industries, given the relatively high importance of knowledge sharing and access to specialized inputs and services, rather than to standardized manufacturing industries, such as textiles, wood, and food processing. Clustering of small and medium enterprises has been beneficial for quality upgrading of Chinese firms (Fleisher et al., 2010; Huang et al., 2008). Summary evidence of productivity spillovers associated with agglomeration in developing countries is provided by Newman and Page (2017), who conclude that the localization effect is also important for industries in less developed countries.

Apart from the above-described benefits, agglomeration can be associated with negative externalities. For instance, Lall et al. (2004) emphasize the burden of costs of locating in dense urban

areas, such as high wages, rents, and congestion. Furthermore, stronger competition in product and factor markets can decrease mark-ups and productivity through higher input costs and lower final product prices (Hu et al., 2015). The chances of capturing the benefits of agglomeration are more limited for small manufacturing firms, located in areas dominated by a few large firms from their own industry, than for firms located in areas with a less concentrated industrial structure (Drucker and Feser, 2012). It therefore remains an empirical issue whether the benefits of agglomeration for productivity exceed the diseconomies.

2.1. Measures of Industrial Agglomeration

To identify industrial agglomeration, a number of different measures and indices have been established. Two of them have become the standard in most of the literature. The first is the Ellison-Glaeser Index (Ellison and Glaeser, 1997). It measures agglomeration as geographical concentration compared to a completely random distribution of firm locations, thereby capturing the excess concentration beyond the random distribution.² The Ellison-Glaeser Index uses (i) a raw index of geographic concentration, aggregating the employment shares of a certain industry over different regions; (ii) the Herfindahl-Hirschman Index, which measures firm sizes relative to a specific industry; and (iii) a reverse of the Herfindahl-Hirschman Index. Zero concentration here means that a firm's decision to locate is completely independent of other firm locations. Ellison and Glaeser (1997) originally used their index to describe agglomeration patterns in the US, and it has since then been used, for instance, to measure industrial agglomeration in Belgium (Bertinelli and Decrop, 2005), Spain (Alonso-Villar et al., 2004), and Sweden (Braunerhjelm and Borgman, 2004), and to investigate the relationship between globalization and industrial agglomeration in China (Ge, 2009). Other studies used, for instance, the Herfindahl-Hirschman Index for sector level concentration (Chhair and Newman, 2014), which is a sub-index for the calculation of the Ellison-Glaeser Index, Moran's spatial correlation coefficient (e.g. Gibbs and Bernat, 1997), which measures the concentration of firms from a given industry within a geographical unit, or other measures based on geographic distance (Billings and Johnson, 2016; Duranton and Overman, 2005).

² Positive values of this index indicate that agglomeration is larger than what would be expected by pure randomness, while negative values indicate that agglomeration is lower than expected by randomness (Casseo and Smith, 2014).

The second, similarly prominent measure of agglomeration, is an indicator first suggested by Florence (1939), which is a measure of an industry's employment share in a certain location relative to that industry's employment share in the whole economy, called the employment location quotient (or just location quotient). This measure is attractive because it offers an obvious and easily interpretable threshold for industrial clustering. It has since become one of the most widely used measures to quantify industrial concentration (Guimarães et al., 2009), and has been used in a variety of contexts. These include studying the relationship between industrial agglomeration and firm size in the US (Holmes and Stevens, 2002) and in China (Li et al., 2012), a core-periphery analysis of the European Union (Mack and Jacobsen, 1996), and a study of the relationship between industrial agglomeration and economic development in East Asia and China (Fan and Scott, 2009).

All these indices require detailed data on sectors and employment on a sub-national level. They were therefore mainly applied in developed countries, especially the US, where data availability is higher, thereby leaving a gap on measuring industrial agglomeration in developing countries (with some exceptions in Asia, especially China).

2.2. Industrial Zones in Myanmar

The first industrial zones in Myanmar were established in the mid-1990s, in an attempt to strengthen industrial development, create jobs, attract foreign investments, and keep industrial activity away from residential areas. In other South-East Asian countries such as China or Vietnam, authorities have usually created financial incentives such as tax cuts to attract firms to locate inside industrial zones (Tsuji et al., 2007). In Myanmar, however, firms were initially forced to locate or relocate into industrial zones. To avoid that, some firms temporarily shut down their business, only to reopen again after a while when attention had shifted away (Abe and Dutta, 2014; Robertson and Taung, 2015).

The planning of the industrial zones was initiated under the Private Industrial Enterprise Law from 1990 and the Promotion of Cottage Industrial Law from 1991. In 1995, the Myanmar Industrial Development Committee was founded, under which the industrial zones were established, for which the military government delegated competences to local and regional authorities that implemented the installation of the zones. Initially, there were 18 different zones, all autonomously acting

and with different rules and regulations (Linn et al., 2014). Additional industrial zones have been established or planned since then, even though different sources provide differing estimates. A high number of industrial zones, or sub-zones, has been established in Yangon alone (Min and Kudo, 2013; Zaw and Kudo, 2011). The zones are principally managed by the Industrial Zone Management Committee, which is a sub-subcommittee of the Industrial Development Committee. Some zones are joint ventures between foreign investors, such as from Japan, and domestic firms or authorities. As such, domestic firms are also involved in the management of the zones in some cases (Min and Kudo, 2013).

Most industrial zones lie in and around the two major urban centers, Mandalay and Yangon, where they are connected to markets and transportation (Min and Kudo, 2013). In Yangon, the industrial zones are dominated by larger firms, while in Mandalay and in other industrial zones outside of the two urban centers, the majority of firms are rather small (Robertson and Taung, 2015).

Industrial zones in Myanmar have been the subject of much debate. In the first place, there was no proper assessment of the feasibility of industrial zones in the respective locations. In addition, there was no clear strategy on how to promote industrial development within the zones, or how to integrate them in the overall economic development strategy (Linn et al., 2014). Accordingly, the industrial zones are often criticized for a lack of infrastructure, investment, regulation, efficiency, and especially a lack of sufficient and steady electricity access and supply (Min and Kudo, 2013; Robertson and Taung, 2015). Even though electricity grid access is high, many firms in industrial zones suffer from regular power shortages, especially in and around the densely populated areas of Yangon and Mandalay, where the majority of firms and people are located, and where the demand is high. A part of the problem has been that private households have in some cases been prioritized in terms of energy supply, while the industrial zones only received the surplus (Min and Kudo, 2013).

Numerous industrial zones are found to run substantially below full capacity. Many plots within the zones are vacant or used for other purposes such as warehousing rather than production. Further, in many industrial zones there are registered firms that are in fact not operating (Linn et al., 2014). These problems indicate a lack of incentives or potential benefits for firms to operate in the zones. Moreover, the zones face problems of land grabbing and of high land prices due to land speculation (Min and Kudo, 2013).

3. DATA

3.1. Quantitative Data

The main data used in this paper comes from a nationally representative enterprise survey, conducted in 2017 and 2019, known as the Myanmar Enterprise Monitoring Survey (MEMS), which focuses on the manufacturing small and medium enterprise sector (UNU-WIDER, 2017). The sample includes 2,496 non-state enterprises in 2017 and 2,497 enterprises in 2019 from all 15 regions and states in Myanmar, including the Nay Pyi Taw Union Council. The data are representative of about 70,000 firms in the manufacturing sector.

A stratified sampling approach was used to select enterprises. First, the sampling frame was created from the lists of active enterprises in each municipality. Only enterprises from the manufacturing sector (2-digit codes 10–33 in the Myanmar Standard Industrial Classification (MSIC)) were kept in the sampling frame. As around one third of all registered enterprises were registered as rice mills (MSIC code 1063), the population of firms was stratified into rice mills and other manufacturing. This resulted in a stratified sampling frame of 19,783 rice mills and 51,443 other manufacturing firms. Second, using probability proportional to size sampling, a number of townships within each region were selected for survey implementation.³ Third, enterprises were drawn randomly from the list, accounting for different townships to obtain a representative sample at the region or state level.⁴ Apart from the officially registered enterprises, the survey includes informal firms identified randomly on-site which comprise about 15 percent of the sample. More detailed information about sampling is available in Berkel et al. (2018) and Hansen et al. (2020).

The main questionnaire was administered to enterprise owners or managers. It included information on enterprise characteristics and practices, such as the number and the structure of the workforce, technology and innovation, revenues and costs, customers, owner characteristics, and economic constraints. Questions refer to the situation in the previous calendar year, whereas the economic accounts contain information on two consecutive years before the survey. Employees

³ The number of townships selected in each state or region was determined in proportion to the total number of townships in each state or region.

⁴ To determine the sample size, we used the sample-to-population ratio for the state or region with the lowest number of registered enterprises and applied a square root rule to scale to other states and regions.

were interviewed following the employee questionnaire with questions on wages, non-wage benefits, education, and the recruitment process.

We base the present analysis on firms that were operating and agreed to be surveyed in both years. Compared to 2017, 1.6 percent of enterprises were temporarily closed in 2019 and 9 percent are replacements for those exiting the survey. We also excluded from the analysis enterprises that do not employ any workers, those that did not report the value of the total wage bill, and those without the wage information (279 enterprises). A number of enterprises had incomplete or implausible economic accounts information. For example, 50 firms did not report the value of assets, 164 did not report the value of intermediates or indirect costs while a further 27 firms had negative value added and profit values. These enterprises were excluded from the analysis, as it is likely that other information on these firms is also unreliable. This left 1,743 firms (and 3,486 observations) in the final sample used here.

In addition to using the municipal lists of manufacturing firms to draw the survey sample, we used these lists to calculate our measures of industrial agglomeration as the lists contain a registry of all enterprises to the level of a township and their 2-digit industry. The lists were merged with the MEMS survey data, which contains information on firm size. The sample information is weighted and used to estimate total employment at the industry, township, and country level required to calculate agglomeration indices.

3.2. Qualitative Data

The analysis also relies on qualitative data from 108 in-depth interviews with enterprise owners, managers, employees, and officers from relevant government institutions. The interviews were conducted over the course of six weeks in the period from March to September 2018, thus between the two survey rounds, in 14 townships from eight regions/states.

The respondents were chosen from the quantitative survey data based on a combination of purposeful and random sampling. The purposeful part of the sampling comprised identifying enterprise categories of interest to maximize the breadth and relevance of information. The categories were established based on industry size, township, firm size, and being located in an industrial zone. In terms of industries, we focused on the largest ones: textiles, apparel, wood, and food, and we also

included industries that are linked to the largest ones through supply chains. For example, leather is linked with the apparel industry. Other industries, such as pharmaceutical, were included due to their uniqueness.⁵ In terms of location, we selected townships with the highest concentration of firms in a particular industry-size category and then sampled randomly within the industry-size category in each township. This means that if there were several firms in the same industry-size category, we would select one or two at random, depending on the total number available. Similarly, if there were several firms of the same industry-size category in the industrial zone, we would select some at random. The same approach was applied for firms located outside the industrial zones. Nine of the enterprises sampled in this way needed to be replaced by others as they either had stopped operating, or could not be reached on the appointed day. The replacement firms were always from the same location (village or town) and from the same industry and size category.

The interviews took the semi-structured form of conversation with key topics prepared in advance: production characteristics, supply chains, employment conditions, access to finance, formalization, business associations, informal payments, and business environment. The interviews always started with questions about production characteristics. Sensitive questions, such as those about informal payments and challenges were asked at the end. The order in which other questions were asked varied from one respondent to another to prevent the fatigue effect. The interview team comprised both Myanmar and English language speakers, so the interviews were conducted with two-way interpretation. The interviews were recorded and the research team also took hand-written notes. The interview material was subjected to thematic and interpretative analysis, mainly to supplement the findings from the quantitative analysis.

3.3. Summary Statistics

Table 1 shows the basic summary statistics for the key quantitative variables in both 2017 and 2019. Exactly 20 percent of the surveyed enterprises are located in industrial zones. The real value added⁶ was 67 million Myanmar Kyats (MMK) in 2017 and 295 million MMK in 2019.⁷ The real average production worker wage was around 145,000 MMK per month in 2017, which increased to about

⁵ See Appendix Table A1 for an overview.

⁶ Value added is calculated as revenue from sales minus total costs, including expenses on intermediate goods and raw materials as well as indirect costs.

⁷ 1 USD was around 1,242 Kyats in 2016 and 1,430 Kyats in 2018 (the questions refer to these two years).

168,000 in 2019. As usual, larger firms are able to pay higher wages (shown in Figure 1). The official minimum wage was 450 MMK per hour, or 3,600 MMK per day in 2016, according to the International Labor Organization’s guide to Myanmar labor law (ILO, 2017). The rate increased in January 2018 to 600 Kyats per hour or 4,800 Kyats per day. This rate is based on an eight-hour per day pay, excluding overtime, bonuses, incentives, or any other allowances. While the surveyed enterprises, on average, pay higher than the monthly minimum wage, 20 percent do not.⁸ The share of enterprises paying below the minimum wage increased from 14 percent in 2017 to 26 percent in 2019.

Table 1: Summary Statistics

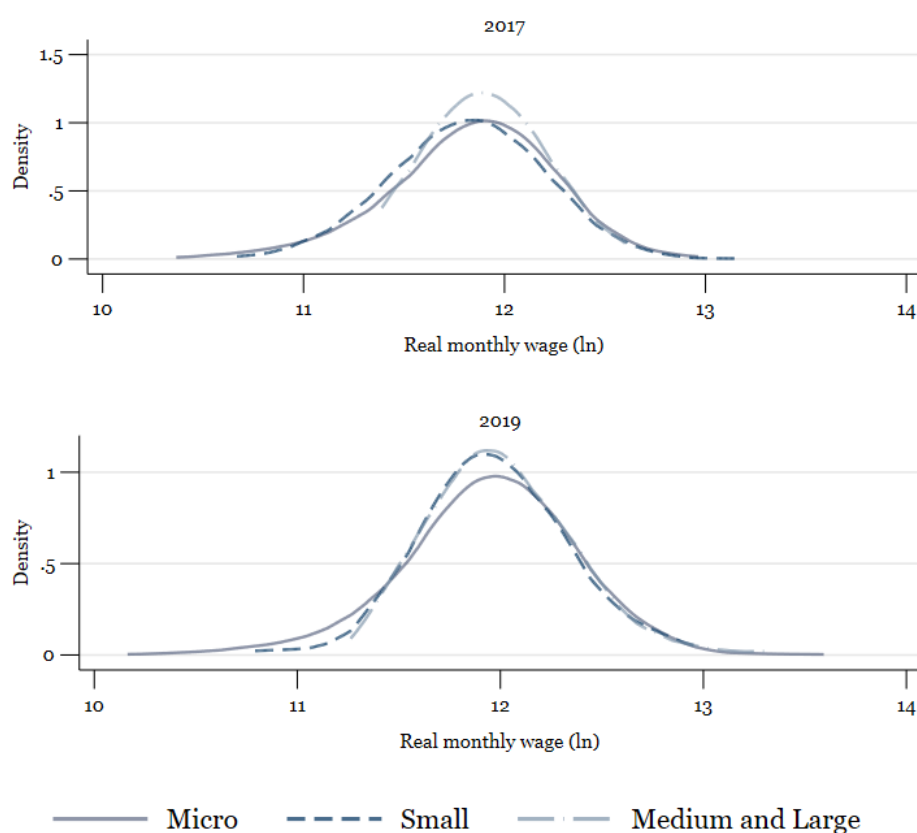
Variables	2017	2019	Both Years
<u>Enterprise variables</u>			
Value added (real, million MMK)	66.76	295.14	180.95
Worker monthly wage (real, MMK)	146,915	167,705	157,310
Industrial zone	0.20	0.20	0.20
Capital (real, million MMK)	48.73	273.82	161.28
Firm size	12.57	12.56	12.56
Firm age	18.55	17.60	18.07
Family firm	0.28	0.28	0.28
Ellison-Glaeser Index	0.16	0.16	0.16
Share of skilled workers	0.87	0.78	0.82
Share of women workers	0.24	0.27	0.26
Share of workers with formal contracts	0.02	0.02	0.02
Share of workers receiving fringe benefits	0.36	0.29	0.32
<u>Owner/manager characteristics</u>			
Owner female	0.29	0.32	0.30
Respondent age	49.82	48.73	49.28
No education	0.01	0.05	0.03
Primary school	0.22	0.22	0.22
Middle school	0.24	0.24	0.24
High school	0.17	0.15	0.16
College/Bachelor’s/Master’s degree	0.32	0.32	0.32
Other degree	0.04	0.02	0.03
Observations	1,743	1,743	3,486

Notes: Value added, wages, and input values are adjusted for price differences among regions.

⁸ The minimum monthly wage in 2017 was 93,600 MMK, calculated as 3,600 Kyats times 26 days. The minimum wage in 2019 was 124,800 MMK, calculated as 4,800 MMK times 26 days.

Enterprises owned on average assets valued at around 160 million MMK. While the surveyed enterprises employ 12 workers on average, three-quarters employ less than nine. Only 5 percent of the sample are enterprises with more than 50 employees. The sample comprises a mix of old and newly established firms. The majority has been operating for around 18 years. Just above one quarter of firms are family-owned.

Figure 1: Average Production Worker Wage by Firm Size



Notes: Authors own calculations based on MEMS data.

While the employment numbers did not change between the two survey years, the employment structure did. The proportion of skilled production workers in total labor force decreased from 87 percent to 78 percent, which reflects owners' perceptions about the level of skills of their workers. The share of women workers increased from 24 percent in 2017 to 27 percent in 2019. While only 2 percent of workers have formal signed employment contracts, 32 percent of workers receive some

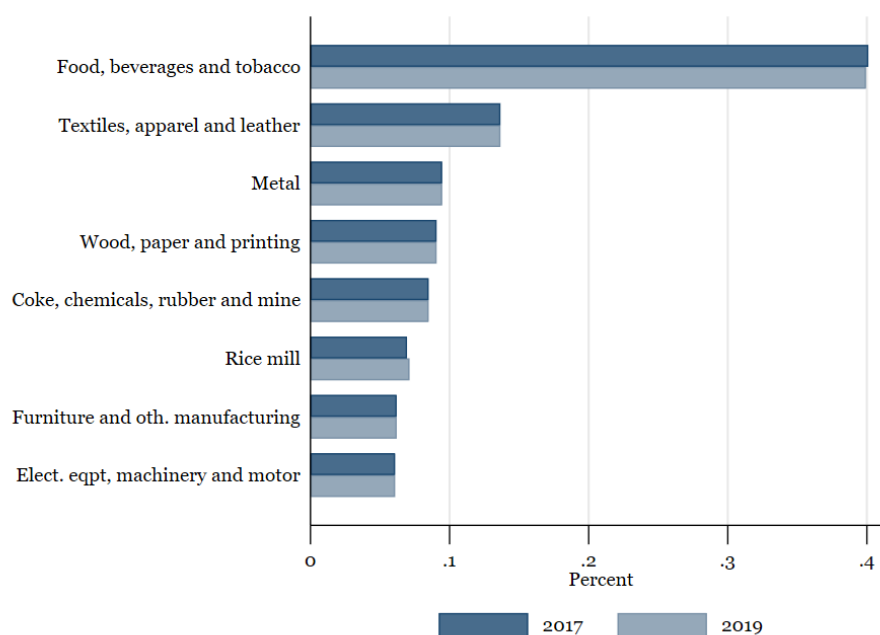
form of fringe benefits, such as sick leave with pay, insurance against accidents, paid maternity leave, severance pay, or annual leave with pay.

Enterprise owners or managers are on average 49 years old. One third of enterprises are owned by women who tend to be slightly younger than male owners. About one quarter of enterprise owners or managers have not progressed beyond primary school, which corresponds to only four years of schooling. Almost the same proportion of respondents has stopped their education at the middle-school level, which corresponds to eight years of schooling. About 30 percent of enterprise owners or managers have completed a Bachelor's degree.

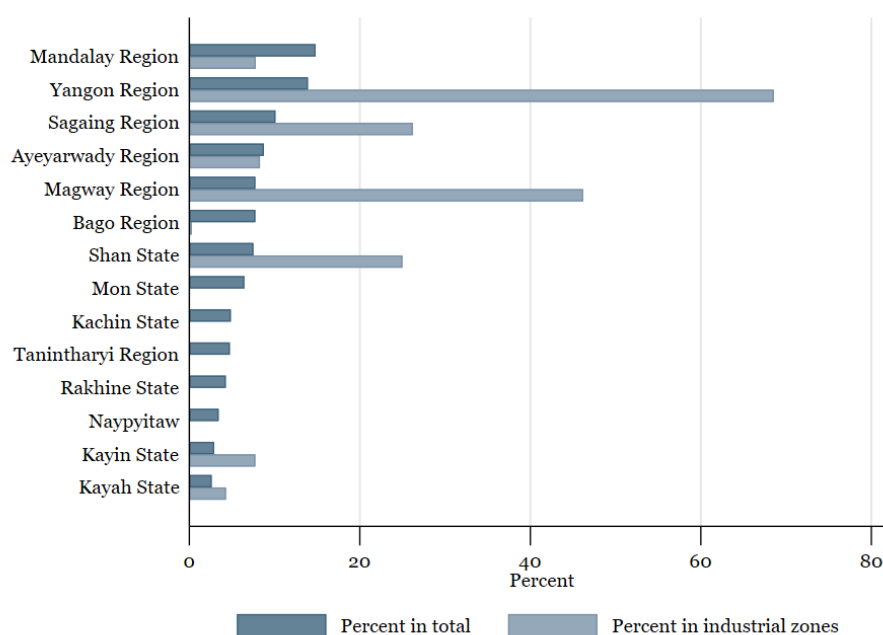
The selected measures of industrial clustering indicate a moderately dispersed agglomeration profile of the manufacturing sector in Myanmar. Ellison and Glaeser (1997) suggested that the degree of geographical concentration can be classified into highly concentrated (values above 0.05), relatively concentrated (values between 0.02 and 0.05), and not very concentrated (values below 0.02). We obtain a value of 0.16, indicating a highly concentrated manufacturing sector. However, the values of the Ellison-Glaeser Index range from 0.015 to 1.16, indicating quite a dispersed agglomeration profile. This level of dispersion suggests that different forces of concentration may affect different industries and that natural advantage may play a role for some (Ellison and Glaeser, 1997). The alternative agglomeration measures we considered also suggest beyond-random industry concentration. The location quotient shows an average of 3.81, which, even though high, appears to be lower than in China, where the average recorded value is 4.66 (Li et al., 2012). The relative enterprise density variable, measured as an industry's enterprise share in a township over the total number of enterprises averages 2.83, but the values range from zero to 40.

The most common industries in our data are food, textiles, and wood, as illustrated in Figure 2. Following the general distribution of population and country geography, most of the firms are based in Mandalay, Yangon, and Sagaing regions, and only few in the areas bordering China, India, and Thailand, as illustrated in Figure 3. Figure 3 further shows that the Yangon, Magway, and Sagaing regions have the highest prevalence of firms located in industrial zones. There are six regions, on the other hand, without any industrial zones.

Table 2 compares the key characteristics of enterprises located in and outside of industrial zones. We obtain that enterprises located in industrial zones have higher value added and capital in 2017, while worker wages appear to be equal in and outside industrial zones. The industrial

Figure 2: Prevalence of Different Sectors in the Sample

Notes: Authors own calculations based on MEMS data.

Figure 3: Proportion of Enterprises in Each Region/State and Industrial Zones

Notes: Authors own calculations based on MEMS data.

zone enterprises are much more likely to provide social benefits and offer formal written contracts, probably because there is more pressure to operate in accordance with the labor law. Our qualitative

interviews with owners and workers suggest that the higher level of social benefits in industrial zones could be related to higher chances of labor and other inspections. Those located in industrial zones, especially in the Yangon Region, were highly aware of labor laws and mentioned that they regulate employment according to the law.

Table 2: Differences in Key Characteristics of Enterprises Inside and Outside Industrial Zones

	2017		2019		Both Years	
	Out	In	Out	In	Out	In
<u>Dependent variables</u>						
Value added (real, million MMK)	47.17	146.46***	299.65	276.77	173.41	211.61
Worker monthly wage (real, MMK)	147,295	145,370	167,993	166,534	157,644	155,952
<u>Independent variables</u>						
Capital (real, million MMK)	34.23	107.72***	283.89	232.90	159.06	170.31
Firm size	8.68	28.42***	8.52	28.99***	8.60	28.70***
Firm age	18.69	17.96	17.77	16.89	18.23	17.42
Family enterprise	0.30	0.21***	0.30	0.20***	0.30	0.20***
Ellison-Glaeser Index	0.17	0.13***	0.17	0.13***	0.17	0.13***
Skilled workers share	0.87	0.84	0.79	0.75*	0.83	0.80**
Women share	0.25	0.21*	0.29	0.22**	0.27	0.21***
Labor contract share	0.02	0.02	0.00	0.07***	0.01	0.04***
Fringe benefits	0.31	0.54***	0.24	0.49***	0.28	0.52***
Respondent female	0.31	0.22**	0.34	0.22***	0.32	0.22***
Respondent age	50.03	48.97	48.89	48.10	49.46	48.53
No education	0.01	0.00*	0.06	0.01***	0.04	0.00***
Primary school	0.25	0.07***	0.23	0.15**	0.24	0.11***
Middle school	0.24	0.24	0.24	0.25	0.24	0.24
High school	0.16	0.21*	0.15	0.17	0.16	0.19*
College/Bachelor's/Master's	0.29	0.45***	0.29	0.41***	0.29	0.43***
Other	0.05	0.03	0.02	0.01	0.03	0.02
Observations	1,743		1,743		3,486	

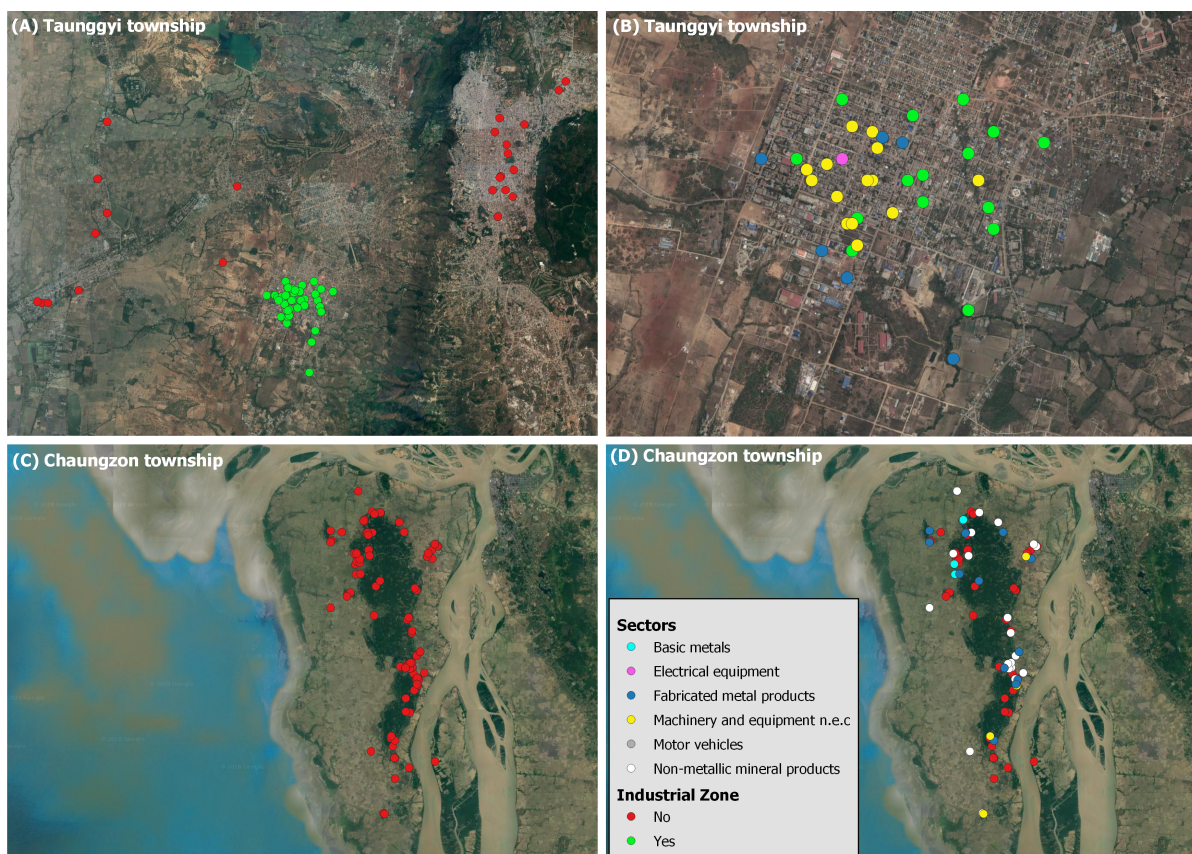
Notes: Value added, wages and, input values are adjusted for price differences between regions. The asterisk denote the significance levels from t-tests of 1, 5, and 10 percent (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Enterprises in industrial zones are on average three times larger in terms of employment than enterprises located elsewhere. They are less likely to be family owned than other enterprises. Industrial zone enterprises employ fewer skilled and women workers. They are also less likely to be owned or managed by women, but much more likely to have highly educated owners than other enterprises.

Even though industrial zone enterprises appear to be more efficient, they face higher costs of running business. The interviewees revealed during qualitative interviews that in addition to municipal fees compulsory for all enterprises, those located in an industrial zone need to pay fees to the industrial zone management. In addition, there are joint industrial zone contributions to local projects, such as employee clinics or infrastructure development. In some cases, being forced to operate in an industrial zone decreases the number of customers passing by, so owners feel the need to operate a more centrally located retail space as well.

3.4. Illustration of Industrial Zones and Innate Clusters

Figure 4: Industrial Zones and Natural Technical Clusters



Notes: Authors own calculations based on MEMS 2017 data.

Figure 4 shows examples of industrial zones and innate industry clustering. For example, Taunggyi Township, illustrated in panels A and B, has a clearly identifiable industrial zone with a high number of firms producing metals, electrical, machinery, and other similar products inside. These sectors are likely to have backward and forward linkages towards each other, as is the case in input-output-tables for other East and Southeast Asian countries such as China, Cambodia, Thailand, and Vietnam (OECD, 2017). Therefore, a high incidence of firms from these sectors in the same location may indicate the existence of a technical cluster. While Chaungzon Township, illustrated in panels C and D, does not have a formal industrial zone, we notice agglomeration of technical industries too, illustrating the formation of innate clusters without the existence of industrial zones.

4. EMPIRICAL SPECIFICATION

Given our interest in the contribution of industrial zones in facilitating agglomeration and increased human capital returns, we follow Konings and Vanormelingen (2015) and take as point of departure the following Cobb-Douglas production function: $Y_{jt} = A_{jt} \hat{L}_{jt}^{\beta_l} K_{jt}^{\beta_k}$, where Y_{jt} is value added of firm j in year t , $\hat{L}_{jt}^{\beta_l}$ is the aggregate effective labor input, $K_{jt}^{\beta_k}$ is physical capital, and A_{jt} is technical efficiency. In the spirit of Konings and Vanormelingen (2015), we assume that the labor aggregate can be written as:

$$\hat{L}_{jt} = \sum_{i=1}^{L_i} \exp[Z_{jt} + \beta_{IZ} IZ_{jt} + (1 + \rho IZ_{jt})(\beta_{HC} \bar{H}_{jt})] \quad (1)$$

where IZ_{jt} are observed (an industrial zone indicator) and Z_{jt} unobserved factors that increase the marginal product of labor, which we assume to be firm-specific. \bar{H}_{jt} represents the average of various available human capital variables of worker i (training, skill level, and women workers share). Contrary to Konings and Vanormelingen (2015), we distinguish between observed and unobserved firm factors and allow for heterogeneity (slope effects) in returns to human capital investments in addition to allowing for differences in intercepts (level effects) along the observed firm factors. In this paper, the observable firm dimensions in focus are measures of industrial clustering and agglomeration, which in the literature have been argued to influence returns to human capital investments. Taking logs, the production function can be written as:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_I \beta_{IZ} IZ_{jt} + \beta_I \beta_{HC} \bar{H}_{jt} + \beta_I \beta_{IZHC} IZ_{jt} \bar{H}_{jt} + \alpha_{jt} + \beta_I Z_{jt} + \varepsilon_{jt} \quad (2)$$

where $\beta_{IZHC} = \rho \beta_{HC}$. Letting $\omega_{jt} = \alpha_{jt} + \beta_I Z_{jt}$ denote labor-quality-adjusted firm productivity, and adding a set of observed firm characteristics X_{jt} which include the Ellison-Glaeser Index (Ellison and Glaeser, 1997) as a measure of agglomeration, as well as the interactions of location (p), sector (s),⁹ and year (t) fixed effects (σ_{pst}),¹⁰ yields:

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_{INZ} IZ_{jt} + \beta_H \bar{H}_{jt} + \beta_{IH} IZ_{jt} \bar{H}_{jt} + \beta_X X_{jt} + \omega_{jt} + \sigma_{pst} + \varepsilon_{jt} \quad (3)$$

where $\beta_{INZ} = \beta_I \beta_{IZ}$ measures the effect of industrial zones on value added in firms with average human capital and agglomeration levels compared to enterprises with the same average characteristics located outside industrial zones; $\beta_H = \beta_I \beta_{HC}$ measures the effect of a specific human capital variable on value added in enterprises located outside industrial zones, and $\beta_{IH} = \beta_I \beta_{IZHC}$ measures the interaction effect of industrial zone and human capital variables on value added. The marginal product of workers employed in industrial zones is defined as $\partial y_{it} / \partial IZ_{jt} = \beta_{INZ} / \beta_I$; the marginal product of specific human capital characteristics of workers outside industrial zones is defined as $\partial y_{it} / \partial \bar{H}_{jt} = \beta_H / \beta_I$, while the marginal product of specific human capital characteristics of the industrial zone workforce is obtained from $\partial y_{it} / \partial IZ_{jt} \bar{H}_{jt} = \beta_{IH} / \beta_I$, where H stands for a number of labor force characteristics: the share of skilled workers (denoted by S in the results shown in Tables 3 and 4), the share of trained workers (denoted by T), and the share of women workers (denoted by W). The impact of the Ellison-Glaeser Index (denoted by E) is estimated and interpreted analogously.

X_{jt} are firm-specific and respondent-specific (owner or manager) control variables, such as firm

⁹ We group several 2-digit industries to reduce large differences in the number of observations between specific industries. The groups are as follows: food, beverages, and tobacco; wood, paper and printing; coke, chemicals, rubber, and minerals; basic metals and fabricated metal products; electrical equipment, machinery, and motor vehicles; furniture and other manufacturing. As almost one-third of the sample are rice mills (MSIC code 1063), we keep them as a separate category, even though they are a part of the food industry.

¹⁰ The term σ_{pst} denotes multiple fixed effects, namely state/region, sector, year, the interaction between state/region and sector, the interaction between state/region and year, and the interaction between sector and year fixed effects.

size (measured as the natural logarithm of the number of permanent employees), real value of capital (in natural logarithms), firm age (the number of years since the enterprise was established), being a family-owned firm, as well as sex, education, and age of the owner or manager. The included controls for firm characteristics, such as firm size, firm age, and the value of inputs, enable us to account for differences in the production cost structure and the nature of technology between firms. Demographic characteristics of the enterprise owner are generally seen as important factors in determining enterprise growth and profitability (McPherson, 1996), so the estimations control for respondent's gender, education, and age. Education is measured as a series of dummy variables for the highest level of education of the respondent. Two controls for working conditions are included as well, including the share of workers with formal contracts, and the share of workers receiving in-kind wage benefits. The location and industry fixed effects account for common factors affecting all firms within industries and states/regions. Such factors could be, for example, differences in governance and institutions, as well as possible differences in industry-specific regulations that can affect enterprise productivity. Finally, ε_{jt} represents statistical noise.

Turning to the individual wage equation, we assume it can be written as:

$$W_{it} = \exp[\alpha_Z Z_{jt} + \alpha_{INZ} IZ_{jt} + (1 + \rho I_{jt})(\alpha_H H_{it})] \quad (4)$$

where α_H represent the wage premium of human capital variables of worker i (training, skill level, and sex), the coefficients α_Z and α_{INZ} measure how unobserved and observed firm-specific factors that alter the marginal product of workers are rewarded through wages. Equation 4 is a modification of Mincer's (1974) specification in which we use firm-level averages instead of individual worker-level wages and human capital characteristics. Instead of individual worker wages, we use the monthly wage of production workers as reported by firm owners or managers.¹¹ Similarly, instead of worker sex, we control for the share of women workers in the total workforce, and instead of worker's schooling, we control for the share of skilled workers. Again, contrary to Konings and Vanormelingen (2015), we assume that non-linear wage premiums of training may exist along observable firm dimensions (in our case industrial zone and agglomeration indicators). As such,

¹¹The aggregation is inspired by the approach in Hellerstein et al. (1999) and Van Biesebroeck (2011). Focusing on production workers is not a particularly important drawback as production workers comprise the largest share of the labor force in manufacturing firms.

$\exp(\alpha_Z Z_{jt})$ is the firm-specific average wage of an untrained worker in firm j outside an industrial zone ($IZ_{jt} = 0$). Summing over all workers i in firm j , applying a first-order Taylor expansion, and adding a set of observed firm characteristics, X_{jt} , as well as the interactions of location (p), sector (s), and year (t) fixed effects (ρ_{pst}), the average wage in the firm can be written as:

$$\bar{w}_{it} = w_0 + \alpha_Z Z_{jt} + \alpha_{INZ} IZ_{jt} + (1 + \rho IZ_{jt}) \alpha_H \bar{H}_{jt} + \beta_X X_{jt} + \sigma_{pst} + \varepsilon_{jt} \quad (5)$$

where \bar{H}_{jt} again represents average labor quality (skills, training level, share of women workers) in firm j and time t . Estimating wage Equation 5 could result in biased estimates of the wage premiums since training is likely to be correlated with unobserved labor quality Z_{jt} . To control for unobserved labor quality, we follow Konings and Vanormelingen (2015) and control for the estimate for ω_{jt} from the production function obtained from Equation 3. If the main component of ω_{jt} is labor quality, then adding estimated total factor productivity to the wage equation will result in the following equation to be estimated:

$$\bar{w}_{it} = w_0 + \hat{\omega}_{jt} + \alpha_{INZ} IZ_{jt} + \alpha_H \bar{H}_{jt} + \alpha_{IZH} IZ_{jt} \bar{H}_{jt} + \beta_X X_{jt} + \rho_{pst} + \varepsilon_{jt} \quad (6)$$

where $\alpha_{IZH} = \rho \alpha_H$ and Equation 6 will provide consistent estimates of the wage premium α_H .

Following Van Biesebroeck (2011), we estimate Equations 3 and 6 jointly using the seemingly unrelated regression (SUR) framework, which allows for correlation between the error terms in the two equations. In an additional set of estimates, we estimate Equations 3 and 6 separately. As the labor input choices are likely to be correlated with unobserved productivity, when estimating the production function (Equation 3), we apply the estimation procedure proposed by Akerberg et al. (2015) and assume that labor inputs (number of workers, training, and choice of skill composition) are determined before choosing material inputs. As highlighted in Konings and Vanormelingen (2015), the identifying assumption then becomes that the choice of labor inputs and labor force quality (e.g. the amount of training given) does not depend on the innovation in firm productivity.

All the reported standard errors are clustered at the township-industry level following the sampling stratification. Monetary variables used in the estimations, such as wages and value added, are corrected for spatial cost of living differences between regions. We used regional CPI deflators such that the values are deflated to 2018 prices for an average over the regions and states.

We probe the validity and robustness of our estimates with a couple of additional specifications. First, we vary the definition of industrial clusters and in addition to the Ellison-Glaeser Index, we identify clusters by the employment location quotient as proposed in Holmes and Stevens (2002) and by the enterprise density measured as the number of enterprises in a particular township. Second, we repeat the estimations without the informal firms as they tend to pay lower wages than formal firms (Meghir et al., 2015).

Some additional endogeneity bias may be present in estimations following equations (3) and (6). There could be unobserved characteristics that influence firm decisions whether to locate in industrial zones that are also correlated with overall firm performance. For example, the estimates would be biased upwards if more productive firms located in industrial zones also have owners or managers with higher unobserved ability. Forces that could bias the estimates downwards could also be at play if, for instance, less efficient firms locate in industrial zones to save taxes. General evidence on the effect of tax benefits on business location decisions is, however, ambiguous (Duranton et al., 2011; Rathelot and Sillard, 2008). A downward estimation bias may also arise if less efficient enterprises move to an industrial zone after a policy prescription. While township and industrial sector fixed effects control for unobserved geographic or sector-specific factors, they cannot fully eliminate biases related to which enterprises decide to locate in an industrial zone for policy reasons. Strict causality is therefore difficult to ascertain in this setup, so the results are best understood as correlations.

5. RESULTS

Table 3 shows the results of estimating the production function (Equation 3) and the wage equation (Equation 6). Odd-numbered columns show the estimates of the productivity premium and even-numbered columns show the estimates of the wage premium in industrial zones. Columns 1-4 show SUR estimates, while Columns 5-8 show GMM estimates based on the procedure described in Akerberg et al. (2015). In Columns 1 and 5, the natural logarithm of value added is regressed on the industrial zone dummy interacted with the share of skilled and trained workers in a firm. In Columns 2 and 6, this is repeated with the natural logarithm of the average wage of production workers as a

dependent variable. The remaining columns contain an interaction of the industrial zone dummy with the Ellison-Glaeser Index. All productivity estimations include additional controls, such as the natural logarithm of capital and labor, firm age, a family firm dummy, share of women workers, share of workers with formal labor contracts, share of workers receiving fringe benefits, and a range of owner characteristics, such as owner sex (woman), education level (from no to tertiary education) and age. All estimations also include year, sector, and location fixed effects, as well as their interactions.

Table 3: Skills-Driven Productivity and Wage Premiums in Industrial Zones

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS				ACF			
	Value added	Wage	Value added	Wage	Value added	Wage	Value added	Wage
IZ ($\beta_{INZ}, \alpha_{INZ}$)	0.069*** (0.022)	0.175*** (0.026)	0.070*** (0.022)	0.178*** (0.026)	0.060*** (0.007)	0.156*** (0.026)	0.041*** (0.015)	0.159*** (0.027)
IZ x Skill ($\beta_{IZS}, \alpha_{IZS}$)	-0.060 (0.063)	-0.159** (0.069)	-0.058 (0.063)	-0.169** (0.068)	-0.091*** (0.007)	-0.141* (0.073)	-0.080*** (0.008)	-0.151** (0.072)
IZ x Train ($\beta_{IZT}, \alpha_{IZT}$)	0.164 (0.156)	-0.244* (0.138)	0.159 (0.153)	-0.237* (0.138)	0.148*** (0.007)	-0.257* (0.153)	0.144*** (0.013)	-0.249 (0.153)
IZ x EGI ($\beta_{IZE}, \alpha_{IZE}$)			0.037 (0.085)	0.177 (0.151)			0.017 (0.011)	0.163 (0.164)
Skilled (β_S, α_S)	0.048* (0.025)	0.012 (0.034)	0.047* (0.025)	0.016 (0.034)	0.028*** (0.005)	0.016 (0.034)	0.025*** (0.006)	0.021 (0.034)
Training (β_T, α_T)	-0.086 (0.071)	0.306*** (0.109)	-0.086 (0.071)	0.301*** (0.106)	-0.115*** (0.007)	0.251** (0.114)	-0.107*** (0.008)	0.247** (0.111)
EGI (β_E, α_E)			0.054 (0.098)	-0.272*** (0.091)			0.040*** (0.014)	-0.259*** (0.093)
Labor (β_l)	0.329*** (0.013)		0.328*** (0.013)		0.326*** (0.006)		0.322*** (0.013)	
Capital (β_k)	0.703*** (0.011)		0.703*** (0.011)		0.709*** (0.008)		0.710*** (0.010)	
TFP ($\hat{\omega}$)		0.213*** (0.027)		0.211*** (0.027)		0.213*** (0.028)		0.211*** (0.028)
State/Region x Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Region x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes		Yes		Yes		Yes	
Observations	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486

Notes: β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's and α 's indicate an interaction, where *IZ* denotes industrial zone, *S* denotes skills, *T* denotes training, and *E* denotes the Ellison-Glaeser Index. $\beta_{INZ} = \beta_l \beta_{IZ}$ as explained in Equations 2 and 3. α_{INZ} denotes the industrial zone coefficient from Equation 6. Control variables include labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. ACF denotes Akerberg-Caves-Frazer model. Columns 5-8 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level are in parentheses. The asterisk denote the significance levels from t-tests of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

All specifications show a positive association between industrial zones and value added, and a positive association between industrial zones and worker wages.¹² Column 1 shows about 7 percent higher value added and Column 2 shows 18 percent higher wages in industrial zone enterprises with average skills and training compared to enterprises with average skills and training located outside industrial zones. As shown in Columns 3 and 4, controlling for the agglomeration level in industrial zones does not affect the size of the industrial zone coefficient very much. Controlling for endogeneity of inputs and human capital characteristics decreases the coefficient for the industrial zone indicator slightly. The coefficients in Columns 5 and 6 show 6 percent higher value added and 16 percent higher wages in industrial zone enterprises with a labor composition having average skills and training compared to their counterparts located outside industrial zones. Industrial zone enterprises that can increase the share of skilled and trained workers by 10 percentage points can achieve 1.2 percent higher value added than enterprises outside industrial zones with the same worker composition ($\beta_{INZ} + \beta_{IZS} + \beta_{IZT}$), but these gains come with a 2.4 percent decrease in worker wages ($\alpha_{INZ} + \alpha_{IZS} + \alpha_{IZT}$).

Controlling for agglomeration, we obtain that industrial zone location associates with 4.1 percent higher value added (Column 7), which implies that the marginal product of workers in industrial zones is 12.8 percent ($\beta_{IZ} = 0.041/0.332$) higher compared to enterprises outside industrial zones with the same worker composition (shown in Table A4). The coefficients in Column 8 show 16 percent higher wages for enterprises with average skills, training, and agglomeration level compared to enterprises with the same average characteristics located outside industrial zones. If the values of skills, training, and agglomeration in industrial zones were to increase by 10 percentage points, enterprises could achieve 1.2 percent higher productivity ($\beta_{INZ} + \beta_{IZS} + \beta_{IZT} + \beta_{IZE}$) with 0.8 percent lower worker wages ($\alpha_{INZ} + \alpha_{IZS} + \alpha_{IZT} + \alpha_{IZE}$).

We explore further how the returns to labor and capital change with an increase in the share of skilled workers. Results in Column 7 show that increasing the share of skilled workers by 10 percentage points in enterprises outside industrial zones is associated with 0.3 percent higher value added (β_S). We do not find any significant wage premium associated with skills outside industrial zones when controlling for productivity differences between firms, as shown in Column 8. Inside

¹²Tables A2 and A3 show how the coefficients measuring the relationships between key independent and outcome variables change with the model specification.

industrial zones, the results in Columns 7 and 8 show productivity and wage losses of 0.6 percent ($\beta_{IZS} + \beta_S$) and 1.3 percent ($\alpha_{IZS} + \alpha_S$), respectively, associated with a 10 percentage point increase in the share of skilled workers in industrial zones, as opposed to the same increase in the share of skilled workers in enterprises outside industrial zones. A 10 percentage point increase in the share of skilled workers above the industrial zone average is associated with 0.4 percent lower productivity ($\beta_{IZS} + \beta_{INZ}$). As shown in Column 8, no statistically significant differences in the wage level can be attributed to increasing the share of skilled workers in industrial zones above the average skill level ($\alpha_{IZS} + \alpha_{INZ}$). There is a 0.8 percent penalty to productivity and a 1.5 percent penalty to wages associated with a 10 percentage point increase in the share of skilled industrial zone workers compared to keeping the labor skills at the average level outside industrial zones (β_{IZS} and α_{IZS}).

The returns to worker training depend very much on firm location. Training outside industrial zones is negatively related to firm productivity, and positively related to worker remuneration, suggesting that workers are overcompensated for their training efforts as compared to the productivity gains training provides. This is expected if the training contents are largely general in nature, if benefits are heterogeneous across workers, or if workers hold private information about the actual gains. This could shift the bargaining power (increase the outside option) towards workers relative to employers. In industrial zones, we get the opposite picture: positive productivity and insignificant wage gains associated with a higher share of trained workers. A 10 percentage point increase in the share of trained workers in industrial zone enterprises is associated with 1.4 percent higher productivity, compared to keeping the share of trained workers in enterprises outside industrial zones at the average level (β_{IZT}). The marginal product of a trained worker in an industrial zone enterprise is 45 percent ($\beta_{IT} = \beta_{IZT}/\beta_I = 0.144/0.332$) higher compared to an untrained worker in enterprises operating outside industrial zones (shown in Table A4). In contrast to productivity estimates, increasing the share of trained workers by 10 percentage points in industrial zones is associated with 2.6 percent lower worker wages (Column 6), but the coefficient is not precisely determined when controlling for agglomeration effects (Column 8). When compared to the average training level in industrial zone enterprises, a 10 percentage point increase in the share of trained workers is associated with 1.9 percent higher value added ($\beta_{IZT} + \beta_{INZ}$) and no significant wage changes ($\alpha_{IZT} + \alpha_{INZ}$). The results imply that, on average, the marginal product of a trained

worker in industrial zones is 57 percent ($\beta_{IT} + \beta_{IZ} = 0.144/0.332 + 0.041/0.332 = 0.446 + 0.128$) higher compared to untrained workers in industrial zones (see Table A4). Increasing the share of trained workers in industrial zones by 10 percentage points is associated with 0.4 percent higher enterprise productivity compared to the same increase in the share of trained workers outside industrial zones ($\beta_{IZT} + \beta_T$). The marginal product of a trained worker in industrial zones is 11 percent ($\beta_{IT} + \beta_T/\beta_l = 0.446 - 0.107/0.322$) higher than the marginal product of a trained worker in enterprises operating outside industrial zones (Table A4).

Next, we focus on the influence of competitive pressure, measured by the Ellison-Glaeser Index of industrial agglomeration, on enterprise performance in industrial zones. The index is positively associated with value added and negatively associated with worker wage, indicating that agglomeration may increase enterprise productivity while exerting a downward pressure on worker wages. This is consistent with earlier findings that agglomeration externalities occur through both productivity and wage effects. For example, Fafchamps and Hamine (2017) find that firms pay lower wages when their sector of activity employs a lot of workers. When contrasting the performance of industrial zones and innately occurring industrial clusters measured by the Ellison-Glaeser Index, our results indicate that productivity gains associated with industrial zones persist after controlling for the level of innately occurring industrial agglomeration in particular townships and industries. Moreover, worker wages in industrial zones can increase as a function of competitive pressure. The results in Columns 7 and 8 show that a 10 percentage point increase in the level of agglomeration above the industrial zone average is associated with 0.6 percent higher enterprise productivity ($\beta_{IZE} + \beta_{INZ}$) and 3.2 percent higher worker wages ($\alpha_{IZE} + \alpha_{INZ}$). Increasing the level of agglomeration in industrial zones by 10 percentage points is associated with 0.6 percent higher enterprise productivity compared to the same increase in the level of agglomeration outside industrial zones ($\beta_{IZE} + \beta_E$).

Table 4 adds an interaction term between the industrial zone dummy and the share of women workers in the enterprise. The estimates imply that productivity gains associated with industrial zones remain after additionally controlling for the share of women workers. The coefficients in Columns 3 and 4 show 3.1 percent higher value added (β_{INZ}) and 16 percent higher wages (α_{INZ}) in industrial zone enterprises with average skills, training, agglomeration, and share of women workers compared to enterprises with the same average characteristics located outside industrial zones.

Table 4: Productivity and Wage Gains from Women Working in Industrial Zones

	(1)	(2)	(3)	(4)
	SUR		ACF	
	Value added	Wage	Value added	Wage
IZ ($\beta_{INZ}, \alpha_{INZ}$)	0.061*** (0.021)	0.174*** (0.027)	0.031** (0.013)	0.155*** (0.022)
IZ x Skill ($\beta_{IZS}, \alpha_{IZS}$)	-0.057 (0.060)	-0.181*** (0.067)	-0.088*** (0.007)	-0.160*** (0.056)
IZ x Train ($\beta_{IZT}, \alpha_{IZT}$)	0.192 (0.151)	-0.196 (0.133)	0.163*** (0.006)	-0.208 (0.139)
IZ x EGI ($\beta_{IZE}, \alpha_{IZE}$)	0.055 (0.093)	0.100 (0.152)	0.045*** (0.009)	0.101 (0.111)
IZ x Women ($\beta_{IZW}, \alpha_{IZW}$)	-0.056 (0.046)	0.165*** (0.046)	-0.063*** (0.011)	0.128** (0.060)
Skilled (β_S, α_S)	0.055** (0.025)	0.028 (0.033)	0.022*** (0.006)	0.032 (0.025)
Training (β_T, α_T)	-0.104 (0.074)	0.277*** (0.102)	-0.115*** (0.012)	0.224** (0.097)
EGI (β_E, α_E)	0.099 (0.099)	-0.151* (0.090)	0.077*** (0.008)	-0.146* (0.078)
Women (β_{WS}, α_{WS})	-0.102*** (0.024)	-0.213*** (0.034)	-0.124*** (0.011)	-0.206*** (0.025)
Labor (β_l)	0.336*** (0.013)		0.340*** (0.015)	
Capital (β_k)	0.701*** (0.011)		0.709*** (0.009)	
TFP ($\hat{\omega}$)		0.205*** (0.029)		0.205*** (0.021)
State/Region x Sector FE	Yes	Yes	Yes	Yes
State/Region x Year FE	Yes	Yes	Yes	Yes
Sector x Year FE	Yes	Yes	Yes	Yes
State/Region FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes		Yes	
Observations	3,486	3,486	3,486	3,486

Notes: β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's and α 's indicate an interaction, where *IZ* denotes industrial zone, *S* denotes skills, *T* denotes training, *E* denotes the Ellison-Glaeser Index, and *WS* denotes the share of women workers. $\beta_{INZ} = \beta_l \beta_{IZ}$ as explained in Equations 2 and 3, and $\beta_{WS} = \beta_l \beta_W$. α_{INZ} denotes the industrial zone coefficient from Equation 6. Control variables include labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. SUR denotes seemingly unrelated regression. ACF denotes Akerberg-Caves-Frazer model. Columns 3 and 4 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level are in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Increasing the share of skilled workers, training, agglomeration, and women workers in industrial zones by 10 percentage points is associated with 0.9 percent higher value added ($\beta_{INZ} + \beta_{IZS} + \beta_{IZT} + \beta_{IZE} + \beta_{IZW}$) and no significant effect on worker wages ($\alpha_{INZ} + \alpha_{IZS} + \alpha_{IZT} + \alpha_{IZE} + \alpha_{IZW}$). The negative association between industrial zones and wages shown in Table 3 no longer holds when controlling for the women workers share, indicating that women workers in industrial zones potentially receive higher wages than women workers elsewhere. This is confirmed by a significantly positive coefficient $\alpha_{INZ} + \alpha_{IZW}$.

The results in Table 4 also show a negative association between productivity and wages and the share of women workers in enterprises outside industrial zones (β_{WS} and α_{WS}). However, the difference between productivity and wage premiums is not statistically different from zero, as indicated by the low χ^2 test value (see Table A5). In industrial zones, we find that the wage premium exceeds the productivity premium in enterprises with an above-average share of women workers (compare β_{IZW} and α_{IZW}). Increasing the share of women workers by 10 percentage points is associated with 0.6 percent lower value added (β_{IZW}) and 1.3 percent higher wage (α_{IZW}), compared to enterprises outside industrial zones with the average share of women workers. Compared to the average women worker share in industrial zone enterprises, a 10 percentage point increase in the share of women workers is associated with 0.3 percent lower value added ($\beta_{IZW} + \beta_{INZ}$) and 2.8 percent higher wage ($\alpha_{IZW} + \alpha_{INZ}$). The result implies that, on average, the marginal product of a woman worker in an enterprise with an average share of women workers in industrial zones is 9.4 percent ($\beta_{IW} + \beta_{IZ} = -0.186 + 0.092$) lower than the marginal product of a woman worker in an enterprise with a below-average share of women workers in industrial zones (see Table A5). Increasing the share of women workers in industrial zones by 10 percentage points is associated with 1.9 percent lower value added ($\beta_{IZW} + \beta_{WS}$) and no changes in worker wages compared to the same increase in the share of women workers outside industrial zones. The marginal product of a woman worker in an enterprise with an average share of women workers in industrial zones is 5.5 percent ($\beta_{IW} + \beta_W = -0.186 - 0.363$) lower than the marginal product of a woman worker in an enterprise with an average share of women workers operating outside industrial zones (shown in Table A5).

6. ROBUSTNESS CHECKS AND DISCUSSION

Our results show that industrial zone enterprises with an average composition of skills, training, agglomeration, and women workers have higher value added than their counterparts located elsewhere. We also find a significant gap between productivity and wage premiums in industrial zones. Productivity gains are associated in particular with competition and hiring of trained workers, while no statistically significant wage gain can be attributed to enterprises in industrial zones with the average worker composition facing the average level of agglomeration. This result indicates the presence of imperfections in local labor markets whereby employers enjoy rents and face costly replacement of workers, which further suggests that workers may gain from changing employment. This is consistent with our qualitative interviews with enterprise owners, who mentioned high labor turnover as one of the main challenges they face in running their business. Overall, our results indicate that enterprises in industrial zones need to carefully consider the quality of their labor input to achieve productivity gains. Increasing the share of women and skilled workers to above the average level could result in productivity losses. In what follows, we probe the robustness and discuss several important aspects of our results.

First, the results on productivity and wage premiums in industrial zones account for key enterprise, owner, and labor characteristics, location, industry, and time effects, as well as their interactions. Consistent with the Akerberg et al. (2015) estimation procedure, our identifying assumption is the timing of decisions related to production inputs (we assume that the choice of labor inputs and labor quality does not depend on innovation in firm productivity). However, this assumption does not preclude all types of estimation bias. It could be that some unobservable factors make firms both more productive and more likely to choose to start operating in an industrial zone, so our results could also be explained by better management practices or by better labor-technology matches in firms located in industrial zones. The level of upward bias in the estimated coefficient is probably not too large, as one part of the firms has located in industrial zones after a policy prescription, but we also cannot exclude that one part of the firms could have chosen to locate in an industrial zone due to adverse selection.

Second, as shown in Table A6, the results are sensitive to the applied measure of industrial agglomeration. The main difference is that the enterprise location quotient and enterprise density

indicate a negative association between productivity and agglomeration in industrial zones, whereas the Ellison-Glaeser Index indicates a positive interaction in Table 4 (β_{IZE} and $\beta_{IZE} + \beta_E$). However, the additive term $\beta_{IZE} + \beta_{INZ}$ in Table A6 is consistent with the results in Table 4, indicating a positive association between industrial agglomeration and wages, and between agglomeration and value added in industrial zones when they are compared to outcomes in enterprises outside industrial zones. These differences in the results could be driven by the way in which different agglomeration measures are constructed. Whereas the location quotient focuses on industry- and location-level employment, the Ellison-Glaeser Index also accounts for the share of individual enterprises in industry employment. Both the enterprise location quotient and the enterprise density are determined less precisely in our estimations than the Ellison-Glaeser Index, indicating the importance of accounting for the contribution of individual enterprises to industrial concentration.

Third, as wages in formal firms tend to be higher than wages in informal firms (Meghir et al., 2015), we repeat the main estimations after excluding informal firms from the sample. The results shown in Table A8 in general confirm our main findings. The differences in coefficient size are as expected: we obtain higher coefficients for the industrial zone dummy in the production function estimates and lower coefficients in the wage equation. One particular difference between the estimates based on the full sample and those based on the sample without informal firms is in the results for women workers. Increasing the share of women workers in industrial zone enterprises no longer has a significantly negative association with value added, while increasing the share of women workers by 10 percentage points is now associated with 1.5 percent higher wages, which is larger than the originally detected 1.3 percent. Even though informal enterprises have a higher share of women workers than formal enterprises (41 percent compared to 24 percent, respectively), our results indicate that women in Myanmar have obtained better employment conditions in formal firms.

