From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2021

Essays on Granularity and Machine Learning in Macroeconomics

Vogler, Maximilian

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp013r074z08x

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays on Granularity and Machine Learning in Macroeconomics

Maximilian Jochen Vogler

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance by
the Department of
Economics

Advisor: Richard Rogerson

September 2021

© Copyright by Maximilian Jochen Vogler, 2021.

All rights reserved.

Abstract

This dissertation consists of three independent chapters on granularity in macroeconomics and machine learning methods designed to solve the methodological challenges imposed by granular models, i.e. models with many different individual agents, firms or countries.

In the first chapter, which is co-authored with Jesús Fernández-Villaverde, Galo Nuño and George Sorg-Langhans, we develop a deep-learning algorithm to globally solve high-dimensional dynamic programming problems that result from granular macroeconomic models. We evaluate our methodology in a standard neoclassical growth model and then demonstrate its power in two high-dimensional granular applications – a model of dynamic capital allocation and a model of migration and labor mobility.

In the second chapter, which is co-authored with Cecile Gaubert and Oleg Itskhoki, we focus on the importance of considering a granular firm distribution for government policy in an international trade setting, highlighting three implications: (i) In antitrust regulation, governments face an incentive to be overly lenient towards domestic mergers in comparative advantage sectors. (ii) In trade policy, targeting individual foreign exporters shifts the burden of tariffs from domestic consumers towards foreign producers. (iii) In industrial policy, while generally suboptimal in a closed economy subsidizing 'national champions' can be unilaterally welfare improving in an open economy.

In the third chapter, I demonstrate the importance of considering granularity at the country level by demonstrating that economic crises drive cross-country migration. I show in an event study setting that net migration caused by the Euro crisis accounts for roughly 21% of all migration from EU countries into Germany between 2010-2019. In addition I highlight three salient facts: (i) This increase is mostly due to inflows rather than outflows. (ii) The migration response evolves gradually and achieves its maximum size only after five years. (iii) There is strong evidence for hysteresis.

Acknowledgements

My time in Princeton has been among one of the most enjoyable periods of my life, which would have not been possible without all the people who made it such a special experience.

I am deeply indebted to my advisor Richard Rogerson, who always pushed me forward and spent an incredibly amount of time and effort at every stage. Oleg Itskhoki has not only been a superb co-author, but also a valuable mentor in both academic and non-academic settings. Jesús Fernández-Villaverde helped me belief in the potential of our paper that now constitutes the first chapter of this thesis and was always supportive when I saw no way forward. Gianluca Violante offered invaluable help throughout my PhD and has been the best and most open-minded placement director I could ask for. Ben Moll gave me the opportunity to work for him in an early stage of my program, thus opening my eyes to the world of computational work that has fascinated me every since. Beyond that I want to thank all those faculty members and co-authors who have offered an incredibly amount of support over the years: Mark Aguiar, Markus Brunnermeier, Cecile Gaubert, Nobu Kiyotaki, Galo Nuño, Ezra Oberfield, Steve Redding, Esteban Rossi-Hansberg and Chris Sims.

I cannot imagine my life in Princeton without the amazing fellow students that I had the privilege to meet. George Sorg-Langhans has been the best co-author, roommate and friend I could have hoped for and I am deeply grateful to have met him here. Joseph Abadi, Patrick Agte, Bruno Baranek, Stefano Baratuche, Riccardo Cioffi, Benny Kleinman, Pellumb Reshidi, Simon Schmickler and Chris Wolf have become close friends with whom I have shared many memorable moments and who I will dearly miss.

I am very grateful for all the generous support I have received from the Department of Economics. Laura Hedden has been the kindest and most wonderful help in all administrative matters. Both her and Jennifer Bello were always there when needed, even in the middle of the Covid Pandemic.

I want to thank my family, who mean so much to me and without whom this thesis would not have been possible. My girlfriend Kate has given me every possible help despite having to tolerate endless discussions about economics. I am very happy we found each other and I cannot wait for our future together. All of my aunts, uncles and cousins provided an incredible bedrock of support. Especially my godfather, Jochen Paul, and my aunt, Barbara Paul-Jahanshahi, have always been there for me in good and in bad days. My grandparents, Hildegard and Jörg Paul, have been a great inspiration and have always encouraged me. Finally, I owe everything I am and everything I have achieved to my mother, Susanne Vogler, and my late father, Christian Vogler.

To my parents.

Contents

Abstract iii

Acknowledgements iv

1 Solving High-Dimensional Dynamic Programming Problems using Deep Learning **1**

1.1 Introduction 1

1.2 Solution Method 6

1.2.1 Outline of Problem 6

1.2.2 Solution Algorithm 8

1.2.3 Discussion 11

1.2.4 Comparison to Related Methods 21

1.3 Neoclassical Growth Model 22

1.3.1 Neoclassical Growth Model with Closed-Form Policy Function 23

1.3.2 Neoclassical Growth Model with Policy Approximation 25

1.4 Dynamic Capital Allocation Model 27

1.4.1 Model Outline 27

1.4.2 Complexity 29

1.4.3 Calibration 29

1.4.4 Implementation 29

1.4.5 Results 31

1.5 Migration and the Propagation of Shocks 34
1.5.1 Model Outline 36
1.5.2 Complexity 42
1.5.3 Calibration 42
1.5.4 Implementation 43
1.5.5 Results 45
1.6 Conclusion 47

2 Government Policies in a Granular Global Economy 50
2.1 Introduction 50
2.2 A Quantified Granular Model 55
2.2.1 Theoretical Framework 55
2.2.2 Fundamental and granular comparative advantage 60
2.2.3 Model quantification 62
2.3 Evaluating Granular Policies 63
2.4 Mergers, Acquisitions and Antitrust 68
2.4.1 Merger analysis: setup 68
2.4.2 Welfare implications of a merger 70
2.4.3 Optimal M&A policy 74
2.5 Granular Tariffs 79
2.6 Industrial Policy in a Granular World 84
2.7 Conclusion 88

3 The Migration Response to the Euro Crisis - An Empirical Investigation 90
3.1 Introduction 90
3.2 Data 95
3.2.1 Data Sources 95

3.2.2 Descriptive Statistics . . . 98
3.3 Identification Strategy . . . 101
3.3.1 A Selection on Observables Framework . . . 102
3.3.2 A Treatment vs Control Framework . . . 105
3.3.3 A Closer look at Treatment and Control Groups . . . 108
3.4 Main Results . . . 112
3.4.1 Setup . . . 112
3.4.2 Results . . . 115
3.4.3 Discussion . . . 120
3.5 Robustness . . . 123
3.6 Decomposition . . . 127
3.6.1 Inflows . . . 128
3.6.2 Outflows . . . 131
3.7 Conclusion . . . 134

A Appendix for Chapter 1 **136**
A.1 Proof for continuous HJB . . . 136
A.2 Implementation Details . . . 137
A.3 Simulation Details for the Ergodic Distribution . . . 142
A.4 Steady States . . . 145

B Appendix for Chapter 2 **147**
B.1 Model Quantification . . . 147
B.2 Granular Tariff . . . 150

C Appendix for Chapter 3 **153**

Bibliography **173**

Chapter 1

Solving High-Dimensional Dynamic Programming Problems using Deep Learning[1]

I thank my co-authors Jesús Fernández-Villaverde, Galo Nuño and George Sorg-Langhans for their permission to reproduce our work here.

1.1 Introduction

To answer a wide range of important economic questions, researchers must solve high-dimensional dynamic programming problems. Examples include consumption-saving problems with many assets, business cycle models with numerous sectors or countries, multi-product menu-cost models, corporate finance models with various types of capital goods and bonds of different maturities, models of household formation and evolution, models of industry dynamics and production networks, learning models, dynamic games with multiple players, and many others.

In particular, during the last few years, the flourishing of survey evidence and detailed micro data has generated a growing interest in models with granular agents (e.g., models with

[1]I thank the participants at 2021 ASSA Meetings for helpful comments.

multiple locations, sectors, goods, etc.) that can incorporate and match this rich observational data. These models with granular agents stand between models with a representative agent and models with a continuum of agents. Up to this point, a fundamental impediment to the study of such granular models has been that the number of states usually grows linearly with the number of agents, as we need to keep track of the individual states of each agent in addition to aggregate states. Thus, even when dealing with just 19 locations (e.g., the 19 member states of the eurozone) or 24 industries (the 24 industry groups in the Global Industry Classification Standard, GICS), we can have dynamic programming problems with dozens, if not hundreds, of state variables.

Solving these high-dimensional dynamic programming problems is exceedingly difficult due to the well-known "curse of dimensionality" (Bellman, 1958, p. ix). As the number of states in the dynamic programming problem grows linearly, the computational burden grows exponentially, both in terms of the number of floating-point operations to be undertaken and regarding memory requirements. Whilst perturbation approaches have proven powerful to tackle approximately linear models and much progress has been made with novel approaches such as sparse grids during the last decade (Brumm & Scheidegger, 2017), globally solving dynamic programming problems with over 30 state variables and inherent non-linearities remains an open challenge.

This paper presents a deep-learning algorithm that tackles the "curse of dimensionality" and efficiently provides a global solution to high-dimensional dynamic programming problems. At the heart of our algorithm lies the use of deep neural networks. Neural networks are an attractive approach because they are universal nonlinear approximators (Hornik et al., 1989; Cybenko, 1989; Bach, 2017) and break the "curse of dimensionality" (Barron, 1993; Bach, 2017). In comparison, the traditional functional approximation approaches, including polynomials, piecewise polynomials, wavelets, splines, and radial functions, cannot break the "curse of dimensionality."

The intuition is simple: neural networks are compositional (they compose simple activation functions to approximate complex nonlinear functions) while traditional functional approximation methods are additive (they build linear combinations of basis functions). The compositional approach handles high dimensions much better than the additive approach. This is because additional activation functions in the compositional approach only move a small range of frequencies, allowing for a fine-tuning of the approximation even in high dimensions.

Among all the different neural network architectures, we use deep neural networks because they have been shown to work surprisingly well in many contexts. Most relevant for us, deep neural networks work very well when dealing with high dimensional PDEs, a class of mathematical problems that includes dynamic programming (Han et al., 2018; Sirignano & Spiliopoulos, 2018). However, our approach could easily accommodate alternative architectures.

The first part of this paper presents our three-step algorithm in the context of a generic recursive problem formulated in continuous time. First, we build four deep neural networks to approximate i) the value function of the problem, ii) the policy function, iii) the associated Karush-Kuhn-Tucker multipliers for constraints that hold with inequality and iv) the Lagrange multipliers for constraints that hold with equality. Equivalently, one can think about these four deep neural networks as just one large deep neural network with multiple outputs. However, presenting the algorithm in terms of four separate networks simplifies the exposition. Second, we construct an appropriate error criterion in order to train the networks, i.e. in order to find the parameters that properly approximate the target functions. To this end, we plug the four deep neural networks into the Hamilton-Jacobi-Bellman (HJB) equation that formally defines our dynamic programming problem. This allows us to define our error criterion as the sum of the HJB error, the policy function error, and the constraint error. Third, we minimize this error criterion through mini-batch gradient descent

over points drawn from the ergodic distribution of the state vector. The whole process is highly efficient and perfectly suitable for parallelization.

Note that we can simulate training data directly from the underlying model given the current state of the neural networks. This obviates the need for a validation and training set as well as regularization techniques designed to reduce overfitting. Whilst in practice we train our model on more than one billion data points to achieve convergence, in principle we could generate arbitrary amounts of data.

Importantly, we cast our problem in continuous time. Although our algorithm applies to discrete time with minor changes of notation, our formulation in continuous time offers one key advantage: we get rid of integrals. An often overlooked aspect of the "curse of dimensionality" is the need to evaluate high-dimensional integrals in the continuation value function.

The relevant set to evaluate an integral is the so-called typical set of the distribution of interest (Cover & Thomas, 2012). The typical set of high-dimensional distributions feature two crucial but surprising properties. First, in general, the typical set is not the region where the associated density is the highest. Thus, naively searching for quadrature points or simulating around that high-density region will result in low accuracy in the approximation of the integrals. Second, by concentration of measure, the typical set will be a narrower and narrower band as the number of states grows. Thus, in most situations with high dimensions, we will waste nearly all computations while approximating integrals. Unless we have information about where this typical set lays, we will be dealing with regions of the distribution that are (nearly) irrelevant for the computation of the integral.

Unfortunately, characterizing the typical set is hard. Algorithms that do so, such as the Hamiltonian Monte Carlo are costly to implement (see, for details, Fernández-Villaverde & Guerrón-Quintana, 2020). A much more fruitful approach is to cast the dynamic programming problem in continuous time, taking advantage of Itô's Lemma, and therefore dropping

the integrals. Computing the derivatives implied by Itô's Lemma is considerably easier than computing any integral, in particular given our neural network architecture and the properties of the backpropagation algorithm used for gradient computation.

Having said that, we reiterate that our algorithm can be applied to discrete time problems and keep all of its advantages with respect to traditional functional approximation procedures (all of which suffer from the exact same need of evaluating high-dimensional integrals).

In the second part of this paper, we illustrate our algorithm with three applications. First, we solve a standard neoclassical growth model with a conventional calibration. Computationally, this is a straightforward problem but it provides us with a sharp testbed to show how our method works and gauge its accuracy.

Next, we show how to solve a dynamic capital allocation model with our algorithm. The model investigates optimal consumption, investment, and cross-location capital flows in a model of aggregate productivity shocks and capital movement frictions. We calibrate the model to 25 locations. There are two continuous state variables in each location, leading to a total of 50 continuous aggregate state variables.

Finally, we turn to a setting of inherent non-linearity in order to illustrate the global solution properties of our method in a state-of-the art macroeconomic model. Specifically, we study the propagation of economic shocks in a multi-country model featuring fully rational expectations, migration linkages, and 75 continuous aggregate state variables.

Our paper is related to the contributions of Maliar et al. (2019) and Azinović et al. (2019), who propose discrete-time algorithms. While similar in spirit in their use of neural networks as global approximators, these papers face the hurdle of approximating the expectations integral as explained above. Our approach elegantly circumvents this issue by using the analytic solutions provided by the continuous-time limit. Another related approach to ours is based on the reinforcement learning literature in the spirit of Sutton & Barto (2018) as applied by Duarte (2018) in a continuouts-time context. In contrast to this approach, we

utilize the fact that in economic problems we have additional knowledge of the underlying problem structure. Specifically, we can draw on the first-order conditions associated with the problem to train our neural networks. This additional information helps us greatly in pinning down the policy functions and extending the dimensionality of the problems we can solve.

The rest of the paper is organized as follows. Section 1.2 introduces our deep learning algorithm, discusses some of its most salient aspects, and compares it with other related methods. Section 1.3 presents our first application: the standard neoclassical growth model. Section 1.4 describes our second application: a dynamic capital allocation model. Section 1.5 introduces our third application: a migration model. Section 1.6 concludes.

1.2 Solution Method

We now present the details of our solution algorithm. We start with an outline of the generic dynamic programming problem we want to handle and discuss the challenges associated with its solution. Next, we introduce our deep-learning algorithm, followed by an in-depth discussion of its elements. We close this section by discussing the differences of our approach to alternative algorithms recently proposed in the literature.

1.2.1 Outline of Problem

To start with, we standardize notation for the rest of the paper. We will denote scalar values as lower case letters, vectors as bold lower case letters, and matrices as bold upper case letters. Vectors are column vectors. Thus, transposed vectors indicate row vectors.

Our goal is to solve the recursive continuous-time HJB equation globally:

$$\begin{aligned} \rho V(\mathbf{x}) = \max_{\boldsymbol{\alpha}} r(\mathbf{x}, \boldsymbol{\alpha}) + \nabla_x V(\mathbf{x}) f(\mathbf{x}, \boldsymbol{\alpha}) + \frac{1}{2} tr(\sigma(\mathbf{x}))^T \Delta_x V(\mathbf{x}) \sigma(\mathbf{x})) \\ \text{s.t.} \quad G(\mathbf{x}, \boldsymbol{\alpha}) \leq \mathbf{0} \quad \text{and} \quad H(\mathbf{x}, \boldsymbol{\alpha}) = \mathbf{0}, \end{aligned} \tag{1.1}$$

where $V : \mathbb{R}^N \to \mathbb{R}$ is a value function depending on a state vector $\boldsymbol{x} \in \mathbb{R}^N$ with $N \in \mathbb{R}$, $\nabla_x V(\mathbf{x}) \in \mathbb{R}^N$ is the gradient of V, and $\Delta_x V(\mathbf{x}) \in \mathbb{R}^{N \times N}$ is the Hessian of V.

In problem (1.1), ρ denotes the discount rate, $r : \mathbb{R}^{N \times M} \to \mathbb{R}$ is the instantaneous return function, that depends on the states and the control vector $\boldsymbol{\alpha} \in \mathbb{R}^M$, where $M \in \mathbb{R}$. The state vector follows a diffusion process with drift $f : \mathbb{R}^{N \times M} \to \mathbb{R}^N$ and diffusion $\sigma : \mathbb{R}^N \to \mathbb{R}^N$. The problem is constrained by the constraint functions $G : \mathbb{R}^M \to \mathbb{R}^{L_1}$ and $H : \mathbb{R}^M \to \mathbb{R}^{L_2}$, where $L_1, L_2 \in \mathbb{R}$ are the number of inequality and equality constraints respectively. We denote by $\alpha : \mathbb{R}^N \to \mathbb{R}^M$ the policy function, which is the maximized control vector given some states (notice the absence of bold font to distinguish the policy function from the control vector itself).

We are interested in solving the problem (1.1) when N is high, that is when we are facing a high-dimensional problem. It is well known that, in this case, problem (1.1) suffers from an acute "curse of dimensionality." Grid-based approaches quickly become infeasible since the number of required grid points grows exponentially with N for a given precision. Local solution methods such as linearization or higher-order polynomial expansions are not suitable for problems that display irregular behavior, such as kinks or other strong nonlinearities.

In contrast, our solution algorithm copes well with both of those challenges. Our approach is capable of i) globally approximating irregularly behaved solutions to ii) high-dimensional problems.

1.2.2 Solution Algorithm

Our approach to solving problem (1.1) builds on the ideas of deep learning. We use neural networks as global nonlinear approximators for both value and policy functions. We define an error criterion, which is composed of the HJB error and the first-order conditions (FOC) errors. Then, we train the neural networks (i.e., update the weights in the networks) by picking points in the state space from their ergodic distribution and minimize our error criterion using mini-batch gradient descent until convergence. Let us consider each of these steps in detail.

First, we define four neural networks $\tilde{V}(\mathbf{x};\boldsymbol{\Theta}^V) : \mathbb{R}^N \to \mathbb{R}$, $\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha) : \mathbb{R}^N \to \mathbb{R}^M$, $\tilde{\mu}(\mathbf{x};\boldsymbol{\Theta}^\mu) : \mathbb{R}^N \to \mathbb{R}^{L_1}$, and $\tilde{\lambda}(\mathbf{x};\boldsymbol{\Theta}^\lambda) : \mathbb{R}^N \to \mathbb{R}^{L_2}$. These four neural networks are parameterized by weights $\boldsymbol{\Theta}^V$, $\boldsymbol{\Theta}^\alpha$, $\boldsymbol{\Theta}^\mu$, and $\boldsymbol{\Theta}^\lambda$ to approximate i) the value function $V(\mathbf{x})$, ii) the policy function $\boldsymbol{\alpha}$, iii) the Karush-Kuhn-Tucker (KKT) multipliers $\boldsymbol{\mu}$ and iv) the Lagrange multipliers $\boldsymbol{\lambda}$. To simplify notation, we accumulate all neural network weights in the matrix $\boldsymbol{\Theta} = (\boldsymbol{\Theta}^V, \boldsymbol{\Theta}^\alpha, \boldsymbol{\Theta}^\mu, \boldsymbol{\Theta}^\lambda)$.[2] If the reader is unfamiliar with neural networks, it suffices for now to treat a neural network as any parameterized function approximator. We will provide a brief exposition along with a discussion of its advantages for this problem below.

Second, we define our error criterion. It is composed of three elements: the HJB error, the policy function error, and the constraint error. For any point $\mathbf{x} \in \mathbb{R}^N$ in the state space, the HJB error is defined as the difference between the right and the left-hand sides of the HJB when we substitute the exact value and policy functions by their approximations:

$$err_{HJB}(\mathbf{x};\boldsymbol{\Theta}) \equiv r(\mathbf{x},\tilde{\alpha}(\mathbf{s};\boldsymbol{\Theta}^\alpha)) + \nabla_x \tilde{V}(\mathbf{x};\boldsymbol{\Theta}^V) f(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha)) + \\ + \frac{1}{2} tr[\sigma(\mathbf{x})^T \Delta_x \tilde{V}(\mathbf{x};\boldsymbol{\Theta}^V)\sigma(\mathbf{x})] - \rho \tilde{V}(\mathbf{x};\boldsymbol{\Theta}^V)$$

[2]We will reserve the term "parameter" for elements defining the economic problem (e.g., risk aversion, discount rate) and "weights" for the elements that parameterize the neural network but that, in general, do not have any sharp economic interpretation.

Analogously, the policy function error is defined as the difference of the FOC from zero when we plug in the approximated value and policy functions:

$$\begin{aligned} err_\alpha(\mathbf{x};\boldsymbol{\Theta}) \equiv & \frac{\partial r(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha)}{\partial \boldsymbol{\alpha}} + D_\alpha f(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha))^T \nabla_x \tilde{V}(\mathbf{x};\boldsymbol{\Theta}^V) \\ & - D_\alpha G(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha))^T \tilde{\mu}(\mathbf{x};\boldsymbol{\Theta}^\mu) - D_\alpha H(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha))\tilde{\lambda}(\mathbf{x};\boldsymbol{\Theta}^\lambda), \end{aligned}$$

where $D_\alpha G \in \mathbb{R}^{L_1 \times M}$, $D_\alpha H \in \mathbb{R}^{L_2 \times M}$, and $D_\alpha f \in \mathbb{R}^{N \times M}$ are the submatrices of the Jacobian matrices of G, H and f respectively containing the derivatives with respect to $\boldsymbol{\alpha}$.

Finally, the constraint error is itself composed of four different errors: the primal feasibility errors:

$$err_{PF_1}(\mathbf{x};\boldsymbol{\Theta}) \equiv \max\{0, G(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha))\}, \qquad err_{PF_2}(\mathbf{x};\boldsymbol{\Theta}) \equiv H(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha)),$$

the dual feasibility error:

$$err_{DF}(\mathbf{x};\boldsymbol{\Theta}) = \max\{0, -\tilde{\mu}(\mathbf{x};\boldsymbol{\Theta}^\mu\},$$

and the complementary slackness error:

$$err_{CS}(\mathbf{x};\boldsymbol{\Theta}) = \tilde{\mu}(\mathbf{x};\boldsymbol{\Theta})^T G(\mathbf{x},\tilde{\alpha}(\mathbf{x};\boldsymbol{\Theta}^\alpha)).$$

We combine these four errors by using the squared error as our loss criterion:

$$\begin{aligned} \mathcal{L}(\mathbf{x};\boldsymbol{\Theta}) \equiv & \left|\left|err_{HJB}(\mathbf{x};\boldsymbol{\Theta})\right|\right|_2^2 + \left|\left|err_\alpha(\mathbf{x};\boldsymbol{\Theta})\right|\right|_2^2 + \left|\left|err_{PF_1}(\mathbf{x};\boldsymbol{\Theta})\right|\right|_2^2 + \\ & + \left|\left|err_{PF_2}(\mathbf{x};\boldsymbol{\Theta})\right|\right|_2^2 + \left|\left|err_{DF}(\mathbf{x};\boldsymbol{\Theta})\right|\right|_2^2 + \left|\left|err_{CS}(\mathbf{x};\boldsymbol{\Theta})\right|\right|_2^2 \end{aligned} \tag{1.2}$$

We train our neural networks by minimizing the above error criterion through mini-batch gradient descent over points drawn from the ergodic distribution of the state vector.

Algorithm 1: Deep-Learning Solution Algorithm for HJB

Parameterize value and policy functions through neural networks $\tilde{V}$, $\tilde{\alpha}$, $\tilde{\mu}$, and $\tilde{\lambda}$;

Define the loss function $\mathcal{L}$ as in (1.2);

Initialize neural network weights $\mathbf{\Theta}_1$;

for $k = 1, ..., K$ **do**

Draw I sample points $\{\mathbf{x}_i\}_{i=1}^I$ from their ergodic distribution $\nu(\mathbf{x})$;

Randomly split these points into B mini-batches of size S: $\{\{\mathbf{x}_{s,b}\}_{s=1}^S\}_{b=1}^B$;

Use previous weights as starting weights for new step: $\mathbf{\Theta}_{k,1} = \mathbf{\Theta}_k$;

for $b = 1, ..., B$ **do**

Define batch error: $\bar{\mathcal{L}}(\{\mathbf{x}_{s,b}\}_{s=1}^S; \mathbf{\Theta}_{k,b}) \equiv \frac{1}{S}\sum_{s=1}^S \mathcal{L}(\mathbf{x}_{s,b}; \mathbf{\Theta}_{k,b})$;

Train $\tilde{V}$: $\mathbf{\Theta}^V_{k,b+1} = \mathbf{\Theta}^V_{k,b} - \eta_k \nabla_{\mathbf{\Theta}^V_{k,b}} \bar{\mathcal{L}}(\{\mathbf{x}_{s,b}\}_{s=1}^S; \mathbf{\Theta}_{k,b})$;

Train $\tilde{\alpha}$: $\mathbf{\Theta}^\alpha_{k,b+1} = \mathbf{\Theta}^\alpha_{k,b} - \eta_k \nabla_{\mathbf{\Theta}^\alpha_{k,b}} \bar{\mathcal{L}}(\{\mathbf{x}_{s,b}\}_{s=1}^S; \mathbf{\Theta}_{k,b})$;

Train $\tilde{\mu}$: $\mathbf{\Theta}^\mu_{k,b+1} = \mathbf{\Theta}^\mu_{k,b} - \eta_k \nabla_{\mathbf{\Theta}^\mu_{k,b}} \bar{\mathcal{L}}(\{\mathbf{x}_{s,b}\}_{s=1}^S; \mathbf{\Theta}_{k,b})$;

Train $\tilde{\lambda}$: $\mathbf{\Theta}^\lambda_{k,b+1} = \mathbf{\Theta}^\lambda_{k,b} - \eta_k \nabla_{\mathbf{\Theta}^\lambda_{k,b}} \bar{\mathcal{L}}(\{\mathbf{x}_{s,b}\}_{s=1}^S; \mathbf{\Theta}_{k,b})$;

end

Update step weights: $\mathbf{\Theta}_{k+1} = \mathbf{\Theta}_{k,B+1}$;

end

The efficient implementation of this last step is the key to the success of our algorithm. We start by initializing our network weights (a detailed discussion of how to pick good initial weights is provided in the Appendix). Then, we perform K learning steps called epochs. For each epoch, we draw I points from the state space according to its ergodic

distribution. This is computationally very inexpensive when a closed-form expression for the ergodic distribution is available and still relatively inexpensive if the ergodic distribution has to be simulated. Then, we randomly split this sample into B mini-batches of size S. For each mini-batch, we define the mini-batch error, by averaging the loss function over the batch. Finally, we perform mini-batch gradient descent for all network weights, with η_k being the learning rate in the k-th epoch, the choice of which is again discussed in the Appendix. Algorithm 1 summarizes our algorithm.

1.2.3 Discussion

It is worthwhile to discuss the individual elements of the algorithm. In particular, we need to explain the advantages of formulating our problem in continuous time, the choice of neural networks as our global nonlinear approximator, the updating algorithm, and the generation of training data. We conclude this section by highlighting the main differences in training neural networks on model-generated data compared to their traditional application to real-world generated data.

1.2.3.1 Advantages of Continuous Time

One crucial element for our algorithm is that we formulate problem (1.1) in continuous time. Technically, this means that we have to handle partial differential equations instead of partial difference equations. In comparison with the existing neural network algorithms existing in the literature, framed in discrete time, the continuous-time formulation enjoys one crucial advantage.

To understand why this is the case, consider the discrete-time value function analogous to the continuous-time HJB, which has the following form:

$$V(\mathbf{x}) = \max_{\boldsymbol{\alpha}} r(\mathbf{x}, \boldsymbol{\alpha}) + \beta \mathbb{E}[V(\mathbf{x}')]. \tag{1.3}$$

Notice that now the continuation value includes an expectation term, i.e., an integral over all possible realizations of $\mathbf{x}'$. This is a fundamental challenge for solution algorithms, since for each point of the state space, one needs to approximate this integral – a potentially time-expensive and inaccurate step. In particular, the solution algorithm needs to know (or cover) the typical set, the relevant region of the state variables for the computation of the expectation of $\mathbf{x}$.

For $\epsilon > 0$ and any I, we define the typical set A_ϵ^I with respect to a density $p(\mathbf{x})$ as:

$$\left\{ (\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_I) : \left| -\frac{1}{I+1} \sum_{i=0}^{I} \log p(\mathbf{x}_i) - h(\mathbf{x}) \right| \leq \epsilon \right\},$$

where $h(\mathbf{x}) = -\int p(\mathbf{x}_i) \log p(\mathbf{x}_i) d\mathbf{x}$ is the differential entropy of the vector $\mathbf{x}$ with respect to its density (a measure of the concentration of a density; Cover & Thomas, 2012, ch. 8).

By a weak law of large numbers, $Pr(A_\epsilon^I) > 1-\epsilon$ for I sufficiently large. In other words, we call A_ϵ^I "typical" because it includes "most" sequences of $\mathbf{x}_i$'s that are distributed according to $p(\mathbf{x})$. This result shows why A_ϵ^I is, indeed, the relevant region to compute expectations of $\mathbf{x}$ or of functions of $\mathbf{x}$.

Two properties of the typical set are counterintuitive, but crucial. First, in general, the typical set is not the region where the density $p(\mathbf{x})$ is the highest. Second, by concentration of measure, the typical set will be a narrower and narrower band as the dimensionality of $\mathbf{x}$ grows.

Since, in general, we will not know where the typical set of the density of the state $\mathbf{x}$ is, we will either spend too many resources computing the integral in the Bellman operator or get an inaccurate evaluation of the expectation. Additionally, this challenge becomes exponentially worse as the number of state variables with respect to which we need to take the integral grows.

As shown in the Appendix, the discrete-time problem (1.3) converges towards the HJB problem (1.1) as the time step $\Delta t \to 0$. But, importantly, once we reach continuous time, we do not face the integral-approximation issue, since due to the application of Ito's Lemma, we get an analytical expression for this expectation, which depends only on the gradient and Hessian of V. Furthermore, the efficient computation of gradients and Hessians is one of the key advantages of neural networks.

In summary, jumping to continuous time allows us to replace a challenging problem, the evaluation of integrals, with a much simpler one, the computation of gradients and Hessians with neural networks training algorithms.

1.2.3.2 Neural Networks

Definition (Artificial) neural networks are a class of nonlinear functions, mapping $\mathbb{R}^N \to \mathbb{R}^J$, that have gained much popularity over the past two decades thanks to breakthroughs in training algorithms (Rumelhart et al., 1986; Glorot & Bengio, 2010) as well as the vast increase in the availability of both data and computing power.

We will work with a simple version of a neural network, namely a multi-layer feedforward network. We have experimented with more complex network structures such as Long Short-Term Memory (LSTM) networks but we have found that a simple feedforward structure achieves the best results in our settings. Also, for simplicity, we will present the case where the final outcome of the neural network is a single output ($J = 1$), but the approach is easy

to generalize to multiple outputs ($J > 1$). In fact, we use that idea below where we create layers.

At the lowest level a neural network is composed of a"neuron", which is a function $\mathbb{R}^N \to \mathbb{R}$ of the form:

$$m(\mathbf{x}; \boldsymbol{\theta}) \equiv \phi\left(\theta_0 + \sum_{i=1}^{N} \theta_i x_i\right)$$

The function takes an input $\mathbf{x} \in \mathbb{R}^N$ and is parameterized by the weight vector $\boldsymbol{\theta} \in \mathbb{R}^{N+1}$, which contains both the signal weights $\{\theta_i\}_{i=1}^N$ and the bias θ_0. The activation function $\phi(\cdot)$ is an appropriate nonlinear function, for example the hyperbolic tangent $\phi(x) = \tanh(x)$ or the rectifier $\phi(x) = \max\{0, x\}$. Intuitively, a neuron is just a nonlinear transformation of a linear combination of its inputs.

We can create a "layer" by stacking N_1 neurons on top of each other. That is, a layer is a function $\mathbb{R}^N \to \mathbb{R}^{N_1}$:

$$M(\mathbf{x}; \tilde{\boldsymbol{\Theta}}) \equiv (m(\mathbf{x}; \boldsymbol{\theta}_1), ..., m(\mathbf{x}; \boldsymbol{\theta}_{N_1}))^T,$$

where $\tilde{\boldsymbol{\Theta}}$ is the matrix of all parameters of the neurons in the layer, given by $\tilde{\boldsymbol{\Theta}} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{N_1})$. All neurons in a layer share the same input $\mathbf{x}$. However, since each neuron i has a different parameterization $\boldsymbol{\theta}_i$, this yields N_1 different outputs of the layer.

A neural network now combines multiple such layers, by feeding the output of the previous layer as the input into the next layer.[3] Finally, the output of the last layer is fed into an output layer. This is a single neuron with an identity activation function, i.e. it is a simple

[3]To improve performance, the inputs into the first layer are usually normalized, a detailed discussion of which can be found in the Appendix.

linear combination of its inputs.[4] A neural network is therefore the composition of different layers. For example a neural network with two layers is a function $\mathbb{R}^N \to \mathbb{R}$:

$$\tilde{N}(\mathbf{x}; \boldsymbol{\Theta}) \equiv \theta_0^{Out} + \boldsymbol{\theta}_{1:N_2}^{Out} M(M(\mathbf{x}; \tilde{\boldsymbol{\Theta}}_1); \tilde{\boldsymbol{\Theta}}_2)),$$

where N_2 is the number of output neurons of the second layer and $\boldsymbol{\Theta}$ is the matrix of all parameters of the neural network, comprised of the parameters of all hidden layers as well as of the output layer $\boldsymbol{\Theta} = (\tilde{\boldsymbol{\Theta}}_1, \tilde{\boldsymbol{\Theta}}_2, \boldsymbol{\theta}^{Out})$, with $\boldsymbol{\theta}^{Out} = (\theta_0^{Out}; \boldsymbol{\theta}_{1:N_2}^{Out})$. Since the outputs of all layers except for the output layer are generally not of direct interest, those layers are commonly referred to as "hidden layers". A neural network with more than two hidden layers is called a "deep" neural network.

A visualization of such a network with two hidden layers is depicted in figure 1.1. Here, we feed three inputs, which are represented by the three red circles and dneoted as 'input layer', into the network. All inputs are then fed into each neuron in the first hidden layer. Each neuron in this layer nonlinearly combines these inputs into one output as described above. All of those outputs are then in turn fed as inputs into the neurons of the second hidden layer, which again transform them. Finally, all the outputs of the second hidden layer are combined into one final output, represented by the green circle in the output layer.

Neural Networks as a Global Nonlinear Approximator Besides neural networks, there exists a wide array of alternative classes of approximators. For example, we could have chosen Chebyshev, Gegenbauer, or Legendre polynomials, of which Chebyshev polynomials are most widely used in economics.[5] It is well known that for Chebyshev polynomials the ap-

[4]The last layer does not necessarily have to use an identity activation function. Depending on the application, we can use different activation functions for the output layer. For example, logistic or softmax functions are frequently used for classification tasks.

[5]They appear to be the best suited for the conventional small state space problems found in economics, as discussed by Boyd & Petschek (2014).

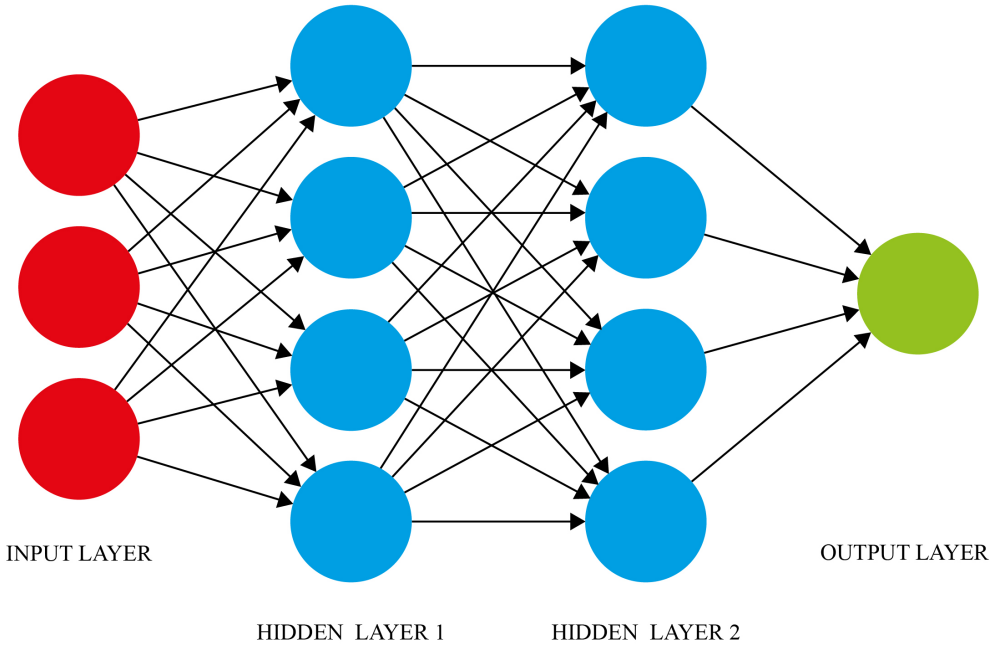


Figure 1.1: A neural network with two hidden layers.

proximation error becomes arbitrarily small as the number of degrees goes to infinity (Erdös and Turan, 1937), and that said error can be approximated for finite degrees (Boyd, 2001). However, an equivalent theorem exists for neural networks. The Universal Approximation Theorem (Hornik et al., 1989; Cybenko, 1989; Bach, 2017) shows that a neural network with at least one layer can approximate any Borel-measurable function of finite dimensionality to any degree of accuracy.

The main theoretical strength of neural networks, however, stems from a theorem by Barron (1993). It states that under some technical conditions a one-layer neural network achieves integrated square errors of order $O(1/n)$, where n is the number of neurons. In comparison, for series approximations, such as polynomials, splines, or trigonometric expansions, the integrated square error is of order $O(1/(n^{2/d}))$, with d denoting the number of

dimensions of the function to be approximated.[6] In other words, holding the approximation error fixed, the number of neurons needed to approximate the function grows linearly with the number of dimensions for neural networks, whilst it grows exponentially for series approximations. This theorem is of fundamental importance for our focus on high-dimensional problems, since it essentially states that the "curse of dimensionality" does not apply to the same extent to neural networks. Note that this theorem makes no statement about the intercept, so it is well possible that for low dimensions series approximations require fewer terms. However, for the high dimensions that we are interested in neural networks are the better approximators.

Note that this power does not come for free. Neural networks have very high data requirements compared to other approximators, which has prevented them from being in widespread use until the recent rise of big data. However, for our algorithm in principle we have access to an arbitrarily large amount of data as we will discuss below. Nevertheless, this data requirement stresses that our algorithm is tailored towards high-dimensional problems and does not aim to compete with alternatives on low-dimensional problems.

In addition to these theoretical strengths, neural networks have also achieved great empirical success over a wide range of problems. They have been observed to extrapolate significantly better beyond their training set than alternatives such as Chebyshev polynomials (for an economic illustration see Fernández-Villaverde et al., 2019).

The final strength of neural networks is the relative ease and efficiency with which they can be estimated. As noted above, our algorithm requires an extremely efficient computation of neural network gradients with respect to both its parameters Θ needed for the gradient descent step and the states $\mathbf{x}$ required for the computation of the loss function. The backpropagation algorithm discussed in the Appendix provides an extremely efficient way to

[6]These results are extended by Bach (2017) to cover nondecreasing positively homogeneous activation functions, such as the rectifier, and to derive approximation and estimation errors.

compute all those gradients for neural networks (Rumelhart et al., 1986). In particular, it can be shown that the computation of all gradients requires the same computation time as one evaluation of the neural network itself (Baydin et al., 2018).

1.2.3.3 Mini-batch Gradient Descent

As alluded to above, one crucial advantage of neural networks is that basic gradient-descent algorithms yield efficient and reliable updates. In our algorithm, we have opted for mini-batch gradient descent as the basic training algorithm. Let us briefly discuss this decision.

Fundamentally, there are two different approaches for gradient descent. First, one could perform one training step per epoch, with the loss function averaged over all I data points generated in one epoch, an approach known as deterministic gradient descent. Second, one could have one training step per data point, which is referred to as stochastic gradient descent. The rationale for deterministic gradient descent is that the large number of points used in the average will lead to a very good approximation of the real gradient, so that the algorithm will always follow the steepest slope down the error surface. Whilst this works well if the error surface is convex, this is problematic if it is less well-behaved as the algorithm might get stuck in local minima. Stochastic gradient descent on the other hand, only poorly approximates the gradient with one observation, which slows down its learning. On the other hand, it is able to cope with local minima, as the poor approximation can catapult the algorithm out of these local minima. In line with most of the modern neural network literature, we use mini-batch gradient descent as a compromise between these two extremes, which exhibits both fast convergence and robustness to local minima.

Note that we have written our algorithm in terms of plain mini-batch gradient descent for simplicity of exposition. In our code, we use momentum-based updating algorithms, which are more robust than mini-batch gradient descent. However, since these algorithms

are based on mini-batch gradient descent, the underlying logic and challenges are the same, so that we delegate a discussion of the optimizing details to the Appendix.

At this point the reader might wonder why we split the simulated data into different mini-batches instead of only simulating the number of points needed in the current batch. The rationale for this is that the simulation step often comes with a fixed cost, like. a burn-in period described below, so that it is computationally advantageous to simulate large amounts of data in one step.

1.2.3.4 Generation of Training Data

One crucial element of our algorithm is to draw our training points from their ergodic distribution over the state space. In particular, the ergodic distribution of the state variables in many economic models has mass over only a small subspace of the entire state-space hypercube. We can thus focus our algorithm on this small subspace, enabling us to shrink the hypercube to a significantly smaller subspace. This therefore simplifies the problem for our neural-network approximators as their training points are concentrated in a smaller space, making the underlying function easier to approximate.

Importantly, this reduction in complexity can be achieved at a very low cost. In particular the ergodic distribution can be simulated based on the current approximation of policy functions. This simulation requires only the evaluation of the policy neural networks, which can be done very efficiently, so that this is again an inexpensive process. For every epoch we can thus simulate a large number of points at low cost.[7]

To maximize the probability that our draws are actually made from the ergodic distribution, we use the following simulation procedure. For the first K_1 epochs, we draw points uniformly at random from the entire hypercube. This is necessary since in the beginning the

[7]Note that in some problems the ergodic distribution can even be expressed through analytical expressions, in which case points can be drawn from it even more easily.

neural networks are still ill-converged, so using them for simulation could possibly lead us to pick points only from far-out regions of the state-space. For every subsequent epoch we then draw I random starting points uniformly at random from the state-space hypercube This is important to overcome potential local convergence issues. From each of those points we iterate T steps forward. We drop the first B_1 iterations as a burn-in, since the starting points are random and might thus be outside of the ergodic distribution. After B_1 iterations, it is likely that the simulation has converged towards the ergodic distribution. We keep all remaining iterations, shuffle them randomly and use the resulting set as our training point for the kth epoch. Note that it is important to draw only a limited number of points per epoch so that updates of the ergodic distribution based on updates of the current approximation are quickly incorporated. We provide a detailed exposition of this procedure for one of our applications in the Appendix.

1.2.3.5 Training on Model-Generated Data

Finally, we will discuss differences between our approach and more traditional neural network applications. In particular, readers familiar with neural networks might have noted the absence of training, validation, and test samples as well as the absence of a discussion of overfitting.[8]

The reason for this is that we train our neural network on model-generated data, which means that in principle we can generate arbitrarily large amounts of data to train our model. As shown in the above algorithm, this enables us to use each data point only once in the training process, so that we do not need to be overly concerned about overfitting. In other words, since the neural network does not repeatedly face the same training data, it is im-

[8]Overfitting is a common problem when training on a finite data set, where the network might fit too well to its training data. By doing so, it concentrates more on fitting the individual data points than finding the underlying structure, leading to a weak predictive power beyond the training set.

possible for it to fit every individual data point. The absence of the overfitting problem thus also obviates the need for a validation and test set, which are used to alleviate said problem.

Another consequence of this is that regularization is not as important in our context as in data-science applications. We have experimented with a wide array of regularization techniques, but as expected this did not lead to increase in accuracy but instead to a slight increase in running time.

Beyond this crucial difference, our approach is very similar to data-science applications of neural networks - we train our neural networks using our model-generated data on a carefully-designed loss function.

1.2.4 Comparison to Related Methods

Recently, there has been an increased interest in the use of deep-learning techniques to solve large-state space problems globally. First, recent contributions by Maliar et al. (2019), as well as by Azinović et al. (2019) propose discrete-time algorithms. While similar in spirit in their use of neural networks as global approximators, they face the hurdle of approximating the expectations integral as explained above. Our approach elegantly circumvents this issue by using the analytic solutions provided by the continuous-time limit.

Next, Sirignano & Spiliopoulos (2018) propose a related algorithm in the applied mathematics literature. They are concerned with the general solution to partial differential equations in a wide range of applications. Compared to their method, our algorithm is tailored towards the kind of problems faced in economics. In particular, one of our main contributions is the ability to solve policy functions for which no closed-form solutions are available, an issue very relevant for most standard economic models but not explored in that literature.

Finally, Duarte (2018) is the paper most closely related to ours in that he uses neural network approximators as well as the advantages of the continuous-time formulation. However,

his specific algorithm is different from ours. Whilst we perform gradient descent directly on a loss function composed of HJB and first-order condition errors, he employs a method inspired by the reinforcement-learning literature not utilizing the first-order conditions directly. In addition, he chooses his training points uniformly at random instead of from the simulated ergodic distribution. These changes enable us to solve significantly more complex models. As shown below, one of our applications solves a model that has fifty state variables and requires the solution of multiple policy functions for which no closed-form solutions are available. His most complex application on the other hand includes 10 state variables while also having closed-form solutions available for the single policy function which significantly simplifies the problem.

To illustrate its power, let us now turn to applying our approach to three different settings of increasing complexity. First, a standard neoclassical growth model, then a dynamic capital allocation model and finally a model of migration in the face of economic shocks and uncertainty.

1.3 Neoclassical Growth Model

First, we solve a standard continuous-time neoclassical growth model with closed-form solutions for policy functions.[9] This allows us to illustrate our algorithm clearly in a simple setting. Next, we compare the solutions of this model to the same model but with a neural network representation of the policy functions. We compare both of these solutions to a reference solution attained with traditional finite difference methods as in Achdou et al. (2020).

[9]Closed-form solutions in this context means that conditional on knowing the value function we can derive a closed-form expression for the policy function.

1.3.1 Neoclassical Growth Model with Closed-Form Policy Function

We start with a standard benchmark model, the continuous-time neoclassical growth model. This features closed-form solutions for the policy functions, which allows us to focus our attention on the analysis of the value function approximation. We can then back out the policy function from this approach and compare it to the results of the next step in which we approximate the policy functions themselves with a neural net.

The model takes the canonical form, with a single agent deciding to either save in capital or consume. The HJB equation takes the following form:

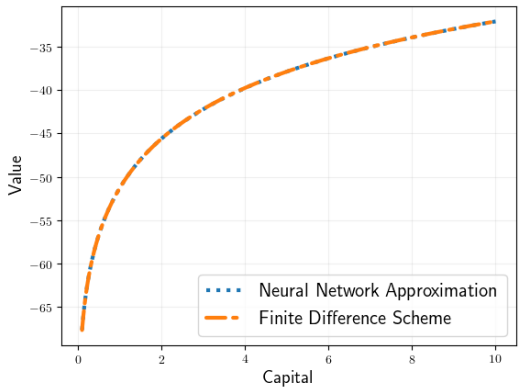
$$\rho V(k) = \max_c U(c) + V'(k)[F(k) - \delta * k - c] \tag{1.4}$$

In this context, one can easily derive an analytical expression for consumption that only depends on the value function and the utility function, namely $c = (U')^{-1}(V'(k))$. With CRRA utility this simplifies further to depend solely on the value function.

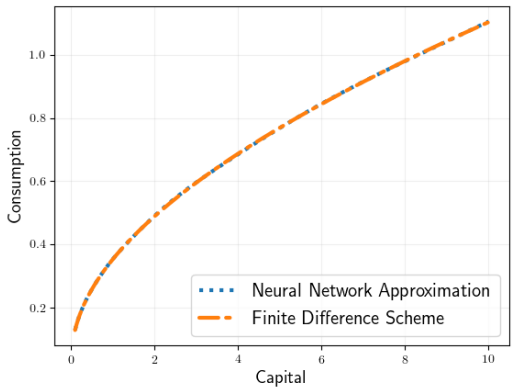
Since the model is purely for illustration of our algorithm we use a standard calibration for all parameters in the model. Utility is CRRA with a constant relative risk aversion of $\gamma = 2$. The discount rate ρ is set to 0.04. The production function takes the form $F(k) = A * k^{\alpha}$. Here total factor productivity A is set to 0.5 and α, which governs returns to scales, is 0.36. Finally, depreciation is set to 0.05, i.e. a 5 percent annual depreciation rate.

We approximate the value function $V(k)$ with a neural network, $\tilde{V}(k;\Theta)$. Our error criterion is the "HJB error" which is defined as:

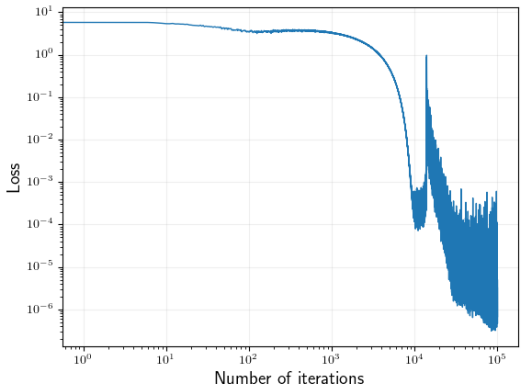
$$err_{HJB} = \rho\tilde{V}(k;\Theta) - U\left((U')^{-1}\left(\frac{\partial\tilde{V}(k;\Theta)}{\partial k}\right)\right)$$



(a) Value Function



(b) Consumption Policy Function



(c) HJB Error

Figure 1.2: Neoclassical Growth Model with Closed-Form Solutions - Results

Notes: Panels (a) and (b) compare the solution of our algorithm for value and policy functions to that of a finite difference scheme, where we use the closed-form solution available for the consumption policy function.. Panel (c) displays the convergence error for the HJB equation.

$$-\frac{\partial \tilde{V}(k;\Theta)}{\partial k}\left[F(k)-\delta * k-(U')^{-1}\left(\frac{\partial \tilde{V}(k;\Theta)}{\partial k}\right)\right]$$

The results are shown in Figure 1.2. In particular, panels (a) and (b) contrast the results for value and policy functions from our neural network approximation with results obtained through a traditional finite difference scheme as in Achdou et al. (2020). Panel (a) displays the value function, which takes the expected concave shape. Panel (b) displays the

consumption policy function, which in this case is directly derived from the value function via the closed-form solution. Importantly, both value and policy function are virtually identical to the reference solution, displaying the high accuracy of our solution method. This is consistent with the low convergence error shown in panel (c). Note that the error does not decrease monotonously, and indeed increases for a subtantial number of iterations. This is in line with the behavior of most machine learning methods and can be intuitively explained as the algorithm finding a promising but wrong solution, which is thus only abandoned after substantial exploration.

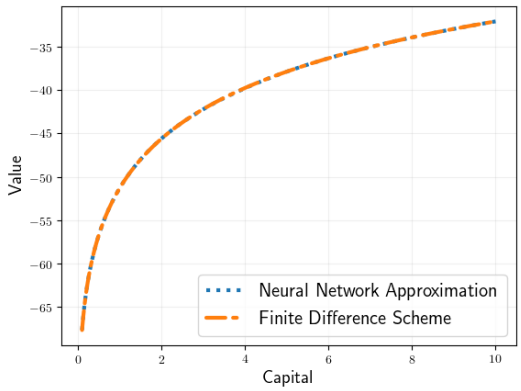
1.3.2 Neoclassical Growth Model with Policy Approximation

As discussed above our algorithm allows us to solve models that do not have closed-form expressions for the policy functions. In order to test the accuracy of our policy-function approximation, we use the same neoclassical growth model as above. However, this time we do not use the closed-form consumption policy function but rather approximate said policy function directly with a policy neural network $\tilde{C}(k;\Theta^C)$. Adjusting our notation for our value function neural network therefore leads to the following HJB error:

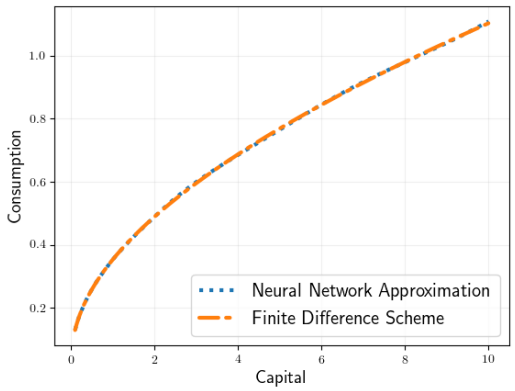
$$err_{HJB} = \rho\tilde{V}(k;\Theta^V) - U\left(\tilde{C}(k;\Theta^C)\right) - \frac{\partial\tilde{V}(k;\Theta^V)}{\partial k}\left[F(k) - \delta * k - \tilde{C}(k;\Theta^C)\right]$$

Similarly we need to include an error that disciplines the policy function for consumption. From the first-order condition for consumption we therefore derive the consumption policy error:

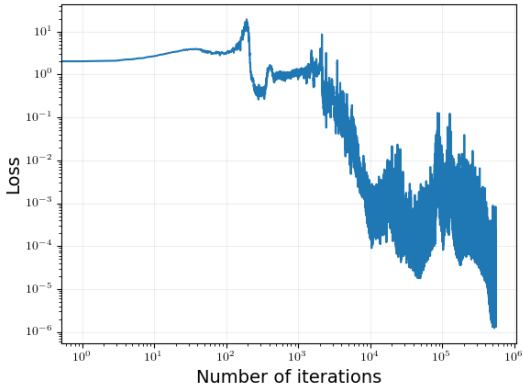
$$err_C = (U')^{-1}\left(\frac{\partial\tilde{V}(k;\Theta)}{\partial k}\right) - \tilde{C}(k;\Theta^C)$$



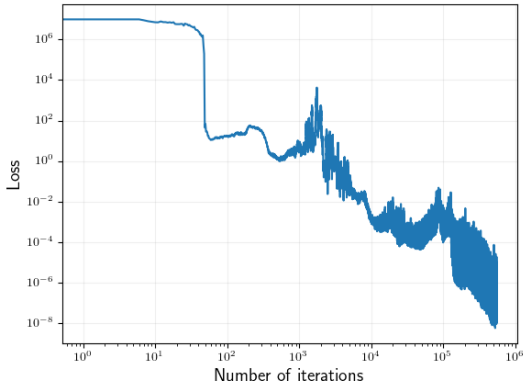
(a) Value Function



(b) Consumption Policy Function



(c) HJB Error



(d) Consumption Policy Error

Figure 1.3: Neoclassical Growth Model with Policy Approximation - Results

Notes: Panels (a) and (b) compare the solution of our algorithm to that of a finite difference scheme, where we approximate the consumption policy function through a neural network. Panel (c) and (d) display the convergence error for the HJB equation and the consumption policy function respectively.

The results for this model are shown in Figure 1.3. Panels (a) and (b) show that the results for this approach are virtually identical to both the finite difference solution and by extension to our previous results. Panels (c) and (d) display the evolution of both the HJB and the consumption policy convergence errors and again confirm the excellent convergence of our algorithm.

1.4 Dynamic Capital Allocation Model

In order to evaluate our algorithm and to showcase its strength in large state spaces let us now turn to a multi-location capital allocation model. The model investigates optimal consumption, investment, and cross-location capital flows in a model of aggregate productivity shocks and capital movement frictions. Here, a social planner who maximizes utility over consumption in L locations faces a dynamic capital allocation problem. Each location exhibits idiosyncratic productivity shocks and capital can be moved across locations subject to a capital and investment adjustment cost. Note that this is a multi-country version of the neoclassical growth model above, where we allow for interactions between the different countries through capital mobility. Indeed, without the adjustment costs the solutions would be trivial as locations would be identical and the aggregate capital stock would be a sufficient statistic for the state of the economy. By introducing the capital frictions we thus create an ideal setting for the application of our algorithm in a simple and well-understood model that is yet untractable with conventional methods.

1.4.1 Model Outline

Time is continuous and there are L locations, each with a representative household. Households have CRRA utility based on local consumption c_ℓ, governed by the relative risk aversion parameter γ, $u(c_\ell) = \frac{c_\ell^{1-\gamma}}{1-\gamma}$. Each location ℓ is characterized by an idiosyncratic and exogenous productivity z_ℓ, as well as capital k_ℓ that jointly determine output y_ℓ through a standard decreasing returns to scale production function:

$$y_\ell = z_\ell k_\ell^\alpha.$$

Productivity evolves according to the following diffusion process:

$$dz_\ell = \left[\left(-\nu \log(z_\ell) + \frac{\sigma^2}{2}\right) z_\ell\right] dt + \sigma\, z_\ell dW_t^\ell$$

Capital depreciates at rate δ and both the movement of capital across locations and capital investment is costly, with the costs given by:

$$C(i_\ell, x_\ell) = \kappa_1 i_\ell^2 + \kappa_2 x_\ell^2.$$

The social planner maximizes the sum of the discounted household utilities in each location, which leads to the following HJB for the social planner's problem:

$$\begin{aligned}\rho V\left(\{k_\ell, z_\ell\}_{\ell=1}^L\right) = \max_{\{c_\ell, i_\ell, x_\ell\}_{\ell=1}^L} \sum_{\ell=1}^L \Bigg[& u(c_\ell) + \partial_{k_\ell} V\left(\{k_\ell, z_\ell\}_{\ell=1}^L\right)\, (i_\ell - \delta k_\ell + x_\ell) \\ & + \partial_{z_\ell} V\left(\{k_\ell, z_\ell\}_{\ell=1}^L\right)\, \mu_\ell(z) + \partial_{z_\ell, z_\ell} V\left(\{k_\ell, z_\ell\}_{\ell=1}^L\right)\, \frac{\sigma_\ell^2}{2}\Bigg]\end{aligned}$$

$$\text{s.t.} \begin{cases} c_\ell + i_\ell + C(i_\ell, x_\ell) = z_\ell k_\ell^\alpha & \forall\, \ell \\ 0 = \sum_{\ell=1}^L x_\ell \end{cases}$$

The flow value of the social planner depends on the instantaneous utility from consumption in each location as well as the evolution of the states. The first constraint represents the feasibility constraint in each location that the social planner has to observe. Output can either be consumed, invested or shipped to another location, subject to adjustment costs. x_ℓ represents the net inflow of capital into location ℓ. Finally, the second constraint forces the social planner's net capital flows to balance, which is required since the global economy

is closed. Note that due to the decreasing returns and consumption smoothing motives, the social planner will move capital between locations.

1.4.2 Complexity

Since we are dealing with finitely many locations, all idiosyncratic shocks are aggregate shocks. Therefore, the social planner needs to keep track of the productivity and capital stock in each location, leading to a state space hypercube of $2 * L$ dimensions.

Moreover, the social planner has to decide how much capital to invest and how much to export for each location. Based on the constraints this amounts to L consumption choices and $L-1$ net capital movement policy functions, each of which have to be approximated by a neural network due to the absence of closed-form solutions.

1.4.3 Calibration

As this model is mainly for illustrative purposes the calibration is not of central importance. Table 1.1 gives an overview of the calibration of the parameters. Risk aversion, time preference and returns to scale are set to standard values. Depreciation is set to 7 percent annually. The adjustment costs for investment and capital movements are symmetric and set to 1, purely for illustrative purposes. The process for productivity is adapted from Moll and Itskhoki (2018).

1.4.4 Implementation

We start by defining our neural networks. First, we approximate the value function of the social planner with a neural network, $V\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta^V\right)$. Similarly, we approximate the optimal policy functions through two neural networks, an investment network $i_\ell\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta_\ell^I\right)$) with L, and a net capital network $x_\ell\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta_\ell^X\right)$ with $L-1$ outputs. We parameterize

Parameter	Value	Description
γ	2.0	risk-aversion parameter
ρ	0.04	time preference parameter
α	0.33	returns to scale
δ	0.07	capital depreciation
κ_1	1.0	adjustment costs for investment
κ_2	1.0	adjustment costs for capital movement
$e^{-\nu}$	0.8	autocorrelation of productivity
σ	0.33	innovation variance of productivity

Table 1.1: Calibration for multi-location capital allocation model

these three networks with Θ_ℓ^V Θ_ℓ^I, and Θ_ℓ^X. Note that we can solve for consumption from the budget constraint.

As above let us next define the error criterion on which we train our learning algorithm. We therefore define the individual errors on each of the key equations of our model. The HJB error in the context of our model is given by:

$$\begin{aligned} err_{HJB} = &- \rho \tilde{V}\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta^V\right) + \sum_{\ell=1}^{L} \Bigg[u(c_\ell) + \partial_{k_\ell} \tilde{V}\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta^V\right) \dot{k}_\ell \\ &+ \partial_{z_\ell} \tilde{V}\left(\{k_\ell, z_\ell\}_{\ell=1}^L, \Theta^V\right) \mu_\ell(z) + \partial_{z_\ell, z_\ell} \tilde{V}\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta^V\right) \frac{\sigma_\ell^2}{2} \Bigg] \end{aligned}$$

Note that consumption and $\dot{k_\ell} \equiv i_\ell - \delta k_\ell + x_\ell$ depend on the policy choices for investment and net capital movements, for convenience of notation we omitted the dependence here. Next we can turn to the errors that discipline the policy functions. Here we rely on the first-order conditions as well as the capital market clearing condition.

$$err_{i_\ell} = c_\ell^{-\gamma}(1 + \kappa_1 * i_\ell\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta_\ell^I\right)) - \partial_{k_\ell} \tilde{V}\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta^V\right) \qquad \forall \ell$$

$$err_{x_\ell} = c_\ell^{-\gamma} \kappa_2 x_\ell \left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta_\ell^X\right) - \partial_{k_\ell} \tilde{V}\left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta^V\right) - c_L^{-\gamma} \kappa_2 x_L \left(\{k_\ell, z_\ell\}_{\ell=1}^L; \Theta_\ell^X\right)$$

$$- \partial_{k_\ell} \tilde{V} \left(\{k_\ell, z_\ell\}_{\ell=1}^{L}; \Theta^V \right) \qquad \forall \ell \neq L$$

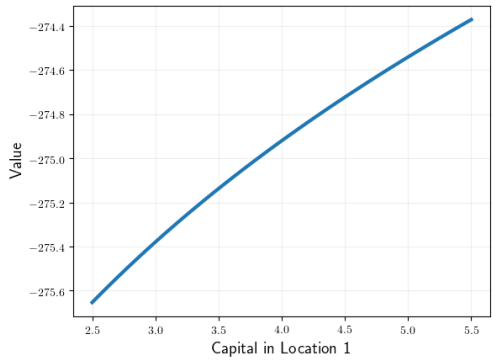
$$err_{clearing} = \sum_{\ell=1}^{L} x_\ell \left(\{k_\ell, z_\ell\}_{\ell=1}^{L}; \Theta_\ell^X \right)$$

1.4.5 Results

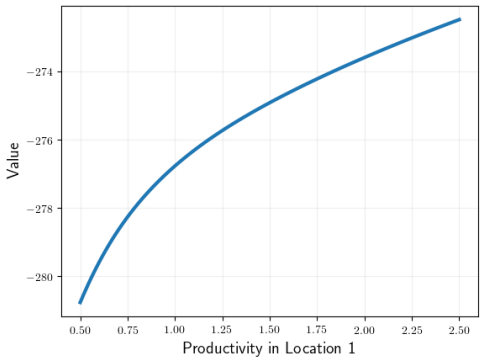
Figure 1.4 displays the basic results of the model for $L = 25$, which corresponds to 50 continuous state variables, namely capital and productivity for each location. Since the value and all policy functions are 50-dimensional objects, we display partial plots by fixing capital and productivity in all dimensions that are not displayed at the midpoint of their display intervals.

Panel (a) plots the value function in location 1 over capital in location 1, panel (b) displays the same value function over productivity in location 1 and panel (c) shows consumption in location 1. All three curves exhibit the expected concave shape, due to diminishing marginal utility. Investment in panel (d) is also concave, but flattens out for large capital values as the incentive to further accumulate capital decreases with an increase in the capital stock. In comparison to the standard neoclassical growth model, this is further enhanced in this setting by the presence of additional insurance through capital movement from all other locations. Panel (e) displays the net capital movement into location 1. Note that since capital in all other locations is fixed at 4.0, location 1 imports capital for $k < 4.0$, and exports capital for $k > 4.0$ due to the diminishing marginal returns of capital. The exact crossing of the net capital policy function at 4.0 is thus a very reassuring confirmation for our results. In the same vein but more importantly, panel (f) confirms convergence as the overall convergence error decreases to about 10^{-5}.

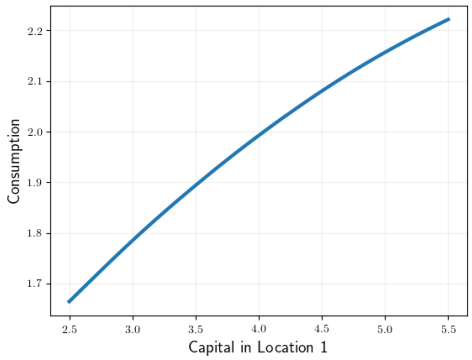
Based on these functions, we can turn to analyzing the dynamics of the model. To do so, let us first compute the deterministic and stochastic steady states. The former is defined



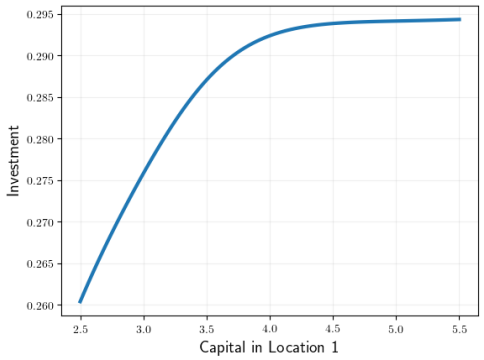
(a) Value over k



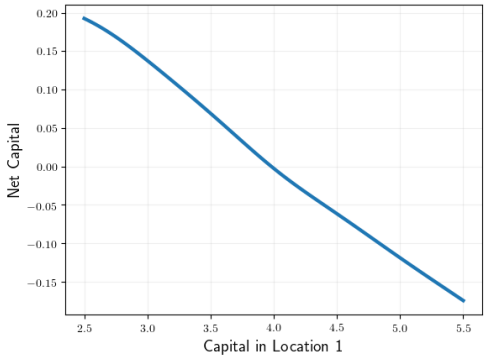
(b) Value over z



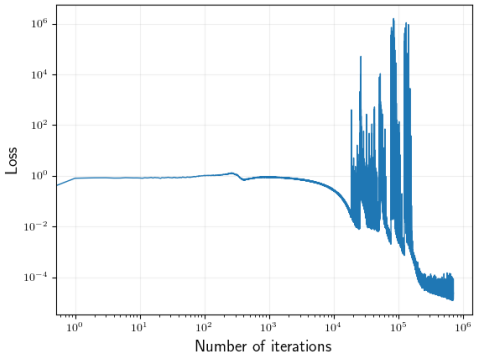
(c) Consumption



(d) Investment



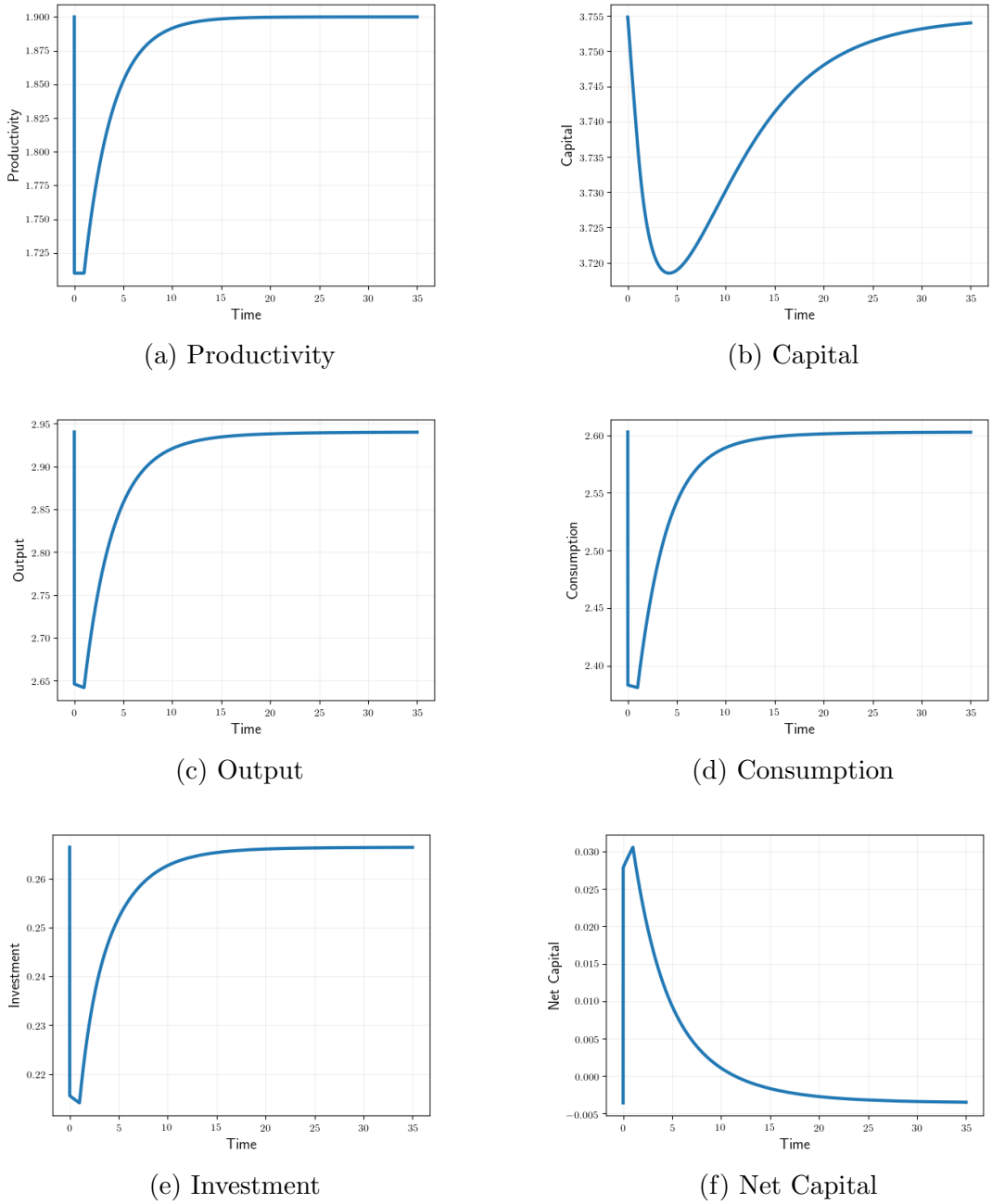
(e) Net Capital



(f) Convergence Error

Figure 1.4: Dynamic Capital Allocation Model - Results

Notes: Results for the 25 location dynamic capital allocation model. We plot partial figures for location 1, with all dimensions not on display fixed at the midpoint of their display intervals.



(a) Productivity

(b) Capital

(c) Output

(d) Consumption

(e) Investment

(f) Net Capital

Figure 1.5: Dynamic Capital Allocation Model - Impulse Reponses

Notes: Impulse responses for the 25 location dynamic capital allocation model to a 10 percent negative productivity shock, which is depicted in panel (a). Panels (b) - (f) dispaly the responses of capital, output, consumption, investment and net capital inflows respectively.

as the steady state if there is not uncertainty in the economy, i.e. $\sigma = 0$. On the other hand the stochastic steady state takes uncertainty into account but the realization of the shock happens to always be zero, i.e. $dW = 0$. In the Appendix we derive that the respective steady states for productivity in each location are $\bar{z}_{DSS} = 1$ and $\bar{z}_{SSS} = 1.90$, whilst the steady states for capital are $\bar{k}_{DSS} = 3.04$ and $\bar{k}_{SSS} = 3.75$. This stresses the value of our global solution model, since the underlying uncertainty significantly drives behavior.

To further analyze model behavior, we consider the impulse responses to a 10 percent negative productivity shock in one location. These are displayed in figure 1.5. Panel (a) displays the evolution of productivity, which drops by 10 percent as the shock hits and then converges back to its steady state with a half-life of two years. Panel (b) shows the response of capital which drops rapidly as the household invests less in order to smooth consumption. After the shock starts to reverse, capital increases back to its steady state, but more slowly as capital accumulation is a gradual process. Panel (c) displays output, which drops in line with productivity and then recovers slightly more slowly due to the slower capital accumulation. Panel (d) displays the drop in consumption, which is significant despite the slowdown in investment and the availability of insurance through capital movement. Panel (e) shows investment which falls as the household smooths consumption. Finally panel (f) displays net capital inflow, which jumps up as the social planner redistributes capital from other locations to make up for the productivity fall and then falls back to the steady-state value of zero.[10]

1.5 Migration and the Propagation of Shocks

As a final application we study the propagation of economic shocks through a multi-country economy with mobile labor and fully rational expectations. This provides an excellent il-

[10]Note that to graphically represent the productivity shock in a continuous time setting, we actually extend the initial productivity shock over a small time interval $[t, t + \tau]$.

lustration for our method, as the model is highly nonlinear and displays kinks in policy functions, which highlights the global solution properties of our method. Additionally, this application is also of great economic interest, as the realistic modeling of migration decisions is understood to be important but has so far eluded the literature due to numerical constraints. Whilst in this paper we focus on this model to demonstrate our method's ability to handle nonlinearities, we return to the economic implications in Fernández-Villaverde et al. (2021).

The importance of including migration decisions stems from the fact that migration is strongly driven by economic factors. This has been shown in the empirical literature (e.g. Molloy et al., 2011, Vogler, 2021). Migration in turn affects a large number of economic variables of interest, so that it is essential to realistically model migration decisions when studying the propagation of shocks.

Due to the inherent complexity of the problem, the previous literature on migration patterns had to contend with simplifying assumptions. In particular, the literature faces four main difficulties. First, it is difficult to model fully rational expectations. For example, Kennan & Walker (2011) restrict agents to only know their economic prospect in one other location. In contrast, our approach allows us to let agents keep track of the entire menu of destinations and their states. This is an important contribution, as Kennan & Walker stress the importance of rational expectations to account for the inherent uncertainty of economic prospects surrounding migration decisions. Second, as argued by Kennan and Walker for a plausible quantification one needs to model a finite number of destinations to be able to speak to the granularity of the data. Previous migration models with fully rational expectations such as Kaplan & Schulhofer-Wohl (2017a) were not able to include a large number of locations due to the curse of dimensionality. Our method is designed for exactly this challenge, and so we are able to include up to 25 different locations, which corresponds to 75 aggregate continuous state variables. Third, as locations are intrinsically heterogeneous,

it is important to accommodate locations that are ex ante heterogeneous. Again, our large state space approach enables this without further difficulty. Lastly, migration choices are inherently forward-looking and are associated with fixed costs of movement. Therefore, it is important to allow for full dynamics in the model. Our model includes this by exhibiting fully forward-looking agents in a dynamic setting.

In conclusion, in order to realistically model migration, a quantitative model should incorporate all four of these aspects - rational expectations, a finite number of destinations, intrinsically heterogeneous locations and forward-looking location decisions. Since these conditions naturally lead to the emergence of large state spaces, other papers had to rely on shortcuts to get around these problems due to methodological constraints. In contrast, our algorithm allows us to include all four elements in our model. To the best of our knowledge, we are thus the first to include all four elements in a fully dynamic rational expectations migration model with a large number of locations.

1.5.1 Model Outline

We now present our multi-country search and matching model with migration frictions, which serves as an illustration of the power of our algorithm as outlined above.

1.5.1.1 Countries

Time is continuous and there are L locations indexed by $\ell \in \{0, ..., L\}$. In the overall economy there is a mass L of workers that are working in the different countries. In particular, there are n_ℓ employed and u_ℓ unemployed workers in country ℓ. The productivity z_ℓ in each country follows a country-specific productivity process that develops according to:

$$dz_\ell = \left[\left(-\nu \log(z_\ell) + \frac{\sigma^2}{2}\right) z_\ell\right] dt + \sigma\, z_\ell dW_t^\ell.$$

A country ℓ is thus characterized by a triple of state variables (z_ℓ, u_ℓ, n_ℓ) and we summarize the overall state of the economy by $\mathbf{s} = \{z_\ell, u_\ell, n_\ell\}_{\ell=1}^{L}$.

1.5.1.2 Migration

Households have the option to migrate between the L countries. In particular, with Poisson rate α a household receives the option to migrate. If a household moves from location ℓ to location ℓ' she incurs welfare costs $\phi + \zeta_{\ell'}$, where ϕ is common across all agents, representing fixed migration costs. $\zeta_{\ell'}$ on the other hand is an idiosyncratic migration cost, which we model as an iid random variable distributed according to the distribution $F(\zeta_{\ell'})$. In order to simplify the algebra later on, we assume a logistic distribution F with mean μ_F and scale s_F. This assumption is very similar to the standard assumptions commonly made in the trade literature, for example in Caliendo et al. (2019). If a household decides to move, she arrives at the new location without a job.[11] We denote the equilibrium rates of movement from location ℓ to location ℓ' by $p^u_{\ell,\ell'}(\mathbf{s})$ and $p^w_{\ell,\ell'}(\mathbf{s})$ depending on the initial employment state of the moving household.