Fourth, our results are consistent with Wang (2013) who finds that the benefits of China's SEZ policies extend to higher enterprise productivity. But, in contrast to Wang (2013), we do not find that the benefits of industrial zones extend to higher workers' earnings, except for women. This is different from the rest of the country (Hansen et al., 2020), but aligned with sporadic evidence of higher wages for women than men noted in the early stages of the development of the manufacturing sector in Myanmar (Francisco, 2007). We speculate that this finding combined with the negative

wage-productivity gap related to a higher share of women workers signals a shortage of women labor in industrial zones. As industrial zones are not equally distributed across the country – the majority is located near major cities, such as Yangon and Mandalay, with high local costs of living and limited transport options – especially women from remote areas can be constrained in seeking employment in industrial zones. Some women may also be reluctant to migrate for employment for fear that this could have negative welfare implications for other household members (Malapit and Quisumbing, 2015; Mueller et al., 2020). Consequently, an additional expansion of manufacturing jobs for women could be achieved by establishment of industrial zones outside the largest cities.

7. CONCLUSION

Many low- and lower-middle income countries have opted for extensive expansion of the manufacturing sector by establishing SEZs of industrial activity. The underlying premise is that locating firms in designated areas can decrease production costs, create market linkages, attract foreign investment, and improve access to labor and knowledge. In this paper, we addressed whether the potential benefits of industrial zones translate into improved enterprise and worker outcomes.

Our paper is among the first to analyze performance (in terms of both labor productivity and worker wages) of private manufacturing enterprises located in industrial zones in Myanmar. We focused on the contribution of industrial zones in facilitating agglomeration and labor quality returns, measured as interactions between the industrial zone indicator and the shares of skilled, trained, and women workers, and the Ellison-Glaeser Index. Both agglomeration effects (Ellison and Glaeser, 1997; Fafchamps and Hamine, 2017; Greenaway and Kneller, 2008; Greenstone et al., 2010; Holmes and Stevens, 2002) and SEZ/industrial zones (Alder et al., 2016; Alkon, 2018; Bräutigam and Tang, 2014; Lu et al., 2019; Wang, 2013) have been studied extensively in the literature. However, no studies have examined the different agglomeration profiles of industrial zones and innately occurring industrial clusters, or have considered the direct contribution of labor quality in affecting the distribution of gains to capital and labor.

Our analysis made use of the data from the first two representative rounds of the Myanmar Enterprise Monitoring Survey (MEMS) conducted in 2017 and 2019, on the one hand, and qualitative

interviews with enterprise owners, workers, and key institutional actors, on the other. Thanks to the panel dimension of the data, we can account for endogeneity of human capital variables using the control function approach when estimating the production function. The production function estimates obtained in this way were then used as measures of unobserved worker characteristics in the wage equation to obtain consistent estimates of the impact of the industrial zone location, agglomeration, and human capital variables on production worker wages. While our analytical approach is similar to that of Konings and Vanormelingen (2015), we allow for slope effects in returns to human capital investments and for level effects in the observed firm factors.

Our results show that, overall, the value added increase associated with industrial zones is larger than the wage increase in enterprises with average composition of skills, training, and women workers facing the average level of industrial agglomeration. We note that workers in industrial zones are more productive than workers in enterprises located elsewhere, though this is not entirely reflected in their wages (except in case of women workers). Further, we find that productivity effects related to agglomeration and hiring of trained workers are larger in industrial zones than elsewhere. When contrasting the performance of industrial zones and naturally occurring industrial clusters, we obtain results in favor of industrial zones, indicating that the benefits from industrial zones extend beyond agglomeration. We also uncovered that increasing the share of women and skilled workers to above the average level could result in productivity losses, implying that labor quality and composition determine the gains from industrial agglomeration. While increasing the share of women workers can lead to productivity losses, our estimates show that wages increase with the share of women workers. The negative wage-productivity gap related to a higher share of women workers indicates that there may be a shortage of female labor in industrial zones due to unobserved factors constraining entry (or discrimination).

While earlier research brings out the benefits to firm performance from industrial agglomeration in developing countries undergoing political and economic transformation (see, e.g., Alkon, 2018; Chhair and Newman, 2014; Hu et al., 2015; Wang, 2013), there is in general limited evidence on both the enterprise and worker outcomes in industrial zones in developing countries. An exception is Wang (2013), who uses municipal data to investigate the benefits of the SEZs policy program in China. Similar to our study, the results from China document an increase in firm productivity after the construction of SEZs. This is reassuring given the wide contextual differences between our

case and China. In addition, our results are consistent with Tanaka's (2019) findings that the general opening of the Myanmar economy to foreign trade until the recent coup d'état was beneficial for working conditions in manufacturing firms.

While workers capture fewer efficiency gains from working in industrial zones, our results are overall supportive of an industrial policy promoting the establishment of industrial zones. They also call attention to a need for careful consideration of the role of efficient management of human resources in assuring enterprise growth in industrial zones.

REFERENCES

- Abe, M., & Dutta, M. K. (2014).** A New Policy Framework for Myanmar's SME Development. ARTNeT Working Paper Series No. 142. Bangkok: ARTNeT.
- Akerberg, D. A., Caves, K., & Frazer, G. (2015).** Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6), 2411–2451.
- Alder, S., Shao, L., & Zilibotti, F. (2016).** Economic Reforms and Industrial Policy in a Panel of Chinese Cities. *Journal of Economic Growth*, 21(4), 305–349.
- Alkon, M. (2018).** Do Special Economic Zones Induce Developmental Spillovers? Evidence from India's States. *World Development*, 107, 396–409.
- Alonso-Villar, O., Chamorro-Rivas, J.-M., & González-Cerdeira, X. (2004).** Agglomeration Economies in Manufacturing Industries: The Case of Spain. *Applied Economics*, 36(18), 2103–2116.
- Audretsch, D. B., & Feldman, M. P. (1996).** Innovative Clusters and the Industry Life Cycle. *Review of Industrial Organization*, 11(2), 253–273.
- Baptista, R. (2001).** Geographical Clusters and Innovation Diffusion. *Technological Forecasting and Social Change*, 66(1), 31–46.
- Berkel, H. M., Cardona, M., Hansen, H., Rand, J., Castro Rodriguez, P., Trifković, N., de Witte, E., Zille, H., Latt, K. S., & Tarp, F. (2018).** Myanmar Micro, Small, and Medium Enterprise Survey 2017: Descriptive Report. Central Statistical Organization Myanmar and UNU-WIDER.
- Bertinelli, L., & Decrop, J. (2005).** Geographical Agglomeration: Ellison and Glaeser's Index Applied to the Case of Belgian Manufacturing Industry. *Regional Studies*, 39(5), 567–583.
- Billings, S. B., & Johnson, E. B. (2016).** Agglomeration Within an Urban Area. *Journal of Urban Economics*, 91, 13–25.

-
- Braunerhjelm, P., & Borgman, B. (2004).** Geographical Concentration, Entrepreneurship and Regional Growth: Evidence from Regional Data in Sweden, 1975-99. *Regional Studies*, 38(8), 929–947.
- Bräutigam, D., & Tang, X. (2014).** “Going Global in Groups”: Structural Transformation and China’s Special Economic Zones Overseas. *World Development*, 63, 78–91.
- Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2007).** Urban Density and the Rate of Invention. *Journal of Urban Economics*, 61(3), 389–419.
- Cassey, A. J., & Smith, B. O. (2014).** Simulating Confidence for the Ellison–Glaeser Index. *Journal of Urban Economics*, 81, 85–103.
- Chhair, S., & Newman, C. (2014).** Clustering, Competition, and Spillover Effects: Evidence from Cambodia. WIDER Working Paper Series No. 065. World Institute for Development Economic Research (UNU-WIDER).
- Ciccone, A., & Hall, R. E. (1996).** Productivity and the Density of Economic Activity. *American Economic Review*, 86(1), 54–70.
- Cirera, X., & Lakshman, R. W. D. (2017).** The Impact of Export Processing Zones on Employment, Wages and Labour Conditions in Developing Countries: Systematic Review. *Journal of Development Effectiveness*, 9(3), 344–360.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012).** The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6), 2543–2594.
- CSO. (2017).** Myanmar Statistical Yearbook 2017. Central Statistical Organization. http://www.mmsis.gov.mm/sub_menu/statistics/fileDb.jsp. Accessed 29.03.2022.
- Deichmann, U., Lall, S. V., Redding, S. J., & Venables, A. J. (2008).** Industrial Location in Developing Countries. *World Bank Research Observer*, 23(2), 219–246.

-
- Delmas, M. A., & Pekovic, S. (2013).** Environmental Standards and Labor Productivity: Understanding the Mechanisms that Sustain Sustainability. *Journal of Organizational Behavior*, 34(2), 230–252.
- Drucker, J., & Feser, E. (2012).** Regional Industrial Structure and Agglomeration Economies: An Analysis of Productivity in Three Manufacturing Industries. *Regional Science and Urban Economics*, 42(1–2), 1–14.
- Duranton, G., Gobillon, L., & Overman, H. G. (2011).** Assessing the Effects of Local Taxation using Microgeographic Data. *The Economic Journal*, 121(555), 1017–1046.
- Duranton, G., & Overman, H. G. (2005).** Testing for Localization Using Micro-Geographic Data. *The Review of Economic Studies*, 72(4), 1077–1106.
- Ellison, G., & Glaeser, E. L. (1997).** Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105(5), 889–927.
- Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010).** What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(3), 1195–1213.
- Fafchamps, M., & Hamine, S. E. (2017).** Firm productivity, wages, and agglomeration externalities. *Research in Economics*, 71(2), 291–305.
- Fan, C. C., & Scott, A. J. (2009).** Industrial Agglomeration and Development: A Survey of Spatial Economic Issues in East Asia and a Statistical Analysis of Chinese Regions. *Economic Geography*, 79(3), 295–319.
- Feldman, M. P., & Audretsch, D. B. (1999).** Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition. *European Economic Review*, 43(2), 409–429.
- Fleisher, B., Hu, D., McGuire, W., & Zhang, X. (2010).** The Evolution of an Industrial Cluster in China. *China Economic Review*, 21(3), 456–469.
- Florence, P. S. (1939).** Report on the Location of Industry. London: Political and Economic Planning.

- Francisco, J. S. (2007).** Gender Inequality, Poverty and Human Development in South East Asia. *Development*, 50(2), 103–114.
- Frick, S. A., & Rodríguez-Pose, A. (2019).** Are Special Economic Zones in Emerging Countries a Catalyst for the Growth of Surrounding Areas? *Transnational Corporations Journal*, 26(2), 75-94.
- Ge, Y. (2009).** Globalization and Industry Agglomeration in China. *World Development*, 37(3), 550–559.
- Gibbs, R., & Bernat, G. A. JR. (1997).** Rural Industry Clusters Raise Local Earnings. *Rural Development Perspectives*, 12(3), 18–25.
- Giuliano, G., Kang, S., & Yuan, Q. (2019).** Agglomeration Economies and Evolving Urban Form. *Annals of Regional Science*, 63(3), 377–398.
- Greenaway, D., & Kneller, R. (2008).** Exporting, Productivity and Agglomeration. *European Economic Review*, 52(5), 919–939.
- Greenstone, M., Hornbeck, R., & Moretti, E. (2010).** Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3), 536–598.
- Guimarães, P., Figueiredo, O., & Woodward, D. (2009).** Dartboard Tests for the Location Quotient. *Regional Science and Urban Economics*, 39(3), 360–364.
- Hansen, H., Rand, J., & Win, N. W. (2020).** The Gender Wage Gap in Myanmar: Adding Insult to Injury? DERG Working Paper Series No. 20–05. University of Copenhagen. Department of Economics. Development Economics Research Group (DERG).
- Hellerstein, J. K., Neumark, D., & Troske, K. R. (1999).** Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations. *Journal of Labor Economics*, 17(3), 409–446.
- Henderson, J. V. (2003).** Marshall’s Scale Economies. *Journal of Urban Economics*, 53(1), 1–28.

-
- Holmes, T. J. (1999).** Localization of Industry and Vertical Disintegration. *Review of Economics and Statistics*, 81(2), 314–325.
- Holmes, T. J., & Stevens, J. J. (2002).** Geographic Concentration and Establishment Scale. *Review of Economics and Statistics*, 84(4), 682–690.
- Howard, E., Newman, C., & Tarp, F. (2016).** Measuring Industry Coagglomeration and Identifying the Driving Forces. *Journal of Economic Geography*, 16(5), 1055–1078.
- Hu, C., Xu, Z., & Yashiro, N. (2015).** Agglomeration and Productivity in China: Firm Level Evidence. *China Economic Review*, 33, 50–66.
- Huang, Z., Zhang, X., & Zhu, Y. (2008).** The Role of Clustering in Rural Industrialization: A Case Study of the Footwear Industry in Wenzhou. *China Economic Review*, 19(3), 409–420.
- ILO. (2017).** ILO Guide to Myanmar Labour Law. International Labour Organization. https://www.ilo.org/yangon/publications/WCMS_577563/lang-en/index.htm. Accessed 29.03.2022-
- Ito, B., Xu, Z., & Yashiro, N. (2015).** Does agglomeration promote internationalization of Chinese firms? *China Economic Review*, 34, 109–121.
- Jacobs, J. (1969).** *The Economy of Cities* (First Pr. edition). Random House.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993).** Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, 108(3), 577–598.
- Kelley, M. R., & Helper, S. (1999).** Firm Size And Capabilities, Regional Agglomeration, And The Adoption Of New Technology. *Economics of Innovation and New Technology*, 8(1–2), 79–103.
- Konings, J., & Vanormelingen, S. (2015).** The Impact of Training on Productivity and Wages: Firm-Level Evidence. *Review of Economics and Statistics*, 97(2), 485–497.
- Krugman, P. R. (1991).** Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3), 483–499.
- Krugman, P. R. (1993).** *Geography and Trade*. MIT Press.

-
- Lall, S. V., Shalizi, Z., & Deichmann, U. (2004).** Agglomeration Economies and Productivity in Indian Industry. *Journal of Development Economics*, 73(2), 643–673.
- Li, D., Lu, Y., & Wu, M. (2012).** Industrial Agglomeration and Firm Size: Evidence from China. *Regional Science and Urban Economics*, 42(1), 135–143.
- Lin, H.-L., Li, H.-Y., & Yang, C.-H. (2011).** Agglomeration and Productivity: Firm-Level Evidence from China's Textile Industry. *China Economic Review*, 22(3), 313–329.
- Linn, S. N., Ni Lin, M. Z., & Weidert, D. (2014).** Pilot Assessment of Industrial Zones in Myanmar. Centre for Economic and Social Development (MDRI-CESD) and Deutsche Gesellschaft fuer Internationale Zusammenarbeit (GIZ).
- Lu, Y., Wang, J., & Zhu, L. (2019).** Place-Based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program. *American Economic Journal: Economic Policy*, 11(3), 325–360.
- Mack, R. S., & Jacobsen, D. S. (1996).** Core Periphery Analysis of the European Union: A Location Quotient Approach. *Journal of Regional Analysis & Policy*, 26(1), 3–21.
- Malapit, H. J. L., & Quisumbing, A. R. (2015).** What Dimensions of Women's Empowerment in Agriculture Matter for Nutrition in Ghana? *Food Policy*, 52, 54–63.
- Marshall, A. (1890).** *Principles of Economics*. MacMillan.
- McPherson, M. A. (1996).** Growth of Micro and Small Enterprises in Southern Africa. *Journal of Development Economics*, 48(2), 253–277.
- Meghir, C., Narita, R., & Robin, J.-M. (2015).** Wages and Informality in Developing Countries. *The American Economic Review*, 105(4), 1509–1546.
- Min, A., & Kudo, T. (2013).** New Government's Initiatives for Industrial Development in Myanmar. In H. Lim & Y. Yamada (eds.) *Economic Reforms in Myanmar: Pathways and Prospects* (pp. 39–84). Bangkok Research Center.
- Mincer, J. (1974).** *Schooling, Experience, and Earnings* [NBER Books]. Cambridge: National Bureau of Economic Research.

-
- Mueller, V., Schmidt, E., & Kirkleeng, D. (2020).** Structural Change and Women's Employment Potential in Myanmar. *International Regional Science Review*, 43(5), 450–476.
- Myint, M. M., & Rasiah, R. (2012).** Foreign Capital and Garment Export from Myanmar: Implications for the Labour Process. *Institutions and Economies*, 4(3), 151–172.
- Newman, C., & Page, J. (2017).** Industrial Clusters: The Case for Special Economic Zones in Africa. WIDER Working Paper Series No. 015. World Institute for Development Economic Research (UNU-WIDER).
- Newman, C., Page, J., Tarp, F., Rand, J., Shimeles, A., & Söderbom, M. (2016).** *Manufacturing Transformation: Comparative Studies of Industrial Development in Africa and Emerging Asia*. Oxford University Press.
- OECD. (2017).** Input-Output Tables. Organisation for Economic Cooperation and Development. <http://stats.oecd.org/Index.aspx?DataSetCode=IOTS>. Accessed 29.03.2022.
- Page, J., & Tarp, F. (eds.). (2017).** The Practice of Industrial Policy: Government—Business Coordination in Africa and East Asia (Illustrated edition). Oxford University Press.
- Rand, J., Tarp, F., Trifković, N. & Zille, H. (2019).** Industrial Agglomeration in Myanmar. WIDER Working Paper Series 2019/3. World Institute for Development Economic Research (UNU-WIDER).
- Rathelot, R., & Sillard, P. (2008).** The Importance of Local Corporate Taxes in Business Location Decisions: Evidence from French Micro Data. *Economic Journal*, 118(527), 499–514.
- Robertson, B., & Taung, M. S. (2015).** Industrial Zones in Myanmar: Diagnostic Review and Policy Recommendations. Friedrich Naumann Stiftung.
- Rosenthal, S. S., & Strange, W. C. (2004).** Evidence on the Nature and Sources of Agglomeration Economies. In J. V. Henderson & J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2119–2171). Elsevier.
- Rovigatti, G., & Mollisi, V. (2018).** Theory and Practice of Total-Factor Productivity Estimation: The Control Function Approach using Stata. *The Stata Journal*, 18(3), 618–662.

-
- Sorenson, O., & Audia, P. G. (2000).** The Social Structure of Entrepreneurial Activity: Geographic Concentration of Footwear Production in the United States, 1940–1989. *American Journal of Sociology*, 106(2), 424–462.
- Tanaka, M. (2019).** Exporting Sweatshops? Evidence from Myanmar. *Review of Economics and Statistics*, 102(3), 442–456.
- Trifković, N. (2017).** Spillover Effects of International Standards: Working Conditions in the Vietnamese SMEs. *World Development*, 97, 79–101.
- Tsuji, M., Giovannetti, E., & Kagami, M. (eds.). (2007).** *Industrial Agglomeration and New Technologies: A Global Perspective*. Edward Elgar.
- UNU-WIDER. (2017).** Towards Inclusive Development in Myanmar. World Institute for Development Economic Research (UNU-WIDER). <https://www.wider.unu.edu/project/towards-inclusive-development-myanmar>. Accessed 29.03.2022.
- Van Biesebroeck, J. (2011).** Wages Equal Productivity. Fact or Fiction? Evidence from Sub Saharan Africa. *World Development*, 39(8), 1333–1346.
- Wang, J. (2013).** The Economic Impact of Special Economic Zones: Evidence from Chinese municipalities. *Journal of Development Economics*, 101, 133–147.
- Wen, M. (2004).** Relocation and agglomeration of Chinese industry. *Journal of Development Economics*, 73(1), 329–347.
- Zaw, M., & Kudo, T. (2011).** A Study on Economic Corridors and Industrial Zones, Ports and Metropolitan and Alternative Roads in Myanmar. In Masami Ishida (eds.) *Intra- and Inter-City Connectivity in the Mekong Region* (pp. 241–287). Bangkok: Bangkok Research Center.
- Zeng, D. Z. (2021).** The Past, Present, and Future of Special Economic Zones and Their Impact. *Journal of International Economic Law*, 24(2), 259–275.

APPENDIX

Table A1: Overview of Qualitative Interviews

Sector	Number	Percent
Wood	40	37.0
Textiles	21	19.4
Food and beverages	12	11.1
Apparel	9	8.3
Furniture	5	4.6
Other manufacturing	10	9.3
Forestry department	5	4.6
Banks	2	1.9
Other institutions	4	3.8
Total	108	100

Notes: Authors' calculations based on qualitative interviews.

Table A2: Skills-Driven Productivity and Wage Premiums in Industrial Zones (SUR)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Value added	Wage added	Value added	Wage added	Value added	Wage added	Value added	Wage added	Value added	Wage added	Value added	Wage added	Value added	Wage added
IZ (β_{NZ}, α_{NZ})	0.023 (0.024)	0.911*** (0.266)			0.022 (0.024)	0.025 (0.036)	0.069*** (0.022)	0.171*** (0.026)	0.021 (0.024)	0.028 (0.036)	0.069*** (0.022)	0.175*** (0.026)	0.070*** (0.022)	0.178*** (0.026)
IZ x Skill ($\beta_{IZS}, \alpha_{IZS}$)									-0.053 (0.059)	0.032 (0.058)	-0.060 (0.063)	-0.159*** (0.069)	-0.058 (0.063)	-0.169*** (0.068)
IZ x Train ($\beta_{IZT}, \alpha_{IZT}$)									0.151 (0.159)	-0.269** (0.110)	0.164 (0.156)	-0.244* (0.138)	0.159 (0.153)	-0.237* (0.138)
IZ x EGI ($\beta_{IZE}, \alpha_{IZE}$)													0.037 (0.085)	0.177 (0.151)
Skilled (β_S, α_S)			0.022 (0.027)	-0.034 (0.032)	0.022 (0.027)	-0.033 (0.032)	0.036 (0.026)	-0.019 (0.038)	0.032 (0.028)	-0.040 (0.038)	0.048* (0.025)	0.012 (0.034)	0.047* (0.025)	0.016 (0.034)
Training (β_T, α_T)			0.037 (0.081)	0.196*** (0.058)	0.032 (0.080)	0.185*** (0.059)	-0.005 (0.076)	0.197*** (0.071)	-0.045 (0.071)	0.320*** (0.096)	-0.086 (0.071)	0.306*** (0.109)	-0.086 (0.071)	0.301*** (0.106)
EGI (β_E, α_E)													0.054 (0.098)	-0.272*** (0.091)
Labor (β_L)	0.316*** (0.015)		0.319*** (0.014)		0.317*** (0.016)		0.329*** (0.013)		0.316*** (0.016)		0.329*** (0.013)		0.328*** (0.013)	
Capital (β_K)	0.703*** (0.012)		0.704*** (0.012)		0.703*** (0.012)		0.702*** (0.011)		0.703*** (0.012)		0.703*** (0.011)		0.703*** (0.011)	
TFP ($\hat{\omega}$)		0.993*** (0.084)		0.300*** (0.029)		0.300*** (0.029)		0.208*** (0.027)		0.301*** (0.029)		0.213*** (0.027)		0.211*** (0.027)
Location x Sector FE	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Location x Year FE	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Sector x Year FE	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Location FE	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Sector FE	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	No	No	No	No	No	Yes	No	No	No	Yes	No	Yes	No
Observations	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486

Notes: β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's and α 's indicate an interaction, where IZ denotes industrial zone, S denotes skills, T denotes training, and E denotes the Ellison-Glaeser Index. $\beta_{NZ} = \beta_I \beta_Z$ as explained in Equations 2 and 3. α_{NZ} denotes the industrial zone coefficient from Equation 6. Control variables include labor, capital, firm age, a family-ownership dummy, sex, education, and age of the owner or manager, as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. SUR denotes seemingly unrelated regression. Location denotes in the fixed effects denotes State/Region. Standard errors clustered at township-industry level are in parentheses. The asterisk denote the significance levels from t-tests of 1, 5, and 10 percent (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table A3: Skills-Driven Productivity and Wage Premiums in Industrial Zones (ACF)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Value added	Wage	Value added	Wage	Value added	Wage	Value added	Wage	Value added	Wage
IZ ($\beta_{NZ,INZ}$)	0.036*** (0.007)	0.154*** (0.021)			0.049*** (0.009)	0.152*** (0.021)	0.060*** (0.007)	0.156*** (0.021)	0.041*** (0.015)	0.159*** (0.022)
IZ x Skill ($\beta_{IZS,IZS}$)							-0.091*** (0.007)	-0.141*** (0.056)	-0.080*** (0.008)	-0.151*** (0.056)
IZ x Train ($\beta_{IZT,IZT}$)							0.148*** (0.007)	-0.257** (0.140)	0.144*** (0.013)	-0.249** (0.140)
IZ x EGI ($\beta_{IZE,IZE}$)								0.017 (0.109)	0.017 (0.109)	0.163 (0.109)
Skilled ($\beta_{S,S}$)			0.014* (0.007)	-0.014 (0.023)	0.012*** (0.004)	-0.011 (0.023)	0.028*** (0.005)	0.016 (0.026)	0.025*** (0.006)	0.021 (0.026)
Training ($\beta_{T,T}$)			-0.027*** (0.009)	0.156** (0.071)	-0.033*** (0.008)	0.133* (0.07)	-0.115*** (0.007)	0.251** (0.097)	-0.107*** (0.008)	0.247*** (0.097)
EGI ($\beta_{E,E}$)								0.040*** (0.014)	0.040*** (0.014)	-0.259*** (0.077)
Labor (β_I)	0.314*** (0.002)		0.326*** (0.004)		0.332*** (0.010)		0.326*** (0.006)		0.322*** (0.013)	
Capital (β_k)	0.708*** (0.008)		0.711*** (0.008)		0.709*** (0.009)		0.709*** (0.008)		0.710*** (0.010)	
TFP ($\hat{\omega}$)		0.208*** (0.019)		0.210*** (0.019)		0.208*** (0.019)		0.213*** (0.020)		0.211*** (0.020)
State/Region x Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Region x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Controls	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Observations	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486

Notes: β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's and α 's indicate an interaction, where IZ denotes industrial zone, S denotes skills, T denotes training, and E denotes the Ellison-Glaeser Index. $\beta_{INZ} = \beta_I \beta_{IZ}$ as explained in Equations 2 and 3. α_{INZ} denotes the industrial zone coefficient from Equation 6. Control variables include labor, capital, firm age, a family-ownership dummy, sex, education, and age of the owner or manager, as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. ACF denotes Akerberg-Caves-Frazier model. All columns are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level are in parentheses. The asterisk denote the significance levels from t-tests of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4: Marginal Effects of Skills-Driven Productivity and Wage Premiums in Industrial Zones (Extension to Table 3)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS				ACF			
	Value added	Wage	Value added	Wage	Value added	Wage	Value added	Wage
$\beta_{IZ} (= \beta_{INZ}/\beta_I)$	0.210		0.214		0.183		0.128	
$\beta_{IZ} = 0$ (<i>p</i> -value)	0.002		0.001		0.000		0.005	
χ^2 ($\beta_{IZ} = \alpha_{INZ}$)	0.236		0.255					
<i>p</i> -value	0.627		0.613					
$\beta_{IS} (= \beta_{ISZ}/\beta_I)$	-0.184		-0.178		-0.278		-0.25	
$\beta_{IS} = 0$ (<i>p</i> -value)	0.339		0.357		0.000		0.000	
χ^2 ($\beta_{IS} = \alpha_{ISZ}$)	0.016		0.002					
<i>p</i> -value	0.899		0.962					
$\beta_{IT} (= \beta_{ITZ}/\beta_I)$	0.500		0.483		0.452		0.446	
$\beta_{IT} = 0$ (<i>p</i> -value)	0.292		0.301		0.000		0.000	
χ^2 ($\beta_{IT} = \alpha_{ITZ}$)	2.157		2.104					
<i>p</i> -value	0.142		0.147					
$\beta_{IE} (= \beta_{IZE}/\beta_I)$			0.112				0.054	
$\beta_{IE} = 0$ (<i>p</i> -value)			0.667				0.111	
χ^2 ($\beta_{IE} = \alpha_{IZE}$)			0.048					
<i>p</i> -value			0.826					
$\beta_{IS} + \beta_{IZ}$	0.026	0.015	0.036	0.009	-0.095	0.015	-0.121	0.009
$\beta_{IS} + \beta_{IZ} = 0$ (<i>p</i> -value)	0.891	0.815	0.850	0.886	0.000	0.831	0.042	0.900
χ^2 ($\beta_{IS} + \beta_{IZ} = \alpha_{ISZ} + \alpha_{INZ}$)	0.003		0.021					
<i>p</i> -value	0.955		0.886					
$\beta_{IT} + \beta_{IZ}$	0.710	-0.069	0.696	-0.059	0.635	-0.101	0.574	-0.090
$\beta_{IT} + \beta_{IZ} = 0$ (<i>p</i> -value)	0.135	0.632	0.136	0.680	0.000	0.531	0.000	0.571
χ^2 ($\beta_{IT} + \beta_{IZ} = \alpha_{ITZ} + \alpha_{INZ}$)	2.289		2.303					
<i>p</i> -value	0.130		0.129					
$\beta_{IE} + \beta_{IZ}$			0.326	0.355			0.183	0.323
$\beta_{IE} + \beta_{IZ} = 0$ (<i>p</i> -value)			0.220	0.030			0.000	0.069
χ^2 ($\beta_{IE} + \beta_{IZ} = \alpha_{IZE} + \alpha_{INZ}$)			0.009					
<i>p</i> -value			0.925					
$\beta_{IS} + \beta_S$	-0.039	-0.148	-0.035	-0.152	-0.193	-0.125	-0.172	-0.13
$\beta_{IS} + \beta_S = 0$ (<i>p</i> -value)	0.834	0.030	0.849	0.023	0.000	0.079	0.000	0.062
χ^2 ($\beta_{IS} + \beta_S = \alpha_{ISZ} + \alpha_S$)	0.382		0.446					
<i>p</i> -value	0.537		0.504					
$\beta_{IT} + \beta_T$	0.237	0.062	0.222	0.064	0.098	-0.006	0.112	-0.002
$\beta_{IT} + \beta_T = 0$ (<i>p</i> -value)	0.584	0.471	0.603	0.467	0.001	0.950	0.003	0.979
χ^2 ($\beta_{IT} + \beta_T = \alpha_{ITZ} + \alpha_T$)	0.137		0.115					
<i>p</i> -value	0.712		0.735					
$\beta_{IE} + \beta_E$			0.276	-0.095			0.177	-0.095
$\beta_{IE} + \beta_E = 0$ (<i>p</i> -value)			0.268	0.395			0.017	0.435
χ^2 ($\beta_{IE} + \beta_E = \alpha_{IZE} + \alpha_E$)			1.912					
<i>p</i> -value			0.167					
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE}$			0.631	-0.051			0.378	-0.078
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} = 0$ (<i>p</i> -value)			0.283	0.792			0.000	0.701
χ^2 ($\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} = \alpha_{INZ} + \alpha_{ISZ} + \alpha_{ITZ} + \alpha_{IZE}$)			1.787					
<i>p</i> -value			0.181					
Observations	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486

Notes: This table shows the marginal effects of the coefficients and their interactions as estimated in Table 3. β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The letters in the subscripts denote industrial zone (*IZ*), skills (*S*), training (*T*), labor (*I*), and the Ellison-Glaeser Index (*E*). The three letter combinations in the subscript denote the interaction parameters. Columns 1-4 and 7-8 use the following fixed effects: state/region, sector, year, state/region-sector, state/region-year, and sector-year. Columns 1, 3, and 7 use additionally the following control variables: labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. ACF denotes Akerberg-Caves-Frazer model. Columns 5-8 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level are in parentheses. Bold numbers indicate statistical significance from t-tests at least at the 10 percent level.

Table A5: Marginal Effects of Productivity and Wage Gains from Women Workers in Industrial Zones (Extension to Table 4)

	(1)	(2)	(3)	(4)
	SUR		ACF	
	Value added	Wage	Value added	Wage
$\beta_{IZ} (= \beta_{INZ} / \beta_I)$	0.181		0.092	
$\beta_{IZ} = 0$ (p-value)	0.004		0.017	
χ^2 ($\beta_{IZ} = \alpha_{INZ}$)	0.011			
p-value	0.916			
$\beta_W (= \beta_{WS} / \beta_I)$	-0.304		-0.363	
$\beta_W = 0$ (p-value)	0.000		0.000	
χ^2 ($\beta_W = \alpha_{WS}$)	1.350			
p-value	0.246			
$\beta_{IS} (= \beta_{IZS} / \beta_I)$	-0.171		-0.257	
$\beta_{IS} = 0$ (p-value)	0.343		0.000	
χ^2 ($\beta_{IS} = \alpha_{IZS}$)	0.003			
p-value	0.956			
$\beta_{IT} (= \beta_{IZT} / \beta_I)$	0.572		0.479	
$\beta_{IT} = 0$ (p-value)	0.202		0.000	
χ^2 ($\beta_{IT} = \alpha_{IZT}$)	2.668			
p-value	0.102			
$\beta_{IE} (= \beta_{IZE} / \beta_I)$	0.163		0.132	
$\beta_{IE} = 0$ (p-value)	0.555		0.000	
χ^2 ($\beta_{IE} = \alpha_{IZE}$)	0.042			
p-value	0.838			
$\beta_{IW} (= \beta_{IZW} / \beta_I)$	-0.165		-0.186	
$\beta_{IW} = 0$ (p-value)	0.218		0.000	
χ^2 ($\beta_{IW} = \alpha_{IZW}$)	5.474			
p-value	0.019			
$\beta_{IS} + \beta_{IZ}$	0.010	-0.007	-0.165	-0.006
$\beta_{IS} + \beta_{IZ} = 0$ (p-value)	0.956	0.904	0.000	0.925
χ^2 ($\beta_{IS} + \beta_{IZ} = \alpha_{IZS} + \alpha_{INZ}$)	0.009			
p-value	0.923			
$\beta_{IT} + \beta_{IZ}$	0.753	-0.023	0.571	-0.054
$\beta_{IT} + \beta_{IZ} = 0$ (p-value)	0.095	0.869	0.000	0.702
χ^2 ($\beta_{IT} + \beta_{IZ} = \alpha_{IZT} + \alpha_{INZ}$)	2.684			
p-value	0.101			
$\beta_{IE} + \beta_{IZ}$	0.343	0.273	0.224	0.256
$\beta_{IE} + \beta_{IZ} = 0$ (p-value)	0.226	0.100	0.000	0.027
χ^2 ($\beta_{IE} + \beta_{IZ} = \alpha_{IZE} + \alpha_{INZ}$)	0.047			
p-value	0.828			
$\beta_{IW} + \beta_{IZ}$	0.015	0.338	-0.094	0.283
$\beta_{IW} + \beta_{IZ} = 0$ (p-value)	0.911	0.000	0.027	0.000
χ^2 ($\beta_{IW} + \beta_{IZ} = \alpha_{IZW} + \alpha_{INZ}$)	4.774			
p-value	0.029			
$\beta_{IS} + \beta_S$	-0.007	-0.153	-0.192	-0.128
$\beta_{IS} + \beta_S = 0$ (p-value)	0.967	0.019	0.000	0.011
χ^2 ($\beta_{IS} + \beta_S = \alpha_{IZS} + \alpha_S$)	0.773			
p-value	0.379			
$\beta_{IT} + \beta_T$	0.263	0.081	0.141	0.016
$\beta_{IT} + \beta_T = 0$ (p-value)	0.515	0.349	0.002	0.872
χ^2 ($\beta_{IT} + \beta_T = \alpha_{IZT} + \alpha_T$)	0.175			
p-value	0.676			
$\beta_{IE} + \beta_E$	0.457	-0.051	0.358	-0.045
$\beta_{IE} + \beta_E = 0$ (p-value)	0.074	0.042	0.000	0.673
χ^2 ($\beta_{IE} + \beta_E = \alpha_{IZE} + \alpha_E$)	3.440			
p-value	0.064			
$\beta_{IW} + \beta_W$	-0.470	-0.049	-0.549	-0.078
$\beta_{IW} + \beta_W = 0$ (p-value)	0.000	0.297	0.000	0.158
χ^2 ($\beta_{IW} + \beta_W = \alpha_{IZW} + \alpha_{WS}$)	11.039			
p-value	0.001			
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW}$	0.580	0.061	0.260	0.015
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW} = 0$ (p-value)	0.373	0.758	0.000	0.935
χ^2 ($\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW} = \alpha_{INZ} + \alpha_{IZS} + \alpha_{IZT} + \alpha_{IZE} + \alpha_{IZW}$)	0.548			
p-value	0.459			
Observations	3,486	3,486	3,486	3,486

Notes: This table shows the marginal effects of the coefficients and their interactions as estimated in Table 4. β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The letters in the subscripts denote industrial zone (IZ), skills (S), training (T), labor (I), the Ellison-Glaeser Index (E), and the share of women workers (WS). The three letter combinations in the subscript denote the interaction parameters. All columns use the following fixed effects: state/region, sector, year, state/region-sector, state/region-year, and sector-year. Columns 1 and 3 use additionally the following control variables: labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. ACF denotes Akerberg-Caves-Frazer model. Columns 3 and 4 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level. Bold numbers indicate statistical significance from t-tests at least at the 10 percent level.

Table A6: Skills-Driven Productivity and Wage Premiums in Industrial Zones with Alternative Measures of Industrial Agglomeration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Enterprise Density				Enterprise Location Quotient			
	Value added	Wage	Value added	Wage	Value added	Wage	Value added	Wage
IZ ($\beta_{INZ}, \alpha_{INZ}$)	0.056*** (0.021)	0.174*** (0.027)	0.031** (0.013)	0.157*** (0.023)	0.055** (0.022)	0.170*** (0.027)	0.037*** (0.011)	0.152*** (0.022)
IZ x Skill ($\beta_{IZS}, \alpha_{IZS}$)	-0.055 (0.060)	-0.176*** (0.068)	-0.075*** (0.017)	-0.156*** (0.056)	-0.059 (0.060)	-0.175*** (0.067)	-0.082*** (0.012)	-0.155*** (0.056)
IZ x Train ($\beta_{IZT}, \alpha_{IZT}$)	0.178 (0.152)	-0.198 (0.133)	0.165*** (0.014)	-0.207 (0.139)	0.192 (0.154)	-0.203 (0.133)	0.172*** (0.007)	-0.213 (0.139)
IZ x E ($\beta_{IZE}, \alpha_{IZE}$)	-0.006* (0.003)	0.002 (0.002)	-0.018* (0.010)	0.003 (0.005)	-0.003 (0.002)	-0.001 (0.001)	-0.030*** (0.010)	-0.001 (0.002)
IZ x Women ($\beta_{IZW}, \alpha_{IZW}$)	-0.045 (0.043)	0.179*** (0.053)	-0.058** (0.029)	0.142** (0.058)	-0.049 (0.043)	0.179*** (0.055)	-0.074*** (0.016)	0.143** (0.058)
Skilled (β_S, α_S)	0.052** (0.025)	0.026 (0.033)	0.024 (0.015)	0.031 (0.025)	0.055** (0.025)	0.025 (0.033)	0.038*** (0.010)	0.030 (0.025)
Training (β_T, α_T)	-0.083 (0.069)	0.277*** (0.101)	-0.101*** (0.013)	0.222** (0.097)	-0.093 (0.070)	0.283*** (0.101)	-0.103*** (0.009)	0.228** (0.097)
E (β_E, α_E)	0.006*** (0.001)	-0.001 (0.001)	0.005 (0.005)	-0.001 (0.002)	0.004 (0.002)	0.001 (0.001)	0.002 (0.003)	0.001 (0.002)
Women (β_{WS}, α_{WS})	-0.102*** (0.024)	-0.222*** (0.033)	-0.109*** (0.031)	-0.214*** (0.025)	-0.099*** (0.024)	-0.223*** (0.033)	-0.112*** (0.018)	-0.216*** (0.025)
Labor (β_l)	0.336*** (0.013)		0.302*** (0.013)		0.335*** (0.013)		0.322*** (0.013)	
Capital (β_k)	0.701*** (0.011)		0.707*** (0.011)		0.702*** (0.011)		0.693*** (0.012)	
TFP ($\hat{\omega}$)		0.203*** (0.030)		0.202*** (0.021)		0.202*** (0.029)		0.201*** (0.021)
State/Region x Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Region x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State/Region FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes		Yes		Yes		Yes	
Observations	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486

Notes: β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's and α 's indicate an interaction, where *IZ* denotes industrial zone, *S* denotes skills, *T* denotes training, *E* denotes the different agglomeration measures, and *W* denotes the share of women workers. $\beta_{INZ} = \beta_l \beta_{IZ}$ as explained in Equations 2 and 3, and $\beta_{WS} = \beta_l \beta_W$. α_{INZ} denotes the industrial zone coefficient from Equation 6. In Columns 1-4, *E* is the enterprise density, in Columns 5-8 it is the enterprise location quotient (see Section 2.1). Control variables include labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. ACF denotes Akerberg-Caves-Frazier model. Columns 1, 2, 5, and 6 use Seemingly Unrelated Regression (SUR). Columns 3, 4, 7, and 8 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level are in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Marginal Effects of Skills-Driven Productivity and Wage Premiums in Industrial Zones with Alternative Measures of Industrial Agglomeration (Extension to Table A6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Enterprise Density				Enterprise Location Quotient			
	Value added	Wage	Value added	Wage	Value added	Wage	Value added	Wage
$\beta_{IZ} (= \beta_{INZ} / \beta_I)$	0.167		0.101		0.165		0.115	
$\beta_{IZ} = 0$ (p-value)	0.008		0.024		0.010		0.001	
χ^2 ($\beta_{IZ} = \alpha_{INZ}$)	0.010				0.006			
p-value	0.920				0.936			
$\beta_{IS} (= \beta_{IZS} / \beta_I)$	-0.164		-0.249		-0.177		-0.254	
$\beta_{IS} = 0$ (p-value)	0.359		0.000		0.325		0.000	
χ^2 ($\beta_{IS} = \alpha_{IZS}$)	0.005				0.000			
p-value	0.945				0.992			
$\beta_{IT} (= \beta_{IZT} / \beta_I)$	0.530		0.548		0.574		0.533	
$\beta_{IT} = 0$ (p-value)	0.243		0.000		0.211		0.000	
χ^2 ($\beta_{IT} = \alpha_{IZT}$)	2.292				2.537			
p-value	0.130				0.111			
$\beta_{IE} (= \beta_{IZE} / \beta_I)$	-0.017		-0.060		-0.009		-0.092	
$\beta_{IE} = 0$ (p-value)	0.064		0.086		0.167		0.003	
χ^2 ($\beta_{IE} = \alpha_{IZE}$)	3.545				1.385			
p-value	0.060				0.239			
$\beta_{IW} (= \beta_{IZW} / \beta_I)$	-0.133		-0.192		-0.146		-0.228	
$\beta_{IW} = 0$ (p-value)	0.289		0.041		0.248		0.000	
χ^2 ($\beta_{IW} = \alpha_{IZW}$)	5.632				5.884			
p-value	0.018				0.015			
$\beta_{IS} + \beta_{IZ}$	0.003	-0.002	-0.148	0.001	-0.012	-0.005	-0.140	-0.003
$\beta_{IS} + \beta_{IZ} = 0$ (p-value)	0.985	0.970	0.035	0.986	0.945	0.934	0.007	0.961
χ^2 ($\beta_{IS} + \beta_{IZ} = \alpha_{IZS} + \alpha_{INZ}$)	0.001				0.002			
p-value	0.975				0.968			
$\beta_{IT} + \beta_{IZ}$	0.697	-0.024	0.649	-0.049	0.738	-0.033	0.647	-0.061
$\beta_{IT} + \beta_{IZ} = 0$ (p-value)	0.128	0.861	0.000	0.725	0.109	0.811	0.000	0.664
χ^2 ($\beta_{IT} + \beta_{IZ} = \alpha_{IZT} + \alpha_{INZ}$)	2.205				2.467			
p-value	0.138				0.116			
$\beta_{IE} + \beta_{IZ}$	0.150	0.175	0.042	0.161	0.155	0.169	0.023	0.151
$\beta_{IE} + \beta_{IZ} = 0$ (p-value)	0.020	0.000	0.379	0.000	0.018	0.000	0.648	0.000
χ^2 ($\beta_{IE} + \beta_{IZ} = \alpha_{IZE} + \alpha_{INZ}$)	0.134				0.038			
p-value	0.715				0.846			
$\beta_{IW} + \beta_{IZ}$	0.034	0.352	-0.091	0.299	0.019	0.349	-0.114	0.295
$\beta_{IW} + \beta_{IZ} = 0$ (p-value)	0.795	0.000	0.416	0.000	0.884	0.000	0.021	0.000
χ^2 ($\beta_{IW} + \beta_{IZ} = \alpha_{IZW} + \alpha_{INZ}$)	5.103				5.253			
p-value	0.024				0.022			
$\beta_{IS} + \beta_S$	-0.008	-0.15	-0.169	-0.125	-0.013	-0.150	-0.138	-0.125
$\beta_{IS} + \beta_S = 0$ (p-value)	0.963	0.025	0.024	0.013	0.940	0.024	0.002	0.013
χ^2 ($\beta_{IS} + \beta_S = \alpha_{IZS} + \alpha_S$)	0.745				0.688			
p-value	0.388				0.407			
$\beta_{IT} + \beta_T$	0.284	0.079	0.213	0.015	0.295	0.080	0.212	0.015
$\beta_{IT} + \beta_T = 0$ (p-value)	0.495	0.373	0.000	0.881	0.483	0.363	0.000	0.880
χ^2 ($\beta_{IT} + \beta_T = \alpha_{IZT} + \alpha_T$)	0.207				0.222			
p-value	0.649				0.637			
$\beta_{IE} + \beta_E$	0.002	0.001	-0.044	0.002	0.001	0.000	-0.086	0.000
$\beta_{IE} + \beta_E = 0$ (p-value)	0.812	0.576	0.113	0.637	0.102	0.881	0.003	0.799
χ^2 ($\beta_{IE} + \beta_E = \alpha_{IZE} + \alpha_E$)	0.013				2.468			
p-value	0.910				0.116			
$\beta_{IW} + \beta_W$	-0.438	-0.043	-0.554	-0.072	-0.440	-0.043	-0.576	-0.072
$\beta_{IW} + \beta_W = 0$ (p-value)	0.000	0.430	0.000	0.183	0.000	0.432	0.000	0.181
χ^2 ($\beta_{IW} + \beta_W = \alpha_{IZW} + \alpha_{WS}$)	11.021				10.352			
p-value	0.001				0.001			
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW}$	0.383	-0.020	0.148	-0.060	0.406	-0.030	0.073	-0.073
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW} = 0$ (p-value)	0.431	0.883	0.233	0.712	0.403	0.827	0.300	0.653
χ^2 ($\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW} = \alpha_{INZ} + \alpha_{IZS} + \alpha_{IZT} + \alpha_{IZE} + \alpha_{IZW}$)	0.655				0.762			
p-value	0.418				0.383			
Observations	3,486	3,486	3,486	3,486	3,486	3,486	3,486	3,486

Notes: This table shows the marginal effects of the coefficients and their interactions as estimated in Table A6. β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's indicate an interaction, where IZ denotes industrial zone, S denotes skills, T denotes training, E denotes the different agglomeration measures, and W denotes the share of women workers. $\beta_{INZ} = \beta_I \beta_{IZ}$ as explained in Equations 2 and 3, and $\beta_{WS} = \beta_I \beta_W$. In Columns 1-4, E is the enterprise density, in Columns 5-8 it is the enterprise location quotient (see Section 2.1). Columns 1-4 use the following fixed effects: state/region, sector, year, state/region-sector, state/region-year, and sector-year. Columns 1 and 3 use additionally the following control variables: labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. ACF denotes Akerberg-Caves-Frazer model. Columns 1, 2, 5, and 6 use Seemingly Unrelated Regression (SUR). Columns 3, 4, 7, and 8 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level. Bold numbers indicate statistical significance from t-tests at least at the 10 percent level.

Table A8: Estimates of Skills-Driven Productivity and Wage Gains in Industrial Zones Without Informal Firms

	(1)	(2)	(3)	(4)
	SUR		ACF	
	Value added	Wage	Value added	Wage
IZ ($\beta_{INZ}, \alpha_{INZ}$)	0.064*** (0.022)	0.147*** (0.025)	0.051** (0.012)	0.134*** (0.022)
IZ x Skill ($\beta_{IZS}, \alpha_{IZS}$)	-0.082 (0.064)	-0.154** (0.060)	-0.122*** (0.007)	-0.134*** (0.056)
IZ x Train ($\beta_{IZT}, \alpha_{IZT}$)	0.194 (0.157)	-0.148 (0.130)	0.171*** (0.008)	-0.164 (0.139)
IZ x EGI ($\beta_{IZE}, \alpha_{IZE}$)	0.082 (0.099)	0.038 (0.149)	0.071*** (0.008)	0.051 (0.119)
IZ x Women ($\beta_{IZW}, \alpha_{IZW}$)	-0.007 (0.048)	0.179*** (0.051)	-0.016*** (0.011)	0.150** (0.060)
Skilled (β_S, α_S)	0.066** (0.027)	0.010 (0.027)	0.034*** (0.009)	0.008 (0.027)
Training (β_T, α_T)	-0.101 (0.078)	0.250** (0.099)	-0.124*** (0.008)	0.208** (0.095)
EGI (β_E, α_E)	0.144 (0.110)	-0.086 (0.093)	0.120*** (0.007)	-0.100 (0.082)
Women (β_{WS}, α_{WS})	-0.115*** (0.026)	-0.252*** (0.034)	-0.146*** (0.013)	-0.247*** (0.027)
Labor (β_l)	0.330*** (0.014)		0.339*** (0.018)	
Capital (β_k)	0.704*** (0.012)		0.710*** (0.018)	
TFP ($\hat{\omega}$)		0.201*** (0.032)		0.201*** (0.022)
State/Region x Sector FE	Yes	Yes	Yes	Yes
State/Region x Year FE	Yes	Yes	Yes	Yes
Sector x Year FE	Yes	Yes	Yes	Yes
State/Region FE	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes		Yes	
Observations	3,094	3,094	3,094	3,094

Notes: β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The three letters next to each other in the subscript of the β 's and α 's indicate an interaction, where *IZ* denotes industrial zone, *S* denotes skills, *T* denotes training, *E* denotes the Ellison-Glaeser Index, and *WS* denotes the share of women workers. $\beta_{INZ} = \beta_I \beta_{IZ}$ as explained in Equations 2 and 3, and $\beta_{WS} = \beta_I \beta_W$. α_{INZ} denotes the industrial zone coefficient from Equation 6. Control variables include labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. SUR denotes seemingly unrelated regression. ACF denotes Akerberg-Caves-Frazer model. Columns 3 and 4 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level are in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table A9: Marginal Effects of Skills-Driven Productivity and Wage Gains in Industrial Zones Without Informal Firms (Extension to Table A8)

	(1)	(2)	(3)	(4)
	SUR		AFC	
	Value added	Wage	Value added	Wage
$\beta_{IZ} (= \beta_{INZ} / \beta_I)$	0.195		0.151	
$\beta_{IZ} = 0$ (<i>p</i> -value)	0.003		0.000	
χ^2 ($\beta_{IZ} = \alpha_{INZ}$)	0.475			
<i>p</i> -value	0.491			
$\beta_{IS} (= \beta_{ISZ} / \beta_I)$	-0.250		-0.329	
$\beta_{IS} = 0$ (<i>p</i> -value)	0.202		0.000	
χ^2 ($\beta_{IS} = \alpha_{ISZ}$)	0.264			
<i>p</i> -value	0.607			
$\beta_{IT} (= \beta_{ITZ} / \beta_I)$	0.588		0.505	
$\beta_{IT} = 0$ (<i>p</i> -value)	0.218		0.000	
χ^2 ($\beta_{IT} = \alpha_{ITZ}$)	2.230			
<i>p</i> -value	0.135			
$\beta_{IE} (= \beta_{IEZ} / \beta_I)$	0.250		0.209	
$\beta_{IE} = 0$ (<i>p</i> -value)	0.405		0.000	
χ^2 ($\beta_{IE} = \alpha_{IEZ}$)	0.375			
<i>p</i> -value	0.541			
$\beta_{IW} (= \beta_{IYW} / \beta_I)$	-0.020		-0.048	
$\beta_{IW} = 0$ (<i>p</i> -value)	0.890		0.150	
χ^2 ($\beta_{IW} = \alpha_{IYW}$)	1.840			
<i>p</i> -value	0.175			
$\beta_{IS} + \beta_{IZ}$	-0.055	-0.008	-0.178	0.000
$\beta_{IS} + \beta_{IZ} = 0$ (<i>p</i> -value)	0.768	0.896	0.000	0.996
χ^2 ($\beta_{IS} + \beta_{IZ} = \alpha_{ISZ} + \alpha_{INZ}$)	0.070			
<i>p</i> -value	0.792			
$\beta_{IT} + \beta_{IZ}$	0.783	-0.002	0.656	-0.030
$\beta_{IT} + \beta_{IZ} = 0$ (<i>p</i> -value)	0.102	0.990	0.000	0.823
χ^2 ($\beta_{IT} + \beta_{IZ} = \alpha_{ITZ} + \alpha_{INZ}$)	2.493			
<i>p</i> -value	0.114			
$\beta_{IE} + \beta_{IZ}$	0.445	0.185	0.360	0.184
$\beta_{IE} + \beta_{IZ} = 0$ (<i>p</i> -value)	0.153	0.251	0.000	0.138
χ^2 ($\beta_{IE} + \beta_{IZ} = \alpha_{IEZ} + \alpha_{INZ}$)	0.511			
<i>p</i> -value	0.475			
$\beta_{IW} + \beta_{IZ}$	0.175	0.326	0.103	0.284
$\beta_{IW} + \beta_{IZ} = 0$ (<i>p</i> -value)	0.252	0.000	0.016	0.000
χ^2 ($\beta_{IW} + \beta_{IZ} = \alpha_{IYW} + \alpha_{INZ}$)	0.957			
<i>p</i> -value	0.328			
$\beta_{IS} + \beta_S$	-0.051	-0.165	-0.227	-0.142
$\beta_{IS} + \beta_S = 0$ (<i>p</i> -value)	0.783	0.008	0.000	0.005
χ^2 ($\beta_{IS} + \beta_S = \alpha_{ISZ} + \alpha_S$)	0.421			
<i>p</i> -value	0.517			
$\beta_{IT} + \beta_T$	0.282	0.102	0.139	0.044
$\beta_{IT} + \beta_T = 0$ (<i>p</i> -value)	0.511	0.233	0.000	0.653
χ^2 ($\beta_{IT} + \beta_T = \alpha_{ITZ} + \alpha_T$)	0.155			
<i>p</i> -value	0.694			
$\beta_{IE} + \beta_E$	0.688	-0.048	0.564	-0.049
$\beta_{IE} + \beta_E = 0$ (<i>p</i> -value)	0.014	0.145	0.000	0.663
χ^2 ($\beta_{IE} + \beta_E = \alpha_{IEZ} + \alpha_E$)	5.299			
<i>p</i> -value	0.021			
$\beta_{IW} + \beta_W$	-0.369	-0.073	-0.479	-0.097
$\beta_{IW} + \beta_W = 0$ (<i>p</i> -value)	0.005	0.119	0.000	0.079
χ^2 ($\beta_{IW} + \beta_W = \alpha_{IYW} + \alpha_{WS}$)	5.344			
<i>p</i> -value	0.021			
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW}$	0.763	0.061	0.489	0.036
$\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW} = 0$ (<i>p</i> -value)	0.312	0.761	0.000	0.849
χ^2 ($\beta_{IZ} + \beta_{IS} + \beta_{IT} + \beta_{IE} + \beta_{IW} = \alpha_{INZ} + \alpha_{ISZ} + \alpha_{ITZ} + \alpha_{IEZ} + \alpha_{IYW}$)	0.743			
<i>p</i> -value	0.389			
Observations	3,094	3,094	3,094	3,094

Notes: This table shows the marginal effects of the coefficients and their interactions as estimated in Table A8. β 's refer to the productivity equation (Equation 3), α 's refer to the wage equation (Equation 6). The letters in the subscripts denote industrial zone (*IZ*), skills (*S*), training (*T*), labor (*I*), the Ellison-Glaeser Index (*E*), and the share of women workers (*WS*). The three letter combinations in the subscript denote the interaction parameters. All columns use the following fixed effects: state/region, sector, year, state/region-sector, state/region-year, and sector-year. Columns 1 and 3 use additionally the following control variables: labor, capital, firm age, a family-ownership dummy; sex, education, and age of the owner or manager; as well as the share of workers with a formal contract and the share of workers receiving in-kind benefits. SUR denotes seemingly unrelated regression. ACF denotes Akerberg-Caves-Frazer model. Columns 3 and 4 are GMM estimates based on the procedure described in Akerberg et al. (2015) with standard errors based on 101 bootstrap replications. Standard errors clustered at township-industry level. Bold numbers indicate statistical significance from t-tests at least at the 10 percent level.

Chapter 2

Brothers in Arms, Brothers in Trade? Measuring the Effect of Violent Conflicts on Trade with Third-Party Countries

Brothers in Arms, Brothers in Trade? Measuring the Effect of Violent Conflicts on Trade with Third-Party Countries*

Helge Zille

Department of Economics, University of Copenhagen

The question about the relationship between violent conflicts and international trade has a long tradition, and empirical research in the 1990s and early 2000s has established that violent interstate conflicts do harm international trade. While most of this literature dates back at least 10 to 20 years, the effect of interstate conflicts on trade with third-party countries has been neglected for most of the time in the literature. In this paper, I attempt to fill this gap. A period of 46 years is covered in the analysis, using more than 500 thousand dyad-year observations. The third-party country dimension is derived from a triadic data set, which covers all possible country-triad combinations for the studied period. I find that violent interstate conflicts reduce trade with third-party countries, and that they cause a shift in trade towards allied countries and away from the enemy's allies. Countries increase imports from members of the same security alliance by between 1 and 4 percent, and trade more with countries that have the same enemies by between 5 and 7 percent. They reduce trade with the formal allies of their enemies by between 9 and 14 percent. This negative trade shifting effect is further amplified by the size of the respective conflict country. This paper contributes to the literature on conflict and trade in two ways: First, by adding to the scarce literature introducing a third-country dimension into standard gravity models and into the literature on conflict and trade. And second, by showing the importance of a spatially dynamic perspective on interstate conflicts. (JEL F14, F51)

*The current version of this paper is to be revised and resubmitted to the Review of International

Organizations. This chapter builds on and reuses text from my master thesis with the title *How does violent conflict affect international trade? Measuring the effect of bilateral conflict on third-country trade in a triadic country setting with different relationship constellations*. I would like to thank my supervisor John Rand as well as Henrik Hansen for their comments and excellent guidance throughout the process of the paper. Further, I want to thank Magnus Tolum Buus, Bertel Teilfeldt Hansen, Ferdinand Rauch, Eliana la Ferrara, Kasper Brandt, Benjamin Wache, Bjørn Bo Sørensen, and Inês Ferreira for their useful comments and suggestions. Finally, I want to thank the participants of the Nordic Conference on Development Economics 2018, of the Danish Graduate Programme in Economics Workshop 2018, of the Dynamic Economics, Growth, and International Trade Conference 2019, of the European Trade Study Group Conference 2019, of the Columbia University International Trade Colloquium 2020, and of the different internal seminars at the University of Copenhagen.

1. INTRODUCTION

The idea that interstate conflict and trade influence each other dates back to the 18th century and has been extensively studied in the 1980s, 1990s, and the early 2000s. Some authors have found that bilateral trade decreases the likelihood of engaging in a violent conflict with each other (for example Polacheck, 1980; Mansfield and Peavehouse, 2000; Oneal et al., 2003). Others, such as Sayers (1990), Barbieri (1996), or Martin et al. (2008), found that trade can both lead to cooperation or to conflict, depending on other political and economic factors.

Moreover, violent conflicts have been found to decrease bilateral trade flows between the conflict parties, even though the estimated magnitudes vary significantly. Keshk et al. (2010) find reductions in trade by between 4 and 10 percent, Li and Sacko (2002) by between 6 and 12 percent, and Long (2008) by more than 70 percent. Between 1968 and 2013 alone, there have been 380 violent interstate conflicts globally. Over the same period the share of trade in global GDP has risen from around 26 to 60 percent (World Bank, 2021). The estimated trade reductions, even at the lower bound, are thus economically significant, and have become much more so over time.

So far, only few papers have studied the economic effects of violent interstate conflicts on third-party countries. Feldman and Sadeh (2018) and Korovkin and Makarin (2021) are the only ones using econometric models to show how trade with third-party countries is affected by dyadic conflicts. My study attempts to contribute to this scarce literature by globally studying the effect of interstate violent conflicts on trade with third-party countries and by directly testing the influence of specific relationship characteristics between the countries of interest on the effect of interstate conflict on trade.