1.5.1.3 Labor Markets

The labor markets in each country are central to our model, since they determine both the wage rate and unemployment in the face of migration. As is conventional in the search literature, we define labor market tightness in country ℓ as

$$\theta_\ell = \frac{v_\ell}{u_\ell},$$

[11] Note that different modeling choices are possible here, e.g. households could also arrive in their destination country with a new job or alternatively with some probability of a new job. We opt for our approach since it simplifies the problem and in the data most migrating households engage in a job search after arrival in the target country.

where v_ℓ denotes the vacancies posted in location ℓ. Given a matching function $m(u_\ell, v_\ell)$, we define the rate of filling a vacancy as

$$q(\theta_\ell) = m\left(\frac{1}{\theta_\ell}, 1\right).$$

We use the commonly employed matching function from Den Haan et al. (2000), that has the convenient property that the matching probabilities are between zero and one:

$$m(u, v) = \frac{uv}{(u+v)^{1/\iota}},$$

where $\iota > 0$ is the matching elasticity. In addition to endogenous job destruction due to migration, jobs are also destroyed exogenously at rate δ. Finally, wages are set according to Nash Bargaining, which we will discuss in more detail below.

1.5.1.4 Household Problem

Households living in this economy discount the future at rate ρ and receive instantaneous CRRA utility $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$ from consumption c. In case a household is unemployed, she generates income b from home production.

Based on the above setup, unemployment in country ℓ evolves according to the following law of motion:

$$\dot{u}_\ell(\mathbf{s}) = \delta_\ell n_\ell + \sum_{\ell' \neq \ell} \left(p^w_{\ell',\ell}(\mathbf{s}) n_{\ell'} + p^u_{\ell',\ell}(\mathbf{s}) u_{\ell'}\right) - \left(\sum_{\ell' \neq \ell} p^u_{\ell,\ell'}(\mathbf{s}) + q(\theta_\ell)\theta_\ell\right) u_\ell.$$

Similarly, employment in country ℓ evolves according to:

$$\dot{n}_\ell(\mathbf{s}) = q(\theta_\ell) v_\ell - \left(\sum_{\ell' \neq \ell} p^w_{\ell,\ell'}(\mathbf{s}) + \delta_\ell\right) n_\ell.$$

The recursive problems of the unemployed and the employed households summarize the household problem. In particular, denoting by $V^u(\hat{\ell};\mathbf{s})$ the value function of the unemployed household in location $\hat{\ell}$ with the economy being in state $\mathbf{s}$, the HJB of an unemployed household is given by:

$$\begin{aligned}\rho V^u(\hat{\ell};\mathbf{s}) =& U(b) + \underbrace{\sum_{\ell=1}^{L}\left[V^u_{z_\ell}(\hat{\ell};\mathbf{s})\mu(z_\ell) + \frac{\sigma(z_\ell)^2}{2}V^u_{z_\ell z_\ell}(\hat{\ell};\mathbf{s})\right]}_{\text{evolution of productivities}} + \underbrace{\sum_{\ell=1}^{L}\left[V^u_{u_\ell}(\hat{\ell};\mathbf{s})\dot{u}_\ell(\mathbf{s}) + V^u_{n_\ell}(\hat{\ell};\mathbf{s})\dot{n}_\ell(\mathbf{s})\right]}_{\text{evolution of (un)employed population}} \\ &+ \underbrace{q(\theta_\ell)\theta_\ell\Big(V^u(\hat{\ell};\mathbf{s}) - V^w(\hat{\ell};\mathbf{s})\Big)}_{\text{job finding}} + \underbrace{\alpha\sum_{\ell'\neq\hat{\ell}}\int_{-\infty}^{\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s})}\Big[V^u(\ell';\mathbf{s}) - V^u(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'} - \zeta_{\ell'}\Big]\,dF\left(\zeta_{\ell'}\right)}_{\text{migration decision}} \\ &\text{s.t. } \bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s}) = V^u(\ell';\mathbf{s}) - V^u(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'}, \\ &\quad \dot{u} \text{ and } \dot{n} \text{ follow their laws of motion}\end{aligned}$$

The equivalent problem for the employed household, with $V^w(\hat{\ell};\mathbf{s})$ denoting her value function, is:

$$\begin{aligned}\rho V^w(\hat{\ell};\mathbf{s}) =& U(w(\hat{\ell};\mathbf{s})) + \underbrace{\sum_{\ell=1}^{L}\left[V^w_{z_\ell}(\hat{\ell};\mathbf{s})\mu(z_\ell) + \frac{\sigma(z_\ell)^2}{2}V^w_{z_\ell z_\ell}(\hat{\ell};\mathbf{s})\right]}_{\text{evolution of productivities}} + \underbrace{\sum_{\ell=1}^{L}\left[V^w_{u_\ell}(\hat{\ell};\mathbf{s})\dot{u}_\ell(\mathbf{s}) + V^w_{n_\ell}(\hat{\ell};\mathbf{s})\dot{n}_\ell(\mathbf{s})\right]}_{\text{evolution of (un)employed population}} \\ &+ \underbrace{\delta_{\hat{\ell}}\Big(V^u(\hat{\ell};\mathbf{s}) - V^w(\hat{\ell};\mathbf{s})\Big)}_{\text{job loss}} + \underbrace{\alpha\sum_{\ell'\neq\hat{\ell}}\int_{-\infty}^{\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s})}\Big[V^u(\ell';\mathbf{s}) - V^w(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'} - \zeta_{\ell'}\Big]\,dF\left(\zeta_{\ell'}\right)}_{\text{migration decision}} \\ &\text{s.t. } \bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s}) = V^u(\ell';\mathbf{s}) - V^w(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'}, \\ &\dot{u} \text{ and } \dot{n} \text{ follow their laws of motion}\end{aligned}$$

1.5.1.5 Firm Problem

In each country there is a forward-looking atomistic representative firm that posts vacancies v_ℓ subject to vacancy-posting costs $\kappa(\theta_\ell)$, where $\kappa(\theta_\ell) = \kappa_0 + \kappa_1 q(\theta_\ell)$ as in Pissarides (2009).[12]

[12]Here $\kappa_1 > 0$ represents the fixed costs of hiring, whilst $\kappa_0/q(\theta_\ell)$ are the marginal costs of hiring which are increasing in the mean duration of vacancies $1/q(\theta_\ell)$. We implement these convex vacancy-posting costs in order to limit firm size in the absence of decreasing returns to scale.

Since firms are owned by households we assume that they also discount the future at rate ρ.[13] The firm produces according to a constant returns to scale production function taking labor as its sole input $f(z_\ell, n_\ell) = z_\ell n_\ell$. Denoting by $\tilde{n}_\ell$ the employment of the firm, its law of motion is given by

$$\dot{\tilde{n}}_\ell(\mathbf{s}) = q(\theta_\ell)\tilde{v} - \left(\sum_{\ell' \neq \ell} p^w_{\ell,\ell'}(\mathbf{s}) + \delta_\ell\right)\tilde{n}_\ell.$$

The firm problem is then summarized by its HJB equation:

$$\rho J(\hat{\ell}, \tilde{n}; \mathbf{s}) = \max_{\tilde{v}\geq 0} z_{\hat{\ell}}\tilde{n}_{\hat{\ell}} - w_{\hat{\ell}}(\hat{\ell}; \mathbf{s})\tilde{n}_{\hat{\ell}} - \kappa\left(\theta_{\hat{\ell}}\right)\tilde{v} + \underbrace{\sum_{\ell=1}^{L}\left[J_{z_\ell}(\hat{\ell}, \tilde{n}; \mathbf{s})\mu(z_\ell) + \frac{\sigma(z_\ell)^2}{2}J_{z_\ell z_\ell}(\hat{\ell}, \tilde{n}; \mathbf{s})\right]}_{\text{evolution of productivities}} +$$

$$+ \underbrace{\sum_{\ell=1}^{L}\left[J_{u_\ell}(\hat{\ell}, \tilde{n}; \mathbf{s})\dot{u}_\ell(\mathbf{s}) + J_{n_\ell}(\hat{\ell}, \tilde{n}; \mathbf{s})\dot{n}_\ell(\mathbf{s})\right]}_{\text{evolution of (un)employed population}} + \underbrace{J_{\tilde{n}_{\hat{\ell}}}(\hat{\ell}, \tilde{n}; \mathbf{s})\dot{\tilde{n}}_{\hat{\ell}}(\mathbf{s})}_{\text{evolution of firm employment}}.$$

1.5.1.6 Equilibrium

The Recursive Competitive Equilibrium of this economy is defined by value and policy functions $\{\tilde{v}_\ell, \tilde{n}_\ell, V^w(\ell), V^u(\ell), J(\ell), \bar{\zeta}^w_{\ell,\ell'}, \bar{\zeta}^u_{\ell,\ell'}\}_{\ell=1}^L$, wages $w(\ell)$, and moving rates $p^w_{\ell,\ell'}$ and $p^u_{\ell,\ell'}$ over the state vector $\mathbf{s} = \{z_\ell, u_\ell, n_\ell\}_{\ell=1}^L$ such that:

1. Given $\{w(\ell; \mathbf{s}), p^w_{\ell,\ell'}, p^u_{\ell,\ell'}\}$ for all $\ell, \ell' \in 1, ..., L$: V^w, V^u and J solve the employed household HJB, the unemployed household HJB and the firm HJB.

2. Labor market clear: $n_\ell = \tilde{n}_\ell$ for all ℓ.

3. Firm and aggregate vacancies are consistent: $v_\ell = \tilde{v}_\ell$ for all ℓ.

[13] Formally, the firm discounts the future at the stochastic discount factor of its owners. Since the firm is fully owned by households and we have representative households without idiosyncratic states, this stochastic discount factor is equal to ρ.

4. The moving rates are given by: $p^w_{\ell,\ell'}(\mathbf{s}) = \alpha F(\bar{\zeta}^w_{\ell,\ell'}(\mathbf{s}))$ and $p^u_{\ell,\ell'}(\mathbf{s}) = \alpha F(\bar{\zeta}^u_{\ell,\ell'}(\mathbf{s}))$.

5. The wage rate $w(\ell;\mathbf{s})$ is consistent with Nash Bargaining.

6. The laws of motion for $\mathbf{s}$ hold.

1.5.1.7 Equilibrium Implications

We can take first-order conditions in the firm problem, which yields

$$\kappa\left(\theta_{\hat{\ell}}\right) = q(\theta_{\hat{\ell}}) J_{n_{\hat{\ell}}}(\hat{\ell};\mathbf{s}).$$

Plugging in our functional form for $\kappa(\theta_\ell)$, this is equivalent to

$$\frac{\kappa_0}{q(\theta_{\hat{\ell}})} + \kappa_1 = J_{n_{\hat{\ell}}}(\hat{\ell};\mathbf{s}),$$

which implicitly defines $\theta_{\hat{\ell}} = \theta(\hat{\ell};\mathbf{s})$. Turning to the determination of wages, we can express the standard Nash Bargaining condition in this context as:

$$w_{\hat{\ell}} = \arg\max_w \left(V^w(\hat{\ell};\mathbf{s}) - V^u(\hat{\ell};\mathbf{s})\right)^\beta \left(J(\hat{\ell};\mathbf{s})\right)^{1-\beta},$$

which yields the following first-order condition:

$$V^w(\hat{\ell};\mathbf{s}) - V^u(\hat{\ell};\mathbf{s}) = \frac{\beta}{1-\beta} J(\hat{\ell};\mathbf{s}).$$

Finally, due to $n_\ell = \tilde{n}_\ell$ for all ℓ in equilibrium, we have $J(\ell,\tilde{n};\mathbf{s}) = J(\ell,\mathbf{s})$, so that the last term in the firm HJB drops out.

1.5.2 Complexity

As above, since we are dealing with a finite set of locations, all idiosyncratic shocks are aggregate shocks. In addition to their own states, agents with fully rational expectations need to keep track of the productivities as well as the employment and unemployment states in all other locations in order to make fully-informed and optimal migration decisions. Therefore, the state space has $3 \times L$ dimensions. Furthermore, this model is highly nonlinear, featuring kinks as we will show below. It is thus not only an interesting economic application but also a great showcase for the global solution capabilities of our method.

1.5.3 Calibration

Besides showcasing the power of our algorithm, the purpose of this model is to study the effects of migration on the propagation of shocks in the presence of multiple countries. Since this application is general instead of being focused on one particular set of countries, we aim for a generally reasonable calibration instead of a country-specific one. Table 1.2 displays our calibrated values for the parameters in our model. As in the dynamic capital allocation model above, we set the risk-aversion parameter γ to 2.0, the time preference parameter ρ to 0.04, and calibrate the productivity process such that it has an autocorrelation of 0.8 and an innovation variance of 0.33. We set the exogenous job destruction rate δ to 4% in the range of values estimated by Davis et al. (2006). The mean of the logistic movement cost distribution μ_F is 0.6 and its scale s_F 1, to match the moving cost distribution in Kennan & Walker (2011). The migration opportunity shock α is set to 0.1. We set the common migration cost ϕ to 0.9 to be 50% higher than the mean idiosyncratic migration shock. The elasticity of the matching function ι is 1.25 as in Den Haan et al. (2000). The Nash bargaining weight of workers β is set to 0.04 as in Petrosky-Nadeau et al. (2018). Regarding home production we set b in line with the labor wedge literature. In particular, we choose an intermediate value

Parameter	Value	Description
γ	2.0	risk-aversion parameter
ρ	0.04	time preference parameter
δ	0.04	exogenous job destruction
$e^{-\nu}$	0.8	autocorrelation of productivity
σ	0.33	innovation variance of productivity
μ_F	0.6	mean of migration cost distribution
s_F	0.6	std of migration cost distribution
α	0.1	migration opportunity shock
ϕ	0.9	common migration cost
ι	1.25	elasticity of matching function
β	0.04	Nash bargaining power of workers
b	0.31	Home Production
κ_0	0.5	Vacancy posting costs 1
κ_1	0.5	Vacancy posting costs 2

Table 1.2: Calibration for multi-location capital allocation model

of 0.31, which corresponds to 0.85 once one normalizes for mean productivity. We choose this value since it is a reasonable value commonly used, that is between the extreme values 0.4 of Shimer (2005) and 0.96 in Hagedorn & Manovskii (2008). Finally, we choose κ_0 and κ_1 to be equal to 0.5 as in Petrosky-Nadeau et al. (2018).

1.5.4 Implementation

We have now assembled all the ingredients to solve the model. First, note that thanks to the assumption of a logistic distribution for the idiosyncratic migration cost, we can derive closed-form solutions for the two integrals in the employed and unemployed household HJBs:

$$\int_{-\infty}^{\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s})} \left[V^u(\ell';\mathbf{s}) - V^u(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'} - \zeta_{\ell'}\right] dF\left(\zeta_{\ell'}\right) = \frac{V^u(\ell';\mathbf{s}) - V^u(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'}}{1 + e^{-\frac{\left(\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s}) - \mu_F\right)}{s_F}}} - \left(\frac{\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s}) e^{\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s})/s_F}}{e^{\mu_F/s_F} + e^{\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s})/s_F}} + \log\left(\frac{e^{\mu_F}}{e^{\mu_F} + e^{\bar{\zeta}^u_{\hat{\ell},\ell'}(\mathbf{s})}}\right)\right),$$

$$\int_{-\infty}^{\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s})} \left[V^u(\ell';\mathbf{s}) - V^w(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'} - \zeta_{\ell'}\right] dF\left(\zeta_{\ell'}\right) = \frac{V^u(\ell';\mathbf{s}) - V^w(\hat{\ell};\mathbf{s}) - \phi_{\hat{\ell},\ell'}}{1 + e^{-\frac{\left(\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s}) - \mu_F\right)}{s_F}}} - \left(\frac{\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s}) e^{\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s})/s_F}}{e^{\mu_F/s_F} + e^{\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s})/s_F}} + \log\left(\frac{e^{\mu_F}}{e^{\mu_F} + e^{\bar{\zeta}^w_{\hat{\ell},\ell'}(\mathbf{s})}}\right)\right).$$

This is crucial, since otherwise we would face an integral approximation problem. As discussed in the Solution section above this is a tremendously difficult problem in high dimensions, which our algorithm elegantly circumvents.

To implement the algorithm we now only need to determine which objects to approximate with neural networks and define an appropriate loss function.

The first step is straightforward. Since we do not have any policy functions without closed-form solutions, we only need to represent the three value functions $J(\ell;\mathbf{s})$, $V^w(\ell;\mathbf{s})$ and $V^u(\ell;\mathbf{s})$ with neural networks $\tilde{J}(\ell,\mathbf{s};\Theta^J_\ell)$, $\tilde{V}^w(\ell,\mathbf{s};\Theta^W_\ell)$ and $\tilde{V}^u(\ell,\mathbf{s};\Theta^U_\ell)$.

The definition of the loss function requires a little more work. First, we use the firm's first-order condition

$$\frac{\kappa_0}{q(\theta_\ell)} + \kappa_1 = \tilde{J}_{n_\ell}(\ell,\mathbf{s};\Theta^J_\ell)$$

to compute $\theta_\ell(\mathbf{s})$, employing $\tilde{J}$. Second, we can invert the HJB of the employed household to derive a closed-form solution for $w(\ell;\mathbf{s})$:

$$w(\ell;\mathbf{s}) = U^{-1}[\rho\tilde{V}^w(\ell,\mathbf{s};\Theta^W_\ell - R)],$$

where R is equal to the entire right-hand side of the employed household's value function with the exception of $U(\cdot)$. Note that everything in this remainder depends only on either θ_ℓ or the three neural networks and can thus be computed efficiently.

Finally, we can define three error criteria that together form the loss function. In particular, we employ the HJB error of the unemployed household $HJB^u(\ell, \mathbf{s}; \Theta_\ell^W, \Theta_\ell^U, \Theta_\ell^J)$, i.e. the difference between its left and right-hand side, and equivalently the HJB error of the firm $HJB^J(\ell, \mathbf{s}; \Theta_\ell^W, \Theta_\ell^U, \Theta_\ell^J)$. The final error term comes from the Nash Bargaining Condition. We can thus define the overall loss function as follows:

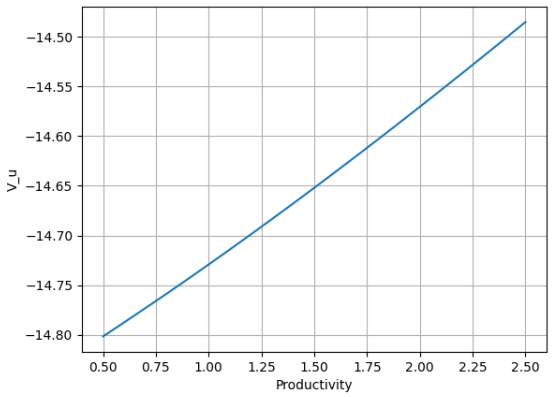
$$\mathcal{L}(\mathbf{s}; \Theta^W, \Theta^U, \Theta^J) = \sum_{\ell=1}^{L} \Bigg[||HJB^u(\ell, \mathbf{s}; \Theta_\ell^W, \Theta_\ell^U, \Theta_\ell^J)||_2 + ||HJB^J(\ell, \mathbf{s}; \Theta_\ell^W, \Theta_\ell^U, \Theta_\ell^J)||_2 + \Big|\Big| \tilde{V}^w(\ell, \mathbf{s}; \Theta_\ell^W) - \tilde{V}^u(\ell, \mathbf{s}; \Theta_\ell^V) - \frac{\beta}{1-\beta} \tilde{J}(\ell, \mathbf{s}; \Theta_\ell^J) \Big|\Big|_2 \Bigg]$$

1.5.5 Results

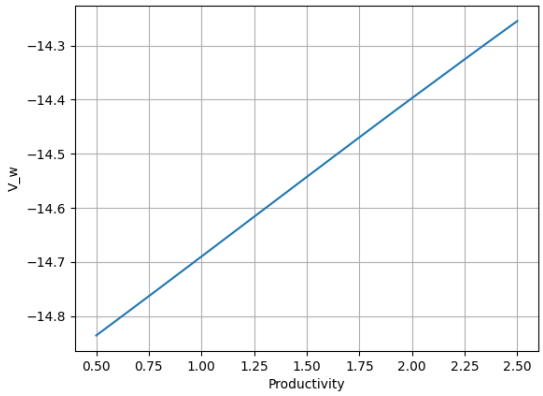
Figure 1.6 displays the results of this model for $L = 25$, which corresponds to 75 continuous aggregate state variables. With this number of locations, this model is significantly more complex than comparable papers in the literature, e.g. Kaplan & Schulhofer-Wohl (2017a) who only use two locations. Therefore, in order to solve it, our solution method is required as the complexity is beyond the scope of other methods.

Panel (a) shows the unemployed household's value function, panel (b) the employed household's value function and panel (c) the firm's value function. These all take the expected shape, namely being close to linear and increasing in productivity. This is the result of the constant returns to scale production function in the firm's problem. This implies that the marginal value of an extra worker to the firm is equal to productivity, which extends to the households' value functions through Nash Bargaining.

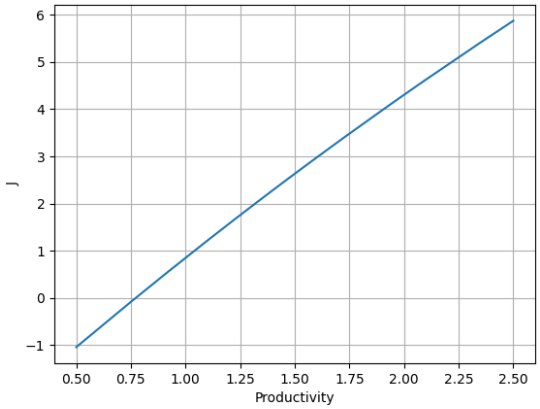
The most interesting graph, however, is panel (d), which displays labor market tightness θ plotted over productivity. Note that since u is fixed at its steady-state value, this plot is equivalent to the number of vacancies posted v. The number of vacancies is equal to zero up to a threshold value and is then increasing and concave. Again, this is the expected



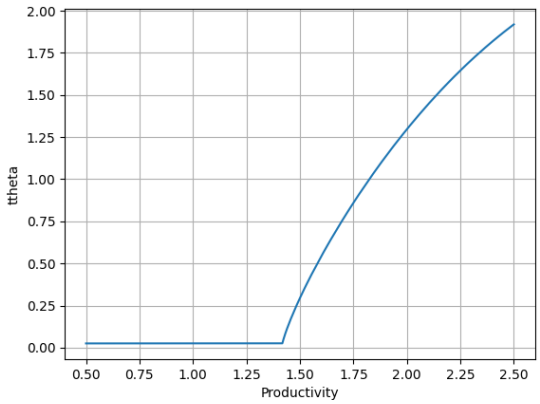
(a) Unemployed household value function



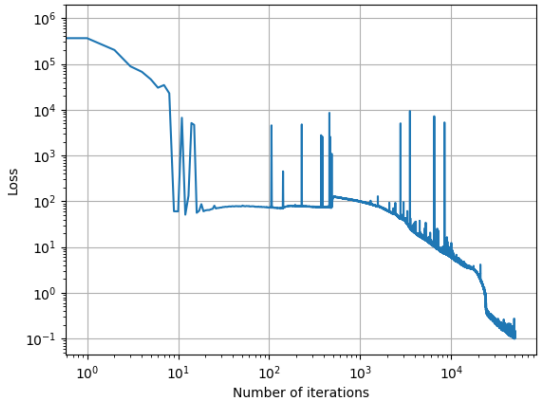
(b) Employed household value function



(c) Firm Value Function



(d) Market Tightness



(e) Loss Function

Figure 1.6: Multi-Country Migration Model - Results

Notes: Results of the multi-country migration model with $L = 25$, corresponding to 75 continuous aggregate state variables. The three value functions and market tightness are plotted over productivity in location 1, with the other states being fixed at their steady-state values.

shape and its underlying logic is intuitive. Given the fixed costs of vacancy creation, firms do not post vacancies for low values of productivity. Once the threshold is crossed, they start posting an increasing number of vacancies. Due to the increasing marginal costs of posting a vacancy this increase is not linear in productivity but instead concave. Critically, this plot displays the inherent nonlinearity in the model. In particular, we have a kink in this essential policy function. For this reason, this model can only be solved with a global solution method, demonstrating the power of our approach to solve very large state space models globally.

Finally, panel (e) displays the evolution of the loss function. Notice that it drops to below 10^{-1}. Analyzing this error, we first note that it is mechanically increasing in the number of locations, since we add more non-negative HJB and Nash Bargaining errors as the number of locations increases. Once we look at the individual HJB errors, they drop to approximately 10^{-4}, which amounts to a remarkable precision given the complexity of the problem.

Overall, these results thus demonstrate that our model is capable of solving a highly nonlinear model with a very large state space with high precision. Whilst we mostly use this model as an illustration of the power of our method here, in future work we will investigate the underlying economics more closely in Fernández-Villaverde et al. (2021).

1.6 Conclusion

In this paper we propose a novel approach for globally solving large state space models. In order to overcome the curse of dimensionality we rely on neural networks to approximate both value and policy functions. By utilizing the advantages of a continuous-time formulation, our algorithm sidesteps the computationally intensive and potentially inaccurate approximation of expectations over the large state space hypercube.

Our algorithm draws on the HJB to discipline the value function, while using the FOCs and additional constraints as convergence errors for the policy functions. This approach is feasible since the otherwise expensive calculation of the various derivatives in these errors is provided efficiently in neural network implementations.

Finally, our algorithm draws on an approximation of the ergodic distribution based on the policy functions to further curb the curse of dimensionality. By focusing only on the manifold in the state space hypercube that is part of the support of the ergodic distribution we can sharpen our approximation of the value and policy functions in the relevant area.

Next we apply our algorithm to three different settings. First we solve the standard neoclassical growth model and compare our results to those of a reference approach using a finite difference scheme. This allows us to investigate the accuracy of our general approach and show that our policy function approximation works well. Next, in order to demonstrate the ability of our approach to handle large scale problems we apply our algorithm to a dynamic model of capital allocation. Here we solve a model with 50 continuous state variables that requires the approximation of 50 policy functions in addition to the value function. Finally, we solve a large scale migration model with 75 continuous aggregate state variables. The inherent non-linearity of this model showcases the global properties of our method. This illustrates the power of our method as this class of complexity was previously unsolvable.

Based on our success in this context our method lends itself to a variety of future applications. Specifically, there is a broad class of models that our algorithm addresses, namely models concerned with granular data and frictions. We refer to these models as "Node" models, where a node is a distinct unit of economic interest, such as an agent, a region, an industry or a firm. Importantly, as long as there are finitely many nodes, shocks to an individual node do not wash out in the aggregate. Consequently, each agent has to keep track of the state of each node in her forward-looking decision problem. This gives rise to large state spaces for which our method is well-suited. We therefore see our methodology

as an avenue to address questions such as how do productivity shocks propagate through production networks or how links between financial institutions can amplify shocks through a financial network.

In future work we will examine the propagation of shocks in a multi-country model with migration more closely. In particular, whilst in the present paper we mainly use the above migration model as an illustration of the power of our solution method, the model is also of inherent economic interest. Specifically, we are interested in understanding what implications a free movement zone like Schengen has on the propagation of shocks across member countries. In addition, we aim to investigate to what extent the size and composition of such a zone matter. With this application and the general class of node models identified above as examples, we are convinced that our new solution method enables future research linking macroeconomic models and granular data.

Chapter 2

Government Policies in a Granular Global Economy[1]

I thank my co-authors Cecile Gaubert and Oleg Itskhoki for their permission to reproduce our work here. This paper is forthcoming at the Journal of Monetary Economics. Below I reproduce the latest working paper version of the article.

2.1 Introduction

Large firms shape the landscape of national economies and even more so the patterns of international trade. Gaubert & Itskhoki (2020) show that "granular" forces (i.e., exports driven by large individual firms) contribute significantly to comparative advantage and trade patterns of countries. In this context, a number of micro-level policies targeted at individual firms may end up having non-trivial aggregate consequences. This is true in particular when these policies impact large exporters. We observe that such policies are common with governments taking policy actions that are sometimes very narrow and appear tailor-

[1]I thank the participants at the Carnegie-Rochester-NYU Conference on Public Policy and the Virtual ITM seminar for helpful comments.

made to target individual firms rather than industries. In particular, antitrust regulation, antidumping policies, and international sanctions all target large individual foreign firms.[2]

In this paper, we study the rationale for and consequences of such "granular" policies. We are particularly interested in the impact these policies may have in a global economy. To do so, we adopt the quantitative model of trade with granular firms developed and estimated in Gaubert & Itskhoki (2020). We use the model to study the general equilibrium impact of national policies in a granular open economy by evaluating their effect on both the welfare of the home country and that of its trade partners. Across sectors, the model features classic Ricardian forces as in Dornbusch et al. (1977); crucially, within sectors, it features a discrete (but potentially large) number of heterogeneous firms that compete oligopolistically. The model is quantified to match patterns of domestic sales and exports of French manufacturing firms. In particular, the distribution of firm size is very skewed, so that some large firms have sectoral market shares well above single digits even in sectors comprised of a large number of home and foreign firms. We use this framework to characterize the qualitative and quantitative properties of various government interventions.

We first shed light on merger policy, by studying the potential merger of two domestic firms who are leaders in their sector. We find that the desirability of such a merger, from the perspective of the home government, depends crucially on how open the economy is. In general, mergers lead to some cost savings and productivity gains; however, they also lead to an increase in monopoly power, which reduces overall efficiency and leads to a transfer of surplus from consumers to producers. In a relatively closed economy, the efficiency loss due to an increase in monopoly power tends to make mergers undesirable.

[2]Recent examples of international antitrust regulations are the 2007 case of the European Commission (EC) against Microsoft Corporation and the 2017 fine imposed by the EC on Google. A very recent case of a granular trade war is the 292% tariff imposed by the US on a particular jet produced by the Canadian Bombardier. "Granular" tactics are particularly widespread in antidumping retaliation (see Blonigen & Prusa, 2008) and international sanctions (as in the recent case of the US against the Chinese ZTE).

However, in a more open economy, mergers lead to a transfer of surplus to the merged home firm which is done, in part, at the expense of the foreign consumer surplus. Therefore, the more open the economy is, the more a given home merger is desirable from the point of view of the home government, especially in sectors where home firms are export champions. Contrasting the attitudes of the home and foreign governments towards the same domestic mergers reveals an international conflict, whereby the home antitrust is excessively lenient towards mergers in comparative advantage sectors. Finally, we compare unilateral optimal decisions in merger policy to the coordinated global planner's solution. Our estimated model suggests that the negative spillover effects on the foreign country and aggregate welfare are significant quantitatively, and are particularly pronounced in the most granular and open sectors. This underlines the need for international cooperation over M&A policies to avoid an excessive build-up of market power.

In a second exercise, we turn to trade policy. We are interested in the incentives faced by governments to impose an import tariff on a single large foreign exporter, rather than imposing it on all firms in a given sector. Narrow trade restrictions and antidumping duties that target a narrow set of firms have indeed been regularly emphasized in the policy debate. To make progress on this question, we assume that the foreign country is passive and does not retaliate. In this context, all unilateral import tariffs shift surplus towards the home economy, and the corresponding cost is born by the foreign country (including the deadweight loss of the policy).

The question that we focus on is whether the breadth of tariff imposition matters for revenue-equivalent sets of tariffs. Specifically, we compare the welfare effects of imposing two different tariff schemes designed to raise the same revenue — a granular tariff and a sector-wide tariff. We find that, in a granular world, a country prefers to impose an import tariff on the largest foreign exporter, rather than imposing a uniform tariff on all sectoral imports. This is particularly true in sectors where its trade partner enjoys a granular

comparative advantage. The reason is that by taxing the largest foreign firm, a country takes advantage not only of the general equilibrium terms-of-trade effect, operating via a reduction in the foreign wage rate, but also of the industry level terms-of-trade improvement due to the incomplete pass-through of the tariff as a result of markup adjustment by the large foreign firm.

In the last exercise, we study industrial policies that subsidize national champions. We show that they are generally suboptimal in closed economies due to an excessive build-up of market power, yet may become welfare improving when used unilaterally in open economies.

Related literature We contribute to the literature that studies the influence of large individual firms on macro aggregates, referred to as "granularity" in the macro literature following Gabaix (2011). This literature typically focuses on the positive question of what share of aggregate fluctuations is driven by idiosyncratic productivity shocks (see Acemoglu et al., 2012; Carvalho & Gabaix, 2013; Carvalho & Grassi, 2019; Grassi, 2017) or what share of trade flows can be traced to firm-level shocks (see di Giovanni & Levchenko, 2012; di Giovanni et al., 2014; Gaubert & Itskhoki, 2020). We complement these studies by investigating, in contrast, the normative policy implications of granularity.

Our study is related to the vast literature on trade policy and market structure, summarized in Helpman & Krugman (1989), Brander (1995) and Bagwell & Staiger (2004). In particular, early contributions that study profit-shifting motives for trade policy under oligopoly include Dixit (1984), Brander & Spencer (1984) and Eaton & Grossman (1986). These papers focus on stylized models with homogeneous firms. A more recent literature explores how optimal trade policy is impacted by the presence of heterogeneous firms that self-select into exporting as in Melitz (2003) (see Demidova & Rodríguez-Clare, 2013; Felbermayr et al., 2013; Bagwell & Lee, 2020; Haaland & Venables, 2016). These analyses all rely on monopolistically competitive models and study a uniform tariff imposed on all firms.

Closer to what we do, Costinot et al. (2020) study non-uniform tariffs imposed on heterogeneous firms in a Melitz (2003) framework. They keep the focus on a setup with monopolistic competition with constant markups; in contrast, our quantitative analysis features strategic interactions between firms and variable markups. Our contribution to this literature is to study granular trade policy in a quantitative model of oligopolistic competition with many firms, which captures the salient features of the market structure of modern manufacturing industries.

We also contribute to the literature studying merger policy. In international trade, merger policy is often viewed as part of the toolkit that policymakers use to affect foreign market access (see e.g. Bagwell & Staiger, 2004, Chapter 9), yet systematic studies of merger policies in this context still remain scarce. Early contributions, focusing typically on homogeneous firms, have expanded the merger analysis from a closed economy context studied in IO to open economies (Barros & Cabral, 1994; Head & Ries, 1997), as well as considered merger and trade policy jointly (Horn & Levinsohn, 2001; Richardson, 1999; De Stefano & Rysman, 2010). Closer to our quantitative analysis, Breinlich et al. (2019) use a two-country trade model with oligopolistic competition to study, in partial equilibrium, when merger policy designed to minimize local consumer prices is too tough or too lenient from the viewpoint of the foreign country. We set up our analysis in general equilibrium and consider a more general welfare function, which accounts for producer surplus in addition to consumer surplus. We decompose the contribution of each channel to the total welfare effects of mergers.

The rest of the paper is structured as follows. Section 2.2 reviews the theoretical framework of Gaubert & Itskhoki (2020) and its quantification, which serve as a basis for the policy analysis that follows. Section 2.3 lays out how we evaluate granular policies in general equilibrium in this context. Section 2.4 studies merger policy, Section 2.5 examines import tariffs and Section 2.6 discusses industrial policy. Section 2.7 concludes.

2.2 A Quantified Granular Model

This section sets up a granular model of international trade, following Gaubert & Itskhoki (2020), and discusses its quantification.

2.2.1 Theoretical Framework

We study a two-country multi-sector model, which combines a Ricardian Dornbusch et al. (1977) model across sectors with the Eaton et al. (2012) model of granular firms within each sector.[3] The two countries are Home and Foreign, which represents the rest of the world. Households inelastically supply L and L^* units of labor at Home and in Foreign, respectively, with L/L^* measuring the relative size of the home country. We describe first the laissez-faire equilibrium in an economy without government policies.

Preferences There is a unit continuum of sectors $z \in [0,1]$, with a *finite* number of product varieties $i \in \{1,..,K_z\}$ in each sector. The numbers of varieties offered in each country, K_z at home and K_z^* abroad, are endogenous, as described further below.

Households have Cobb-Douglas preferences over consumption in each sector, which is itself a CES aggregate of each product variety, that is:

$$Q = \exp\left\{\int_0^1 \alpha_z \log Q_z \, \mathrm{d}z\right\} \qquad \text{with} \qquad Q_z = \left[\sum\nolimits_{i=1}^{K_z} q_{z,i}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{2.1}$$

and $\int_0^1 \alpha_z \mathrm{d}z = 1$. Parameters α_z measure expenditure shares on sectors $z \in [0,1]$ and $\sigma > 1$ is the within-sector elasticity of substitution, common across sectors.

[3] The Eaton et al. (2012) model is a granular version of the Melitz (2003) model in its Chaney (2008) formulation. The model nests as special cases both the DFS-Melitz model, as firms become infinitesimal, as well as the Ricardian DFS model, as varieties of products become perfect substitutes and fixed costs tend to zero.

Using the properties of the CES demand aggregator, consumer expenditure on variety i in sector z in the home market is given by:

$$r_{z,i} \equiv p_{z,i}\, q_{z,i} = s_{z,i}\, \alpha_z Y \qquad \text{with} \qquad s_{z,i} \equiv \left(\frac{p_{z,i}}{P_z}\right)^{1-\sigma}, \tag{2.2}$$

where $p_{z,i}$ is the price of variety i in sector z, $P_z = \left[\sum_{i=1}^{K_z} p_{z,i}^{1-\sigma}\right]^{1/(1-\sigma)}$ is the sectoral price index, $s_{z,i}$ is the within-sector market share of the product variety, and Y is aggregate income (expenditure) at Home.

Production Each firm produces a distinct product variety using local labor as input. The technology has constant returns to scale, $y_{z,i} = \varphi_{z,i}\, \ell_{z,i}$, where $\varphi_{z,i}$ denotes the productivity of the firm that produces product i in sector z. The output of the firm can be marketed domestically as well as exported. Exports incur an iceberg trade cost $\tau \geq 1$. The marginal cost of supplying the home market is therefore constant and equal to:

$$c_{z,i} = \begin{cases} w/\varphi_{z,i}, & \text{for a home variety,} \\ \tau w^*/\varphi^*_{z,i}, & \text{for a foreign variety,} \end{cases} \tag{2.3}$$

where w and w^* are the home and foreign wage rates, respectively. Symmetrically, the marginal cost of serving the foreign market is denoted $c^*_{z,i}$. In each market, we sort all potential sellers in increasing order of their marginal cost $c_{z,i}$ ($c^*_{z,i}$ in foreign). The index i therefore refers to the marginal cost ranking of a firm in a given market, so that the same firm is in general represented by different indexes in different markets.

To access a given market, firms have to pay a fixed cost F in units of labor of the destination country. The fixed cost is independent of the origin of the firm. As a result, the differential selection of domestic and foreign firms into the local market is driven only by

iceberg trade costs, not by a differential fixed cost to access the market, in order to simplify equilibrium characterization.

Productivity draws We denote with M_z a potential (shadow) number of domestic products in sector z. M_z is the realization of a Poisson random variable with parameter $\bar{M}_z$. Each of the M_z potential entrants takes an *iid* productivity draw $\varphi_{z,i}$ from a Pareto distribution with shape parameter θ and lower bound $\underline{\varphi}_z$. Given this structure, the combined parameter:

$$T_z \equiv \bar{M}_z \cdot \underline{\varphi}_z^{\theta} \tag{2.4}$$

is a sufficient statistic that determines the expected productivity of a sector. Intuitively, a sector is more productive either if it has more potential entrants ($\bar{M}_z$) or if the average productivity of a potential entrant (proportional to $\underline{\varphi}_z$) is high. Symmetrically, foreign expected sectoral productivity in sector z is T_z^*, so that the ratio T_z/T_z^* determines the expected relative productivity of the two countries in sector z, a measure of the *fundamental* comparative advantage.

Market structure In a given market, entrants play an oligopolistic price setting game, following Atkeson & Burstein (2008). Firms are large in their own sectors, so that they internalize their influence on the sectoral price index P_z when they maximize profits. On the other hand, they are infinitesimal for the economy as a whole, hence take economy-wide aggregates (w, Y) as given.

We consider both a Bertrand and a Cournot oligopolistic setting.[4] The Nash equilibrium in these oligopolistic games is a markup price setting rule:

$$p_{z,i} = \mu_{z,i} \cdot c_{z,i}, \qquad \text{where} \qquad \mu_{z,i} \equiv \frac{\varepsilon_{z,i}}{\varepsilon_{z,i} - 1} \tag{2.5}$$

$$\text{and} \qquad \varepsilon_{z,i} \equiv \varepsilon(s_{z,i}) = \begin{cases} \sigma(1 - s_{z,i}) + s_{z,i}, & \text{under Bertrand,} \\ \left[\sigma^{-1}(1 - s_{z,i}) + s_{z,i}\right]^{-1}, & \text{under Cournot,} \end{cases} \tag{2.6}$$

with the market share of the firm $s_{z,i}$ defined in (2.2), and $\varepsilon_{z,i} \in [1, \sigma]$ measuring the effective elasticity of residual demand for the product of the firm. Both Bertrand competition in prices and Cournot competition in quantities result in the same qualitative patterns: the elasticity of residual demand $\varepsilon_{z,i}$ is decreasing in the firm's market share $s_{z,i}$, so that in turn the markup $\mu_{z,i}$ increases with the firm's market share.

Given the set of entrants and their marginal costs $\{c_{z,i}\}_{i=1}^{K_z}$, the equilibrium, characterized by each firm's price and market share, is unique and has the property that prices $p_{z,i}$ increase with marginal costs $c_{z,i}$, while markups $\mu_{z,i}$ and market shares $s_{z,i}$ decrease with $c_{z,i}$. We turn next to characterizing the endogenous set of entrants in each market.

Entry Firms enter a market if they make a positive profit there. Profits from serving the home market, for instance, are given by:[5]

$$\Pi_{z,i} \equiv \Pi_z(s_{z,i}) = \frac{s_{z,i}}{\varepsilon(s_{z,i})} \alpha_z Y - wF, \tag{2.7}$$

[4] Formally, in the case of Bertrand competition, for example, the profit maximization problem of firm i in the home market is to choose its price $p_{z,i}$ such that $\Pi_{z,i} = \max_{p_{z,i}} \left\{ (p_{z,i} - c_{z,i}) \frac{p_{z,i}^{-\sigma}}{\sum_{j=1}^{K_z} p_{z,j}^{1-\sigma}} \alpha_z Y - wF \right\}$, taking as given the prices of its competitors $\{p_{z,j}\}_{j \neq i}$ and (w, Y). Under Cournot, the firm instead takes the quantities of its competitors $\{q_{z,j}\}_{j \neq i}$ defined by (2.2) as given.

[5] Notice that due to the linearity of the production function, each firm's profit maximization problem is separable across markets, and hence can be considered one market at a time.

where given the markup pricing (2.5), the elasticity $\varepsilon(s_{z,i})$ equals the ratio of revenues $s_{z,i}\alpha_z Y$ to operating profits (before subtracting the fixed cost), and $\Pi_{z,i}$ monotonically increases in $s_{z,i}$.

The setup detailed above opens the possibility of multiplicity in entry patterns. In order to select a unique equilibrium, we consider a sequential entry game. Specifically, in each market separately, firms with lower marginal costs of serving the market move first and decide whether or not to enter. With this equilibrium selection, the entry game has a unique cutoff equilibrium, so that only firms with marginal costs below some cutoff enter the market. This entry game results in the equilibrium number of entrants K_z, with $\Pi_{z,i} \geq 0$ for all $i \in \{1, .., K_z\}$, and such that $\Pi_{z,K_z+1} < 0$ if the next firm were to enter. Similarly, K_z^* is the number of firms that enter in the foreign market, yet the sets of firms entering in the two markets are in general not the same.

General equilibrium The general equilibrium vector contains wage rates and aggregate incomes (w, w^*, Y, Y^*) such that labor markets clear in both countries and aggregate incomes equal aggregate expenditures (which implies trade balance). In particular, in the home country the latter condition holds if:

$$Y = wL + \Pi, \tag{2.8}$$

where Π are aggregate profits of all home firms distributed to home households:

$$\Pi = \int_0^1 \left[\sum\nolimits_{i=1}^{K_z} \iota_{z,i} \Pi_z(s_{z,i}) + \sum\nolimits_{i=1}^{K_z^*} (1 - \iota_{z,i}^*) \Pi_z^*(s_{z,i}^*) \right] \mathrm{d}z, \tag{2.9}$$

with the profit function $\Pi_z(s_{z,i})$ defined in (2.7); the indicator function $\iota_{z,i} \in \{0, 1\}$ is 1 if firm i is of local origin in the home market, while $\iota_{z,i}^*$ plays the same role for the foreign market.

Labor market clearing requires that aggregate labor income wL equals the total expenditure of all firms on domestic labor:

$$wL = \int_0^1 \left[\alpha_z Y \sum\nolimits_{i=1}^{K_z} \iota_{z,i} \frac{s_{z,i}}{\mu(s_{z,i})} + \alpha_z Y^* \sum\nolimits_{i=1}^{K_z^*} (1 - \iota_{z,i}^*) \frac{s_{z,i}^*}{\mu(s_{z,i}^*)} + wFK_z \right] \mathrm{d}z. \tag{2.10}$$

The three terms on the right-hand side of (2.10) correspond, respectively, to expenditure on domestic labor for (i) production for the domestic market, (ii) production for the foreign market, and (iii) entry of firms in the domestic market (of both home and foreign origin). A parallel market clearing condition to (2.10) holds in the foreign country. We normalize $w = 1$ as the numeraire.

Conditional on the sectoral equilibrium vectors $\mathbf{Z} \equiv \left\{ K_z, \{s_{z,i}\}_{i=1}^{K_z}, K_z^*, \{s_{z,i}^*\}_{i=1}^{K_z^*} \right\}_{z \in [0,1]}$, the vector of general equilibrium quantities $\mathbf{X} \equiv (w, w^*, Y, Y^*)$ solves conditions (2.8), (2.10) and their foreign counterparts.[6] In turn, given the aggregate equilibrium vector $\mathbf{X}$, the solution to the entry and price-setting game in each country-sector yields the sectoral equilibrium vector $\mathbf{Z}$. The resulting fixed point $(\mathbf{X}, \mathbf{Z})$ is the equilibrium in the granular economy.

2.2.2 Fundamental and granular comparative advantage

In our analysis, we focus on two sector-level characteristics: (i) a measure of the overall sectoral comparative advantage, and (ii) a measure of its granular component that extracts the contribution to comparative advantage of the idiosyncratic productivities of large firms. The first is captured by the sectoral *home share* abroad:

$$\Lambda_z^* \equiv \frac{X_z}{\alpha_z Y^*} = \sum\nolimits_{i=1}^{K_z^*} (1 - \iota_{z,i}^*) s_{z,i}^*, \tag{2.11}$$

[6] One of the four aggregate equilibrium conditions is redundant by Walras Law, and is replaced by the numeraire normalization. Also note that in the closed economy conditions (2.8) and (2.10) are equivalent, and amount to $Y/w = \bar{\mu}[L - FK]$, where $K = \int_0^1 K_z \mathrm{d}z$ is the total number of firms serving the home economy and $\bar{\mu} = \left[\int_0^1 \alpha_z \sum_{i=1}^{K_z} s_{z,i}/\mu(s_{z,i}) \right]^{-1}$ is the (harmonic) average markup.

where X_z is total home exports and $\alpha_z Y^*$ is total foreign absorption in sector z. Therefore, Λ_z^* equals the cumulative market share of all home firms serving the foreign market, capturing the export stance of home in sector z. It is a random variable that depends on the market shares (hence the realized productivity draws) of Home firms in sector z in the foreign market. Conveniently, this statistic is an observable measure of the effective comparative advantage of home in sector z, irrespective of the source of this comparative advantage.

In our granular model, this sectoral outcome Λ_z^* is driven by two forces: an expected value based on sectoral characteristics and the contribution of idiosyncratic firm draws around this expected value. Specifically, the expectation of firms' productivities in the sector is formally pinned down by the *fundamental* comparative advantage of the sector, T_z/T_z^*. Given T_z/T_z^*, the *expected* home share abroad is:

$$\mathbb{E}\{\Lambda_z^*\} = \frac{1}{1 + (\tau\omega)^\theta \cdot T_z^*/T_z}. \tag{2.12}$$

Across sectors, this expected share is increasing in T_z/T_z^*, while in all sectors it decreases with trade costs τ and the relative home wage $\omega \equiv w/w^*$.

Furthermore, as our model features granularity, the home share Λ_z^* is shaped in part by idiosyncratic realizations of firm productivities, especially when the underlying productivity distribution is fat-tailed. As a result, the realized export stance of a country-sector may differ markedly from its expected value, driven by a handful of firms with outsized productivity draws. We therefore decompose Λ_z^* into its expected value based on sectoral characteristics and the contribution of idiosyncratic firm draws around this expected value:

$$\Lambda_z^* = \mathbb{E}\{\Lambda_z^*\} + \Gamma_z^* \tag{2.13}$$

We refer to $\Gamma_z^* \equiv \Lambda_z^* - \mathbb{E}\{\Lambda_z^*\}$ as the *granular residual* that captures departures of realized comparative advantage from its expectation, driven by outstanding firms (or the lack thereof). In what follows, we study how welfare benefits of government policies vary across sectors with different overall and granular comparative advantage.

2.2.3 Model quantification

The model is estimated to match salient features of French firm-level data on domestic and export sales, and in particular their variation across 119 manufacturing industries, as discussed in detail in Gaubert & Itskhoki (2020). To quantify the model, we first parameterize the distribution of fundamental comparative advantage across sectors as a log-normal distribution:

$$\log\left(T_z/T_z^*\right) \sim \mathcal{N}(\mu_T, \sigma_T^2), \tag{2.14}$$

where the mean parameter μ_T controls home's absolute advantage and the dispersion parameter σ_T controls the extent of the fundamental comparative advantage. With this parametrization, in order to quantify the model, we need to estimate the six parameters of the model, $\Theta \equiv (\sigma, \theta, \tau, F, \mu_T, \sigma_T)$, as well as calibrate the Cobb-Douglas shares α_z. We describe the estimation procedure and moment fit in Appendix B.1, and we report the estimated parameters in Table 2.1, which we use as the benchmark in our quantitative policy counterfactuals.

Table 2.1: Estimated parameters

Parameter	Estimate	Std. error	Auxiliary variables	
σ	5	—	$\kappa = \frac{\theta}{\sigma-1}$	1.096
θ	4.382	0.195		
τ	1.342	0.101	w/w^*	1.130
F $(\times 10^5)$	1.179	0.252	L^*/L	1.932
μ_T	0.095	0.150	Y^*/Y	1.710
σ_T	1.394	0.190	Π/Y	0.180

We point out a few features of the estimated parameters. First, the combined parameter $\kappa \equiv \theta/(\sigma - 1)$ controls the Pareto shape of the sales distribution, and hence the strength of granular forces. It is estimated to equal 1.096, corresponding to a somewhat *thinner* tail relative to Zipf's law (see Gabaix, 2009). We estimate μ_T to be positive, albeit small, implying an overall mild productivity advantage of French firms relative to their average foreign competitor, consistent with a French wage rate w which is 13% higher than that in a typical French trade partner w^*. The estimated value of $\sigma_T = 1.39$ is large, suggesting that a one standard deviation increase in fundamental comparative advantage across sectors corresponds to a four-fold increase in the relative productivity T_z/T_z^*. We find that the iceberg trade costs are $\tau = 1.34$, broadly in line with the estimates in the literature (see Anderson & van Wincoop, 2004).[7] Finally, the model implies an aggregate share of profits in GDP, Π/Y, equal to 18%, broadly in line with the national income accounts.

Armed with the estimated model, we are ready to study the impact of various policies in this granular environment.

2.3 Evaluating Granular Policies

A range of policies specifically target large firms, and our model is well-suited to study the impact of such policies on trade flows and the welfare of countries. Indeed, this question cannot be analyzed using standard 'continuous' trade models where, even in the presence of heterogeneity, every firm is infinitesimal. In contrast, in a granular model, firms are large and their response to policy can affect sectoral productivity and trade flows. In the rest of the paper we explore in turn three different granular policies, while in this section we outline

[7] The estimated model implies that France is about two times smaller than the rest of the world in terms of population. This is, of course, an abstraction of a two-country model with a common iceberg trade cost τ separating the two regions. The appropriate interpretation of L^*/L in the model is the relative size of the rest of the world in which the individual countries are discounted by their economic distance to France (i.e., if countries trade little with France, their population weight is heavily discounted).

the general methodology we follow to compute and decompose the welfare effects of these policies.

Welfare decomposition The welfare of a representative consumer at home is given by $\mathbb{W} = Y/P$, where Y is aggregate home income and $P = \exp\left\{\int_0^1 \alpha_z \log P_z \mathrm{d}z\right\}$ is the home price index. In general, aggregate income can be decomposed as $Y = wL + \Pi + TR$, where wL is labor income, Π are aggregate profits defined in (2.9), and TR are government policy revenues distributed lump-sum to workers. Since labor is supplied inelastically and the home wage is the numeraire, the log-change in home welfare in response to a policy can be expressed as follows:[8]

$$\hat{\mathbb{W}} \equiv \mathrm{d}\log\frac{Y}{P} = \frac{\mathrm{d}\Pi}{Y} + \frac{\mathrm{d}TR}{Y} - \int_0^1 \alpha_z \mathrm{d}\log P_z \mathrm{d}z. \tag{2.15}$$

The three components in (2.15) correspond to the respective changes in producer surplus, government revenues, and consumer surplus in general equilibrium.

We are interested in the general equilibrium impact of policies targeted at large firms, and in particular, in contrasting the effects they have in granular versus non-granular sectors.[9] The direct partial equilibrium effect of a policy change in sector z is $\left[\frac{\mathrm{d}\Pi_z + \mathrm{d}TR_z}{\alpha_z Y} - \mathrm{d}\log P_z\right]\alpha_z \mathrm{d}z$ and has an order of magnitude $\alpha_z \mathrm{d}z$. In addition, the general equilibrium impact of the policy on (w, w^*, Y, Y^*) is also of the order $\alpha_z \mathrm{d}z$, which in turn affects every sector $z' \in [0,1]$, and hence needs to be taken into account on par with the direct effect.[10]

[8] Note that the change in the real wage is fully accounted for by the changes in the price level P since the nominal wage $w = 1$ by our choice of the numeraire; otherwise, there would be an additional term $\frac{wL}{Y}\mathrm{d}\log w$.

[9] Note that all sectors are *ex ante* symmetric (aside from their expenditure size α_z); however, *ex post*, sectors differ both in their fundamental comparative advantage T_z/T_z^* and in the realized granular productivity draws, which makes some sectors more granular than others (i.e., featuring relatively larger top firms).

[10] Formally, the general equilibrium system composed of (2.8)–(2.10) and their foreign counterparts can be written as a system of linear equations in the equilibrium vector (w, w^*, Y, Y^*). The coefficients in this system are integrals of the sectoral allocations, and they shift by on order of magnitude $\alpha_z \mathrm{d}z$ when sector z is shocked, resulting in a corresponding shift in the equilibrium vector.

Given that the model features a continuum of sectors, a policy that impacts a sector in isolation has an infinitesimal aggregate effect (indeed, proportional to $\alpha_z \mathrm{d}z$). Thus, to capture the overall quantitative general equilibrium effect of a policy, we study a given policy change in a positive measure of sectors Z that have similar levels of granularity and comparative advantage. Concretely, we bin sectors into percentiles of the granular residual Γ_z^* defined in (2.13) or into percentiles of the realized export share Λ_z^* defined in (2.11). We then compute the corresponding welfare impact $\hat{\mathbb{W}}_Z = \mathrm{d}\log(Y/P)$ of the policy in bin Z, and report its *average* aggregate welfare effect, normalized by the size of set Z, given by:

$$\hat{W}_Z = \frac{1}{\int_{z\in Z} \alpha_z \mathrm{d}z} \hat{\mathbb{W}}_Z. \tag{2.16}$$

In the limit, as sets Z become tight around individual sectors z, the aggregate welfare change $\hat{\mathbb{W}}$ can be decomposed into sectoral contributions $\hat{W}_z$:[11]

$$\hat{\mathbb{W}} = \int_0^1 \alpha_z \hat{W}_z \mathrm{d}z. \tag{2.17}$$

In this sense, $\hat{W}_z$ are the general-equilibrium aggregate welfare gradients of sectoral policies.

Finally, we consider the decomposition of the average welfare effects $\hat{W}_Z$ into the contribution of changes in the consumer and producer surplus, according to (2.15). Quantitatively, given the linearity of (2.15), a welfare impact $\hat{W}_Z$ equal to 0.01 is equivalent to the welfare effect of a 1% sectoral productivity improvement or, equivalently, a 1% reduction in the sectoral price level holding income constant. We report the welfare effects in these units.

Welfare approximation To guide intuition, we complement the quantitative evaluation of policies with an analytical approximation, which for tractability we carry out in partial

[11] Formally, consider the aggregate effect $\hat{\mathbb{W}}$ of a policy profile $\boldsymbol{\varsigma} \equiv \{\varsigma_z\}_{z\in[0,1]}$, where ς_z corresponds to a policy implemented in sector z; (2.17) is the decomposition of $\hat{\mathbb{W}}$ into the weighted average of the general equilibrium welfare effects $\hat{W}_z$ of individual sectoral policies ς_z.

equilibrium holding w and Y constant and taking the set of active firms K_z as given. Typically, both the entry/exit margin and general equilibrium effects partially mute the aggregate response to policy changes, without changing the direction of the overall effect. We set up the welfare approximation in a closed economy, and then show how to accommodate an open economy and the various specific policies studied below.

We consider a shock (e.g., a merger, an industrial policy, or a tariff) which leads productivity $\varphi_{z,i}$ and/or markup $\mu_{z,i}$ to change for a subset of firms, which in turn triggers an equilibrium adjustment of prices by all firms in the market. We denote by $\hat{\varphi}_{z,i}$ the proportional change in productivity. The markup is endogenous and we write its proportional change as:

$$\hat{\mu}_{z,i} = \hat{\varepsilon}_{z,i} - \frac{\varkappa_{z,i}}{1-\varkappa_{z,i}}(\hat{p}_{z,i} - \hat{P}_z), \tag{2.18}$$

where by (2.6) and (2.2) the markup $\mu_{z,i}$ is a decreasing function of the relative price $p_{z,i}/P_z$, and we denote its elasticity with $\frac{\varkappa_{z,i}}{1-\varkappa_{z,i}} \geq 0$, and $\hat{\varepsilon}_{z,i}$ is an exogenous markup shifter.[12] Combining (2.18) with the price setting equation (2.5) in log changes, $\hat{p}_{z,i} = \hat{\mu}_{z,i} + \hat{c}_{z,i}$, we can solve for the equilibrium log deviation of the firm's price:

$$\hat{p}_{z,i} = (1-\varkappa_{z,i})\big[\hat{\varepsilon}_{z,i} - \hat{\varphi}_{z,i}\big] + \varkappa_{z,i}\hat{P}_z, \tag{2.19}$$

where we used $\hat{c}_{z,i} = -\hat{\varphi}_{z,i}$, since the wage rate w is held constant. Thus, $\varkappa_{z,i} \in [0,1)$ is the strategic complementarity elasticity with respect to competitor prices P_z, and $1-\varkappa_{z,i}$ is the cost pass-through elasticity (see Amiti et al., 2019).

The log differential of the sectoral price index is given by:

$$\hat{P}_z = \sum\nolimits_{i=1}^{K_z} s_{z,i}\hat{p}_{z,i} = \frac{1}{1-\bar{\varkappa}_z}\sum\nolimits_{i=1}^{K_z} s_{z,i}(1-\varkappa_{z,i})\big[\hat{\varepsilon}_{z,i} - \hat{\varphi}_{z,i}\big], \tag{2.20}$$

[12] For example, a merger between producers i and j (studied in Section 2.4) leads the new firm to set markup based on the cumulative market share of the two products, which changes the perceived elasticity in (2.6) from $\varepsilon(s_{z,i})$ to $\varepsilon(s_{z,i}+s_{z,j})$. Then, formally, we have $\hat{\varepsilon}_{z,i} = \varepsilon'(s_{z,i})\cdot s_{z,j}$ and symmetrically $\hat{\varepsilon}_{z,j} = \varepsilon'(s_{z,j})\cdot s_{z,i}$.

where the second equality substitutes in (2.19) and solves the fixed point for $\hat{P}_z$ where $\bar{\varkappa}_z \equiv \sum_{i=1}^{K_z} s_{z,i}\varkappa_{z,i}$. Note that $-\hat{P}_z$ captures the change in consumer surplus, which declines if the weighted average of exogenous markup shifts outweighs productivity increases, resulting in a higher sectoral price level.

The producer surplus can in turn be approximated as:[13]

$$\frac{\mathrm{d}\Pi_z}{\alpha_z Y} \approx \sum\nolimits_{i=1}^{K_z} \frac{s_{z,i}}{\mu_{z,i}}\hat{\mu}_{z,i} = \sum\nolimits_{i=1}^{K_z} \frac{s_{z,i}}{\mu_{z,i}}\left[\hat{\varphi}_{z,i} + (1-\varkappa_{z,i})[\hat{\varepsilon}_{z,i} - \hat{\varphi}_{z,i}] + \varkappa_{z,i}\hat{P}_z\right]. \tag{2.21}$$

That is, producers capture a portion of the surplus from increased productivity, as they pass-through the rest with lower prices to consumers, and also benefit from increased markups. Note that the exogenous increase in markups for the firms directly affected by the shock is captured by $(1-\varkappa_{z,i})\hat{\varepsilon}_{z,i}$, while $\varkappa_{z,i}\hat{P}_z$ is the indirect markup effect from strategic complementarities in price setting for every firm in the industry.

Combining (2.20) and (2.21), we can write the change in the combined surplus (welfare) as:

$$\hat{W}_z = \frac{\mathrm{d}\Pi_z}{\alpha_z Y} - \hat{P}_z = \sum\nolimits_{i=1}^{K_z} \frac{s_{z,i}}{\mu_{z,i}}\hat{\varphi}_{z,i} - \frac{\bar{\mu}_z - 1}{\bar{\mu}_z}\hat{P}_z, \tag{2.22}$$

with $\hat{P}_z$ given by (2.20) and $\bar{\mu}_z > 1$ measuring the average markup in the economy;[14] and if the policy leads to government revenues, $\hat{W}_z$ additionally features a $\frac{\mathrm{d}TR_z}{\alpha_z Y}$ term. Intuitively, (2.22) splits the overall change in welfare into the direct effect of the average productivity improvement (weighted by the cost shares $s_{z,i}/\mu_{z,i}$) and the redistributive effect from the movement in the price level in the presence of markup pricing $\bar{\mu}_z > 1$ (a wedge between consumer prices and producer costs resulting in a Harberger's triangle deadweight

[13] Indeed, using (2.7), we have $\frac{\mathrm{d}\Pi_z}{\alpha_z Y} = \mathrm{d}\left[\sum_{i=1}^{K_z} s_{z,i}\left(1-\frac{1}{\mu_{z,i}}\right)\right] = \sum_{i=1}^{K_z}\pi_{z,i}\hat{s}_{z,i} + \sum_{i=1}^{K_z}\frac{s_{z,i}}{\mu_{z,i}}\hat{\mu}_{z,i}$, where $\pi_{z,i} \equiv \frac{\Pi_{z,i}+wF}{\alpha_z Y} = \frac{s_{z,i}}{\varepsilon_{z,i}}$. By the definition of market shares, $\sum_{i=1}^{K_z} s_{z,i}\hat{s}_{z,i} = 0$, and therefore $\sum_{i=1}^{K_z}\pi_{z,i}\hat{s}_{z,i}$ is a second-order covariance term, which we omit. Then (2.21) is obtain with the use of (2.18) and (2.19).

[14] Specifically, $\bar{\mu}_z = \left[\frac{\bar{\varkappa}_z}{\bar{\mu}'_z} + \frac{1-\bar{\varkappa}_z}{\bar{\mu}''_z}\right]^{-1}$, where $\bar{\mu}'_z \equiv \frac{\sum_{i=1}^{K_z} s_{z,i}\varkappa_{z,i}}{\sum_{i=1}^{K_z}\frac{s_{z,i}}{\mu_{z,i}}\varkappa_{z,i}}$ and $\bar{\mu}''_z = \frac{\sum_{i=1}^{K_z} s_{z,i}(1-\varkappa_{z,i})[\hat{\varepsilon}_{z,i}-\hat{\varphi}_{z,i}]}{\sum_{i=1}^{K_z}\frac{s_{z,i}}{\mu_{z,i}}(1-\varkappa_{z,i})[\hat{\varepsilon}_{z,i}-\hat{\varphi}_{z,i}]}$.

loss). Indeed, a change in $\hat{P}_z$ has a non-zero welfare effect, as it redistributes surplus between consumers and producers, yet starting from an already distorted equilibrium. In an economy without markups ($\mu_{z,i} \equiv 1$), the overall welfare effect simply reduces to the average improvement in productivity, $\hat{W}_z = \sum_{i=1}^{K_z} s_{z,i}\hat{\varphi}_{z,i}$. This summarizes the approach we take to the welfare approximation that we use in the context of specific policies in an open economy below.

2.4 Mergers, Acquisitions and Antitrust

We are now ready to analyze quantitatively a series of policies typically targeted at large firms, with antitrust policy regulating mergers of firms with significant market power being an obvious example. Specifically, we are interested in the merger of *large* domestic firms and under which conditions they can increase welfare at Home. Since these large domestic firms are typically also large exporters (Melitz, 2003), such a merger likely has nontrivial implications for the foreign country as well. Does Foreign face an incentive to block the mergers of domestic superstar firms? How much does trade openness shape these considerations?

2.4.1 Merger analysis: setup

Typically, firms engage in merger and acquisition activities in order to realize cost synergies, to increase efficiency by transferring knowledge and best practices between entities, but also to increase their market power. When evaluating the desirability of such mergers, the policymaker typically trades-off the risk of increased market power in the economy against the efficiency benefits associated with the merger. We capture these channels as follows. First, we assume that, upon merging, the new firm continues to produce the two distinct product

lines previously produced by the two separate firms, but it now sets prices to maximize the combined profits. As a consequence, the new firm's market power and markups increase.[15]

Second, we allow the merger to result in efficiency gains, as the new firm can reduce both fixed and variable costs. Specifically, we assume that the merged firm incurs only one fixed cost rather than two; we note however that this assumption is largely inconsequential quantitatively, as fixed costs are a very small fraction of revenues for the largest firms. Finally, we allow the merger to generate productivity spillovers between the merged entities: the less productive product inherits some of the efficiency of the more productive one, with the strength of the spillover governed by the parameter $\varrho \in [0, 1]$. Specifically, the productivity of the post-merger product, $\varphi'_{z,2}$, is parameterized as:

$$\varphi'_{z,2} = \varrho\varphi_{z,1} + (1 - \varrho)\varphi_{z,2},$$

where $\varphi_{z,i}$ is the productivity of the pre-merger firm $i \in \{1, 2\}$.

Given this post-merger market structure and productivity distribution, we solve for the new entry game and price-setting equilibrium in each sector. To get at the full welfare effect of a merger, we simulate it for a positive measure of sectors $z \in Z$, and recompute the corresponding general equilibrium, as discussed in Section 2.3.

In our baseline analysis, we consider an economy where $\tau = 1.34$, as estimated in Table 2.1 for France, hence reflecting a fairly high level of trade openness typical of modern developed economies. Harder to calibrate is the value of the productivity spillover: we choose a value of $\varrho = 0.35$, that is a merger closes a third of the productivity gap between the first and the second product. With this value of the spillover parameter we can illustrate some of the most interesting policy trade-offs at play. We report the sensitivity of the analysis to

[15]Given CES demand, the optimal markups are the same for both products and depend on their cumulative market share $s'_{z,1} + s'_{z,2}$ in the new equilibrium, according to the same functional relationship as in (2.5) and (2.6).

alternative values of the merger spillovers as well as the trade openness below. The rest of the model parameters correspond to their benchmark values described in Table 2.1. We report results under Cournot competition for concreteness; the results under Bertrand competition are qualitatively similar, albeit somewhat attenuated quantitatively.

2.4.2 Welfare implications of a merger

The welfare consequences — at home and abroad — of merging the top two domestic firms in a given sector are considerably different across sectors, as we report in Figure 2.1. We split all sectors into deciles $z \in Z$ based on their overall comparative advantage (Λ_z^*; left panel) and its granular component (Γ_z^*; right panel). We then report the welfare impact $\hat{W}_Z$ (as defined in (2.16)) from a merger of the top two home firms in each sector in these deciles for Home and Foreign.

For the bottom 80% of sectors, be it in terms of comparative advantage Λ_z^* or in terms of granularity Γ_z^*, both Home and Foreign benefit from these top mergers, due to the productivity spillover. In each case, welfare gains are more modest for Foreign than for Home, as the market shares of home firms are smaller in the foreign market due to trade costs. In stark contrast, in sectors with the strongest (top 20%) comparative advantage or level of granularity, welfare gains are considerably larger for Home and significantly negative for Foreign.

What are the mechanisms behind the starkly heterogeneous results of Figure 2.1? A merger has two effects on economic outcomes. First, it increases productivity due to the spillover ϱ, which results in lower consumer prices. However, it also increases market power and markups, which results in higher consumer prices and higher firm profits. This second effect also has an international distributive consequence in an open economy, as part of the reduction in Foreign consumer surplus is redistributed towards Home through increased

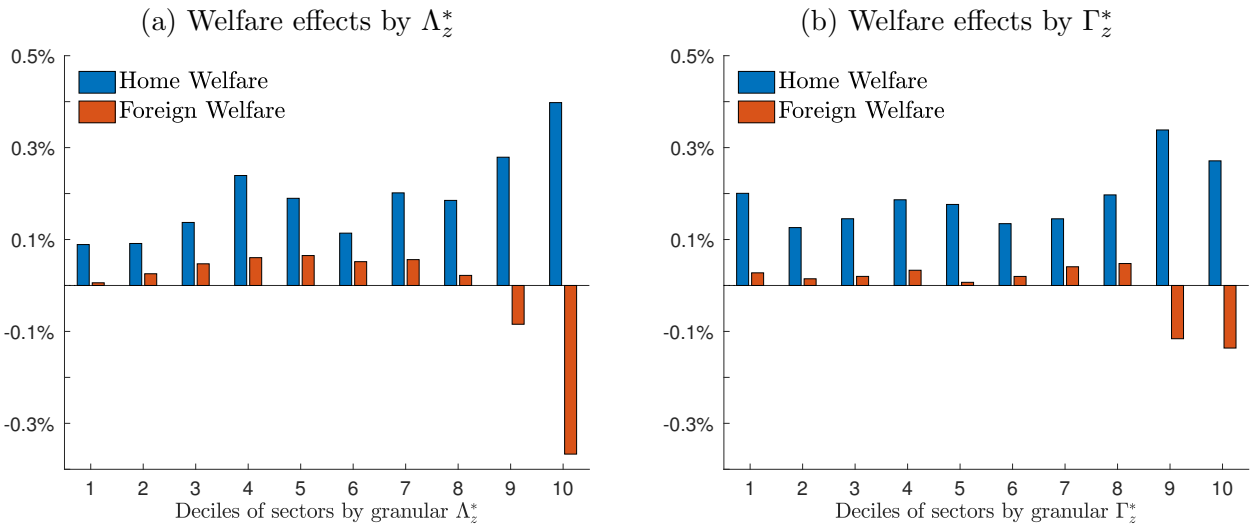


Figure 2.1: Welfare effects of mergers

Note: Welfare impact $\hat{W}_Z$, as defined in (2.16), at home and abroad, of a merger of the top two domestic firms, by deciles Z of sectors $z \in [0,1]$ sorted by overall home comparative advantage Λ_z^* and its granular component Γ_z^*.

profits (producer surplus). This is one reason for the differential welfare effects of the same merger in the two countries.

More specifically, in low-Λ_z^* or Γ_z^* sectors, the largest domestic firms tend to account for relatively small market shares, and especially so in the foreign market. In the model, consistent with the empirical evidence on decreasing cost pass-through with firm size (see Amiti et al., 2019), markups are a convex function of market shares. As a result, the market power increase from a merger is less important in sectors where merged firms have relatively small market shares. Overall, the positive productivity effect (that scales proportionally with market shares) dominates the relatively more modest increase in market power (which is convex in market shares). Furthermore, the net gains are felt more strongly at Home, where the top two domestic firms have a more significant presence than abroad.

Matters are different in the top sectors in terms of both granularity and overall comparative advantage. In the foreign country, top domestic mergers now have significant negative welfare effects. The increased monopoly power of the top domestic firm destroys consumer

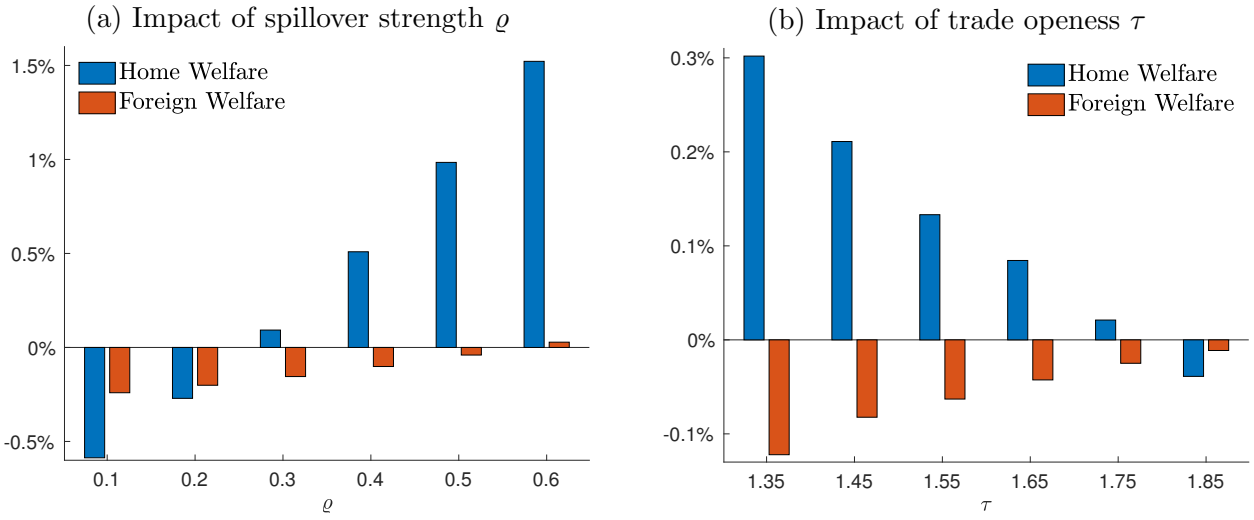


Figure 2.2: Impact of top mergers in comparative advantage sectors: the role of ϱ and τ

Note: Welfare impact of mergers in the top 20% of sectors in terms of home granularity Γ_z^*. See notes to Figure 2.1.

surplus, and is not sufficiently compensated for by productivity gains of the merged firm. When the increase in markups dominates the reduction in costs, foreign consumers lose surplus due to increasing prices. The same effect plays out in the home market as well, and Home's consumer surplus declines too, but, crucially, it is more than offset by the increase in the producer surplus of the merged domestic firms. Indeed, the merged entity increases profits both in the home and the foreign market but channels them back to domestic workers. Through this mechanism, mergers in an open economy have a "beggar-thy-neighbor" spillover on the trading partners. This suggests a rationale for governments in open economies to be overly lenient towards domestic mergers, especially in more granular industries with strong comparative advantage, a topic we explore further below.

Overall, the welfare consequences of a top domestic merger are not trivial. In this baseline calibration, a merger between the top two firms in sectors with the strongest comparative advantage has the same welfare effect as a uniform 0.4% sectoral productivity increase at home and a 0.35% sectoral productivity reduction abroad.

Of course, these numbers hinge importantly on the strength of the productivity spillover ϱ. We report their sensitivity to the intensity of the cost savings realized by a merger in the left panel of Figure 2.2, focusing on mergers in the top 20% of sectors in terms of Home's granularity Γ_z^*.[16] In short, greater ϱ increases welfare gains from mergers, both for Home and Foreign, but it does so much more strongly for Home, where the market shares of these firms are higher. If spillovers are limited, the mergers are particularly detrimental at Home, as the dominant impact comes from the increased market power and markups of the combined entity. The effect is also negative abroad, but is less pronounced. In contrast, as productivity spillover grows larger, Home starts to strongly benefit from mergers, while Foreign benefits much less and only when spillovers are very strong. The reason again is the transfer of foreign consumer surplus into the profits of domestic firms.

The level of trade openness, governed by the iceberg trade cost τ, is also crucial to understand the results in Figure 2.1. At the current level of openness, or for even more open economies, the mergers in high export intensive sectors tend to be beneficial at home but detrimental abroad. In contrast, if we study countries that are sufficiently closed (large τ), we see in the right panel of Figure 2.2 that the consumer loss at home will outweigh the producer gain, resulting in a net loss in domestic welfare, while effects on foreign welfare are still negative yet very limited, because countries trade little. This highlights the essential role of trade openness for the welfare analysis of granular mergers. Domestic mergers are particularly contentious abroad if they happen in granular comparative advantage sectors, especially if countries have a strong trade relationship.

Welfare approximation The quantitative effects discussed above can be conveniently illustrated using the welfare approximation introduced in (2.20)–(2.21) in Section 2.3. Since a merger "shocks" the productivity and markups of only the largest combined home firm

[16]The results are qualitatively the same when we focus on the top 20% of sectors in terms of Home's comparative advantage Λ_z^*.

directly, the consumer surplus effect is given by $-\hat{P}_z = s'_{z,1}\frac{1-\varkappa_{z,1}}{1-\bar{\varkappa}_z}[\hat{\varphi}_{z,1} - \hat{\varepsilon}_{z,1}]$, where $s'_{z,1}$ corresponds to the combined market share of the merged firm, and $\hat{\varepsilon}_{z,1}$ and $\hat{\varphi}_{z,1}$ capture the corresponding markup and productivity shifts.[17] A parallel expression captures the change in the consumer surplus abroad, with $s^{*\prime}_{z,1}$ and $\hat{\varepsilon}^*_{z,1}$ now corresponding to the combined market share and markup change in the foreign market. In general, due to trade costs, $s^{*\prime}_{z,1} < s'_{z,1}$, and thus we expect $\hat{\varepsilon}_{z,1} > \hat{\varepsilon}^*_{z,1}$, explaining why the effects on consumer surplus in Foreign are typically both muted and less adverse in most sectors.