I develop an analytical framework combining insights from both economics and political science and derive five hypotheses to be tested. The relationship constellations that are tested in the different hypotheses and that are expected to determine how trade with third-party countries is affected by conflict, are geographic contiguity, formal security alliances, and common conflicts. Therefore, available political, geographical, and macroeconomic data is used to create a dyadic panel data set with information on third-party countries. The particularity of this paper is that it uses a dyadic data set which is tailored to test for very specific relationship constellations with countries outside of the respective dyad. Using available data on all sovereign countries over a period of 46 years, I

generate a panel data set of more than 500,000 unique dyad-year observations. I use a gravity model of trade with dyad-specific fixed effects and a large number of country-specific and dyad-specific control variables. The fixed effects control for observed and unobserved time-invariant dyad-specific characteristics, while the country-specific and dyad-specific control variables minimize the risk of bias due to omitted variables.

The results show that violent interstate conflicts do affect trade with third-party countries, and that the different relationship constellations significantly matter for this relationship. I find robust evidence that violent interstate conflicts reduce trade with third-party countries by between 1 and 4 percent. Further, I find evidence that trade is used as an indirect measure to support allies and to harm enemies. If two countries have a violent conflict with the same third-party country, they increase their bilateral trade by between 5 and 7 percent. Furthermore, states decrease their bilateral trade with the formal allies of the countries they are in conflict with by between 9 and 14 percent. A slight increase in imports of between 1 and 4 percent is experienced from formal security allies. These effects are already net of the stand-alone negative effect of the conflict on an economy's trade level. Given the importance of international trade in GDP, these magnitudes are significant. Neighbors of countries that experience a conflict do not appear to significantly gain in terms of trade. Further results show that the size of the conflict countries amplifies the negative effects on trade with third-party countries, but it does not appear to increase the positive trade shifting effect to allies or countries with the same enemy.

The remainder of the paper is structured as follows. Section 2 gives a more detailed overview of the theoretical and empirical findings within the nexus of conflicts and trade. In Section 3, the conceptual setup and the different relationship constellations are introduced, and the corresponding hypotheses are derived. Section 4 describes the identification strategy and the data, before Section 5 then presents and discusses the results. Finally, Section 6 briefly summarizes the findings and explains their possible implications for future research.

2. THE POLITICAL ECONOMY OF CONFLICT AND TRADE

As early as in the 18th century has the idea been discussed that interstate conflicts and trade influence each other. Scholars such as Adam Smith, Immanuel Kant, Vladimir Lenin, and Montesquieu have studied this nexus (Reuveny, 2000). But it is within the studies of international political economy (IPE), investigating the role of economic activities, interests, and politics in the international sphere of states, that the conflict-trade nexus has found its academic leanings. Three of the predominant IPE theories are Mercantilism, Marxism, and Liberalism.

Liberalism's stance on interstate conflict and trade is mainly based on classic economic trade theory and the concept of the comparative advantage. Liberals see international trade as a positive-sum game and focus on absolute rather than on relative gains from trade. Given that, entering an interstate conflict with a trade partner would be costly due to potential reductions in trade flows. Following an economic rationality, this would make states reluctant to enter a violent conflict with other states (Gilpin, 1987). Moreover, Liberals see economics and politics as two different fields. Political and economic decisions might be influenced by each other, but neither of them is ultimately superior to the other. Accordingly, war and imperialism are caused by political, rather than economic factors and decisions. Consistent with these ideas, Polachek (1980) and Gasiorowski and Polachek (1982) find a strong inverse relationship between trade and interstate conflict. Polachek (1980) develops an expected utility model, in which interstate conflict is seen as a source of costs and risks, especially in regard to trade, which would reduce profits and wealth. Therefore, conflictive behavior is irrational as it has a lower expected utility than cooperation. Other empirical studies confirm these results while setting different foci, for instance by focusing on democratic norms (Oneal et al., 1996 and Oneal and Russett, 1997), on preferential trade agreements (Mansfield and Pevehouse, 2000), on different goods (Oneal et al., 2003 and Dorussen, 2006), on trade symmetries (Hegre, 2004), or on military alliances (Jackson and Nei, 2015). Recent literature emphasized the role of third-party trade and trade networks, and found that third-party trade ties reduce the likelihood of conflict (Cranmer et al., 2015; Kinne, 2014; Lupu and Traag, 2013).

Mercantilists and Marxist both have a more pessimistic view on the role of trade for interstate conflict. In both theories, trade is seen as a zero-sum game and trade relations are always asymmetrical. That produces trade dependencies, rather than the potentially pacifying interdependencies.

Both theories predict that over these unequal trade relations, tensions and eventually conflicts would arise. Either among the developed and imperialistic countries over controversies about protectionist measures or access to markets (Mercantilist view), or between developing countries and those countries, they are dependent on (Marxist view). Mercantilism further builds on the idea that economics is subordinated to politics and economic relations are determined by political factors. Marxism, on the other hand, argues that economics drive politics (Gilpin, 1987). Both theories perceive trade as inherently conflictive and interstate conflict as a result of trade as inevitable (Sayrs, 1990). Some of these Mercantilist and Marxist arguments have been used in empirical studies to show that trade can promote interstate conflict (e.g. Hirschman, 1945, Uchitel, 1993, and Barbieri, 2002), even though the evidence is more limited. Other authors reject the view that trade is the cause exclusively for one of the two extremes (peace or conflict), and argue that trade can have a conflictive as well as a cooperative effect. Sayrs (1990), for instance, criticizes and rejects the assumption that conflict and cooperation are the two “opposite outcomes of the same continuum” (Sayrs, 1990: 23). De Vries (1990) argues that trade intensifies the interaction between two countries in general, thereby generating gains but also potentially controversies. Following this approach, Uchitel (1993), Barbieri (1996, 2002), Gasiorowski (1986), and Li and Reuveny (2008), among others, find similarly mixed results. Martin et al. (2008) find that while bilateral trade openness reduces the likelihood of violent conflicts, multilateral trade openness increases the conflict probability, as it decreases bilateral dependence and thereby the potential costs of a conflict. Peterson (2011) shows that third-party trade ties can aggravate dyadic conflicts.

In those IPE theories, international trade is characterized by specialization. The products in which countries specialize are normally determined by differences in factor endowments. Developing countries are usually characterized by abundance in natural resources and relatively unskilled labor, and specialize in labor intense products and raw materials, while developed countries are abundant in technology and relatively high-skilled labor (Bernard et al., 2012). The benefits that arise from specialization are creating interdependencies and are the reason why, according to Liberals, trading states will not start a conflict with each other. And it is these interdependencies that, as Marxists and Mercantilists argue, are always asymmetrical and lead to exploitation, vulnerability, and eventually to conflict.

However, more recent work such as Melitz (2003), Bernard et al. (2003, 2012), and Melitz and

Trefler (2012) has focused on firm heterogeneity and product diversification rather than specialization as a driver of trade. Melitz and Trefler (2012) argue that the gains from trade stem from three sources: Love-of-variety gains associated with intra-industry trade, productivity gains associated with trade-related innovation, and efficiency gains stemming from the reallocation of capital and labor from small and less productive to large and more productive firms. These findings have implications for the trade and conflict nexus and for the introduced IPE theories. Trade dependencies and interdependencies can be expected to be less intense when countries diversify trade and participate in intra-industry trade. If a country trades more different goods, it is also more likely that it trades with more different partners or at least has the possibility to do so. Mercantilism's and Marxism's main argument, that trade leads to interstate conflict because of dependencies, thus loses power. If sufficiently diversified and heterogeneous, trade does not have to lead to dependency, exploitation, vulnerability, and aggression. A more diversified range of importers and exporters also makes it easier for an economy to substitute trade losses due to conflict. Thus, the economic cost of interstate conflict could decrease compared to a state of specialization and the pacifying effect of trade would decrease. New trade theories therefore weaken the arguments for both the pessimistic view of Mercantilism and Marxism that trade promotes interstate conflict, and for the optimistic view of Liberalism that trade promotes peace.

Pollins (1989) was one of the first to empirically study the conflict-trade nexus in the reverse direction. He argues that there are three different categories of actors who could potentially influence trade: Governments, interest groups, and individuals. Using a gravity model, his findings confirm the assumed effect, namely that an increase in bilateral cooperation increases bilateral trade. Bergeijk (1994) and Gowa (1994) build up on a similar model and find that interstate conflict is negatively correlated with bilateral trade, and that alliances positively affect trade. Kim (1998) uses three different trade measures to find for all of them a reduction with the militarized interstate dispute variable. Morrow et al. (1998), on the other hand, deploy again the idea that cooperation and conflict are not necessarily two values of the same variable. They include separate variables for both alliances and military conflict. Their results are rather mixed and cannot confirm previous results, which might stem from their very limited selection of countries into their sample. Li and Sacko (2002) add measures for the severity, duration, and unexpectedness of a military conflict to their analysis. They find that all these three factors intensify the negative effect that interstate

conflict has on trade. Keshk et al. (2004, 2010), motivated by concerns about the direction of causality, use a simultaneous equations model to find that conflict inhibits trade, while the effect of economic interdependence on the likelihood of a military conflict is statistically insignificant, thereby questioning the liberal trade-promotes-peace argument. Long (2008) finds that also violent conflicts short of war as well as the expectation of these reduce trade. Further, he finds that also inner-state conflicts and the expectations of these reduce trade. Bayer and Rupert (2004) look specifically at the effect of civil wars on international trade. Their analysis indicates that a civil war strongly reduces bilateral trade between the country where the civil war occurs and its trade partners.

Only very few studies have touched upon how trade is shifted because of conflicts and how third parties are involved in this nexus. Korovkin and Makarin (2021) use Ukrainian firm-level data to study the impact of the Russian-Ukrainian Crimea conflict on trade in non-conflict areas. They find that Ukrainian counties in non-conflict areas with a higher proportion of ethnic Russians experience a lower trade decrease. They further show that firms from areas with fewer ethnic Russians increase their trade with Turkey and Poland, thereby showing trade-shifting to third-party countries on the firm-level. Theoretically and methodologically somewhat related (even though trade is not the outcome variable here), Li et al. (2017) apply network theory to study military alliance networks and the pacifying effect of higher-degree alliances. They find that the likelihood of a violent conflict is lower between up to third-degree allies, meaning allies of allies of allies.

De Groot (2010), building upon previous studies by Murdoch and Sandler (2002a, 2002b, 2004), investigates the spillover effects of conflicts on economic growth in neighboring countries in Africa. His results show that conflicts do have a negative effect on growth in primary neighbors, but that secondary neighbors (neighbors of neighbors) can actually benefit in terms of growth. As one possible channel, De Groot names trade. First, because primary neighbors would possibly move some of their trade, away from the state that experiences the conflict to their other neighbors, thus the secondary neighbors of the state with the conflict. Second, because richer (non-African) nations would turn to new trade partners with similar resources or industries. However, the article does not test these assumptions empirically, leaving them open for further research.

Caruso (2003) studies the effect of unilateral and multilateral economic sanctions on bilateral trade. He finds that both have a negative impact on bilateral trade. However, when it comes

to unilateral sanctions imposed by the US, he finds that while bilateral trade between the US and the receiver country decreases, it increases between the receiver country and the other G7 countries. Even though there has not necessarily to be an underlying military conflict (as it will be the case in this work), Caruso finds evidence for trade shifting due to foreign policy affairs and for “sanction-busting” by other states (Caruso, 2003: 19).

Fuchs (2018), in his work on labor mobility in Spain during WWI, shows case study evidence that when France entered the war against Germany, imports from Spain increased substantially. Fuchs argues that the deployment of French troops led to a loss of domestic production which had to be made up for by imports.

Finally, the study by Feldman and Sadeh (2018) has the closest relation to my study. Feldman and Sadeh use data from 1885 to 2000 to show that combatant countries trade more with third-party countries with higher alliance similarity, and less so with countries with higher alliance similarity with their enemy in the conflict. While Feldman and Sadeh consider alliance similarity as a facilitator to third-party trade, my analysis includes geographical contiguity and common conflicts. I moreover extend the data to include a more recent period (until 2013) and to study both imports and exports, and, finally, I extend the analysis to consider the size of the conflict countries.

3. CONCEPTUAL SETUP

3.1. Dyadic Setting

There are several plausible reasons why violent conflicts inhibit trade. As a result from a reduction of labor and/or capital, a violent conflict can cause a reduction in the production possibilities of an economy and thereby in its potential ability to trade. Violent conflicts can cause damage and destruction to production facilities like factories, but also to agricultural land, natural resources, or infrastructure. Similarly, casualties among civilians cause a loss in the labor force. Another reason for the reduction of capital and labor is the reallocation and reprioritization of both. At least in a severe military conflict, labor and public (and private) spending is allocated increasingly to military purposes (Feldman and Sadeh, 2018). Foreign firms and investors might reallocate their

resources too, as a country with an ongoing conflict is a less favorable destination for foreign direct investments.

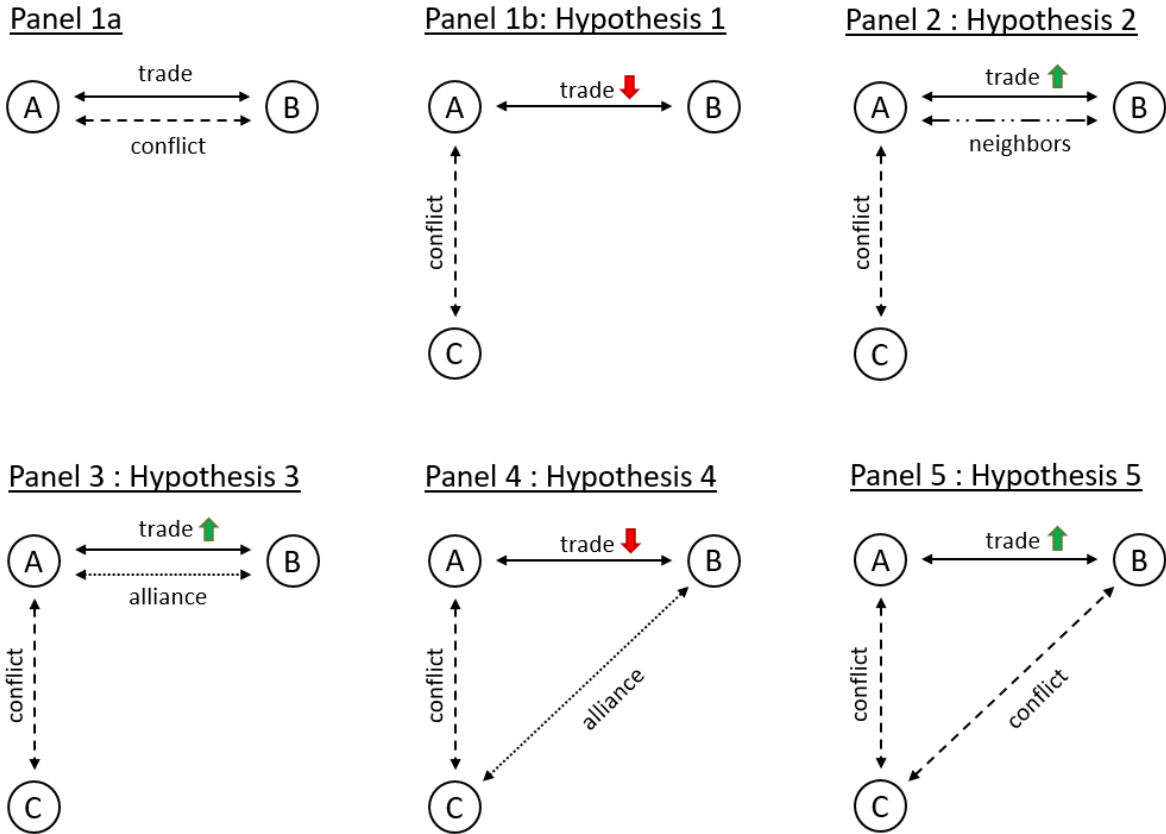
The above-mentioned mechanisms depend on the type, intensity, and location of the conflict as well as on the conflict parties. A country is more likely to experience destruction if the conflict takes place on its own territory. If, on the other hand, the conflict does not take place on the own territory at all, it is unlikely that the country will face destruction of production facilities or infrastructure. Also in a highly asymmetric conflict, it can happen that one side barely faces any destruction. Countries with sufficient military forces would also not necessarily need to allocate additional resources to its military. This is often the case for supporter or intervener countries. In the Gulf War (1990/1991) for instance, a total of 35 countries deployed troops to respond to Iraq's invasion and annexation of Kuwait. However, none of these faced any destruction at home or had to reallocate large amounts of resources to their defense capabilities.

Besides these mechanical effects, a reduction in trade can be the result of an active decision. This decision can be made on different levels: By politicians/governments, companies, interest groups, or individuals (Pollins, 1989)¹. A government can impose sanctions such as a trade embargo, thereby forbidding and punishing companies and individuals trading with the sanctioned country (Long, 2008; Whang et al., 2013). But also without a political decision, companies can decide to reduce or disrupt trade with the adversary country. They may want to support their government or country in general and thus act for ideological or nationalistic reasons (Michaels and Zhi, 2010). On the other hand, they may have economic reasons. Especially in highly violent conflicts, it is difficult and dangerous to transport goods (Long, 2008). Infrastructure can be destroyed, and, in certain war-affected zones, transportation is simply too risky (Feldman and Sadeh, 2018). Moreover, the enforcement of law may not be secured, thus a higher risk of loss is imposed (Korovkin and Makarin, 2021). All these factors constitute additional costs for a company, making a reduction of trade likely.

Individuals have an impact through their consumption decisions. The motives are very similar to those of companies. It can be a political or ideological decision not to buy products from the adversary country in order to weaken it and support the own country (Katz, 2013). It can also be an economic decision, because products from the conflict country will be more expensive (as

¹ Pollins (1989) only mentions three levels, as he treats companies as part of the interest groups.

Figure 1: Triadic Relationship Constellations



explained for companies). Especially if many individuals bundle their efforts to boycott certain products, this can have a severe impact. Interest groups can influence the decision of whether to reduce trade at all the three levels. Company interest groups, for instance, can have a strong concern that the government does not impose a trade embargo if their income highly relies on this particular trade. Unions may have an interest in maintaining trade too, if many of their members rely on it. Other groups may have an interest in the opposite. Human rights groups, for instance, can have an interest in sanctioning a country for its human rights violations. Nationalist groups can be motivated in weakening the other country by reducing trade with it. Company interest groups representing companies that are competitors of foreign firms can also have an interest in trade reduction. These different mechanisms have been put to empirics by a large body scholars as outlaid in Section 2. As a setting of the scene, I will test whether violent conflicts reduce bilateral trade. This baseline scenario is illustrated in Panel 1a of Figure 1.

3.2. Triadic Setting

One mechanism of how third-party countries can be affected is trade shifting. That is, one country may deliberately (or forcibly) reduce its trade with a partner country, and increase it with another (thus, shifts it to that country). In this way, it tries to compensate for some of the lost trade with the country in conflict. To illustrate that, assume that a violent conflict erupts between country A and country C. As explained in the derivation of the baseline dyadic case, trade between country A and country C will assumedly decrease. The question is now, how this will affect trade between country A and a third country, B. Panel 1b illustrates the baseline setting for the triadic framework on which the remaining hypotheses are built. Again, decisions and actions taken by different actors play a role here. Country A's government can use instruments of trade policy to favor trade with country B, such as trade subsidies or reductions of trade barriers. Country A's companies will try to build up new trade relationships or extend their existing relationships with trade partners in country B to compensate for their loss of trade with country C. Individuals, here probably to a lesser extent, can contribute to trade shifting as well by turning their demand to imports from country B. Whether this leads to a significant increase in trade between country A and country B depends to a large extent on country B's behavior (or more precisely, on the behavior of country B's actors). From an economic point of view, it could be beneficial for country B to intensify its trade with country A. For consumers, this could extend the consumption variety and, in case of favorable trade policies, lower the prices. Companies could benefit from access to new markets and better prices too. Hence, for the government of country B, it makes sense to reciprocate the intensification of trade relations with country A. Beside economic arguments, also political and ideological decisions affect whether trade is shifted. Individuals and companies in country B as well as country B's government could simply want to support country A in its conflict with C. However, for similar reasons country B could also act the opposite way. Out of sympathy or alignment with country C, or out of enmity with country A, country B could take measures to interrupt its trade with country A, thereby increasing the harm that country A is already suffering through reduced trade with country C. How country B will react depends on the relationship between the three countries. I will introduce three different relationship characteristics that are important to determine how third country trade will develop. These are contiguity, formal security alliances, and common conflicts. Holding everything else

equal, however, I expect that a conflict between countries A and C will, due to the overall negative economic impact of conflict, reduce trade between A and B. Thus:

Hypothesis 1: If country A and country C have a violent interstate conflict with each other, bilateral trade between country A and country B will decrease.

3.2.1. Contiguity

Geographic distance has proven to be one of the most stable and reliable determinants of foreign trade. Gravity models of trade, in which geographic proximity and country sizes are the major determinants, have been applied in many different settings and modifications. The result that trade increases with shorter geographical distance between countries remains stable for most of these studies (Chaney, 2018).

There are different reasons why geographic proximity favors trade. Chaney (2018) argues that for international trade, there need to be stable networks of importers and exporters. Geographic distance affects these networks by two mechanisms: Direct costs of creating a foreign contact, and costs of direct interaction with an existing foreign contact. Both these costs are smaller over shorter distances and it is easier to set up trade networks in direct neighboring countries. For that reason, it seems likely that companies turn first to geographically close countries to set up new connections in order to shift trade. I therefore argue:

Hypothesis 2: If country A and country C have a violent interstate conflict with each other, country A will increase its trade with its direct neighbors (Panel 2 in Figure 1).

In this hypothesis, I distinguish between two different effects. First, a pure trade shifting effect, that is the stand-alone change in trade between country A and country B if country A and country C have a violent interstate conflict conditional on country A and B being neighbors. Second, the full effect of a violent interstate conflict between countries A and C on trade between countries A and B, conditional on countries A and B being neighbors. This covers the trade shifting effect and the stand-alone effect of the conflict. Both, the trade shifting effect and the full effect are

hypothesized to be positive.

3.2.2. Formal Security Alliances

Formal security alliances refer to alliances with an explicit security purpose, such as defense treaties. Members of these alliances commit each other to cooperate and support each other in security questions. It therefore seems likely that member states have a general interest in their partners' wellbeing. Often, alliances have more than only an interest in common security. Their members share similar values, norms, and interests, such as political systems or human rights. Thus, alliance members have an interest not only to support each other militarily, but also in other forms. Trade can be such a non-military support of an allied country. Some alliances even explicitly state economic cooperation as their goals alongside with security and defense (Long and Leeds, 2006). Trade with alliance members has moreover positive security externalities. In fact, international trade, following Liberalism and classical trade theory, leads to greater total wealth, and thus potentially increases the available resources for military spending, which is beneficial for the whole alliance (Long and Leeds, 2006). Feldman and Sadeh (2018) have moreover found that combatant states export more to third-party countries with higher similarity in their alliance portfolio. I therefore argue:

Hypothesis 3: If country A and country B are formal allies, they will increase their trade with each other if country A has a violent interstate conflict with country C (Panel 3).

Again, the hypothesized change applies for both, the trade shifting effect and the full effect, as for the remaining hypotheses as well.

A different situation occurs when country B and country C are formal allies (Panel 4). Feldman and Sadeh (2018) showed that combatant countries reduce exports to states with a high alliance similarity as their enemies. As explained above, alliance members have an interest in supporting each other. In this case, country B could support country C by harming country A. It could do so with trade embargos or other economic sanctions and measures against country A. Economic sanctions have become a quite common instrument of foreign policy and often replaced the use of military force (Caruso, 2003). Thus, I argue:

Hypothesis 4: If country B and country C are formal allies, country B will reduce its trade with country A if country A has a violent interstate conflict with country C (Panel 4).

3.2.3. Common Conflict

This argument follows a similar logic as the one in the previous section. If two countries are in a conflict with the same country (Panel 5), it might be in their common interest to support each other as a measure to harm their common enemy. An increase in bilateral trade would be a plausible consequence. A concern is, however, whether two conflict-affected countries have the capabilities to increase trade with each other. If both countries, A and B, are heavily affected in terms of destruction by their conflict with country C, it seems questionable whether companies in one of these countries would try to build up foreign contacts in a country that is similarly war-affected. Destruction of infrastructure, instability of institutions and the rule of law, or additional risk in war-affected zones – all these consequences are already heavy barriers to build up trade contacts if they occur in only one country. If both countries are affected by some of the aforementioned consequences, the difficulties will be even bigger. It might be easier to set up new contacts or intensify existing ones in a rather stable and peaceful country. This, again, depends on the type, location, and intensity of the conflict. Nevertheless, I argue:

Hypothesis 5: Trade between country A and country B will increase, if they are both engaged in a violent interstate conflict with country C (Panel 5).

3.3. Heterogeneity

With regards to the outlined hypotheses, I will test for two types of heterogeneous effects. First, in the type of trade flow, i.e. imports or exports. Second, in the size of the respective C-country. While in the above-stated hypotheses trade was described and treated as one aggregated type of flow, it is not obvious why imports and exports should behave the same with respect to reaction to conflict. It is, in fact, plausible to argue that the hypothesized effects will affect imports and exports to different extents. Presumably, imports might react stronger to the expected trade increase from

neighboring and allied countries (Hypotheses 2, 3, and 5 respectively), while, following the findings of Korovkin and Makarin (2021), exports react stronger to the expected trade decreases from the allies of the conflict country C and stronger to the overall negative conflict effect (Hypotheses 1 and 4). Export capacities are more affected by the conflict in general, as the risk of destruction and higher uncertainty about contract enforcement increase the transportation and transaction costs for exporters, raising exporter prices and reducing competitiveness (Long, 2008). On the other hand, the demand for imports might increase more (or decrease less) during conflict due to reduced domestic production capabilities.

For the second type of heterogeneity in the effect on trade, I will incorporate a GDP weight for the respective C-country. The argument behind this is that the impact of a conflict with a C-country has a differently strong impact depending on that C-country's size. A conflict with a very large country, say the US, is likely to have stronger impact on trade than a conflict with a relatively smaller country, say Belgium. Similarly, the hypothesized trade shifting effects from the different hypotheses are likely to have a different magnitude depending on the size of the C-country.

4. EMPIRICAL STRATEGY

4.1. Identification Strategy and Models

The baseline model for the analysis in this paper is a structural gravity model of trade, which is most commonly used and reliable in the international trade literature (Chaney, 2018; Head and Mayer, 2014), and particularly in the trade and conflict literature (e.g. Li and Sacko, 2002; Long, 2008; Keshk et al., 2010). The very basic idea of that model has been brought up by Tinbergen (1962) and is that bilateral trade is proportional to the product of any two countries' sizes (proxied here by their GDPs), divided by their geographical distance. This relationship is usually expressed as

$$Y_{ab} = \frac{\alpha X_a^{\beta_1} X_b^{\beta_2} H_{ab}}{X_{ab}^{\beta_3}} \quad (1)$$

where Y_{ab} represents the total trade between the countries A and B, X_a and X_b represent country A's and country B's GDP, respectively, and X_{ab} represents the distance between both countries. H_{ab} is

what is referred to as multilateral resistance terms (Anderson and van Wincoop, 2003; Head and Mayer, 2014). As the formulation is multiplicative, the β -parameters can be estimated taking the natural logarithm:

$$\ln(y_{ab}) = \ln(\alpha) + \beta_1 \ln(x_a) + \beta_2 \ln(x_b) - \beta_3 \ln(x_{ab}) + \ln(\eta_{ab}) \quad (2)$$

Since the interest lies in the effect of time-varying variables, i.e. conflict, parameter estimates will be obtained by using fixed effects estimation (Equation 3). Thereby, I control for observed and unobserved time-invariant country- and dyad-specific characteristics, including the multilateral resistance terms (Head and Mayer, 2014). In order to account for heterogeneity in the type of trade flow, the dependent variable in Equation 3, y_{abt} , represents either total imports of country A from country B, or total exports from country A to country B in year t in the respective specifications. X_{at} is a vector of country-A-specific covariates, X_{bt} is accordingly a vector of country-B-specific covariates. These country-specific covariates are country A's and country B's GDP, respectively, and their export intensities of oil and oil related products, and of diamonds, gems, and precious stones. X_{abt} is a vector of dyad-AB-specific covariates, including an indicator for violent conflict between A and B in order to test the baseline scenario. Furthermore, it includes an indicator for a formal security alliance between A and B, for geographic contiguity between A and B, for a common free trade agreement, and a measure of the polity-score difference between the two countries. Those country- and dyad-specific covariates will be discussed more in the remainder of this and in the following section. η_{ab} is a dyad-specific fixed effects term, covering time-invariant dyad-specific characteristics. Accordingly, μ_a and θ_b are country-A-specific and country-B-specific fixed effects terms, respectively, covering time-invariant country-A-specific and country-B-specific characteristics. In order to control for general macroeconomic trends and shocks, year dummies τ_t are included. Furthermore, trade from the previous period, y_{abt-1} , is included in order to control for dyad-specific time trends in trade.² Finally, ε_{abt} is the dyad-specific heteroscedasticity-consistent error term.

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (3)$$

² While in dynamic models with a lagged dependent variable structure there will always be an asymptotic bias in the within-group estimator, unless T tends to infinity and thus N/T tends to zero (Nickell, 1981; Alvarez and Arellano, 2003), this bias decreases and becomes reasonably negligible with large T.

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

One of the particularities that distinguishes this paper from others is that it incorporates a third-country dimension into a dyadic data set. Therefore, a set of additional variables containing C-country information are included (Equation 4). Most importantly, a conflict indicator $C_conflict_{at}$ is included, which indicates whether country A has a violent conflict with any third-party country C in year t . Using fixed effect estimation, this model exploits within-dyad variation of $C_conflict_{at}$ over time in order to measure the effect of a change in violent conflict with third-party C-countries from period $t-1$ to t on trade between countries A and B in period t . This corresponds to Hypothesis 1.

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (4)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

In order to measure Hypothesis 2, the interaction term $C_conflict_{at} * contiguity_{abt}$ is included in the model (Equation 5).³ Recall that $contiguity_{abt}$ is included in the vector of covariates X_{abt} , and thus included in the specification. This interaction term is the additional or joint effect of a conflict between A and C and the respective relationship variable, in this case geographic contiguity between A and B. The corresponding coefficient, ζ_1 , therefore depicts the trade shifting effect. Moreover, this interaction term is required to calculate the full effect that a violent conflict between countries A and C has on trade between A and B conditional on country A and country B being neighbors. The coefficient of the contiguity indicator between countries A and B would only provide the estimate for the effect of contiguity between those two countries, regardless of whether there is a conflict between countries A and C or not. The coefficient for the interaction term between conflict between A and C and contiguity between A and B corresponds to the additional effect of a conflict between countries A and C on trade between countries A and B, if the latter two are neighbors. The sum of the coefficient of the conflict between A and C, the stand-alone effect of that conflict, and the coefficient of the interaction, is then the full effect.

³ While in this equation I use the same coefficients α , β , and γ_1 as in Equation 4, the coefficient estimates will obviously differ in the different model specifications, as they vary in the independent variables included. For simplicity, though, I restrain from using additional subscripts to distinguish the coefficients for the same variables in the different model specifications.

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \zeta_1 (C_conflict_{at} * contiguity_{abt}) + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (5)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

Similarly, and in order to test Hypothesis 3, Equation 6 includes an interaction term between conflict between A and C and the indicator for a formal security alliance between A and B (which is included in the vector of covariates X_{abt}), $C_conflict_{at} * alliance_{abt}$. Again, the corresponding coefficient κ_1 represents the trade shifting effect, and the full effect can be calculated by adding up γ_1 and κ_1 .

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \kappa_1 (C_conflict_{at} * alliance_{abt}) + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (6)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

In order to test Hypothesis 4, two additional variables are required. First, an indicator for whether country B has a formal security alliance with any C-country, $C_alliance_{bt}$, similar to the conflict AC indicator $C_alliance_{at}$. Now for the trade shifting effect, one cannot just interact those two with each other. Consider country A having a conflict with a certain C-country, C_1 , and thus $C_alliance_{at} = 1$. At the same time, country B has an alliance with a *different* C-country, C_2 , so that $C_alliance_{bt} = 1$. Interacting those two, would obviously result in $C_conflict_{at} * C_alliance_{bt} = 1$. In Hypothesis 4, however, I make a claim about the effect of a conflict between A and C, if country B has an alliance with that exact *same* C-country. Therefore, in order to generate the appropriate variable for the trade shifting effect here, I generate a true triadic data set first.⁴ That is a data set, where the observation is a country triad, thus a combination of three different countries. With a total of 200 countries over a period of 46 years, this triadic data has $200 * 199 * 198 * 46 = 362,498,400$

⁴ Poast (2010, 2016) and Fordham and Poast (2016) point out the shortcomings of using dyadic data in international relations and in alliance formation in particular. They stress for the use of k -adic data, that is an observation is a combination of a certain set k countries, and demonstrate that approach using simulations and empirical data on alliance formation. However, a k -adic approach is challenging to implement into regression estimation due to the problem of repeated values in the dependent variable and thus repeated equations in the estimation.

observations. It contains information on every possible ABC relationship. From this data, I obtain all the variables with a C-dimension in my models by aggregating it to dyadic data. In order to generate $C_conflict_{at}$, for instance, I can simply aggregate $conflict_{act}$ to the dyad-year-level, so that

$$C_conflict_{at} = 1 \text{ if } \Sigma_C conflict_{act} > 0 \quad (7)$$

Similarly, $C_alliance_{bt}$ is constructed as

$$C_alliance_{bt} = 1 \text{ if } \Sigma_C alliance_{bct} > 0 \quad (8)$$

Their interaction was then obtained from the triadic data as

$$C_conflict_alliance_{abt} = 1 \text{ if } [\Sigma_C conflict_{act} * alliance_{bct}] > 0 \quad (9)$$

Thus, in order to make sure to capture the exact effect required to test the posed hypothesis, the interaction was in fact generated already in the triadic data, so that it can be aggregated to a dyadic variable following my data structure. Hypothesis 4 can then be tested with Equation 10, where λ_1 is the coefficient of interest.

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} \\ + \lambda_1 C_conflict_alliance_{abt} + \nu C_alliance_{bt} + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (10)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

Finally for Hypothesis 5, the interaction term was constructed in a similar way from the triadic data. That is, the interaction variable $C_conflict_conflict_{abt}$ was obtained as in Equation 11 so that it is equal to one if both country A and country B have a conflict with the *same* C-country. In line with the previous models, the coefficient of interest is then again ξ_1 as estimated in Equation 12. Equation 13 includes all relationship variables and interaction terms as outlined above in a full model.

$$C_conflict_conflict_{abt} = 1 \text{ if } [\Sigma_C conflict_{act} * conflict_{bct}] > 0 \quad (11)$$

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \xi_1 C_conflict_conflict_{abt} + \rho C_conflict_{bt} + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (12)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \zeta_1 (C_conflict_{at} * contiguity_{abt}) + \kappa_1 (C_conflict_{at} * alliance_{abt}) + \lambda_1 C_conflict_alliance_{abt} + \nu C_alliance_{bt} + \xi_1 C_conflict_conflict_{abt} + \rho C_conflict_{bt} + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (13)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

Besides country-A-specific, country-B-specific, and dyad-AB-specific fixed effects, capturing all time-invariant characteristics of country A, country B, and of country A and B as a dyad, a number of covariates is included in the estimation model in order to reduce the risk of omitted variable bias. This type of bias can arise if there are variables affecting both the regressand (trade between countries A and B) and the regressor (conflict between countries A and a C-country). Contiguity and security alliances between countries A and B are already included in order to test for Hypotheses 2 and 3. GDP is included as the above-mentioned proxy for country size in the gravity equation. The rationale behind it is that the size of an economy (as measured by GDP) increases a country's potential to trade. At the same time, GDP could affect the military capabilities of a country, which in turn might affect the likelihood of being engaged in a violent conflict. A measure of similarity in the political system, that is the difference in democracy scores of two countries, is included to capture additional attitudes and international policy preferences towards each other. A variable indicating whether countries A and B are members of the same free trade agreement (bilateral and multilateral) is included too. Free trade agreements between countries positively affect trade between countries (Baier and Bergstrand, 2007), but could potentially also affect the likelihood of conflict. On the one hand, increased interaction and negotiations through and for the free trade agreement could also make negotiations on other topics more likely, potentially improving diplomatic ties and reducing the likelihood of a conflict (Mansfield and Pevehouse, 2000). On the

other hand, countries with a more intense conflict history are more likely to sign trade agreements (Martin et al., 2012). Apart from relationship variables, bias due to unobserved heterogeneity could stem from omitted country-A-specific variables, since country A is involved in both trade between A and B and conflict between A and C. Numerous country characteristics are already covered by the fixed effect term (all time-invariant observed and unobserved country characteristics). Thus, only time-varying country-specific factors that can affect both foreign politics and economics of a country can pose a problem as omitted variables. The overall economic performance and capabilities of an economy are controlled for by GDP. The involvement in international political and economic organizations, regimes, and treaties, is further controlled for by the different above-explained relationship variables. However, different types of economic and political shocks could potentially play a role here. One such factor that is often brought in connection with conflicts is the discovery or abundance of natural resources (e.g. Berman et al., 2017). If country A experiences such a discovery (of significant size), could this potentially increase trade flows with other countries, both in terms of raw materials but also of processed or manufactured products. At the same time, this could make country A more attractive as a target for a foreign aggression by country C, since the prospect of natural resource extraction could be an attractive gain. In reality though, only few violent interstate conflicts were actually caused, or partly caused, by natural resources. The vast majority of evidence on the effect of natural resources on the likelihood of conflict is limited to inner-state conflicts. Caselli et al. (2013), however, find that also interstate conflicts can be fueled by natural resources (in particular by oil fields), if they are located sufficiently proximate to the state border. To proxy for the abundance of natural resources, the country specific control vectors include measures for the amount of oil and oil related exports as well as of exports of diamonds, pearls, and precious stones, which are the most common natural resources related to conflict in the literature (e.g. De Soysa and Neumayer, 2007, or Sorens, 2011).

Another typical concern in the field of conflict and trade studies is the risk of reverse causality, thus the risk that it is actually trade that determines the likelihood of conflict, rather than the other way around. The large body of literature studying this direction (as outlined in the Section 2), gives support to this concern. These previous studies, however, exclusively study the dyadic case without a third-party country dimension. In that framework, it seems plausible that the causality could go in both directions: A violent conflict between countries A and B has an effect on trade between

A and B; or trade between countries A and B has an effect on the likelihood of violent conflict between countries A and B. In the framework of this paper, however, this logic is not as straight forward. While I argue that a conflict between country A and a third-party country C will have an effect on trade between countries A and B, reverse causality would in this case imply that trade between countries A and B has an effect on the likelihood of conflict between countries A and C. The argument of the IPE literature (e.g. Polachek, 1980) that the potential loss of trade benefits would restrain countries from entering a conflict, does not hold here: For country A to enter a conflict with C would not necessarily reduce the trade ties with B (except for the general negative effect a conflict has on the economy of the country, which should, however, restrain a country from entering *any* conflict then). There is therefore no reason to believe that trade ties between A and B should cause a reduced (or increased) likelihood of a conflict between A and C. Moreover, country C is not involved in the bilateral trade between AB, thus, there is even less reason to believe that trade between A and B should influence the likelihood of country C entering a conflict with country A.

Finally, bias could arise due to systematic measurement error. The concern is that trade could systematically contain a higher number of missing values during conflicts. Especially in high intensity conflicts, it is plausible that official trade records are not taken or are less reliable. If this was the case, a number of observations containing a conflict would be missing from the data set, and thus the coefficient estimates would be biased. While the effect of the violent conflict between countries A and C would likely be downward biased, since it is the high intensity conflicts that inflict the highest damage to an economy and are more likely to cause missing trade data, the direction of bias for the trade shifting effect, thus the interaction term, would remain unclear. It is the high intensity conflicts that would generate a greater urge to shift trade to other countries. But at the same time, it is also the high intensity conflicts that reduce the overall trade capabilities of an economy the most. The bias could thus go into both directions. While this risk of a bias due to measurement error remains, and the possibility of omitted variable bias cannot be entirely ruled out, the above mentioned strategy using C-country variation and country- and dyad-specific covariates and fixed effects, makes me confident that a causal effect is identified.

As stated in Section 3.3, I will not only look at heterogeneity in terms of the type of trade flow (i.e. imports and exports), but also in terms of the size of the respective C countries. In order to

test whether the GDP size of country C matters for the tested trade shifting effects, the respective interaction terms are weighted by country C's GDP. As the C-country dimension is involved for this, the required variables are generated in the true triadic data, similar as the other C-country variables. For Hypothesis 1, where just the effect of conflict between A and C is measured, a weighted measure for conflict between A and C is constructed as

$$C_conflict_gdp_share_{at} = \Sigma_C [conflict_{act} * GDP_{ct}] \quad (14)$$

where GDP_{ct} is measured as country C's GDP as a share of world GDP in year t . The share of GDP of all countries that country A has a violent conflict with in year t is then given by $C_conflict_gdp_share_{at}$. This weighted conflict measure, additionally to the conflict dummy, is then included in Equation 4. In Equation 15, γ_2 then measures how the effect of a violent conflict between countries A and C varies with the GDP size of those C countries, controlled for whether there is a violent conflict between A and C.

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \gamma_2 C_conflict_gdp_share_{at} + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (15)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

For Hypothesis 2, both $C_conflict_{at}$ and $C_conflict_gdp_share_{at}$ are then interacted with $contiguity_{abt}$ in order to measure how the size of country C's GDP affects the trade shifting effect, given that countries A and C have a conflict and that countries A and B are neighbors (Equation 16). The former will simply be a binary indicator, and thus ζ_1 will measure the effect of whether countries A and C have a conflict and countries A and B are neighbors. The later will be a continuous measure, and ζ_2 will depict how country C's size will affect trade between A and B if the two are neighbors and country A and C have a violent conflict.

$$y_{abt} = \alpha + \beta_a X_{at} + \beta_b X_{bt} + \beta_{ab} X_{abt} + \delta y_{abt-1} + \gamma_1 C_conflict_{at} + \zeta_1 (C_conflict_{at} * contiguity_{abt}) + \zeta_2 (C_conflict_gdp_share_{at} * contiguity_{abt}) + \eta_{ab} + \mu_a + \theta_b + \tau_t + \varepsilon_{abt} \quad (16)$$

$$y_{abt} \in \{import_{abt}, export_{abt}\}$$

For Hypothesis 3 (alliance between countries A and B), the model is set up in the same way. For Hypotheses 4 and 5, the C-country GDP share is again constructed separately in the triadic data, as I want a measure for the GDP share of only those C-countries that have a conflict with country A *and* are allies with country B (Hypothesis 4) or in conflict with country B (Hypothesis 5). Those are generated by Equations 17 and 18. The estimation models are then set up equivalently as in Equation 16.

$$C_conflict_gdp_share_{at}^{Hyp4} = \sum_C [conflict_{act} * alliance_{bct} * GDP_{ct}] \quad (17)$$

$$C_conflict_gdp_share_{at}^{Hyp5} = \sum_C [conflict_{act} * conflict_{bct} * GDP_{ct}] \quad (18)$$

4.2. Data

For this paper, a country-dyadic data set with a third-party country dimension is used, where an observation is a pair of countries in a certain year. Given the total number of countries in the data (all sovereign states, as listed by the United Nations, plus some quasi-sovereign but not fully recognized states and some former states), the potential number of country combinations is 39,800 directed country dyads. For the studied period from 1968 to 2013, this adds up to 1,830,800 dyad-year observations. However, given that some of the following data sources are not complete, the panel is highly unbalanced. Constraining this to a balanced panel will reduce the number of observations to 543,040 (if a basic set of covariates is used) and to 475,130 (if the full set of covariates is used).

The particularity of this paper is that it adds a third-party country (or C-country) dimension to the dyadic data. This C-country dimension is obtained from a triadic data set. That is a data set where an observation is a triplet of countries in a certain year. Given the total number of countries and years studied, this adds up to a potential number of 362,498,400 triad-year observations in the studied period. Certain C-country variables are then obtained by aggregating this triadic data set to a dyadic one, as explained in the previous section.

The crucial data for this analysis are information on conflict and on trade. For conflict, I use the data of the International Crisis Behavior (ICB) Project (Brecher et al., 2016), which provides a large variety of information about interstate conflicts on a dyadic level and covers a period of 96

years (1918 to 2013), of which I will utilize the latest 46 years.⁵ The total number of dyad-year observations with conflict in the ICB data set for this period is 545, of which 165 are non-violent conflicts. This leaves a share of 5.80 percent of the dyad-year observations of the balanced data set with a violent interstate conflict with a C-country (see Table 1). The remaining violent conflicts consist of three different intensities, of which I am using the lowest one as the threshold to define a violent conflict dummy. That is, a violent conflict is defined by any use of directed kinetic force by one country with the potential to kill people in the other one.

The trade data comes from the Correlates of War (COW) database (Barbieri and Keshk, 2016). The COW data covers imports and exports over the period from 1870 to 2014 and is based on the Directions of Trade Statistics of the International Monetary Fund (IMF, 2017), complemented with data from the UN and from national statistic bureaus. All sovereign as well as some former and quasi-sovereign states are included. The data is far from being complete. Many countries do not have trade data reported for all years and/or all partner countries. The total number of dyad year observations is 886,828 (678,116 in the studied period 1968-2013). In some dyads, there is only one flow – import or export – reported, while the other is missing. In some of these cases, however, the missing trade flow is reported in the opposite direction, and thus the missing information could be filled in (for instance exports from country A to country B are missing, but imports from country A into country B are reported). In some other dyads, both import and export data are missing, but at least one of them is reported in the opposite direction. It was again possible to fill up some missing values with reported values. This lifts up the number of dyadic observations with both import and export data reported to 688,828 in the studied period. As the dependent variables, the natural

⁵ Two other conflict databases that are often used in research on trade and conflict are the Correlates of War (COW) Project, and the Uppsala Conflict Data Program (UCDP). The main disadvantage of the COW conflict data is that it does not come in a clear dyadic format. In conflicts with more than two states involved, it is often not clear who is actually fighting whom. Using this data would therefore bear the risk of reporting violent conflicts between countries who were not actually violently involved with each other. For instance, many South and Central American states declared war on Japan or Germany during World War II, while most of them were not actually involved in any combat action. Others declared war to all of the Axis powers, but were only involved in combats with some certain states. The UCDP data has a similar flaw. It does not come in a clear dyadic format either. As this dataset includes non-state actors, other states are often only listed as supporters, without distinguishing between military or just political or verbal support. Other conflict data-bases cover shorter periods (Conflict and Peace Data Bank (COPDAB), Data on Armed Conflict and Security (DACS), World Event/Interaction Survey (WEIS)), are limited to certain regions (Armed Conflict Location Event Data Project (ACLED)), or do not provide their full data in an accessible data format (Heidelberg Institute for International Conflict Research (HIK)).

logarithm⁶ of imports of country A from country B and of exports from country A to country B in year t , expressed in millions of 2010 US dollars, is then used.

As this paper draws upon a gravity model of international trade, data on country sizes, proxied by the natural logarithm of GDP, is required. I use the data from the World Development Indicators (WDI) databank of the World Bank (2017), where GDP is reported in millions of 2010 US dollars and covers the period from 1968 on. The data that is required for the formation of the different relationship constellations, i.e. geographic contiguity and formal military alliances, is extracted from the COW project too (Correlates of War Project, 2016; Stinnett et al., 2002; Gibler, 2009). Based on that, dummy variables are created which denote whether two countries are geographical neighbors in the broadest sense, or whether there is any form of formal security-related alliance between two countries.

The measure of similarity of the political system of two countries is based on the Polity IV data (Integrated Network for Societal Conflict Research, 2018), which provides a score on how democratic or autocratic a country is. The data ranges from (depending on the country) 1800 to 2015, and the score ranges from -10 (maximum autocratic) to 10 (maximum democratic). The similarity is calculated by subtracting the lower democracy score from the higher democracy score within a dyad. The minimum distance is 0 (same democracy score), the maximum is 20. Data on free trade agreements comes from the Design of Trade Agreements Database (Baccini et al., 2014), and covers data on bilateral and multilateral trade agreements over the period of 1945 to 2019. Finally, data on natural resources exports is extracted from the Observatory of Economic Complexity (OEC) (Hidalgo and Simoes, 2011), which contains bilateral and unilateral import and export data from 1962 to 2017 on a 4-digit product level. The OEC data is based on the data from The Center for International Data and from the UN COMTRADE database. The oil variable is then created by taking the natural logarithm of the sum of exports of petroleum and petroleum related products (SITC4 Rev. 2 codes 3330 – 3354). The stone variable is generated in the same way by using exports of worked and unworked pearls, and precious and semi-precious stones (SITC4 Rev. 2 codes 6671 – 6674), which covers diamonds and other precious stones.

⁶ As taking the logarithm of imports and exports in the presence of zero trade flows leads to missing values in the data, I re-run the main analyses with $\log(1+\text{imports})$ and $\log(1+\text{exports})$ as the dependent variables. While magnitudes change, the direction and precision of the main estimates remain mostly unchanged, as can be seen in the Tables A1, A2, and A3 in the appendix.

Table 1: Summary Statistics

Variable	Description	Scale	N	Mean	SD	Min	Max
log imports	log of imports (in millions of 2010 US dollars) of country A from country B	continuous	475,130	1.951	3.554	-22.06	12.99
log exports	log of exports (in millions of 2010 US dollars) from country A to country B	continuous	475,130	1.951	3.554	-22.06	12.99
conflict AB	violent interstate conflict between countries A and B	binary	475,130	0.0005	0.022	0	1
conflict AC	violent interstate conflict between countries A and C	binary	475,130	0.0580	0.234	0	1
conflict BC	violent interstate conflict between countries B and C	binary	475,130	0.0576	0.233	0	1
contiguity AB	geographic contiguity between countries A and B	binary	475,130	0.0554	0.229	0	1
alliance AB	formal security alliance between countries A and B	binary	475,130	0.0997	0.300	0	1
alliance BC	formal security alliance between countries B and C	binary	475,130	0.7358	0.441	0	1
FTA AB	preferential trade agreement between countries A and B	binary	475,130	0.0108	0.103	0	1
polity AB	polity score difference between countries A and B	ordinal	475,130	7.6699	6.384	0	20
GDP A	log of GDP (in millions of 2010 US dollars) of country A	continuous	475,130	11.1793	2.068	5.10	16.58
GDP B	log of GDP (in millions of 2010 US dollars) of country B	continuous	475,130	11.0341	2.139	5.10	16.58
oil A	log of total exports (in current US dollars) of petroleum and related products of country A	continuous	475,130	9.3862	9.854	0.00	26.45
oil B	log of total exports (in current US dollars) of petroleum and related products of country B	continuous	475,130	9.3375	9.817	0.00	26.45
stones A	log of total exports (in current US dollars) of diamonds, gems, and precious stones of country A	continuous	475,130	6.8512	8.083	0.00	23.97
stones B	log of total exports (in current US dollars) of diamonds, gems, and precious stones of country B	continuous	475,130	6.7056	8.028	0.00	24.21
conflict AC x contiguity AB	interaction term between conflict AC and contiguity AB	binary	475,130	0.0034	0.058	0	1
conflict AC x alliance AB	interaction term between conflict AC and alliance AB	binary	475,130	0.0065	0.080	0	1
conflict AC x alliance BC	interaction term between conflict AC and alliance BC	binary	475,130	0.0049	0.070	0	1
conflict AC x conflict BC	interaction term between conflict AC and conflict BC	binary	475,130	0.0031	0.055	0	1

Table 1 provides summary statistics for the variables used in the baseline estimation models. It also includes the relevant interaction terms for Hypotheses 2 to 5. The constellation tested in Hypothesis 2, namely that country A has a violent conflict with country C and is neighbors with country B, occurs in 0.34 percent of the of the dyad-year observations. Country A having a violent conflict with country C while being in a security alliance with country B (Hypothesis 3) occurs in 0.65 percent of the observations. For Hypothesis 4, country A and country C having a violent conflict while countries B and C are neighbors, the share is 0.49 percent. And for the final case in Hypothesis 5, where both countries A and B have a violent conflict with country C, the share 0.31 percent.

5. RESULTS

5.1. Baseline Results

Table 2: Estimates of Dyadic and Third-Party Country Conflicts

	(1)	(2)	(3)	(4)	(5)	(6)
	log import 1968-2013	log import 1968-2013	log import 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013
conflict AB	-0.7571*** (0.1142)	-0.8231*** (0.1221)	-0.8180*** (0.1223)	-0.7571*** (0.1142)	-0.8231*** (0.1221)	-0.8131*** (0.1221)
conflict AC			-0.0211*** (0.0079)			-0.0398*** (0.0082)
Observations	543,040	475,130	475,130	543,040	475,130	475,130
Year FE	yes	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes	yes
Controls	basic	full	full	basic	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use GDP and contiguity AB as covariates. Columns 2, 3, 5, and 6 use additionally alliance AB, FTA AB, polity AB, as well as the oil and diamond covariates for A and B. Columns 1 to 3 include imports lagged by one year, Columns 4 to 6 exports lagged by one year. Columns 1, 2, 4, and 5 are estimated using Equation 3, Columns 3 and 6 are estimated using Equation 4. Fixed effects, covariates, R-squared, and constants are not reported.

Table 2 displays the results from Equations 3 and 4. Columns 1 and 2 show the results for imports as the dependent variable without any C-country dimension. The estimation reports a negative and statistically highly significant coefficient for *conflict AB*. The estimated coefficients of -0.7571 (if only GDP and contiguity are used as covariates) and -0.8231 (if the full set of covariates is used⁷) correspond to reductions in imports of 53.10 and 56.09 percent, confirming what has been found in the literature before. The magnitudes lie closer to what has been estimated by Long (2008), with trade reductions of more than 70 percent, or Feldman and Sadeh (2018), with reductions of 64 percent, than by Li and Sacko (2002; between 6 and 12 percent) or Keshk et al. (2010; between

⁷ For better readability, most of the covariates are not displayed in Tables 2, 3, and 5. Only those that are directly relevant to the outcomes of the different hypotheses are shown. Tables A4, A5, and A6 in the appendix show the same tables with all covariates displayed.

4 and 10 percent). Columns 3 and 4 show the same estimation with exports as the dependent variable. As there is no C-country dimension yet in these estimations, the results for exports are just the mirrored results of imports and the estimates are in fact the same. For Columns 3 and 6, the C-country dimension is added and Equation 4 is used. Both columns report the respective coefficient for conflict AC, which is the indicator variable $C_conflict_{at}$ from Equation 4. Both show a negative and statistically significant effect of *conflict AC* on the dependent variable. This corresponds to a reduction in imports of 2.11 percent (Column 3) and of 3.98 percent in exports (Column 6). The magnitudes are, unsurprisingly, weaker than for *conflict AB*. This is also in line with what has been found by Feldman and Sadeh (2018), even though their estimates are larger (between 14 and 23 percent), potentially due to the difference in the sample period. These results confirm Hypothesis 1, namely that a violent interstate conflict between countries A and C reduces bilateral trade (both imports and exports) between countries A and B. The results further show that country A's exports appear to decrease more than its imports.

Table 3 shows the results for testing the remaining four hypotheses on imports as the dependent variable. All columns use the full set of covariates as well as year, country, and dyad fixed effects. Column 1 uses Equation 5, thus includes the interaction term between *conflict AC* and *contiguity AB* in order to test Hypothesis 2. The coefficient for *conflict AC* continues to be negative and statistically significant (slightly weaker though in terms of magnitude and precision compared to Column 3 of Table 2). The coefficient estimate for the interaction term between *conflict AC* and *contiguity AB* is negative and statistically insignificant. Similarly, in Column 5, where all relevant interaction terms are included in a full model and thus Equation 13 is used, the interaction coefficient is negative and statistically insignificant too. Thus, different than hypothesized, country A does not appear to import more from its neighbors when it has a conflict with country C.

In Column 2, the interaction between *conflict AC* and *alliance AB* is included, using Equation 6. The coefficient estimate is positive and statistically significant (even though only weakly, at the 10 percent level), corresponding to an increase in imports of 3.90 percent. This confirms Hypothesis 2 regarding the trade shifting effect: Country A increases imports from its allies if it has a conflict with country C. This trade shifting effect appears to be larger than the stand-alone negative effect of *conflict AC*, thus the full effect of *conflict AC* conditional on countries A and B being allies is an increase in imports by 1.42 percent (see Table 4, Column 2), even though an F-test for joint

significance has to be rejected. In the full model (Table 3, Column 5), the trade shifting effect for Hypothesis 3 appears to be larger both in terms of magnitude and precision, representing an increase in imports by 6.29 percent. The full effect is stronger too, with an increase in imports of 4.01 percent (Table 4, Column 4), jointly significant at the 10 percent level.

Table 3: Estimates of Trade Shifting Effects on Imports

	(1) log import 1968-2013	(2) log import 1968-2013	(3) log import 1968-2013	(4) log import 1968-2013	(5) log import 1968-2013
conflict AC	-0.0206** (0.0082)	-0.0251*** (0.0084)	-0.0139* (0.0081)	-0.0265*** (0.0081)	-0.0231*** (0.0087)
conflict AC x contiguity AB	-0.0078 (0.0247)				-0.0184 (0.0258)
conflict AC x alliance AB		0.0382* (0.0199)			0.0610*** (0.0213)
conflict AC x alliance BC			-0.0858*** (0.0275)		-0.1051*** (0.0282)
conflict AC x conflict BC				0.0910*** (0.0228)	0.0918*** (0.0236)
Observations	475,130	475,130	475,130	475,130	475,130
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 3 and 5 additionally use alliance BC, Columns 4 and 5 use conflict BC. Column 1 is estimated using Equation 5, Column 2 using Equation 6, Column 3 using Equation 10, Column 4 using Equation 12, and Column 5 using Equation 13. Fixed effects, covariates, R-squared, and constants are not reported.

Column 3 of Table 3 uses Equation 10 and shows the results for testing Hypothesis 4. The coefficient estimate for *conflict AC* is weaker in terms of magnitude and significance than in the previous models, yet still negative and weakly statistically significant. The trade shifting effect, thus the coefficient estimate for the interaction term *conflict AC x alliance BC*, is negative, statistically significant and large in terms of magnitude. The coefficient size is -0.0858, corresponding to a

decrease in imports by 8.22 percent. Adding this to the stand-alone effect of *conflict AC*, the full effect appears to be even stronger: Country A reduces its imports from country B by 9.61 percent if it has a violent interstate conflict with country C *and* if countries B and C are allies (see Table 4, Column 2). This confirms Hypothesis 4, showing that countries reduce imports from the allies of the countries they have a conflict with. Both the trade shifting and the full effect are even stronger in full model (Column 5), amounting in a total reduction in imports of 12.26 percent (Table 4, Column 4). The full effect is statistically significant both in the single model and in the full model.

Table 4: Effect Sizes on Imports in %

	(1) log import Single Model Trade Shifting	(2) log import Single Model Full Effect	(3) log import Full Model Trade Shifting	(4) log import Full Model Full Effect
conflict AC x contiguity AB	-0.78 (0.025)	-2.81 (0.024)	-1.82 (0.026)	-4.11 (0.026)
conflict AC x alliance AB	3.90* (0.020)	1.42 (0.018)	6.29*** (0.021)	4.01* (0.021)
conflict AC x alliance BC	-8.22*** (0.028)	-9.61*** (0.027)	-9.98*** (0.028)	-12.26*** (0.028)
conflict AC x conflict BC	9.53*** (0.023)	6.92*** (0.022)	9.62*** (0.024)	7.33*** (0.024)

Notes: The displayed numbers are the percentage changes in imports corresponding to the coefficient estimates from Table 3. Columns 1 and 2 of this table refer to the coefficient estimates from Columns 1-4 of Table 3, and Columns 3 and 4 of this table refer to the coefficient estimates of Column 5 of Table 3. Columns 2 and 4 of this table report the sum of the coefficient estimates of *conflict AC* in the respective columns and of the respective interaction terms. Standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). For Columns 2 and 4, the significance levels are based on an F-statistic for joint significance.