The fall in consumer surplus in Home is partially or fully compensated by the increase in the producer surplus of home firms serving both home and foreign markets:

$$\frac{\mathrm{d}\Pi_z}{\alpha_z Y} \approx \frac{s_{z,1}}{\mu_{z,1}}[\varkappa_{z,1}\hat{\varphi}_{z,1} + (1-\varkappa_{z,1})\hat{\varepsilon}_{z,1}] + \frac{s^*_{z,1}}{\mu^*_{z,1}}[\varkappa^*_{z,1}\hat{\varphi}^*_{z,1} + (1-\varkappa^*_{z,1})\hat{\varepsilon}^*_{z,1}].$$

The presence of this direct producer gain for home firms, absent for foreign firms, explains the international distributional conflict over domestic mergers between the home and the foreign governments. This conflict is particularly pronounced when $s^*_{z,1}$ and $\hat{\varepsilon}^*_{z,1}$ are large in the foreign market. Provided that trade costs are low, this occurs in sectors with pronounced domestic comparative advantage and granularity, consistent with the quantitative patterns in Figures 2.1 and 2.2.

2.4.3 Optimal M&A policy

We have seen above that countries may be impacted quite differently by the mergers of large firms. In practice, large mergers of multinational firms may be evaluated by the antitrust authority of each country in which those firms operate, not only in the country that hosts

[17] Formally, in this case, $\hat{\varphi}_{z,1} = \frac{s_{z,2}}{s_{z,1}+s_{z,2}}\log\frac{\varphi'_{z,2}}{\varphi_{z,2}}$ and $\hat{\varepsilon}_{z,1} = \frac{s_{z,1}s_{z,2}}{s_{z,1}+s_{z,2}}[\varepsilon'(s_{z,1}) + \varepsilon'(s_{z,2})]$, where $(s_{z,1}, s_{z,2})$ are the pre-merger market shares of the two products. Thus, $\hat{\varphi}_{z,1}$ and $\hat{\varepsilon}_{z,1}$ are the sales weighted averages of changes in productivity and markups of the two merged products. Note that the direct effect on consumer surplus, $s'_{z,1}[\hat{\varphi}_{z,1} - \hat{\varepsilon}_{z,1}]$, is modified by the endogenous markup adjustment of all firms, captured by the multiplier $\frac{1-\varkappa_{z,1}}{1-\bar{\varkappa}_z}$.

the headquarters of these firms. We now examine the incentives and optimal policies of each country when assessing merger proposals.

We call a *merger policy* a simple binary policy choice of whether or not to allow a merger. We consider a general objective function $\hat{W}_z + \lambda \hat{W}_z^*$, indexed by $\lambda \geq 0$, the relative weight of foreign welfare. Recall that $\hat{W}_z$ is the domestic welfare impact of a merger defined in (2.16) and $\hat{W}_z^*$ is the corresponding welfare impact abroad. The case with $\lambda = 0$ captures the objective function of the home government, $\lambda = \infty$ the objective of Foreign, and $\lambda = L^*/L$ corresponds to a utilitarian global planner. We define a corresponding merger policy function:

$$m_\lambda(z) \equiv \mathbf{1}\{\hat{W}_z + \lambda \hat{W}_z^* > 0\},$$

i.e. an indicator function defined over the set of sectors $z \in [0,1]$ with $m_\lambda(z) = 1$ if the merger of the top two domestic firms in z is welfare-increasing for a given value of λ.[18] We study the properties of $m_\lambda(z)$ for different values of λ, tracing out the international spillover effects of domestic antitrust policy, focusing on the baseline case with productivity spillover $\varrho = 0.35$.

Figure 2.3 plots a first broad summary of the results. Specifically, it shows the fraction of sectors where a merger is beneficial, $\int_0^1 m_\lambda(z)\mathrm{d}z$, for Home, Foreign and for a utilitarian global planner, $\lambda \in \{0, \infty, L^*/L\}$. These statistics change as a function of trade openness τ, ranging from a closed economy to free trade. Interestingly, in an economy that is sufficiently closed (high τ), the foreign country benefits most from domestic mergers, as their market power impact is very limited abroad.[19] In contrast, the home government blocks the majority

[18] We only focus on a potential merger of two home firms, mostly for technical reasons: it is computationally easier to evaluate a single fixed point problem in terms of $m_\lambda(z)$ and the world GE vector (w, w^*, Y, Y^*), rather than simultaneously solving for a Nash equilibrium in both $m_0(z)$ at home and $m_\infty(z)$ abroad. However, such a focus likely leads to little loss of generality in our analysis, as we anticipate the optimal M&A policy to be approximately a dominant strategy independently of the M&A policy abroad. We leave this conjecture for future quantitative evaluation. Another interpretation is that we focus on M&A in a large country, which trades with a continuum of small open economies, where M&A within each individual country is quantitatively inconsequential.

[19] Note the difference with Figure 2.2, where we focused on the top-20% of sectors in terms of home export share, while in the current analysis we look at merger policy across *all* sectors.

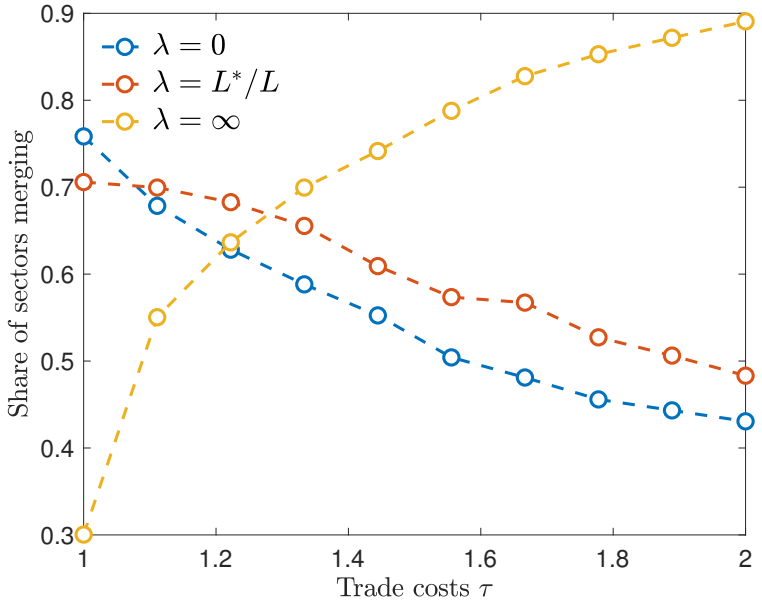


Figure 2.3: Optimal M&A policy and trade openness

Note: Share of sectors where a merger at the top leads to positive welfare effects — for the domestic government, the foreign government, and a utilitarian world planner, that is $\int_0^1 m_\lambda(z)\mathrm{d}z$ for $\lambda \in \{0, \infty, L^*/L\}$.

of mergers when the economy is closed to international trade, as mergers in the closed economy have a particularly strong market power effect — market shares are high due to a lack of foreign competition. As trade costs decline and Home and Foreign trade more with each other, the domestic government is more favorable towards the mergers, while they become increasing less welcome abroad. In particular, a utilitarian global planner would approve fewer mergers than the domestic government when economies are very open.

The summary statistics in Figure 2.3, however, hide dramatic heterogeneity in the subsets of sectors in which the various planners would see mergers with a favorable eye. This is the main source of international conflicts of interest over merger policies, which we illustrate in Figure 2.4. In this figure, we sort sectors into deciles of comparative advantage Λ_z^* in the left panel and deciles of granularity Γ_z^* in the right panel, using the baseline value of trade costs $\tau = 1.34$. We combine together the bottom six deciles as there is little variation

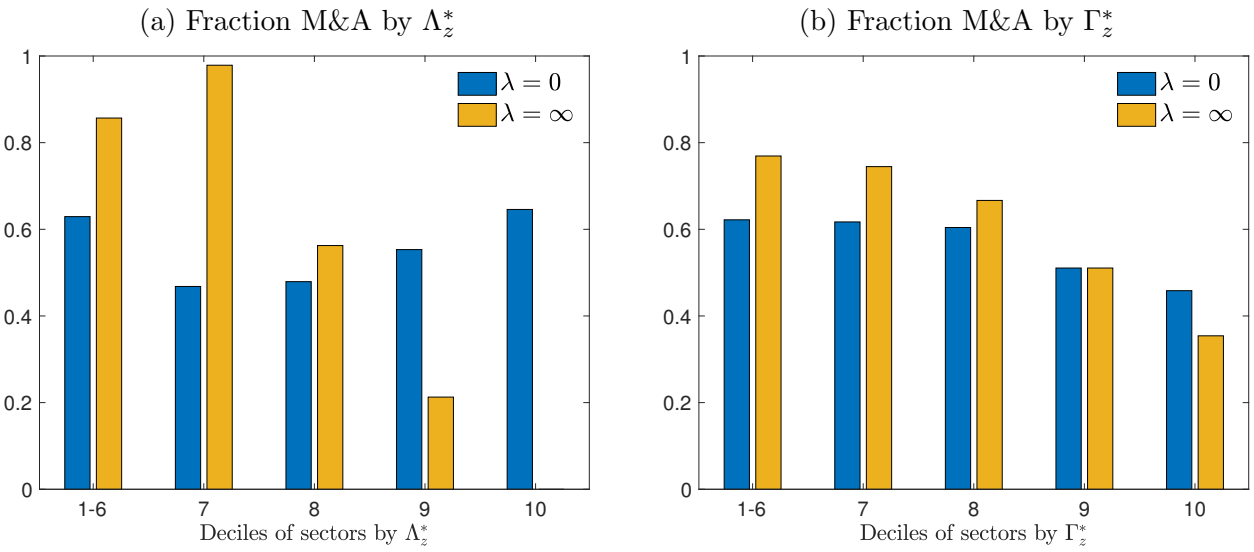


Figure 2.4: Optimal M&A policy across sectors

Note: Figure plots the fraction of sectors in which home mergers are welfare improving at home ($\lambda = 0$) and abroad ($\lambda = \infty$), by deciles of sectors in terms of home comparative advantage Λ_z^* and its granular component Γ_z^*. The bottom five deciles are binned together. Baseline parameters $\varrho = 0.35$ and $\tau = 1.34$.

across these lower deciles. We then plot the fractions of sectors within each bin for which the domestic ($\lambda = 0$) and foreign ($\lambda = \infty$) planners would favor a merger.

Two main insights emerge from the cross sectional analysis. The right panel indicates that both Home and Foreign dislike mergers in granular sectors compared to mergers in non-granular sectors. While they green-light the majority of mergers in sectors without large granular firms, Home would only allow mergers in 40% and Foreign in 25% of cases in the top decile of sectors by granularity. The reason is the excessive market power that firms generally hold in such sectors, which is particularly costly abroad as the loss in consumer surplus there is not compensated by any direct gains in producer surplus.

The pattern is somewhat different in the left panel of Figure 2.3. From the perspective of home welfare, the proportion of favorable mergers to a first approximation does not depend on sectoral comparative advantage — the home planner favors roughly 50–60% of mergers independently of Λ_z^*. In contrast, the foreign approval of domestic mergers decreases sharply

with Home's comparative advantage — while Foreign favors most mergers in low-Λ^* sectors, it would want to block *every* merger in the top decile of home comparative advantage.

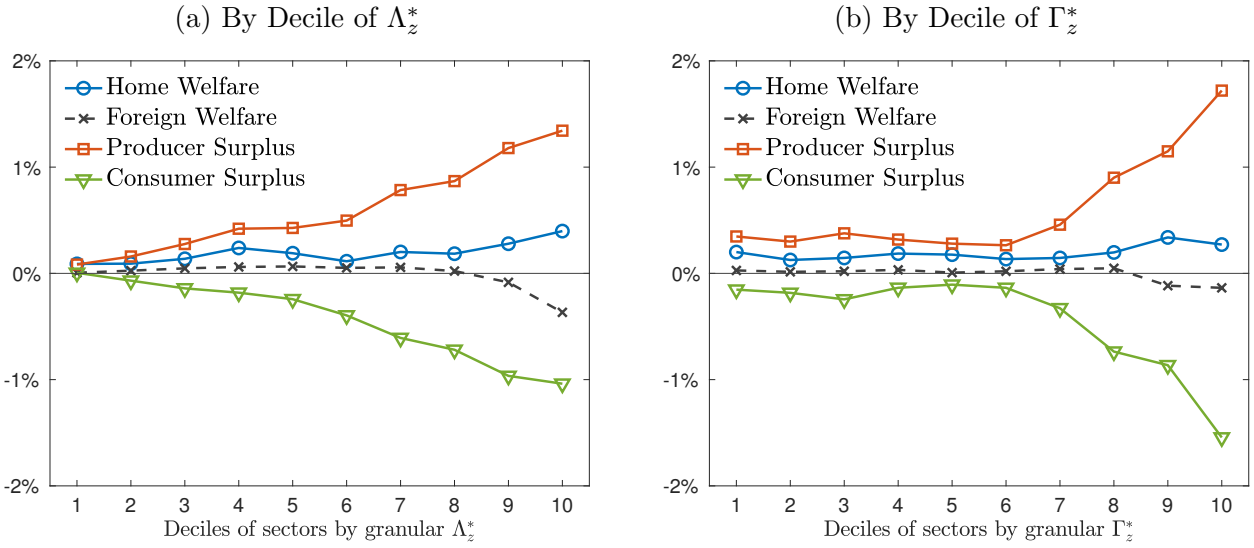


Figure 2.5: Welfare effects of a merger: decomposition into producer and consumer surplus

Note: Decomposition of home welfare effects from the merger policy according to (2.15).

This pattern is not surprising: as we discussed, home mergers in comparative advantage sectors disproportionately hurt Foreign, due to a transfer of consumer surplus. This, in turn, is the reason why Home favors many mergers in high comparative advantage sectors. Nonetheless, such mergers can also be costly in the domestic economy due to their excessive concentration of market power, and this is the reason why the home planner would not favor all such mergers. Figure 2.5 illustrates these conflicting welfare effects — sectors with strong comparative advantage have large consumer surplus losses at Home, which are however offset by large producer surplus gains.

Overall, this analysis suggests an important role for international cooperation over M&A policies in open economies to avoid excessive build-up of market power. Absent cooperation, each country will pursue excessive mergers, in particular in sectors with strong comparative advantage, resulting in a Prisoner-dilemma-like equilibrium. Furthermore, the beggar-

thy-neighbor distributional conflict makes some mergers unfavorable for Foreign, even in situations when a utilitarian global planner may favor them. This suggests that the home decision-maker is typically too lenient to domestic mergers in comparative advantage sectors, while the foreign decision-maker would try to always block such mergers, sometimes at a cost to multilateral efficiency.

2.5 Granular Tariffs

We turn to another aspect of government policies that may affect and target large firms — trade policy. In fact, trade policy is often so narrow that it appears tailor-made to target individual firms rather than industries. Such "granular" tactics are particularly widespread in antidumping retaliation, international sanctions and trade wars. Our quantified model allows us to analyze the economic incentives faced by the home government when considering imposing such granular tariffs rather than industry-wide ones, putting aside the legal ramifications of such a decision.

Specifically, we study two alternative tariff policies in an open granular economy. We contrast a *uniform* tariff $\bar{\varsigma}_z$ levied on all imports in sector z and a *granular* tariff $\hat{\varsigma}_{z,1}$ levied exclusively on the largest foreign exporter in the same sector. For concreteness, we compare a 1% uniform tariff with a granular tariff $\hat{\varsigma}_{z,1}$ that generates the same tariff revenue at the sectoral level. A government may prefer a granular tariff over a uniform one for two reasons. First, it may be more attractive in terms of domestic political economy, though perhaps more complex to impose legally. We leave aside these considerations. Second, it may be a more effective policy tool at extracting surplus from foreign producers and improving the home country's terms of trade. As we shall shortly see, this latter consideration is indeed at play in our granular model with oligopolistic competition.

General setup Consider firm-specific tariffs $\{\varsigma_{z,i}\}$ imposed by the home government on foreign firms i in sector z. In particular, if a foreign firm generates revenues $r_{z,i} = s_{z,i}\alpha_z Y$ in the home market, it needs to pay $\varsigma_{z,i} r_{z,i}$ to the home government, so that its net revenues are $(1-\varsigma_{z,i})r_{z,i}$. The foreign firm's profit maximization problem in the home market is then:

$$\Pi'_{z,i} = \max_{p_{z,i}} \left\{ \left[(1-\varsigma_{z,i})p_{z,i} - c_{z,i}\right] p_{z,i}^{-\sigma} \frac{\alpha_z Y}{\sum_{j=1}^{K_z} p_{z,j}^{1-\sigma}} - wF \right\} = (1-\varsigma_{z,i})\alpha_z Y \frac{s'_{z,i}}{\varepsilon(s'_{z,i})} - wF,$$

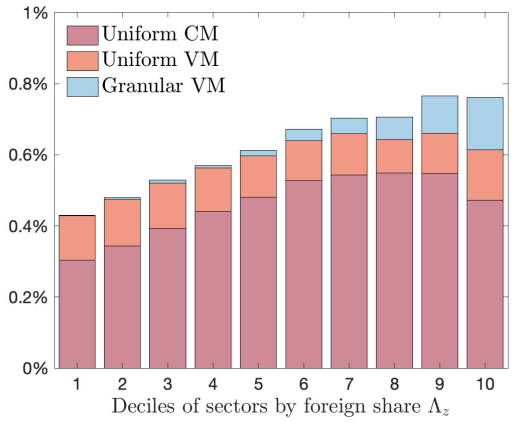
with the solution for prices and markups as if its costs were increased to $c'_{z,i} = c_{z,i}/(1-\varsigma_{z,i})$. We denote the resulting market shares by $\{s'_{z,i}\}$ with the residual demand elasticity $\varepsilon(s)$ still defined in (2.6). The foreign share in the home market is still given by $\Lambda'_z = \sum_{i=1}^{K_z}(1-\iota_{z,i})s'_{z,i}$. However, the home government now collects tariff revenues $TR_z = \alpha_z Y \sum_{i=1}^{K_z}(1-\iota_{z,i})\varsigma_{z,i}s'_{z,i}$, with the rest, $\Lambda'_z\alpha_z Y - TR_z$, being the export revenue of foreign firms. We describe the resulting changes to the general equilibrium conditions in Appendix B.2.

The implied change in home welfare from the tariff policy $\{\varsigma_{z,i}\}$ is described by the general expression in (2.15), which allows us to decompose the overall welfare effect into tariff revenue, and changes in consumer and producer surplus respectively. We again calculate the "welfare derivatives" $\hat{W}_z$, as described in (2.16), by studying the tariff policy in a subset of sectors with similar characteristics. This is necessary to appropriately capture the general equilibrium effects from the sectoral tariffs on overall welfare, which are of the same order of magnitude as the direct sectoral effect.[20]

Results Using the general framework above, we now compare two alternative tariff policies: (a) a uniform tariff with $\varsigma_{z,i} = \bar{\varsigma}_z = 0.01$ for all foreign firms i selling in the home market in sector z; and (b) a granular tariff $\hat{\varsigma}_{z,1}$ levied only on the largest foreign exporter to the home

[20]The direct effect is first-order within the sector, but comes with a weight $\alpha_z \mathrm{d}z$ in aggregation, while the indirect general equilibrium effect (on wages and price levels) is also of the order $\alpha_z \mathrm{d}z$, via general equilibrium conditions. We implement the policy in a subset of sectors $z \in Z$ with cumulative expenditure weight $\int_{z\in Z} \alpha_z \mathrm{d}z$, and scale the resulting welfare effect by $\int_{z\in Z} \alpha_z \mathrm{d}z$, as in (2.16).

(a) Change in home welfare



(b) Differential welfare effect

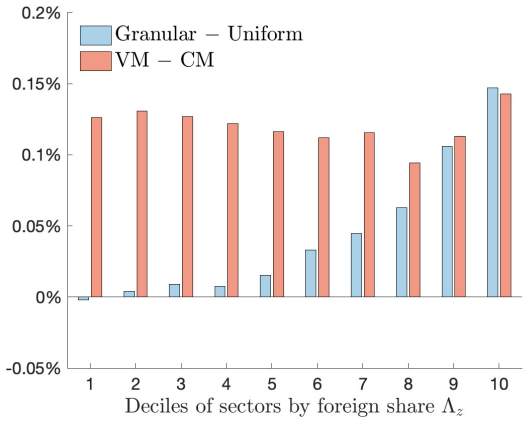


Figure 2.6: Granular tariff: welfare effect

Note: CM stands for constant markup counterfactual and VM is the variable markup baseline. The figures plot the welfare effects $\hat{W}_Z$ of tariff policy by deciles of foreign comparative advantage (foreign share Λ_z in the home market). The left panel plots $\hat{W}_Z$ in the case of a uniform tariff under CM (pink bars), under VM (pink+red bars), and a granular tariff under VM (pink+red+blue bars). The right panel zooms in on the differential welfare effects of the uniform tariff over the counterfactual case with constant markups (red bars), as well as the differential welfare effect of the granular vs the uniform tariff (both VM; blue bars).

market in sector z, with the value of the tariff given by $\hat{\varsigma}_{z,1}s'_{z,1} = \bar{\varsigma}_z\bar{\Lambda}'_z$, that is, the same as in case (a). We report here the results when competition takes the Cournot form, with the results for Bertrand competition being qualitatively similar, albeit quantitatively attenuated as markups are less variable in that case. Figure 2.6 compares the domestic welfare effects of each policy, contrasting sectors with different comparative advantage of the foreign country, Λ_z.[21] We find that imposing a granular tariff has clear welfare benefits from the point of view of Home, all the more in the sectors in which Foreign has a larger share of the home market.

To understand the forces behind these results, we decompose these domestic welfare effects in Figure 2.7: namely, we report its three components, following equation (2.15) — tariff revenues, consumer surplus, and producer surplus. A clear picture emerges. First, by construction, tariff revenues are equivalent between a granular and a uniform tariff in all

[21] Results are very similar when we group sectors by their level of foreign granularity Γ_z.

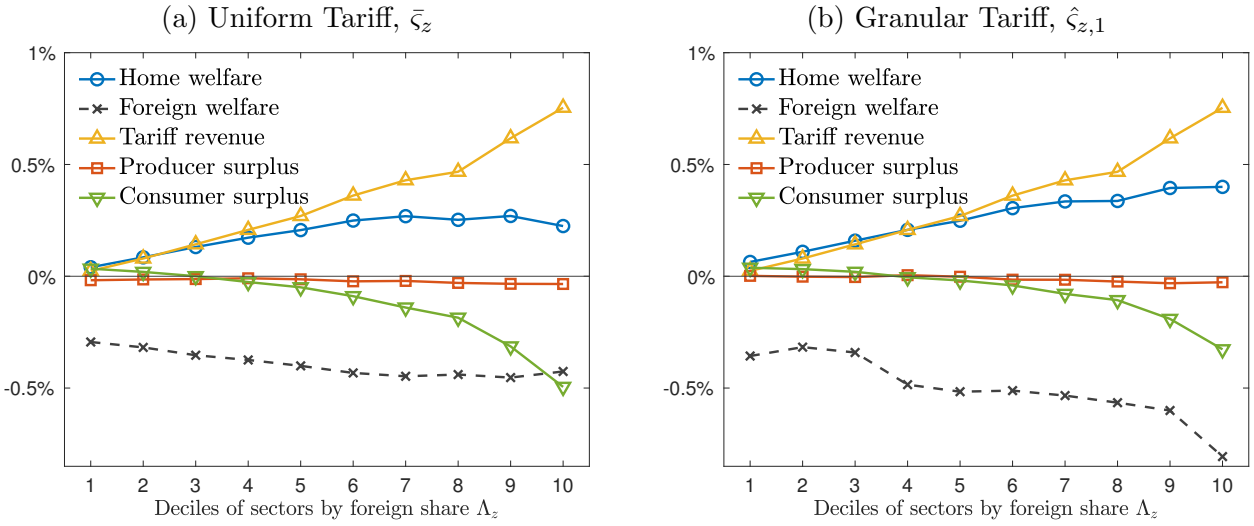


Figure 2.7: Welfare effect of a tariff: Decomposition

Note: Decomposition of the welfare effect $\hat{W}_Z$ at home into three components according to (2.15), and the welfare effect abroad, from a uniform and a granular tariff, by deciles of sectors split by foreign comparative advantage.

sectors and increase proportionally with the ex post sectoral import share Λ'_z. Second, the impact of these tariffs on domestic producer surplus is small, as these effects are indirect. Therefore, third, the major driver of the net welfare effects is the loss in domestic consumer surplus that these policies induce, offsetting a large part of the gains from tariff revenues. This is where the difference between uniform and granular tariffs plays a key role. Home consumers are hurt more by the uniform tariff levied on all foreign exporters than by the granular tariff concentrated on a single largest foreign exporter.

The reason for this result is that with a granular tariff, the pass-through of the import tariff to consumer prices at home is much lower. The tariff hits the largest foreign firm, which typically has a double-digit market share in the home market in high Λ_z (or Γ_z) sectors. This firm exerts significant power in the home market, and absorbs part of the increase in marginal costs coming from the tariff by lowering its markups (see Amiti et al., 2019). Overall, the increase in prices is therefore lower for home consumers than if the tariff were levied uniformly an all foreign firms in the sector. Put differently, the leading firm

prices strategically to keep its market share high and reduces its markup significantly in response to a loss in relative efficiency. If, instead, all foreign firms are subject to a (smaller) uniform tariff, their relative stance is largely unchanged, so that markup adjustments are smaller. This makes granular tariff an efficient policy tool for the home government as it allows to extract the same tariff revenue at a minimal cost in terms of the loss in consumer surplus due to the lower tariff pass-through into prices.[22]

Notice that this effect is not present in Costinot et al. (2020), who study a constant markup case with a continuum of heterogeneous firms that all exhibit a complete pass-through of the tariff. We find that the pass-through effect, which operates even in partial equilibrium, is quantitatively large — the loss in consumer surplus is cut by more than half in the most import-intensive sectors compared to the constant markup case. Notice further that this pass-through effect is present even when we apply a uniform tariff to the whole sector, so long as markups are variable and the foreign share $\Lambda_z < 1$. This can be seen in Figure 2.6 where we compare the welfare impact of a uniform tariff against a counterfactual case with constant markups set at the monopolistically competitive level, $\frac{\sigma}{\sigma-1}$. The incomplete pass-through effect is maximized by imposing a granular tariff, which targets the firm with the lowest pass-through rate in the industry, thus minimizing the loss of consumer surplus for a given amount of tariff revenues.

Quantitatively, Figure 2.6 shows that the home welfare gain from a uniform tariff in a variable markup environment is about a third higher than in a constant markup counterfactual. This gap is roughly constant across all sectors, independently of the foreign share Λ_z. In contrast, the additional gains from setting a non-uniform granular tariff are only present in sectors with substantial foreign comparative advantage. In sectors with large foreign firms,

[22] In Appendix B.2, we provide a formal analysis of these effects, using the welfare approximation introduced in Section 2.3. In particular, we show that the main difference between the tariffs is driven by the loss in consumer surplus, which in general is given by $\hat{P}_z = \frac{1-\bar{\varkappa}_z^F}{1-\bar{\varkappa}_z} \bar{\varsigma}_z \Lambda'_z$, where $1 - \bar{\varkappa}_z^F$ is the average pass-through of the tariff among the affected foreign firms. A granular tariff minimizes this pass-through, and thus dominates a uniform tariff (or any other tariff profile which yields the same tariff revenue).

using a granular tariff increases the home welfare gain by another third. In other sectors, even the largest foreign firm has a small market share in the domestic market, insufficient to result in a significantly lower pass-through of the granular tariff.

This analysis suggests that variable markups and the incomplete pass-through of large firms are crucial quantitative components of the optimal tariff analysis. Moreover, it shows that governments likely have strong economic incentives to target only the largest foreign firms with tariffs as an effective way of extracting foreign producer surplus with minimal consequences for domestic consumer surplus.

2.6 Industrial Policy in a Granular World

To conclude, we briefly consider an extension of the model in which firms can make a one-time investment in boosting their productivity. We explore under which circumstances the planner wants to subsidize such investments in comparative advantage sectors and in particular by the industrial champions — the largest firms in the economy.

Setup economy Consider an extension of the granular open economy in which a firm in sector z can invest

$$v_{z,i} = \kappa \alpha_z \frac{\varphi_{z,i}^{\delta}}{\sum_{j=1}^{K_z} \varphi_{z,j}^{\delta}} x_{z,i} \tag{2.23}$$

to boost its productivity from $\varphi_{z,i}$ to $\varphi'_{z,i} = \varphi_{z,i}(1 + x_{z,i})^{1/\zeta}$ for some $\kappa, \delta > 0$ and $\zeta > \sigma - 1$. We are interested in the home planner's allocation of investment $\{v_{z,i}\}$ and associated productivity boosts $\{x_{z,i}\}$, and for concreteness limit the planners budget to V, so that:

$$\int_{z\in[0,1]} \left(\sum_{i=1}^{K_z} \iota_{z,i} v_{z,i} \right) \mathrm{d}z \leq V. \tag{2.24}$$

In a counterfactual closed economy with constant markups, a planner is simply maximizing aggregate productivity given by:

$$\Phi = \exp \int_{z\in[0,1]} \frac{\alpha_z}{\sigma-1} \log\left(\sum\nolimits_{i=1}^{K_z} \varphi_{z,i}^{\sigma-1} (1+x_{z,i})^{\frac{\sigma-1}{\zeta}} \right) \mathrm{d}z. \tag{2.25}$$

In an open economy with variable-markup pricing, the planner chooses $\{v_{z,i}, x_{z,i}\}$ to maximize the change in welfare $\hat{\mathbb{W}} = \mathrm{d}\log(Y/P)$, as defined in (2.15), subject to (2.23) and (2.24).

For concreteness we consider a total budget $V = \kappa\zeta/100$ so that a uniform 1% productivity improvement ($x_{z,i} \equiv \bar{x} = \zeta/100$) is feasible for every firm in every sector.[23] We ask the question under which circumstances the planner would indeed favor such a uniform allocation of investment and when she would prefer to skew the investment towards certain sectors and firms, in particular national champions.

Closed economy Consider first a closed economy, using our welfare approximation developed in Section 2.3. The log productivity improvement is $\hat{\varphi}_{z,i} \equiv \log(\varphi'_{z,i}/\varphi_{z,i}) \approx x_{z,i}/\zeta$. Using (2.20) and (2.21), the reduction in sectoral price levels and the increase in profits is given by:

$$\hat{P}_z = -\frac{1}{1-\bar{\varkappa}_z} \sum\nolimits_{i=1}^{K_z} s_{z,i}(1-\varkappa_{z,i})\hat{\varphi}_{z,i} \quad \text{and} \quad \frac{\mathrm{d}\Pi_z}{\alpha_z Y} = \sum\nolimits_{i=1}^{K_z} \frac{s_{z,i}}{\mu_{z,i}} \varkappa_{z,i}(\hat{\varphi}_{z,i} + \hat{P}_z), \tag{2.26}$$

as the firms increase their markups by $\hat{\mu}_{z,i} = \varkappa_{z,i}(\hat{\varphi}_{z,i} + \hat{P}_z)$ according to (2.19), absent any exogenous shifts in markups. Therefore, the overall welfare effect can be evaluated as:[24]

$$\hat{W}_z = \frac{\mathrm{d}\Pi_z}{\alpha_z Y} - \hat{P}_z = \sum\nolimits_{i=1}^{K_z} s_{z,i}\hat{\varphi}_{z,i} - \mathrm{cov}\left(\hat{\varphi}_{z,i}, \frac{\varkappa_{z,i}}{\varepsilon_{z,i}} + \frac{(\overline{\varkappa_z/\varepsilon_z})}{1-\bar{\varkappa}_z}\varkappa_{z,i} \right), \tag{2.27}$$

[23] Note that in this case, $\int_{z\in[0,1]} \left(\sum_{i=1}^{K_z} v_{z,i}\right)\mathrm{d}z = \kappa\bar{x}$, as $\sum_{i=1}^{K_z} \frac{\varphi_{z,i}^{\delta}}{\sum_{j=1}^{K_z}\varphi_{z,j}^{\delta}} = 1$ and $\int_{z\in[0,1]} \alpha_z \mathrm{d}z = 1$.

[24] To derive this result, combine the expressions in (2.26) and simplify by using the fact that $\mu_{z,i} = \varepsilon_{z,i}/(\varepsilon_{z,i}-1)$.

where the covariance term is market-share weighted and $(\overline{\varkappa_z/\varepsilon_z}) \equiv \sum_{i=1}^{K_z} s_{z,i}\varkappa_{z,i}/\varepsilon_{z,i}$.

Equation (2.27) has a number of implications. First, in a counterfactual economy with constant markups, that is with complete pass-through $1-\varkappa_{z,i} \equiv 1$ and thus with $\varkappa_{z,i} \equiv 0$, the distribution of productivity boosts across firms (and sectors) does not matter, and a sufficient statistic is the average productivity growth, $\sum_{i=1}^{K_z} s_{z,i}\hat{\varphi}_{z,i}$. This is a version of Hulten's theorem. As a result, the planner should simply maximize aggregate productivity (2.25) subject to her budget constraint (2.24). In this case, it is optimal to have $\hat{\varphi}_{z,i}$ common across all firms ($x_{z,i} = \bar{x}$ for all firms) when $\delta = \sigma - 1$ in (2.23), while a smaller (larger) δ favors investment into the largest (smallest) firms. In what follows, we focus on the case of $\delta = \sigma - 1$, which establishes an indifference benchmark in a counterfactual closed economy with constant markups, whereby the planner is indifferent on the margin which firms should receive the productivity boost. Indeed, per unit of investment expenditure, this results in the same aggregate productivity gains independent of the details of the allocation.

When markups are variable, both $\varkappa_{z,i}$ and $\varkappa_{z,i}/\varepsilon_{z,i}$ are increasing in the firm's market share $s_{z,i}$, as larger firms have both a greater strategic complementarity elasticity $\varkappa_{z,i}$ and a smaller elasticity of demand $\varepsilon_{z,i}$. Therefore, the planner can make the covariance term in (2.27) negative by allocating $\hat{\varphi}_{z,i}$ towards the smaller firms. When $\delta = \sigma - 1$, the increase in the average productivity does not depend on the allocation of investments on the margin, as discussed above, and thus the preferences of the planner are shaped by the negative of the covariance term in (2.27). As a result the planner favors investment to increase productivity of the smallest firms. This result is intuitive, as such investment allows the smaller firms to catch up with the larger firms, thus eroding the monopoly power of the larger firms and bringing down the average markup in the closed economy. This reduces the deadweight loss and results in greater welfare gains for the same increase in the average productivity across firms. Another way to see this from (2.26) is that smaller firms have a greater pass-through $1-\varkappa_{z,i}$ of their productivity gains into a reduction of consumer prices, while the larger firms

retain a larger share $\varkappa_{z,i}$ of the productivity improvement by increasing their markups, which has a proportionally smaller (by a factor $\mu_{z,i} > 1$) effect on aggregate welfare.

Open economy Matters are different in an open economy as the benefits of domestic productivity gains only partially accrue to domestic consumers, who buy a basket of home and foreign products, and also partially accrue to foreign consumers by means of their terms of trade appreciation. This provides an incentive for the home country to subsidize more intensively its comparative advantage sectors, which maximize the benefits from productivity improvements. In contrast, in comparative disadvantage sectors, the home country has an incentive to "free ride" on the foreign high productivity level, which translates into favorable terms of trade for the home country.[25]

Furthermore, the same force that made it disadvantageous to subsidize large firms in a closed economy, may render it desirable in an open economy. Specifically, the low pass-through $1 - \varkappa_{z,i}$ of productivity improvements into consumer prices at home translates into greater profitability of home firms abroad. In other words, subsidizing productivity improvements for large exporters has disadvantages in the home market, yet allows to capture a greater share of enhanced competitiveness in the foreign market in the form of greater producer surplus. Indeed, the direct effect of productivity improvement on the producer surplus from exports is given by $\sum_{i=1}^{K_z^*}(1-\iota_{z,i}^*)\frac{s_{z,i}^*}{\mu_{z,i}^*}\varkappa_{z,i}^*\hat{\varphi}_{z,i}$, and it is increasing in both the comparative advantage of a sector $\Lambda_z^* = \sum_{i=1}^{K_z^*}(1-\iota_{z,i}^*)s_{z,i}^*$ and in the average markup elasticity $\varkappa_{z,i}^*$ of the subsidized firms, which is increasing in firm size.

[25] Formally, the increase in consumer surplus is given by:

$$-\hat{P}_z = \frac{1}{1-\bar{\varkappa}_z}\sum_{i=1}^{K_z}\iota_{z,i}s_{z,i}(1-\varkappa_{z,i})\hat{\varphi}_{z,i} = (1-\Lambda_z)\frac{1-\bar{\varkappa}_z^H}{1-\bar{\varkappa}_z}\frac{\sum_{i=1}^{K_z}\iota_{z,i}s_{z,i}(1-\varkappa_{z,i})\hat{\varphi}_{z,i}}{\sum_{i=1}^{K_z}\iota_{z,i}s_{z,i}(1-\varkappa_{z,i})},$$

where $1-\Lambda_z = \sum_{i=1}^{K_z}\iota_{z,i}s_{z,i}$ is the home share and $1-\bar{\varkappa}_z^H = \frac{1}{1-\Lambda_z}\sum_{i=1}^{K_z}\iota_{z,i}s_{z,i}(1-\varkappa_{z,i})$ is the average pass-through rate among domestic firms. If the home government subsidizes proportionately every home firm in sector z, so that $\hat{\varphi}_{z,i} = \hat{\varphi}_z$ for all i, then $-\hat{P}_z = (1-\Lambda_z)\frac{1-\bar{\varkappa}_z^H}{1-\bar{\varkappa}_z}\hat{\varphi}_z$, which is increasing in the home share $1-\Lambda_z$.

To summarize, the planner favors productivity gains by smaller firms in a closed economy, as it helps to bring down the average markup distortion. That is, in a closed economy, the planner is likely to adopt policies opposing the creation of "national champions" and granular firms. In an open economy, however, there exists a rationale to subsidize productivity gains in comparative advantage sectors and, furthermore, by the largest firms in such sectors, thus creating national champions in sectors of comparative advantage. The gains in producer surplus abroad may become the goal of the policy, in which case large firms with a low pass-through of productivity gains into foreign consumer prices may become the target of industrial policies. We leave the comprehensive quantitative evaluation of this hypothesis to future research.

2.7 Conclusion

Granular firms play a pivotal role in international trade. We analyze a series of policies aimed specifically at large individual firms in an open economy context, relying on the theoretical and quantitative framework developed in Gaubert & Itskhoki (2020). We study merger policies as well as 'granular' trade policy. In both cases, we find that the granular structure of the world economy offers powerful incentives for governments to adopt policies targeted at individual firms, and that they tend to create negative international spillovers.

The key force underlying these results is that these large individual firms tend to enjoy substantial market power. In this context, lenient merger policy towards local firms allows national governments to help domestic firms build producer surplus at the expense of foreign consumers, a form of 'beggar-thy-neighbor' policy. Turning to trade policy, granular tariffs targeting large foreign exporters allow to leverage the fact that these firms have high markups, hence absorb a large fraction of import tariffs rather than passing them on to home

consumers. As a result, granular trade policy can achieve the same tariff revenues at lower cost in terms of consumer surplus, compared to imposing a tariff on all firms in a sector.

Negative spillovers created by such unilateral policies could be addressed with international cooperation mechanisms. In particular, the analysis puts forth a clear rationale for international cooperation on merger and antitrust policy. Our work has analyzed the incentives faced by governments when deciding on unilateral antitrust and trade policies, assuming that its trade partners do not retaliate. This leaves a set of interesting questions for future research. Next on the agenda should be the analysis of incentives to retaliate by the foreign country, and the Nash equilibrium that ensues. Another one is to distinguish between two sources of large firms' market power, which may have very different implications for these open-economy policies: market power on the output market and market power on the input markets.

Chapter 3

The Migration Response to the Euro Crisis - An Empirical Investigation

3.1 Introduction

To what extent does cross-country migration respond to temporary economic shocks, such as business cycle fluctuations? The economic migration literature has so far ignored this question and instead focused on the role of persistent economic differences in driving migration as summarized in Greenwood (1997) and Lucas (1997). This paper fills this gap and provides causal evidence for the effect of a temporary economic crisis on inter-country migration, using the Euro crisis as a natural experiment.

Why would one be interested in answering this question? After all, migration caused by persistent income differences, such as that from eastern to western Europe or from southern to northern America, appears to be quantitatively dominant and most impactful in the political discourse. Whilst this is undoubtedly true, the response to temporary shocks is crucial to assess the strength of migration as an adjustment mechanism to business cycle fluctuations. With the emergence of the Euro currency area, the importance of and interest in such an adjustment mechanism has grown strongly. In fact, in the optimal currency area

theory going back to Mundell (1961), this role of labor mobility is one of the preconditions for the success of a currency union. The necessity of such a mechanism has become apparent with the advent of the recent Euro crisis that has demonstrated the vulnerability of the Euro zone to diverging economic fortunes of its member countries. In response to this crisis, a series of macroeconomic papers outlined below has attempted to capture and explain the role of migration in smoothing out business-cycle fluctuations through the use of large-scale quantitative DSGE models.

However, the absence of empirical evidence on the response of migration to temporary shocks has been a major impediment to these efforts. In particular, this kind of causal evidence is crucial to properly connect quantitative DSGE models to the data in the spirit of Nakamura & Steinsson (2018). In showing the existence and providing a quantification of this causal effect, this paper provides exactly that kind of evidence. It thus not only provides policy makers with an understanding of the significance of crisis-driven migration, but also aids macroeconomists in better understanding the underlying mechanisms.

Identifying the migration response to the Euro crisis is not straightforward, since migration is not a phenomenon specific to crises and one thus needs to exclude all migration that would have happened even in the absence of a crisis. To solve this problem, I utilize one key characteristic of the Euro crisis, which makes it a great natural experiment for my purposes. In particular, not all EU countries were equally affected by the Euro crisis. Instead, whilst some, mostly southern European countries, were hit hard, others, mostly in central and northern Europe, were barely affected at all. This allows me to estimate the effect of the Euro crisis on migration in an event study context, with unaffected countries serving as the control group for those suffering from the Euro crisis. In this way, I can isolate crisis-driven migration by differencing out migration from countries in the control group. This reduces causal identification to the question whether non-crisis countries are a good control group

for crisis countries - the familiar common trends assumption which is empirically testable in an event study setting.

To implement this approach, I rely on a unique data set collected by the German Federal Statistics Office that separately keeps track of inflows and outflows of migrants for each EU country over a twenty year window. In order to generate my treatment and control groups, I classify countries into two groups - those affected by the crisis and those unaffected. Since net migration to Germany from both groups of countries follows the same trend pre-crisis, the crucial identification assumption holds. In contrast to this common pre-crisis behavior, I document a strong divergence in migration between treatment and control groups as soon as the crisis hits, and describe the exact time profile of these differential migration flows. To cope with the data scarcity inherent in this macroeconomic panel data set, I not only rely on a traditional nonparametric event study but also construct parametric event studies that reduce the number of parameters to be estimated. I conclude that the Euro crisis led to an additional 935,000 migrants to Germany over the time period 2010-2019 accounting for roughly 21% of all EU migration into Germany. In a final step, I decompose the net migration response into inflow and outflow responses and show that the majority of the increase in net migration is due to rapidly increasing inflows with outflows staying constant.

Overall, the Euro crisis thus caused a large amount of inter-country migration, suggesting a significant role for migration as an insurance mechanism against country-specific shocks. For macroeconomists studying the underlying economic forces, I not only provide this quantification and hence a whole array of calibration targets but also three key salient facts that their migration models should match. First, besides the sheer size of the effect, it is crucial that models generate large inflows without a simulatenous response in outflows. Second, the migration response to a crisis takes time to unfold as migration flows reach their maximum only around five years after the crisis. It is thus not an immediate but rather a medium-term adjustment mechanism. Third, after the conclusion of the crisis, migration

flows merely return to their pre-crisis levels but do not reverse, leading to hysteresis in the cross-country population distribution. Policy makers should take those numbers and facts into consideration when considering how to make the Euro zone more robust for future crises.

Related literature I contribute to the quantitative literature on migration by providing calibration targets and salient facts these models should match. The canonical model in this literature is Kennan & Walker (2011), who focus on expected income as the main source of migration. Influenced by this, a large literature in economic geography (see for an overview Redding, 2010, 2013; Redding & Rossi-Hansberg, 2017) has arisen that attributes a strong role for economic considerations underlying migration. This literature has been very fruitful, demonstrating for example the circular dependence of migration and economic development in Desmet et al. (2018) and the importance of migration for the propagation of economic shocks in Caliendo et al. (2017).

Closer to the focus of this paper are recent quantitative migration models of the Euro zone and the Euro crisis. House et al. (2018) aim to quantify the benefits of labor mobility in Europe. Oikonomou (2020) investigates migration as an adjustment mechanism for country-specific shocks, arguing for the roll of the skill composition. Bandeira et al. (2019) focus on the interplay between migration and fiscal policies in a curreny zone and Hauser & Seneca (2019) on that between migration and monetary policy. Lozej (2019) and Smith & Thoenissen (2019) study the effects of migration shocks on destination countries, focussing on Ireland and New Zealand respectively.

A different literature hase emerged that empirically investigates the relationship between migration and economic factors. In particular, there is a large literature investigating the link of economic development and inter-state migration in the US (see Molloy et al., 2011, 2016, 2017; Kaplan & Schulhofer-Wohl, 2017b; Dao et al., 2017). Regarding Europe Izquierdo et al. (2016) and Bandeira et al. (2019) descriptively show the strength of the migration response

in Spain and Greece to the Euro crisis. They conclude that 1% of the total population in Spain and 7% of the active population in Greece exited their home countries between 2010 and 2015.

Closer to this paper, a large literature on the response of migration to economic shocks is inspired by the VAR approach in Blanchard & Katz (1992) that aims to understand the joint behavior of employment, unemployment and the participation rate in response to positive regional labor demand shocks in the US. For the European case this analysis is replicated in Decressin & Fatás (1995), Huart & Tchakpalla (2015) and Beyer & Smets (2015), who all conclude that the migration response to economic shocks is significantly weaker in Europe than in the US. Maffei-Faccioli & Vella (2020) and Furlanetto & Robstad (2019) focus on the reverse causality and study the impact of migration on investment in Germany and unemployment in Norway respectively. Whilst we thus share the focus on a response to shocks rather than permanent differences, these papers differ in their focus on the response to one large crisis rather than multiple small shocks. Due to potential nonlinearities in the migration response, their VAR results do not reveal the causal impact of a large economic crisis on inter-country migration.

The papers closest to my study are Bandeira et al. (2019) and in particular Bertoli et al. (2016), who investigate the relationship between migration and unemployment in the EU. In contrast to me, they do not aim to identify the migration response to a large crisis shock but instead aim to find the main source of migration in the Euro zone by comparing the strength of the correlation of migration and various factors. They conclude that economic conditions in the origin country are a very strong predictor of migration.

The rest of the paper is structured as follows. In Section 3.2 I first introduce the data sources I rely on and provide a detailed assessment of their reliability. Section 3.3 then discusses my identification strategy in detail. Section 3.4 presents the main results and

Section 3.5 provides various robustness checks. Finally, Section 3.6 decomposes net flows into inflows and outflows. Section 3.7 concludes.

3.2 Data

In this section I first describe my data sources and the main variables that I use to investigate the causal link between an economic crisis and cross-country migration in the EU in subsequent sections. I then show and discuss descriptive statistics of those variables.

3.2.1 Data Sources

Population Data In order to normalize migration for the population sizes of the origin countries, I use population data from the European Statistical Office (Eurostat, 2021c). In particular, in line with the standard in the migration literature I use what is known as "average" populations. That is, I define the population p_{jt} for each origin country j and year t as the average of the population on January 1 of year t and year $t+1$. The population measures are collected from EU Member States and are based on recent local censuses and population registers.[1]

Migration Data To draw insights on migration flows, I rely on a unique administrative data set collected by the German Federal Statistics Office (Statistisches Bundesamt, 2021), which records inflows into Germany from EU countries and outflows into other EU countries at a yearly frequency. In particular, for inflows the country of origin is recorded as the last country of residence before moving to Germany and equivalently for outflows the destination

[1] In particular, the "usual resident population" is counted, where one qualifies as a usual resident by either living in the country for 12 months or by arriving within the past 12 months with an intention to stay there for at least one year.

country is the new country of residence.[2] I use data from 2000 to 2019, which allows me to have both extensive pre and post-event time series.

By German law[3] each person moving in or out of Germany is required to immediately register the move at the municipal office.[4] The law does not subject registration requirements to a minimum duration or the scope of the stay, although an exception for people who stay less than three months exists, so that tourists are not counted. The data from municipal population registers are then forwarded from the municipal registers to the German Federal Statistics Office according to the to the Bevölkerungsstatistikgesetz (BevStatG), which thoroughly checks the data for double reporting, plausibility and other inconsistencies.

Regarding migration outflows, municipal governments follow up with those migrants who miss their reporting obligations and impute the destination country as the country of citizenship for those who cannot be contacted.[5]

Concerning migration inflows, migrants face strong additional incentives to immediately register, since registration is required to receive an income tax card. This in turn is needed to sign any contract and is thus necessary to gain employment (also for seasonal work), issue an invoice if self-employed, rent an apartment, gain health insurance, open a bank account,

[2]Note that while asylum seekers are included in these records, immigrants of German heritage from Eastern Europe and the former Soviet Union, so-called "Aussiedler" and "Spätaussiedler", are counted as domestic migrants instead of international migrants and thus do not show up in the cross-border migration numbers.

[3]Since 2015 this is legislated at the federal level through the Bundesmeldegesetz (BMG), whilst it was legislated at the state level before that. However, the federal Melderechtsrahmengesetz (MRRG) defined minimum standards that all states had to meet since 1980, so that state regulations did not differ with respect to the registration requirements for foreign migrants. Hence, for the purposes of this paper, registration requirements have been uniform in Germany for the entire period of interest, guaranteeing comparability across time and space.

[4]There are other groups such as diplomatic personnel and their families that are exempt from registration requirements, but those numbers are neglible compared to the observed migration streams.

[5]This did not apply in 2016, 2017 and 2018 for countries such as Syria and Iraq, since the assumption that refugees would return to their home countries is unreasonable. Instead, they are listed as destination unknown. However, since my focus lies on the Euro crisis and EU countries, these issues lead to no substantial challenges for this paper.

get a mobile phone contract, etc.[6] Thus migrants reliably comply with their mandatory reporting obligations and the timing of their registration is an accurate representation of the actual cross-border move. Data quality is further improved by the fact that the share of tax revenues that municipalities receive depend on their populations, incentivizing them to further enforce registrations through late fees.

Overall, all of these factors lead to extremely high data quality, with almost all remaining mistakes being due to those outflowing migrants who neglect their reporting obligations, do not respond to follow-up inquiries and do not move to their country of citizenship.[7]

Since there were 27 countries in the EU in 2019 for each of which I observe migration inflows and outflows in all 20 years, there are $27 \times 20 = 540$ observations. The overall inflows from these countries into Germany over this 20-year window were 11,385,325, whilst there were 8,899,955 outflows leading to a total net inflow of 2,485,370 immigrants. This net inflow from EU countries represents 45.17% of the overall net migration inflow into Germany during this time period. The three largest origin countries are Poland (695,058), Romania (619,050) and Bulgaria (303,414) with countries hit particularly hard during the crisis also featuring highly with Italy (159,925) in fifth place, Spain (94,833) in seventh place and Greece (72,884) in eight place.

The dependent variables for the analysis below are defined as follows. Let y_{jt}^{in} and y_{jt}^{out} be the total inflows and the total outflows into/from Germany from/into country j in year t. The total net migration inflow from country j to Germany in year t, y_{jt}^{net}, is then defined

[6]Schneider (2012) and Vogel & Aßner (2011) document that illegal immigration in Germany appears to be limited and none of the EU countries are included in the 10 largest origin countries of illegal migrants, even before the lifting of intra-EU immigration restrictions. Whilst the number of illegal migrants might have increased in light of the refugee crisis, the free movement granted to all EU citizens strongly suggests that illegal immigration from EU countries is negligible.

[7]Note, however, that as usual with administrative migration data, registrations and thus the number of moves are measured instead of the number of movers. It is therefore possible that the same person is measured multiple times.

as

$$y_{jt}^{net} \equiv y_{jt}^{in} - y_{jt}^{out}.$$

Analogously, I define the inflow rates $\tilde{y}_{jt}^{in}$ and outflow rates $\tilde{y}_{jt}^{out}$ by dividing y_{jt}^{in} and y_{jt}^{out} by p_{jt}, so that the net migration rate $\tilde{y}_{jt}^{net}$ is defined by[8]

$$\tilde{y}_{jt}^{net} \equiv \tilde{y}_{jt}^{in} - \tilde{y}_{jt}^{out} = \frac{y_{jt}^{in} - y_{jt}^{out}}{p_{jt}}.$$

Unemployment Data I rely on yearly unemployment data from the European Statistical Office (Eurostat, 2021d). In particular, Eurostat bases their unemployment numbers on the European Union Labor Force Survey, which ensures comparability through the use of a harmonized definition of unemployment. In particular, an unemployed person is defined in line with the guidelines of the International Labor Organization as "someone aged 15 to 74[9] without work during the reference week who is available to start work within the next two weeks and who has actively sought employment at some time during the last four weeks" (Eurostat, 2021b). The unemployment rate u_{jt} in country j in year t is then defined as the number of unemployed people as a percentage of the labor force.[10]

3.2.2 Descriptive Statistics

Table 3.1 presents descriptive statistics for the different variables. In particular, it displays the mean, minimum and maximum of the total net migration y_{jt}^{net}, the migration rate $\tilde{y}_{jt}^{net}$ and the unemployment rate u_{jt} for each of the 27 origin countries. Several interesting observations emerge from this table.

[8] In order to simplify exposition I multiply the migration rates by one million, so that the migration rates are the number of migrants per one million inhabitants instead of per inhabitant.

[9] 16 to 74 in Italy, Spain and the UK.

[10] The labor force is defined as the active population, i.e it includes both employed and unemployed people but not the economically inactive such as pre-school children, school children, students and pensioners (Eurostat, 2021a).

First, note how there are persistent differences in the mean levels of net migration, with eastern European countries such as Poland, Romania and Bulgaria leading and smaller western European countries such as Malta an Ireland trailing. Similar patterns emerge when considering migration rates, though smaller countries like Luxembourg and Hungary join the top ranks. Second, in addition to the above-mentioned eastern European countries, southern European countries such as Greece, Italy, Spain and Portual also have a high variability of migration (reflected by the differences between maxima and minima). Third, looking at unemployment, both eastern and southern European countries top the list with respect to both the average and the variability of unemployment, with the maxima being achived by Greece with 27.5%, Spain with 26.1% and Bulgaria with 20.3%. Based on these preliminary statistics there thus appears to be a positive relationship between high unemployment and high migration rates, in line with the results of Bertoli et al. (2016).

The last two rows in the table show the population-weighted average[11] statistics across all origin countries as well as the unemployment statistics for Germany as a comparison. The first of these rows once again show the large differences between individual countries, whilst the second row seems to indicate that Germany had lower unemployment than the average country during the period of interest. This, in addition to the fact that Germany is the largest European country and a traditional host country to immigrants from all European regions and countries, validates the use of Germany as the single host country in my analysis that is necessitated by data availability.

Appendix tables C.2 and C.3 provide analogous descriptive statistics for total inflows y_{jt}^{in} and outflows y_{jt}^{out} as well as inflow rates $\tilde{y}_{jt}^{in}$ and outflow rates $\tilde{y}_{jt}^{out}$. The results for inflows mirror those for net migration, though countries with high net migration on average also appear to have higher outflows. These results suggest that high net migration is predomi-

[11] To be precise, each country j is weighted by $\frac{\frac{1}{20}\sum_{t=2000}^{2019} p_{jt}}{\frac{1}{27}\sum_{j=1}^{27}\frac{1}{20}\sum_{t=2000}^{2019} p_{jt}}$.

Table 3.1: Descriptive Statistics Net Flows

	y_{jt}^{net}			$\tilde{y}_{jt}^{net}$			u_{jt}		
Country	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Austria	-1903	-7221	4403	-230.0	-867.8	495.8	4.9	3.5	6.0
Belgium	278	-653	1377	24.3	-61.0	124.0	7.6	5.4	8.5
Bulgaria	15170	123	37850	2091.3	16.1	5273.1	10.5	4.2	20.3
Croatia	8112	-3415	36727	1952.3	-792.2	8727.9	12.9	6.6	17.3
Cyprus	187	-56	724	220.6	-71.2	840.0	7.8	3.6	16.1
Czechia	2209	-810	4581	211.2	-78.0	434.4	6.1	2.0	8.8
Denmark	-154	-1518	794	-28.2	-276.3	136.6	5.7	3.7	7.8
Estonia	279	-127	683	208.2	-95.0	514.5	9.0	4.4	16.7
Finland	-12	-792	477	-3.3	-150.4	87.7	8.3	6.4	9.8
France	632	-8084	5464	8.9	-128.3	83.1	9.2	7.4	10.4
Greece	3644	-9634	22923	337.0	-879.4	2075.4	15.7	7.8	27.5
Hungary	7941	-464	26208	800.9	-45.8	2641.8	7.2	3.4	11.2
Ireland	84	-834	1067	12.0	-204.9	232.0	8.2	4.2	15.5
Italy	7996	-14851	37057	130.0	-257.4	609.6	9.4	6.1	12.7
Latvia	1721	293	5007	834.4	134.6	2430.9	11.4	6.1	19.5
Lithuania	2600	357	5189	851.4	111.6	1713.6	10.7	4.3	17.8
Luxembourg	748	130	1343	1417.4	298.0	2545.9	4.8	2.3	6.7
Malta	7	-81	175	15.2	-177.9	347.2	6.0	3.6	7.6
Netherlands	2676	-12	5143	162.6	-0.7	315.1	4.8	2.3	7.4
Poland	34752	-1130	71610	912.6	-29.6	1882.5	11.2	3.3	20.0
Portugal	482	-2873	6858	46.3	-274.0	655.8	9.2	4.1	16.4
Romania	30952	2242	86274	1552.2	101.3	4353.8	6.6	3.9	8.6
Slovakia	2245	-655	5112	415.5	-121.8	945.3	13.3	5.8	19.3
Slovenia	827	-602	3209	399.5	-297.8	1555.1	6.9	4.4	10.1
Spain	4741	-3604	23795	100.3	-84.0	510.4	15.9	8.2	26.1
Sweden	-405	-1855	634	-44.1	-201.2	61.7	6.9	4.9	8.6
UK	-565	-5810	5322	-10.5	-96.9	80.6	5.7	3.8	8.1
All origins	7150	-5212	23202	301.0	-150.6	960.1	9.2	5.3	13.6
Germany	-	-	-	-	-	-	6.9	3.1	11.2

Notes: 20 yearly observations between 2000 and 2019. Total migration y_{jt}^{net} in total number of migrants. Migration rate $\tilde{y}_{jt}^{net}$ in migrants per one million inhabitants. Unemployment rate u_{jt} in percentage points.

nantly driven by inflows. The fact that outflows are higher in the same countries could be due to the fact that high-immigration countries have a large stocks of migrants already in the host country.

Overall, these statistics thus suggest interesting relationships, which I will now further explore in a rigorous causal inference setting focused on isolating the impact of a crisis shock on migration.

3.3 Identification Strategy

In this section, I describe and discuss my identification strategy. To start with, I discuss a selection on observables framework to showcase why that strategy does not provide the information I am interested in. I then discuss the benefits of a treatment and control comparison and introduce the natural experiment that I base this comparison on. Next, I define the criterion that I use to split the sample into treatment and control. Finally, I compare the treatment and control groups along several dimensions and display time series graphs that reveal some interesting patterns essential to answering my question of interest.

Note that to ease exposition I will focus on net migration, so that from now on I will use the terms migration and net migration interchangably. This is because net migration summarizes both inflows and outflows and is hence the crucial component to assess overall migration flows. As such it is also the central calibration object in spatial economic models. In Section 3.6 I will decompose the analysis into inflows and outflows. Note further that I report all of my main results for both total net migration y_{jt}^{net} and the net migration rate $\tilde{y}_{jt}^{net}$, whenever both dependent variables lead to well-specified empirial settings. The first one gives an indication of the absolute size of the real-world migration movements observed in the Euro crisis given the particular set of countries and shock sizes. It thus yields a very precise estimate of the actual effect size of the Euro crisis. The latter one normalizes for population size and thus insulates against the possibility that the number of migrants increased more in crisis countries due to higher migration potentials. Furthermore, it possesses greater external validity as it can be understood to better uncover the true underlying patters, i.e.

the fundamental relationship between a crisis and migration in general. Yet another way to view this is that the first specification weights countries by their population assigning more importance to larger countries whereas the second does not. Furthermore, providing both sets of results already demonstrates the robustness of my analysis to different specifications and thus complements the sensitivity checks in Section 3.5.

3.3.1 A Selection on Observables Framework

Setup A first approach to my question could be a standard Selection on Observables identification strategy. In particular, using unemployment as a proxy for economic conditions, one could pool the data over all years and all countries and regress total migration or the migration rate on unemployment. That is, one could run the following regression:

$$\tilde{y}_{jt}^{net} = \beta_0 + \beta_1 u_{jt} + \varepsilon_{jt}$$

Column (1) in Table 3.2 presents the results to this regression. According to these results an increase in unemployment by one percentage point increases migration from a country to Germany by roughly 45.3 people per one million inhabitants. With a p-value below 0.01 this result is strongly statistically significant. Column (2) shows that these results retain their statistic significance if standard errors are clustered at the country level. Finally, column (3) includes year and country fixed effects thus exploiting the panel structure of the data, which reduces the coefficient of unemployment 25.5.[12]

Results Note that the coefficients are all positive and thus have the expected sign - an increase in unemployment leads to an increase in migration to Germany. In particular, a deterioration of employment prospects at home leads to an increased net migration. This could

[12] Appendix Table C.1 presents analogous results for inflows and outflows.

Table 3.2: Linear Regressions Results

	$\tilde{y}_{jt}^{net}$		
	(1)	(2)	(3)
u_{jt}	45.3***	45.3**	25.5**
	(4.52)	(2.24)	(2.13)
Constant	63.3	63.3	-676.9***
	(0.65)	(0.30)	(-2.91)
Year FE			✓
Country FE			✓
Clustered SE		✓	
Adjusted R^2	0.0348	0.0348	0.4618
Countries	27	27	27
Observations	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
This table shows regressions with the net migration rate $\tilde{y}_{jt}^{net}$ as the dependent variable in all columns. Column (2) clusters standard errors at the country level and column (3) includes year and country fixed effects.

be either through an increased outflow of people and/or a reduced backflow of immigrants who emigrated to Germany previously. Again, I will not focus on whether and to what extent both of these channels contribute to the net migration result now, but will revisit this question in Section 3.6. In terms of magnitude the numbers imply that on average a one percentage point increase in unemployment in every European country leads to an increase in net migration of roughly $25.5 * 420 \approx 10,700$ people to Germany per year as there were on average 420 million people in EU countries other than Germany over the sample period. To assess economic significance it is useful to set this in relationship to the average yearly net

migration into Germany from all countries over the sample period, which roughly numbers 125,000. A back-of-the-envelope calculation thus suggests that an increase of one percentage point in the unemployment rate leads to a roughly $\frac{10{,}700}{125{,}000} = 8.6\%$ increase in net migration, suggestive of a highly economically significant value. These results are broadly in line with those in Bertoli et al. (2016), who also include lags of the unemployment rate, additional covariates and run their regression at a monthly frequency, albeit over a shorter time period.

Discussion It might thus be tempting to conclude that the basic linear regression above already answers the question this paper investigates. However, this is premature as a more thorough analysis reveals. First, the regression above is subject to the general criticisms applied to a selection on observables identifaction strategy. In particular, there is no way to exclude the possibility of endogeneity. For example, it could be the case that increased migration is caused by a decrease in local amenities in the origin countries, which might be correlated with unemployment.

Even more importantly, however, the selection on observables framework above does not actually address the research question I pose. In particular, I aim to establish to what extent a crisis, i.e. a sudden deterioration of economic conditions, impacts migration. The regression above does not only pick up the effect on migration from a crisis-induced increase in unemployment, but also the effect caused by low-frequency changes in unemployment. It is generally acknowledged that persistent differences in economic conditions that change only slowly over time are a strong predictor for migration, as evidenced by the persistently large migration movements from poorer eastern European countries to Germany described in Section 3.2.2.[13] The goal of this paper, however, is not to investigate to what extent persistent differences in living standards or low-frequency changes in unemployment lead to migration but rather to capture to what extent migration is used as a vehicle to smooth

[13]Another example is the constant migration flow from poorer central and south American countries to the US.

short-term divergences in economic conditions caused by an economic crisis. The selection on observables framework is therefore not suited to analyze this question, which motivates the identification strategy I will introduce now.

3.3.2 A Treatment vs Control Framework

In order to avoid the flaws of the Selection on Observables Framework, I use a Treatment vs Control Framework that is well suited to answer my research question, as I now demonstrate. This framework is derived from the cleanest causal identification setting - an ideal experiment. In such an experiment I could determine the causal effect of an economic crisis by randomly selecting a subset of countries, the "treatment group", expose them to a crisis and compare their migration flows to the non-selected countries, the "control group". In this way I could be certain that due to the random selection, any differences between the two groups are due to the treatment, i.e. the crisis, thus identifying the causal effect of a crisis on migration.

A Natural Experiment In reality, of course, such an ideal experiment is not feasible. Neither is it possible to expose a chosen set of countries to an economic crisis, nor would it be moral to do so. This means that I cannot rely on an ideal experiment and must instead look for a so-called "natural experiment", that is a real-world situation similar to an ideal experiment. In particular, for my question, I require two conditions to hold. First, a situation where some countries were exposed to a crisis whilst others were not. Second, the "treated" and "non-treated" countries need to be similar to each other with respect to their migration behavior. This second criterion, which is also known as the "parallel trends assumption", arises as my identification assumption. In particular, it embodies the idea that in the absence of the crisis, the crisis countries would have behaved exactly like the non-crisis

countries, so that I can interpret the difference in their migration streams as the causal effect of a crisis on migration.

Fortunately, the Euro crisis that started in 2010[14], satifies these conditions of a natural experiment. In particular, on the one hand some countries (mostly in southern Europe) were exposed to a large crisis featuring rapidly increasing unemployment and bond yields as well as collapsing economic growth. On the other hand, other countries (mostly in central and northern Europe) did not experience this crisis. This therefore satisfies the first requirement of a natural experiment. The second requirement, whether the crisis and non-crisis countries are actually sufficiently comparable with respect to their migration behavior remains to be assessed. In particular, one could argue that southern European crisis countries are very different from northern European non-crisis countries in many dimensions, so that the non-crisis countries are not a good comparison group for the crisis countries. Fortunately, the event-study framework introduced in Section 3.4 makes this an econometrically testable assumption.

Definition of Treatment and Control Groups The question now remains how exactly to group the 27 EU countries in the sample into treatment and control groups. First, a crisis could be characterized by various characteristics, such as falling GDP growth, or - in particular with respect to the Euro crisis - surging bond yields. However, since this paper is concerned with migration, I choose the unemployment rate as my proxy of the crisis state of a country. This is because the vast field of migration theory, to which this paper aims to contribute, has identified unemployment as the main driver of migration among the set of economic conditions. Importantly, note that this paper does not aim to determine whether it is actually the unemployment rate that drives migration in a crisis, but instead to what

[14]The exact start date of the crisis is hard to pin down but it is generally agreed to start at the very end of 2009. From the perspective of yearly data, this is very fortunate timing, as I can thus consider 2010 as the year the crisis started. It would be more difficult had the crisis started in the middle of a year as in that case one would need to account for a year that was both a crisis and a non-crisis year.

extent a crisis, with all its different features, drives migration. Nevertheless, in light of the theoretical literature this paper aims to aid, the unemployment rate emerges as an ideal proxy for the crisis state of a country. Also note that from now on I will use "crisis countries" and "treatment group" as well as "non-crisis countries" and "control group" interchangably as their crisis state determines which group they belong to.

This leaves the question of how exactly I use the unemployment rate to divide the 27 countries into treatment and control groups. The measure I use is the difference between the unemployment rate in 2012 and that in 2009, i.e.

$$\Delta_j^u = u_{j,2012} - u_{j,2009}.$$

The reasoning behind this definition is that it compares the unemployment rate at the height of the Euro crisis with that before the start of the Euro crisis and thus makes sure that only the additional unemployment resulting from the Euro crisis is taken into account. However, in Section 3.5 I show that my main results are robust to various other definitions of Δ_j^u.

I then define a country to be a crisis country if this difference is larger than some threshold $\bar{\Delta}$. Mathematically, I express the crisis state of country j as an indicator variable C_j, defined as follows

$$C_j = \begin{cases} 1, & \text{if } \Delta_j^u \geq \bar{\Delta} \\ 0, & \text{if } \Delta_j^u < \bar{\Delta}. \end{cases}$$

Finally, I use as a threshold $\bar{\Delta} = 1.5$, that is a country is considered to be a crisis country if its unemployment rate was at least 1.5 percentage points higher at the height of the Euro crisis than at its start. I am aware that this threshold appears arbitrary at this point and in fact I discuss in detail in Section 3.5 what happens as I vary this threshold. I show there that

the results presented in this text are robust to a large range of thresholds. However, I choose this threshold as my default threshold as its classification of crisis countries corresponds best to a subjective assignment of crisis status.

3.3.3 A Closer look at Treatment and Control Groups

In order to gain a better understanding of the treatment and control groups, it is useful to investigate their characteristics. First, consider the countries constituting each group. For the control group these are Austria, Belgium, Czechia, Denmark, Finland, France, Hungary, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Romania, Sweden and the UK. Analogously the treatment group consists of Bulgaria, Croatia, Cyprus, Greece, Ireland, Italy, Poland, Portugal, Slovakia, Slovenia and Spain. As expected the treatment group thus contains all major southern European countries that were at the center of the Euro crisis. On the other hand, the control group contains most northern and central European countries. Interestingly, note that the eastern European countries that have traditionally had large migration flows to Germany are split roughly evenly between the two groups. This is validation for the fact that I am now investigating crisis-driven migration instead of migration driven through persistent differences in economic conditions. Furthermore, as noted above, this validates the splitting criterion as it overlaps with a subjective assignment of crisis status. One might argue that there are some marginal countries in either group that could have been classified either way, though as I demonstrate in Section 3.5 these marginal countries do not drive the results.

Next, Table 3.3 shows some overview statistics for both crisis and non-crisis countries. Importantly, in each group the number of countries as well as the total and the average population are similar. This fact validates my splitting criterion as the overall migration potential, as measured by the people who could potentially move, is similar between the two

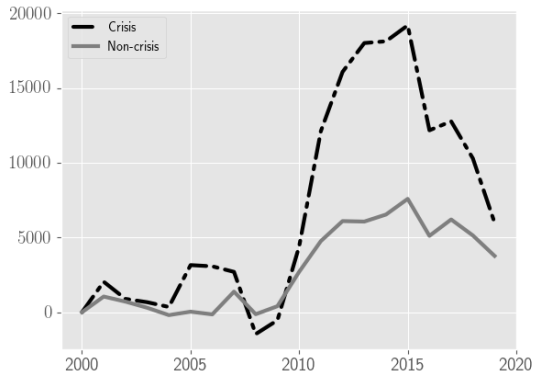
groups. Note that the average total and net migration as well as the unemployment rate are somewhat larger in the crisis countries. As evident from the discussion that follows now, these differences are mostly driven by the states of the two groups after the crisis occurred.

Table 3.3: Comparison of Crisis and Non-Crisis Groups

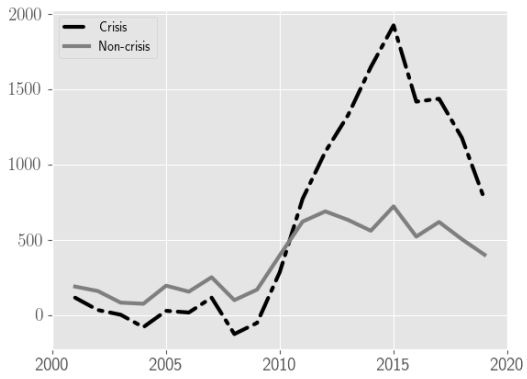
	# Countries	Total p_{jt}	Average p_{jt}	Average y_{jt}^{net}	Average $\tilde{y}_{jt}^{net}$	Average u_{jt}
Crisis	11	187.8	17.1	7113.3	601.6	11.0
Non-Crisis	16	231.8	14.5	2937.9	360.7	7.2

Notes: Total and average population in millions. Average unemployment in percentage points. Columns 2 and 3 defined as the sum and mean (over countries) of the (over time) average population of countries respectively. Columns 4, 5 and 6 defined as the average over both countries and time of y_{jt}^{net}, $\tilde{y}_{jt}^{net}$ and u_{jt} respectively.

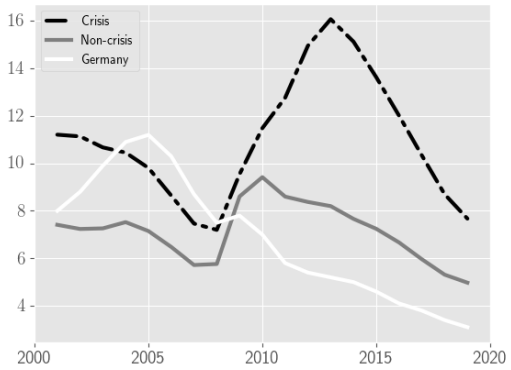
To further investigate this, Figure 3.1 displays time series graphs of the central variables for the entire time period 2000-2019. In particular, panel (a) displays the average total migration y_{jt}^{net}, panel (b) the average migration rate $\tilde{y}_{jt}^{net}$ and panel (c) the average unemployment rate u_{jt} across treatment and control groups. The patterns that emerge are very interesting. First, consider panel (c) which shows that even though unemployment rates were in generally slightly higher in the early 2000s in crisis countries, they converged towards those in non-crisis countries until 2008. At this point the unemployment rates in both groups increased strongly and in parallel with each other at the onset of the financial crisis. Crucially, however, unemployment in non-crisis countries flattened out in 2010 and started a steady decline. In contrast, unemployment in crisis countries started growing strongly again in 2010, reaching its peak only in 2013. Importantly, this shows that whilst both groups of countries were roughly evenly exposed to the adverse consequences of the financial crisis, unemployment profiles diverged strongly at the onset of the Euro crisis. This in turn validates the identification strategy as one can difference out the effects of the financial crisis that affected both groups similarly.



(a) Total Migration y_{jt}^{net}



(b) Migration Rate $\tilde{y}_{jt}^{net}$



(c) Unemployment Rate u_{jt}

Figure 3.1: Time Series Graphs

Notes: Crisis countries are those for which $C_j = 1$ and non-crisis countries are those with $C_j = 0$. All graphs show the yearly averages for all countries in each group. Germany is not included in either group and its unemployment rate is plotted for comparison in panel (c).

Since Germany holds a central position in my analysis as the only host country for migration, I have also included the German unemployment rate for comparison. Germany follows an unusual pattern in that its unemployment peaked in the early 2000s and has been steadily decreasing ever since. Its unemployment rate is thus significantly lower than that of any other major EU country, which validates the choice of Germany as the host country.

Next, panels (a) and (b) reveal a similar pattern for migration streams in both absolute numbers and per capita. Before the Euro crisis migration was very similar in both groups. It then started to increase to similar extents in light of the financial crisis in 2008. However, mirroring unemployment, whilst migration from non-crisis countries plateaued after 2011, migration from crisis countries continued its unabated and strong increase until 2015.

These graphs thus yield three very important insights for my further analysis. First, note that it is not surprising that the unemployment profiles between crisis and non-crisis countries diverge at the start of the crisis, as this is the exact criterion according to which they were split. Second and most importantly, before the Euro crisis both sets of countries are very similar with respect to both their unemployment and migration patterns. Crucially, this includes the financial crisis occurring shortly before the Euro crisis. In econometric terms, these pictures thus suggest that the "parallel trends" assumption discussed above is reasonable.[15] Again, for the subsequent analysis I will not rely on a visual inspection of time series graphs to verify this assumption, but will subject it to econometric testing in an event study framework. Nevertheless, it is very reassuring to see the parallel trends assumption seemingly holding in these plots. Third, the fact that migration diverges so strongly at the onset of the Euro crisis provides the first suggestive evidence that the impact of a crisis on migration flows might be potentially strong. Indeed, whilst the visual inspection of time series graphs again does not provide any formal results, the shape and magnitude of these curves does suggest the presence of a potentially strong causal effect.

Overall, the results thus suggest that the Euro crisis might well fit the conditions required of a natural experiment and indeed the presence of an economically significant causal effect of a crisis on migration. To further analyze this natural experiment I will now formalize the setting to derive econometrically rigorous results to my research question.

[15]To be precise, the parallel trends assumption only requires parallel trends in migration. The fact that the levels of migration and the trends in unemployment are also very similar provide further evidence for the comparability of treatment and control but are not necessary from an econometric point of view.

3.4 Main Results

In this section I present the main results of this paper. First, I introduce the main analytical tool used in this paper - an event study of the Euro crisis. In particular, I analyze both a non-parametric and various parametric event studies, which allow me to test the parallel trends assumption and derive a detailed time profile of the migration response to the Euro crisis. I then present and discuss the results of these various event studies.

3.4.1 Setup

I first set up the formal event study. The basic idea behind an event study framework is very similar to the popular difference-in-differences framework. In particular, it compares the difference between migration before and after the crisis in the treatment group to the same difference in the control group. The basic requirement as outlined above is that non-crisis countries are a good comparison for crisis countries.

In comparison to a simple difference-in-differences estimator, the event study setting provides three advantages. The first one is that it makes the identifying common pre-trends assumption empirically testable in a Granger-causality setting. Secondly, it allows a detailed analysis of the time profile of the effect. In particular, whilst a difference-in-differences analysis only yields one effect size, which is the sum of the effects over all post-event time periods, the event study gives an estimate for the size of the effect in each post-event time period separately. Thirdly, an event study does not require the events to happen at the same time in every country.

This last characteristic is less relevant in this application, where the event, i.e. the crisis, happened in all countries simultaneously. However, I will stick to the standard event study terminology which originates from this characteristic. In particular, I will name the periods

relative to the event timing instead of the actual years. This means that from now on I will define the year 2009 as the pre-event year t^*, and define all other time periods relative to it. For example, year 2010, i.e. the event year, is denoted as t^*+1, year 2011 as $t=t^*+2$, year 2008 as $t=t^*-1$ etc.