Finally, Column 4 of Table 3 uses Equation 12 and tests the case of common conflicts. The included interaction term is between *conflict AC* and *conflict BC*. The coefficient estimate is positive, statistically significant, and high in terms of magnitude, corresponding to an increase in imports of 9.53 percent, thus confirming Hypothesis 5. As the positive trade shifting effect is stronger than the negative stand-alone effect of *conflict AC*, the full effect is positive too, with an increase in imports of 6.92 percent (see Table 4, Column 2). Thus, if country A and country B both have a violent

interstate conflict with the *same* country C, country A increases its imports from that country B. In the full model (Table 3, Column 5), the trade shifting and the full effect are slightly stronger, with increases in imports of 9.62 and 7.33 percent, respectively. Again, the full effect is statistically significant in both the single and the full model.

Table 5 now turns to exports as the dependent variable, testing the different Hypotheses separately and together in a full model. Again, all columns include the full set of covariates and use time, country, and dyad fixed effects. The *conflict AC* coefficient is negative and statistically significant in all columns. As already indicated in Table 2, the effect appears to be stronger for exports than for imports, with reductions in exports of between 3.00 and 4.37 percent, as compared to reductions in imports of 1.39 to 2.61 percent. Column 1 of Table 5 shows the results for testing Hypothesis 2. The coefficient estimate for the interaction term between *conflict AC* and *contiguity AB* is negative and statistically insignificant. Thus, as already shown for imports, exports to neighboring countries do not increase as a result of violent interstate conflicts with C-countries, and Hypothesis 2 can accordingly not be confirmed. This result remains the same in the full model (Column 5).

In Column 2, the interaction between *conflict AC* and *alliance AB* is included. While this interaction had a positive and weakly significant coefficient estimate on imports as the dependent variable (see Table 3, Column 2), the estimate on exports is negative and statistically not significant. Hypothesis 3 can thus not be confirmed for exports. In other words, while countries import more from their allies if they have a violent interstate conflict with a third-party country C, they do not export more to those allies, contrary to what has been found by Feldman and Sadeh (2018). Again, this result does not change in the full model (Table 5, Column 5).

The estimate of the interaction term between *conflict AC* and *alliance BC* (Column 3) is negative and statistically significant. The coefficient of -0.1083 corresponds to a decrease in exports of 10.27 percent. The full effect of *conflict AC* conditional on countries B and C being allies is a decrease in exports from country A to country B by 13.27 percent and is statistically significant (see Table 6, Column 2). In the full model (Table 5, Column 5) both the trade shifting effect as well as the stand-alone conflict effect are stronger. Accordingly, the full effect is stronger too, with a decrease in exports of 14.09 percent (Table 6, Column 4). Those results give strong support for Hypothesis 4: Countries do decrease exports to the allies of the countries they are in conflict with, in line with previous findings of Feldman and Sadeh (2018). Note that both the trade shifting as well as the full

Table 5: Estimates of Trade Shifting Effects on Exports

	(1)	(2)	(3)	(4)	(5)
	log export 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013
conflict AC	-0.0377*** (0.0086)	-0.0379*** (0.0088)	-0.0305*** (0.0085)	-0.0447*** (0.0085)	-0.0347*** (0.0092)
conflict AC x contiguity AB	-0.0340 (0.0259)				-0.0232 (0.0269)
conflict AC x alliance AB		-0.0176 (0.0194)			0.0043 (0.0204)
conflict AC x alliance BC			-0.1083*** (0.0263)		-0.1130*** (0.0268)
conflict AC x conflict BC				0.0910*** (0.0228)	0.1031*** (0.0237)
Observations	475,130	475,130	475,130	475,130	475,130
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 3 and 5 additionally use alliance BC, Columns 4 and 5 use conflict BC. Column 1 is estimated using Equation 5, Column 2 using Equation 6, Column 3 using Equation 10, Column 4 using Equation 12, and Column 5 using Equation 13. Fixed effects, covariates, R-squared, and constants are not reported.

effect are stronger on exports than on imports.

Column 4 of Table 5, finally, shows the results for testing Hypothesis 5, thus the case where both country A and country B have a violent interstate conflict with the same country C. The interaction term *conflict AC x conflict BC* is positive and statistically significant. Its size corresponds to an increase in exports by 9.52 percent. In the full model (Column 5), this trade shifting effect is even larger, with an increase in exports by 10.86 percent. Those positive trade shifting effects are larger than the negative stand-alone effects of *conflict AC*, thus the full effect is positive both in the single (5.15 percent increase in exports) and in the full model (7.45 percent), and is statistically significant in both. Thus, Hypothesis 5 can be confirmed for exports as the dependent variable too: Country A's exports to country B increase if they both have a violent interstate conflict with the same country

C. The trade shifting effect appears to be slightly stronger for exports than for imports. However, since the stand-alone negative effect of *conflict AC* is stronger for exports as well, the full effect does not differ much between imports and exports.

Table 6: Effect Sizes on Exports in %

	(1) log export Single Model Trade Shifting	(2) log export Single Model Full Effect	(3) log export Full Model Trade Shifting	(4) log export Full Model Full Effect
conflict AC x contiguity AB	-3.34 (0.026)	-7.04*** (0.025)	-2.30 (0.027)	-5.70** (0.027)
conflict AC x alliance AB	-1.75 (0.019)	-5.47*** (0.018)	0.43 (0.020)	-2.98 (0.020)
conflict AC x alliance BC	-10.27*** (0.026)	-13.27*** (0.025)	-10.69*** (0.027)	-14.09*** (0.027)
conflict AC x conflict BC	9.52*** (0.023)	5.15** (0.022)	10.86*** (0.024)	7.45*** (0.024)

Notes: The displayed numbers are the percentage changes in exports corresponding to the coefficient estimates from Table 4. Columns 1 and 2 of this table refer to the coefficient estimates from Columns 1-4 of Table 4, and Columns 3 and 4 of this table refer to the coefficient estimates of Column 5 of Table 4. Columns 2 and 4 of this table report the sum of the coefficient estimates of *conflict AC* in the respective columns and of the respective interaction terms. Standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). For Columns 2 and 4, the significance levels are based on an F-statistic for joint significance.

5.2. Additional Specifications

This section will show the results regarding heterogeneity in the size of the respective C-countries, as outlined in Sections 3.3 and 4.1. Table 7 shows the results for the five different hypotheses with imports as the dependent variable and with the *GDP share C* measure included. Column 1 shows the results for the baseline case without any additional interaction terms, meaning that it measures the effect of *conflict AC* (using Equation 16). Unlike in main results (see Tables 2, 3, and 4), the coefficient estimate of *conflict AC* is statistically insignificant. The *GDP share C* estimate, however, is negative and statistically significant. Note that *GDP share C* only covers the GDP of countries that country A has a violent conflict with. Thus, the two different coefficients can be interpreted as

the extensive margin (*conflict AC*, indicating *whether* there is a conflict between countries A and C) and the intensive margin (*GDP share C*, indicating *the size* of that C-country). As for Column 1, the intensive margin seems to be the relevant effect, as the coefficient of *GDP country C* captures most of the negative relationship between *conflict AC* and imports between A and B.⁸

In Column 2, the interaction between *conflict AC* and *contiguity AB* is included. As in the main results, this interaction has a statistically insignificant coefficient estimate. *GDP country C* has a negative and insignificant coefficient estimate too. Column 3 shows the result for testing Hypothesis 3. In the baseline model (without *GDP share C* included), the interaction of *conflict AC* and *alliance AB* had a positive and statistically significant coefficient estimate (see Table 3, Column 2). Now, with *GDP share C* included, the interaction term coefficient remains positive and becomes even larger in magnitude and more precisely estimated. *GDP share C* itself, however, is statistically not significant in this estimation. Thus, in the presence of the positive trade shifting effect to allied B-countries, the size of the conflict country C does not impede or amplify this effect.

Column 4 of Table 7 includes the interaction between *conflict AC* and *alliance BC* as well as *GDP share C*. Similar as in Column 1, the negative trade shifting effect that was estimated in the model without *GDP share C* (Table 3, Column 3), seems to be covered to a large degree by country C's size now: The coefficient estimate for *conflict AC x alliance BC* is negative but statistically not significant, while the estimate for *GDP share C* is negative and highly statistically significant. Thus, imports of country A from country B decrease more the higher country C's GDP is, if country A and country C have a violent interstate conflict with each other and country B and country C are allies. Finally, Column 5 of Table 7 includes the interaction term between *conflict AC* and *conflict BC*. As in the results without *GDP share C* (Table 3), the coefficient estimate for the interaction is positive, statistically highly significant and strong in magnitude, even stronger here than in Table 3. The coefficient estimate for *GDP share C* is highly negative and statistically weakly significant. That means that while there is a positive trade shifting effect for imports of country A from country B if both countries have a conflict with the same country C, and while this effect is large enough to outweigh the negative stand-alone effect of *conflict AC*, it decreases with the size of country C's

⁸ Note that the GDP share measure is not centered, as the interaction between country C's GDP and violent conflict between A and C takes place before the aggregation from the triadic to the dyadic data. The variable can thus not simply be centered as usually done. The magnitudes in Table 7 and 8 should thus be interpreted with caution.

Table 7: Estimates of Trade Shifting Effects on Imports, GDP C Weighted

	(1)	(2)	(3)	(4)	(5)
	log import 1968-2013	log import 1968-2013	log import 1968-2013	log import 1968-2013	log import 1968-2013
conflict AC	-0.0072 (0.0080)	-0.0206** (0.0082)	-0.0251*** (0.0084)	-0.0140* (0.0081)	-0.0265*** (0.0081)
GDP share C	-1.2998*** (0.2018)				
conflict AC x contiguity AB		0.0023 (0.0255)			
GDP share C		-0.8047 (0.6202)			
conflict AC x alliance AB			0.0452** (0.0197)		
GDP share C			-0.4570 (0.4557)		
conflict AC x alliance BC				-0.0441 (0.0288)	
GDP share C				-0.9221*** (0.2977)	
conflict AC x conflict BC					0.1184*** (0.0271)
GDP share C					-24.1379* (14.0877)
Observations	475,130	475,130	475,130	475,130	475,130
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 4 additionally uses alliance BC, Column 5 uses conflict BC. Column 1 is estimated using Equation 15, Column 2 using Equation 16, Column 3 using Equation 6 with Equation 14 included additionally, Column 4 using Equation 10 with Equation 17 included additionally, and Column 5 using Equation 12 with Equation 18 included additionally. Fixed effects, covariates, R-squared, and constants are not reported.

GDP.

Table 8 shows the results for the analysis with exports as the dependent variable. Unlike as for imports, one can see in Column 1 that for exports both the extensive and intensive margin matter.

Table 8: Estimates of Trade Shifting Effects on Exports, GDP C Weighted

	(1)	(2)	(3)	(4)	(5)
	log export 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013
conflict AC	-0.0314*** (0.0083)	-0.0378*** (0.0086)	-0.0379*** (0.0088)	-0.0306*** (0.0085)	-0.0447*** (0.0085)
GDP share C	-0.8269*** (0.2710)				
conflict AC x contiguity AB		-0.0141 (0.0264)			
GDP share C		-1.5736* (0.8108)			
conflict AC x alliance AB			-0.0173 (0.0195)		
GDP share C			-0.0249 (0.4846)		
conflict AC x alliance BC				-0.0810*** (0.0266)	
GDP share C				-0.6353 (0.4359)	
conflict AC x conflict BC					0.1184*** (0.0271)
GDP share C					-24.1373* (14.0879)
Observations	475,130	475,130	475,130	475,130	475,130
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 4 additionally uses alliance BC, Column 5 uses conflict BC. Column 1 is estimated using Equation 15, Column 2 using Equation 16, Column 3 using Equation 6 with Equation 14 included additionally, Column 4 using Equation 10 with Equation 17 included additionally, and Column 5 using Equation 12 with Equation 18 included additionally. Fixed effects, covariates, R-squared, and constants are not reported.

Both *conflict AC* and *GDP share C* have negative and statistically significant coefficient estimates, indicating that both the fact *whether* there is a conflict with a C-country matters as well as *the size* of this C-country does. The results in Column 2 resemble those of imports. The interaction

between *conflict AC* and *contiguity AB* remains statistically insignificant, while the estimate for *GDP share C* is negative and even weakly statistically significant here. Thus, the hypothesized positive trade shifting effect cannot be found here either, and the size of the respective C-country further intensifies the decrease in exports from country A to B.

For the interaction with *alliance AB* (Column 3), the results resemble those of Table 5. The hypothesized positive trade shifting effect can only be found on imports, but not on exports. Interestingly, *GDP share C* does not have a significant coefficient in this estimation. For the interaction between *conflict AC* and *alliance BC* (Column 4), the coefficient estimate remains negative and statistically significant as it is in Column 3 of Table 5, thus confirms the hypothesized negative trade shifting effect. Country C's size does not seem to amplify this effect, as the coefficient estimate is statistically insignificant.

Finally, for the interaction between *conflict AC* and *conflict BC* (Column 5), the results mirror those for imports: The interaction term estimate is strongly positive and statistically significant, while the *GDP share C* estimate is strongly negative and statistically weakly significant. The magnitude of increased exports from country A to B thus depends on the size of the conflict country C. Overall, the results from Tables 7 and 8 indicate that the GDP size of the conflict country C only matters for trade reduction. It does not appear that the positive trade shifting effect increases with the size of the adversary country.

6. CONCLUSION

The aim of this paper was to examine the effect of violent interstate conflicts on trade with third-party countries, and which role the relationship constellations of the three countries play. I derived five hypotheses based on four different country-triad relationship constellations, and tested them separately for imports and exports as the outcome variable. For three out of the five hypotheses I find support for both imports and exports. For one hypothesis I only find support on imports, while one hypotheses has to be rejected for both. Countries do shift trade to other countries to outweigh trade losses, and shift it away from other countries to harm them.

The results show that violent interstate conflicts between two countries reduce both imports

from as well as exports to third-party countries by between 1.38 and 2.61 percent (imports) and by between 3.00 and 4.37 percent (for exports), confirming Hypothesis 1. The results for Hypotheses 3, 4, and 5 show that countries use trade to a certain degree as a measure of foreign policy when it comes to supporting their allies by harming their enemies. Countries net-increase imports by between 6.92 and 7.33 percent from countries that have a common enemy (Hypothesis 5), and by between 1.42 and 4.01 percent from members of the same formal security alliance (Hypothesis 3). On the other hand, countries do net-decrease imports by between 9.61 and 12.26 percent from their enemy's allies (Hypothesis 4). While an increase in exports to members of the same security alliance cannot be found, countries net-increase exports by between 5.15 and 7.45 percent to countries that have the same enemy, and reduce exports by between 13.27 and 14.09 percent to their enemy's allies. That can be seen as support to both the ideas of Liberalism, that politics and economics are entangled and influence each other, and to those of Mercantilism, that economics are subordinated to politics. However, trade shifting to neighboring countries, as hypothesized in Hypothesis 2, could not be found, neither for imports nor for exports. An additional heterogeneity analysis revealed that the size of the C-conflict countries in terms of GDP does amplify the overall stand-alone trade decrease of a violent conflict with C-countries, but does not amplify the positive trade shifting effect.

Table 9: Overview of Results

Hypotheses	Expected effect on trade AB	Imports Trade Shifting	Imports Full Effect	Exports Trade Shifting	Exports Full Effect
<u>Hypothesis 1</u> conflict AC	negative	negative	negative	negative	negative
<u>Hypothesis 2</u> conflict AC x contiguity AB	positive	insignificant	insignificant	insignificant	negative
<u>Hypothesis 3</u> conflict AC x alliance AB	positive	positive	positive	insignificant	insignificant
<u>Hypothesis 4</u> conflict AC x alliance BC	negative	negative	negative	negative	negative
<u>Hypothesis 5</u> conflict AC x conflict BC	positive	positive	positive	positive	positive

Notes: This table gives a summary of the results with regard to the different hypotheses. Results written in **bold** letters mean they confirm the respective hypotheses.

The results add to both the existing trade as well as to the trade and conflict literature. They confirm and extend the existing findings on the effect of violent conflicts on trade, and they show the application of the gravity equation with a third-party country dimension incorporated. The results are mostly in line with previous work on third-party trade by Feldman and Sadeh (2018), even though most estimates in this paper appear to be much smaller. I extended the existing evidence by including more specific relationship characteristics, considering both imports and exports, extending the sample period to more recent years, and taking into account the size of the conflict countries.

Besides adding to the conflict and trade literature, this paper provides a more dynamic perspective to the understanding of international economics and politics by looking specifically at third-party country effects. It added a unique and new analytical framework to the fields of international economics and politics, which could be applied to gain a similar understanding of other research questions within these fields, such as non-violent conflicts, political tensions, development assistance, or foreign direct investments.

REFERENCES

- Alvarez, J. & Arellano, M. (2003).** The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators. *Econometrica*, 71(4), 1121-1159.
- Arad, R. W. & Hirsch, S. (1981).** Peacemaking and Vested Interests: International Economic Transactions. *International Studies Quarterly*, 25(3), 439-468.
- Baccini, L., Dür, A., & Elsig, M. (2014).** The Design of International Trade Agreements: Introducing a New Database. *Review of International Organizations*, 9(3), 353-375. Data retrieved from <https://www.designoftradeagreements.org/downloads/>. Accessed 29.04.2019.
- Baier, S. L. & Bergstrand, J. H. (2007).** Do Free Trade Agreements Actually Increase Members' International Trade? *Journal of International Economics*, 71(1), 72-95.
- Barbieri, K. (1996).** Economic Interdependence: A Path to Peace or a Source of Interstate Conflict? *Journal of Peace Research*, 33(1), 29-49.
- Barbieri, K. (2002).** *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor: University of Michigan Press.
- Barbieri, K. & Keshk, O. M. G. (2016).** Correlates of War Project Trade Data Set Codebook, Version 4.0. <http://www.correlatesofwar.org/>. Accessed 12.06.2017.
- Bayer, R. & Rupert, M. C. (2004).** Effects of Civil Wars on International Trade, 1950-92. *Journal of Peace Research*, 41(6), 699-713.
- Bergeijk, P. A. G. (1994).** *Economic Diplomacy, Trade, and Commercial Policy: Positive and Negative Sanctions in a New World Order.* Vermont: Edward Elgar.
- Berman, N., Couttenier, M., Rohner, D., & Thoenig, M. (2017).** This Mine Is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review*, 107(6), 1564-1610.
- Bernard, A. B., Eaton, J., Jensen, J. B., & Kortum, S. (2003).** Plants and Productivity in International Trade. *The American Economic Review*, 93(4), 1268-1290.

-
- Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K. (2012).** The Empirics of Firm Heterogeneity and International Trade. *Annual Review Economics*, 4(1), 283-313.
- Brecher, M., Wilkenfeld, J., Beardsley, K., James, P., & Quinn, D. (2016).** International Crisis Behavior Data Codebook, Version 11. <https://sites.duke.edu/icbdata/data-collections/>. Accessed 08.06.2017.
- Caruso, R. (2003).** The Impact of International Economic Sanctions on Trade: An Empirical Analysis. *Peace Economics, Peace Science and Public Policy*, 9(2), 1-34.
- Caselli, F., Morelli, M., & Rohner, D. (2013).** The Geography of Inter-State Resource Wars. NBER Working Paper No. 18978. Cambridge: National Bureau of Economic Research.
- Chaney, T. (2018).** The Gravity Equation in International Trade: An Explanation. *Journal of Political Economy*, 126(1), 150-177.
- Correlates of War Project (n. d.).** Direct Contiguity Data, 1816-2016. Version 3.2. <http://www.correlatesofwar.org/data-sets/direct-contiguity>. Accessed 04.08.2017.
- Correlates of War Project (2017).** National Material Capabilities (v5.0). <http://www.correlatesofwar.org/data-sets/national-material-capabilities>. Accessed 02.08.2018.
- Cranmer, S. J., Menninga, E. J., & Mucha, P. J. (2015).** Kantian Fractionalization Predicts the Conflict Propensity of the International System. *Proceedings of the National Academy of Sciences*, 112(38), 11812-11816.
- De Soysa, I. & Neumayer, E. (2007).** Resource Wealth and the Risk of Civil War Onset: Results from a New Dataset of Natural Resource Rents, 1970—1999. *Conflict Management and Peace Science*, 24(3), 201-218.
- Dorussen, H. (2006).** Heterogeneous Trade Interests and Conflict: What You Trade Matters. *Journal of Conflict Resolution*, 50(1), 87-107.
- Feldman, N. & Sadeh, T. (2018).** War and Third-Party Trade. *Journal of Conflict Resolution*, 62(1), 119-142.

-
- Fordham, B. & Poast, P. (2016).** All Alliances are Multilateral: Rethinking Alliance Formation. *Journal of Conflict Resolution*, 60(5), 840-865.
- Fuchs, S. (2018).** The Spoils of War: Trade Shocks during WWI and Spain's Regional Development. 2018 Meeting Papers 1172, Society for Economic Dynamics.
- Gilpin, R. (1987).** *The Political Economy of International Relations*. Princeton: Princeton University Press.
- Gasiorowski, M. J. (1986).** Economic Interdependence and International Conflict: Some Cross-National Evidence. *International Studies Quarterly*, 30(1), 23-38.
- Gasiorowski, M. J. & Polachek, S. W. (1982).** Conflict and Interdependence: East-West Trade and Linkages in the Era of Détente. *Journal of Conflict Resolution*, 26(4), 709-729.
- Gibler, D. M. (2009).** International Military Alliances, 1648-2008. CQ Press. <http://www.correlatesofwar.org/data-sets/formal-alliances>. Accessed 07.08.2017.
- Goenner, C. F. (2011).** Simultaneity between Trade and Conflict: Endogenous Instruments of Mass Destruction. *Conflict Management and Peace Science*, 28(5), 459-477.
- Gowa, J. (1994).** *Allies, Adversaries, and International Trade*. Princeton: Princeton University Press.
- De Groot, O. J. (2010).** The Spillover Effects of Conflict on Economic Growth in Neighbouring Countries in Africa. *Defence and Peace Economics*, 21(2), 149-164.
- Head, K. & Mayer, T. (2014).** Gravity Equations: Workhorse, Toolkit, and Cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (eds.) *Handbook of International Economics* (Vol. 4, pp. 131-195). Elsevier.
- Hegre, H. (2004).** Size Asymmetry, Trade, and Militarized Conflict. *Journal of Conflict Resolution*, 48(3), 403-429.
- Hidalgo, C. A. & Simoes, A. J. G. (2011).** The Economic Complexity Observatory: An Analytical Tool for Understanding the Dynamics of Economic Development. Workshops at the Twenty-Fifth

AAAI Conference on Artificial Intelligence. Data retrieved from <https://atlas.media.mit.edu/en/>. Accessed 01.05.2019.

Hirschman, A. O. (1945). *National Power and the Structure of Foreign Trade*. Berkeley: University of California Press.

Integrated Network for Societal Conflict Research (2018). Polity IV Annual Time-Series, 1800-2015. <http://www.systemicpeace.org/inscrdata.html>. Accessed 24.07.2018.

International Monetary Fund (2017). IMF Data. <http://www.imf.org/en/Data>. Accessed 24.07.2017.

Jackson, M. O. & Nei, S. (2015). Networks of Military Alliances, Wars, and International Trade. *Proceedings of the National Academy of Sciences*, 112(50), 15277-15284.

Katz, R. (2013). Mutual Assured Production: Why Trade Will Limit Conflict Between China and Japan. *Foreign Affairs*, 92(4), 18-24.

Keshk, O. M. G., Pollins, B. M., & Reuveny, R. (2004). Trade Still Follows the Flag: The Primacy of Politics in a Simultaneous Model of Interdependence and Armed Conflict. *Journal of Politics*, 66(4), 1155-1179.

Keshk, O. M. G., Pollins, B. M., & Reuveny, R. (2010). Trade and Conflict: Proximity, Country Size, and Measures. *Conflict Management and Peace Science*, 27(1), 3-27.

Kim, S. Y. (1998). Ties that Bind: The Role of Trade in International Conflict Processes, 1950-1992 (Doctoral Dissertation, Yale University).

Kinne, B. J. (2014). Does Third-Party Trade Reduce Conflict? Credible Signaling Versus Opportunity Costs. *Conflict Management and Peace Science*, 31(1), 28-48.

Korovkin, V. & Makarin, A. (2021). Conflict and Inter-Group Trade: Evidence from the 2014 Russia-Ukraine Crisis. SSRN Working Paper.

Li, Q. & Reuveny, R. (2008). Trading for Peace? Disaggregated Bilateral Trade and Interstate Military Conflict Initiation. *Journal of Peace Research*. 812, 1-41.

- Li, Q. & Sacko, D. H. (2002).** The (Ir)Relevance of Militarized Interstate Disputes for International Trade. *International Studies Quarterly*, 46(1), 11-43.
- Li, W., Bradshaw, A. E., Clary, C. B., & Cranmer, S. J. (2017).** A Three-Degree Horizon of Peace in the Military Alliance Network. *Science advances*, 3(3), e1601895.
- Long, A. G. (2008).** Bilateral Trade in the Shadow of Armed Conflict. *International Studies Quarterly*, 52(1), 81-101.
- Long, A. G. & Leeds, B. A. (2006).** Trading for Security: Military Alliances and Economic Agreements. *Journal of Peace Research*, 43(4), 433-451.
- Lupu, Y. & Traag, V. A. (2013).** Trading Communities, the Networked Structure of International Relations, and the Kantian Peace. *Journal of Conflict Resolution*, 57(6), 1011-1042.
- Mansfield, E. D. & Pevehouse, J. C. (2000).** Trade Blocs, Trade Flows, and International Conflict. *International Organization*, 54(4), 775-808.
- Martin, P., Mayer, T., & Thoenig, M. (2008).** Make Trade Not War? *Review of Economic Studies*, 75(3), 865-900.
- Martin, P., Mayer, T., & Thoenig, M. (2012).** The Geography of Conflicts and Regional Trade Agreements. *American Economic Journal: Macroeconomics*, 4(4), 1-35.
- Mayer, T. & Zignago, S. (2011).** Notes on CEPII's Distance Measures: The GeoDist Database. CEPII Working Paper, 2011-25.
- Melitz, M. J. (2003).** The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, 71(6), 1695-1725.
- Melitz, M. J. & Trefler, D. (2012).** Gains from Trade when Firms Matter. *Journal of Economic Perspectives*, 26(2), 91-118.
- Michaels, G. & Zhi, X. (2010).** Freedom Fries. *American Economic Journal: Applied Economics*, 2(3), 256-81.

-
- Morrow, J. D., Siverson, R. M., & Tabares, T. E. (1998).** The Political Determinants of International Trade: The Major Powers, 1907–1990. *American Political Science Review*, 92(3), 649-661.
- Murdoch, J. C. & Sandler, T. (2002a).** Economic Growth, Civil Wars, and Spatial Spillovers. *Journal of Conflict Resolution*, 46(1), 91-110.
- Murdoch, J. C. & Sandler, T. (2002b).** Civil Wars and Economic Growth: A Regional Comparison. *Defence and Peace Economics*, 13(6), 451-464.
- Murdoch, J. C. & Sandler, T. (2004).** Civil Wars and Economic Growth: Spatial Dispersion. *American Journal of Political Science*, 48(1), 138-151.
- Nickell, S. (1981).** Biases in Dynamic Models with Fixed Effects. *Econometrica*, 1417-1426.
- Oneal, J. R., Oneal, F. H., Maoz, Z., & Russett, B. (1996).** The Liberal Peace: Interdependence, Democracy, and International Conflict, 1950-85. *Journal of Peace Research*, 33(1), 11-28.
- Oneal, J. R. & Russett, B. (1997).** The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985. *International Studies Quarterly*, 41(2), 267-294.
- Oneal, J. R., Russett, B., & Berbaum, M. L. (2003).** Causes of Peace: Democracy, Interdependence, and International Organizations, 1885–1992. *International Studies Quarterly*, 47(3), 371-393.
- Papayouanou, P. A. (1996).** Interdependence, Institutions, and the Balance of Power: Britain, Germany, and World War I. *International Security*, 20(4), 42-76.
- Peterson, T. M. (2011).** Third-Party Trade, Political Similarity, and Dyadic Conflict. *Journal of Peace Research*, 48(2), 185-200.
- Poast, P. (2010).** (Mis)Using Dyadic Data to Analyze Multilateral Events. *Political Analysis*, 18(4), 403-425.
- Poast, P. (2016).** Dyads are Dead, Long Live Dyads! The Limits of Dyadic Designs in International Relations Research. *International Studies Quarterly*, 60(2), 369-374.

-
- Polachek, S. W. (1980).** Conflict and Trade. *Journal of Conflict Resolution*, 24(1), 55-78.
- Pollins, B. M. (1989).** Conflict, Cooperation, and Commerce: The Effect of International Political Interactions on Bilateral Trade Flows. *American Journal of Political Science*, 33(3), 737-761.
- Reuveny, R. (2000).** The Trade and Conflict Debate: A Survey of Theory, Evidence and Future Research. *Peace Economics, Peace Science and Public Policy*, 6(1), 23-49.
- Sayrs, L. W. (1990).** Expected Utility and Peace Science: An Assessment of Trade and Conflict. *Conflict Management and Peace Science*, 11(1), 17-44.
- Singer, J. D. (1987).** Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985. *International Interactions*, 14(2), 115-132.
- Sorens, J. (2011).** Mineral Production, Territory, and Ethnic Rebellion: The Role of Rebel Constituencies. *Journal of Peace Research*, 48(5), 571-585.
- Stinnett, D. M., Tir, J. Schafer, P., Diehl, P. F., & Gochman, C. (2002).** The Correlates of War Direct Contiguity Data, Version 3. *Conflict Management and Peace Science*, 17(1), 93-112.
- Tinbergen, J. (1962).** *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.
- Uchitel, A. (1993).** Interdependence and Instability. In J. Snyder and R. Jervice (eds.) *Coping with Complexity in the International System*. Boulder: Westview Press.
- De Vries, M. S. (1990).** Interdependence, Cooperation and Conflict: An Empirical Analysis. *Journal of Peace Research*, 27(4), 429-444.
- Whang, T., McLean, E. V., & Kuberski, D. W. (2013).** Coercion, Information, and the Success of Sanction Threats. *American Journal of Political Science*, 57(1), 65-81.
- World Bank (2017).** World Development Indicators. <http://data.worldbank.org/data-catalog/world-development-indicators>. Accessed 22.07.2017.
- World Bank (2021).** World Development Indicators. <http://data.worldbank.org/data-catalog/world-development-indicators>. Accessed 20.07.2021.

APPENDIX

Table A1: Estimates of Dyadic and Third-Party Country Conflicts, log(1+trade)

	(1)	(2)	(3)	(4)	(5)	(6)
	log 1+import 1968-2013	log 1+import 1968-2013	log 1+import 1968-2013	log 1+export 1968-2013	log 1+export 1968-2013	log 1+export 1968-2013
conflict AB	-0.5811*** (0.0729)	-0.6131*** (0.0759)	-0.6076*** (0.0760)	-0.5811*** (0.0729)	-0.6131*** (0.0759)	-0.6044*** (0.0760)
conflict AC			-0.0189*** (0.0032)			-0.0289*** (0.0033)
Observations	841,326	702,314	702,314	841,326	702,314	702,314
Year FE	yes	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes	yes
Controls	basic	full	full	basic	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use GDP and contiguity AB as covariates. Columns 2, 3, 5, and 6 use additionally alliance AB, FTA AB, polity AB, as well as the oil and diamond covariates for A and B. Columns 1 to 3 include imports lagged by one year, Columns 4 to 6 exports lagged by one year. Columns 1, 2, 4, and 5 are estimated using Equation 3, Columns 3 and 6 are estimated using Equation 4. Fixed effects, covariates, R-squared, and constants are not reported.

Table A2: Estimates of Trade Shifting Effects on Imports, log(1+trade)

	(1)	(2)	(3)	(4)	(5)
	log 1+import 1968-2013	log 1+import 1968-2013	log 1+import 1968-2013	log 1+import 1968-2013	log 1+import 1968-2013
conflict AC	-0.0178*** (0.0033)	-0.0198*** (0.0034)	-0.0152*** (0.0033)	-0.0212*** (0.0033)	-0.0181*** (0.0034)
conflict AC x contiguity AB	-0.0253 (0.0168)				-0.0293* (0.0176)
conflict AC x alliance AB		0.0103 (0.0107)			0.0300** (0.0120)
conflict AC x alliance BC			-0.0472*** (0.0132)		-0.0551*** (0.0137)
conflict AC x conflict BC				0.0463*** (0.0134)	0.0495*** (0.0138)
Observations	702,314	702,314	702,314	702,314	702,314
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 3 and 5 additionally use alliance BC, Columns 4 and 5 use conflict BC. Column 1 is estimated using Equation 5, Column 2 using Equation 6, Column 3 using Equation 10, Column 4 using Equation 12, and Column 5 using Equation 13. Fixed effects, covariates, R-squared, and constants are not reported.

Table A3: Estimates of Trade Shifting Effects on Exports, log(1+trade)

	(1)	(2)	(3)	(4)	(5)
	log 1+export 1968-2013	log 1+export 1968-2013	log 1+export 1968-2013	log 1+export 1968-2013	log 1+export 1968-2013
conflict AC	-0.0272*** (0.0034)	-0.0274*** (0.0035)	-0.0254*** (0.0034)	-0.0310*** (0.0034)	-0.0262*** (0.0036)
conflict AC x contiguity AB	-0.0365** (0.0168)				-0.0313* (0.0176)
conflict AC x alliance AB		-0.0177* (0.0105)			-0.0028 (0.0114)
conflict AC x alliance BC			-0.0443*** (0.0128)		-0.0433*** (0.0131)
conflict AC x conflict BC				0.0463*** (0.0134)	0.0550*** (0.0138)
Observations	702,314	702,314	702,314	702,314	702,314
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 3 and 5 additionally use alliance BC, Columns 4 and 5 use conflict BC. Column 1 is estimated using Equation 5, Column 2 using Equation 6, Column 3 using Equation 10, Column 4 using Equation 12, and Column 5 using Equation 13. Fixed effects, covariates, R-squared, and constants are not reported.

Table A4: Estimates of Dyadic and Third-Party Country Conflicts, Covariates Reported

	(1)	(2)	(3)	(4)	(5)	(6)
	log import 1968-2013	log import 1968-2013	log import 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013
log import AB, t-1	0.5315*** (0.0021)	0.5466*** (0.0023)	0.5466*** (0.0023)			
log export AB, t-1				0.5315*** (0.0021)	0.5466*** (0.0023)	0.5466*** (0.0023)
conflict AB	-0.7571*** (0.1142)	-0.8231*** (0.1221)	-0.8180*** (0.1223)	-0.7571*** (0.1142)	-0.8231*** (0.1221)	-0.8131*** (0.1221)
conflict AC			-0.0211*** (0.0079)			-0.0398*** (0.0082)
log GDP A	0.4961*** (0.0086)	0.4796*** (0.0088)	0.4792*** (0.0088)	0.5895*** (0.0093)	0.5625*** (0.0096)	0.5618*** (0.0096)
log GDP B	0.5895*** (0.0093)	0.5625*** (0.0096)	0.5624*** (0.0096)	0.4961*** (0.0086)	0.4795*** (0.0088)	0.4796*** (0.0088)
contiguity AB	-0.0793 (0.1054)	-0.0676 (0.1044)	-0.0691 (0.1044)	-0.0793 (0.1054)	-0.0677 (0.1044)	-0.0700 (0.1043)
alliance AB		0.0626*** (0.0135)	0.0628*** (0.0135)		0.0626*** (0.0135)	0.0630*** (0.0135)
FTA AB		-0.0450*** (0.0166)	-0.0450*** (0.0166)		-0.0450*** (0.0166)	-0.0450*** (0.0166)
polity difference AB		0.0006 (0.0005)	0.0006 (0.0005)		0.0006 (0.0005)	0.0006 (0.0005)
log oil A		-0.0010** (0.0004)	-0.0010** (0.0004)		0.0031*** (0.0005)	0.0032*** (0.0005)
log oil B		0.0031*** (0.0005)	0.0031*** (0.0005)		-0.0010** (0.0004)	-0.0010** (0.0004)
log diamonds A		-0.0018*** (0.0005)	-0.0018*** (0.0005)		0.0036*** (0.0006)	0.0037*** (0.0006)
log diamonds B		0.0036*** (0.0006)	0.0036*** (0.0006)		-0.0017*** (0.0005)	-0.0017*** (0.0005)
Observations	543,040	475,130	475,130	543,040	475,130	475,130
Year FE	yes	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes	yes
Controls	basic	full	full	basic	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use GDP and contiguity AB as covariates. Columns 2, 3, 5, and 6 use additionally alliance AB, FTA AB, polity AB, as well as the oil and diamond covariates for A and B. Columns 1 to 3 include imports lagged by one year, Columns 4 to 6 exports lagged by one year. Columns 1, 2, 4, and 5 are estimated using Equation 3, Columns 3 and 6 are estimated using Equation 4. Fixed effects, R-squared, and constants are not reported. This table contains the same estimation results as Table 2, but reports all covariates.

Table A5: Estimates of Trade Shifting Effects on Imports, Covariates Reported

	(1)	(2)	(3)	(4)	(5)
	log import 1968-2013	log import 1968-2013	log import 1968-2013	log import 1968-2013	log import 1968-2013
log import AB, t-1	0.5466*** (0.0023)	0.5466*** (0.0023)	0.5465*** (0.0023)	0.5466*** (0.0023)	0.5465*** (0.0023)
conflict AB	-0.8172*** (0.1222)	-0.8199*** (0.1222)	-0.7988*** (0.1225)	-0.8033*** (0.1222)	-0.7809*** (0.1224)
conflict AC	-0.0206** (0.0082)	-0.0251*** (0.0084)	-0.0139* (0.0081)	-0.0265*** (0.0081)	-0.0231*** (0.0087)
conflict BC				-0.0447*** (0.0085)	-0.0448*** (0.0085)
contiguity AB	-0.0683 (0.1044)	-0.0685 (0.1043)	-0.0724 (0.1044)	-0.0702 (0.1043)	-0.0712 (0.1043)
alliance AB	0.0628*** (0.0135)	0.0603*** (0.0135)	0.0758*** (0.0137)	0.0627*** (0.0135)	0.0720*** (0.0137)
alliance BC			-0.0368*** (0.0104)		-0.0371*** (0.0104)
conflict AC x contiguity AB	-0.0078 (0.0247)				-0.0184 (0.0258)
conflict AC x alliance AB		0.0382* (0.0199)			0.0610*** (0.0213)
conflict AC x alliance BC			-0.0858*** (0.0275)		-0.1051*** (0.0282)
conflict AC x conflict BC				0.0910*** (0.0228)	0.0918*** (0.0236)
log GDP A	0.4792*** (0.0088)	0.4791*** (0.0088)	0.4785*** (0.0088)	0.4792*** (0.0088)	0.4781*** (0.0088)
log GDP B	0.5624*** (0.0096)	0.5624*** (0.0096)	0.5623*** (0.0096)	0.5617*** (0.0096)	0.5615*** (0.0096)
FTA AB	-0.0450*** (0.0166)	-0.0450*** (0.0166)	-0.0453*** (0.0165)	-0.0448*** (0.0166)	-0.0450*** (0.0165)
polity difference AB	0.0006 (0.0005)	0.0006 (0.0005)	0.0006 (0.0005)	0.0006 (0.0005)	0.0006 (0.0005)
log oil A	-0.0010** (0.0004)	-0.0010** (0.0004)	-0.0010** (0.0004)	-0.0010** (0.0004)	-0.0010** (0.0004)
log oil B	0.0031*** (0.0005)	0.0031*** (0.0005)	0.0031*** (0.0005)	0.0032*** (0.0005)	0.0032*** (0.0005)
log diamonds A	-0.0018*** (0.0005)	-0.0018*** (0.0005)	-0.0018*** (0.0005)	-0.0018*** (0.0005)	-0.0018*** (0.0005)
log diamonds B	0.0036*** (0.0006)	0.0036*** (0.0006)	0.0036*** (0.0006)	0.0037*** (0.0006)	0.0037*** (0.0006)
Observations	475,130	475,130	475,130	475,130	475,130
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 3 and 5 additionally use alliance BC, Columns 4 and 5 use conflict BC. Column 1 is estimated using Equation 5, Column 2 using Equation 6, Column 3 using Equation 10, Column 4 using Equation 12, and Column 5 using Equation 13. Fixed effects, R-squared, and constants are not reported. This table contains the same estimation results as Table 3, but reports all covariates.

Table A6: Estimates of Trade Shifting Effects on Exports, Covariates Reported

	(1)	(2)	(3)	(4)	(5)
	log export 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013	log export 1968-2013
log export AB, t-1	0.5466*** (0.0023)	0.5466*** (0.0023)	0.5466*** (0.0023)	0.5466*** (0.0023)	0.5466*** (0.0023)
conflict AB	-0.8091*** (0.1221)	-0.8121*** (0.1221)	-0.7893*** (0.1208)	-0.8032*** (0.1222)	-0.7755*** (0.1209)
conflict AC	-0.0377*** (0.0086)	-0.0379*** (0.0088)	-0.0305*** (0.0085)	-0.0447*** (0.0085)	-0.0347*** (0.0092)
conflict BC				-0.0265*** (0.0081)	-0.0265*** (0.0081)
contiguity AB	-0.0663 (0.1043)	-0.0701 (0.1043)	-0.0722 (0.1043)	-0.0703 (0.1043)	-0.0698 (0.1044)
alliance AB	0.0630*** (0.0135)	0.0642*** (0.0135)	0.0561*** (0.0136)	0.0627*** (0.0135)	0.0555*** (0.0137)
alliance BC			0.0285*** (0.0093)		0.0283*** (0.0093)
conflict AC x contiguity AB	-0.0340 (0.0259)				-0.0232 (0.0269)
conflict AC x alliance AB		-0.0176 (0.0194)			0.0043 (0.0204)
conflict AC x alliance BC			-0.1083*** (0.0263)		-0.1130*** (0.0268)
conflict AC x conflict BC				0.0910*** (0.0228)	0.1031*** (0.0237)
log GDP A	0.5618*** (0.0096)	0.5619*** (0.0096)	0.5620*** (0.0096)	0.5617*** (0.0096)	0.5618*** (0.0096)
log GDP B	0.4796*** (0.0088)	0.4796*** (0.0088)	0.4798*** (0.0088)	0.4790*** (0.0088)	0.4793*** (0.0088)
FTA AB	-0.0450*** (0.0166)	-0.0450*** (0.0166)	-0.0445*** (0.0166)	-0.0448*** (0.0166)	-0.0444*** (0.0166)
polity difference AB	0.0006 (0.0005)	0.0006 (0.0005)	0.0005 (0.0005)	0.0006 (0.0005)	0.0005 (0.0005)
log oil A	0.0032*** (0.0005)	0.0032*** (0.0005)	0.0032*** (0.0005)	0.0032*** (0.0005)	0.0032*** (0.0005)
log oil B	-0.0010** (0.0004)	-0.0010** (0.0004)	-0.0010** (0.0004)	-0.0009** (0.0004)	-0.0009** (0.0004)
log diamonds A	0.0037*** (0.0006)	0.0037*** (0.0006)	0.0037*** (0.0006)	0.0037*** (0.0006)	0.0037*** (0.0006)
log diamonds B	-0.0017*** (0.0005)	-0.0017*** (0.0005)	-0.0017*** (0.0005)	-0.0017*** (0.0005)	-0.0017*** (0.0005)
Observations	475,130	475,130	475,130	475,130	475,130
Year FE	yes	yes	yes	yes	yes
Country FE	yes	yes	yes	yes	yes
Dyad FE	yes	yes	yes	yes	yes
Controls	full	full	full	full	full

Notes: Robust standard errors are reported in parentheses. The asterisk denote the significance levels of 1, 5, and 10 percent (***) $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All columns use year, country, and dyad fixed effects. All columns use the full set of AB covariates: GDP, contiguity AB, alliance AB, FTA AB, polity AB, oil and diamond exports, conflict AB, as well as imports lagged by one year. Column 3 and 5 additionally use alliance BC, Columns 4 and 5 use conflict BC. Column 1 is estimated using Equation 5, Column 2 using Equation 6, Column 3 using Equation 10, Column 4 using Equation 12, and Column 5 using Equation 13. Fixed effects, R-squared, and constants are not reported. This table contains the same estimation results as Table 3, but reports all covariates.



Chapter 3

The Ownership of Pioneer Firms: The Role of State-Owned and Multinational Enterprises in Industrial Diversification in Vietnam

The Ownership of Pioneer Firms: The Role of State-Owned and Multinational Enterprises in Industrial Diversification in Vietnam

Henrik Hansen, John Rand, Bjørn Bo Sørensen, and Helge Zille

Department of Economics, University of Copenhagen

We study the ownership of pioneer firms that establish market activities in provinces where such activities did not previously exist. Using the Vietnamese Enterprise Survey (2001-2017), which allows us to track close to one million formal firms over time, we explore whether Vietnam's remarkable industrial diversification during the past two decades was driven by state-owned, multinational, or domestic and privately owned enterprises (SOEs, MNEs, or PDEs). We document significantly higher ownership-specific pioneering frequencies in SOEs and MNEs compared to PDEs, also after controlling for a broad spectrum of observable characteristics. Using a simple conceptual framework, we attribute these differences to variations in access to capital, knowledge of firm- and industry productivity, and firm-specific strategies and policy objectives. Second, we investigate the dynamic employment effects in industries pioneered by different firms. We find a large employment premium in MNE-pioneered industries. The premium is driven entirely by employment in pioneering MNEs and other MNEs following the pioneer in subsequent years. In contrast, there is no employment premium in industries pioneered by SOEs relative to PDEs. Pioneering SOEs are systematically larger than pioneering PDEs, but this direct employment effect is counterbalanced by less employment in PDEs and MNEs in SOE-pioneered industries in the long run. (JEL F23, L21, L32, O25, O31, P26)

1. INTRODUCTION

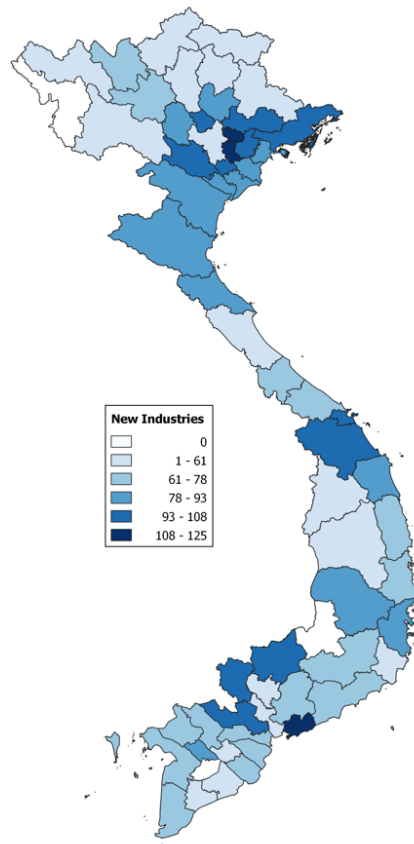
Economic growth is tightly linked to an economy's ability to diversify into new and more productive and sophisticated industries (Grossman and Helpman, 1989; Hidalgo and Hausmann, 2009; Jovanovic and Nyarko, 1996). At the micro-level, regional diversification depends on "pioneer firms" that establish new industries that did not previously exist (Jara-Figueroa et al., 2018). Pioneers starting infant industries that prove locally viable will attract emulators, and their industries will grow. Understanding who these pioneers are and what determines their success are first-order questions in the study of economic development. Yet, pioneer firms have received surprisingly little attention in the empirical literature.¹

In this paper, we study the ownership of pioneer firms and thereby revisit a long-standing debate about the role of the state *vis-à-vis* the market in the development process. Diverging explanations of the causes of the East Asian growth miracle constitute prominent examples of this debate. Some attribute the rise of the so-called East Asian Tigers to a strong and interventionist "developmental state" (Wade, 1990), while others argue that these states primarily succeeded due to prudent market-friendly policies (World Bank, 1993). Echoes of this debate are now repeated in explanations of the performance of China and Vietnam (Malesky and London, 2014). We add a new angle to this debate by studying how pioneering private domestic enterprises (PDEs), state-owned enterprises (SOEs), and multinational enterprises (MNEs) have affected the diversification process in Vietnam during the past two decades. Specifically, we test which firms have the highest propensity to pioneer, and we identify the ownership of the firms that pioneer successful industries that grow and thereby move people into new economic activities.

A unique combination of context and data-coverage makes Vietnam ideal to study the role of different firms in the pioneering process. First, Vietnam underwent a remarkable economic diversification during this period. Figure 1 shows the astonishing number of unique 4-digit industries that Vietnam's 63 provinces added to their production portfolio from 2001 to 2017, with some provinces adding more than 100 industries during the period under study. The fact that Vietnam's economy consisted in total of 231 unique industries during this entire period puts into perspective the significance of this figure. PDEs, MNEs, and SOEs have all been integral parts of the Vietnamese

¹ Notable exemptions are Hausmann and Neffke (2019) and Jara-Figueroa et al. (2018).

Figure 1: Number of New Unique 4-Digit Industries by Province, 2001-2017



Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. Industry codes follow the 4-digit Vietnamese Standard Industrial Classification system. We use a slightly modified version of the classification system (see Section 3). During the sample period, firms operate in a total of 231 unique 4-digit industries. The categorization follows Jenks natural breaks classification method (Jenks, 1967).

economy during this transformation. The formal private domestic sector flourished and is now the largest sector in Vietnam's economy, employing around 60 percent of all formal workers. Meanwhile, foreign direct investments (FDI) into the Vietnamese economy soared, peaking at almost 10 percent of GDP in 2008, just after the WTO accession in 2007, and averaging around 6 percent of GDP in the past decade.² In 2017, MNEs employed about one-third of all formal workers in Vietnam. At the same time, the relative size of the state-sector has declined, but formal employment in SOEs have remained substantial and relatively stable, despite continued reforms and equitization of SOEs.

Second, data from the Vietnam Enterprises Survey (VES) allows us to track the near-universe of close to one million formal firms registered under Vietnam's Enterprise Law from 2001 to 2017.

² Source: Authors' calculations based on data from World Development Indicators (World Bank, 2022).

The data contains detailed information on ownership structures, geo-location, and firms' (multiple) economic activities, and it thereby makes it possible to identify pioneers with different ownership structures at a granular level of detail. Importantly, the long time span of the data also enables us to track the evolution of industries pioneered by different firms.

Equipped with this data, our analysis proceeds in three steps. We first present a set of stylized facts about the pioneering process in Vietnam, focusing on how it is shaped by firms with different ownership structures. We show that the majority of pioneering is undertaken by PDEs, but that the propensity to pioneer is higher in both SOEs and MNEs compared to PDEs. We also note some qualitative differences in the pioneering activities of firms. In particular, we show that MNEs tend to pioneer in complex industries in Vietnam's economic centers, while SOEs tend to pioneer in unsophisticated industries in rural or less developed provinces. Finally, we exploit the longitudinal dimension of our data to highlight that industries pioneered by different firm types exhibit different dynamics. The monopoly period of the pioneer tends to be longer in SOE-pioneered industries, and these industries also exhibit lower post-pioneering employment growth compared to industries pioneered by domestic and foreign private firms. Conversely, there seems to be a substantial employment premium in industries pioneered by multinationals.

We then outline a simple conceptual framework to make sense of the findings and systematically connect them to the related theoretical and empirical literature. Our framework combines the set-up in Verhoogen (2021) with theoretical models of pioneering (Hausmann and Rodrik, 2003; Hausmann and Klinger, 2006; Hausmann et al., 2007; Hoff, 1997; Wei et al., 2021). In our framework, firms choose between four mutually exclusive options in each time period. They can either shut down operations, stay in the same industry, switch to a new industry that is already established, or pioneer an industry that did not previously exist. Firms choose the option that maximizes their value function. With reference to the literature, we hypothesize that firms' decision to pioneer, and the growth of their pioneered industries, are systematically related to their ownership structure. In models of pioneering such as Wei et al. (2021), pioneering has a fixed "discovery cost". The cost may vary across firms, for instance, because MNEs and SOEs have better access to cheap finance (Baccini et al., 2019; Cao et al., forthcoming). We also hypothesize that fixed costs, productivity, and prices of follower firms are partly determined by the characteristics of the pioneer. There may, for instance, be political and regulatory barriers to entry in SOE-pioneered industries (Baccini et al., 2019; Brandt

et al., 2020), while followers can potentially benefit from productivity spillovers from pioneering MNEs (Görg and Greenway, 2004; Javorcik, 2004; Newman et al., 2015). Finally, SOEs may maximize over profits *plus* the value they assign to the welfare of other firms. Our conceptual framework therefore allows SOEs to consider the externalities of their activities, which is in line with the literature stating that SOEs are tasked with multiple objectives or "policy burdens" (Bai et al., 2000, 2006; Berkowitz et al., 2017; Kornai, 1986; Lin et al., 1998; Lin and Tan, 1999; Shleifer and Vishny, 1994; Wen, 2022). It is especially appropriate in our setting because a key conclusion from theoretical models of pioneering is that the pioneer cannot appropriate the entire value of its investment due to information externalities (Hausmann and Rodrik, 2003; Hausmann et al., 2007; Hoff, 1997) or capability spillovers (Hausmann and Klinger, 2006).

In the final part of the paper, we test the extent to which MNEs and SOEs are special. We cannot explicitly test the mechanisms outlined in the conceptual framework, but we can to some extent isolate the contribution of sets of the proposed mechanisms to the ownership differences we observe. As a first step, we focus on the ownership differences in pioneering frequencies. We use Oaxaca-Blinder decompositions to disentangle the influence of observable characteristics (composition effect) from the effect of unobservable characteristics (ownership effect) on pioneering frequencies. In the decompositions, we incrementally add controls to systematically quantify the contribution of different characteristics to the difference in pioneering frequencies. With a difference of 0.55 percentage points (ppt), the pioneering frequency of SOEs is more than three times larger than that of PDEs. A large part of this difference can be attributed to composition effects, but even in our most restrictive specification, more than one-fifth of the difference remains in the unexplained component. The raw difference in pioneering frequencies between MNEs and PDEs is smaller (0.28 ppt), but the ownership effect accounts for about half of this difference. Using our conceptual framework, we hypothesise that the differences arise from better access to finance; better information about productivity gains; and/or SOE-specific policy objectives and MNE-specific knowledge and strategies.

We then turn to study the evolution of industries pioneered by firms with different ownership structures. Using the imputation-based estimator proposed by Borusyak et al. (2022) in an event-study set-up with staggered adoption, we focus on the dynamic evolution of total firm counts and employment in local industries in different provinces to test if there is a difference in ownership of

pioneers that establish industries that attract followers and grow. In a static difference-in-differences set-up, we estimate a big employment premium in MNE-pioneered industries of 640-770 percent, corresponding to roughly 650 additional workers. Almost two-thirds of the effect is attributable to employment in the pioneer firm itself, while one-third is attributable to employment in follower firms. We also estimate fully dynamic event studies, and document that the crowding-in effect in MNE-pioneered industries is entirely driven by the entrance of other MNEs. We find no significant difference in employment between industries pioneered by SOEs and PDEs. In SOE-pioneered industries, the pioneer itself employs around 60 workers more, but employment in emulating MNEs and PDEs is significantly lower in these industries compared to PDE-pioneered industries. This is not surprising given the findings in other studies of significant entry-barriers in state-dominated sectors (Baccini et al., 2019; Brandt et al., 2020).

In a final exercise, we assign all workers in the formal Vietnamese economy to employment in either incumbent industries or infant industries, which were pioneered during our sample period. We further decompose infant industry employment into employment in local industries started by PDEs, SOEs, or MNEs. Almost 40 percent of workers employed in infant industries in 2017 were working in industries pioneered by MNEs. This is remarkable since only about 5.5 percent of all pioneering activity is conducted by multinationals.

Our analysis is situated within a broad and diverse literature studying the role of firms in industrial diversification of regions and countries.³ We build directly on the literature studying pioneer firms, where influential theoretical work has shown that the pioneering process may be hindered by the existence of externalities, justifying state intervention. In Hausmann and Klinger (2007), Hausmann and Rodrik (2003), and Hoff (1997), investment in new industries is risky because the underlying cost structure of the local economy is unknown. Successful entrepreneurship therefore signals market opportunities for emulators. In Hausmann and Klinger (2006), diversification into new industries requires investment in new productive capabilities that may spill over to emulators entering the same or related industries. These externalities cannot be fully appropriated by the incumbent entrepreneur and will lead to underinvestment in new industries in a competitive private

³ While an extensive body of work has already studied diversification and industry switching *within firms* (Bernard et al., 2010; Brambilla, 2009; Newman et al., 2013), we are addressing a different literature studying the role of pioneer firms in *regional* diversification.

market.⁴ In this paper, we consider the possibility that SOEs' objective functions incorporate these externalities, presumably due to political pressure, making them more likely to pioneer.

Only a few studies analyse the micro dynamics of the pioneer process empirically. Due to the availability of granular, product-level customs data in many countries, most of these studies focus on pioneering in export markets. Evidence from Peru (Freund and Pierola, 2010), Mexico (Iacovone and Javorcik, 2010), and Chile (Wagner and Zahler, 2015) suggests that the addition of a new product to a country's export basket is a rare event, indicating high sunk costs of exploring the profitability of new products. The observation that other firms tend to follow pioneers (Iacovone and Javorcik, 2010), especially if the pioneers are successful and survive (Wagner and Zahler, 2015), suggests pioneer-to-follower spillovers as predicted by theoretical models. In this paper we seek to assess whether pioneers' ability to "pick winners" and attract followers varies with their ownership.

We also build on Hausmann and Neffke (2019) and Jara-Figueroa et al. (2018), who study the workforce of pioneer firms in Eastern Germany and Brazil, respectively. The two studies empirically confirm that human capital is critical for the diffusion of industries within a country. Jara-Figueroa et al. (2018) show that the success of pioneer firms, measured by growth and survival rates, depends on the industry-specific and location-specific knowledge of their workforce. Hausmann and Neffke (2019) show that the re-industrialization of Eastern Germany after the reunification was driven by pioneer firms hiring highly skilled labour with industry-related knowledge from Western Germany. While these pioneer plants relied heavily on non-local hires, they did also create jobs for local workers. We add to this literature by studying the role of ownership in the success of pioneer firms and in the employment growth of their pioneered industries. While we do not have detailed information on the workforce of Vietnam's pioneer firms, we can quantify the employment that pioneers create in new activities by tracking the employment growth of the industries they establish.

Our study also adds new evidence to a large literature documenting the effects of foreign direct investment (FDI) on host economies. A large body of work argues that FDI inflows causes productivity improvements, because MNEs bring superior technology, skills, and managerial practices that can spill over to domestic private competitors within the same industry (horizontal spillovers) or to domestic private suppliers and customers through the supply chain (vertical linkages)

⁴ A recent paper by Wei et al. (2021) uses a structural estimation framework to show that "missing pioneers" due to the cost-discovery externalities is a low-probability event in Chinese electronics export data. The externality may thus be less influential than previously believed.

(Aitken and Harrison, 1999; Görg and Greenway, 2004; Javorcik, 2004; Markusen and Venables, 1999; Newman et al., 2015, 2020; Rodríguez-Clare, 1996). However, it is not just that FDI inflows help domestic firms operate more efficiently, MNEs may also help them do entirely new things, overcome "idea gaps" (Romer, 1992, 1993), and thereby contribute to structural transformation (Moran, 2011).

Several articles have already explored whether FDI affects quality upgrading (Bajgar and Javorcik, 2020; Poupakis, 2020) and diversification into complex products (Eck and Huber, 2016; Javorcik et al., 2018; Sørensen, 2020) *within* domestic firms. In contrast, we are in this paper interested in the effect of MNEs on *regional* diversification. A small literature shows that MNEs may speed up, and even break, the path-dependent diversification process of regions and countries⁵ – either directly by founding new industries or indirectly when spillovers from their operations enable other firms to pioneer.⁶ We add to this literature by studying whether MNEs in Vietnam have contributed to industrial diversification by pioneering new industries that crowd-in domestic emulators.

We also contribute to a rich literature exploring the role of SOEs in economic development. It is well documented that SOEs operate with lower productivity and profitability than private firms (Megginson and Netter, 2001) and that privatization is associated with increased firm-level productivity (Brown et al., 2006; La Porta and Lopez-de-Silanes, 1999; Loc and Tran, 2016). Moreover, preferential access to resources also makes SOEs sources of misallocation (Hsieh and Song, 2015; Nigmatulina, 2021) and their public connections contribute to the erection of barriers to entry for private firms (Baccini et al., 2019; Brandt et al., 2020). It is therefore not surprising that state sector reforms have contributed to macro-level productivity growth in both China (Brandt et al., 2008; Hsieh and Song, 2015) and Vietnam (McCaig and Pavcnik, 2017).

Other studies within this literature try to understand why SOEs continue to exist, especially in many developing countries. The *economic raison d'être* for SOEs is well-known and revolves around the existence of market failures (Megginson and Netter, 2001). Political economy explanations view

⁵ On the path dependency of regions and countries see Neffke et al. (2011) and Hidalgo et al. (2007).

⁶ Elekes et al. (2019) show that MNEs drive unrelated, regional diversification through the *direct channel* in Hungary. Lo Turco and Maggioni (2019) find that the presence of MNEs also facilitates pioneering through the *indirect channel* via knowledge spillovers to local firms in Turkey, and MNEs enhance domestic firms' ability to introduce new product-destination export varieties in Chinese regions (Mayneris and Poncet, 2015).

SOEs as vehicles for rent seeking (Shleifer and Vishny, 1994) or as tools to implement government objectives (Bai et al., 2006, 2000; Lin and Tan, 1999). Several studies show that SOEs are used to ensure a low unemployment rate (Azmat et al., 2012; Berkowitz et al., 2017), and that SOE employment surges prior to elections (Cahan, 2019; Li et al., 2020), in tandem with managers' promotion cycles (Gu et al., 2020), and in response to threats of civil unrest (Wen, 2022). We relate to this literature by suggesting a so far under-explored function of SOEs. Specifically, we show that SOEs may have contributed to industrial diversification, especially in peripheral Vietnamese provinces.

Finally, only a few studies have previously looked at the relationship between SOEs and innovation. Bortolotti et al. (2019) find that European SOEs innovate more on aggregate—measured as the number of patents—but produce fewer patents per dollar of R&D spending. In contrast, Cao et al. (2020) show that Chinese SOEs invest more in R&D, produce more patents, and are more efficient in terms of patents per R&D spending. Yet, no study has thus far directly investigated the role of SOEs in industrial diversification.

The remainder of this paper is structured as follows: Section 2 provides context on the economic reform process in Vietnam. In Section 3, we introduce our data, and in Section 4, we provide a set of stylized facts on the pioneering process in Vietnam. Section 5 presents our conceptual framework. We test a number of implications and mechanisms derived from the framework in Sections 6 (focusing on the pioneers) and 7 (focusing on industry dynamics). Section 8 closes the paper with a discussion and conclusion.

2. ECONOMIC REFORMS AND DEVELOPMENT IN VIETNAM

Vietnam's economy has undergone a remarkable transformation over the past century. In 1954, the country was divided into a communist North and a capitalist South. In the North, private enterprises and private ownership were forbidden. After the reunification in 1975, the communist system was extended to the South and all private enterprises were nationalized. In 1986, the government initiated the *Doi Moi* ["renovation" or "innovation"] reforms, following disappointing agricultural and industrial progress during the five-year plans of 1976-1980 and 1981-1985 (Van Arkadie

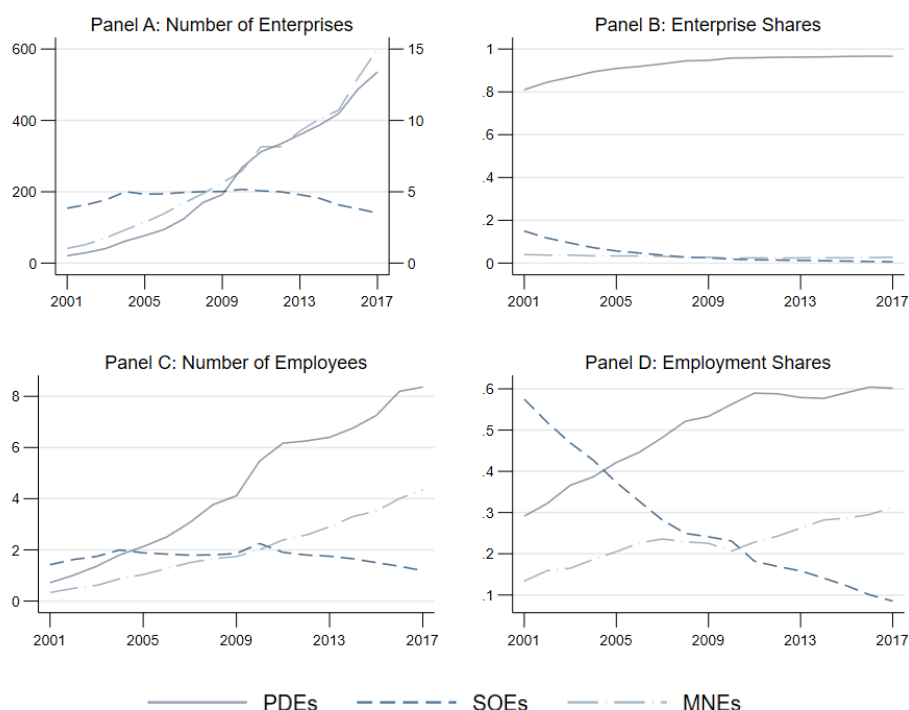
and Mallon, 2004). The reforms aimed to create a socialist-orientated market economy, and thus gradually introduced free-market elements into the thinking of the country's development strategy. At the heart of the restructuring of Vietnam's economy were a number economic reforms that redefined the role of the state, the private, and the foreign sector.

With the initiation of *Doi Moi*, the role of the private sector was acknowledged as an important element in the country's economic development for the first time. The enterprise law of 1990 allowed and encouraged the establishment of private enterprises and its successor from 1999 further opened the economy to private enterprises by lowering bureaucratic and legal requirements (Van Arkadie and Mallon, 2004). The enterprise law of 2005 eventually aligned the legal and bureaucratic regulations faced by private and state-owned enterprises – *de jure* leveling the playing field (Brundenius and Le Dang, 2014; McCaig et al., 2020; Tran and Tien, 2008). *De facto*, however, it was not before the minimum wage reform in 2012 that PDEs could compete with SOEs on (somewhat) equal terms (Hansen et al., 2016).

Accompanying the reforms of the private domestic sector, a series of reforms to the laws governing foreign investment and foreign ownership were implemented. The first law on foreign investment came into effect in 1988. It was amended several times, gradually reducing the administrative burdens for foreign investors, granting tax exemptions, allowing joint-ventures with private domestic firms, and decreasing the number of sectors barred from foreign investment (Tran and Tien, 2008; Van Arkadie and Mallon, 2004). While MNEs were initially concentrated mainly in a few strategic sectors, the reforms enabled them to diversify into other export-intensive sectors. By 2000, MNEs accounted for about 30 percent of output in the manufacturing sector (Van Arkadie and Mallon, 2004). Since then, MNEs have been shown to be a contributor to the successful economic growth process (Anwar and Nguyen, 2010) and poverty reduction (Bui et al., 2019; Do et al., 2021; Hemmer and Nguyen, 2002), and have significantly contributed to the technological upgrading of enterprises through knowledge transfers and spillovers to domestic firms (Newman et al., 2015; Nguyen and Nguyen, 2008).

The gradual equitization of the state-sector constituted another pillar in Vietnam's reform process. The initial goal of the equitization process was to increase competitiveness and spur manufacturing development, but also to reduce the fiscal burden of unprofitable SOEs (Loc and Tran, 2016). Between 1989 and 1992, Vietnam reduced the number SOEs from over 12,000 to around 6,500. The

Figure 2: Enterprises and Employment by Ownership Type



Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. The scales of the y-axes in Panel A are expressed in thousands of enterprises. The left axis refers to PDEs, the right axis to SOEs and MNEs. In Panel C, the y-axis is expressed in millions of employees.

reduction happened mainly through mergers and consolidation of small and locally owned SOEs (Anh, 2005; McCaig et al., 2020; Van Arkadie and Mallon, 2004). Subsequently, the government began a pilot stage of equitization with the goal of equitizing small and *profitable* SOEs (Loc et al., 2006). In practice, however, it was mainly small and *unprofitable* SOEs that were equitized. The interest of outside investors was therefore limited, and the tendered shares were often sold under-priced to managers and employees in the equitized SOEs (Anh, 2005; Loc et al., 2006). The number of equitized enterprises, the revenue for the state, and the desired efficiency gains remained limited in the pilot stage. After the slow pilot stage, the equitization process gained pace in 1998 and equitized enterprises became more efficient and profitable (Loc and Tran, 2016; Loc et al., 2006; Ngo et al., 2015). Equitization, however, remained partial. In many cases, fully state-owned enterprises were transformed into joint-stock companies with the state as a share holder.

The reforms to the private, foreign, and state sectors completely changed the relative composition of formal firms in the economy.⁷ Figure 2 shows the development in the number and shares of

⁷ See McCaig and Pavcnik (2017) for a detailed portrait of the structural transformation process in Vietnam.

enterprises and employees accounted for by SOEs, PDEs, and MNEs. The figures clearly show the rise of the private domestic sector. The number of PDEs grew from around 25,000 in 2001 to more than 550,000 in 2017 (Panel A) and the number of employees grew from around 700,000 to more than 7.3 million (Panel C). The number of enterprises and employment of MNEs grew with a similar pace from around 1,000 firms and 325,000 employees in 2001 to almost 15,000 firms and more than 4.2 million employees in 2017. Accordingly, the share of total employment accounted for by PDEs and MNEs have increased dramatically (Panel D). At the same time, the relative size of the state-sector has shrunk, despite the fact that state employment and the total number of SOEs have remained relatively constant. While 56.9 percent of all formal employment was held by SOEs in 2001, this share reduced to only 8.9 percent by 2017.

Vietnam's reform process has been a remarkable success story. With the initiation of *Doi Moi* in 1986, the Vietnamese GDP per capita grew with an average of 6.6 percent per year. Different versions of the Vietnamese success story attribute the success to the role of the private, foreign, and state sectors. One narrative argues that Vietnam is a case of state-led development: government interventions and a strong state-sector have pushed Vietnam towards the technological frontier and economic growth (Malesky and London, 2014). In fact, official policy declarations in Vietnam directly argue that SOEs should drive innovation, promote strategic industries, and facilitate integration into global value chains (Anh, 2014).⁸ An alternative narrative attributes Vietnam's success to the rise of the private sector following the *Doi Moi*, which legalized private ownership of enterprises and removed the red tape of doing business. An extension of this market-driven development hypothesis posits that Vietnam has benefited from its "backwardness" through the inflow of FDI and foreign technologies, which have played a key role in creating jobs and facilitating productivity growth (Malesky and London, 2014; Newman et al., 2015). In the rest of this paper, we add new evidence to this debate by studying the role of PDEs, SOEs, MNEs in Vietnam's diversification process.

⁸ For example, Vietnam's Master Plan on Economic Restructuring 2013-2020 (Decision No. 339/QĐ-TTg) states that the activities of SOEs should be centered in "*sectors like military industries, fields and industries of natural monopoly, provision of primary goods and services, and some fundamental industries and high technologies of strong pervasiveness*" (Government of Vietnam, 2013).

3. DATA

We use data from the Vietnamese Enterprise Surveys (VES) for 2001-2017, collected by the General Statistics Office (GSO).⁹ The VES data allow us to follow nearly one million firms over time. It covers the universe of registered businesses under Vietnam's Enterprise Law from 2001-2004 and in the census years of 2007, 2012, and 2017.¹⁰ In all other years, GSO also surveyed the universe of SOEs, MNEs, and PDEs with more than 10 (2005-2010), 20 (2011-2015), or 100 (2016-2017) employees. These enterprises accounted for around 55.8 percent of all firms, 97.4 percent of total employment, and they operated in 98.2 percent of all industries in 2001, and for 5.2, 66.1, and 92.5 percent, respectively, in 2017.¹¹ In addition, all enterprises within a time-varying selection of provinces and sectors are surveyed each year, along with 15-20 percent of smaller private enterprises falling below the employment thresholds. In some years and some provinces, there are small deviations from the sampling procedure outlined here. For a detailed description, see Appendix A.4.

3.1. Key Variables and Firm-Level Data Set

We use detailed information from the VES data on the ownership structure of firms, their administrative location, and their economic activity to create indicators for ownership and pioneering activity. Below, we provide a brief description of these key variables.

We construct three ownership types: SOEs, MNEs, PDEs. We define SOEs as enterprises that are either *(i)* fully state-owned, *(ii)* limited liability companies (LLCs) with at least one state member, *(iii)* joint-stock companies (JSCs) with state capital, or *(iv)* joint ventures (JVs) with state-ownership. We define MNEs as being either fully foreign owned or a JV between a foreign enterprise and a private sector partner. We classify all remaining companies as PDEs.¹²

⁹ Appendix A.4 has a detailed description of the data.

¹⁰ All SOEs, MNEs, and collectives are legally obliged to register under the Enterprise Law. Private businesses that have more than 10 employees or operate in more than one location are also required to register as an enterprise to legally operate. Smaller businesses may choose to register either as a household business or an enterprise (Government of Vietnam, 2000, 2004).

¹¹ Calculations are made by the authors and based on a cleaned VES data set.

¹² Our definition of SOEs deviates slightly from the official definition from GSO. After GSO's definition, JSCs only count as SOEs if they have more than 50 percent state capital. GSO further does not include JVs with state-membership into their SOE definition (GSO, 2019). See Appendix A.1.1 for a detailed discussion on our SOE definition.

The VES data has province-level geo-locations of firms. Vietnam has 63 provinces.¹³ Provinces constitute the second-level administrative units in Vietnam, after regions. They have historically enjoyed significant political autonomy and have been encouraged to implement and test regulatory experiments locally since the early 2000s. This means that similar firms can face very different regulatory conditions depending on the province in which they are located (Jaax, 2020; Rama, 2008). For these reasons, we use provinces as our geographical unit for defining pioneers.

The economic activity of firms is available at the 4-digit level using the Vietnamese Standard Industrial Classification (VSIC). VSIC codes loosely follow the International Standard Industrial Classification (ISIC) system. The VSIC system was updated in 2007. From 2001-06, firms only reported their activities under the old classification system; from 2006-10, they reported under both the new and the old system; and since 2010 they only reported under the new system. We use descriptions of each industry to manually create a cross-walk between the two classification systems. For some firms, the old and the new industry codes do not align in years where they report under both systems. To deal with these inconsistencies, we assign firms to industries based on the new codes, if the old and the new codes align at the 2-digit level as a minimum (see Appendix A.4 for details). Further, there is a disproportionate amount of pioneering in certain industries due to an unrealistically high rate of short-termed, temporary industry switching. This pattern is especially prevalent in 2006 and 2010, when firms start and stop reporting under multiple classification systems. We take this to indicate confusion among the firms' respondents and address it in two ways. First, the switching is particularly high between sub-sectors within retail; wholesale; other transport; electricity distribution, and auction and brokerage; education; and monetary and financial activities. We aggregate these sub-sectors. Second, we remove temporary, one-year industry switches to further reduce noise.¹⁴

We use the firms' geo-location and VSIC codes to identify their pioneering activity. With inspiration from Jara-Figueroa et al. (2018), we define a firm as a pioneer when it establishes a new 4-digit industry in a province, given that there has been no activity in that industry in that province

¹³Vietnam had 64 provinces before Hà Tây province was merged with Hanoi in 2008. We collapse the two province codes into one for the entire period of our analysis. We also drop all firms (1,577) in Vietnam's military districts.

¹⁴This is straightforward for single-industry firms, but slightly more involved for multi-industry firms. See Appendix A.1.2 for details.

for at least three years prior to the pioneering event. We allow pioneers to enter industries that existed more than three years prior to the pioneering event since it is likely that these industries disappeared again because they were not marketable at that point in time. Trying to establish them again thus still requires a pioneer to face the risk and uncertainty associated with cost-discovery (Hausmann and Rodrik, 2003). The indicator varies within firms over time. We change this somewhat arbitrary definition of pioneering in a number of checks.

We use employment, not revenue, to classify whether an industry is pioneered, because the employment information in the VES data is better populated and more reliable than the financial information. When identifying pioneering, we further use the fact that we can observe firms' employment across multiple industries. The multi-industry aspect turns out to be very important: Approximately 8 percent of firms operate in more than one industry and they account for almost 23 percent of all pioneering activity. While we use multi-industry information to identify pioneering firms, we conduct most of our analysis at the firm-level. We use the industry with the highest revenue to classify the main activity of multi-product firms.

Summarized, we end up with a firm-level data set of 918,880 unique firms and 4,101,013 firm-year observations.¹⁵ For the analysis in Section 6, we exclude the first three years from the sample, as they can, by definition, not have any pioneers. We further take the logarithm of revenue, capital, wages, labor, and intermediaries, which generates missing entries due to zeros in the data. This reduces the number of observations to 3,011,767.

3.2. Aggregated Data for Pioneered Industries

Besides the firm-level data set, we also construct an aggregated data set of local industries that have been pioneered at some point during our sample period. We use the province-industry-level panel to analyse the evolution of industries started by different pioneers using a difference-in-differences set-up with staggered adoption. We therefore exclude all non-pioneered industries, because they are irrelevant for counterfactual analysis of pioneering dynamics.¹⁶ The data contains 101,116

¹⁵Note that this number deviates from the numbers shown in Table A22, which is due to the more involved industry and pioneer cleaning as described above.

¹⁶Some local industries are never pioneered and could be used as a never-treated control group. As we show below, however, there is economic activity in some pioneered industries prior to the pioneering event. Using never-pioneered industries as controls will likely bias results upwards because prior economic activity in

observations. Pioneering events have occurred in all 63 provinces and across 227 unique 4-digit VSIC industries.

In the analysis, we focus on the evolution of total employment and number of firms in pioneered industries. In the three years before the pioneering event, there is (by definition) no activity in the industry-province pairs and these observations are therefore missing. Because there may be activity in the local industries at earlier event times, we replace the missing observations with zeroes, which is the natural lower bound for both employment and number of firms. In 133 instances, a local industry is simultaneously pioneered by firms with different ownership structures. These hybrid industries pose a classification problem when assigning industries to pioneers. As a baseline, we assume that the political connections of SOEs make them influential lead firms. We therefore classify all hybrid industries, where we observe an SOE pioneer, to be SOE-pioneered industries. Following a similar logic, we assign industries to be MNE-pioneered if they are simultaneously pioneered by a PDE and an MNE. Our results are robust to leaving out these hybrid industries from our analysis.

4. THE PIONEERING PROCESS IN VIETNAM: SOME STYLIZED FACTS

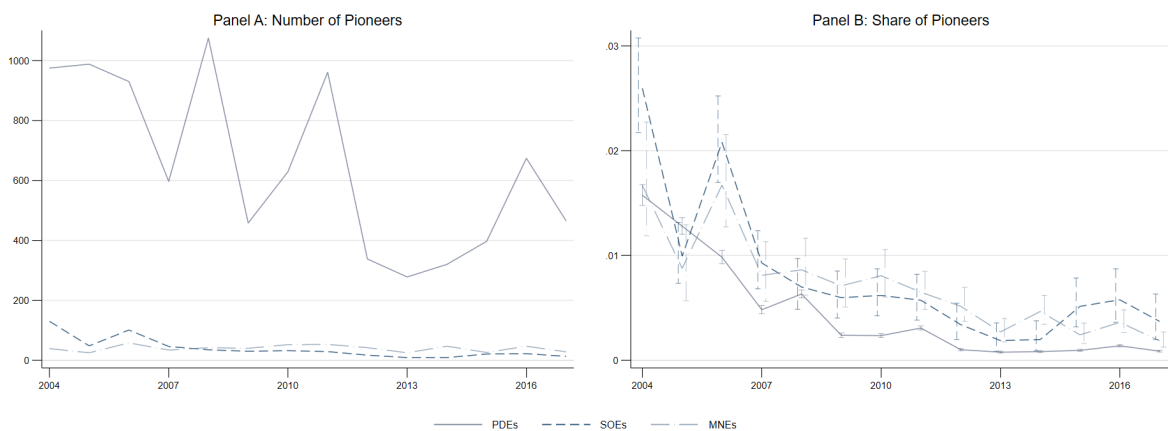
This section presents a number of stylized facts about the pioneering process in Vietnam. We focus in particular on how this process is shaped by firms with different ownership structures. First, we show that the raw propensity to pioneer is higher for SOEs and MNEs compared to PDEs. Next, we show qualitative differences in the pioneering activities of firms: MNEs pioneer in complex industries in Vietnam's economic centers, while SOEs tend to pioneer in unsophisticated industries in rural and poor areas. Finally, we focus on the dynamics of the pioneering process and show that industries pioneered by different firm types exhibit different dynamics. The monopoly period tends to be longer in SOE-pioneered industries, and these industries also have lower post-pioneering employment growth. In contrast, there seem to be a substantial employment premium in MNE-pioneered industries.

these industries will be lower. It is also not feasible to use already existing industries as controls in the presence of dynamic treatment effects (Goodman-Bacon, 2021).

4.1. Ownership and Pioneering

Figure 3 plots the number of pioneering SOEs, MNEs, and PDEs from 2004 to 2017 in Panel A, and the share of pioneers within each ownership category in Panel B. In absolute terms, the majority of pioneering takes place in the private domestic sector. Yet, the share of pioneers among SOEs and MNEs is consistently higher than among PDEs (except in 2005).¹⁷ We show in Appendix A.2 that these results are robust to a wide range of different pioneering definitions. Taken at face value these statistics help to reconcile different explanations of Vietnam's industrial transformation. The continued reforms of the private sector have enabled Vietnam's private firms to take a leading role in the local industrial diversification process. Moreover, the high propensity to pioneer among SOEs and MNEs is both consistent with the state-led development hypothesis and the idea that Vietnam has benefited from its backwardness through the inflow of FDI and adoption of foreign technologies.¹⁸

Figure 3: Number Share of Pioneers by Ownership Type



Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. The vertical bars in Panel B represent 95 percent confidence intervals.

A part of these differences in pioneering is likely driven by differences in observables. The upper part of Table 1 shows differences between pioneering and non-pioneering firms. Pioneers are larger

¹⁷The VES data starts in 2001. As pioneers are defined based on the past three years, 2004 is naturally the first year in which we can observe any pioneers. Despite significant fluctuations over time, there is a decreasing trend in the number of pioneers. This trend is, at least to some extent, mechanic. Following our definition, the rate of pioneering has to decline in the long run as diversification eventually exhausts the number of not-yet-existing 4-digit industries in each province.

¹⁸See Malesky and London (2014) for a discussion of these diverging narratives.

Table 1: Summary Statistics by Pioneering Status and Ownership Type

	Pioneer	Ever Pioneer	Never Pioneer
Number of Observations	10,185	64,547	4,036,466
Number of Enterprises	9,670	9,670	909,210
Mean Number of Employees	67	103	33
Mean Value Added (in mio VND)	12,750	15,111	3,388
Mean Capital (in mio VND)	73,988	93,289	51,237
Mean Intermediaries (in mio VND)	120,370	127,677	35,212
Mean Total Wage (in mio VND)	3,750	6,144	1,950
Share Start-Up	0.665	0.141	0.219
Share Multi Industry	0.192	0.187	0.039
	SOEs	PDEs	MNEs
Number of Observations	78,052	3,914,716	108,245
Number of Enterprises	11,449	897,091	20,452
Mean Number of Employees	376	19	319
Mean Value Added (in mio VND)	59,854	1,236	33,766
Mean Capital (in mio VND)	563,277	35,976	227,377
Mean Intermediaries (in mio VND)	437,983	20,575	235,188
Mean Total Wage (in mio VND)	25,554	935	20,797
Share Start-Up	0.066	0.222	0.158
Share Multi Industry	0.288	0.037	0.036
Share Pioneers	0.007	0.002	0.005
Share Shut Down	0.030	0.091	0.040
Share Industry Exit	0.127	0.156	0.082
Share Incumbents	0.844	0.702	0.783
Share Followers	0.029	0.053	0.036

Notes: In the upper panel, Column 1 gives statistics on pioneers in the year they pioneer. Column 2 presents characteristics of pioneering firms over their entire lifespan. Column 3 reports statistics on firms that never pioneer. In the lower panel, *Shut Down* refers to firms who stop operating in year t . *Industry Exit* refers who to firms stop operating in a given industry, without shutting down their business entirely. *Incumbents* refers to firms that operating in the same industry as they have been operating in the previous period, $t-1$. *Followers* refers to firms who stop operating in an industry and switch to a different, already existing industry.

than non-pioneers in terms of the number of employees, value added, capital, intermediaries, and total wage. This is true both in the year where they pioneer as well as over the whole observation period. The vast majority of pioneers are either start-ups (66.5 percent) or multi-industry enterprises (MIEs) (19.2 percent).

In the lower part of Table 1, we report similar statistics by ownership type. On average, SOEs and MNEs are substantially larger than PDEs, and SOEs are larger than MNEs.¹⁹ Moreover, SOEs

¹⁹This is further displayed in Figure A1 in Appendix A.2, where we show density plots for employment, revenue, capital, and intermediaries. The figure shows that MNEs and SOEs are, on average, larger than

are more likely to operate in multiple (4-digit) industries, but are least likely to be start-ups. The highest start-up rate is found among PDEs. In Table A5 in Appendix A.2, we show that PDEs and MNEs also tend to pioneer as start-ups, while the majority of pioneering SOEs are multi-industry firms. We also note that SOEs are least likely to shut down, but more likely to exit their industry than MNEs.

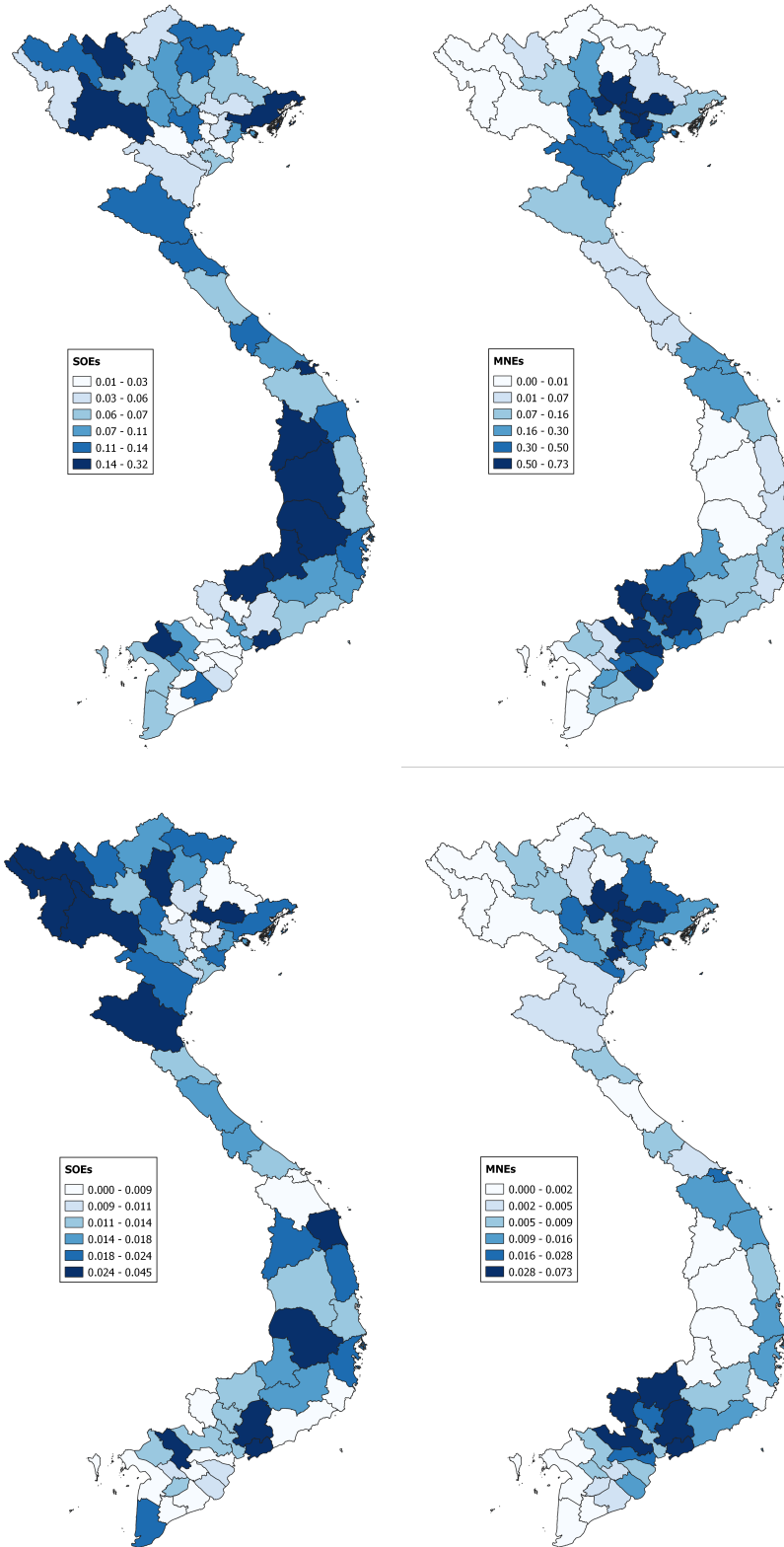
Besides differences in pioneering propensities, there are important qualitative differences in the pioneering activities of firms with different ownership structures. MNEs tend to pioneer in complex industries in Vietnam's economic centers, while SOEs tend to pioneer in unsophisticated industries in rural and poor areas. These patterns are consistent with the idea that MNEs push the knowledge frontier, while SOEs diffuse knowledge to peripheral regions.

Figure 4 shows that SOEs and MNEs pioneer in different parts of Vietnam. The figure plots the provincial employment share of SOEs (left) and MNEs (right) in 2017 in the top panel. The bottom panel shows the geographical distribution of pioneering activity for each ownership type, where the shares in each province sum to one. The concentration of firms' pioneering activity largely mirrors their relative economic dominance. MNEs are highly concentrated in and around the two economic centers of Hanoi and Ho Chi Minh City, where they account for more than 50 percent of all formal employment. The spatial distribution of SOEs is less concentrated. They make up a larger share of total employment in the less developed central highlands and in the mountainous northern border region (Bui et al., 2019). These regions are also home to a large fraction of Vietnam's minority groups. The placement of SOEs may reflect policy objectives of stimulating local growth and reducing unemployment rates to maintain social stability.²⁰

Figure 5 shows that SOEs, MNEs, and PDEs pioneer in different industries. The figure illustrates the novelty and complexity of the industries pioneered by firms of different ownership type. The leftmost figure reports the number of other provinces in which an industry is already present, when it is pioneered by an SOE, MNE, or PDE. The figure illustrates that the pioneering process in Vietnam is a story of diffusion rather than innovation: most pioneering occurs in industries that are already established in several other provinces. Yet, the novelty of pioneering varies significantly by firm type. About 75 percent of the local industries started by MNEs were present in less than 30

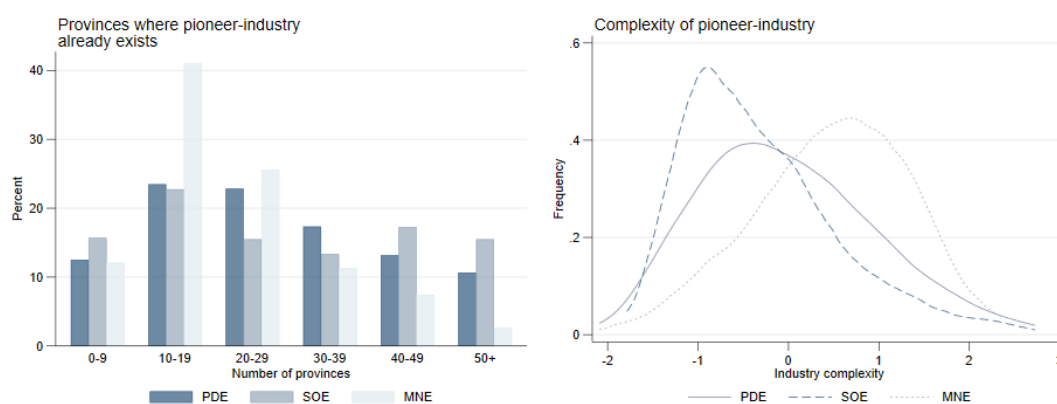
PDEs (as seen in Table 1), but also that their size distributions overlap to a large extent.

²⁰Studies on the multitask role of the state sector include Azmat et al. (2012); Bai et al. (2000, 2006); Lin and Tan (1999); Berkowitz et al. (2017); and Wen (2022).

Figure 4: SOE and MNE Employment and Pioneering Shares by Province (2017)

Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. The categorization follows Jenks natural breaks classification method (Jenks, 1967).

Figure 5: Novelty and Complexity of Pioneered Industries by Ownership Type



Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. Economic Complexity is calculated using as described in the appendix using the approach in (Hidalgo and Hausmann, 2009)

other provinces at the time of pioneering. These novel industries account for only about 60 and 50 percent of the pioneering activity for PDEs and SOEs, respectively.²¹

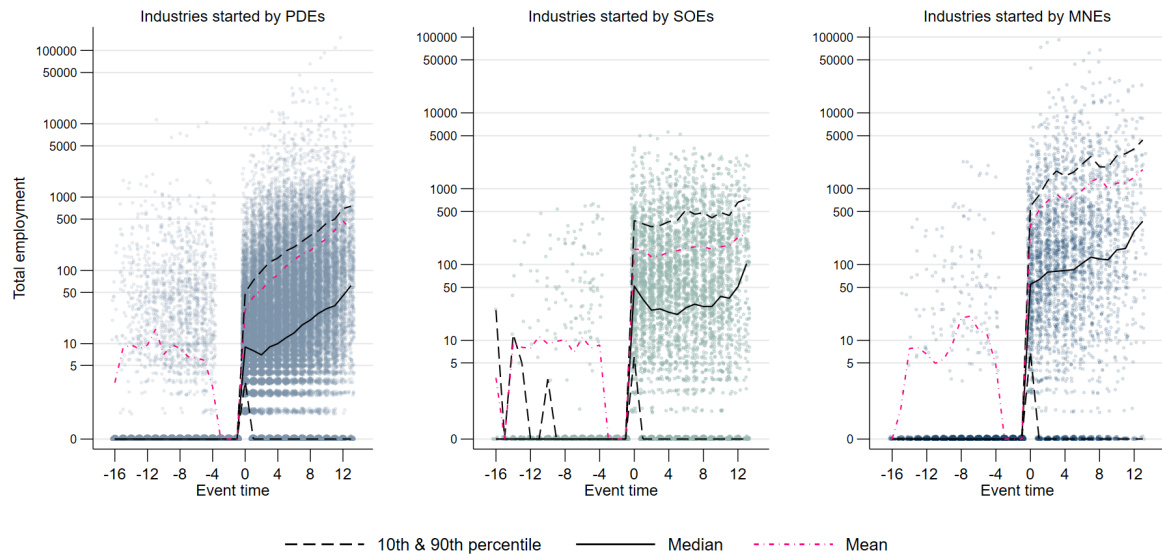
In the right panel of Figure 5, we show kernel density plots of the complexity of industries pioneered by MNEs, SOEs, and PDEs. We calculate industry complexity using the Method of Reflection from Hidalgo and Hausmann (2009). Where they apply the method to the country-product network encoded in international trade data, we apply it to the province-industry network in Vietnam.²² Intuitively, the measure of industry complexity is based on the assumption that complex industries are difficult to establish and therefore are only present in a few, complex provinces. The density plot corroborates the general finding from the FDI literature that MNEs operate with superior knowledge and technology (Javorcik, 2004; Newman et al., 2015) and enter complex industries in the host economy (Javorcik et al., 2018).

4.2. Ownership and Industry Dynamics

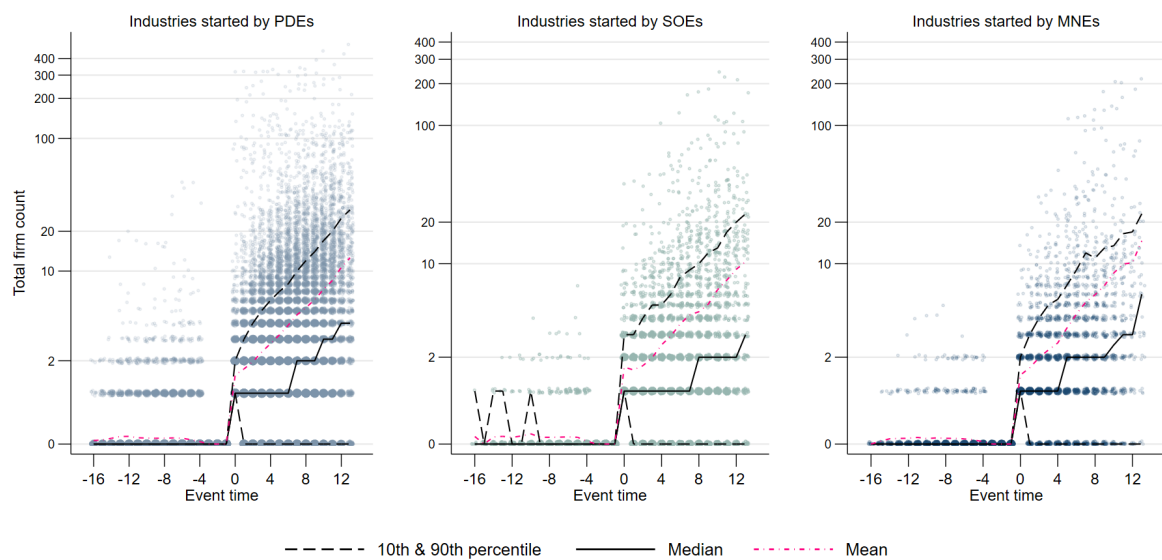
We now turn to the dynamics of the pioneering process. To investigate which pioneers establish industries that grow, we use the aggregated data set containing all local industries that were pioneered at some point between 2004 and 2017. Figures 6 and 7 can be understood as descriptive event study

²¹In Figure A2 in Appendix A.2, we show that pioneering events tend to occur in industries that are entirely new to the provinces where they are started. There are no notable differences between different firms' tendency to pioneer in novel industries.

²²A detailed description is available in Appendix A.2.3.

Figure 6: Employment Dynamics in Infant Industries by Ownership of Pioneer

Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. Each dot represents a province-industry in relative time to pioneering. Total employment numbers are transformed using the inverse hyperbolic transformation.

Figure 7: Firm Dynamics in Infant Industries by Ownership of Pioneer

Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. Each dot represents a province-industry in relative time to pioneering. Total firm counts are transformed using the inverse hyperbolic transformation.

plots. They are scatter plots with relative time to the pioneering event on the horizontal axes and the inverse hyperbolic transformation of employment and firm counts on the vertical axes. Each dot in the figure represents a province-industry-year observation. We show dynamics in industries started by PDEs in the leftmost figure, SOEs in the middle, and MNEs in the rightmost figure. In each figure, we plot the mean and median along with the 10th and 90th percentiles.

A number of messages can be taken from these plots. First, they show that there are activities in some industries before they die completely and are re-established by pioneers. Eight years before the pioneering event, however, less than 10 percent of the industries are active. There are no substantial differences in these pre-trends across industries pioneered by different firms. Second, the graphs clearly show that pioneering is risky. At least 10 percent of all industries die again within the first year of pioneering. This applies irrespective of ownership of the pioneer firm. Third, SOEs and MNEs are large when they pioneer. The median number of workers in year zero is around 50 in SOE- and MNE-pioneered industries compared to 10 in PDE-pioneered industries. Fourth, industries started by SOEs appear less competitive and exhibit lower growth rates than industries pioneered by PDEs and MNEs. We show in Figure 7 that the median industry is a (local) monopoly for several years after the pioneering event – irrespective of the ownership of the pioneer firm. Yet, the monopoly period is substantially longer in industries pioneered by SOEs. Figure 6 shows that the employment growth rate in SOE-pioneered industries is much lower than in PDE- and MNE-pioneered industries. In the latter industries, the median, mean, and 90th percentile grow at a factor of around 10 over 13 years. Fifth, because MNE-pioneered industries are larger than PDE-pioneered industries from then beginning, and they exhibit similar growth paths, there seem to be a substantial employment premium in MNE-pioneered industries. Finally, SOEs do not seem to hit ‘big winners’. Yearly employment reaches 5,000 in the most successful SOE-pioneered industries, while MNEs and PDEs have started industries employing around 100,000 people in a single province.

5. CONCEPTUAL FRAMEWORK

We propose a simple conceptual framework to help understand the patterns we observed in Section 4. The framework combines the set-up in Verhoogen (2021) with theoretical models of pioneering (Hausmann and Rodrik, 2003; Hausmann and Klinger, 2006; Hausmann et al., 2007; Wei et al., 2021), and it is general enough to accommodate the main mechanisms highlighted in the literature linking firms' characteristics to their choice of industry (re)location. Importantly, it also allows for different market failures and that firms operate with different objective functions.

5.1. A Simple Framework

Consider an economy consisting of K potential industries, $k \in \Omega$, and N firms, $i \in \Psi$. It is useful to think of i as indexing three representative firm types – PDE, SOE, and MNE – to capture the main ideas in the framework. Firms produce in a subset of industries $\Omega^P \in \Omega$, while the remaining industries are unpopulated $\Omega^U = \Omega - \Omega^P$. For simplicity, assume that firms operate under bounded rationality: they are partially forward-looking and base their current choices on the expected pay-off in the next period only.²³ Because this set-up is highly simplified, we will leave out time subscripts in the following notation to prevent clutter. In any period, firm i must choose between four mutually exclusive strategies in order to maximize its value in the next period (V_i). The firm can choose (V_i^I) and stay as an incumbent in its current industry, it can pioneer a new industry and get (V_i^P), it can choose (V_i^F) and follow other firms into an established industry, or it can exit and get ($V_i^E = 0$). Formally, the firm's value in the next period can be written as:

$$V_i = \max (V_i^I, V_i^P, V_i^F, 0). \quad (1)$$

To understand the firm's maximization problem, it is instructive to explicitly specify the value functions associated with each choice. We build up from below. First, assume that firms produce output with an industry-specific production function:

²³The assumption abstracts from any strategic planning by the firm that goes several periods into the future. The assumption greatly simplifies the firm's problem, while it still allows us to describe the main points from the literature.

$$Y_{ik} = F_k(M_{ik}, \delta_{ik}, \alpha_{.k}), \quad (2)$$

where Y_{ik} is physical output and M_{ik} is a vector of inputs. We further assume that the firm's productivity has a firm-industry-specific component, δ_{ik} , and an industry-specific component, $\alpha_{.k}$. δ has subscript k because we allow the firm-specific productivity to vary depending on the industry in which the firm chooses to operate. α has subscript $.k$. The dot in the place of i makes explicit that the industry-level productivity component not only depends on factors intrinsic to the industry, such as technology, but also on attributes of other firms through spillovers and agglomeration economies. Each firm only knows $F_k(\cdot)$ for a subset $k \in \Omega$. Following Verhoogen (2021), we denote this subset of k 's and δ_{ik} as a firm's know-how.

Let $P_{ik} = D_k(Y_{ik}, Z_{.k}^y, \Lambda_{ik})$ be the firm's inverse demand curve. Because the price depends on the output of firm i , we assume that it has some price-setting power (this will be especially pronounced when the firm is pioneering). The demand curve also depends on output and pricing of other firms in the industry $Z_{.k}^y$, to which firm i responds strategically (Verhoogen, 2021). Λ_{ik} denote firm-specific shifters in price, for instance related to quality of their output. Λ_{ik} may also denote the extent to which the firm values consumer surplus. We shall return to this issue below. Further, let $W_{ik} = S_k(M_{ik}, Z_{.k}^m, \kappa_{ik})$ be the corresponding inverse input-supply curve facing firm i . As in Verhoogen (2021), the input price vector W_{ik} is a function of external factors in input markets $Z_{.k}^m$, such as prices paid by other firms. In addition, let κ_{ik} denote firm-industry-specific shifters in input prices such as differences in minimum wage requirements or availability of external versus internal financing of capital inputs. κ_{ik} also denotes the extent to which the firm cares about the producer surplus of other firms, to which we will turn now.

It is commonly recognized in the literature that SOEs are tasked with multiple objectives or 'policy burdens' (Bai et al., 2000, 2006; Berkowitz et al., 2017; Lin et al., 1998; Lin and Tan, 1999; Wen, 2022). To capture this, we add a policy term to the firm's maximization problem:

$$B_{ik} = \sum_{\substack{j=1 \\ j \neq i}}^N \Phi_{ij} V_j(k_i, W_{ik}' M_{ik}, P_{ik} Y_{ik}), \quad (3)$$

where k_i is firm i 's industry. The term is a weighted sum of the value of all other firms in the

economy, expressed as a function of firm i 's choice of industry, production inputs and outputs, and prices charged/paid. The weights Φ_{ij} determine the value firm i assigns to the value of firm j . This makes explicit that some firms may put a value on the externalities of their activities. The firm's choice-variables do not only affect other firms within the same industry, but also firms in vertically connected industries. For instance, if the firm puts value on the consumer surplus in downstream industries, it will set prices below the profit maximizing level. Likewise, if the firm cares about the producer inputs in upstream industries, it will pay above-market prices on intermediaries. Φ_{ij} will be zero for all js for any profit maximizing firm, whose maximization problem boils down to choosing the optimal M_{ik} given a set of input and output prices.

The value for firm i of staying in its current industry $k = c \in \Omega^P$ is:

$$V_i^I = P_{ic}F_c(M_{ic}, \delta_{ic}, \alpha_c) - W_{ic}'M_{ic} + B_{ic}, \quad (4)$$

If the firm instead chooses to be a pioneer, it will need to pay an additional fixed cost $f_{ik}^P = \rho_{ik} + \phi_k$ to begin operating in that new industry. f_{ik}^P depends on both a firm-specific and an industry-specific component. The industry-specific component, ϕ_k , captures fixed costs in the traditional sense, like investments in fixed capital, but it also denotes expenses related to initial market segmentation and similar discovery costs (Hausmann and Rodrik, 2003; Wei et al., 2021), investments and bribes required to overcome regulatory barriers to entry, and so forth. The firm-specific component, ρ_{ik} , incorporates, for instance, differential access to cheap finance. Following Hausmann and Klinger (2006), it can also be thought of as the distance between a firm's current capabilities and the capabilities necessary for mastering $F_k(\cdot)$. As costs are increasing in capability distance, ρ_{ik} denotes the cost outlay necessary to master k 's production function $F_k(\cdot)$. ρ_{ik} varies by firm and industry because the applicability of firm's current capabilities varies for different industries. ρ_{ik} is low if the firm's know-how already includes $F_k(\cdot)$. When the firm is a pioneer, it also has the possibility of charging monopoly prices $P_{ik} = D_k(Y_{ik}, 0, \Lambda_{ik})$. In other words, it does not have to take into account the behaviour of other firms. Under monopoly, the price-elasticity of consumer surplus is high and a monopolist valuing overall welfare is likely to lower the price. Assuming that similar inputs are used in different industries, the same is not true for W_{ik} .

Because the firm will only pioneer in the industry that maximizes its value, we write its value

function as the solution to a maximization problem over $k \in \Omega^U$:

$$V_i^P = \max_{k \in \Omega^U} [V_{ik}^P] = \max_{k \in \Omega^U} [P_{ik}F_k(M_{ik}, \delta_{ik}, \alpha_k) - W_{ik}^l M_{ik} - (\rho_{ik} + \phi_k) + B_{ik}]. \quad (5)$$

Note that α no longer has dot-subscript because pioneers' productivity are not influenced by other firms already in industry k .

If the firm chooses to follow other firms into an established industry, it will also need to pay a fixed cost, $f_{ik}^F = \rho_{ik} + \phi_{.k}$. We let ϕ have subscript $.k$ to denote that the industry-specific fixed cost is dependent on the behaviour of incumbent firms. For switchers, α will also depend on other firms in their new industry, so we add the subscript $.k$ again. Due to cost-discovery externalities (Wei et al., 2021), the fixed cost f_{ik}^F associated with being a follower is likely smaller than the fixed cost f_{ik}^P associated with being a pioneer. Yet, it does not have to be like this as incumbents may erect barriers to entry (Brandt et al., 2020). If the firm chooses to become a follower and switch industry, its choices are limited to the set of industries that have already been pioneered, $k \in \Omega^P$. The firm will choose the industry that maximizes V_i^F :

$$V_i^F = \max_{k \in \Omega^P} [V_{ik}^F] = \max_{k \in \Omega^P} [P_{ik}F_k(M_{ik}, \delta_{ik}, \alpha_{.k}) - W_{ik}^l M_{ik} - (\rho_{ik} + \phi_{.k}) + B_{ik}], \quad (6)$$

Eqs. (1)-(6) clarify the ways in which firm-heterogeneity may impact the process and choice of pioneering. Heterogeneity stems from differences in the value function that maximises firm value. In turn, these differences are determined by differences in the underlying production function, profit function and policy burden parameters.

5.2. Discussion

In this section, we use the conceptual framework in conjunction with related literature to highlight the potential mechanisms that lead firms with different ownership to shape the pioneering process in a way that is consistent with the patterns observed in Vietnam.

Understanding Differences in Pioneering Probabilities

In Figure 3, we show that SOEs and MNEs have a higher tendency to pioneer than PDEs. Our framework highlights two ways in which ownership may affect firm's pioneering probability. First, firms will decide whether to pioneer or not by evaluating the fixed costs of doing so, $f_{ik}^P = \rho_{ik} + \phi_k$, against the possibility of earning monopoly profits. Differences in ρ_{ik} can either be driven by differences in know-how (knowledge of $F_k(\cdot)$) or by differential access to finance. It is well established in the literature that MNEs operate with superior know-how and technology and can access knowledge and finance from abroad (Görg and Greenway, 2004; Javorcik et al., 2018; Newman et al., 2015; Setzler and Tintelnot, 2021). It is therefore not surprising that we observe a higher propensity of pioneering among MNEs, especially in new and complex industries (Figure 5). It has also been repeatedly documented in Vietnam and elsewhere that SOEs enjoy preferential access to credit from state-owned banks (Baccini et al., 2019; Cao et al., forthcoming; Girma et al., 2008; Malesky and Taussig, 2009). All else equal, this should make it cheaper for SOEs to undertake the necessary investments for pioneering.

Firms will weight the fixed cost of pioneering against potential monopoly profits. As in Hausmann and Rodrik (2003), this ties preferences for pioneering to the existence of entry barriers for emulators. In our framework, higher barriers to entry are captured by a higher fixed cost, ϕ_k , and the pricing behaviour of incumbent firms, Z_k^y . In Figure 7, we find longer monopoly periods in SOE-pioneered industries. This is in line with Baccini et al. (2019), who provide evidence that SOEs operate in industries with higher political/regulatory barriers to entry in Vietnam. Brandt et al. (2020) find similar results in China. In contrast, we find the shortest median monopoly periods in MNE-pioneered industries. This is perhaps surprising, since one would expect credit constrained domestic firms to find it hard to raise the capital necessary to learn the complex production functions in MNE-pioneered industries. But the FDI literature studying barriers to entry and crowding-out is in fact not conclusive. Some studies find a positive effect of FDI on firm-survival (Kosová, 2010; Lu et al., 2017), while others find that the presence of MNEs discourages entry and stimulates exit of domestic firms (De Backer and Sleuwaegen, 2003).

Second, firms' propensity to pioneer depends on the extent to which the policy factor is incorporated in their value function. In theory, B_{ik} should only be non-zero for SOEs, who are

tasked with multiple government objectives that go beyond profit-maximization (Bai et al., 2006, 2000; Lin et al., 1998; Wen, 2022). The policy factor may help explain why SOEs are likely to pioneer. Models on pioneering have, for example, shown that market exploration is associated with positive externalities (Hausmann and Rodrik, 2003; Hausmann and Klinger, 2006; Hausmann et al., 2007; Hoff, 1997; Wei et al., 2021). In Hausmann and Rodrik (2003), firms are uncertain about the local cost structure (and hence productivity, α) of non-pioneered industries. Pioneering brings value to potential emulators because it spreads information about the viability of the pioneered industry through the success of the pioneer. In Wei et al. (2021), some of the industry-specific fixed costs associated with exploring an industry, ϕ_k , is paid only by the pioneer.²⁴ The main conclusion from these models is that pioneering lies below the socially optimal level without state-intervention. If SOEs incorporate (parts of) the pioneering externalities into their value function, they will be more likely to pioneer. Other externalities, such as learning-by-doing (Arrow, 1962), may also be internalized by SOEs. In combination with long-term investment horizons of governments and a stable access to finance from state-owned banks, this should make SOEs more likely to pioneer in industries with a dynamic comparative advantage (Bortolotti et al., 2019; Redding, 1999). Finally, by creating a market that did not exist before and charging below-profit-maximizing monopoly prices, SOEs can create a substantial consumer surplus, which they may value through B_{ik} .

Understanding Differences in Industry Dynamics

In Figures 6 and 7, we showed that industry growth after pioneering varies with the ownership of the pioneer. SOE-pioneered industries experience, as an example, slower employment growth and are characterized by longer monopoly periods. How can one make sense of these patterns? Below, we discuss how the characteristics of the pioneers may affect industry dynamics through a direct and an indirect channel.

Size and the growth of the pioneer itself influence the size and growth of the pioneered industry *directly*. The direct effect will be more dominant in concentrated industries and in the short run where the market share of the pioneer is bound to be high. As discussed above, SOE-pioneered industries are – for various reasons – characterized by longer monopoly periods and the direct effect

²⁴These discovery costs entail, for instance, part of the costs related to obtaining knowledge about suppliers, location and segmentation of customers, and institutions and regulations governing the industry.

will thus be more dominant in these industries. When, on the other hand, pioneering crowds-in follower firms, the importance of the direct effect of industry size and growth will decrease over time.

A-priori uncertainty about the productivity parameters δ_{ik} and α_k may explain heterogeneity in pioneer growth and entry size. In Jovanovic (1982), firms are uncertain about their individual productivity δ_{ik} . All firms share the same beliefs about δ_{ik} and only realise their true productivity post entry. In this setting, productive firms stay and grow, while unproductive firms decline and eventually exit. Empirical literature has persistently documented clear productivity differences between SOEs, MNEs, and PDEs. It is a well-established fact that SOEs operate with lower productivity than private firms (Brown et al., 2006; La Porta and Lopez-de-Silanes, 1999; Loc and Tran, 2016; Megginson and Netter, 2001) and that MNEs tend to be more productivity than domestic firms (Arnold and Javorcik, 2009; Javorcik, 2004; Newman et al., 2015). This set-up is consistent with the lower growth rates observed in SOE-pioneered industries, provided that all firms share the same prior beliefs about their productivity and that the direct effect of the pioneer on overall industry growth is sizable.

It is arguably a strong assumption that firms share similar prior beliefs about their productivity. Different degrees of uncertainty about α_k and δ_{ik} can alternatively translate into different entry sizes and growth rates of pioneers. When potential pioneers faces less uncertainty about the viability of industry k , and their own relative productivity in that industry, they are more likely to enter big and grow fast, all else equal. The employment dynamics observed in Figure 7 are consistent with the idea that MNEs with experience from the same industry in other countries have precise knowledge about their competitiveness in the industries they pioneer. This intuition is also line with Jara-Figueroa et al. (2018), who show that pioneer firms hiring workers with industry-specific knowledge are more likely to survive.

Alternatively, differences in the size and growth of pioneers may stem from the policy factor B_{ik} rather than uncertainty about productivity. A government may value the establishment of a sizeable new industry, and not only the act of pioneering itself. It may be, for instance, that a large SOE-pioneered industry is deemed important because it sustains a local supply of, or demand for, central products in the local production network. It is also likely that the Vietnamese state uses SOEs to maintain employment by establishing large local industries. A number of studies

have already documented that SOEs are used as tools to ensure employment and social stability in other contexts (Azmat et al., 2012; Bai et al., 2006; Berkowitz et al., 2017; Lin et al., 1998; Wen, 2022). This idea is consistent with the geographical distribution in Figure 4, documenting that SOEs tend to pioneer in remote areas populated by minority groups. The high initial employment in SOE-pioneered industries also support this hypotheses (see Figure 6).

Second, pioneers may impact industry dynamics *indirectly* through productivity spillovers, entry barriers, pricing behaviour, and their ability to ‘pick winner industries’ with an inherent growth potential. It is useful to think of spillovers, entry barriers, and pricing behaviour as constituting an *explicit* channel, because they are directly attributable to the characteristics and behaviour of the pioneer. In our framework, this channel is captured by the *dot*-subscripts on industry productivity ($\alpha_{.k}$), fixed costs ($\phi_{.k}$), and input- and output prices (P_{ik}, W_{ik}) via ($Z_{.k}^y$) and ($Z_{.k}^m$). The pioneer’s industry pick can alternatively be conceptualized as an *implicit* channel, because it influences entry and growth dynamics through the characteristics of the industry and is not directly attributable to the behaviour the pioneer. This channel works through every parameter with a subscript k in followers’ value function from Equation 6.

Followers may take potential productivity spillovers from the pioneer firm into account when deciding whether to switch industry. Productivity spillovers are incorporated through $\alpha_{.k}$ in our framework. Such spillovers have been studied extensively in the FDI literature and their existence has often been confirmed. The effects are, however, highly dependent on context and they typically appear through vertical linkages and not horizontally, which is the kind of spillover we have in mind in this paper (Aitken and Harrison, 1999; Görg and Greenway, 2004; Javorcik, 2004; Newman et al., 2015; Setzler and Tintelnot, 2021). Yet, the short monopoly-period in MNE-pioneered industries may indicate horizontal spillovers from pioneering multinationals in Vietnam. In contrast, productivity and technology spillovers from SOEs have received little attention in the literature.²⁵ On one hand, SOEs tend to be less productive and we should therefore not expect positive spillovers from them. On the other hand, SOEs may be less proprietary about their technology and managerial knowledge.

Alternatively, pioneers may negatively affect entry and growth of emulators through (unfair) competition or due to erection of barriers to entry. As discussed in detail above, our framework

²⁵See Beraja et al. (2021) for a notable exception.

captures these channels via the fixed cost parameter ϕ_k and through the pricing behaviour of incumbent firms, Z_k^y . We shall not repeat the discussion here on why these parameters might vary systematically with firm ownership. We simply note that long monopoly periods and low growth in SOE-pioneered industries are consistent with other studies documenting the existence of these crowding-out mechanisms in state-dominated sectors (Baccini et al., 2019; Brandt et al., 2020).

Finally, the pioneer type can affect industry dynamics indirectly through the characteristics of the industries in which they choose to pioneer. On one hand, firms may be uncertain about the industry-level productivity parameter, α_k , before they pioneer, as in models on cost-discovery. The patterns we observe in Vietnam are consistent with a story where MNEs and PDEs have a higher ability to pick winners with inherently higher potential for productivity growth. On the other hand, there are several reasons to believe that MNEs' and SOEs' choice of industry may be negatively correlated with entry of followers. Their chosen industries may, for instance, require scale economies that firms in the private sector do not have. MNE-pioneered industries will also have higher barriers to entry because MNEs pioneer in industries with complex production functions $F_k(\cdot)$, as we document in Figure 5. It is likely that domestic firms do not know these production functions and simultaneously cannot access finance to invest in learning them. Note that the high growth rates in MNE-pioneered industries are not sufficient to reject that these industry-specific entry barrier mechanisms are present in Vietnam. They may simply be countered by positive spillovers or strong direct pioneer effects.

In the remainder of this paper we formally test the main conclusions from the Section 4 and, to the extent possible, the mechanisms behind them as discussed here.

6. TESTING DIFFERENCES IN PIONEERING

Our conceptual framework highlights important unobserved characteristics that can lead to differences in pioneering behavior, even for enterprises with similar observable characteristics. In this section, we therefore seek to disentangle the influence of observable attributes from the effect of unobservable characteristics on the pioneering frequencies. We do so using the classical Oaxaca-Blinder decomposition of the difference in pioneering frequencies in which we condition on an

expanding set of observable attributes. Even though this is algebraically equal to regression based estimates of a causal difference in pioneering probabilities conditional on observables (Fortin et al., 2011; Kline, 2011), it should not be interpreted as a strictly causal estimate of ownership effects. The reason is that the hypothetical causal experiment asks what would happen to the probability of pioneering if an average SOE (or MNE) was changed to a PDE with the same observable attributes. Several SOEs were equitized in the period we investigate, so the hypothesized change in ownership has certainly happened. However, our estimates of the difference in pioneering probabilities do not take account of the selection of SOEs being equitized. Hence, the estimates presented below are meant to quantify the relative importance of differences in the observable attributes compared to the unobserved, such as differential access to financial capital, different know-how, differences in fixed costs, and possible effects of the policy factor for SOEs.