Having set this basic notation, the first event study that I run is the traditional event study, which I denote as the "nonparametric" event study henceforth. In particular, this nonparametric event study takes the following form:

$$y_{jt}^{net} = \delta_j + \lambda_t + \sum_{\tau=k_{min}}^{k_{max}} \beta_\tau T_{j,t^*+\tau} + \varepsilon_{jt}, \tag{3.1}$$

where δ_j are country fixed effects and λ_t year fixed effects that isolate the results from both country and year averages. $T_{j,t}$ are dummy variables that equal 1 in year t if country j is in the treatment group and 0 otherwise. $k_{min}<0$ and $k_{max}>0$ are the maximum number of pre- and post-crisis periods included in the event study. I set $k_{max}=10$ to trace out the full post-event time profile and $k_{min}=-8$, to grasp the entire pre-event time profile. That is, my event study includes an indicator for each year in the sample. The only exception is the pre-event year t^*, which is excluded, so that all other coefficients are comparisons relative to this pre-event year.[16] In this way the coefficients β_τ allow me to trace out the differential behavior of crisis countries relative to control countries in each year t compared to the pre-event year t^*. This setting thus has the above mentioned two advantages: One can use the coefficients $\beta_{k_{min}},...,\beta_{-1}$, called leads, to see if there are any systematic differences in the trends between the two sets of countries before the crisis. This allows me to test econometrically for the common pre-trend assumption through an F-test with the following

[16] That is, the treatment indicator $T_{j,0}$ is excluded from equation 3.1.

null-hypothesis:

$$H_0 : \beta_{k_{min}} = ... = \beta_{-1} = 0 \tag{3.2}$$

On the other hand, the coefficients $\beta_1, ..., \beta_{k_{max}}$, commonly referred to as lags, show the differential behavior of crisis and non-crisis countries after the crisis and trace out the entire time profile (the "impulse response" in macroeconomic terms) of the effect of the crisis on migration.

I call this event study "nonparametric" since it does not impose any parametric structure on pre or post trends and instead estimates an independent parameter for each time period. Whilst this flexibility can certainly be advantageous, it does not come for free. In particular, it requires the estimation of a large number of parameters, namely $k_{min}+k_{max} = 19$ treatment parameters, whilst there are only 27 observations per year in the data. To cope with the inherent data scarcity in this macroeconomic data set, I also run parametric event studies as in Lafortune et al. (2018). The idea here is to impose a parametric jump-structure on the trends and thus reduce the number of parameters to be estimated. This means that the estimate of each parameter is based on significantly more data, thus potentially increasing the confidence in those estimates.

The first specification models the crisis event as a permanent immediate shift in migration relative to non-affected countries:

$$y_{jt}^{net} = \delta_j + \lambda_t + \beta^{jump} \mathbb{1}\{t > t^*\} T_j + \varepsilon_{jt}, \tag{3.3}$$

where δ_j and λ_t again represent country and year fixed effects respectively, T_j is an indicator for whether country j is in the treatment group and the coefficient β^{jump} represents the change in migration following the crisis.

In addition to an immediate effect, the migration response to a crisis might have a dynamic component. For example, it is possible that migration does not rise immediately but that the increase takes time to unfold. Similarly, it is perceivable that after an initial jump the migration pressure is lower as the most affected people have already moved. In fact, both of these effects could be present over time, which is what the raw time series graph Figure 3.1 suggests. To take this into account, I incorporate a parabolic shape in my second event study. Furthermore, as a way to test for the common pre-trend assumption in a parametric setting I include a pre-event trend:

$$\begin{aligned} y_{jt}^{net} =& \delta_j + \lambda_t + \beta^{pre}\mathbb{1}\{t < t^*\}(t^* - t)T_j + \beta^{jump}\mathbb{1}\{t > t^*\}T_j + \\ &+ \beta^{post}\mathbb{1}\{t > t^*\}(t - t^*)T_j + \beta^{square}\mathbb{1}\{t > t^*\}(t - t^*)^2 T_j + \varepsilon_{jt}. \end{aligned} \tag{3.4}$$

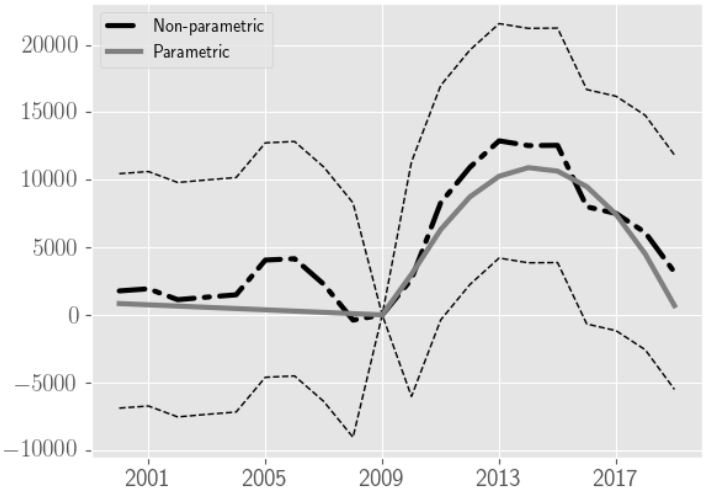
In addition to the elements present in the first event study, β^{post} now accounts for a post trend and β^{square} allows for this trend to be time-varying. Furthermore, β^{pre} accounts for a potential pre trend, so that testing $\beta^{pre} = 0$ with a t-test is the parametric way of testing for the identifying common trends assumption.

3.4.2 Results

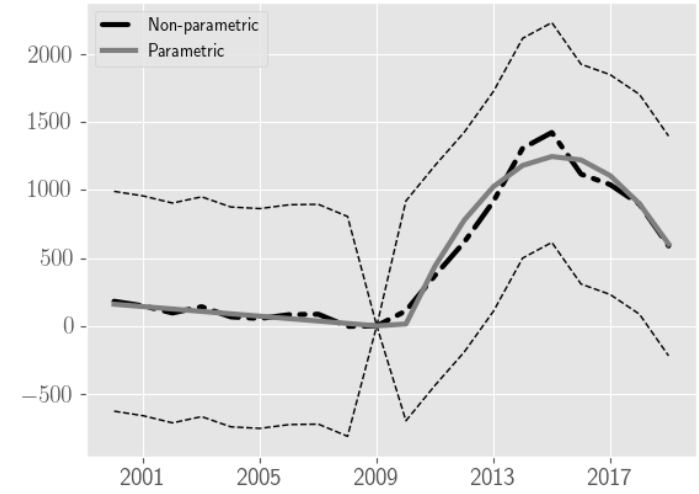
Having described the setup, let me turn to the results of these event studies, which in turn constitute the main results of this paper. To start with, Figure 3.2 displays the results of the nonparametric event studies in graphical form.

Consider first panel (a), which displays the results for total net migration y_{jt}^{net}. The non-parametric parameters $\{\beta_\tau\}_{\tau=k_{min}}^{k_{max}}$ are displayed as the black dash-dotted line, with their 95% confidence intervals being displayed as the thin dashed lines.[17] First, note that before the event, i.e. before 2009, the non-parametric estimates oscillate closely around 0, which

[17] Appendix Table C.6 displays the tabular results for the nonparametric events studies.



(a) Total Migration y_{jt}^{net}



(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure 3.2: Event Studies

Notes: Notes: Panel (a) displays nonparametric and parametric event studies for total net migration y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification.

is roughly at the center of the confidence intervals. This is strong evidence for the common trends assumption, which is further confirmed by the F-test in Equation 3.2. In particular, the p-value for that test equals 0.98, so that one cannot reject the common-trends assumption at any level of significance. Next, consider the effect of the crisis on migration. As the crisis hits in 2010, the pattern from the previous years changes suddenly and decisively with migration from crisis countries reaching roughly 13,000 net migrants compared to non-crisis countries in 2013. The yearly extra migration attributable to the Euro crisis then hovers at around that number until 2015 before declining back towards 0 at the end of the time period under consideration. Note that these effects are statistically significant at the 95% confidence level from 2012 to 2015, a remarkable fact in light of the data scarcity discussed above.

Regarding the parametric event studies, their results are displayed in column (1) of Table 3.4. Panel A refers to the first parametric specification in Equation 3.3 and panel B to the second parametric specification in Equation 3.4. The results of the second specification are also visualized as the solid grey line in Figure 3.2. Consider first this second specification, which confirms the results from the nonparametric event study to a remarkable extent. Again, one cannot reject the common pre-trend assumption, which has a p-value of 0.78. Whilst the immediate jump at the onset of the crisis of 3,000 extra migrants is not statistically significant, the parameters indicating a bell-shaped curve are strongly significant, even at the 1% confidence level. This means that this parametric event study serves its purpose in deriving similar results as the nonparametric event study with fewer and more precisely estimated parameters. In particular, it confirms both the initial upward trend up to 2013-2015 which is followed by a steady decline back towards zero.

In contrast, by design the first parametric event study attributes all the increase to an initial jump. This study thus tells us that in an average post-crisis year migration from a crisis-country increased by between 6,700-8,400, a useful overview statistic.

Table 3.4: Parametric Event Studies

	y_{jt}^{net}		$\tilde{y}_{jt}^{net}$	
	(1)	(2)	(3)	(4)
Panel A. Parametric Event Study 1				
Post event	6689.6***	6689.6	754.6***	754.6
	(4.80)	(1.54)	(5.79)	(1.65)
Constant	-3445.5	-3445.5**	-385.2*	-385.2***
	(-1.47)	(-2.23)	(-1.76)	(-3.82)
Panel B. Parametric Event Study 2				
Post event	3001.9	3001.9	13.5	13.5
	(0.93)	(1.16)	(0.05)	(0.05)
Post event × years elapsed	3761.3***	3761.3*	473.3***	473.3*
	(2.99)	(1.78)	(4.05)	(1.71)
Post event × (years elapsed)^2	-446.5***	-446.5*	-45.3***	-45.3**
	(-3.32)	(-1.87)	(-3.62)	(-1.99)
Pre event × years elapsed	-93.3	-93.3	-17.7	-17.7
	(-0.27)	(-0.51)	(-0.56)	(-0.83)
Constant	-3616.5	-3616.5**	-417.7*	-417.7***
	(-1.51)	(-2.10)	(-1.87)	(-3.48)
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Clustered SE		✓		✓
Countries	27	27	27	27
Observations	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. Column (1) displays results for total net migration y_{jt}^{net} and column (2) the same but with standard errors clustered at the country level. Analogously, column (3) presents results for the net migration rate $\tilde{y}_{jt}^{net}$ with standard errors clustered at the country level in column (4).

Regarding the aggregate crisis-driven increase in net migration over the entire time period one can derive this number in four ways. In particular, one can sum up all the parameters from the nonparametric event study or alternatively compute the aggregate number based

on the parametric specifications. As discussed, in this case the nonparametric and the second parametric event study give the best fit to the data, so that I will rely on their estimates which are very similar at around 85,000 additional net migrants per country over the entire post-crisis period. This yields a total additional net migration of $85,000 \times 11 = 935,000$ to Germany across all 11 crisis countries over the post crisis period 2010-2019.

Next, consider the same event studies for the migration rate $\tilde{y}_{jt}^{net}$ instead of total migration y_{jt}^{net}. Panel (b) in Table 3.4 displays the nonparametric event study as before. Once again the common trends assumption appears to hold very well graphically, with all nonparametric parameters being almost identical to 0. In particular the F-test in Equation 3.2 yields a p-value of around 1, so that once again the common trends assumption cannot be rejected.

The graph is very similar to the one from panel (a), with the profile just marginally moved to the right and the effects being again significant at the 95% level from 2012 until 2018. That means that whilst the initial increase occurs over a longer time period, reaching its maximum only in 2015, the decline is also less pronounced later on. This suggests that crisis countries with a larger population had a higher initial increase in migration compared to the average country in the treatment group and then also a larger decrease after the peak of the crisis. Both effects reach their maximum in year 2015 and decline from then onwards. The parametric event studies in column (3) of Table 3.4 confirm this. First, the pre-trend in the second specification is almost identically equal to 0 with a p-value of 0.58, confirming the validity of the common trends assumption. Again the specification does not suggest a jump on impact but instead a steady rise followed by a decreases, staying close to the nonparametric estimates. As above the first parametric specification attributes the entire increase to the initial jump term.

Overall, the results for both total migration y_{jt}^{net} and the migration rate $\tilde{y}_{jt}^{net}$ thus demonstrate a strong causal effect of the Euro crisis on migration. First, they strongly confirm the common trends assumption necessary for causal identification. Second, they show that

the time profile is generally characterized by a strong initial increase in extra migration until 2015 and a slow decrease afterwards. The total number of additional net migrants to Germany attributable to the crisis is on the order of 935,000.

3.4.3 Discussion

In order to assess the economic significance of those results, it is important to set these numbers in perspective. In particular, the cumulative net migration into Germany from other Euro countries after the crisis, i.e. between 2010 and 2019 stood at roughly 4.5m. My results show that 935,000 of those, and therefore roughly 20.7% of those immigrated as a result of the Euro crisis. In other words, the crisis increased net migration from other European countries into Germany by roughly 20.7%. This is certainly a seizable increase,in line with the newspaper coverage at the time and other anecdotal evidence. Of course, as discussed above, the increase was not homogeneous but the extra migration attributable to the crisis is significant from an economic point of view over the entire post-crisis period . I therefore conclude that the effect of the Euro crisis on EU-internal net migration was strongly positive raising migration by roughly 21%.

An additional result from this specification is that migration is not an immediate phenomenon but that migration pressures are released with a lag and over multiple years. In particular, my results suggest that the maximum migration occurs roughly five years after the start of the crisis and two years after the height of the crisis, which occurred in 2013 as measured by the maximum unemployment rates in Figure 3.1. As a matter of fact, even ten years after the crisis, migration is only just now converging back towards the pre-crisis trends. To derive realistic quantitative results it is thus crucial for migration models to take this lagged behavior into account.

Furthermore, note that after the increased net migration caused by the crisis, migration converges back to its pre-crisis state. This means that I do not observe a reversal of migration after the crisis, which would be required for populations to settle back into their pre-crisis steady states. My results thus provide strong evidence for hysteresis, meaning that the temporary crisis shock has permanent effects on the population distribution across countries.

Of course, these results only have meaning if identification is ensured. I have demonstrated above that the common-trends assumption holds very well, removing the biggest threat to identification.[18] However, let me discuss some other potential issues. First, one might argue that the crisis countries were selected endogenously, as after all the Euro crisis did not affect a random subset of countries. In fact, countries were more affected if they had higher debt or more fragile economies. Does this endogenous selection invalidate my research design? Fortunately, the answer is no. Whilst the subset of countries affected by the crisis is indeed not random, this does not matter as long as the common trends assumption holds. This is because the common trends assumption assures that the countries are comparable in all factors relevant for their migration behavior. All their pre-crisis differences that might have affected their selection into the crisis were therefore inconsequential with respect to their migration interactions with Germany.

In order for the research design to be invalidated, the crisis would need to be accompanied by a separate shock that occurs simultaneously and affects only the countries in the treatment group. Arguably, many things happened at once at the start of the crisis. However, note that my research question does not aim to examine the impact of one element of the crisis, such as increased unemployment, but instead that of the crisis overall, i.e. of all of its components in conjunction. To the best of my knowledge there was no event apart from the Euro crisis happening only to the crisis countries in 2012, so the identification stands.

[18]Notice here the importance of a similar behavior of treatment and control groups in the financial crisis that preceded the Euro crisis. One can see that both sets of countries behaved very similarly, which is part of the identification assumption.

Finally, one might argue that there exist network effects. The fact that some countries were affected by the crisis might have also affected migration from non-crisis countries. For example, one could argue that due to large migration from crisis countries, migration from non-crisis countries decreased as a result of the increased competition on the German labor market. However, given the large size of the German labor market and the continuously low and decreasing unemployment rate during the crisis, this story seems implausible. Alternatively, one might argue that there were spillover effects from crisis countries into non-crisis countries given the close interconnectedness of the European economy. As a result, non-crisis countries might have also suffered more and thus increased their migration. This story seems implausible given the unemployment profile in Figure 3.1. If anything, unemployment decreased in non-crisis countries during the Euro crisis. However, even if this story were to be true, this would imply a downward bias for my results, so that they could still be regarded as a conservative lower bound for the actual increase.

Apart from the threats to identification I want to finish this section with a discussion of the appropriate handling of standard errors. Note that the results presented so far do not cluster standard errors. It would be a well-used strategy to cluster standard errors at the country level in my event studies to try to control for country specific time-series correlation of the error term. In fact, I do this in columns (2) and (4) of Table 3.4 and in columns (2) and (4) of Appendix Table C.6. Note that while the results are still mostly significant, the standard errors are significantly larger.

The reason why I don't cluster my standard errors in the main specifications is again related to the small data sample available. In particular, two issues arise. First, by clustering at the country level, each cluster is composed of only 20 observations (20 yearly observations for each country), a small sample which significantly increases the standard erros. Second, there are only 27 countries and thus 27 clusters. As is well established in the econometric literature (see Rokicki et al., 2018; Cameron & Miller, 2015; Donald & Lang, 2007; McCaffrey

& Bell, 2006) clustered standard error estimates are consistent but biased and it is generally not recommended to use them with less than 50 clusters.

For both of these reasons I rely on non-clustered standard errors. However, whilst the nonparametric estimates become somewhat less sharp with clustered standard errors, it is very reassuring that the parametric results are still statistically significant at the 1% level, confirming their usefulness as a way of handling the scarcity of the data.

3.5 Robustness

In this section, I demonstrate that my results are robust to a wide range of alternative specifications. In particular, I check for different definitions of the unemployment differential Δ_j^u, different thresholds $\bar{\Delta}$ and the omission of certain countries.

Unemployment differential Δ_j^u Remember that in the baseline definition I defined the unemployment differential as the difference between the unemployment rate in 2012 and that in 2009. The rational for this was to compare unemployment at the height of the Euro crisis to that just before its start to isolate the unemployment stemming from the Euro crisis from that stemming from the previously occurring financial crisis.

The first robustness check is based on the observation that 2012 does not necessarily constitute the highpoint of the Euro crisis, since unemployment in crisis countries did not reach its maximum until 2010 as seen in panel (c) of Figure 3.1. As a result I repeat the exercises with a different definition:

$$\Delta_j^u = u_{j,2013} - u_{j,2009}$$

The results for the parametric event studies for the migration rate are displayed in column (1) of Table 3.5 and those for total migration in Appendix Table C.9. Furthermore, I display

the event study graphs in Appendix Figure C.8 and the exact nonparametric results in Appendix Table C.7. Note that the results do not change compared to the baseline number. The reason for this is that there are only 27 countries in the data set and this different definition of Δ_j^u does not move any country from one group into another. Therefore, the results are completely robust to this different definition.

Next, some might argue that one should include the migration gain from the financial crisis. To take this into account I again redefine the unemployment differential as the difference between unemployment at the height of the Euro crisis and that before the financial crisis:

$$\Delta_j^u = u_{j,2012} - u_{j,2008}$$

Note that this different definition means that the average unemployment differential is substantially higher, since the financial crisis caused large unemployment in all countries.[19] The results for the parametric event studies for the migration rate can be found in column (2) of Table 3.5 and the remaining results in the Appendix. Again the results are robust to this change. Note however, that the baseline definition is more suited for my analysis since it neatly identifies those countries affected by the Euro crisis. In contrast, the treatment group here are those countries most strongly affected by the sum of Euro and financial crisis. The reason for the similarity of the results is that the financial crisis affected most countries to a very similar extent, so that this criterion also mostly divides based on the strength of the Euro crisis unemployment shock.

[19] For this reason, I set a new threshold $\bar{\Delta} = 2.9$, since otherwise all countries would be classified as crisis countries. This threshold again corresponds to a classification of countries most in line with a subjective assessment of their crisis status.

Table 3.5: Robustness Event Studies

	$\tilde{y}_{jt}^{net}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Parametric Event Study 1						
Post event	701.9***	643.8***	515.3***	332.9**	415.0***	315.7***
	(5.36)	(4.89)	(3.95)	(2.44)	(4.76)	(5.03)
Constant	-395.9*	-348.2	-414.9*	-434.1*	-405.4***	-329.4***
	(-1.80)	(-1.56)	(-1.86)	(-1.91)	(-2.83)	(-3.28)
Panel B. Parametric Event Study 2						
Post event	-36.9	394.3	-35.4	-10.7	190.8	147.9
	(-0.12)	(1.29)	(-0.12)	(-0.03)	(0.95)	(1.03)
Post event × years elapsed	475.7***	276.0**	372.1***	299.8**	223.8***	216.4***
	(4.05)	(2.31)	(3.16)	(2.43)	(2.84)	(3.86)
Post event × (years elapsed)^2	-45.7***	-31.5**	-35.2***	-31.0**	-25.7***	-25.1***
	(-3.64)	(-2.47)	(-2.80)	(-2.35)	(-3.05)	(-4.18)
Pre event × years elapsed	-17.8	-17.3	-21.8	-22.0	-9.38	-16.4
	(-0.56)	(-0.54)	(-0.68)	(-0.66)	(-0.44)	(-1.08)
Constant	-428.7*	-394.2*	-462.0**	-496.3**	-421.6***	-357.1***
	(-1.91)	(-1.66)	(-1.99)	(-2.02)	(-2.87)	(-3.50)
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Countries	27	27	27	27	26	24
Observations	540	540	540	540	520	480

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. In all columns the dependent variable is the net migration rate $\tilde{y}_{jt}^{net}$. The first two columns vary the definition of Δ_j^u with column (1) $\Delta_j^u = u_{j,2013} - u_{j,2009}$ and in column (2) $\Delta_j^u = u_{j,2012} - u_{j,2009}$. The next columns vary the threshold $\bar{\Delta}$ where in column (3) $\bar{\Delta} = 1$ and in column (4) $\bar{\Delta} = 0$. The final two columns vary the countries included, with column (5) excluding Croatia and column (6) excluding Croatia, Romania and Bulgaria.

Threshold $\bar{\Delta}$ Next, I will investigate to what extent changing the threshold $\bar{\Delta}$ affects the results. First, I reduce $\bar{\Delta}$ from 1.5 to 1.0, the parametric results of which are shown in column (3) of Table 3.5, with more details provided in the Appendix. This change in the threshold means that countries whose unemployment differential is betwen 1.0 and 1.5 move from the control group into the treatment group. In particular, those are two countries, Denmark and the Netherlands, so that the treatment group grows from 11 to 13 countries.

Importantly, the common trends assumption still holds well and the results do not change qualitatively. In particular, in terms of the second parametric event study one can see that the linear post-event coefficient β^{post} is still positive whilst the squared post-event coefficient β^{square} is negative. Regarding the quantitative repercussions, it is interesting to see that the coefficients are somewhat smaller, indicating a lower migration response. This makes sense in light of the average unemployment differences between treatment and control (and any other crisis characteristics) being not as pronounced as before, thus leading to a lower average migration response. However, by looking at the nonparametric coefficients in Appendix Figure C.5 and Appendix Table C.13 it is clear that the quantitative differences are small - in line with what one would expect by moving two marginal countries into the control group.

Next, I reduce the threshold further to 0.0, the parametric results of which are displayed in column (4) of Table 3.5. I am particulary curious about this case, since it classifies all countries as crisis countries that experienced any unemployment increase, a definition that might seem the most intuitive. Note that in this case the treatment group consists of 17 countries, with Czechia, France, Hungary and the UK joining the treatment group. As expected the pattern still holds, though it is even more subdued quantitatively. This is to be expected as with such a low threshold countries with very low increases in unemployment enter the treatment group and thus further dilute the migration response.

Notice that further lowering the threshold continues this trend, though eventually the pre-trend ceases to hold as only very few and small countries remain in the control group. Conversely, increasing the threshold initially leads to larger results as only the most strongly affected countries remain in the treatment group, though again results do not differ by much. Overall, this second sequence of robustness checks thus confirms my main results. Importantly, it demonstrates the strength of my identification assumption which holds in all of these configurations, demonstrating that EU countries are valid controls for each other,

independently of their vulnerability to the Euro crisis. However, the baseline threshold of 1.5 seems to deliver the most reliable quantitative results as the countries classified in the treatment group correspond closely to those commonly associated with the Euro crisis and lowering the threshold thus dilutes the results by adding unaffected or barely affected countries into the treatment group.

Exclusion of countries Among the countries in the data set, three stand out from the rest as they joined the EU much later. In particular, Bulgaria and Romania entered the EU in 2007 and Croatia in 2013. This means that data from those countries might be less comparable as they faced different migration hurdles for a sustained period. Since all three countries are among the main sources of immigration into Germany in terms of absolute numbers over the sample period, this could be reason for concern. Therefore, in order to make sure that this exceptional situation does not drive the results, my last two robustness checks control for this, by redoing the experiments without data from those countries. In particular, column (5) in Table 3.5 drops all observations from Croatia and column (6) additionally drops observations from Bulgaria and Romania. As it turns out, the only effect this has is to slightly lower the migration response in quantitative terms, which makes sense in light of the outsized importance those countries hold in terms of overall migration. I thus conclude that my results are not driven by the unique circumstances of these countries.

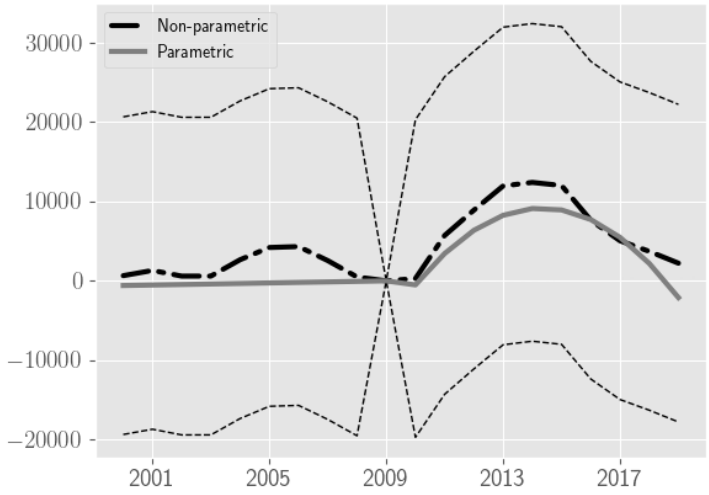
3.6 Decomposition

In this final section, I decompose net migration, the study object up to this point, into its components - inflows and outflows. Whilst arguably net flows are the most important single statistic that summarizes movements in both directions, it is very interesting to see where exactly the effects described above originate. To what extent are they due to more people

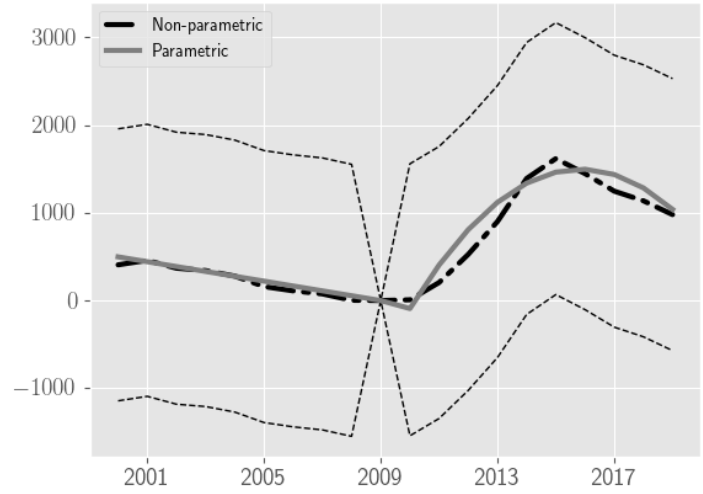
immigrating into Germany versus people leaving Germany at a lower rate? This question is of high importance to policymakers, since new immigrants and old immigrants can differ along a range of dimension. For example, migrants already in the country are more likely to have acquired some language knowledge and have found jobs, increasing their labor market participation relative to new migrants. Furthermore, from the perspective of macroeconomic migration models it can also be important to learn about the decomposition of net migration. Whilst most current models calibrate their parameters based on net migration numbers, it would yield even greater credibility to them were they able to match both inflows and outflows. For these reasons, I now decompose net migration, first considering the effects of the Euro crisis on inflows and then on outflows.

3.6.1 Inflows

To start with, consider a comparison of treatment and control groups akin to that in Section 3.3.3. The first two columns in Appendix Table C.4 show that average migration is higher in crisis countries than in non-crisis countries both in total numbers and per capita. Looking at Appendix Figure C.1 it becomes apparent that the reason for this divergence differs when considering total inflows and the inflow rate. In particular, in terms of total inflows, crisis countries have consistently lower inflows during the entire sample period, whilst in per capita terms the difference only emerges at the start of the crisis. Crucially for the purpose of the analysis, the common trends assumption seems to hold in the data for both total inflows and the inflow rate. Whilst for total inflows, the level differs but the trend is very similar, both level and trend are very similar up to the crisis for the inflow rate. As soon as the crisis hits, inflows, as measured in both total inflows and the inflow rate, increase rapidly, reaching their peak in 2015 and decreasing slowly afterwards.



(a) Inflows y_{jt}^{in}



(b) Inflow Rate $\tilde{y}_{jt}^{in}$

Figure 3.3: Event Studies Inflows

Notes: Panel (a) displays nonparametric and parametric event studies for total inflows y_{jt}^{in} and panel (b) for the inflow rate $\tilde{y}_{jt}^{in}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification.

Encouraged by these initial findings, Figure 3.3 shows the formal nonparametric event studies with total inflows y_{jt}^{in} as the dependent variable in panel (a) and the inflow rate $\tilde{y}_{jt}^{in}$ in panel (b). Note that in both graphs the common trends assumption holds well by inspection and indeed the Null-Hypothesis of all leads equalling zero cannot be rejected at any significance level with p-values roughly equal to one for both event studies.

Having thus validated the main identification assumption, consider what happens as the crisis hits in 2009. In general, there is a strong increase in inflows up to 2015, which is followed by a decline back towards zero afterwards. Quantitatively, the peak is at roughly 12,000 total incoming migrants per crisis country and 1,600 immigrants per one million inhabitants. This means that quantitatively, inflows account for the overwhelming majority of the increase in net migration discussed previously. However, in comparison to the net migration graph, one difference emerges. Whilst net migration starts its rise immediately at the start of the crisis, inflows start to increase with a lag of one year, i.e. in 2011, with the immediate response in 2010 being close to zero.

Whilst the pattern is familiar and strongly economically significant, note that the standard errors are significantly larger than in the net migration case. This is most likely due to the inherent data scarcity. To cope with this, consider next the parametric event studies in columns (1) and (2) of Table 3.6. The first parametric event study yields a statistically significant increase. This is an important results as it indicates that there was a statistically significant change post-crisis compared to pre-crisis when ignoring the exact timing profile. Parametric event study 2 takes the hump-shaped pattern into account and thus indeed replicates the nonparametric pattern quite well as evidence by the solid grey line in Figure 3.3. The coefficients on this pattern are also statistically significant, with the only exception being the linear coefficient post crisis, which while being quantitatively large, narrowly misses the 10% significance level. The parametric event studies thus confirm the nonparametric results while being mostly successful in handling the data scarcity.

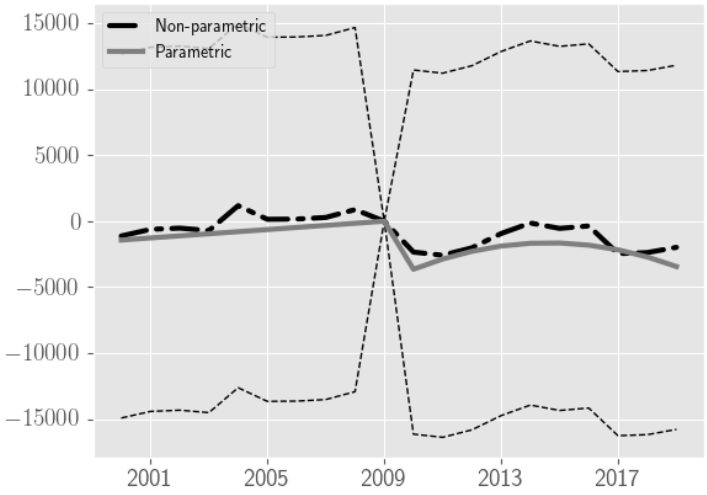
Table 3.6: Parametric Event Studies Inflows and Outflows

	y_{jt}^{in}	$\tilde{y}_{jt}^{in}$	y_{jt}^{out}	$\tilde{y}_{jt}^{out}$
	(1)	(2)	(3)	(4)
Panel A. Parametric Event Study 1				
Post event	5253.2*	726.1***	-1540.5	-39.2
	(1.66)	(2.94)	(-0.71)	(-0.27)
Constant	10431.0*	1436.8***	14259.1***	1865.2***
	(1.96)	(3.46)	(3.89)	(7.52)
Panel B. Parametric Event Study 2				
Post event	-506.1	-91.2	-3641.0	-118.5
	(-0.07)	(-0.16)	(-0.71)	(-0.34)
Post event × years elapsed	4471.5	543.0**	866.7	85.8
	(1.55)	(2.43)	(0.44)	(0.64)
Post event × (years elapsed)^2	-516.6*	-46.4*	-93.8	-3.51
	(-1.67)	(-1.94)	(-0.44)	(-0.24)
Pre event × years elapsed	65.4	-55.4	158.7	-37.6
	(0.08)	(-0.91)	(0.29)	(-1.04)
Constant	10550.9*	1335.4***	14550.0***	1796.2***
	(1.91)	(3.12)	(3.83)	(6.99)
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Clustered SE				
Countries	27	27	27	27
Observations	540	540	540	540

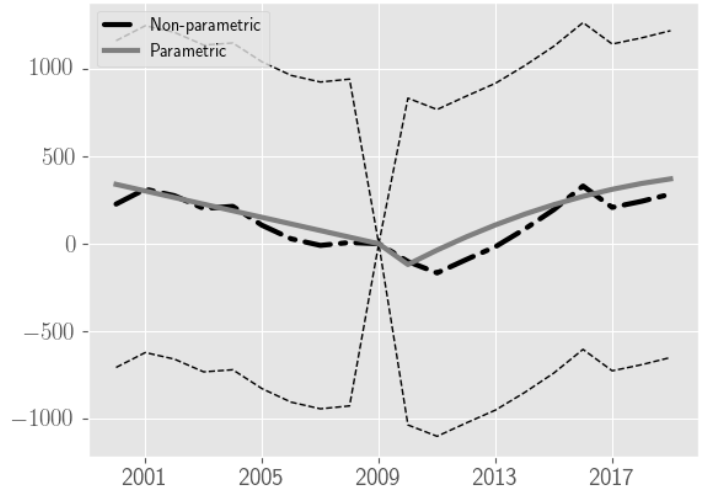
Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. Column (1) displays results for total inflows y_{jt}^{in} and column (2) for the inflow rate $\tilde{y}_{jt}^{in}$. Columns (3) and (4) do the same but for total outflows y_{jt}^{out} and the outflow rate $\tilde{y}_{jt}^{out}$.

3.6.2 Outflows

The last two columns in Appendix Table C.4 show significantly larger average outflows from crisis countries than from non-crisis countries during the sample period. Appendix Figure C.2



(a) Outflow y_{jt}^{out}



(b) Outflow Rate $\tilde{y}_{jt}^{out}$

Figure 3.4: Event Studies Outflows

Notes: Panel (a) displays nonparametric and parametric event studies for total outflows y_{jt}^{out} and panel (b) for the outflow rate $\tilde{y}_{jt}^{out}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification.

decomposes these aggregate numbers into their time profiles. Note that whilst graphically the common trends assumption again seems reasonable, the divergence after the crisis, that was so stark in both net migration and inflows, is not apparent in these graphs.

To further investigate, Figure 3.4 displays the nonparametric event studies for total outflows y_{jt}^{out} in panel (a) and the outflow rate $\tilde{y}_{jt}^{out}$ in panel (b). These graphs show more formally that there is no clear break in the trend at the start of the crisis. As a matter of fact both leads and lags oscillate around zero. This is further confirmed by the confidence intervals that are very wide so that no single nonparametric coefficient is close to statistically significant. To show that this is not due to data scarcity, the parametric event studies in columns (3) and (4) of Table 3.6 confirm the results. The coefficients are both quantitatively small and statistically insignificant with small p-values.

It therefore seems that in contrast to inflows, outflows were not strongly affected by the migration crisis. The only possible exception is revealed by looking closely at the graph and the parametric jump coefficients. Indeed, it appears that at the start of the crisis outflows slightly and temporarily decrease. Whilst this result is also not statistically significant there might be some evidence for a small initial drop.

To what extent are these results surprising? On the one hand, one might have expected outflows to decrease as people already in the host country might be less likely to return to their home country due the deteriorating economic conditions there. On the other hand, it appears feasible that higher inflows lead to higher outflows as a proportion of migrants might not be happy with their situation in the host country, for example by not being able to find a job there either. The fact that overall outflows appear to not change might thus indicate that these two effects cancel each other out.

Overall, whilst being less sharp in terms of statistical significance than the net migration results, the following story emerges from studying the decomposition. As the crisis hits, migrants already in the country are initially more hesitant to return to their crisis-hit home

countries. However, this effect seems to be short lived and there is no evidence for a protracted reduction in outflows. On the other hand, after a lag of one year, inflows rapidly pick up and contribute most of the increased net migration discussed above. The initial lag is interesting, as it suggests some kind of friction in the migration process. In particular, this could be due to people trying to find a job in their home country first before eventually deciding to migrate in search of better opportunities. For macroeconomic models aiming to match migration patterns, it is thus important to generate large and slightly lagged inflows without simulatenously affecting outflows.

3.7 Conclusion

Labor mobility plays a key role for the adjustment to country-specific shocks in a currency union. To investigate the strength of this adjustment mechanism, this paper investigates the responsiveness of migration to a large economic shock. I use the Euro crisis as a natural experiment and compare the net migration into Germany from countries affected by the Euro crisis to those unaffected by it in an event study setting. In particular, I demonstrate that the Euro crisis indeed constitutes a valid natural experiment by showing that the non-crisis countries are a valid control group for the crisis countries. I then study the time profile of the migration response through both parametric and non-parametric event studies. The causal effect of the crisis on migration is strong, accounting for roughly 935,000 extra migrants to Germany between 2010-2019, which accounts for about 21% of total net migration to Germany from EU countries over that period. Beyond this quantification, I highlight three additional facts about the migration response. First, migration is a medium-run rather than short-run adjustment mechanism to economic shocks. Migration starts to increase at the start of the crisis but reaches its peak five years later and reverts to its pre-crisis level only ten years after the crisis. Second, the vast majority of additional net migration is accounted

for by an increased inflow of migrants with outwards migration staying roughly constant. Finally, there is strong evidence for hysteresis in populations across countries, as the end of the crisis does not results in significant reverse migration that would restore pre-crisis populations in origin and destination countries.

This paper thus fills a crucial gap in the literature by helping economists and policy-makers alike assess the quantitative significance and qualitative pattern of the migration response to a crisis. It thus connects to a vast quantitative macroeconomic migration literature by providing these models with causally identified moments that can be used to connect quantitative DSGE models to the data in line with Nakamura & Steinsson (2018).

Since this paper is the first to directly address the question of how an economic crisis affects migration movements, it opens up a range of interesting research questions. First, incorporating data on the skills of migrants could allow interesting conclusions about the mobility of different skill groups in response to a crisis. Second, it would be interesting to see to what extent the numbers in this paper depend on the freedom of movement in the European Union, i.e. to what extent open borders increase the role of migration as an adjustment mechanism to economic shocks. Finally, an exact quantification of the insurance value of migration would be of great value to policy makers. I approach this question next with a large-scale nonlinear quantitative model in Fernández-Villaverde et al. (2021).