In the comparison of pioneering behavior across ownership types, we estimate linear probability models for each ownership type indexed by g . Hence, the outcome, y , is an indicator variable taking the value one for pioneering firms and zero for other firms. The pioneering is linearly related to (functions of) the covariates, x , and an error term, v .

$$y_{gi} = \beta_{g0} + \sum_{j=1}^J x_{gij} \beta_{gj} + v_{gi}, \quad (7)$$

The parameters of the model are estimated by OLS.

We make pairwise comparisons of the three ownership types. We first compare PDEs to SOEs, then PDEs to MNEs and finally SOEs to MNEs. For brevity, we use the standard group indicators 0 and 1 below. It is well known that the overall difference in the average outcome (the share of pioneering firm in each group) can be decomposed into two components (the Oaxaca-Blinder decomposition):

$$\bar{y}_1 - \bar{y}_0 = \underbrace{(\hat{\beta}_{10} - \hat{\beta}_{00}) + \sum_{j=1}^J \bar{x}_{1j} (\hat{\beta}_{1j} - \hat{\beta}_{0j})}_{\text{Ownership effect}} + \underbrace{\sum_{j=1}^J (\bar{x}_{1j} - \bar{x}_{0j}) \hat{\beta}_{0j}}_{\text{Composition effect}} \quad (8)$$

The first two terms, which we denote the ownership effect, are the average treatment effect of the treated (ATT) when using a regression based adjustment for bias from imbalance in the observables,

x_{gi} , over the two groups (see, e.g., Wooldridge (2010)). The second term is the bias from differences in the groups' characteristics that we denote the composition effect, following the decomposition literature (see, e.g., Fortin et al. (2011)).

Several of our observable characteristics are categorical (year of pioneering, industry, province). This gives rise to the well known identification problem in the decomposition that arises because the choice of the reference category affects the interpretation of the detailed decomposition in Equation 8.²⁶ Therefore, we only present the overall decompositions below. When computing overall decompositions for an expanding set of observable characteristics, we face the problem of path dependence in the inclusion of observables. We seek to solve this problem by structuring the order of inclusion of variables in accordance with the organizing framework in Section 5.

6.1. Decomposition Results

The decompositions are reported in Table 2. The top part of the table has the pioneering frequencies for each ownership group. The second part of the table presents decompositions of the difference between SOEs and PDEs. The difference in frequencies between the two groups is 0.547 percentage point (ppt), which is substantial in relative terms as the SOE frequency is more than three times larger than the PDE frequency in the sample of firms for which we have all information necessary for the decompositions.

In Column 1 we only include time dummies as controls to account for the difference in the shares of PDEs and SOEs over time. The time dummies do explain a significant and substantial part of the raw difference in pioneering frequencies. In Column 2, we add a large set of attributes that are related to firm operations and profits. Specifically, we include the revenue, the book value of capital equipment, the wage bill, total costs of intermediates, and the total employment – all in logs. Moreover, we include an indicator for multi-industry firms. By this set of regressors, we aim to control for composition effects related to firm size, profits, and operating costs, thereby isolating the effects of firm and sector specific differences in productivity (δ, α); pioneering fixed costs (ρ, ϕ), and the policy factor (B), which will all be part of the estimated ownership effect. Differences in access to capital will also be in the estimated ownership effect. The estimates in Column 2 show

²⁶See, e.g., Oaxaca and Ramson (1999) and Fortin et al. (2011).

Table 2: Decomposition of the Pioneering Frequencies

	(1)	(2)	(3)	(4)
	Share of pioneers in each group (percent)			
SOE	0.802 (0.036)	0.802 (0.036)	0.802 (0.036)	0.802 (0.036)
MNE	0.530 (0.024)	0.530 (0.024)	0.530 (0.024)	0.530 (0.024)
PDE	0.254 (0.003)	0.254 (0.003)	0.254 (0.003)	0.254 (0.003)
SOE vs. PDE				
Difference	0.547** (0.036)	0.547** (0.036)	0.547** (0.036)	0.547** (0.036)
Composition	0.185** (0.004)	0.294** (0.012)	0.408** (0.016)	0.429** (0.016)
Ownership	0.362** (0.037)	0.254** (0.038)	0.139** (0.039)	0.119** (0.039)
MNE vs. PDE				
Difference	0.276** (0.024)	0.276** (0.024)	0.276** (0.024)	0.276** (0.024)
Composition	0.009** (0.001)	-0.139** (0.007)	0.119** (0.012)	0.128** (0.013)
Ownership	0.266** (0.024)	0.415** (0.025)	0.156** (0.027)	0.147** (0.027)
MNE vs. SOE				
Difference	-0.272** (0.043)	-0.272** (0.043)	-0.272** (0.043)	-0.272** (0.043)
Composition	-0.223** (0.022)	-0.536** (0.048)	-0.247** (0.077)	-0.378** (0.089)
Ownership	-0.048 (0.041)	0.265** (0.049)	-0.025 (0.082)	0.107 (0.091)
Obs. SOE	60634	60634	60634	60634
Obs. MNE	92489	92489	92489	92489
Obs. PDE	2858644	2858644	2858644	2858644
Regressors				
Year	yes	yes	yes	yes
Controls	-	yes	yes	yes
Industry	-	-	yes	yes
Province	-	-	-	yes

Notes: Robust standard errors are reported in parentheses. Asterisks denote significance levels: ** $p < 0.01$, * $p < 0.05$. Control variables in Columns 2-4 are $\ln(\text{revenue})$, $\ln(\text{capital})$, $\ln(\text{wage})$, $\ln(\text{intermediaries})$, $\ln(\text{employment})$, and a dummy for multi product firms. Industry fixed effects are at the 2-digit VSIC level and two sectors are included for each firm-period observation: The current and the lagged. In addition, indicators for start-up firms are included.

that the composition effect accounts for about half of the difference in the pioneering frequencies, illustrating the importance of firm size, profit, and multi-product activities for pioneering.

Moving to Column 3, we seek to account for some of the unobserved ownership factors, notably the potential differences in productivity and fixed costs of pioneering ($\delta, \alpha, \rho, \phi$). In the regressions, we include indicators for the industry in which the firm operated prior to pioneering and the industry in which it established the new production. For staying firms, the two indicators are equal, for switchers they are like the pioneers, a prior and a post industry, while the prior industry indicator is zero for start-up pioneers, for which we add an individual indicator. To the extent that the two sets of industry indicators capture the productivity gains and fixed costs of pioneering for PDEs and SOEs, the ownership effect is reduced to an estimate of the differential access to capital and the policy factor. The industry controls increase the estimated composition effect to 0.41 ppt, leaving only 0.14 ppt as an estimate of the ownership effect. Thus, as expected, sector composition and knowledge strongly influence pioneering behavior.

In Column 4, we add province indicators to the set of controls. Our aim is to capture a part of the policy factor, as SOEs may be mandated, or incentivised, to pioneer in low productivity areas or areas in which the fixed costs are high. Such policy factors are not captured by the industry indicators. The province indicators increase the composition effect to 0.43 ppt, giving us a final estimate of a pure SOE ownership effect of 0.12 ppt. This is an estimate of the effect of preferential access to capital and residual (non spatial) policy factors. Thus, by this decomposition we end up with an SOE-specific effect that is only 14 percent of the SOE pioneering frequency. It is, however, clearly statistically significant.

In the third panel of Table 2, we compare MNEs and PDEs. The raw difference between the pioneering frequencies is smaller (0.28 ppt) than the SOE/PDE difference. This is mainly due to smaller differences in the composition effect while the ownership effect is actually larger for MNEs compared to SOE. In particular, when we condition on the firm information (size, revenue, capital, and costs) in Column 2, we find a larger ownership effect than the raw difference. The large effect is reduced when we condition on industry composition and slightly more when we include the province composition. This highlights the endogenous placement of MNEs in industries and locations that are favourable to pioneering. With all controls included, about half of the difference in the pioneering frequencies between MNEs and PDEs can be attributed to elements related to

Table 3: Decomposition of the Pioneering Frequencies for Alternative Measures of Pioneering

	(1) 4 year gap	(2) 5 year gap	(3) First time
Share of pioneers in each group (percent)			
SOE	0.569 (0.031)	0.475 (0.028)	0.665 (0.033)
MNE	0.453 (0.022)	0.397 (0.021)	0.397 (0.021)
PDE	0.204 (0.003)	0.174 (0.002)	0.197 (0.003)
SOE vs. PDE			
Difference	0.365** (0.031)	0.301** (0.028)	0.468** (0.033)
Composition	0.298** (0.014)	0.235** (0.013)	0.368** (0.014)
Ownership	0.068* (0.034)	0.066* (0.031)	0.099** (0.036)
MNE vs. PDE			
Difference	0.249** (0.022)	0.223** (0.021)	0.200** (0.021)
Composition	0.120** (0.012)	0.088** (0.010)	0.096** (0.011)
Ownership	0.129** (0.025)	0.135** (0.023)	0.104** (0.023)
MNE vs. SOE			
Difference	-0.116** (0.038)	-0.078* (0.035)	-0.268** (0.039)
Composition	-0.158* (0.074)	-0.159* (0.065)	-0.405** (0.081)
Ownership	0.042 (0.079)	0.080 (0.069)	0.137 (0.081)
Obs. SOE	60634	60634	60634
Obs. MNE	92489	92489	92489
Obs. PDE	2858644	2858644	2858644

Notes: Robust standard errors are reported in parentheses. Asterisks denote significance levels: ** $p < 0.01$, * $p < 0.05$. All regressions include $\ln(\text{revenue})$, $\ln(\text{capital})$, $\ln(\text{wage})$, $\ln(\text{intermediaries})$, $\ln(\text{employment})$, year fixed effects industry fixed effects and a dummy for multi product firms, and province fixed effects. Industry fixed effects are at the 2-digit VSIC level and two sectors are included for each firm-period observation: The current and the lagged. In addition, indicators for start-up firms are included.

better access to capital, better information about productivity gains, or specific MNE strategies.

The fourth panel of Table 2 presents the comparison of MNEs and SOEs. The main result from this comparison is that once we condition on firm attributes and sector composition, the ownership effect is insignificant for MNEs and SOEs. The comparison of MNEs and SOEs is all the more interesting in conjunction with the information about the geographical placement of MNEs and SOEs shown in Figure 4. Whereas MNEs are located and pioneer in close proximity to the big economic centers, SOEs are much more frequent in the remote areas with lower levels of economic activity.

We test the robustness of the results using alternative ways of defining pioneering. For all three pairwise comparisons (SOEs vs. PDEs, MNEs vs. PDEs, MNEs vs. SOEs), we estimate the raw difference using the full set of controls and fixed effects. The robustness results are given in Table 3. First, we change the three-year period in which an industry must not have been observed to four and five years, respectively, to make sure that our main results are not driven by the choice of three years (Columns 1 and 2 in Table 3). Second, we restrict the definition of pioneering so that an industry must have never been observed before within a province (Column 3). The results are remarkably robust to the different ways of defining a pioneer. Of course, the shares of pioneering firms decrease as we increase the required gap, but we continue to find statistically significant ownership effects for both SOEs and MNEs compared to PDEs. The magnitudes change over the different indicators, but the overall results remain the same.

7. TESTING DIFFERENCES IN INDUSTRY DYNAMICS

In this section, we formally test the industry dynamics documented in Figures 6 and 7. We test which pioneers set up industries that attract followers and grow using a difference-in-differences set-up with staggered adoption. As outcome variables, we focus on total employment and number of firms in pioneered industries. We use the aggregated data set described in Section 3.2 to conduct the analysis. It contains aggregated information for all local industries that were pioneered between 2004 and 2017.

Recent literature on difference-in-differences and event studies has highlighted a number of

issues with the OLS two-way fixed effects (OLS TWFE) estimator in staggered adoption designs with heterogeneous and dynamic treatment effects, including the possibility of negative weights, under-identification, and bias when conditioning on a pre-treatment test passing (Borusyak et al., 2021; Callaway and Sant’Anna, 2021; de Chaisemartin and D’Haultfœuille, 2020; Goodman-Bacon, 2021; Roth, 2022; Sun and Abraham, 2021). To overcome these issues, we use the imputation estimator by Borusyak et al. (2021), which allows for estimation of treatment effect heterogeneity. We both implement the estimator in its standard form, and we apply it in conjunction with a simple matching estimator.²⁷ We describe the estimation approach below.

Let Ψ^{PDE} be the subsample of industry-province pairs that are pioneered by PDEs. Further, in any given year, let Ψ_0^{PDE} be the industry-province pairs within this subsample that are not yet pioneered, while Ψ_1^{PDE} has the pairs for which pioneering has occurred. Define also $\Psi_0^{SOE}, \Psi_1^{SOE}, \Psi_0^{MNE}, \Psi_1^{MNE}$ to be the corresponding subsamples for SOEs and MNEs, respectively, and let $N(\cdot)$ denote the number of observations in each subsample. Further, let $Y(0)_{kpt}$ denote the outcome for all not-yet-pioneered province-industry pairs, and let β_{kpt} be the pioneer effect in industry k , province p , and year t .

The imputation-based estimates of the differential dynamics in PDE and SOE pioneered industries are obtained in three steps. First, we estimate three-way fixed effects from the OLS regressions $Y(0)_{kpt} = \lambda_{kt} + \mu_{pt} + \gamma_{kp} + e_{kpt}$ for all not-yet-treated observations $kpt \in \{\Psi_0^{PDE} \cup \Psi_0^{SOE}\}$. λ_{kt}, μ_{pt} , and γ_{kp} denote industry-year, province-year, and industry-province fixed effects, respectively. The three-way specification is our preferred model, but we also present results from less restrictive specifications. Second, we impute the counterfactual $\hat{Y}(0)_{kpt} = \hat{\lambda}_{kt} + \hat{\mu}_{pt} + \hat{\gamma}_{kp}$ for all treated observations $kpt \in \{\Psi_1^{PDE} \cup \Psi_1^{SOE}\}$. We then estimate the industry-province-specific effect of pioneering as:²⁸

$$\hat{\beta}_{kpt} = Y_{kpt} - \hat{Y}(0)_{kpt} \quad (9)$$

Third, the estimates of interest are obtained as weighted averages of the industry-province-specific effects. Specifically, let β^X be our parameter of interest in the static regression, where X denotes

²⁷For comparison, we report OLS estimates in the appendix.

²⁸Note that the estimate of β_{kpt} changes slightly depending on exact specification we use, because the sample used to impute counterfactuals changes in different specifications. For instance, when we estimate the effect of pioneering by PDEs, our sample only includes PDEs-pioneered industries. Since the counterfactuals always remain close to zero, this does not impact our main conclusions.

different average effects of interest. The estimator for the effect of PDE pioneering is given by:

$$\hat{\beta}^{PDE} = \frac{1}{N(\Psi_1^{PDE})} \sum_{\Psi_1^{PDE}} \hat{\beta}_{kpt}. \quad (10)$$

Thus, $\hat{\beta}^{PDE}$ is the average of the outcomes of all industry-province-year observations for which the pioneering was by a PDE firm. The average is an estimate of the ATT assuming there are no spillover effects between provinces.

An estimator of the differential effect of pioneering by SOEs and PDEs is:

$$\hat{\beta}^{SOE} = \frac{1}{N(\Psi_1^{SOE})} \sum_{\Psi_1^{SOE}} \hat{\beta}_{kpt} - \frac{1}{N(\Psi_1^{PDE})} \sum_{\Psi_1^{PDE}} \hat{\beta}_{kpt}. \quad (11)$$

To obtain estimates of the coefficients of interest in a dynamic specification, $\hat{\beta}_h^X$, we follow a similar procedure, except that we only average over the subset of observations in event time h . Estimates of the different dynamics in MNE vis-a-vis PDE pioneered industries, $\hat{\beta}^{MNE}$ and $\hat{\beta}_h^{MNE}$, are obtained in three analogous steps. Standard errors are calculated as in Borusyak et al. (2021) and clustered at the province-industry level.²⁹ Appendix A.3.3 reports the corresponding OLS estimation for comparison.

Because the fixed effects used to impute counterfactuals are estimated based on non-treated observations only, they do not control for treatment effect heterogeneity. This is important, because there may be inherent differences in the growth potential of different industries and/or provinces, and these differences may be correlated with the pioneer type, as discussed in the conceptual framework. Because the imputation-based estimator computes individual-level treatment effects, it is straightforward to account for the differences in pioneer-effects attributable to pioneering in provinces/industries with different growth potentials. To do so, we compute a simple re-weighting estimator. The idea is to estimate industry-specific differences in ATTs, and then average these ATTs over all industries.

We calculate the re-weighted-imputation estimates of the differential effects of pioneering by SOE vs. PDE as follows. Let k denote the industry and K be the total number of industries where

²⁹In robustness checks, we show that the results are not sensitive to clustering at either industry or province-level.

SOEs and PDEs are simultaneously present. The estimator is:

$$\hat{\beta}_m^{SOE} = \frac{1}{K} \sum_{k=1}^K \left[\frac{1}{N(\Psi_1^{SOE} \cap k)} \sum_{(\Psi_1^{SOE} \cap k)} \hat{\beta}_{kpt} - \frac{1}{N(\Psi_1^{PDE} \cap k)} \sum_{(\Psi_1^{PDE} \cap k)} \hat{\beta}_{kpt} \right], \quad (12)$$

We also re-weight by matching on provinces (p) and finally on both industries and provinces. We estimate β_m^{MNE} analogously.³⁰

One could imagine that PDEs tend to enter industries that have existed before. One could also imagine that SOEs pioneer to re-start dying industries for social-policy purposes. To investigate whether the industry-dynamics *prior* to the pioneering event differ among pioneer types, we implement two tests for pre-trends. The pre-trend tests are conducted on the pre-treatment sample. We test for trends from event time -7 to -1, using all prior time periods as controls to increase power.³¹ The pre-trend test proposed in Borusyak et al. (2021) is an F-test for joint significance of some pre-treatment coefficients estimated by OLS. The test appropriately uses untreated observations only, makes a conceptual distinction between estimation and testing (different estimators are used), and solves the pre-test problem highlighted by Roth (2022) under homoskedasticity (Borusyak et al., 2021). The test for pre-trends in the PDE-specification is the standard test. When testing the differential impact of SOE and MNE pioneering, the reported p-values are from an F-test of joint significance of the interaction terms between pre-trend dummies and dummies indicating whether the industry was pioneered by an SOE or MNE. The estimation follows the strategy described in Appendix A.3.3. We also conduct a placebo pre-trend test explicitly using the imputation estimator from Borusyak et al. (2021). Here, we push the pioneering event seven years into the past and test the joint significance of the coefficients on the placebo treatment dummies between -7 and -1. The test is conducted using a chi-square test.

7.1. Static Results

We organize the analyses based on the conceptual framework from Section 5. In Table 4, we separate the *direct effect* of the pioneers on industry growth from the *indirect effect* of pioneering on

³⁰When matching on industries and provinces, for instance, we only include the industry-province pairs in the estimation sample where we observe both SOEs and MNEs.

³¹When controlling for the Ashenfelter dip, we only test for pre-trends from event time -7 to -3.

entry and growth of emulators. In Tables 5 and A18, we subsequently analyse whether these effects are driven by composition effects in the sense that firms with different ownership are pioneering in different industries and provinces.

Table 4 shows results from static specifications with employment as the outcome variable in the top panel and firm count as the outcome in the bottom panel. We show two sets of results to distinguish the direct and indirect effects of pioneers on industry growth. In Columns 1-4, we include the pioneers themselves and their employment in the dependent variables, while the pioneer contributions are excluded in Columns 5-8.

We include two estimates of β^{PDE} to account for the "Ashenfelter dip" that we observe in Figure 7. Take the case of employment as an example. Our definition of pioneering ensures that the local industry's employment rate is zero in the three years prior to a pioneering event, but in some cases we observe employment prior to event time -3. Ashenfelter's dip then leaves us with an identification problem when estimating β^{PDE} . In our standard specification, part of the effect of PDE-pioneering on employment reported in Columns 1 and 5 will be caused by mean-reversal, and we reject the null of no pre-trends. We address this problem in Columns 2 and 6. In these specifications, we exclude event times three years prior to the pioneering event and thereby ensure that these observations do not generate a downward bias in the imputed counterfactuals. When excluding event times -3 to -1, we cannot reject the null at a 5 percent significance level. In this way, the two estimates of β^{PDE} can be interpreted as an upper and lower bound on the effect of pioneering by PDEs. Note that we do not exclude any pre-event periods when estimating β^{SOE} and β^{MNE} because the dip is netted out by difference operation in Equation 11.

Columns 1 and 2 in the top panel of Table 4 show that new industries started by PDEs employ approximately 97-121 additional workers post-pioneering, depending on the specification. Because there may be employment in the local industries from three years prior to the pioneering event, the coefficient estimate is not a measure of the average workforce post-pioneering. Rather, it measures the *difference* from the pre-pioneering periods. In our sample, PDEs have pioneered in 5,045 local industries. A simple back-of-the-envelope calculation suggests that they have added somewhere between 490 and 610 thousand jobs in infant industries during our sample period. Column 3 shows that industries pioneered by SOEs employ, on average, 40 additional workers compared to PDE-pioneered industries, but the difference is not statistically significant at the 5

Table 4: Total and Indirect Effects of Pioneering on Employment and Firm Counts

	Total Effects (Incl. Pioneer)				Indirect Effects (Excl. Pioneer)			
	(1) PDEs only ($\hat{\beta}^{PDE}$)	(2) PDEs only ($\hat{\beta}^{PDE}$)	(3) SOEs vs. PDEs ($\hat{\beta}^{SOE}$)	(4) MNEs vs. PDEs ($\hat{\beta}^{MNE}$)	(5) PDEs only ($\hat{\beta}^{PDE}$)	(6) PDEs only ($\hat{\beta}^{PDE}$)	(7) SOEs vs. PDEs ($\hat{\beta}^{SOE}$)	(8) MNEs vs. PDEs ($\hat{\beta}^{MNE}$)
Employment								
ATT	121.182*** (15.026)	97.196*** (10.576)	40.224* (22.656)	653.650*** (136.451)	99.037*** (14.755)	73.365*** (10.095)	-24.112 (18.908)	248.112*** (73.717)
Pre-trend (p-value)	0	0.201	0.373	0.649	0	0.226	0.266	0.738
Placebo test (p-value)	0.458	0.331	0.668	0.872	0.511	0.369	0.642	0.918
Firm count								
ATT	3.371*** (0.171)	2.908*** (0.188)	-0.092 (0.434)	0.083 (0.491)	2.610*** (0.143)	2.006*** (0.148)	-0.115 (0.398)	-0.010 (0.473)
Pre-trend (p-value)	0	0	0.616	0.832	0	0	0.677	0.789
Placebo test (p-value)	0	0	0.552	0.841	0	0	0.398	0.835
Event times used for test	-7 to -1	-7 to -4	-7 to -1	-7 to -1	-7 to -1	-7 to -4	-7 to -1	-7 to -1
Observations	79,713	49,942	87,295	86,634	79,713	49,942	87,295	86,634
Treated obs.	37,361	22,725	41,944	40,570	37,361	22,725	41,944	40,570
Control obs.	42,352	27,217	45,351	46,064	42,352	27,217	45,351	46,064

Notes: The dependent variable is province-industry employment (top panel) and province-industry firm count (bottom panel). In Columns 1-4, the dependent variable includes employment/firm of the pioneers, while the pioneer effects are excluded in Columns 5-8. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Standard errors in parentheses are clustered at province-industry level. Results in Columns 1-2 and 5-6 are the estimator presented in Equation 10. Results in Columns 3-4 and 7-8 are from the estimator presented in Equation 11. Counterfactuals are imputed using province-year, industry-year, and province-industry fixed effects, where industries are defined using 2-digit VSIC codes. *** p<0.01, ** p<0.05, * p<0.1.

percent level of significance. The effect of SOE pioneering is dwarfed by the employment boom in MNE-pioneered industries. After entrance by an MNE, these industries employ around 654 additional workers – a premium of 640-770 percent, depending on the PDE benchmark estimate.³² The large standard error on the estimated ATT indicates a substantial heterogeneity behind the estimate. P-values from pre-treatment tests are reported below each estimate. We cannot reject the null, except when testing the effect of PDE-pioneering without accounting for the Ashenfelter dip.

The bottom panel of Table 4 documents that employment in new industries tends to be located in very few firms. PDE-pioneering adds only around three firms to a local industry, on average, and pioneering by SOEs and MNEs is indistinguishable from that of PDEs. The differential employment effect reported above is therefore driven by differences in firm size. The estimates should, however, be interpreted with caution since we find evidence of pre-trends in firm counts.

In Columns 5-8 in the top panel, we exclude employment of the pioneers themselves from the dependent variable. The columns give estimates of the indirect effect of pioneering by firms of

³²We calculate the relative premiums as follows: $\frac{653.7+121.2}{121.2} \times 100 = 639.4$ and $\frac{653.7+97.2}{97.2} \times 100 = 772.5$

different ownership. We find substantial differences between the estimates with and without pioneer employment. The employment-bonus in MNE-pioneered industries is reduced by 60 percent. The majority of employment in infant industries created by MNEs is, in other words, created directly by the pioneering MNE. That being said, average employment in follower firms remains more than 240 heads higher than in PDE-pioneered industries. The effect size of PDE-pioneering in Column 6 is reduced by less than one-third, so a larger portion of employment growth in PDE-pioneered industries comes from entry and growth of followers. Finally, the point estimate on the relative effect of SOE ownership turns negative but remains insignificant.

In Tables A9 and A10 in the appendix, we show that the results above are robust to using different fixed effects to impute counterfactuals. The point estimates are slightly higher in the preferred three-way fixed effects specification, although the difference is not statistically significant. This is because employment and firm counts are close to zero in most industry-province-years prior to pioneering. The counterfactuals are therefore also close to zero regardless of the set of fixed effects we employ. The minor changes in effect sizes indicate that the industries pioneered by firms with different ownership are on similar growth paths prior to pioneering. In Table A11, we also show that the statistical significance of the main estimates are unchanged when we cluster standard errors at the province or industry level instead of clustering at the province-industry level.

The results presented thus far include province-industry pairs that have experienced two or even three pioneering events during our sample period. These are provinces, where an industry has appeared, disappeared, and then reappeared again. So far, we have only considered the first pioneering event, ignoring the second and third. This will bias our estimates if some firm types are more likely to pioneer industries that fail, but that are subsequently pioneered again successfully by other firm types. We test for the presence of such bias in Table A12 by excluding all province-industry pairs that have experienced multiple pioneering incidents. Our main results remain unchanged.

Another potential bias arise from the way we have assigned industries to pioneers. In 133 instances, a local industry is simultaneously pioneered by firms with different ownership structures. As described in Section 3.2, these hybrid industries pose a classification problem. So far, we have classified all hybrid industries to be SOE-pioneered industries, if we observe an SOE pioneer, and we have assigned industries to be MNE-pioneered they are simultaneously pioneered by a PDE

and a MNE. This decision is, of course, somewhat arbitrary and may cause an upwards bias in our estimates of employment premiums in SOE- and MNE-pioneered industries. Reassuringly, the results are robust to excluding all hybrid industries, as we do in Table A13.

Finally, we also provide corresponding three-way FE OLS results in Tables A19 and A20 in the appendix for comparison. The OLS estimates are slightly different (as expected), but generally align with the findings presented here.

Which underlying factors are driving the ownership differences that we observe in Table 4? From the conceptual framework, differences in the *direct effects* may come from differences in pioneers' tendency to enter different industries. The size of pioneering MNEs and SOEs may, for instance, be larger if they are more likely to enter less competitive industries dominated by economies of scale and high entry barriers. Differences in the entry size and growth rates of pioneers may also arise due to differences in a-priori uncertainty about firm- and industry-specific productivity (δ_{ik}, α_k) or because of policy priorities (B_{ik}). On the other hand, differences in *indirect effects* occur because pioneers influence followers differently via spillovers, entry barriers, and pricing behaviour (explicit channel) or via their ability to pick viable industries and provinces (implicit channel). The first channel enters the value function of followers (V_{ik}^F) through *dot*-subscripts on industry productivity ($\alpha_{.k}$), entry costs ($\phi_{.k}$), and input- and output prices (P_{ik}, W_{ik}) via ($Z_{.k}^y, Z_{.k}^m$). The latter channel enters the value function through every parameter with a subscript k .

We cannot estimate the contribution of each individual parameter to the ownership differences presented in Table 4. We can, however, use the re-weighting estimator from Equation 12 to assess the relative importance of groups of parameters. For instance, when we average over industry specific effects, we remove the composition effect related to unbalanced pioneering in different industries. Thereby we control for differences in industry fixed costs. We can thereby attribute the remaining difference in direct effects to within-industry cost and know-how, policy, and access to capital. When we match on industry, we also net out intrinsic industry characteristics (implicit channel) and the remaining difference in indirect effects can be attributed to differences in spillovers, entry barriers, and pricing behaviour (explicit channel).

Table 5 reports estimates of β_m^{SOE} and β_m^{MNE} with employment as the dependent variable. We leave estimates with firm count as the dependent variable to the appendix (Table A18). The first two columns of Table 5 provide estimates of direct effects, while the last two columns report estimates

Table 5: Pioneering Effects on Employment – Accounting for Composition Effects

	Direct Effects (Only Pioneer)		Indirect Effects (Excl. Pioneer)	
	(1) SOEs vs. PDEs ($\hat{\beta}_m^{SOE}$)	(2) MNEs vs. PDEs ($\hat{\beta}_m^{MNE}$)	(3) SOEs vs. PDEs ($\hat{\beta}_m^{SOE}$)	(4) MNEs vs. PDEs ($\hat{\beta}_m^{MNE}$)
No Matching				
ATT	64.070*** (10.651)	403.098*** (99.606)	-24.479 (18.183)	233.063*** (73.856)
Pre-trend (p-value)	.231	.102	.32	.265
Placebo test (p-value)	.25	.06	.46	.301
Observations	88,320	87,520	88,320	87,520
Treated obs.	42969	41456	42969	41456
Control obs.	45351	46064	45351	46064
Matching on Industry				
ATT	61.582*** (9.971)	218.880*** (55.322)	-57.229 (42.602)	6.445 (77.215)
Pre-trend (p-value)	.222	.094	.351	.298
Placebo test (p-value)	.514	.196	.416	.725
Observations	86,352	81,936	86,352	81,936
Treated obs.	42041	38684	42041	38684
Control obs.	44311	43252	44311	43252
Matching on Province				
ATT	67.947*** (11.965)	659.522*** (176.724)	-1.807 (22.502)	237.919 (156.700)
Pre-trend (p-value)	.222	.101	.325	.256
Placebo test (p-value)	.33	.164	.341	.846
Observations	87,056	74,352	87,056	74,352
Treated obs.	42387	35621	42387	35621
Control obs.	44669	38731	44669	38731
Matching on Industry and Province				
ATT	66.984*** (13.150)	418.667*** (156.209)	34.825** (13.882)	71.865 (137.591)
Pre-trend (p-value)	.168	.857	.577	.83
Placebo test (p-value)	.345	.859	.246	.729
Observations	21,536	16,176	21,536	16,176
Treated obs.	10783	7698	10783	7698
Control obs.	10753	8478	10753	8478
Event times used for test	-7 to -1	-7 to -1	-7 to -1	-7 to -1

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level. The dependent variable is province-industry employment. Results are ATTs calculated using the matching-based version of the imputation-based estimator (Borusyak et al., 2021) from Equation 12. Counterfactuals are imputed using year, province, and industry fixed effects, where industries are defined using 2-digit VSIC codes. The last year of the sample (2017) is excluded in all columns since proper counterfactuals cannot be calculated when all province-industries are treated.

of indirect effects. The sum of the estimates are comparable to the total effects reported to the left in Table 4. All estimates are calculated using year, industry, and province fixed effects to impute counterfactuals.³³ Moving from top to bottom, each panel in the table presents estimates using

³³We do not use the full three-way fixed effect specification, because of computational difficulties in estima-

increasingly granular groups to match SOEs and MNEs with PDEs. The top panel presents estimates with no re-weighting.³⁴ In the second panel, we match on 2-digit industry codes. We match on province in the third panel to account for the fact that SOEs and MNEs pioneer in different parts of Vietnam (see Figure 4). In the final panel, we match on both industry and province.³⁵

In Table 5, the direct effect of SOE-pioneering relative to PDE-pioneering on employment remains positive, significant, and stable in magnitude across all specifications (Column 1). This indicates that access to financial resources, a-priori knowledge about productivity parameters or (non-spatial) policy priorities explain most of the difference in employment between pioneering SOEs and MNEs. It is somewhat surprising that the size of the effect does not vary with industry and province matching. If SOEs pioneer in industries that require economies of scale or are mandated to generate *additional* employment in rural areas, we would expect the estimates to respectively decrease and increase when matching on industry and province.³⁶

We also note that a large portion of the direct MNE employment premium is related to industry and province characteristics. Column 2 shows that the employment premium in pioneering MNEs drops from around 403 to 219 workers when comparing pioneering MNEs and PDEs within the same industry. Part of the premium is thus related to MNEs' ability to select viable industries or to guess their own competitiveness within those industries before pioneering. The dramatic increase in the average direct employment when matching on province illustrates how the MNEs focus on industries requiring large plant sizes in the provinces in which they choose to pioneer. Thus access to capital appears to be a decisive factor.

Finally, the indirect employment effects in SOEs and MNEs seem to run through different

tions with few observations. As shown in Tables A9 and A10, and as discussed above, the specification of fixed effects does not have a significant influence on our results.

³⁴These are comparable to the estimates presented in Table 4, although with a slightly different set of fixed effects and where the total effects reported in Table 4 are the sum of the direct and indirect effects from Table 5.

³⁵Note that we do not match on year to account for time trends. This is because different ownership types pioneer at different times. Given the dynamic evolution of post-pioneering employment (documented in Section 7.2 below), year-matching would lead us to compare industry-groups at different event times, thereby biasing the effect downwards for the groups that tend to pioneer at later stages in the sample period. We could match by event time, but this is equivalent to a dynamic specification, and we report such results in Section 7.2.

³⁶In results not reported here, we do see a decline in the estimate from 64 to 55 workers when we match on 4-digit industries.

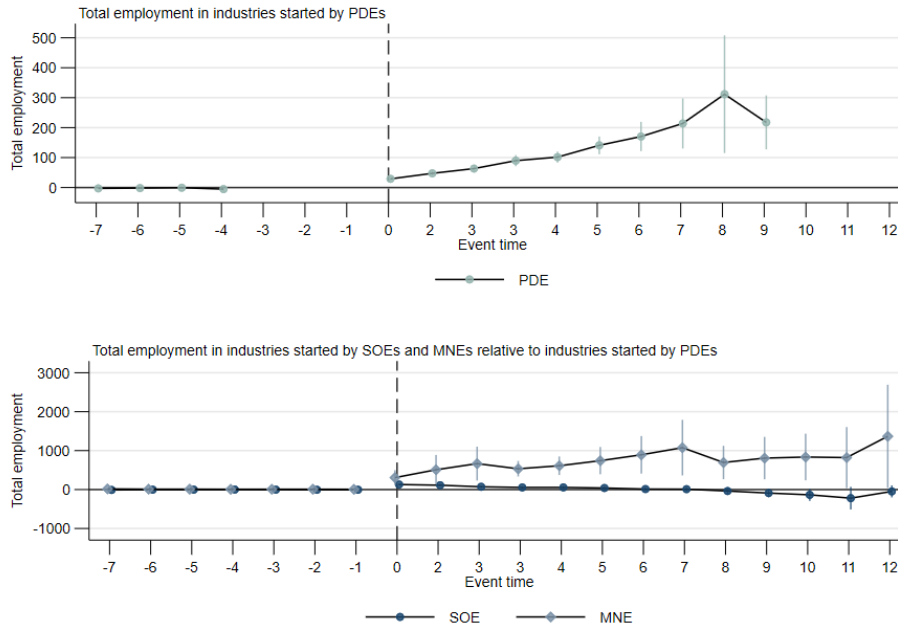
channels. When we control for both province and industry composition effects, the average indirect employment effect turns significant and positive for SOEs (Column 3). This is noteworthy because the estimate captures the combined difference in spillovers, entry barriers, and pricing behaviour between SOEs and PDEs. Previous studies would lead one to expect a negative coefficient due to the political and regulatory barriers to entry in state-dominated sectors (Baccini et al., 2019; Brandt et al., 2020). In contrast, our estimates indicate that SOEs may even crowd-in followers to a larger extent than PDEs when the comparison is restricted to industries and provinces in which both firm types have pioneered. The difference between the estimate with and without matching (35 vs. -24 workers) can thus be attributed to SOEs pioneering in unattractive industries and/or provinces in which PDEs are unlikely to pioneer. Comparing the estimates in panel two and three, the results could indicate that pioneer investment in unproductive provinces is driving the difference.

To sum up, while the direct employment premium in pioneering SOEs does not vary by geography, it does seem that the low indirect effect of SOEs pioneering is a composition effect as SOEs generally tend to pioneer in less attractive areas than PDEs. In contrast, Column 4 shows that MNE-pioneered industries attract followers primarily through MNEs' selective focus on profitable industries and provinces. Once we match on industries, we cannot reject the null of no difference in spillovers, entry barriers, and pricing behaviour between MNEs and PDEs.

7.2. Dynamic Results

We now turn to unpack the dynamic evolution in employment that generates the average static results from Table 4. Corresponding results for firm counts are presented in Appendix A.3.2 and OLS results are given in Appendix A.3.3. In Figure 8, we report dynamic estimates of the total employment effects, while Figure 9 gives estimates of the indirect effects. All estimations are obtained using three-way fixed effects to impute counterfactuals. We report the estimates of employment in PDE-pioneered industries in the top panel, dropping event times -3 to -1 to account for the Ashenfelter dip. Because we leave out three event times, there exist no counterfactuals for industries pioneered later than 2013 so we cannot estimate effects more than nine years from the pioneering event. Estimates of the differential effect in industries started by MNEs and SOEs are presented in the bottom panel of the figures. The plotted pre-trends are from the placebo test using

Figure 8: Total Effects on Employment (Including Pioneers)

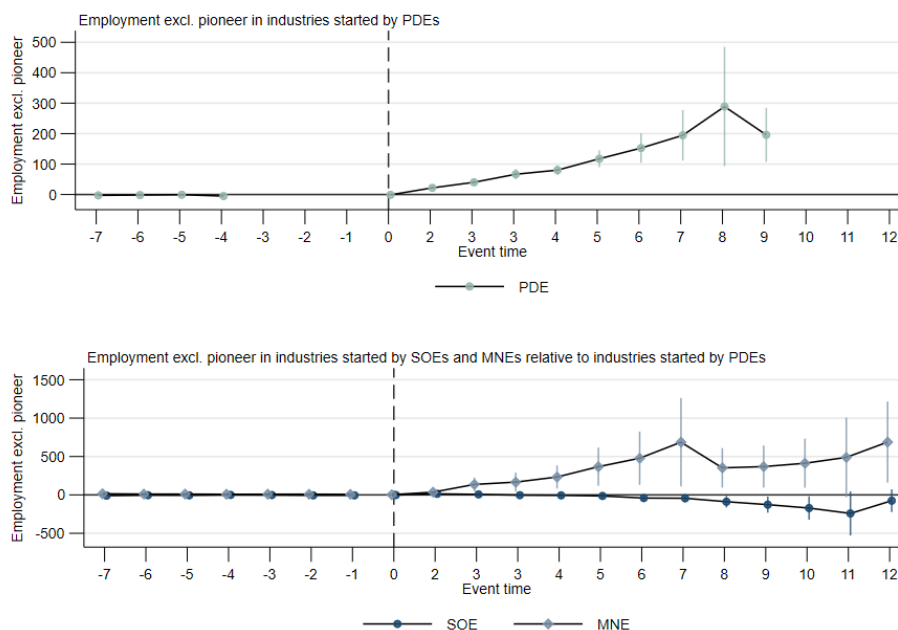


Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. The dependent variable is province-industry employment including employment in the pioneer firms. Results are event time-specific ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Results in the top and bottom panels are from dynamic versions of Equation 10 and Equation 11, respectively. Counterfactuals are imputed using province-year, industry-year, and province-industry fixed effects, where industries are defined using 2-digit VSIC codes.

the Borusyak et al. (2021) estimator.

Figure 8 presents evidence that average employment growth in PDE-pioneered industries is high and continues several years after pioneering. From the bottom-panel, we note that SOE-pioneered industries start large, but are eventually overtaken by PDE-pioneered industries. Industries started by MNEs also tend to grow fast in the beginning before stabilizing at an average employment premium around 1,000 workers.³⁷ These dynamic employment effects mean that the static estimates in Table 4 understate the absolute differences in long-run industry employment. The average employment premium in MNE-pioneered industries is, for instance, estimated to be only 654 workers in the

³⁷One may be concerned that the results reported in Figure 8 may be driven by an unbalanced sample around the pioneering event, because the effects further away from the event are calculated using only a few observations. In Tables A14-A17 we report dynamic total effects of pioneering on employment and firm counts when balancing the sample around the pioneering event. This ensures that an equal number of observations are used to estimate event time-specific ATTs. For robustness, we balance on different windows around the event time. In these estimations, the employment effect of SOE-pioneering is slightly lower and the employment premium of MNE-pioneering is slightly higher, but the exercise does not change our main conclusions.

Figure 9: Indirect Effects on Employment (Excluding Pioneers)

Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. The dependent variable is province-industry employment excluding employment in the pioneer firms. Results are event time-specific ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Results in the top and bottom panels are from dynamic versions of Equation 10 and Equation 11, respectively. Counterfactuals are imputed using province-year, industry-year, and province-industry fixed effects, where industries are defined using 2-digit VSIC codes.

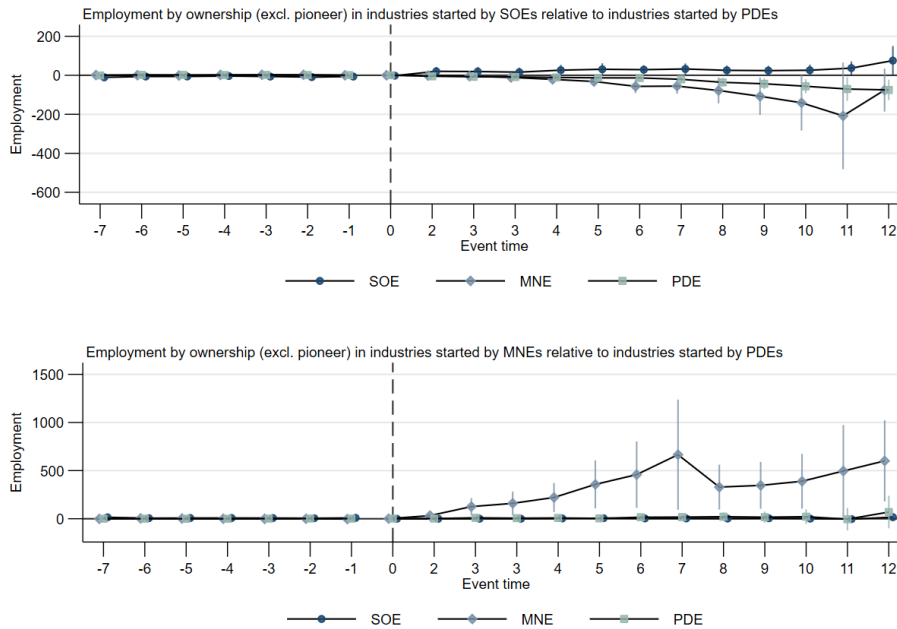
static specification. In relative terms, however, the difference is somewhat lower. The static estimate suggested a relative MNE premium of 770 percent (accounting for Ashenfelter's dip), while we estimate the premium in event times eight and nine to be around 320 and 470 percent, respectively.³⁸

Figure 9 confirm the results from Table 4 showing that a large fraction of the employment-growth in SOE- and MNE-pioneered industries is attributable to the size of the pioneers and not the incumbents. The point estimates of the indirect employment effect in MNE-pioneered industries is roughly cut in half and the SOEs coefficients are negative and even borderline significant in the long run.

Knowing whether foreign multinationals and SOEs crowd-in or crowd-out private domestic entrepreneurship has been a continued discussion, and literature on SOEs shows evidence of the crowding-out hypotheses (Baccini et al., 2019; Brandt et al., 2020), while the FDI literature is

³⁸The exact point estimates for PDEs are 311.7 and 217.5, respectively, eight and nine years after the pioneering event. The corresponding relatively estimates for MNEs are 694.9 and 807.5. We calculate the premium for, say, event time eight as $\frac{311.7+694.9}{311.7} \times 100 = 322.9$.

Figure 10: Decomposition of Indirect Effects on Employment by Ownership of Followers



Notes: Authors' calculations based on data from the Vietnam Enterprise Survey. The dependent variables are province-industry employment in following firms, decomposed by ownership type. The top panels shows results from SOE-pioneered industries, and the bottom panel from industries pioneered by MNEs. Results are event time-specific ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021) and detailed in Equation 11. Counterfactuals are imputed using province-year, industry-year, and province-industry fixed effects, where industries are defined using 2-digit VSIC codes.

less conclusive (De Backer and Sleuwaegen, 2003; Kosová, 2010; Lu et al., 2017). In Figure 10, we decompose the indirect employment effect into employment in follower firms with different ownership. In the top panel, we find evidence that SOE-pioneered industries generate significantly less employment in domestic firms and multinationals compared to PDE-pioneered industries. The negative effect is, surprisingly, largest for the multinational sector. In the bottom panel, we show that the indirect employment premium in MNE-pioneered industries is driven exclusively by crowding-in of other multinationals. There is no additional crowding-in of PDEs. To be clear, this does not mean that the employment spillovers to the domestic sector are completely similar in PDE- and MNE-pioneered industries. While the crowding-in of emulating domestic firms may be of the same magnitudes, the industries pioneered by MNEs are more complex than the ones established by PDEs (see Table 5).

7.3. Aggregate Effects

So far, we have focused on the relative differences in pioneering effects between PDEs, SOEs, and MNEs. The results indicate that both SOEs and MNEs are more likely to pioneer and that the total employment effect is higher when they do so. Yet, we also showed in Figure 3 that the domestic private sector accounts for the vast majority of pioneering events in absolute terms. One may wonder then, whether the relative differences translates into economically meaningful contributions on aggregate.

Figure 11: Aggregate Employment in Infant Industries by Ownership of Pioneer



Notes: Authors' calculations based on data from the Vietnam Enterprise Survey (VES). The top panel shows total employment in the VES data, decomposed by employment in incumbent industries and infant industries started by different pioneers. The bottom panel presents the share of workers in infant industries started by PDEs, MNEs, and SOEs.

In a final exercise, we assign all workers in the formal Vietnamese economy to employment in either incumbent industries, that were already established prior to 2004, and infant industries, which were pioneered during our sample period. We further decompose infant industry employment into employment in local industries started by PDEs, SOEs, or MNEs. The top panel of Figure 11 shows this breakdown of total employment. Trough our sample period, most employees continue to work in industries that existed prior to 2004. These incumbent industries have experienced a significant

employment growth and they accounted for around 90 percent of total employment in 2017.

In the bottom panel of Figure 11, we zoom in on the employment in infant industries and show the share of employment in these industries attributable to pioneering by the three different firm types. In 2017, almost 40 percent of all Vietnamese workers in local industries established after 2004 work in MNE-pioneered industries. This is remarkable considering that only about 5.5 percent of all pioneering is conducted by MNEs. As shown above, this employment effect is mainly driven by employment in MNEs and not by crowding-in of the private domestic sector.

8. CONCLUSION

In this paper, we analyse the role of ownership in pioneer firms in Vietnam. In particular, we document the extent to which firms with different ownership structures have moved people into new economic activities in Vietnam during the past two decades. First, we show that MNEs and SOEs are more likely to pioneer than PDEs. MNEs tend to pioneer in the complex industries in Vietnam's economic centers, while SOEs diffuse existing industries to peripheral regions. The differences in pioneering frequencies are largely attributable to differences in firm characteristics and industry- and province placements. Yet, we find significant ownership-specific differences even after controlling for a broad spectrum of observable characteristics. Using a simple conceptual framework, we attribute the remaining ownership-specific differences to differences in access to capital, knowledge of firm and industry productivity, and firm-specific strategies and policy objectives.

Second, we document a large employment premium in MNE-pioneered industries. The premium is driven entirely by employment in pioneering MNEs and other MNEs following the pioneer in subsequent years. We also show that the additional employment created by the follower firms (the indirect employment premium) is explained by MNEs ability to pick winners, i.e. industries that prove profitable for other MNEs in Vietnam. In 2017, nearly 40 percent of all employment industries established after 2004 are attributable to pioneering by MNEs. We find no evidence of additional crowding-in of domestic firms in MNE-pioneered industries. Our results provide nuances to the unsettled debate on the effect of FDI on entries and exits of domestic firms (De Backer and Sleuwaegen, 2003; Kosova, 2010; Lu et al., 2017). The conceptual framework suggests that the

missing crowding-in effect may be related to the absence of horizontal spillovers from MNEs, which emulating PDEs could take advantage of. This is in line with the empirical FDI literature, where the existence of horizontal spillovers has proved difficult to establish (see, e.g., Moran (2011) and Kugler (2006)).

In contrast, there is no employment premium in industries started by state-owned versus domestic and private-owned pioneers. Pioneering SOEs are systematically larger than pioneering PDEs, but this direct employment effect is counterbalanced by less employment in PDEs and MNEs in the long run. These results align with other studies finding higher entry-barriers in SOE-dominated sectors (Baccini et al., 2019; Brandt et al., 2020).

Our findings also relate to recent literature studying how countries and regions diversify (Hidalgo et al., 2007; Neffke et al., 2011), and who are the agents pushing this diversification process forward (Hausmann and Neffke, 2019; Jara-Figueroa et al., 2018; Javorcik et al., 2018; Lo Turco and Maggioni, 2019; Neffke et al., 2018). In East and South-East Asia, such questions are tightly linked to an old debate of whether the remarkable economic growth rates in this region are outcomes of state-led or market-driven development (Malesky and London, 2014; Wade, 1990; World Bank, 1993). We show that MNEs have played a central role in moving people into new economic activities in Vietnam, but we also find that SOEs have played a role in diffusing industries to peripheral provinces. By studying the ownership of pioneer firms, we add new evidence to the literature on pioneer firms and regional diversification, and our results may simultaneously help to reconcile different explanations of Vietnam's industrial transformation.

REFERENCES

- Aitken, B. J. & Harrison, A. E. (1999).** Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela. *American Economic Review*, 89(3), 605–618.
- Anh, V. T. T. (2005).** Vietnam - The Long March to Equitization. The William Davidson Institute, Policy Brief #33.
- Anh, V. T. T. (2014).** WTO Accession and the Political Economy of State-Owned Enterprise Reform in Vietnam. GEG Working Paper 2014/92.
- Anwar, S. & Nguyen, L. P. (2010).** Foreign Direct Investment and Economic Growth in Vietnam. *Asia Pacific Business Review*, 16(1-2), 183–202.
- Arnold, J. M. and Javorcik, B. S. (2009).** Gifted Kids or Pushy Parents? Foreign Direct Investment and Plant Productivity in Indonesia. *Journal of International Economics*, 79(1), 42–53.
- Arrow, K. J. (1962).** The Economic Implications of Learning by Doing. *Review of Economic Studies*, 29(3), 155–173.
- Azmat, G., Manning, A. & Reenen, J. V. (2012).** Privatization and the Decline of Labour's Share: International Evidence from Network Industries. *Economica*, 79(315), 470–492.
- Baccini, L., Impullitti, G. & Malesky, E. J. (2019).** Globalization and State Capitalism: Assessing Vietnam's Accession to the WTO. *Journal of International Economics*, 119, 75–92.
- Bai, C.-E., Li, D. D., Tao, Z. & Wang, Y. (2000).** A Multitask Theory of State Enterprise Reform. *Journal of Comparative Economics*, 28(4), 716–738.
- Bai, C.-E., Lu, J. & Tao, Z. (2006).** The Multitask Theory of State Enterprise Reform: Empirical Evidence from China. *American Economic Review*, 96(2), 353–357.
- Bajgar, M. & Javorcik, B. (2020).** Climbing the Rungs of the Quality Ladder: FDI and Domestic Exporters in Romania. *The Economic Journal*, 130(628), 937-955.

-
- Beraja, M., Yang, D. & Yuchtman, N. (2021).** Data-Intensive Innovation and the State: Evidence from AI Firms in China. NBER Working Paper No. 27723. Cambridge: National Bureau of Economic Research.
- Berkowitz, D., Ma, H. & Nishioka, S. (2017).** Recasting the Iron Rice Bowl: The Reform of China's State-Owned Enterprises. *The Review of Economics and Statistics*, 99(4), 735–747.
- Bernard, A. B., Redding, S. J. & Schott, P. K. (2010).** Multiple-Product Firms and Product Switching. *American Economic Review*, 100(1), 70–97.
- Bortolotti, B., Fotak, V. & Wolfe, B. (2019).** Innovation at State Owned Enterprises. BAFFI CAREFIN Centre Research Paper No. 2018-72.
- Borusyak, K., Hull, P. & Jaravel, X. (2022).** Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies*, 89(1), 181–213.
- Borusyak, K., Xavier, J. & Spiess, J. (2021).** Revisiting Event Study Designs: Robust and Efficient Estimation. arXiv preprint arXiv:2108.12419.
- Brambilla, I. (2009).** Multinationals, Technology, and the Introduction of Varieties of Goods. *Journal of International Economics*, 79(1), 89–101.
- Brandt, L., Hsieh, C.-t. & Zhu, X. (2008).** Growth and Structural Transformation in China. In Brandt, L. & Rawski, T. G. (eds.) *China's Great Economic Transformation*, (pp. 683-728). Cambridge University Press.
- Brandt, L., Kambourov, G. & Storesletten, K. (2020).** Barriers to Entry and Regional Economic Growth in China. CEPR Discussion Paper No. DP14965.
- Brown, J. D., Earle, J. S. & Telegdy, Á (2006).** The Productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia, and Ukraine. *Journal of Political Economy*, 114(1), 61-99.
- Brundenius, C. & Le Dang, D. (2014).** Innovation, Entrepreneurship, and SMEs: What can Cuba Learn from the Vietnam Reform Process? In C. Brundenius & R. Torres-Pérez (eds.)

-
- No More Free Lunch: Reflections on the Cuban Economic Reform Process and Challenges for Transformation*, (pp. 153–172). Springer International Publishing, Cham.
- Bui, A. T., Nguyen, C. V. & Pham, T. P. (2019).** Impact of Foreign Investment on Household Welfare: Evidence from Vietnam. *Journal of Asian Economics*, 64, 101130.
- Cahan, D. (2019).** Electoral Cycles in Government Employment: Evidence from US Gubernatorial Elections. *European Economic Review*, 111, 122–138.
- Callaway, B. & Sant’Anna, P. H. C. (2021).** Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2), 200–230.
- Cao, X., Cumming, D. & Zhou, S. (2020).** State Ownership and Corporate Innovative Efficiency. *Emerging Markets Review*, 44, 100699.
- Cao, Y., Fisman, R., Lin, H. & Wang, Y. (forthcoming).** SOEs and Soft Incentive Constraints in State Bank Lending. *American Economic Journal: Economic Policy*.
- De Backer, K. & Sleuwaegen, L. (2003).** Does Foreign Direct Investment Crowd Out Domestic Entrepreneurship? *Review of Industrial Organization*, 22(1), 67–84.
- de Chaisemartin, C. & D’Haultfœuille, X. (2020).** Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9), 2964–2996.
- Do, Q. A., Le, Q. H., Nguyen, T. D., Vu, V. A., Tran, L. H. & Nguyen, C. T. T. (2021).** Spatial Impact of Foreign Direct Investment on Poverty Reduction in Vietnam. *Journal of Risk and Financial Management*, 14(7), 292.
- Eck, K. & Huber, S. (2016).** Product Sophistication and Spillovers from Foreign Direct Investment. *Canadian Journal of Economics/Revue Canadienne d’Économique*, 49(4), 1658–1684.
- Elekes, Z., Boschma, R. & Lengyel, B. (2019).** Foreign-Owned Firms as Agents of Structural Change in Regions. *Regional Studies*, 53(11), 1603–1613.
- Fortin, N., Lemieux, T. & Firpo, S. (2011).** Decomposition Methods in Economics. In O. Ashenfelter & D. Card (eds.) *Handbook of Labor Economics* (pp. 1-102, Volume 4, Part A) North Holland.

-
- Freund, C. & Pierola, M. D. (2010).** Export Entrepreneurs: Evidence from Peru. Policy Research Working Paper No. 5407. World Bank.
- Girma, S., Gong, Y. & Görg, H. (2008).** Foreign Direct Investment, Access to Finance, and Innovation Activity in Chinese Enterprises. *World Bank Economic Review*, 22(2), 367–382.
- Goodman-Bacon, A. (2021).** Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2), 254–277.
- Government of Vietnam (2000).** Decree No. 02/2000/ND-CP. <https://luatminhkhue.vn/en/decree-no-02-2000-nd-cp-dated-february-03-2000-of-the-government-on-business-registration.aspx>. Accessed 06.04.2022.
- Government of Vietnam (2004).** Decree No. 109/2004/ND-CP. <https://luatminhkhue.vn/en/decree-no-109-2004-nd-cp-dated-april-02-2004-of-the-government-on-business-registration.aspx>. Accessed 06.04.2022.
- Government of Vietnam (2013).** Decision No. 339/QĐ-TTg. <https://english.luatvietnam.vn/decision-no-339-qd-ttg-dated-february-19-2013-of-the-prime-minister-approving-the-overall-scheme-on-restructuring-economy-in-association-with-transf-76860-Doc1.html>. Accessed 08.04.2022.
- Grossman, G. M. & Helpman, E. (1989).** Product Development and International Trade. *Journal of Political Economy*, 97(6), 1261–1283.
- GSO (2019).** Explanation of Terminology, Content and Methodology of Some Statistical Indicators on Enterprises. <https://www.gso.gov.vn/en/metadata/2019/03/enterprise/>. Accessed 22.10.2021.
- Gu, Z., Tang, S. & Wu, D. (2020).** The Political Economy of Labor Employment Decisions: Evidence from China. *Management Science*, 66(10), 4703–4725.
- Görg, H. & Greenaway, D. (2004).** Much Ado about Nothing? Do Domestic Firms Really Benefit from Foreign Direct Investment? *World Bank Research Observer*, 19(2), 171–197.
- Hansen, H., Rand, J. & Torm, N. (2016).** The Impact of Minimum Wage Adjustments. International Labour Organization.

-
- Hausmann, R., Hwang, J. & Rodrik, D. (2007).** What you Export Matters. *Journal of Economic Growth*, 12(1), 1–25.
- Hausmann, R. & Klinger, B. (2006).** Structural Transformation and Patterns of Comparative Advantage in the Product Space. KSG Working Paper No. RWP06-041, CID Working Paper No. 128.
- Hausmann, R. & Klinger, B. (2007).** The Structure of the Product Space and the Evolution of Comparative Advantage. CID Working Paper No. 146.
- Hausmann, R. & Neffke, F. M. H. (2019).** The Workforce of Pioneer Plants: The Role of Worker Mobility in the Diffusion of Industries. *Research Policy*, 48(3), 628–648.
- Hausmann, R. & Rodrik, D. (2003).** Economic Development as Self-Discovery. *Journal of Development Economics*, 72, 603–633.
- Hemmer, D. H.-R. & Nguyen Thi, P. H. (2002).** Contribution of Foreign Direct Investment to Poverty Reduction: The case of Vietnam in the 1990s. *Entwicklungsökonomische Diskussionsbeiträge*, No. 30.
- Hidalgo, C. & Hausmann, R. (2009).** The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences*, 106(26), 10570–10575.
- Hidalgo, C. A., Klinger, B., Barabasi, A.-L. & Hausmann, R. (2007).** The Product Space Conditions the Development of Nations. *Science*, 317(5837), 482–487.
- Hoff, K. (1997).** Bayesian Learning in an Infant Industry Model. *Journal of International Economics*, 43(3-4), 409–436.
- Hsieh, C.-T. & Song, Z. (2015).** Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China. NBER Working Paper Series No. 21006.
- Iacovone, L. & Javorcik, B. S. (2010).** Multi-Product Exporters: Product Churning, Uncertainty and Export Discoveries. *Economic Journal*, 120(544), 481–499.
- Jaax, A. (2020).** Private Sector Development and Provincial Patterns of Poverty: Evidence from Vietnam. *World Development*, 127, 104747.

-
- Jara-Figueroa, C., Jun, B., Glaeser, E. L. & Hidalgo, C. A. (2018).** The Role of Industry-Specific, Occupation-Specific, and Location-Specific Knowledge in the Growth and Survival of New Firms. *Proceedings of the National Academy of Sciences*, 115(5), 12646–12653.
- Javorcik, B. S. (2004).** Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers Through Backward Linkages. *American Economic Review*, 94(3), 605-627.
- Javorcik, B. S., Lo Turco, A. & Maggioni, D. (2018).** New and Improved: Does FDI Boost Production Complexity in Host Countries? *Economic Journal*, 128(614), 2507–2537.
- Jenks, G. F. (1967).** The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, 7, 186–190.
- Jovanovic, B. (1982).** Selection and the Evolution of Industry. *Econometrica*, 50(3), 649–670.
- Jovanovic, B. & Nyarko, Y. (1996).** Learning by Doing and the Choice of Technology. *Econometrica*, 64(6), 1299-1310.
- Kline, P. (2011).** Oaxaca-Blinder as a Reweighting Estimator. *American Economic Review*, 101(3), 532–537.
- Kornai, J. (1986).** The Soft Budget Constraint. *Kyklos*, 39(1), 3-30.
- Kosova, R. (2010).** Do Foreign Firms Crowd Out Domestic Firms? Evidence from the Czech Republic. *Review of Economics and Statistics*, 92(4), 861–881.
- Kugler, M. (2006).** Spillovers from Foreign Direct Investment: Within or Between Industries? *Journal of Development Economics*, 80(2), 444–477.
- La Porta, R. & Lopez-de-Silanes, F. (1999).** The Benefits of Privatization: Evidence from Mexico. *Quarterly Journal of Economics*, 114(4), 1193–1242.
- Li, Q., Lin, C. & Xu, L. (2020).** Political Investment Cycles of State-Owned Enterprises. *Review of Financial Studies*, 33(7), 3088–3129.

-
- Lin, J. Y., Cai, F. & Li, Z. (1998).** Competition, Policy Burdens, and State-Owned Enterprise Reform. *American Economic Review*, 88(2), 422–427.
- Lin, J. Y. & Tan, G. (1999).** Policy Burdens, Accountability, and the Soft Budget Constraint. *American Economic Review*, 89(2), 426–431.
- Lo Turco, A. & Maggioni, D. (2019).** Local Discoveries and Technological Relatedness: the Role of MNEs, Imports and Domestic Capabilities. *Journal of Economic Geography*, 19(5), 1077–1098.
- Loc, T. D., Lanjouw, G. & Lensink, R. (2006).** The Impact of Privatization on Firm Performance in a Transition Economy: The Case of Vietnam. *Economics of Transition*, 14(2), 349–389.
- Loc, T. D. & Tran, N. M. (2016).** Impact of Equitization on Performance of Enterprises in Vietnam. *Journal of Economic Development*, 23(3), 36–56.
- Lu, Y., Tao, Z. & Zhu, L. (2017).** Identifying FDI Spillovers. *Journal of International Economics*, 107, 75–90.
- Malesky, E. & London, J. (2014).** The Political Economy of Development in China and Vietnam. *Annual Review of Political Science*, 17(1), 395–419.
- Malesky, E. J. & Taussig, M. (2009).** Where Is Credit Due? Legal Institutions, Connections, and the Efficiency of Bank Lending in Vietnam. *The Journal of Law, Economics, and Organization*, 25(2), 535–578.
- Markusen, J. R. & Venables, A. J. (1999).** Foreign Direct Investment as a Catalyst for Industrial Development. *European Economic Review*, 43, 335–356.
- Mayneris, F. & Poncet, S. (2015).** Chinese Firms' Entry to Export Markets: The Role of Foreign Export Spillovers. *World Bank Economic Review*, 29(1), 150–179.
- McCaig, B. & Pavcnik, N. (2017)** Moving out of Agriculture: Structural Change in Viet Nam. In McMillan, M. S., Rodrik, D. & Sepulveda, C. P. (eds.) *Structural Change, Fundamentals, and Growth: A Framework and Case Studies* (pp. 81-124), International Food Policy Research Institute.

-
- McCaig, B., Pavcnik, N. & Wong, W. F. (2020).** Export Markets and Long-Run Industry Adjustment: State, Private, and Foreign Firms in Vietnam. Unpublished Manuscript.
- Meggison, W. L. & Netter, J. M. (2001).** From State to Market: A Survey of Empirical Studies on Privatization. *Journal of Economic Literature*, XXXIX, 321–389.
- Moran, T. (2011).** *Foreign Direct Investment and Development: Launching a Second Generation of Policy Research*. Peterson Institute for International Economics.
- Neffke, F., Hartog, M., Boschma, R. & Henning, M. (2018).** Agents of Structural Change: The Role of Firms and Entrepreneurs in Regional Diversification. *Economic Geography*, 94(1), 23–48.
- Neffke, F., Henning, M. & Boschma, R. (2011).** How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, 87(3), 237–265.
- Newman, C., Page, J., Rand, J., Shimeles, A., Soderbom, M. & Tarp, F. (2020).** Linked-in by FDI: The Role of Firm-Level Relationships for Knowledge Transfers in Africa and Asia. *The Journal of Development Studies*, 56(3), 451–468.
- Newman, C., Rand, J., Talbot, T. & Tarp, F. (2015).** Technology Transfers, Foreign Investment and Productivity Spillovers. *European Economic Review*, 76, 168–187.
- Newman, C., Rand, J. & Tarp, F. (2013).** Industry Switching in Developing Countries. *World Bank Economic Review*, 27(2), 357–388.
- Ngo, M. T., Nonneman, W. & Jorissen, A. (2015).** Privatization of Vietnamese Firms and Its Effects on Firm Performance. *Asian Economic and Financial Review*, 5(2), 202–217.
- Nguyen, A. N. & Nguyen, T. (2008).** Foreign Direct Investment in Vietnam: Is There Any Evidence of Technological Spillover Effects. SSRN.
- Nigmatulina, D. (2021).** Misallocation and State Ownership: Evidence from the Russian Sanctions. SSRN.
- Oaxaca, R. L. & Ransom, M. R. (1999).** Identification in Detailed Wage Decompositions. *Review of Economics and Statistics*, 81(1), 154–157.

-
- OECD (2008).** OECD Benchmark Definition of Foreign Direct Investment, Fourth Edition. Organisation for Economic Co-operation and Development, Paris.
- Poupakis, S. (2020).** Are Inflows of FDI Good for Russian Exporters? World Bank Policy Research Working Paper No. 9201.
- Rama, M. (2008).** Making Difficult Choices: Vietnam in Transition. Working Paper No. 40, Commission on Growth and Development.
- Redding, S. (1999).** Dynamic Comparative Advantage and the Welfare Effects of Trade. *Oxford Economic Papers*, 51(1), 15–39.
- Rodríguez-Clare, A. (1996).** Multinationals, Linkages, and Economic Development. *American Economic Review*, 86(4), 852–873.
- Romer, P. M. (1992).** Two Strategies for Economic Development: Using Ideas and Producing Ideas. *World Bank Economic Review*, 63–91.
- Romer, P. M. (1993).** Idea Gaps and Object Gaps in Economic Development. *Journal of Monetary Economics*, 32(3), 543–573.
- Roth, J. (2022).** Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends. *American Economic Review: Insights*, forthcoming.
- Setzler, B. & Tintelnot, F. (2021).** The Effects of Foreign Multinationals on Workers and Firms in the United States. *Quarterly Journal of Economics*, 136(3), 1943–1991.
- Shleifer, A. & Vishny, R. W. (1994).** Politicians and Firms. *Quarterly Journal of Economics*, 109(4), 995–1025.
- Sun, L. & Abraham, S. (2021).** Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, 225(2), 175–199.
- Sørensen, B. B. (2020).** Turnin’ it up a Notch: How Spillovers from Foreign Direct Investment Boost the Complexity of South Africa’s Exports. WIDER Working Paper 2020/3, Helsinki: UNU-WIDER.

- Tran, T. Q. & Tien, Q. T. (2008).** Reforms in FDI Policy and the Investment Climate in Vietnam. *Journal of World Trade*, 42(6), 1179–1202.
- Van Arkadie, B. & Mallon, R. (2004).** *Viet Nam: A Transition Tiger?* Canberra: The Australian National University Press and Asia Pacific Press.
- Verhoogen, E. (2021).** Firm-Level Upgrading in Developing Countries. NBER Working Paper No. 29461.
- Wade, R. (1990).** *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization.* Princeton University Press.
- Wagner, R. & Zahler, A. (2015).** New Exports from Emerging Markets: Do Followers Benefit from Pioneers? *Journal of Development Economics*, 114, 203–223.
- Wei, S.-J., Wei, Z. & Xu, J. (2021).** On the Market Failure of “Missing Pioneers”. *Journal of Development Economics*, 152, 102705.
- Wen, J. Y. (2022).** State Employment as a Strategy of Autocratic Control in China. Unpublished Working Paper.
- Wooldridge, J. (2010).** *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
- World Bank (1993).** *The East Asian Miracle: Economic Growth and Public Policy*, Oxford University Press, New York.
- World Bank (2022).** World Development Indicators. <https://datatopics.worldbank.org/world-development-indicators/>. Accessed 06.04.2022.

APPENDIX

This appendix is organized as follows. Appendix A.1 and A.4 are data appendices. In Appendix A.1, we provide supporting documentation for Section 3. In that appendix section, we describe how we clean and define relevant variables used in this paper. In contrast, Appendix A.4 is a general data appendix, where we present a detailed description of the VES data and outline the general cleaning procedure we have applied to prepare it for analysis. That section is general and not specific to this paper. It includes, for instance, a description of the VES sampling procedure, how we generate firm IDs so that we can track firms over time, how we clean industries, geo-codes, balance sheet and employment variables, generate deflators, and so forth.

Appendices A.2 and A.3 provide supporting evidence for our analyses. Appendix A.2 presents descriptive statistics discussed in Section 4. The appendix also outlines how we calculate the complexity of different industries. In Appendix A.3, we deliver supporting evidence to, and robustness checks of, the analysis on industry dynamics presented in Section 7.

A.1. Data Appendix

In this section, we outline supporting information for Section 3. For a general description of the how we clean the raw VES data, the reader is referred to Appendix A.4.

A.1.1. SOE Definition

We define SOEs as enterprises that are either *(i)* fully state-owned, *(ii)* limited liability companies (LLCs) with at least one state member, *(iii)* joint-stock companies (JSCs) with state capital, or *(iv)* joint ventures (JVs) with state-ownership. LLCs are legal entities established through capital contributions of one or more members. Those members are liable for the financial obligations up to the corresponding amount of their capital contributions. JSCs are entities established via the distribution of shares. JSCs must have at least three shareholders and the founding members' shares correspond to the amount of their invested capital. In both enterprise forms, the state can serve as a member or shareholder. JVs are usually LLCs or JSCs too, where one or more members or shareholders are foreign companies or individuals. JVs with state ownership are thus LLCs or

JSCs with at least one foreign member or shareholder and with the Vietnamese state as a member or shareholder. Different national or local state entities can act as members or shareholders. Our definition of SOEs deviates slightly from the official definition from GSO. After GSO's definition, JSCs only count as SOEs if they have more than 50 percent state capital. GSO further does not include JVs with state-membership into their SOE definition (GSO, 2019). We deviate from the GSO classification for two reasons: First, we argue that the state can possess a significant influence on the enterprise's decision making even if it only holds a minor share. The state can thus *control* an enterprise even if it does not fully *own* it (McCaig et al., 2020). This idea is also in line with OECD's (2008) benchmark definition of FDI, where it is argued that foreign investors can be expected to have "lasting interest" in a firm, if they own at least 10 percent of the voting rights in that firm. In turn, the objective of establishing a lasting interest is to "*ensure a significant degree of influence by the direct investor in the management of the direct investment enterprise*" (OECD, 2008, pp.17). Second, due to the changing nature of the survey, we do not possess information on the amount of state-ownership in all years. Using our broader SOE definition is therefore the most consistent way of classification in our panel.

A.1.2. Dealing with Temporary Industry Switchers

As described in Section 3, there is a disproportionate amount of pioneering in certain industries because of unrealistically high rates of temporary industry switching. Of concern, the pattern is particularly prevalent in 2006 and 2010, when firms start and stop reporting under multiple VSIC classification systems. We take this as evidence of reporting by "confused firms". One way in which we addressed the issue is by removing temporary, one-year industry switches. That is, when a firm operates in industry A in $t=1$, then switches to industry B in $t=2$, before it returns to A in $t=3$, we replace the firm's industry in $t=2$ with A. To simplify explanations below, we will henceforth refer to this as a "sandwich" procedure. We will refer to industry A as the "bread" in the sandwich, and industry B as the "lettuce", because it is flanked by industry A in $t = 1$ and $t = 3$. The procedure is easy to implement for single-industry firms and illustrated in Table A1.

For multi-industry firms, the cleaning is more involved. Now, we face five different scenarios. In scenario (I), there is one bread and one lettuce industry. In scenario (II), there are the same

Table A1: Cleaning of Temporary Switchers for Single-Industry Firms

Firm ID	t	Raw VSIC	Type	Cleaned VSIC
1	1	A	Bread	A
1	2	B	Lettuce	A
1	3	A	Bread	A

Notes: Example of cleaning procedure for temporary industry switchers for single-industry firms.

amount of lettuce and bread industries, but more than one of each. Here, we need to link each bread industry to a lettuce industry. We only clean the simple case when we observe two bread industries and two lettuce industries. The 2×2 case is extremely rare (46 cases) and other cases are close-to non-existent. Further, the possibility of matching industries incorrectly is much higher when there are more than two lettuce and bread industries since the complexity of the matching problem increases exponentially. In scenario (III), there is one lettuce industry and multiple bread industries, and we need to match one of them to the lettuce industry. Conversely, there are multiple lettuce industries and one bread industry in scenario (IV), and we need to assign the bread industry to one lettuce industry. Finally, in scenario (V), there are multiple, but an unequal number of lettuce and bread industries. This is a very rare case too, and it poses a very complex matching problem, that we do not address in our cleaning. Below, we describe how we deal with scenarios (I)-(IV).

Table A2: Cleaning of Temporary Switchers for Multi-Industry Firms - Scenario (I)

Firm ID	t	Raw VSIC	Type	Cleaned VSIC
1	1	A	-	A
1	1	B	Bread	B
1	2	A	-	A
1	2	C	Lettuce	B
1	3	A	-	A
1	3	B	Bread	B

Notes: Example of cleaning procedure for temporary industry switchers for multi-industry firms under scenario (I), where there is one lettuce and one bread industry.

Scenario (I) is easy to clean. We follow the procedure outlined in Table A2. In scenario (II)-(IV), we link a firm's multiple lettuce and bread industries based on the firm's reported number of workers in those industries. Specifically, we use the employment numbers in $t = 1$ and $t = 3$ in the sandwich

industries to calculate expected employment in $t = 2$, assuming a linear trend. We then match industries based on the smallest difference between the real and expected employment numbers in $t = 2$. Illustrative examples are provided in Table A3.

To ensure that we only make sensible replacements of lettuce industries, we only replace industry codes if the absolute difference between the actual employment in the lettuce industry ($EMP_t^{VSIC^{lettuce}}$) and any of the bread industries ($EMP_{t-1}^{VSIC^{bread}}, EMP_{t+1}^{VSIC^{bread}}$) is not too large. Specifically, we implement the following rule:

Replace VSIC^{lettuce} with VSIC^{bread} if :

$$\max \left(\left| \frac{EMP_t^{VSIC^{lettuce}} - EMP_{t-1}^{VSIC^{bread}}}{EMP_t^{VSIC^{lettuce}}} \right|, \left| \frac{EMP_t^{VSIC^{lettuce}} - EMP_{t-1}^{VSIC^{bread}}}{EMP_{t-1}^{VSIC^{bread}}} \right| \right) < 2.5$$

OR

$$\max \left(\left| \frac{EMP_t^{VSIC^{lettuce}} - EMP_{t+1}^{VSIC^{bread}}}{EMP_t^{VSIC^{lettuce}}} \right|, \left| \frac{EMP_t^{VSIC^{lettuce}} - EMP_{t+1}^{VSIC^{bread}}}{EMP_{t+1}^{VSIC^{bread}}} \right| \right) < 2.5.$$

In the year where a firm starts up, we cannot use the sandwich procedure outlined above, because the industry in the first year is not flanked by two bread industries. Instead, we replace a firm's industry in the first year of operation, if it consistently reports a different industry code (which does not change) in the following three years. Hence, we only apply this procedure to firms that we observe for at least four years. An example for single-industry firms are shown in Table A4. For multi-industry firms, we follow a similar procedure as the one outlined above, except that we now calculate the expected employment in $t = 1$ by fitting a linear line with employment numbers in $t = 2, t = 3$, and $t = 4$.

Table A3: Cleaning of Temporary Switchers for Multi-Industry Firms - Scenario (II)-(IV)

Firm ID	t	Raw VSIC	Type	Employment	Expected employment		Cleaned VSIC
					VSIC A	VSIC B	
Scenario (II)							
1	1	A	Bread	10	-	-	A
1	1	B	Bread	200	-	-	B
1	2	C	Lettuce	250	20	325	B
1	2	D	Lettuce	15	20	325	A
1	3	A	Bread	30	-	-	A
1	3	B	Bread	450	-	-	B
Scenario (III)							
1	1	A	Bread	10	-	-	A
1	1	B	Bread	200	-	-	B
1	1	C	-	40	-	-	C
1	2	D	Lettuce	250	20	325	B
1	2	C	-	15	-	-	C
1	3	A	Bread	30	-	-	A
1	3	B	Bread	450	-	-	B
1	3	C	-	50	-	-	C
Scenario (IV)							
1	1	A	-	10	-	-	A
1	1	B	Bread	200	-	-	B
1	1	C	-	40	-	-	C
1	2	D	Lettuce	250	20	325	D
1	2	E	Lettuce	320	20	325	B
1	2	C	-	15	-	-	C
1	3	B	Bread	450	-	-	B
1	3	C	-	50	-	-	C

Notes: Example of cleaning procedure for temporary industry switchers for multi-industry firms under scenario (II)-(IV).

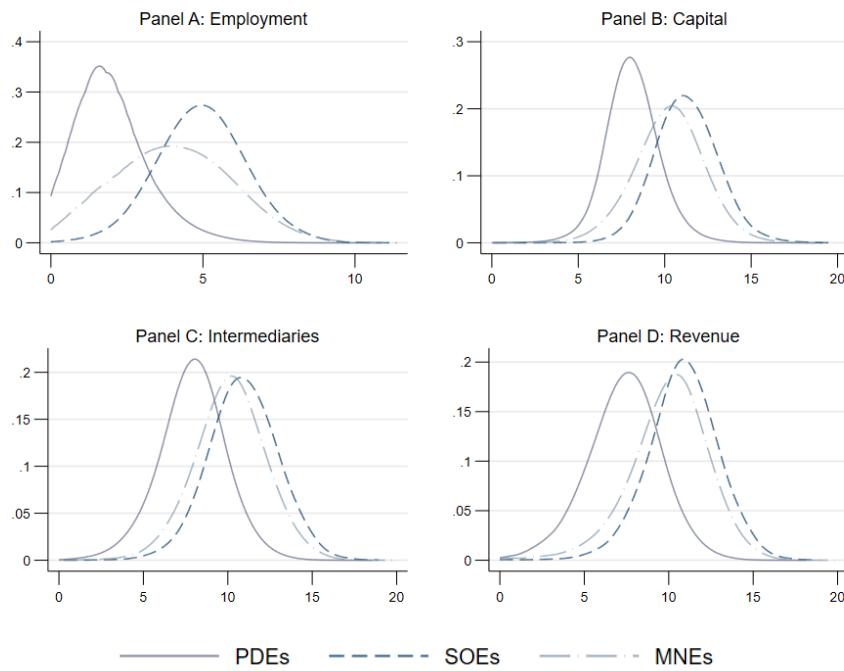
Table A4: Cleaning of Temporary Switchers for Single-Industry Start-Up Firms

Firm ID	t	Raw VSIC	Status	Cleaned VSIC
1	0	-	not active	-
1	1	B	active	A
1	2	A	active	A
1	3	A	active	A
1	4	A	active	A

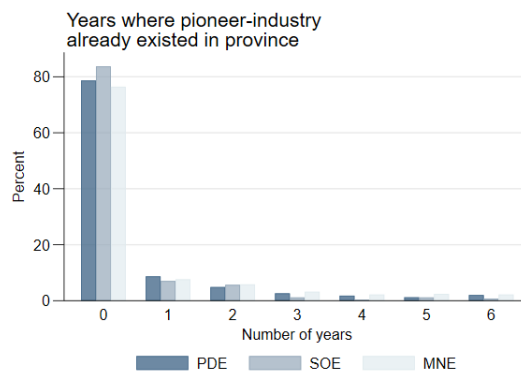
Notes: Example of cleaning procedure for temporary industry switchers for single-industry start-up firms.

A.2. Supporting Results to Section 4: Stylized Facts

A.2.1. Figures

Figure A1: Density Plots by Ownership Type

Notes: Authors own calculations based on data from the Vietnam Enterprise Survey. All variables are expressed in logarithms.

Figure A2: Novelty of pioneered industries in province

Notes: Authors own calculations based on data from the Vietnam Enterprise Survey.

A.2.2. Tables

Tabulation of Ownership Type and Pioneering Status: Table A5 reports a tabulation of ownership type by pioneer type. We distinguish here between three types of pioneers: new firms starting a new industry (start-ups); firms diversifying their existing business into a new industry (multi-industry firms, MIE); and firms that switch from an existing industry to a new one (switchers). Overall, pioneering is dominated by private domestic start-ups. Among pioneering MNEs, start-ups are the most common form of pioneering too. Note, however, that this only means that these firms have not been operating in Vietnam before. So the number of start-up MNEs is likely to be existing enterprises that are starting to operate in Vietnam. Pioneering SOEs, on the other hand, are most often multi-industry firms.

Table A5: Cross Tabulation of Ownership Type and Pioneering Status

Ownership Type	Pioneering Type			Total
	Start-up	MIE	Switcher	
PDE	6,200	1,118	1,767	9,085
SOE	181	277	84	542
MNE	396	44	118	558
Total	6,777	1,439	1,969	10,185

Test Statistics on Pioneering and Ownership: To formally test whether pioneering is dependent on ownership type, we calculate test statistics for state-ownership and pioneering. Table A6 reports year-by-year Pearson's χ^2 (with the corresponding p -values), where the Null-hypothesis is independence between ownership type (SOEs vs. PDEs, excluding MNEs) and pioneering. We additionally report the same test statistics as a robustness check for three alternative ways of defining pioneering. We change the three-year period, in which an industry must not have been observed, to four and five years, respectively, to make sure that our main results are not driven by the somewhat arbitrary choice of three years. We further restrict the definition so that an industry must have never been observed before within a province. In Tables A7 and A8 we report the same test statistics for MNEs vs. PDEs and SOEs vs. MNEs, respectively.

Table A6: Test Statistics: SOEs vs. PDEs

Year	Pioneer 3 years (baseline)	Pioneer 4 years	Pioneer 5 years	Pioneer first time
2004	29.731 (0.000)	.	.	29.731 (0.000)
2005	2.989 (0.084)	0.179 (0.672)	.	0.179 (0.672)
2006	54.141 (0.000)	53.658 (0.000)	46.951 (0.000)	46.951 (0.000)
2007	19.057 (0.000)	18.033 (0.000)	16.158 (0.000)	11.489 (0.001)
2008	0.341 (0.559)	0.082 (0.775)	0.030 (0.862)	0.349 (0.555)
2009	25.542 (0.000)	28.020 (0.000)	31.697 (0.000)	25.661 (0.000)
2010	30.793 (0.000)	35.864 (0.000)	35.093 (0.000)	38.385 (0.000)
2011	11.266 (0.001)	15.891 (0.000)	18.056 (0.000)	8.106 (0.004)
2012	26.825 (0.000)	28.915 (0.000)	21.792 (0.000)	17.905 (0.000)
2013	7.347 (0.007)	4.118 (0.042)	4.549 (0.033)	4.440 (0.035)
2014	7.085 (0.008)	6.375 (0.012)	7.947 (0.005)	3.287 (0.070)
2015	72.252 (0.000)	58.093 (0.000)	52.683 (0.000)	32.955 (0.000)
2016	51.534 (0.000)	55.873 (0.000)	68.279 (0.000)	23.557 (0.000)
2017	31.704 (0.000)	19.976 (0.000)	17.371 (0.000)	15.577 (0.000)

Reported are χ^2 -statistics and their corresponding p -values in parentheses, calculated separately for each year, where the Null-hypotheses is no difference between the likelihood of pioneering for SOEs and PDEs. No test statistics are reported when there is no pioneering observed according to the respective definition in the respective year.

Table A7: Test Statistics: MNEs vs. PDEs

Year	Pioneer 3 years (baseline)	Pioneer 4 years	Pioneer 5 years	Pioneer first time
2004	0.127 (0.722)	.	.	0.127 (0.722)
2005	3.557 (0.059)	0.377 (0.539)	.	0.377 (0.539)
2006	15.970 (0.000)	13.230 (0.000)	13.539 (0.000)	13.539 (0.000)
2007	8.881 (0.003)	8.997 (0.003)	8.583 (0.003)	8.229 (0.004)
2008	3.966 (0.046)	3.585 (0.058)	2.631 (0.105)	3.145 (0.076)
2009	48.496 (0.000)	39.053 (0.000)	36.838 (0.000)	41.416 (0.000)
2010	82.799 (0.000)	81.824 (0.000)	77.366 (0.000)	58.368 (0.000)
2011	29.500 (0.000)	34.543 (0.000)	32.925 (0.000)	24.274 (0.000)
2012	123.126 (0.000)	96.298 (0.000)	95.094 (0.000)	85.667 (0.000)
2013	41.042 (0.000)	31.472 (0.000)	33.602 (0.000)	26.919 (0.000)
2014	156.327 (0.000)	149.420 (0.000)	135.290 (0.000)	134.518 (0.000)
2015	23.248 (0.000)	22.862 (0.000)	17.040 (0.000)	8.254 (0.000)
2016	44.077 (0.000)	43.820 (0.000)	42.931 (0.000)	23.788 (0.000)
2017	16.304 (0.000)	13.472 (0.000)	14.457 (0.000)	9.380 (0.002)

Reported are χ^2 -statistics and their corresponding p -values in parentheses, calculated separately for each year, where the Null-hypotheses is no difference between the likelihood of pioneering for MNEs and PDEs. No test statistics are reported when there is no pioneering observed according to the respective definition in the respective year.

Table A8: Test Statistics: SOEs vs. MNEs

Year	Pioneer 3 years (baseline)	Pioneer 4 years	Pioneer 5 years	Pioneer first time
2004	6.098 (0.014)	.	.	6.098 (0.014)
2005	0.256 (0.613)	0.057 (0.811)	.	0.057 (0.811)
2006	1.809 (0.179)	2.335 (0.127)	1.620 (0.203)	1.620 (0.203)
2007	0.367 (0.545)	0.282 (0.596)	0.206 (0.650)	0.024 (0.876)
2008	0.861 (0.354)	1.142 (0.285)	1.492 (0.222)	2.626 (0.105)
2009	0.518 (0.472)	0.086 (0.769)	0.003 (0.954)	0.224 (0.636)
2010	1.416 (0.234)	0.899 (0.343)	0.749 (0.387)	0.090 (0.764)
2011	0.300 (0.584)	0.143 (0.706)	0.030 (0.863)	0.359 (0.549)
2012	2.146 (0.143)	0.993 (0.319)	1.622 (0.203)	1.607 (0.205)
2013	0.895 (0.344)	1.045 (0.307)	1.045 (0.307)	0.667 (0.414)
2014	5.923 (0.015)	5.672 (0.017)	4.360 (0.037)	5.722 (0.017)
2015	6.913 (0.009)	4.929 (0.026)	5.146 (0.023)	4.015 (0.045)
2016	3.329 (0.068)	3.743 (0.053)	5.190 (0.023)	1.268 (0.260)
2017	4.345 (0.037)	2.421 (0.120)	1.752 (0.186)	1.973 (0.160)

Reported are χ^2 -statistics and their corresponding p -values in parentheses, calculated separately for each year, where the Null-hypotheses is no difference between the likelihood of pioneering for SOEs and MNEs. No test statistics are reported when there is no pioneering observed according to the respective definition in the respective year.

A.2.3. Calculating Industry Complexity (Figure 5)

This section explains how we calculate industry used to produce Figure 5. We follow the Method of Reflection from Hidalgo and Hausmann (2009), but where they apply the method to calculate country and product complexity using international trade data, we apply it to calculate industry and province complexity using the Vietnam Enterprise Survey. We first aggregate employment data to the province-industry-year level, using 4-digit VSIC codes. To prevent the measure of complexity from being impacted by noise, we exclude small industries with average total revenue of less than 50 thousand Vietnamese dong between 2014 and 2017.

We can then measure the complexity of provinces and industries by relying on the assumption that complex industries are difficult to develop and therefore only present in a few, complex provinces. Vice versa, complex provinces are those with the productive capabilities necessary to establish many complex industries.

We first measure the extent to which province p has productive capabilities in industry k in year t through its Revealed Comparative Advantage (RCA) in that industry:

$$RCA_{k,p,t} = \frac{X_{p,k,t}}{\sum_p X_{p,k,t}} / \frac{\sum_k X_{p,k,t}}{\sum_{p,k} X_{p,k,t}}, \quad (13)$$

where $X_{p,k,t}$ is the employment in industry k in province p in year t .

We define province p to have strong capabilities in industry k if its RCA in k in year t is above one. To avoid noise from measurement error, we also define province p to have strong industry capabilities if its average RCA over the last four years was above one:

$$M_{p,k,t} = \begin{cases} 1 & \text{if } RCA_{k,p,t} \geq 1 \quad \text{or} \quad \frac{1}{4} \sum_{i=t-3}^t RCA_{k,p,i} \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

M_{pkt} is the adjacency matrix encoding a bipartite network that links provinces to the industries they have in their production portfolio. Each column of the matrix represents a 4-digit VSIC industry and each row represents a country.

We use the Method of Reflection from Hidalgo and Hausmann (2009) to estimate province and

industry complexity. First, note that the industrial diversity of a province and the ubiquity of an industry constitute rough approximations of respectively province and industry complexity. We find diversity and ubiquity through by summing the columns and rows of the adjacency matrix, M_{pk} :

$$Diversity_p = \alpha_{p,0} = \sum_k M_{p,k} \quad (15)$$

$$Ubiquity_k = \alpha_{k,0} = \sum_c M_{p,k} \quad (16)$$

Diversity and ubiquity are, however, only crude approximations of the productive capabilities that a province have or that is required to establish an industry. For instance, some natural resource-based industries may be rare because the natural resource itself is rare and not because their production function is difficult to learn. one can correct for this bias, by weighting the ubiquity of each industry with the diversity of the provinces that are operating them. If the natural resource-based industries tend to be located in industries that do not host many other rare (and therefore complex) industries, we can assume that they are actually unsophisticated industries. The same method of adjustment can be applied to provinces, by weighting their industrial diversity with the ubiquity of each industry. In other words, the two measures can be improved by jointly and iteratively using one to adjust the other:

$$\alpha_{p,N} = \frac{1}{\alpha_{p,0}} \sum_k M_{p,k} \cdot \alpha_{k,N-1} \quad (17)$$

$$\alpha_{k,N} = \frac{1}{\alpha_{k,0}} \sum_p M_{p,k} \cdot \alpha_{p,N-1}, \quad (18)$$

where $\alpha_{p,N}$ and $\alpha_{k,N}$ denote measures of province and industry complexity after N iterations of the iterative correction procedure.

In practice, we can calculate precise measures of complexity by inserting Equation 18 into Equation 17:

$$\begin{aligned}
k_{p,N} &= \frac{1}{\alpha_{p,0}} \sum_k M_{p,k} \frac{1}{\alpha_{k,0}} \sum_{p'} M_{p',k} \cdot \alpha_{p',N-2} \\
&= \sum_{p'} \alpha_{p',N-2} \sum_k \frac{M_{p,k} M_{p',k}}{\alpha_{p,0} \alpha_{k,0}} \\
&= \sum_{p'} \alpha_{p',N-2} \tilde{M}_{p,p'}, \tag{19}
\end{aligned}$$

where we define the province diversity matrix of interest as:

$$\tilde{M}_{p,p'} = \sum_k \frac{M_{p,k} M_{p',k}}{\alpha_{p,0} \alpha_{k,0}}. \tag{20}$$

The industry ubiquity matrix, $\tilde{M}_{p,p'}$, can analogously be found by substituting Equation 17 into Equation 18. Following Hidalgo and Hausmann (2009), the complexity measures are defined as the eigenvectors with the *second largest* eigenvalues from the two matrices. These vectors capture the largest variance of the system. We standardize the industry complexity measure and use the one-year lag to produce Figure 5. We use the lagged measure in order to avoid that firms' current pioneering activities impact the measure of complexity.

A.3. Supporting Material to Section 7: Industry Dynamics

A.3.1. Robustness Checks

Robustness Check I: Fixed Effects

Table A9: Total and Indirect Effects of Pioneering on Employment with Different Fixed Effects for Imputation

P-values from pre-trend tests	Δ Employment incl. pioneer				Δ Employment excl. pioneer			
	(1) PDEs only ($\hat{\beta}^{PDE}$)	(2) PDEs only ($\hat{\beta}^{PDE}$)	(3) SOEs vs. PDEs ($\hat{\beta}^{SOE}$)	(4) MNEs vs. PDEs ($\hat{\beta}^{MNE}$)	(5) PDEs only ($\hat{\beta}^{PDE}$)	(6) PDEs only ($\hat{\beta}^{PDE}$)	(7) SOEs vs. PDEs ($\hat{\beta}^{SOE}$)	(8) MNEs vs. PDEs ($\hat{\beta}^{MNE}$)
Year FEs								
	116.449*** (14.936)	92.186*** (10.293)	39.498* (22.462)	647.019*** (137.237)	94.319*** (14.667)	68.301*** (9.796)	-24.869 (18.628)	241.428*** (74.647)
Borusyak et al. (2021)	0	0.09	0.038	0.353	0	0.094	0.082	0.482
Placebo	0	0.063	0.126	0.146	0	0.066	0.136	0.222
Province-year FEs								
	116.354*** (14.935)	91.820*** (10.295)	39.423* (22.445)	647.050*** (137.243)	94.225*** (14.665)	67.935*** (9.797)	-24.935 (18.603)	241.438*** (74.647)
Borusyak et al. (2021)	0	.168	.502	.709	0	.174	.55	.801
Placebo	.001	.243	.68	.83	.001	.248	.543	.896
Industry-year FEs								
	116.442*** (14.936)	92.128*** (10.291)	39.515* (22.456)	646.535*** (137.229)	94.311*** (14.667)	68.251*** (9.795)	-24.845 (18.629)	240.975*** (74.635)
Borusyak et al. (2021)	0	0.081	0.52	0.501	0	0.087	0.47	0.565
Placebo	0	0.062	0.433	0.666	0	0.067	0.432	0.659
Three-way FEs								
	121.182*** (15.026)	97.196*** (10.576)	40.224* (22.656)	653.650*** (136.451)	99.037*** (14.755)	73.365*** (10.095)	-24.112 (18.908)	248.112*** (73.717)
Borusyak et al. (2021)	0	0.201	0.373	0.649	0	0.226	0.266	0.738
Placebo	0.458	0.331	0.668	0.872	0.511	0.369	0.642	0.918
Excluded event times	none	-3 to -1	none	none	none	-3 to -1	none	none
Event times used for test	-7 to -1	-7 to -4	-7 to -1	-7 to -1	-7 to -1	-7 to -4	-7 to -1	-7 to -1
Observations	79,713	49,942	87,295	86,634	79,713	49,942	87,295	86,634
Treated obs.	37,361	22,725	41,944	40,570	37,361	22,725	41,944	40,570
Control obs.	42,352	27,217	45,351	46,064	42,352	27,217	45,351	46,064

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level. The dependent variable is province-industry employment. In Columns 1-4, the dependent variable includes employment of the pioneers, while the pioneer effects are excluded in Columns 5-8. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Results from Columns 1-2 and 5-6 are the estimator presented in Equation 10. Results from Columns 3-4 and 7-8 are from the estimator presented in Equation 11. Each panel use a different set of fixed effects to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

Table A10: Total and Indirect Effects of Pioneering on Firms Counts with Different Fixed Effects for Imputation

P-values from pre-trend tests	Δ Firm count incl. pioneer				Δ Firm count excl. pioneer			
	(1) PDEs only ($\hat{\beta}^{PDE}$)	(2) PDEs only ($\hat{\beta}^{PDE}$)	(3) SOEs vs. PDEs ($\hat{\beta}^{SOE}$)	(4) MNEs vs. PDEs ($\hat{\beta}^{MNE}$)	(5) PDEs only ($\hat{\beta}^{PDE}$)	(6) PDEs only ($\hat{\beta}^{PDE}$)	(7) SOEs vs. PDEs ($\hat{\beta}^{SOE}$)	(8) MNEs vs. PDEs ($\hat{\beta}^{MNE}$)
Year FEs								
	3.342*** (0.171)	2.909*** (0.189)	-0.099 (0.434)	0.096 (0.500)	2.581*** (0.143)	2.006*** (0.148)	-0.122 (0.398)	0.003 (0.481)
Borusyak et al. (2021)	0	0	0.035	0.763	0	0	0.061	0.685
Placebo	0	0	0.005	0.222	0	0	0.004	0.195
Province-year FEs								
	3.342*** (0.171)	2.908*** (0.189)	-0.100 (0.433)	0.098 (0.500)	2.581*** (0.143)	2.006*** (0.148)	-0.123 (0.397)	0.005 (0.481)
Borusyak et al. (2021)	0	0	0.672	0.89	0	0	0.782	0.83
Placebo	0	0	0.957	0.863	0	0	0.919	0.815
Industry-year FEs								
	3.342*** (0.171)	2.910*** (0.189)	-0.101 (0.434)	0.098 (0.500)	2.581*** (0.143)	2.008*** (0.148)	-0.123 (0.398)	0.005 (0.481)
Borusyak et al. (2021)	0	0	0.76	0.952	0	0	0.872	0.904
Placebo	0	0	0.867	0.939	0	0	0.921	0.908
Three-way FEs								
	3.371*** (0.171)	2.908*** (0.188)	-0.092 (0.434)	0.083 (0.491)	2.610*** (0.143)	2.006*** (0.148)	-0.115 (0.398)	-0.010 (0.473)
Borusyak et al. (2021)	0	0	0.616	0.832	0	0	0.677	0.789
Placebo	0	0	0.552	0.841	0	0	0.398	0.835
Excluded event times	none	-3 to -1	none	none	none	-3 to -1	none	none
Event times used for test	-7 to -1	-7 to -4	-7 to -1	-7 to -1	-7 to -1	-7 to -4	-7 to -1	-7 to -1
Observations	79,713	49,942	87,295	86,634	79,713	49,942	87,295	86,634
Treated obs.	37,361	22,725	41,944	40,570	37,361	22,725	41,944	40,570
Control obs.	42,352	27,217	45,351	46,064	42,352	27,217	45,351	46,064

Notes: *** p<0.01, ** p<0.05, * p<0.1. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level. The dependent variable is province-industry firm counts. In Columns 1-4, the dependent variable includes the firms of the pioneers, while the pioneer effects are excluded in Columns 5-8. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Results from Columns 1-2 and 5-6 are the estimator presented in Equation 10. Results from Columns 3-4 and 7-8 are from the estimator presented in Equation 11. Each panel use a different set of fixed effects to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

Robustness Check II: Clustered Standard Errors

Table A11: Total and Indirect Effects of Pioneering with Differently Clustered Standard Errors

	(1) SOEs vs. PDEs	(2) MNEs vs. PDEs	(3) SOEs vs. PDEs	(4) MNEs vs. PDEs	(5) SOEs vs. PDEs	(6) MNEs vs. PDEs
Δ Employment, Total Effects (Incl. Pioneers)						
	40.224* (22.656)	653.650*** (136.451)	40.224 (26.109)	653.650*** (214.504)	40.224* (24.363)	653.650*** (143.792)
Borusyak et al. (2021)	0.349	0.528	0.416	0.475	0.484	0.512
Placebo	0.467	0.697	0.504	0.661	0.304	0.568
Δ Employment, Indirect Effects (Excl. Pioneers)						
	-24.112 (18.908)	248.112*** (73.717)	-24.112 (22.057)	248.112*** (90.189)	-24.112 (20.544)	248.112*** (74.992)
Borusyak et al. (2021)	0.141	0.636	0.232	0.563	0.343	0.651
Placebo	0.503	0.769	0.558	0.737	0.341	0.667
Δ Firm Counts, Total Effects (Incl. Pioneers)						
	-0.092 (0.434)	0.083 (0.491)	-0.092 (0.403)	0.083 (0.555)	-0.092 (0.389)	0.083 (0.474)
Borusyak et al. (2021)	0.943	0.633	0.916	0.662	0.905	0.437
Placebo	0.883	0.852	0.891	0.869	0.836	0.675
Δ Firm Counts, Indirect Effects (Excl. Pioneers)						
	-0.115 (0.398)	-0.010 (0.473)	-0.115 (0.376)	-0.010 (0.517)	-0.115 (0.356)	-0.010 (0.451)
Borusyak et al. (2021)	0.855	0.61	0.868	0.637	0.703	0.527
Placebo	0.703	0.834	0.757	0.856	0.663	0.79
SE cluster	Province-industry	Province-industry	Industry	Industry	Province	Province
Excluded event times	none	none	none	none	none	none
Event times used for test	-7 to -1	-7 to -1	-7 to -1	-7 to -1	-7 to -1	-7 to -1
FEs for imputation	three-way	three-way	three-way	three-way	three-way	three-way
Observations	87,295	86,634	87,295	86,634	87,295	86,634
Treated obs.	41,944	40,570	41,944	40,570	41,944	40,570
Control obs.	45,351	46,064	45,351	46,064	45,351	46,064

Notes: *** p<0.01, ** p<0.05, * p<0.1. *** p<0.01, ** p<0.05, * p<0.1. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level in Columns 1-2, at industry level in Columns 3-4, and at province level in Columns 5-6. The dependent variable varies in each panel and is defined as described in the panel heading. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021), and detailed in Equation 11. Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

Robustness Check III: Excluding Industries with Multiple Pioneering Events

Table A12: Total and Indirect Effects of Pioneering Excluding Province-Industries with Multiple Pioneering Events

	Total Effects (Incl. Pioneers)		Indirect Effects (Excl. Pioneers)	
	(1) SOEs vs. PDEs	(2) MNEs vs. PDEs	(3) SOEs vs. PDEs	(4) MNEs vs. PDEs
Δ Employment				
	52.426* (28.108)	717.560*** (153.250)	-26.032 (23.577)	267.643*** (82.752)
Δ Firm Counts				
	0.047 (0.541)	-0.171 (0.553)	-0.011 (0.497)	-0.241 (0.531)
Industries with multiple events	dropped	dropped	dropped	dropped
FEs for imputation	three-way	three-way	three-way	three-way
Observations	74,394	74,589	74,394	74,589
Treated obs.	33,692	32,934	33,692	32,934
Control obs.	40,702	41,655	40,702	41,655

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level. The dependent variable is province-industry employment (top panel) and province-industry firm count (bottom panel). In Columns 1-4, the dependent variable includes employment/firm of the pioneers, while the pioneer effects are excluded in Columns 5-8. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes. All industries experiencing more than one pioneering event are excluded from the regressions.

Robustness Check IV: Clean Assignment of Industry Pioneers

Table A13: Total and Indirect Effects of Pioneering Excluding "Hybrid" Industries with Simultaneous Pioneering of Firms with Different Ownership Structures

	Total Effects (Incl. Pioneers)		Indirect Effects (Excl. Pioneers)	
	(1) SOEs vs. PDEs	(2) MNEs vs. PDEs	(3) SOEs vs. PDEs	(4) MNEs vs. PDEs
Δ Employment				
	36.604 (26.329)	703.258*** (171.336)	-33.054 (20.953)	242.181*** (93.301)
Borusyak et al. (2021)	0.658	0.673	0.563	0.726
Placebo	0.713	0.936	0.612	0.956
Δ Firm Counts				
	-0.935** (0.477)	-0.681 (0.462)	-0.739 (0.457)	-0.662 (0.449)
Borusyak et al. (2021)	0.735	0.623	0.484	0.696
Placebo	0.488	0.808	0.244	0.847
Hybrid industries	dropped	dropped	dropped	dropped
Excluded event times	none	none	none	none
Event times used for test	-7 to -1	-7 to -1	-7 to -1	-7 to -1
FEs for imputation	three-way	three-way	three-way	three-way
Observations	85,290	84,839	85,290	84,839
Treated obs.	40,740	39,706	40,740	39,706
Control obs.	44,550	45,133	44,550	45,133

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level. The dependent variable is province-industry employment (top panel) and province-industry firm count (bottom panel). In Columns 1-2, the dependent variable includes employment/firm of the pioneers, while the pioneer effects are excluded in Columns 3-4. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes. All "hybrid" industries with simultaneous pioneering of firms with different ownership structures are excluded from the sample.

Robustness Check V: Balanced Sample Around Pioneering Event

Table A14: Total Effects of SOE-Pioneering on Employment with Differently Balanced Samples Around the Pioneering Event

	(1) SOEs vs. PDEs	(2) SOEs vs. PDEs	(3) SOEs vs. PDEs	(4) SOEs vs. PDEs
Event time 0	129.261*** (16.108)	107.078*** (21.051)	107.768*** (15.830)	117.100*** (18.558)
Event time 1	112.103*** (19.504)	100.427*** (28.110)	90.706*** (16.665)	94.831*** (19.708)
Event time 2	69.951*** (17.447)	56.592*** (21.750)	59.704*** (14.941)	57.980*** (17.730)
Event time 3	54.768*** (18.045)	56.719** (25.147)	45.351*** (15.672)	42.406** (18.410)
Event time 4	54.555*** (21.065)			46.667** (21.000)
Event time 5	38.539* (23.113)			30.459 (24.132)
Event time 6	13.655 (27.112)			12.740 (31.684)
Event time 7	10.237 (30.565)			5.522 (36.343)
Balanced event window	Not balanced	-8 to +4	-4 to +4	-4 to +8
FEs for imputation	three-way	three-way	three-way	three-way
Observations	79,284	52,864	61,249	67,537
Treated obs.	33,933	7,513	15,898	22,186
Control obs.	45,351	45,351	45,351	45,351

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at province-industry level. Observations are defined at the province-industry-year level. The dependent variable is industry-province employment including employment in pioneer firms. The table only present relative effects of pioneering by SOEs vs. PDEs. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021), and detailed in Equation 11. The sample changes in each column. In Columns 2-3, the sample is balanced around the pioneering event, such that an equal number of observations are used to calculate effects at each event time. The sample is not balanced in Column 1. Column 2 presents the estimates using a sample only including industries, where we have data at least 8 years prior to, and 4 years after, the pioneering event. In Column 3, we require data 4 years prior and 4 years after the pioneering event. In Column 4, we only use industries with data 4 years prior and 8 years after the event. Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

Table A15: Total Effects of MNE-Pioneering on Employment with Differently Balanced Samples Around the Pioneering Event

	(1) MNEs vs. PDEs	(2) MNEs vs. PDEs	(3) MNEs vs. PDEs	(4) MNEs vs. PDEs
Event time 0	304.063*** (97.880)	490.638** (204.444)	346.923*** (119.659)	200.336*** (37.492)
Event time 1	505.339*** (194.705)	791.856** (382.591)	544.200** (223.874)	300.643*** (69.308)
Event time 2	661.662*** (222.193)	943.070** (415.282)	685.178*** (245.471)	464.449*** (125.140)
Event time 3	527.564*** (102.210)	624.784*** (162.582)	521.861*** (109.943)	577.229*** (150.099)
Event time 4	604.256*** (122.488)			669.609*** (171.792)
Event time 5	733.365*** (180.617)			808.637*** (235.851)
Event time 6	885.864*** (248.071)			915.189*** (283.890)
Event time 7	1,067.844*** (366.781)			1,120.373*** (434.933)
Balanced event window	Not balanced	-8 to +4	-4 to +4	-4 to +8
FEs for imputation	three-way	three-way	three-way	three-way
Observations	79,192	53,939	61,980	67,757
Treated obs.	33,128	7,875	15,916	21,693
Control obs.	46,064	46,064	46,064	46,064

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at province-industry level. Observations are defined at the province-industry-year level. The dependent variable is industry-province employment including employment in pioneer firms. The table only present relative effects of pioneering by MNEs vs. PDEs. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021), and detailed in Equation 11. The sample changes in each column. In Columns 2-3, the sample is balanced around the pioneering event, such that an equal number of observations are used to calculate effects at each event time. The sample is not balanced in Column 1. Column 2 presents the estimates using a sample only including industries, where we have data at least 8 years prior to, and 4 years after, the pioneering event. In Column 3, we require data 4 years prior and 4 years after the pioneering event. In Column 4, we only use industries with data 4 years prior and 8 years after the event. Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

Table A16: Total Effects of SOE-Pioneering on Firm Counts with Differently Balanced Samples Around the Pioneering Event

	(1) SOEs vs. PDEs	(2) SOEs vs. PDEs	(3) SOEs vs. PDEs	(4) SOEs vs. PDEs
Event time 0	0.123 (0.141)	0.183* (0.110)	0.084 (0.154)	0.027 (0.203)
Event time 1	-0.191 (0.173)	0.004 (0.128)	-0.142 (0.170)	-0.309 (0.222)
Event time 2	-0.319* (0.189)	-0.040 (0.202)	-0.255 (0.189)	-0.498** (0.245)
Event time 3	-0.431* (0.226)	-0.172 (0.218)	-0.409* (0.220)	-0.639** (0.273)
Event time 4	-0.340 (0.356)			-0.686** (0.289)
Event time 5	-0.371 (0.368)			-0.446 (0.407)
Event time 6	-0.380 (0.517)			-0.486 (0.444)
Event time 7	-0.642 (0.665)			-0.724 (0.510)
Balanced event window	Not balanced	-8 to +4	-4 to +4	-4 to +8
FEs for imputation	three-way	three-way	three-way	three-way
Observations	79,284	52,864	61,249	67,537
Treated obs.	33,933	7,513	15,898	22,186
Control obs.	45,351	45,351	45,351	45,351

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at province-industry level. Observations are defined at the province-industry-year level. The dependent variable is industry-province firms counts including pioneer firms. The table only present relative effects of pioneering by SOEs vs. PDEs. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021), and detailed in Equation 11. The sample changes in each column. In Columns 2-3, the sample is balanced around the pioneering event, such that an equal number of observations are used to calculate effects at each event time. The sample is not balanced in Column 1. Column 2 presents the estimates using a sample only including industries, where we have data at least 8 years prior to, and 4 years after, the pioneering event. In Column 3, we require data 4 years prior and 4 years after the pioneering event. In Column 4, we only use industries with data 4 years prior and 8 years after the event. Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

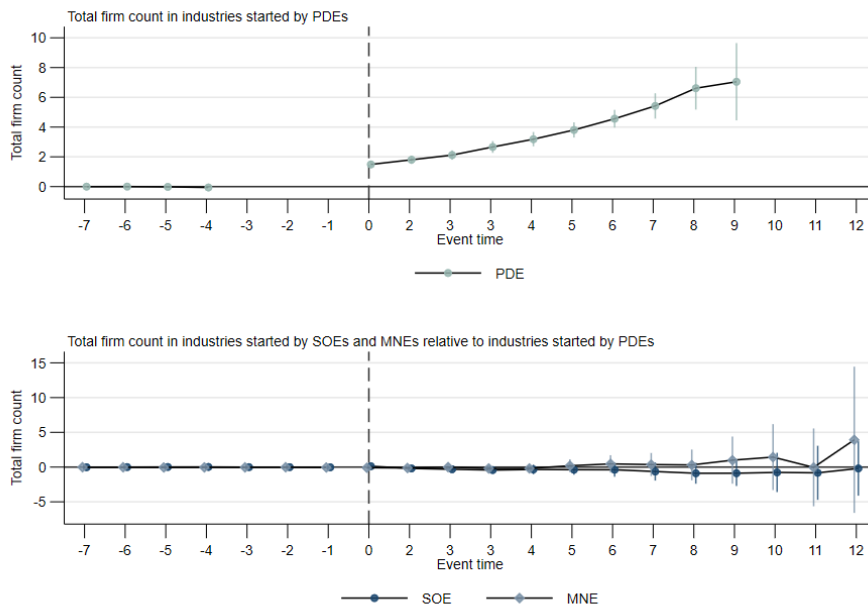
Table A17: Total Effects of MNE-Pioneering on Firm Counts with Differently Balanced Samples Around the Event Time

	(1) MNEs vs. PDEs	(2) MNEs vs. PDEs	(3) MNEs vs. PDEs	(4) MNEs vs. PDEs
Event time 0	-0.082 (0.107)	0.128* (0.069)	-0.125 (0.127)	-0.263 (0.178)
Event time 1	-0.111 (0.150)	0.236* (0.123)	-0.039 (0.158)	-0.172 (0.224)
Event time 2	0.006 (0.197)	0.393* (0.207)	0.068 (0.205)	-0.135 (0.266)
Event time 3	-0.182 (0.223)	0.270 (0.192)	-0.075 (0.232)	-0.185 (0.315)
Event time 4	-0.168 (0.286)			0.040 (0.377)
Event time 5	0.216 (0.481)			0.533 (0.620)
Event time 6	0.478 (0.648)			0.842 (0.685)
Event time 7	0.371 (0.857)			0.736 (0.858)
Balanced event window	Not balanced	-8 to +4	-4 to +4	-4 to +8
FEs for imputation	three-way	three-way	three-way	three-way
Observations	79,192	53,939	61,980	67,757
Treated obs.	33,128	7,875	15,916	21,693
Control obs.	46,064	46,064	46,064	46,064

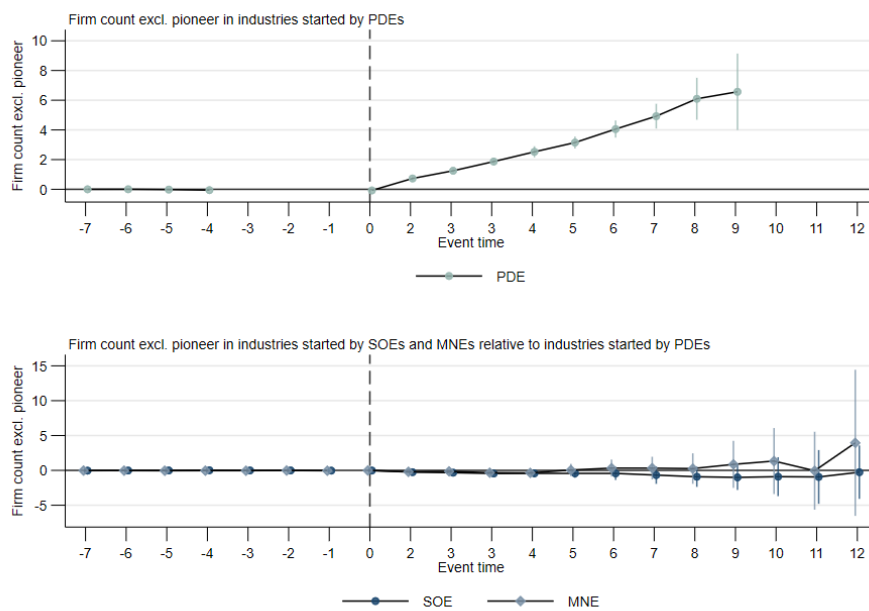
Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at province-industry level. Observations are defined at the province-industry-year level. The dependent variable is industry-province firms counts including pioneer firms. The table only present relative effects of pioneering by MNEs vs. PDEs. Results are ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021), and detailed in Equation 11. The sample changes in each column. In Columns 2-3, the sample is balanced around the pioneering event, such that an equal number of observations are used to calculate effects at each event time. The sample is not balanced in Column 1. Column 2 presents the estimates using a sample only including industries, where we have data at least 8 years prior to, and 4 years after, the pioneering event. In Column 3, we require data 4 years prior and 4 years after the pioneering event. In Column 4, we only use industries with data 4 years prior and 8 years after the event. Three-way fixed effects are used in all specifications to impute counterfactuals. Industry FEs are defined using 2-digit VSIC codes.

A.3.2. Results with Firm Counts as Dependent Variable

Figure A3: Total Effects on Firm Counts (Including Pioneers)



Notes: Authors own calculations based on data from the Vietnam Enterprise Survey. The dependent variable is province-industry firm counts including pioneer firms. Results are event time-specific ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Results in the top and bottom panels are from dynamic versions of Equation 10 and Equation 11, respectively. Counterfactuals are imputed using province-year, industry-year, and province-industry fixed effects, where industries are defined using 2-digit VSIC codes.

Figure A4: Indirect Effects on Firm Counts (Excluding Pioneers)

Notes: Authors own calculations based on data from the Vietnam Enterprise Survey. The dependent variable is province-industry firm counts excluding pioneer firms. Results are event time-specific ATTs calculated using the imputation-based estimator proposed in Borusyak et al. (2021). Results in the top and bottom panels are from dynamic versions of Equation 10 and Equation 11, respectively. Counterfactuals are imputed using province-year, industry-year, and province-industry fixed effects, where industries are defined using 2-digit VSIC codes.

Table A18: Pioneering Effects on Firm Counts – Accounting for Industry and Province

	Direct Effects (Only Pioneer)		Indirect Effects (Excl. Pioneer)	
	(1) SOEs vs. PDEs ($\hat{\beta}_m^{SOE}$)	(2) MNEs vs. PDEs ($\hat{\beta}_m^{MNE}$)	(3) SOEs vs. PDEs ($\hat{\beta}_m^{SOE}$)	(4) MNEs vs. PDEs ($\hat{\beta}_m^{MNE}$)
No Matching				
	0.031 (0.103)	0.099 (0.087)	-0.079 (0.424)	-0.133 (0.478)
Borusyak et al. (2021)	0.146	0.038	0.605	0.952
Placebo	0.134	0.024	0.789	0.943
Observations	88,320	87,520	88,320	87,520
Treated obs.	42,969	41,456	42,969	41,456
Control obs.	45,351	46,064	45,351	46,064
Matching on Industry				
	0.055 (0.093)	0.213** (0.101)	-0.335 (0.308)	0.804 (0.766)
Borusyak et al. (2021)	0.144	0.038	0.589	0.966
Placebo	0.308	0.183	0.984	0.889
Observations	86,352	81,936	86,352	81,936
Treated obs.	42,041	38,684	42,041	38,684
Control obs.	44,311	43,252	44,311	43,252
Matching on Province				
	0.048 (0.102)	0.110 (0.101)	0.029 (0.499)	-0.696 (0.571)
Borusyak et al. (2021)	0.143	0.045	0.593	0.938
Placebo	0.07	0.061	0.595	0.235
Observations	87,056	74,352	87,056	74,352
Treated obs.	42,387	35,621	42,387	35,621
Control obs.	44,669	38,731	44,669	38,731
Matching on Industry and Province				
	-0.067 (0.251)	0.330*** (0.074)	0.838* (0.489)	0.438 (0.452)
Borusyak et al. (2021)	0.103	0.732	0.28	0.418
Placebo	0.124	0.797	0.114	0.133
Observations	21,536	16,176	21,536	16,176
Treated obs.	10,783	7,698	10,783	7,698
Control obs.	10,753	8,478	10,753	8,478
Excluded event times	none	none	none	none
Event times used for test	-7 to -1	-7 to -1	-7 to -1	-7 to -1
Imputation based on:				
Year FE	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
Province FE	yes	yes	yes	yes

Notes: *** p<0.01, ** p<0.05, * p<0.1. Observations are defined at the province-industry-year level. Standard errors clustered at province-industry level. The dependent variable is province-industry firm counts. Results are ATTs calculated using the matching-based version of the imputation-based estimator (Borusyak et al., 2021) from Equation 12. Counterfactuals are imputed using year, province, and industry fixed effects, where industries are defined using 2-digit VSIC codes. The last year of the sample (2017) is excluded in all columns since proper counterfactuals cannot be calculated when all province-industries are treated.

A.3.3. OLS Regressions and Results

The OLS regressions described here are used to generate OLS estimates comparable to the imputation-based estimates reported in the main text. We report these OLS estimates below. The dynamic OLS regression described below is also the regression framework we use to test for pre-trends in the tables in the main text. The approach follows the OLS-based pre-trend test proposed in Borusyak et al. (2021), but is extended with interaction terms.

The static OLS regression is:

$$Y_{kpt} = \gamma D_{kpt} + \tau D_{kpt} SOE_{kp} + \rho D_{kpt} MNE_{kp} + \mu_{pt} + \varepsilon_{kpt}, \quad (21)$$

where Y_{kpt} denotes either total employment or firm count in industry k and province p at time t . D_{kpt} is a dummy variable that equals 1 in the year of the pioneering event and all subsequent periods. SOE_{kp} and MNE_{kp} are dummy variables indicating whether the industry in province p was started by a SOE or a MNE, respectively. γ then captures the average difference in employment/firm count in industries started by PDEs before and after they pioneer. We are particularly interested in the coefficients of the interaction terms, τ and ρ , that measure the differential effect change in outcome variables in industries pioneered by SOEs and MNEs. μ_{pt} is a province-year fixed effect.³⁹

The equivalent dynamic OLS equation is:

$$\begin{aligned} Y_{kpt} = & \sum_{\substack{h=-15 \\ h \neq -4, \dots, -1}}^{12} \gamma_h \mathbf{1}[h = E_{kp} - t] \\ & + \sum_{\substack{h=-15 \\ h \neq -4, \dots, -1}}^{12} \tau_h \mathbf{1}[h = E_{kp} - t] SOE_{kp} \\ & + \sum_{\substack{h=-15 \\ h \neq -4, \dots, -1}}^{12} \rho_h \mathbf{1}[h = E_{kp} - t] MNE_{kp} + \mu_{pt} + e_{kpt}, \end{aligned} \quad (22)$$

where τ_h and ρ_h are estimates of the differential dynamic development in Y_{ipt} between industries started by SOEs and MNEs vis-a-vis industries pioneered by PDEs. E_{kp} denotes the year of the

³⁹Since we interact year and province fixed effects and do not include industry fixed effects in the baseline OLS regression, it is strictly speaking not a TWFE regression. Yet, the issues with the estimator, as outlined in the recent literature, remain.

pioneering event and h is event time. $\mathbf{1}[\cdot]$ is an indicator function taking the value 1 if $h = E_{kp} - t$ is true. Because our data runs from 2001 to 2017, there are up to 16 pre-trend periods. Since an industry has to be absent from a province for three years before it can be pioneered, there is no pioneering from 2001-2003. There are thus 13 post-period events. Yet, we exclude all observations from 2017 since identification in this year can only be made based on "forbidden comparisons" (?): there are no not-yet-treated groups in 2017 because we only keep industry-province pairs in our sample that were treated at some point during the sample period.⁴⁰ The restriction does not prevent forbidden comparisons but will reduce the bias of these. The restriction cuts the number of post-treatment time periods to 12. By definition, the outcome is zero in the three pre-treatment event time periods so we impose $x_{-3} = x_{-2} = x_{-1} = 0$ for $x = \{\gamma, \tau, \rho\}$. Because there is no variation in Y_{kpt} at these event times, the restriction does not solve the under-identification issue in a fully dynamic event study specification (Borusyak et al., 2021). We therefore also impose $x_{-16} = x_{-4} = 0$ for identification.

In the tables and figures below, we present results from OLS estimation corresponding to the results presented in the main text using the imputation-based estimator from Borusyak et al. (2021).

⁴⁰There may in fact be some industries that die again after initial pioneering, and these could be argued to constitute valid comparison groups in 2017. In the baseline regressions, we ignore this and drop all observations from 2017.

Table A19: Total and Indirect Effects of Pioneering on Employment (OLS)

	Total Effects (Incl. Pioneer) (1)	Indirect Effects (Excl. Pioneer) (2)
PDEs (γ)	59.75*** (10.25)	34.76*** (8.697)
SOEs vs. PDEs (τ)	90.01*** (17.50)	17.34 (11.74)
MNEs vs. PDEs (ρ)	353.7*** (129.9)	4.006 (98.86)
Observations	95,120	95,120
R-squared	0.033	0.022
Province-year FE	yes	yes
Industry-year FE	yes	yes
Province-industry FE	yes	yes

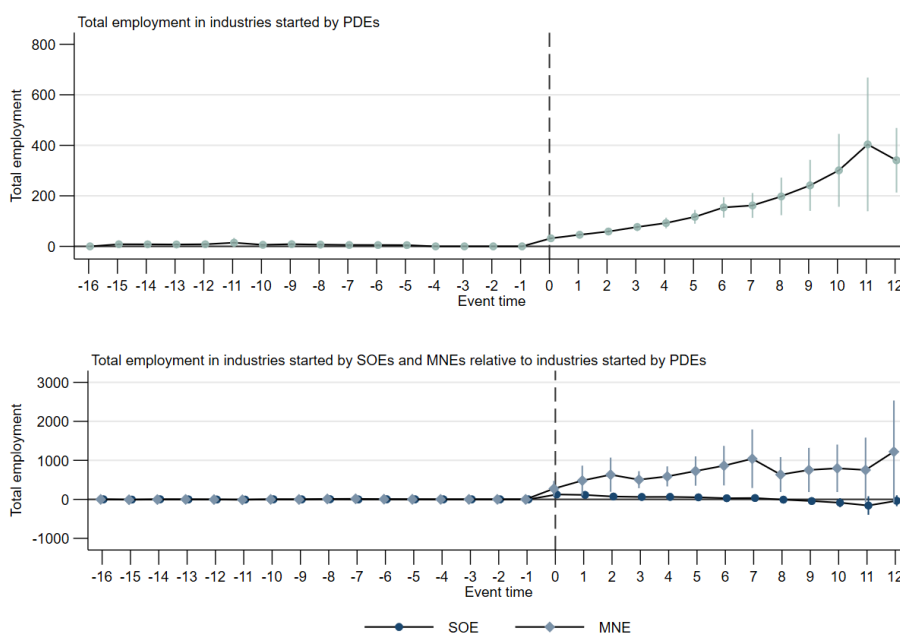
Standard errors clustered at province-industry level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations are defined at the province-industry-year level. The dependent variable is province-industry employment. Results are from OLS estimation as described in text and presented in Equation 21.

Table A20: Total and Indirect Effects of Pioneering on Firm Counts (OLS)

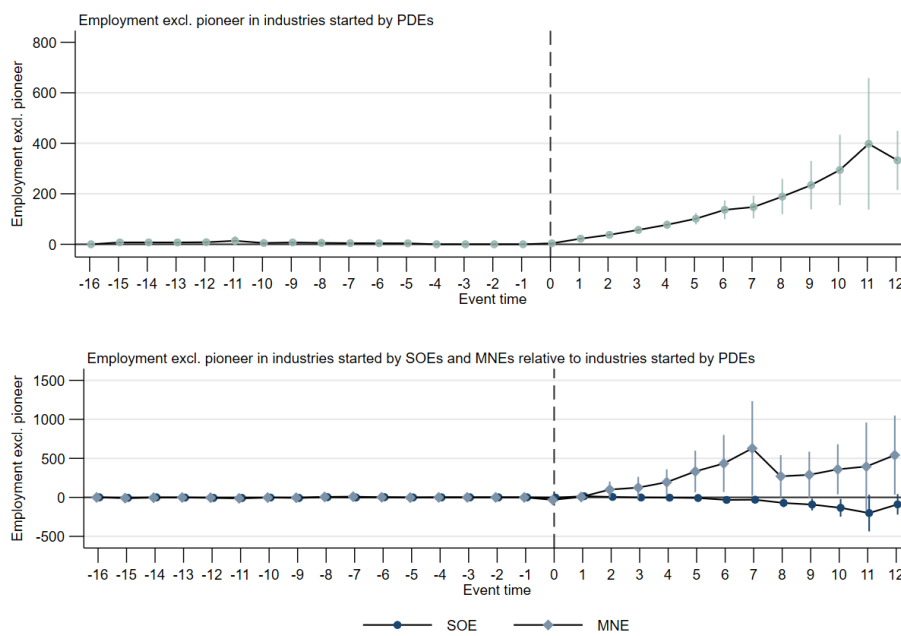
	Total Effects (Incl. Pioneer) (1)	Indirect Effects (Excl. Pioneer) (2)
PDEs (γ)	2.118*** (0.188)	1.118*** (0.153)
SOEs vs. PDEs (τ)	0.257 (0.459)	0.395 (0.365)
MNEs vs. PDEs (ρ)	0.398 (0.381)	0.234 (0.367)
Observations	95,120	95,120
R-squared	0.081	0.084
Province-year FE	yes	yes
Industry-year FE	yes	yes
Province-industry FE	yes	yes

Standard errors clustered at province-industry level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations are defined at the province-industry-year level. The dependent variable is the number of firms in an industry in a province. Results are from OLS estimation as described in text and presented in Equation 21.

Figure A5: Total Effects on Employment (Including Pioneers) - OLS

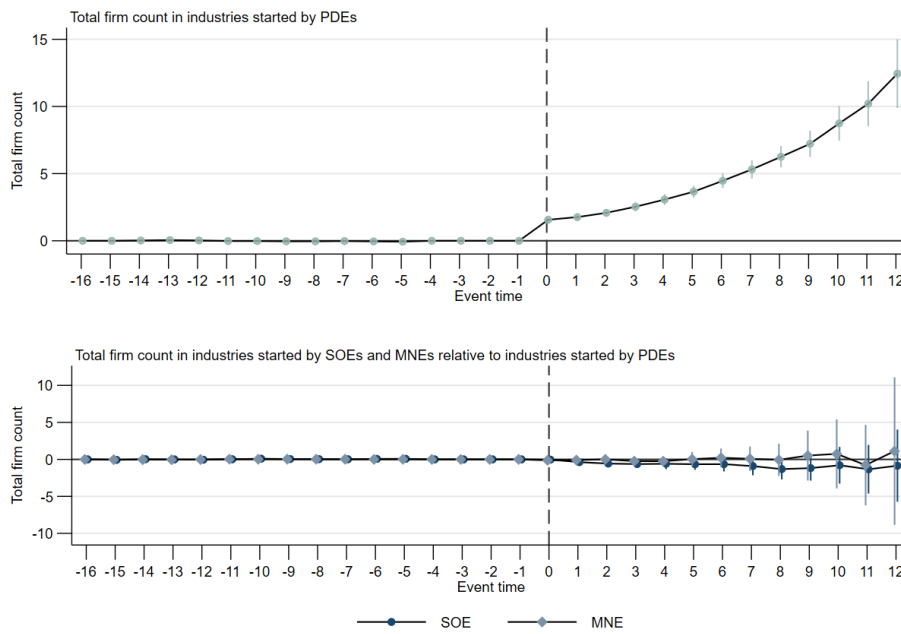


Notes: Authors own calculations based on data from the Vietnam Enterprise Survey and using the OLS estimator as described in text and presented in Equation 22. The top panel presents estimates of γ_h , while the bottom panel gives estimates of τ_h and ρ_h . The dependent variable is province-industry employment excluding including employment in pioneer firms.

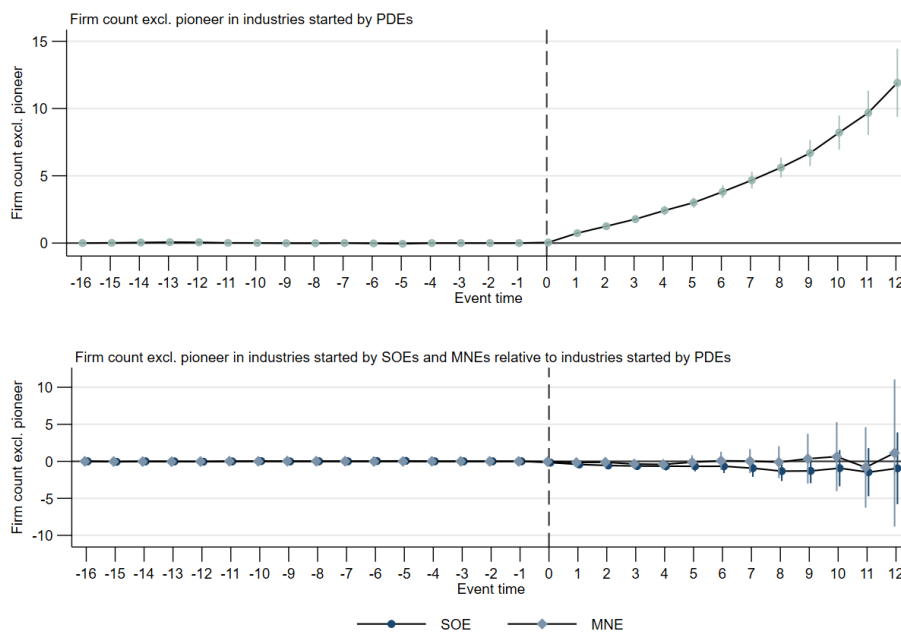
Figure A6: Indirect Effects on Employment (Excluding Pioneers) - OLS

Notes: Authors own calculations based on data from the Vietnam Enterprise Survey and using the OLS estimator as described in text and presented in Equation 22. The top panel presents estimates of γ_h , while the bottom panel gives estimates of τ_h and ρ_h . The dependent variable is province-industry employment excluding employment in pioneer firms.

Figure A7: Total Effects on Firm Counts (Including Pioneers) - OLS



Notes: Authors own calculations based on data from the Vietnam Enterprise Survey and using the OLS estimator as described in text and presented in Equation 22. The top panel presents estimates of γ_h , while the bottom panel gives estimates of τ_h and ρ_h . The dependent variable is province-industry firms counts including pioneer firms.

Figure A8: Indirect Effects on Firm Counts (Excluding Pioneers) - OLS

Notes: Authors own calculations based on data from the Vietnam Enterprise Survey and using the OLS estimator as described in text and presented in Equation 22. The top panel presents estimates of γ_h , while the bottom panel gives estimates of τ_h and ρ_h . The dependent variable is province-industry firms counts excluding pioneer firms.

A.4. Constructing the Vietnam Enterprise Panel

In this data appendix, we describe in detail the construction of the Vietnam Enterprise Panel (2001-2017) based on data collected by the General Statistics Office (GSO) in Vietnam. First, we describe and discuss the sampling methodology and representativeness of the data. Second, we describe how we generate consistent firm IDs over time to link yearly data sets into a panel. Third, we describe in detail how we have cleaned the panel, providing statistics on yearly firm counts through all steps of the cleaning procedure. Specifically, we describe how we clean geo-information, industry codes, and firm categories to ensure consistency over time. We also describe how we clean core economic variables from firms' income statements, balance sheets, and statements on labour compositions. Throughout, the data appendix will include descriptive statistics on the (continuously decreasing) count of firms in each step of the cleaning process.

A.4.1. Sampling and Coverage

The Vietnam Enterprise Survey (VES) data contain information on enterprises that are registered under Vietnam's Enterprise Law. All SOEs, MNEs, and collectives are legally required to register under the Enterprise Law. Private businesses that have more than 10 employees or operate in more than one location are also required to register under the law in order to legally operate. Smaller businesses may choose to register either as a household business or an enterprise (Government of Vietnam, 2000, 2004).

The VES data is constructed through a combination of censuses and surveys. In census years, all firms are surveyed. In other years, 100 percent of a subset of firms are surveyed and 10-20 percent of the remaining firms are sampled. Table A21 provides a detailed description of the sampling procedure for each year. From Table A21, it can be seen that the VES data covers the universe of registered enterprises from 2001-2004 and in the census years of 2007, 2012, and 2017. In all other years, GSO also surveyed the universe of SOEs, MNEs, and private enterprises with more than 10 (2005-2010), 20 (2011-2015), or 100 (2016-2017) employees. In 2001 (2017), these enterprises accounted for around 55.8 (5.2) percent of all firms, 97.4 (66.1) percent of total employment, and they operated in 98.2 (92.5) percent of all industries. In addition, all enterprises within a time-varying selection of provinces and sectors are surveyed each year, along with 15-20

percent of smaller private enterprises falling below the employment thresholds.

Table A21: VES Sampling Procedure

Year	Census (all enterprises surveyed)	Survey (sample of enterprises surveyed)
2001	All enterprises	N/A
2002	All enterprises	N/A
2003	All enterprises	N/A
2004	All enterprises	N/A
2005	Ownership type: SOEs, MNEs, PDEs with 10+ employees Province: Hanoi, Haiphong, Hatay, Quang Ninh, Thanh Hoa, Nghe An, Thua Thien-Hue, Da Nang, Binh Dinh, Khanh Hoa, HCM City, Binh Duong, Dong Nai, Ba Ria-Vung Tau, Long An, Dong Thap, An Giang, Ten Giang, Ben Tre, Kien Giang, Can Tho, Ca Mau.	20 pct. of PDEs with 0-9 employees
2006	Ownership type: SOEs, MNEs, PDEs with 10+ employees Province: Hanoi, Haiphong, Hatay, Bac Ninh, Hai Duong, Nam Dinh, Thai Binh Ninh Binh, Thai Nguyen, Phu Tho, Bac Giang, Quang Ninh, Thanh Hoa, Nghe An, Quang Binh, Thua Da Nang, Quang Nam, Quang Ngai, thien-Hue, Binh Dinh Khanh Hoa, Gia Lai, Dac Lac, Lam Dong, Tp, HCM, Binh Duong, Dong Nai, Binh Thuan, Tay Ninh, Ba Ria-Vung Tau, Long An, Dong Thap, An Giang, Tien Giang, Vinh Long, Ben Tre, Kien Giang, Can Tho, Soc Trang and Ca Mau. Sector: Hotels	20 pct. of PDEs with 0-9 employees
2007	All enterprises	N/A
2008	Ownership type: SOEs, MNEs, PDEs with 10+ employees (Hanoi: PDEs with 20+ employees; Ho Chi Minh City: PDEs with 30+ employees). Province: Lai Chau, Ha Giang, Dien Bien, Bac Can, Son La, Cao Bang, Tyen Quang, Lao Cai, Yen Bai, Ha Nam, Hoa Binh, Ninh Thuan, Kon Tum, Đk Nông and Hu Giang. Sector: Agriculture, forestry, fishery, hotels and restaurant, finance, credits, and insurance.	15 pct. of PDEs with 0-9 employees (Hanoi: 10 pct. of PDEs with 0-9 employees; Ho Chi Minh City: 10 pct. of PDEs with 0-9 employees.)
2009	Ownership type: SOEs (≥ 50 pct. state capital), MNEs, PDEs with 10+ employees (Hanoi: PDEs with 20+ employees; Ho Chi Minh City: PDEs with 30+ employees).	15 pct. of PDEs with 0-9 employees (Hanoi: 15 pct. of PDEs with 0-19 employees; Ho Chi Minh City: 10 pct. of PDEs with 0-29 employees).

Continued on next page

Table A21 – Continued from previous page

Year	Census (all enterprises surveyed)	Survey (sample of enterprises surveyed)
	<p>Province: Hà Nam, Ninh Bình, Hà Giang, Cao Bng, Lào Cai, Bc Kn, Lng Sn, Tuyên Quang, Yên Bái, Lai Châu, Đin Biên, Sn La, Hoà Bình, Qung Tr, Phú Yên, Kon Tum, Gia Lai, Đk Lk, Đk Nông, Ninh Thun, Bình Phc, Hu Giang, Trà Vinh, Bc Liêu.</p> <p>Sector: Agriculture, forestry, fishery, hotels and restaurant, information and communication.</p>	
2010	<p>Ownership type: SOEs (≥ 50 pct. state capital), MNEs, PDEs with 10+ employees (Hanoi and Ho CHI Minh City: PDEs with 30+ employees).</p> <p>Province: Hà Giang, Cao Bng, Bc Kn, Tuyên Quang, Lào Cai, Đin Biên, Lai Châu, Sn La, Yên Bái, Hoà Bình, Lng Sn, Phú Yên, Ninh Thun, Kon Tum, Đk Nông, Trà Vinh, Hu Giang, Sóc Trăng, Bc Liêu.</p> <p>Sector: Agriculture, forestry, fishery, transportation, hotels and restaurant, tourism, finance, credits and insurance.</p>	15 pct. of PDEs with 0-9 employees (Hanoi and Ho Chi Minh City: 15 pct. of PDEs with 0-29 employees).
2011	<p>Ownership type: SOEs (≥ 50 pct. state capital), MNEs, PDEs with 20+ employees (Hanoi and Ho Chi Minh City: PDEs with 50+ employees; Hai Phong, Dong Nai and Binh Duong: PDEs with 30+ employees).</p> <p>Province: Hà Giang, Cao Bng, Bc Kn, Tuyên Quang, Lào Cai, Đin Biên, Lai Châu, Sn La, Yên Bái, Lng Sn, Ninh Thun, Kon Tum, Đk Nông, Trà Vinh, Hu Giang and Bc Liêu.</p> <p>Sector: Agriculture, forestry, fishery, hotels and restaurant, information and communication.</p>	15 pct. of PDEs with 0-19 employees (Hanoi and Ho Chi Minh City: 10 pct. of PDEs with 0-19 employees and 20 pct. of PDEs with 20-49 employees; Hai Phong, Dong Nai and Binh Duong: 20 pct. of PDEs with 0-29 employees).
2012	All enterprises	N/A
2013	<p>Ownership type: SOEs (≥ 50 pct. state capital), state-general corporations, MNEs, cooperatives, union of cooperatives, people's credit funds, PDEs with 20+ employees (Hanoi and Ho Chi Minh City: PDEs with 50+ employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: PDEs with 30+ employees).</p> <p>Province: Hà Giang, Hu Giang, Bc Liêu, Kon Tum, Tuyên Quang, Cao Bng, Lng Sn, Sn La, Đin Biên, Lai Châu, Đk Nông, Bc Kn.</p> <p>Sector: Hotels and restaurant, information and communication, transportation, warehouses and real-estate.</p>	20 pct. of PDEs with 0-19 employees (Hanoi and Ho Chi Minh City: 10 pct. of PDEs with 0-19 employees and 20 pct. of PDEs with 20-49 employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: 20 pct. of PDEs with 0-29 employees).

Continued on next page

Table A21 – Continued from previous page

Year	Census (all enterprises surveyed)	Survey (sample of enterprises surveyed)
2014	<p>Ownership type: SOEs (≥ 50 pct. state capital), PDEs with 20+ employees (Hanoi and Ho Chi Minh City: PDEs with 100+ employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: PDEs with 50+ employees).</p> <p>Province: Hà Giang, Cao Bng, Bc Kn, Tuyên Quang, Lào Cai, Đin Biên, Lai Châu, Sn La, Yên Bái, Lng Sn, Ninh Thun, Kon Tum, Đk Nông, Trà Vinh, Hu Giang and Bc Liêu</p> <p>Sector: Hotels and restaurant, information and communication, transportation (aviation, marine and rail-way transportation), securities, solid waste collection and treatment.</p>	<p>20 pct. of PDEs with 0-19 employees (Hanoi and Ho Chi Minh City: 10 pct. of PDEs with 0-49 employees and 20 pct. of PDEs with 50-99 employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: 20 pct. of PDEs with 0-49 employees).</p>
2015	<p>Ownership type: SOEs (≥ 50 pct. state capital), PDEs with 20+ employees (Hanoi and Ho Chi Minh City: PDEs with 100+ employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: PDEs with 50+ employees).</p> <p>Province: Hà Giang, Cao Bng, Bc Kn, Tuyên Quang, Lào Cai, Đin Biên, Lai Châu, Sn La, Yên Bái, Lng Sn, Ninh Thun, Kon Tum, Đk Nông, Trà Vinh, Hu Giang and Bc Liêu.</p> <p>Sector:Hotels and restaurant, information and communication, transportation (aviation, marine and rail-way transportation), securities.</p>	<p>20 pct. of PDEs with 0-19 employees (Hanoi and Ho Chi Minh City: 10 pct. of PDEs with 0-49 employees and 20 pct. of PDEs with 50-99 employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: 20 pct. of PDEs with 0-49 employees).</p>
2016	<p>Ownership type: SOEs (≥ 50 pct. state capital), enterprises with 2+ branches where at least one is located in a provincial city with a different address from that of the headquarter's, PDEs with 100+ employees.</p> <p>Province: Hà Giang, Cao Bng, Bc Kn, Tuyên Quang, Lào Cai, Đin Biên, Lai Châu, Sn La, Yên Bái, Lng Sn, Ninh Thun, Kon Tum, Đk Nông, Trà Vinh, Hu Giang and Bc Liêu.</p> <p>Sector: Agriculture, forestry, fishery, hotels and accommodation, transportation (aviation, marine and rail-way transportation), securities, insurance.</p>	<p>50 pct. of PDEs with 50-99 employees (Hanoi and Ho Chi Minh City: 20 pct. of PDEs with 50-99 employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: 30 pct. of PDEs with 50-99 employees);</p> <p>20 pct. of PDEs with 10-49 employees operating in manufacturing, construction, insurance, finance (Hai Phong, Da Nang, Dong Nai and Binh Duong: 15 pct.); 15 pct. of PDEs with 10-49 employees operating in trade and service sectors (except insurance and finance) (Hai Phong, Da Nang, Dong Nai and Binh Duong: 10 pct.). In Hanoi and Ho Chi Minh City: 10 pct. of all PDEs with 10-49 employees irrespective of their main economic activity.</p> <p>10 pct. of PDEs with 0-9 employees (Hanoi and Ho Chi Minh City: 3 pct. of PDEs with 0-9 employees; Hai Phong, Da Nang, Dong Nai and Binh Duong: 7 pct. of PDEs with 0-49 employees).</p>
2017	All enterprises	N/A

A.4.2. Generating Firm IDs to Create a Panel

The data was made available separately for each year, and a common identifier had to be used to uniquely identify enterprises over time. For the years 2001 to 2015, this unique identifier is based on a common province specific identifier (uniquely identifying firms within a province - but not across) and on the respective province code. From 2016 on, it was replaced by an identifier that is based on the tax registration number. 2015 contains both the tax registration number and the province-based identifier, and thus serves as a bridge year in the data, allowing us to match those two periods. For firms that existed before 2016, the initial province-based identifier was extended to the later period. For enterprises that did not exist before 2016, new identifiers consistent with the province-based ones were assigned.

Throughout the years, there are instances where the required identifying information is missing, incomplete, or inconsistent (134,763 observations in total), which we dropped from data. In other instances, firms could not be uniquely define based on the province identifier. Some of those instances were exact duplicates, and were thus removed from the data (43,339 observations in total). Other instances turned out to be distinguishable using both the information from the province identifier and the tax registration number, and which cases we did not drop them, if possible. The identified panel, before any further data cleaning, consists of 4,279,668 observations and 942,544 unique firms.

Table A22: Firm Count

	All firms	With province	With ownership type	With sector	With employees	With revenue	With intermediaries
2001	31860	31847	31847	30862	29828	27516	27074
2002	41163	41150	41150	39925	38891	37610	36507
2003	55144	55131	55131	53510	52467	50232	47432
2004	81377	81360	81360	79060	77985	73124	68889
2005	100709	100554	100554	97512	96532	90992	86384
2006	107836	107836	107836	104162	103332	100896	97008
2007	140074	139919	139919	134337	133526	130309	124060
2008	188886	188730	188729	181390	180648	176888	167746
2009	212637	212637	212634	204564	203874	194078	170351
2010	290675	290675	290674	280917	280214	255050	220341
2011	338857	338737	338734	330498	328763	307209	260680
2012	358620	358504	358501	350859	350220	320604	278086
2013	385779	385664	385662	378550	377926	332100	270140
2014	412722	412607	412605	405968	405324	351279	295642
2015	445078	444971	444969	439005	438421	376522	298905
2016	517598	517452	517450	512002	511494	437654	332681
2017	570653	570508	570506	565312	564827	482660	412558
Total	4279668	4278282	4278261	4188433	4174272	3744723	3194484

In the following sections, we explain how the key variables in the panel are cleaned. Table A22 presents yearly firm counts for each step in the cleaning procedure. In its raw state, the VES data has 4,28 million firm-year observations from 2001-2017. Our cleaning reduces this number 3,19 million observations.

A.4.3. Geo-Codes

Vietnam has 63 provinces, indicated by province codes ranging from 1 to 99. The information on provinces in the identified data is complete, and thus no observation had to be dropped due to missing province codes. We did drop though all observations with province codes 97, 98, and 99, as those are military bases and not provinces. Together, they consist of 1,386 observations over the whole period. In 2008, a reform merged Hà Tây province (province code 28) into Hanoi province (province code 1). For consistency reasons, we change all firms located in Hà Tây prior to its dismantling, to being located in Hanoi. This affects 6,757 firm-year observations. 343 observations were moreover coded in the wrong province (province code 26), as could be seen from the district they are located in, and were corrected to be located in Hanoi (province code 1).

The VES data also contains information on districts, which is the next smaller administrative geographical unit. Vietnam has more than 700 districts as well as towns and cities with the same legal status as districts. The exact number varies over time, as districts are subject different administrative reforms. 554,161 observations have missing district information. The district codes for 2001-2003, moreover, are not consistent with the later years. Instead of dropping firms with missing or inconsistent district information, we impute the districts of firms that we observe in other years where the information is not missing. The same applies for a smaller number of inconsistent district codes in 2007-2009. By this procedure, we impute 174,471 district codes.

A.4.4. Ownership Type

The different ownership types available in GSO's questionnaires changed frequently over time, which made it necessary to create consistent ownership categories over time. Table A23 provides an overview over the available categories in the different years, where the rightmost column shows the uniform categorization that we made them consistent to. In this consistent categorization, we end up

with 13 different categories, which is fairly close to the number of different ownership types GSO provides throughout the studied period. In most cases, the wording or formulation of ownership types changes at least once, often from 2011 to 2012. The wording was, however, mostly similar enough that it did not pose a problem to match the correct ownership types. In other cases, where the wording changes more substantially or more often, the trends in numbers of certain categories gave most of the times a clear indication. In 2001, the ownership type 10 (JSC w. State Capital) is used for both consistent ownership types 12 (JSC w. $> 50\%$ State Capital) and 13 (JSC w. $\leq 50\%$ State Capital), as there is a complementary question, asking for the amount of state capital. The ownership type 8 (Ltd. Liability Company with $\leq 50\%$ state capital), on the other hand, is classified solely into the consistent ownership category 6 (LLC Private), as the accompanying question about the amount of state capital is not consistently available over all years. Four observations had no ownership type recorded, and were dropped together with all their occurrences in other years (23 observations in total). In further analyses, these categories can be aggregated into broader categories such as broader definitions of foreign or state-ownership, depending on the research question.

A.4.5. Industry Codes

The economic activity of firms is reported in accordance with Vietnamese Standard Industrial Classification (VSIC). VSIC codes loosely follow the International Standard Industrial Classification system.

Creating consistent VSIC codes over time: The VSIC classification system was originally created in 1993. It was updated in 2007. In the remainder of this section, we will refer to the "old" codes as VSIC93 and the "new" codes as VSIC07. From 2001-2006, firms only used VSIC93 to report their activities; from 2006-2010, they both reported VSIC93 and VSIC07 codes; and since 2010 they only reported VSIC07 codes. The VSIC93 codes are consistently available at the 4-digit level from 2001-2010, whereas the VSIC07 codes are available at the 5-digit level. We use descriptions of the industry codes, obtained from GSO's website, to manually create a concordance table between VSIC93 and VSIC07.⁴¹ Because the VSIC07 coding system is more granular than the VSIC93 system, we have to aggregate a number of sectors to achieve a concise mapping between

⁴¹Due to an update of GSO's website, the descriptions are no longer available online.

Table A23: Ownership Type by Year

Ownership Type	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Consistent
SOE																		1
Central SOE	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	
Local SOE	2	2	2	2	2	2	2	2	2	2	2	-	-	-	-	-	-	
State Company	-	-	-	-	-	-	-	-	-	-	-	4	4	4	4	4	4	
Collective/Cooperative																		2
Collective Enterprise	3	3	6	6	6	6	6	6	6	6	6	-	-	-	-	-	-	
Cooperative (Union)	-	-	-	-	-	-	-	-	-	-	-	5	5	5	5	5	5	
Private Enterprise																		3
Private Enterprise	4	4	7	7	7	7	7	7	7	7	7	-	-	-	-	-	-	
Private Business	-	-	-	-	-	-	-	-	-	-	-	6	6	6	6	6	6	
Partnership Company																		4
LLC with Central State																		5
State Limited Company	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Central State Limited Company	-	6	-	6	6	6	6	6	6	6	6	-	-	-	-	-	-	
State Limited Company with 1 member	-	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
1-Member Liability with 100% central state capital	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	
LLC Private																		6
Private Ltd. with more than 2 members	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Private Limited Company	7	8	9	9	9	9	9	9	9	9	9	-	-	-	-	-	-	
Ltd Liability Company with ≤ 50% state capital	-	-	-	-	-	-	-	-	-	-	-	8	8	8	8	8	8	
JSC w/o State Capital																		7
Foreign-Owned Company																		8
Wholly Foreign-Owned Company	12	12	12	12	12	12	12	12	12	12	12	-	-	-	-	-	-	
Enterprise with 100% Foreign Capital	-	-	-	-	-	-	-	-	-	-	-	11	11	11	11	11	11	
Joint-venture w. State																		9
Joint-Venture (SOE and Foreign Partner)	13	13	13	13	13	13	13	13	13	13	13	-	-	-	-	-	-	
Joint-Venture with Foreign Capital	-	-	-	-	-	-	-	-	-	-	-	12	12	12	12	12	12	
Joint-venture w. Private																		10
JV (Private/Nonstate and Foreign Partners)	14	14	14	14	14	14	14	14	14	14	14	-	-	-	-	-	-	
Other Types with Foreign Capital	-	-	-	-	-	-	-	-	-	-	-	13	13	13	13	13	13	
LLC with Local State																		11
State Ltd. Company with more than 2 members	8	-	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Local State Ltd. Company	-	7	-	4	4	4	4	4	4	4	4	-	-	-	-	-	-	
1-Member Liability with 100% local state capital	-	-	-	-	-	-	-	-	-	-	-	2	2	2	2	2	2	
JSC w. > 50% State Capital																		12
JSC w. State Capital	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
JSC w. > 50% State Capital	-	9	5	5	5	5	5	5	5	5	5	-	-	-	-	-	-	
SC/LLC w. > 50% State Capital	-	-	-	-	-	-	-	-	-	-	-	3	3	3	3	3	3	
JSC w. ≤ 50% State Capital																		13
JSC w. State Capital	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
JSC w. ≤ 50% State Capital	-	10	11	11	11	11	11	11	11	11	11	-	-	-	-	-	-	
SC/LLC w. ≤ 50% State Capital	-	-	-	-	-	-	-	-	-	-	-	8	8	8	8	8	8	

the codes. In a few instances, we also have to aggregate VSIC93 codes too. The result is a mixed 4-digit VSIC coding system with 288 unique industry codes of which there were Vietnamese firms operating in 281 in 2017. Henceforth, we shall simply refer to this coding system as "VSIC".⁴²

Some firms report inconsistent VSIC93 and VSIC07 codes in years where they report both. Some firms also report industry codes that do not exist in the official classification system. To deal with these inconsistencies, we wrote a program that assigns firms to industries based on the following rule outlined in bullets below. In the program, we apply the assumptions that inconsistencies at the 4-digit level does not mean a code is misspecified at the 3-digit level. We do require, however, that industry codes are consistent at the 2-digit level as a minimum.

- **Step 1:** Drop firms with non-existing or inconsistent 2-digit industry codes:
 - For 2001-2005, where firms only report VSIC93 codes: Drop firms whose 2-digit industry codes do not exist in the official VSIC93 classification system.
 - For 2011-2017, where firms only report VSIC07 codes: Drop firms whose 2-digit industry codes do not exist in the official VSIC07 classification system.
 - For 2006-2010, where firms report both VSIC93 and VSIC07 codes: Drop firms whose 2-digit VSIC93 and VSIC07 are inconsistent.
- **Step 2:** Determine the minimum number of digits in each VSIC93 code (2-4) needed to map a firm to a unique 4-digit VSIC code, given a specific number of digits in the corresponding VSIC07 code (2-5).
- **Step 3:** Assign firms to a unique VSIC code based on their reported VSIC93 and/or VSIC07 code(s):
 - **Round 1:**
 - * If firms report both VSIC93 and VSIC07 codes: Assign a firm to a 4-digit VSIC code if the 5-digit VSIC07 code is consistent with the VSIC93 code at the 2, 3, or 4-digit level, provided that the digit-combination of VSIC93 and VSIC07 is adequate to identify a unique VSIC code.
 - * If firms only report either VSIC93 or VSIC07 codes: When the VSIC93 (VSIC07) code is missing, then use the VSIC07 (VSIC93) at the highest possible level of

⁴²The concordance table is available from the authors upon request.

aggregation (up to 2-digits) to assign the firm a unique 4-digit VSIC code.

– **Round 2:** If firms report both VSIC93 and VSIC07 codes and they have not been assigned a VSIC code yet: Repeat the following procedure for 4, 3, and 2-digit VSIC07 codes, respectively:

* Assign a firm to a 4-digit VSIC code if the VSIC07 code is consistent with the VSIC93 code at the 2, 3, or 4-digit level, provided that the digit-combination of VSIC93 and VSIC07 is adequate to identify a unique VSIC code.

• **Step 4:** If possible, assign 2-digit VSIC codes to firms that have not been assigned 4-digit VSIC codes. There are situations, where we cannot identify the 4-digit VSIC code of a firm due to missing or inconsistent information. In some instances, however, it is still possible to assign the firm a 2-digit VSIC code. This is because the digit-combination for which VSIC93 and VSIC07 is consistent, may not be enough to link the firm to a unique 4-digit VSIC code. It may, however, be enough to link the firm to a 2-digit code.

We drop 89,828 firms with missing 2-digit VSIC code (see Table A22, Column 4).

Multi-industry information: The VES data has information on firms' employment and revenue across different industries. Table A24 shows the years where multi-industry information is available. The top six rows of the table show the years for which the main activity and other activities of firms are available in the VES data with VSIC93 and VSIC07 codes, respectively. Tee numbers in the table indicate the number of industry codes firms are allowed to report within each category. In most years, for instance, firms are allowed to report operations in up to four economic activities other than their main activity. The three bottom rows of the table illustrate how we combine the VES information into time-consistent VSIC codes.

We clean the multi-industry information using the same procedure as explained in section above. We save the firm-industry level data in a separate file, that is available to be merged back into the firm-level panel based on firms' id number and year using one-to-many matching.

Table A24: Multi-Industry Information

Industry codes	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	
Main activity (VSIC93)	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	
Other activities, employment*	7	4	4	4	4	4	4	4	4	4	-	-	-	-	-	-	-	
Other activities, revenue	-	4	4	4	4	4	4	4	4	4	-	-	-	-	-	-	-	
Main activity (VSIC07)	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1
Other activities, employment	-	-	-	-	-	4	4	4	4	4	4	4	4	4	13	11	11	11
Other activities, revenue	-	-	-	-	-	4	4	4	4	4	4	4	4	4	13	11	11	11
Main activity (VSIC)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Other activities, employment	7	4	4	4	4	4	4	4	4	4	4	4	4	4	13	11	11	11
Other activities, revenue	-	4	4	4	4	4	4	4	4	4	4	4	4	4	13	11	11	11

Notes: The top six rows of this table show the years for which the main activity and other activities of firms are available in the VES data with VSIC93 and VSIC07 codes. Numbers indicate the number of industry codes firms are allowed to report within each category. The three bottom rows show how we combine this information into time-consistent VSIC codes.

* Employment data is available for both beginning and end of the year from 2001-2010.

A.4.6. Employment

For the years 2001-2010, employment data is available for beginning and end of the year, the latter being better populated though. Employment at the beginning of the year contains 98,240 missing values, while at the end of the year only of 629. In addition, there are nine companies that report negative number of employees both at the beginning and at the end of the year, which we replace with missing values. 189,782 observations report moreover zero employment at the beginning of the year, 3,779 do so at the end. For the years 2002-2006, the data further contains information on employees hired and fired throughout a year, which we use to populate some of the missing employment values: If a firm has missing employment at the beginning of the year, non-missing employment at the end of the year, missing fired, and non-missing hired, we assume this is a new firm, and replace employment at the beginning of the year with zero (14,375 observations). Similarly, we replace 888 missing or zero employment at the beginning of the year by subtracting the net turnover (hired - fired) from non-missing end of the year employment. The other way around, we replaced two missing end of the year employment values with non-missing beginning of the year employment added to the net turnover.

For observations where we don't have employment turnover, we populate missing employment data by exploiting information from the previous or next year. The reason is that employment at the *end* of year t should be the same (or reasonably close) to employment at the *beginning* of year $t+1$. We thus replace missing or zero employment at the beginning of the year by employment at the end of the previous year. Similarly, we replace missing or zero employment at the end of a year by employment at the beginning of the previous year. By that, we are able to populate 7,373 values for beginning of the year employment, and 256 for end of the year. For observations where there is no information from the previous year (because the firm is new or the information is missing) and no turnover information, we replace missing or beginning of the year values with end of the year values of that same year. We proceed accordingly for missing end of the year employment, where information from the next year is not available. Thus allows us to populate another 75,430 observations at the beginning, and 264 at the end of the year.

After cleaning employment at the beginning and end of the year, we create one trusted and consistent employment variable for each year. To be consistent with other variables that are only

recorded for the end of the year, end of the year employment is the number we are naturally more interested in. If end of the year employment in year t is equal to employment at the beginning of year $t+1$, employment at the end of year is used as the overall employment for year t . For a large number of observations (2,301,926), however, those two are not exactly equal (or one of them is missing). For those cases, we calculate average province-year-2-digit-industry labor compensation, which is the total wage (wage + contributions) per employee within a province in a 2-digit industry in a year, if there are at least five firms present. The overall employment value for year t is then end of the year employment of year t if a firm's labor compensation is closer to the province-year-2-digit-industry average than the firm's labor compensation in $t+1$, and vice versa. If that approach is not possible, because end of year t employment and beginning of year $t+1$ employment are not available, beginning of year t employment provide the value for overall employment. After this procedure, there are 117 observations with missing and 2,177 with zero employment, corresponding to 14,161 firm-year observations. Those are removed from the data.

A.4.7. Revenue and Profit Variables

In this section, we describe how we clean the revenue and profit variables in the VES data. After the cleaning procedure described below, there are 3,744,723 firm-year observations in the data with revenue information (see Table A22, Column 6). We do not drop firm with missing data from the panel, but let researchers decide whether to drop these observations in their specific projects.

Table A25 displays the yearly availability of variables from firms' profit-loss statements. Most profit-loss items are directly reported, but it is also possible to calculate them from 2002-2017. We can thus check the reliability of a firms' reported variables by checking whether the math adds up. We can also use the accounting rules (displayed in Table A25) to populate missing values. We construct two programs to help with this accounting exercise.

First, we write a "data recovery"-program, which uses the accounting rules to populate missing items. Items may either be missing because they are not reported in a year, or due to incomplete data. For instance, we can calculate *deductions* in 2001 by subtracting *net revenue from production and services (PS)* from *gross revenue from PS*. Likewise, if *net revenue from PS* is missing in, say 2005, we can populate it by subtracting *deductions* from *gross revenue form PS*. The program conducts a

plethora of these calculation. We loop over the program because the continuous population of items may make it possible to extract further information from calculations conducted in a previous step. The loop continuous until no more information can be directly calculated from the data.

Second, we write a "consistency check"-program. The program generates consistent (or trusted) profit-loss items by comparing the reported items to calculated items and choosing between them, when they deviate. We apply the program at each step of the profit-loss statement shown in Table A25, starting at the top. As a working example, consider the first step, going from *gross revenue from PS* to *net revenue from PS* through the following relationship: $Net\ revenue\ from\ PS = Gross\ revenue\ from\ PS - Deductions$. The program creates a consistent *net revenue from PS* item in few steps:

- **Step 1:** Flag firms if the calculated *net revenue form PS* deviates from the reported or if any of *gross revenue form PS*, *deductions*, or *net revenue form PS* are missing. The flags will be used later.
- **Step 2:** Recover missing information, if one of the three items are missing.
- **Step 3:** Re-calculate *net revenue from PS* using any additional information from Step 2.
- **Step 4:** For each firm, generate a new "consistent" *net revenue from PS* by deciding between the reported and calculated value when they differ. Make the choice based on the following decision rule:
 - **If:** the consistent variable cannot, by definition, be higher than some specific item (for instance, *other net profit* cannot, by definition, be higher than *other revenue*. Note that this is not the case for our working example, where *deductions* may be positive):
 - * If only one of the calculated and reported values are smaller than the specific item, then use the one that is smaller (provided that the value for the consistent variable is missing).
 - * If both the calculated and the reported value are smaller than the specific item, then use the one that is closest to the 2-digit industry-province-year average, provided that there are at least 5 observations within the cell (provided that the value for the consistent variable is missing).
 - **Else:**

- * Use the value that is closest to the 2-digit industry-province-year average, provided that there are at least 5 observations within the cell.
 - * Use the reported value if there are less than 5 observations within the each 2-digit industry-province-year cell (provided that the value for the consistent variable is missing).
 - * Use the calculated (and thereafter the reported) value, if the value for the consistent variable is missing.
- **Step 5:** Re-adjust the "intermediate" value (*deductions*) if necessary, so that the relationship between *gross revenue from PS*, *deductions* and *net revenue from PS* is restored.⁴³

We use the data recovery-program and the consistency check-program to clean the profit-loss section of the VES data. First, we apply the data recovery-program to the uncleaned data.⁴⁴ Second, we apply the consistency check-program at each step of the profit-loss statement shown in Table A25, starting at the top. We then use the new consistent variable as a starting point for next step. For example, when we have generated a consistent *net profit from PS*, we use this as the starting point for calculating *gross profit form PS*, through the relationship $gross\ profit\ from\ PS = consistent\ net\ revenue\ from\ PS - cost\ of\ sales$. Third, we re-run the data recovery-program on the consistent variables. Finally, we generate three flags that indicate inconsistent profit-loss data:

- **Flag 1:** We flag firms if they have too many deviations between reported and calculated items in their profit-loss statement. The flag is created by counting the number of individual flags per firms, generated in Step 1 of the consistency check-program. The amount of possible flags differs between the years due to different data availability (see Table A25). In 2001 and 2017, firms can maximum get flagged once. In 2002-2008, firms can get flagged twice. In 2009-2011, firms can get flagged five times, while they can get flagged 6 times in the years 2012-2016. We activate Flag 1, if firms are flagged one time from 2001-2008 and in 2017, and four times from 2009-2016.
- **Flag 2:** We flag firms if their profit items (*gross profit form PS*; *net profit from PS*; *net profit*

⁴³In case of two intermediary variables, we do not know which one to adjust and we set both equal to missing. This is, however, a minor issue since these "intermediary" variables often are of little interest to researchers.

⁴⁴For multi-industry firms, we also use the sum of net revenue from firms' operations in "other" industries to populate missing *net revenue from PS* for the "main" industry if it is missing.

from PS and financial activities; or total pre-tax profits) exceed the corresponding revenue items.

- **Flag 3:** We flag firms with missing, zero or negative *net revenue from PS and financial activities*.

Based on these flags, we set all profit-loss variables for a firm equal to missing in a year, if Flag 1 or Flag 2 is activated in that year. We set all financial variables, including all balance sheet variables, equal to zero for a firm in a year, if Flag 3 is activated for that firm.

Table A25: Sales and Profit Information

Item	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Gross revenue from PS	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
- Deductions	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Net revenue from PS	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
- Cost of sales	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gross profit from PS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
+ Revenue from fin. act.	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
- Financial costs	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
+ Net profit from fin act.	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
- Operational expenses	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Net profit from PS and fin. act.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
(<i>Net profit from PS only</i>)	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
+ Other revenue	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
- Other costs	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
+ Other net profit	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Total pre-tax profit	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Notes:

PS denotes production and services

Fin. act. denotes financial activities

A.4.8. Value Added and Intermediaries

Value added and intermediaries are not readily available in the VES data. We calculate these variables using the existing information. Value added can be calculated using that $value\ added = wages + gross\ profits$. Unfortunately, we do not observe gross profits of firms from 2001-2008. Therefore, we rely on net profits as proxies for gross profits. Because we have different measures of net profit (see section above), each with their own advantages and drawbacks, we calculate three different value added measures and leave it up to the individual researcher to choose between them. In all calculations, we always measure wage compensation in the same way, including both wages and contributions paid to the employee. The three value added variables are calculated as follows:

- **Value added 1 = total wages + net profits from PS.** While *net profits from PS* is only available from 2002-2008, we can calculate it for 2009-2016 using information from *other net profits* and *net profits from PS and financial activities* (see Table A25). This value added variable is not available in 2001 and 2017.
- **Value added 2 = total wages + net profits from PS and fin. act.** While *net profits from PS and fin. act.* is only available from 2009-2016, we can calculate it 2002-2008 using information from *other net profits* and *net profits from PS* (see Table A25). This value added variable is not available in 2001 and 2017.
- **Value added 3 = total wages + total pre-tax profit.** The total pre-tax profit variable is available in all years and so is the value added variable. While this value added variable is better populated than the above, it include profits and loss from financial and other activities that may not be directly linked to production.

Panel A of Table A26 displays the pairwise correlations of the three value added variables. The variables are highly correlated with coefficients above 0.97. In the bottom row of Panel A, we also show their each variables pairwise correlation with a value added variable which is correctly calculated using gross profits from PS. Since the data only have gross profit information from 2009-2016, the correlations are only based on data for these years. As we would expect, the value added measure calculated based on net profits from PS has the highest correlation with value added measure based on gross profits from PS (0,908).

Table A26: Pairwise Correlations of Value Added and Intermediaries

Panel A: Value added pairwise correlations					
	VA1	VA2	VA3	VA GP	
VA1 = total wages + net profits from PS (2002-17)	1,000				
VA2 = total wages + net profits from PS and fin. act (2002-16)	0,979	1,000			
VA3 = total wages + total pre-tax profit (2001-17)	0,978	0,998	1,000		
VA GP = total wages + gross profit from PS (2009-16)	0,908	0,888	0,713	1,000	
Panel B: Intermediaries pairwise correlations					
	M1	M2	M3	M4	M GP
M1 = net revenue from PS + Δ WC - VA1 (2002-17)	1,000				
M2 = gross revenue from PS and fin. act + Δ WC - VA2 (2002-16)	0,999	1,000			
M3 = total gross revenue + Δ WC - VA3 (2001-17)	0,999	1,000	1,000		
M4 = gross revenue from PS + Δ WC - VA3 (2001-17)	0,999	1,000	1,000	1,000	
M GP = gross revenue from PS + Δ WC - VA GP (2009-16)	0,997	0,996	0,984	0,985	1,000

Notes: We set all financial variables to missing for firms with non-negative intermediaries using the M4 variable. Therefore, pairwise correlations are calculated only for firms with non-negative intermediaries.

PS: production and services.

VA: value added.

M: intermediaries.

GP: gross profit from PS.

Δ WC: Change in working capital.

Equipped with the three value added measures, we calculate four measures of intermediary inputs using that $intermediaries = revenue + change\ in\ working\ capital - value\ added$. The four variables are calculated as specified below:

- **Intermediaries 1 = net revenue from PS + change in working capital - value added 1.** This first measure uses *net* and not *gross* revenue to account for the fact that our value added measure is calculated from *net* profits. The variable is populated from 2002 to 2016.
- **Intermediaries 2 = gross revenue from PS and fin. act. + change in working capital - value added 2.** Our second measure uses revenue and value added from both PS and financial activities to calculate intermediaries. The variable is available from 2002-2016.
- **Intermediaries 3 = total gross revenue + change in working capital - value added 3.** The third intermediary input variable uses total gross revenue and, for consistency, value added calculated from PS, financial activities, and other activities. It is across all years.
- **Intermediaries 4 = gross revenue from PSs + change in working capital - value added 3.** We also calculate a value added variable using gross revenue from PS and value added from all activities. This variable is also available in all years.

A significant part of firms have negative intermediaries. This applies across intermediary input variants. Since we cannot trust the financial information in these firms, we set all of their financial variables equal to missing (based on the third definition of intermediaries). This reduces the number of firms (with financial data) from 3,744,723 to 3,194,484 (see Table A21).

Panel B of Table A26 displays the pairwise correlation between the intermediary input measures. All pairwise correlations shown in Table A26 are calculated without when negative values have been removed. All proxies are very strongly and some are even indistinguishable from each other at the three decimal level. The final row of Table A26 shows the correlation between our different intermediary input proxies and a precise variable calculated based on gross revenue from PS (which is only available from 2009-2016). The results suggest that our different proxies for intermediaries are very closely aligned with the true value (at least in the latter half of the panel). Researchers' choice of intermediary variables should therefore not make a big difference in their analysis. We leave it up to the individual researcher to decide which intermediary variable to use, but we encourage researchers to use an intermediary variable that is consistent with the choice of value added variable.

A.4.9. Balance Sheet Variables

In the VES data, there are six balance sheet variables that are consistently reported over the whole observation periods, with the only exception of liabilities, which is not reported in 2017. Those six variables are related via three accounting identities:

1. total capital = total assets
2. total capital = fixed capital + working capital
3. total assets = equity + liabilities

Moreover, those six variables are also available separately for beginning and end of the year. The above relationships allow us to fill in missing information. Using accounting identities 2 and 3, we are able to populate 932,696 missing values at the beginning of the year, and 944,022 at the end of the year, the vast majority of them being fixed capital and liabilities.

By definition, working capital, total capital, liabilities, and total assets must never be negative. We thus replace negative values for those to missing. Fixed capital and liabilities may be negative, but only if working capital is larger in absolute terms for the former, and if debt is larger in absolute terms for the latter. If those conditions are not fulfilled, we also replace negative values to missing here. By this, we replace 16,552 negative values in the beginning of the year, and 19,419 at the end of the year. More than 96 percent of those changes are liabilities.

Subsequently, we exploit the fact that we have information reported at the beginning and at the end of the year. If total capital at the *end* of year t is missing, we replace it with total capital at the *beginning* of year $t+1$. However, in order to maintain the equality from accounting identity 2, we also replace fixed capital and working capital at end of year t with those at the beginning of year $t+1$. In the same manner, we replace missing total capital at the *beginning* of year t with total capital at the *end* of year $t-1$. Based on this approach, we fill in and adjust the capital information of 5,724 observations at the beginning of the year, and of 461 observations at the end of the year. The same approach is applied to total assets and its two components. 6,578 asset sheets are filled in and adjusted at the beginning of the year, and 559 at the end of the year.

After filling up missing information as described above, we apply a program to deal with inconsistencies in the balance sheet. We therefore check for each observation whether the three

Table A27: Balance Sheet Checks and Adjustments

balance check 1	balance check 2	balance check 3	Adjustment	#beginning	#end
1	1	1	no adjustment needed	3,065,462	3,064,894
1	1	2 or 3	adjust equity	2	6
1	2 or 3	1	adjust working capital	23,823	22,740
1	2 or 3	2 or 3	adjust working capital and equity	0	0
2	1	1 or 2	adjust total assets and equity	25	0
2	2	1	adjust working capital and total capital	0	0
2	2	2	adjust working capital, total assets, and equity	0	0
2	1 or 2 or 3	3	set whole balance sheet missing	0	0
2	3	1 or 2 or 3	set whole balance sheet missing	0	0
3	1 or 2 or 3	1 or 2 or 3	set whole balance sheet missing	1,447	132

Notes:

balance check 1: total capital = total assets

balance check 2: total capital = fixed capital + working capital

balance check 3: total assets = equity + liabilities

1 (exact match), 2 (less than 5 percent deviation), 3 (more than 5 percent deviation).

The two rightmost columns show the number of adjustments made.

accounting identities add up. We start with the first one, i.e. we check whether *total capital = total assets*. Based on this, an observation is assigned one of three different tags (*balance check 1*):

1. if the difference of total capital and total assets is exactly zero
2. if the difference is less than 5 percent of total capital
3. if the difference is larger than 5 percent of total capital

The same tags are assigned for accounting identities 2 (total capital = fixed capital + working capital; *balance check 2*) and 3 (total assets = equity + liabilities; *balance check 3*). Based on the combination of tags, balance sheet information are adjusted the as shown in Table A27. For the vast majority of observations, the three identities perfectly add up, so that adjustment needs to be made. Of those observations where they do not add up exactly, the majority are problems in accounting identity 2, in which cases we adjusted the working capital.

As we now replaced some observations with missings values, we exploit the time dimension of the balance sheet information again as we have done before. We populate missing values at the beginning (end) of year t with values from the end (beginning) if $t-1$ ($t+1$). In order to not mess up the accounting identities 2 and 3 again, we only check for total capital (which is the variable we consider most trustworthy) and then replace the whole balance sheet accordingly. In that way, most of the newly created missing balance sheet observations (1,446 at the beginning and 132 at the end of the year) can be replaced.

After applying all the above measures to populate and adjust missing and inconsistent balance sheet information, we subsequently check for inconsistencies across time. We thus compare information at the end of year t with information at the beginning of year $t+1$. For a given variable, those two should be the same or reasonably close to each other. For cases where the difference between total capital at the end of year t and at the beginning of year $t+1$ is more than 5 percent, we will adjust the balance sheet in one of the two periods. We do so by calculating three ratios: capital over labor, profit over capital, and revenue over capital. For each of those ratios, we then calculate the province-year-2-digit-industry average, using both beginning and end of the year information in a given year. We then adjust according to the following steps in the arranged order:

- **Step 1:** We give priority to balance sheet consistency. That is, if *balance check 1* equals 1 for either end or beginning but not the other, the instance where it equals 1 is used. The whole balance sheet is then replaced.
- **Step 2:** We check whether the three calculated ratios (capital over labor, profit over capital, revenue over capital) are closer to the province-year-2-digit-industry average at the end of year t or at the beginning of year $t+1$. If two out of three of these ratios are closer to their averages, the balance sheet from that period is used to replace the other one.
- **Step 3:** If Step 2 cannot be conducted because one or more of the ratios are missing, we just look at the capital over labor ratio and use the period where that ratio is closer to the average.

In **step 1**, we replace the balance sheets of 28 observations at the beginning of year $t+1$ with those at the end of year t , and zero the other way around. In **step 2**, we replace the balance sheets of 846,862 observations at the beginning of year $t+1$ with those at the end of year t , and 622,012 the other way around. And in **step 3**, the balance sheets of 8 observations at the beginning of year $t+1$ are replaced with those at the end of year t , and 40 the other way around.

A.5. Deflators

All financial variables in the VES data are deflated using the spatial cost of living index (SCOLI), which varies over time and by province. We use Hanoi, 2010, as base. Unfortunately, the SCOLI data available from GSO's website is incomplete, and we need to extrapolate for years 2001-4 and

2016-18. This is years where we have been unable to obtain SCOLI data with temporal and spatial variation. The Extrapolation is conducted as described below.

For years 2001-2004, we predict SCOLI values with a linear regression model. That is, we use the coefficients from a regression of the SCOLI deflator in province p on a spatial GDP deflator in p for years 2005-15 to predict values in 2001-4. The regression is run separately for each province and is of the form:

$$SCOLI_t^{defl} = GDP_t^{defl} + GDP_{t+1}^{defl} + GDP_{t+2}^{defl} + u_t.$$

where we have taken the log of all values. Note that the GDP deflator varies by industry and has a base for each industry in year 2010. Before we run the regression, we aggregate the GDP deflator to the province-level by taking a weighted averaged over provinces, with the weight being the share of output of industry k in province p in year t . We then rebase the deflator to Hanoi, 2010, using spatial variation in 2010 from the SCOLI deflators.

For years 2016-2017: GSO has SCOLI deflators (henceforth SCOLI2) for 2011-2019 on their website, but here Hanoi is a base in each year. We use a national CPI deflator to rebase the SCOLI2 deflators to Hanoi, 2010. To get variation that is consistent with our original SCOLI deflators, we run the following linear model for years 2011-2015:

$$SCOLI2_t^{defl} = SCLOI_t^{defl} + SCLOI_{t-1}^{defl} + SCLOI_{t-2}^{defl} + e_t,$$

and use the coefficients to predict SCOLI in 2016-18. When running the regression, all variables are log-transformed.