Appendix A

Appendix for Chapter 1

A.1 Proof for continuous HJB

Here we provide a heuristic proof of the HJB equation which highlights the link with the discrete-time formulation.

In particular, we show that the discrete-time problem (1.3) is equal to the continuous-time HJB (1.1) as $dt \to 0$. We start with the discrete-time Bellman equation:

$$V(\mathbf{x}_t) = \max_{\alpha} r(\mathbf{x}_t, \alpha) + \beta \mathbb{E}\left[V(\mathbf{x}_{t+1})\right].$$

If we express it in continuous time over a fixed interval of size Δt and define $\rho \equiv -\frac{1}{\Delta t}\log\beta$, we get

$$V(\mathbf{x}_t) = \max_{\{\alpha_{t+s}\}_{s\in[0,\Delta t]}} \int_0^{\Delta t} r(\mathbf{x}_{t+s}, \alpha_{t+s})\, ds + e^{-\rho\Delta t}\mathbb{E}\left[V(\mathbf{x}_{t+\Delta t})\right].$$

Given the optimal policy α^*, if we take the derivative with respect to Δt, we obtain

$$0 = r\left(\mathbf{x}_{t+\Delta t}, \alpha^*_{t+\Delta t}\right) - \rho e^{-\rho\Delta t}\mathbb{E}\left[V(\mathbf{x}_{t+\Delta t})\right] + e^{-\rho\Delta t}\mathbb{E}\left[\frac{dV(\mathbf{x}_{t+\Delta t})}{dt}\right].$$

Applying Itô's lemma and taking expectations,

$$\begin{aligned}\frac{1}{dt}\mathbb{E}\left[dV\left(\mathbf{x}_{t+\Delta t}\right)\right] &= \frac{1}{dt}\mathbb{E}\left[\left(\nabla_x V\left(\mathbf{x}_{t+\Delta t}\right)\right)^T f\left(\mathbf{x}_{t+\Delta t}, \alpha^*_{t+\Delta t}\right)\right. \\ &+ \left.\frac{1}{2}Tr\left[\left(\sigma\left(\mathbf{x}_{t+\Delta t}\right)\right)^T \Delta_x V\left(\mathbf{x}_{t+\Delta t}\right)\sigma\left(\mathbf{x}_{t+\Delta t}\right)\right]\right] d\left(t+\Delta t\right) \\ &+ \underbrace{\frac{1}{dt}\mathbb{E}\left[\left(\nabla_x V\left(\mathbf{x}_{t+\Delta t}\right)\right)^T \sigma\left(\mathbf{x}_{t+\Delta t}, \alpha^*_{t+\Delta t}\right) dZ_{t+\Delta t}\right]}_{=0},\end{aligned}$$

where the last term is zero due to the properties of the Brownian motion. Taking the limit as Δt goes to zero, this yields

$$\begin{aligned}0 &= r\left(\mathbf{x}_t, \alpha^*_t\right) - \rho V\left(\mathbf{x}_t\right) + \left(\nabla_x V\left(\mathbf{x}_t\right)\right)^T f\left(\mathbf{x}_t, \alpha^*_t\right) \\ &+ \frac{1}{2}Tr\left[\left(\sigma\left(\mathbf{x}_t\right)\right)^T \Delta_x V\left(\mathbf{x}_t\right)\sigma\left(\mathbf{x}\right)\right].\end{aligned}$$

Finally, rearranging terms and making use of the fact that the policy is optimal, we obtain the HJB equation:

$$\rho V\left(\mathbf{x}\right) = \max_{\alpha} r\left(\mathbf{x}, \alpha\right) + \left(\nabla_x V\left(\mathbf{x}\right)\right)^T f\left(\mathbf{x}, \alpha\right) + \frac{1}{2}Tr\left[\left(\sigma\left(\mathbf{x}\right)\right)^T \Delta_x V\left(\mathbf{x}\right)\sigma\left(\mathbf{x}\right)\right].$$

A.2 Implementation Details

In this section, we briefly outline the various implementation choices that we made and alternative options available.

Number of Neurons and Layers There is no clear guidance in the deep learning literature as to how many neurons or layers to use. The general recommendation is to start with a relatively low number of layers and increase them only if needed. We have experimented with different numbers of layers for all of our models and whilst the accuracy gains from

two to three layers were substantial, a higher number of layers did not have similar effects. Thus, we employ a neural network with three hidden layers in all our models.

Until a few years ago, it was convention to reduce the number of neurons with each hidden layer. Recently, however, empirical evidence seems to suggest that a reduction of the number of neurons is only advantageous in deep convolutional networks that are mostly used for image recognition. For all other networks, it is now conventional to use the same number of neurons for each layer. We have experimented with both approaches and for our application we have achieved greater accuracy with the same number of neurons in each layer. In particular, for all neural networks in both versions of the neoclassical growth models we employ eight neurons per layer, whilst we employ 128 neurons in each layer in the dynamic capital allocation model and 64 neurons in each layer in the migration model.

Activation Function Considering the activation function, the most commonly used ones are the following:

- Logistic: $\phi(x) = \frac{1}{1+e^{-x}}$
- tanh: $\phi(x) = \tanh(x)$
- ReLU: $\phi(x) = \max\{0, x\}$
- leaky ReLU: $\phi(x) = \begin{cases} \alpha x, & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$

  α is either set to a conventional value of 0.01 or is also learned by the model in which case it is known as parametric leaky ReLU.
- ELU: $\phi(x) = \begin{cases} \alpha(e^x - 1), & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$.

Until a few years ago, sigmoidal functions like logistic or tanh were commonly used, since these are closest to the activation functions found in human brains. However, in many modern applications ReLU-style application functions are preferred. The reason for this is that the gradients of sigmoidal activation functions zero out quickly, so that learning stops, while the gradients of ReLU-class functions do not die out. With sigmoidal activation functions one thus commonly observes "neuron death", meaning that the gradients of a neuron are basically at zero, making further learning of this neuron all but impossible. Xu et al. (2015) evaluated different activation functions and generally find the following:

Logistic $<$ tanh $<$ ReLU $<$ leaky ReLU $<$ parametric leaky ReLU $<$ ELU

For our applications on the other hand, tanh significantly outperforms all other activation functions. This is due to the smooth shape of tanh that corresponds better to the mostly smooth functions we are approximating. For this reason, we use tanh as our activation function in all of our applications.

Weight Initialization It is now widely accepted that the right initialization is fundamental for the convergence of the neural network. The reasoning behind this is quite simple. Since with sigmoidal activation functions gradients disappear for very large and small inputs, it is essential that the inputs are centered around zero. With traditional $N(0, 1)$ initialization schemes, it has been shown empirically that we often observe vanishing (or sometimes exploding) gradients. The deeper in the network we go, the lower the gradients are so that no learning takes place there.

Glorot & Bengio (2010) propose an initialization scheme that counteracts this.[1] In particular, the authors propose that what is needed in order to avoid vanishing gradients is to have the variance of outputs equal the variance of the inputs for each layers and similarly to have the gradients have equal variance before and after flowing through a layer. This

[1] In fact, this paper was so successful that it is one of the key reasons for the renewed interest in neural networks.

ensures a smooth forward and backward flow through the network that is not diminishing the gradients ever further. Denoting by n_{input} and n_{output} the number of input and output connections of a layer, they argue that the optimal initialization is a normal distribution with mean 0 and standard deviation $\sigma = 4\sqrt{\frac{2}{n_{input}+n_{output}}}$. This is widely known as Xavier Initialization.[2]

Input Normalization To further reduce problems with vanishing gradients we employ input normalization, by standardizing each neural network input to have a mean of zero and a standard deviation of one. More advanced input normalization schemes like batch normalization did not lead to improvements for us and instead slowed down convergence significantly. This is due to the fact that with three layers our networks are not as exposed to the vanishing gradient problem as deeper networks and costly fixes are therefore not necessary for us.

Optimizing Algorithm - Adam Whilst we have used stochastic gradient descent for ease of exposition in our main algorithm, in our implementation we follow the literature in using the momentum-based Adam optimizer instead. The updating algorithm follows the five steps below:

1. $\mathbf{m} \leftarrow \beta_1 \mathbf{m} - (1-\beta_1)\nabla_\theta J(\theta)$
2. $\mathbf{s} \leftarrow \beta_2 \mathbf{s} + (1-\beta_2)\nabla_\theta J(\theta) \otimes \nabla_\theta J(\theta)$
3. $\mathbf{m} \leftarrow \frac{\mathbf{m}}{1-\beta_1^t}$
4. $\mathbf{s} \leftarrow \frac{\mathbf{s}}{1-\beta_2^t}$
5. $\theta \leftarrow \theta + \eta \mathbf{m} \oslash \sqrt{\mathbf{s}+\varepsilon}$

[2]Note that this is only the optimal standard deviation for our tanh activation function. For other activation functions other formulas are available, such as He Initialization for ReLU activation functions.

First, let us clear up some notation: t denotes the iteration number (starting at 1). β_1 is the momentum decay hyperparameter, typically initialized to 0.9, whilst β_2 is the scaling decay hyperparameter, typically initialized to 0.999. ε is usually initialized to a tiny number like 10^{-8}. $\otimes$ denotes element-wise multiplication and $\oslash$ element-wise division. J denotes the cost function, so $\nabla_\theta J(\theta)$ denotes the gradient of the cost function with respect to the weights θ.

Next, let us quickly describe the basic principle of a momentum optimizer: Instead of relying solely on the current gradient like a gradient descent optimizer, it also takes into account previous gradients. It does this by subtracting the current gradient from a momentum vector $\mathbf{m}$ and then updating the weights by adding this momentum vector. In contrast to gradient descent, the gradient is thus not used as "speed" but instead as "acceleration".

Turning to Adam we can observe that this is the basic principle behind the algorithm. In step one, the momentum vector is updated by subtracting the gradient. Note that this is modified by using an average between the momentum and the gradient. As a result we care more about recent gradients than gradients further back. In the second step we update another vector $\boldsymbol{s}$. This vector accumulates the square of the gradient of the cost function, i.e. a measure of the steepness of the cost function.[3] Again, we use an average to put more weight on recent gradients. Steps 3 and 4 are more of a technical detail. As $\mathbf{m}$ and $\mathbf{s}$ are initialized to zero, these steps help to boost $\mathbf{m}$ and $\mathbf{s}$ at the beginning of training. Step 5 is the crucial updating step. We update θ by adding the momentum times the learning rate ν. However, the momentum vector is scaled down by the steepness of the accumulated gradients. This is included so that the optimizer does not just follow the steepest direction but instead takes into account all directions when updating. This significantly improves

[3]Note that methods using the full Hessian are not used in neural network learning as they are significantly slower.

convergence by stopping the algorithm from exploring "side valleys" and instead leading it towards its global minimum. ε enters in order to avoid division by zero.

We have tried a variety of optimizers, but Adam yields by far the best results. Regarding the learning rate, $\eta = 0.001$ works well for us, even though the particular learning rate should not be as important here, as it is automatically updated by the algorithm itself over time. The conventional values of β_1 and β_2 work well for us. There are some examples in the literature where changing these parameters substantially improved the results, but our trials have not shown similar effects in our model.

Backpropagation One fundamental breakthrough in the deep learning literature is the backpropagation algorithm by Rumelhart et al. (1986). This algorithm allows an efficient training of the neural network as well as an efficient computation of all derivatives of the neural network with respect to its inputs. It is thus absolutely essential to our approach. The fundamental idea is that in addition to the computation of the neural network in a "forward pass", we can compute all partial derivatives in a "backward pass" and then combine those through the chain rule. For details, we refer the interest reader to the excellent exposition in Goodfellow et al. (2016).

A.3 Simulation Details for the Ergodic Distribution

As outlined in the main text, it can be helpful to simulate the ergodic distribution to narrow down the hypercube that needs to be explored by the algorithm. Whilst closed-form expression for this distribution are occasionally available, in this section we demonstrate how to simulate an ergodic distribution in the case that closed-form expressions do not exist. In particular, we use the dynamic capital allocation model as an example.

Start the procedure by specifying a time step Δt. Start an epoch e with current neural network parameters Θ^e. Now start the simulation at I random starting points $\{\{z_{\ell,i,0}, k_{\ell,i,0}\}_{\ell=1}^{L}\}_{i=1}^{I}$. The simulation length is denoted by T. For each $i = 1 : I$, for each $t = 1 : T$, and for each $\ell = 1 : L$ recursively define:

$$z_{\ell,i,t} = z_{\ell,i,t-1} + (-\nu \log(z_{\ell,i,t-1}) + \frac{\sigma^2}{2}) z_{\ell,i,t-1} \Delta t + \sigma z_{\ell,i,t-1} \varepsilon_{\ell,i,t} \sqrt{\Delta t}, \quad \text{where } \varepsilon_{\ell,i,t} \sim \mathcal{N}(0,1)$$

$$k_{\ell,i,t} = k_{\ell,i,t-1} + [\tilde{i}(\{k_{\ell,i,t-1}, z_{\ell,i,t-1}\}_{\ell=1}^{L}; \Theta^e) + \tilde{nk}(\{k_{\ell,i,t-1}, z_{\ell,i,t-1}\}_{\ell=1}^{L}; \Theta^e) - \delta k_{\ell,i,t-1}] \Delta t$$

Define the location and starting point specific simulation vectors:

$$z_{\ell,i} \equiv \begin{pmatrix} z_{\ell,i,1} \\ z_{\ell,i,2} \\ \vdots \\ z_{\ell,i,T} \end{pmatrix}, \qquad k_{\ell,i} \equiv \begin{pmatrix} k_{\ell,i,1} \\ k_{\ell,i,2} \\ \vdots \\ k_{\ell,i,T} \end{pmatrix}$$

Drop the first B_1 periods as burn-in in order to allow the simulation to move away from the initial random starting points into the ergodic set:

$$z_{\ell,i} \equiv \begin{pmatrix} z_{\ell,i,B_1+1} \\ z_{\ell,i,B_1+2} \\ \vdots \\ z_{\ell,i,T} \end{pmatrix}, \qquad k_{\ell,i} \equiv \begin{pmatrix} k_{\ell,i,B_1+1} \\ k_{\ell,i,B_1+2} \\ \vdots \\ k_{\ell,i,T} \end{pmatrix}$$

Combine the different starting points into two big data matrices:

$$z_\ell \equiv \begin{pmatrix} z_{\ell,1} \\ z_{\ell,2} \\ \vdots \\ z_{\ell,I} \end{pmatrix}, \qquad k_\ell \equiv \begin{pmatrix} k_{\ell,1} \\ k_{\ell,2} \\ \vdots \\ k_{\ell,I} \end{pmatrix}$$

$$Z \equiv \begin{pmatrix} z_1 & z_2 & \cdots & z_L \end{pmatrix}, \qquad K \equiv \begin{pmatrix} k_1 & k_2 & \cdots & k_L \end{pmatrix}$$

Split the data into M mini-batches of size N, where $M = \frac{(T-B)\times I}{N}$:

$$\{Z^m, K^m\}_{m=1}^M \equiv \{Z[(m-1)\times N : m \times N, :], K[(m-1)\times N : m \times N, :]\}_{m=1}^M$$

For each $m \in \{1, ..., M\}$ define:

$$\{z_m^n, k_m^n\}_{n=1}^N = (Z^m[:], K^m[:])$$

Thus, we can update the weights for each $m \in \{1, ..., M\}$ (i.e. perform M learning steps) based on $\{z_m^n, k_m^n\}_{n=1}^N$. After these M learning steps, the neural network parameters will be updated from Θ^e to Θ^{e+1}, with which we can enter the next simulation episode.

Regarding the choice of initial starting points in the next epoch we can take the last points from the previous epoch, since they are likely closer to the ergodic set than random starting points. That is $\{\{z_{\ell,i,0}^{e+1}, k_{\ell,i,0}^{e+1}\}_{\ell=1}^L\}_{i=1}^I = \{\{z_{\ell,i,T}^e, k_{\ell,i,T}^e\}_{\ell=1}^L\}_{i=1}^I$.

Finally, note that we only employ this simulation algorithm after a certain number of epochs. Before that we sample uniformly at random from the entire state space hypercube to get a good approximation for the policy functions. This is necessary since the policy

functions are most likely very poor approximations in the beginning and simulation might thus actually move us away from the ergodic set instead of narrowing in on it.

A.4 Steady States

Deterministic Steady State The deterministic steady state is the steady state reached when eliminating uncertainty from the model, that is $\sigma = 0$. We can then solve for the DSS in closed form. First note that without the stochastic component all locations are completely identical. Thus there is no incentive for the social planner to redistribute any capital between locations. The problem is therefore equivalent to solvings 25 separate models. That is $x_\ell = 0$ for all ℓ and we will focus on one location from now on. First note that with $\sigma = 0$, the steady state for the productivity process is determined by $\mu(z) = 0$, which leads to two steady states, $\bar{z} = 0$ and $\bar{z} = 1$. We will concentrate on the stable steady state $\bar{z} = 1$. To find the steady state for the capital stock, we consider the Hamiltonian of the problem:

$$H(k, i, \lambda) = u(zk^\alpha - i - \kappa_i i^2) + \lambda(i - \delta k)$$

The necessary and sufficient conditions for the solution are given by:

$$u'(c)(1 + 2\kappa_i i) = \lambda$$

$$\dot{\lambda} = \rho\lambda - u'(c)\alpha z k^{\alpha-1} + \lambda\delta$$

$$\dot{k} = i - \delta k$$

In steady state we then have $\bar{z} = 1$ as well as $\dot{k} = \dot{\lambda} = 0$. Combining this with the equations above leads to the following nonlinear equation in $\bar{k}$:

$$\alpha \bar{z} \bar{k}^{\alpha-1} = (\rho + \delta)(1 + 2\kappa_i \delta \bar{k})$$

We can solve this equation with any nonlinear solution techninque such as bisection, which yields the DSS:

$$\bar{k}_{DSS} \approx 3.04$$

Stochastic Steady State The stochastic steady state is the steady state that is attained when uncertainty is present, i.e. $\sigma \neq 0$ and $\varepsilon_{\ell,t} = 0$ for all ℓ and all t. With $\varepsilon_{\ell,t} = 0$ and $\mu(z) = 0$, we can solve for the stochastic steady state for productivity in closed form: $z_{SSS} = e^{\frac{\sigma^2}{2\nu}} \approx 1.90$. In order to find the stochastic steady state for capital we utilize the policy functions for investment and net capital movement from our converged model and feed in a constant z at the level z_{SSS} to the productivity of all locations. We find that capital in each location settles down to the steady state level $\bar{k}_{SSS} \approx 3.75$.

Appendix B

Appendix for Chapter 2

B.1 Model Quantification

The model is estimated to match salient features of French firm-level data, as discussed in detail in Gaubert & Itskhoki (2020). Here we briefly summarize the main steps.

To estimate the model, we rely on French firm-level balance sheet data, which in particular reports sales at home and abroad, as well as the firm industry. This data is merged with international trade data from Comtrade, to get the aggregate imports and exports of France in each industry.

The estimation proceeds in two steps. In the first step, we calibrate Cobb-Douglas shares $\{\alpha_z\}$ to equal the sectoral expenditure shares in the French data. We also calibrate $w/w^* = 1.13$, which corresponds to the ratio of wages in France to the average wage of its trading partners weighted by trade values. Lastly, in our estimation, we find that the elasticity of substitution σ and the productivity parameter θ are only weakly separately identified. Therefore, we choose to fix σ at a conventional value in the trade literature ($\sigma = 5$; see Broda & Weinstein, 2006), and estimate the constrained model with five parameters $\Theta' = (\theta, \tau, F, \mu_T, \sigma_T)$.

Table B.1: Moments used in SMM estimation

	Moments		Data	Model	Loss (%)
1.	Log number of firms, mean	$\log \tilde{M}_z$	5.631	5.429	1.9
2.	— st. dev.		1.451	1.230	3.9
3.	Top-firm sales share, mean	$\tilde{s}_{z,1}$	0.197	0.205	3.0
4.	— st. dev.		0.178	0.148	4.5
5.	Top-3 sales share, mean	$\sum_{j=1}^{3} \tilde{s}_{z,j}$	0.356	0.343	2.0
6.	— st. dev.		0.241	0.176	12.2
7.	Imports/dom. sales, mean	$\tilde{\Lambda}_z$	0.365	0.354	1.5
8.	— st. dev.		0.204	0.266	15.2
9.	Exports/dom. sales, mean	$\tilde{\Lambda}_z^{*\prime}$	0.328	0.345	3.9
10.	— st. dev.		0.286	0.346	7.2
11.	Fraction of sectors with exports>dom. sales	$\mathbb{P}\{\tilde{X}_z > \tilde{D}_z\}$	0.185	0.095	39.7
Regression coefficients:†					
12.	export share on top-firm share	$\hat{b}_1^*$	0.215 (0.156)	0.240 (0.104)	2.2
13.	export share on top-3 share	$\hat{b}_3^*$	0.254 (0.108)	0.222 (0.090)	2.6
14.	import share on top-firm share	$\hat{b}_1$	−0.016 (0.097)	−0.011 (0.079)	0.1
15.	import share on top-3 share	$\hat{b}_3$	0.002 (0.074)	0.008 (0.069)	0.1

Note: Last column reports the contribution of the moment to the SMM loss function.
†Moments 12–15 are regression coefficients of $\tilde{\Lambda}_z^{*\prime}$ and $\tilde{\Lambda}_z$ on $\tilde{s}_{z,1}$ and $\sum_{j=1}^{3} \tilde{s}_{z,j}$ (pairwise), controlling in all cases for the size of the sector with the log domestic sectoral expenditure $\tilde{Y}_z$; OLS standard errors in brackets.

In the second step, we use simulated method of moments (SMM) to estimate the remaining parameters. We search for parameter values that minimize the distance between data moments and their model counterpart. We are particularly interested in the following salient feature of the data: (a) heterogeneity across sectors in top firm concentration, (b) heterogeneity across sectors in export stance and, importantly, (c) the extent to which the two are correlated, capturing granular forces at play in shaping sectoral outcomes.

To that end, we choose to target three types of moments, summarized in Table B.1. The first set of moments is informative about the prevalence of large firms in domestic

sectoral sales (a). Namely, we target the average and standard deviation across sectors of two measures of within-industry concentration (the relative sales shares of the largest and top-3 largest French firms within-industry relative to other French firms). We also target the average (log) number of French firms operating within sectors, as well as its standard deviation. This ensures that the model captures simultaneously the large number of firms operating in French sectors and the high concentration of sales. These moments particularly help to inform the estimation of the fixed costs F, as well as the productivity dispersion parameter θ.

The second set of moments is informative about the intensity of sectoral exports (b). Specifically, we match the average and standard deviation of foreign shares in the French market $\tilde{\Lambda}_z$, and the French export intensity $\tilde{\Lambda}_z^{*\prime}$. We also target the fraction of French sectors in which export sales exceed the overall domestic sales of French firms. These trade moments help inform the estimation of the size of the trade cost τ and the average productivity advantage of France μ_T.

Finally, the third set of moments is informative about the joint distribution of firm concentration and sectoral exports (c). We target four moments describing the correlation between French import and export shares and the sectoral sales concentration at home. Specifically, we target the regression coefficients of $\tilde{\Lambda}_z$ and $\tilde{\Lambda}_z^{*\prime}$ separately on $\tilde{s}_{z,1}$ and $\sum_{j=1}^{3} \tilde{s}_{z,j}$, controlling in all four regressions for the size of the sector (log total domestic expenditure, $\log \tilde{Y}_z$). These moments are instrumental for identifying the relative importance of fundamental and granular forces in shaping trade patterns.

The parameter values resulting from this SMM estimation are reported in Table 2.1. Table B.1 reports the model-based values of the 15 moments used in the estimation, and compares them with their empirical counterparts. The table also reports the percentage contribution of each moment to the SMM loss function. Overall, the model provides a good fit to the data for the 15 moments targeted in the estimation.

B.2 Granular Tariff

Consider the firm-specific tariff profile $\{\varsigma_{z,i}\}$ imposed by the home government on foreign firms i in sector z. Then the foreign firm's profit maximization results in the same prices and markups as if the firm has its costs increased to $c'_{z,i} = c_{z,i}/(1-\varsigma_{z,i})$, or equivalently its productivity draw reduced to $\varphi'_{z,i} = \varphi_{z,i}(1-\varsigma_{z,i})$. We denote the resulting market shares $\{s'_{z,i}\}$, and the resulting profits for foreign firms are $\Pi'_{z,i} = (1-\varsigma_{z,i})\alpha_z Y \frac{s'_{z,i}}{\varepsilon(s'_{z,i})} - wF$, where $\varepsilon(s)$ is as defined in (2.6).[1]

The expenditure on foreign goods in the home market is still given by $s'_{z,i}\alpha_z Y$, and the foreign share is still $\Lambda'_z = \sum_{i=1}^{K_z}(1-\iota_{z,i})s'_{z,i}$. However now, the home government collects $TR_z = \alpha_z Y \sum_{i=1}^{K_z}(1-\iota_{z,i})\varsigma_{z,i}s'_{z,i}$, while the rest $(\Lambda'_z\alpha_z Y - TR_z)$ are the revenues of foreign firms, which are split between production labor $\alpha_z Y \sum_{i=1}^{K_z}(1-\iota_{z,i})(1-\varsigma_{z,i})\frac{s'_{z,i}}{\mu(s'_{z,i})}$, fixed costs $wF\sum_{i=1}^{K_z}(1-\iota_{z,i})$, and profits $\sum_{i=1}^{K_z}(1-\iota_{z,i})\Pi'_{z,i}$, where $\mu(s) = \frac{\varepsilon(s)}{\varepsilon(s)-1}$.

Therefore, there are changes to the three general equilibrium conditions (2.8)–(2.10). In particular, (2.8) becomes:

$$Y = wL + \Pi + TR, \qquad \text{where} \qquad TR = Y\int_0^1 \alpha_z \left[\sum\nolimits_{i=1}^{K'_z}(1-\iota_{z,i})\varsigma_{z,i}s'_{z,i}\right]\mathrm{d}z,$$

and where the profits of home firms Π are still expressed as in (2.9). Foreign aggregate income is still $Y^* = w^*L^* + \Pi^{*\prime}$, where now aggregate foreign profits reflect import tariffs

[1] Note that a non-uniform tax creates a computational challenge for the entry game, as the effective condition for entry becomes $\alpha_z Y \frac{s'_{z,i}}{\varepsilon(s'_{z,i})} \geq \frac{wF}{1-\varsigma_{z,i}}$, and ranking firms on $c'_{z,i}$ (and hence $s'_{z,i}$ does not guarantee monotonicity of $\Pi'_{z,i}$. We assume, however, that for a small enough $\varsigma_{z,i}$ (as is the case in our simulation), the approximation $F/(1-\varsigma_{z,i}) \approx F$ is sufficiently accurate in the entry game. Indeed, recall that entry is a discrete zero-one decision, in which most entering firms are inframarginal, with $\Pi'_{z,i} \gg 0$ due to the Zipf's law.

imposed by the home government:

$$\Pi^{*\prime} = \int_0^1 \left[\sum_{i=1}^{K_z} (1-\iota_{z,i}) \left((1-\varsigma_{z,i}) \alpha_z Y \frac{s'_{z,i}}{\varepsilon(s'_{z,i})} - wF \right) + \sum_{i=1}^{K_z^*} \iota^*_{z,i} \left(\alpha_z Y^* \frac{s^*_{z,i}}{\varepsilon(s^*_{z,i})} - w^* F \right) \right] \mathrm{d}z.$$

Finally, the current account balance becomes:

$$\Lambda' Y - wFK_F - TR = Y^* \Lambda^* - w^* F^* K_H^*,$$

where $\Lambda' = \int_0^1 \alpha_z \big[\sum_{i=1}^{K_z} (1-\iota_{z,i}) s'_{z,i} \big] \mathrm{d}z$, $K_F = \int_0^1 \big[\sum_{i=1}^{K_z} (1-\iota_{z,i}) \big] \mathrm{d}z$ and similarly for Λ^* and K_H^*, as now the foreign income from exporting is reduced by TR.

Welfare approximation We now adapt the welfare approximation introduced in (2.20)–(2.21) in Section 2.3 to the case of the import tariff. In this case, $\hat{W}_z$ has an additional component, namely, government tariff revenue $\frac{\mathrm{d}TR_z}{\alpha_z Y} = \bar{\varsigma}_z \Lambda_z$ (equivalently under the granular tariff $\hat{\varsigma}_{z,1}$), levied on foreign exporters who in turn partially pass it through to home consumers. From the point of view of home consumers, a tariff is equivalent to an adverse productivity shock to foreign products, $\hat{\varphi}_{z,i} = -\varsigma_{z,i}$, and there is no direct effect on Home's producer surplus (only a second-round effect on their markups). As a result, using (2.20), the decline in consumer surplus is given by $\hat{P}_z = \frac{1-\bar{\varkappa}_z^F}{1-\bar{\varkappa}_z} \bar{\varsigma}_z \Lambda_z$, where $1-\bar{\varkappa}_z^F$ is the average pass-through rate of the tariff among the affected foreign firms. Thus, $\bar{\varkappa}_z^F$ is equal to $\frac{1}{\Lambda_z} \sum_{i=1}^{K_z} (1-\iota_{z,i}) s_{z,i} \varkappa_{z,i}$ under the uniform tariff and $\varkappa_{z,1}^F$ under the granular tariff. This erosion of consumer surplus is partially offset by an increase in producer surplus, as home firms increase their markups and prices by $\varkappa_{z,i} \hat{P}_z$ according to (2.19).

Therefore, the net effect of a tariff on home welfare is given by:

$$\hat{W}_z = \bar{\varsigma}_z \Lambda_z - \left[1 - (1-\Lambda_z) \frac{\bar{\varkappa}_z^H}{\bar{\mu}_z^H} \right] \hat{P}_z = \bar{\varsigma}_z \Lambda_z \left[\frac{\bar{\varkappa}_z^F - \bar{\varkappa}_z}{1-\bar{\varkappa}_z} + (1-\Lambda_z) \frac{\bar{\varkappa}_z^H}{\bar{\mu}_z^H} \frac{1-\bar{\varkappa}_z^F}{1-\bar{\varkappa}_z} \right],$$

where $\bar{\varkappa}_z^H$ and $\bar{\mu}_z^H$ are the average strategic complementarity elasticity and the average markup among domestic firms. We see that $\bar{\varkappa}_z^F > \bar{\varkappa}_z$ is sufficient (but not necessary) for a positive welfare effect of an import tariff on home welfare.[2] This condition is equivalent to the pass-through rate being lower among foreign exporters relative to an average firm serving the home market, $1 - \bar{\varkappa}_z^F < 1 - \bar{\varkappa}_z$. We generally expect this condition to hold due to the Melitz selection force of larger firms into the foreign market and pass-through being lower among the firms with greater market shares. Adopting a granular tariff maximizes $\bar{\varkappa}_z^F$ and minimizes the tariff pass-through $1 - \bar{\varkappa}_z^F$, thus delivering larger welfare gains to Home, especially in sectors where the largest foreign firm commands a substantial market share in the home market.

This discussion requires two remarks. First, it focuses exclusively on partial equilibrium effects, thus omitting the standard general equilibrium terms-of-trade effect, which is usually the focus of the optimal tariff argument. The welfare effect of a tariff in our partial equilibrium approximation would be altogether absent if the model were to feature constant markups and complete pass-through, $\varkappa_{z,i} = 0$ for every firm. Our numerical analysis above combines both partial and general equilibrium forces, thus delivering a complete quantitative evaluation of the welfare consequences of different tariffs. Second, we note that the welfare consequences of a tariff are always negative if the impact on foreign countries, in particular on foreign producers, is taken into account. Indeed, the loss in foreign producer surplus is greater than the welfare benefit at home due to the increased Harberger's triangle and deadweight loss, independently of the type of tariff adopted. The loss in foreign producer surplus is given by $\bar{\varsigma}_z \Lambda_z \left[\frac{1}{\bar{\varepsilon}_z^F} + \frac{\bar{\varkappa}_z^F}{\bar{\mu}_z^F} \left(1 - \frac{1-\bar{\varkappa}_z^F}{1-\bar{\varkappa}_z} \Lambda_z \right) \right]$.

[2]For example, when $\bar{\varkappa}_z^H = \bar{\varkappa}_z^F = \bar{\varkappa}_z > 0$ (e.g., if $\varkappa_{z,i} > 0$ and constant across all firms), $\hat{W}_z > 0$ provided that $\Lambda_z \in (0, 1)$. Indeed, tax pass-through is incomplete iff the tax is levied on some but not all firms in the industry *and* firms feature incomplete cost pass-through. In contrast, if $\bar{\varkappa}_z^F = 0$ and $\bar{\varkappa}_z^H > 0$, the partial equilibrium effect of a tariff on home welfare is negative, as consumers bear out the full cost of the tariff on foreign firms and in addition home firms raise their prices in response to increased prices of foreign competitors.

Appendix C

Appendix for Chapter 3

Table C.1: Linear Regressions Results Inflows and Outflows

	$\tilde{y}_{jt}^{in}$			$\tilde{y}_{jt}^{out}$		
	(1)	(2)	(3)	(4)	(5)	(6)
u_{jt}	35.7	35.7	-40.3*	-10.5	-10.5	-65.1***
	(1.53)	(0.66)	(-1.81)	(-0.70)	(-0.29)	(-5.05)
Constant	1571.6***	1571.6**	1506.9***	1520.1***	1520.1***	2225.2***
	(6.90)	(2.43)	(3.49)	(10.42)	(3.32)	(8.90)
Year FE			✓			✓
Country FE			✓			✓
Clustered SE		✓			✓	
Adjusted R^2	0.0025	0.0025	0.6458	-0.0009	-0.0009	0.7093
Countries	27	27	27	27	27	27
Observations	540	540	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Columns (1)-(3) show regressions with the inflow rate $\tilde{y}_{jt}^{in}$ as the dependent variable while in columns (4)-(6) this is the outflow rate $\tilde{y}_{jt}^{out}$. Columns (2) and (4) cluster standard errors at the country level and columns (3) and (6) include year and country fixed effects.

Table C.2: Descriptive Statistics Inflows

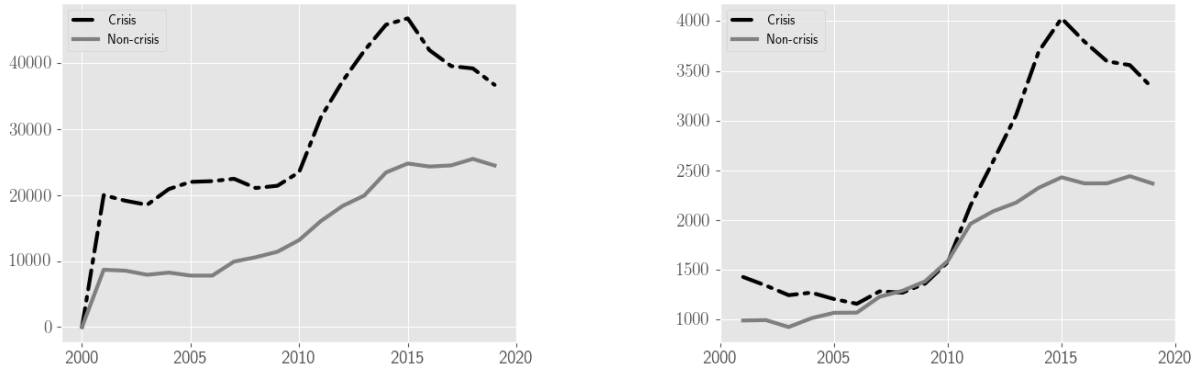
	y_{jt}^{in}			$\tilde{y}_{jt}^{in}$		
Country	Mean	Min	Max	Mean	Min	Max
Austria	17169	13456	20804	2039.1	1647.8	2381.2
Belgium	5025	4115	6099	461.4	390.1	544.1
Bulgaria	42304	7655	83579	5835.4	1007.1	11929.6
Croatia	23685	8543	57476	5632.0	1980.8	13775.1
Cyprus	571	257	1118	693.2	353.0	1297.1
Czechia	10315	7190	13274	990.4	686.4	1258.7
Denmark	3174	2563	3749	571.4	471.4	667.7
Estonia	1000	621	1515	747.0	461.1	1141.3
Finland	2322	1354	3014	433.7	257.1	582.3
France	19653	9873	23307	304.4	156.7	353.6
Greece	20820	8908	35811	1910.4	806.3	3242.3
Hungary	33539	14965	58993	3377.3	1477.4	5963.1
Ireland	2519	1551	3346	566.6	372.8	716.1
Italy	40865	20130	74105	687.1	346.2	1220.2
Latvia	5126	1757	10177	2502.1	798.5	4941.0
Lithuania	6910	3384	11854	2285.7	967.0	4231.2
Luxembourg	2953	1439	4073	5650.0	3298.2	7076.6
Malta	221	84	528	507.4	209.3	1047.5
Netherlands	13385	11007	14393	806.3	691.2	875.2
Poland	146845	94105	197908	3856.1	2459.7	5206.5
Portugal	8941	5608	14494	857.7	533.9	1386.0
Romania	102136	21145	238824	5121.2	955.4	12263.8
Slovakia	11675	8558	15435	2161.4	1588.8	2848.5
Slovenia	3378	1157	7003	1650.5	576.5	3393.7
Spain	23985	14004	44119	526.7	320.3	946.4
Sweden	3780	3124	4354	400.9	338.8	447.1
UK	16955	12611	22526	269.6	208.8	337.0
All origins	36461	19544	58627	1367.9	665.3	2352.1

Notes: 20 yearly observations between 2000 and 2019. Total migration y_{jt}^{in} in total number of migrants. Migration rate $\tilde{y}_{jt}^{in}$ in migrants per one million inhabitants.

Table C.3: Descriptive Statistics Outflows

	y_{jt}^{out}			$\tilde{y}_{jt}^{out}$		
Country	Mean	Min	Max	Mean	Min	Max
Austria	19513	14875	24049	2318.6	1849.6	2890.0
Belgium	4746	4191	5376	437.2	377.3	475.4
Bulgaria	27134	6747	63321	3744.2	825.8	9077.3
Croatia	15572	10283	27706	3679.8	2384.2	6812.0
Cyprus	389	226	832	478.4	306.0	943.4
Czechia	8106	5889	9691	779.2	561.1	950.4
Denmark	3408	2694	4549	613.3	497.1	828.1
Estonia	721	518	936	538.8	384.6	712.0
Finland	2357	2025	2800	441.2	375.8	540.9
France	19021	16703	22158	295.5	255.2	343.5
Greece	17176	11259	21540	1573.4	1013.9	2009.0
Hungary	25837	14973	41236	2600.9	1466.4	4201.7
Ireland	2434	1872	3059	554.6	408.7	803.9
Italy	32869	23164	45136	557.1	390.1	755.7
Latvia	3405	1290	6009	1667.8	551.9	3139.8
Lithuania	4310	1699	9305	1434.3	485.5	3330.2
Luxembourg	2205	1253	3146	4232.6	2837.9	5074.2
Malta	214	89	428	492.3	226.4	883.1
Netherlands	10785	8616	14114	648.1	531.0	813.7
Poland	112094	71409	138680	2943.6	1866.5	3648.3
Portugal	8513	6090	13326	816.7	579.2	1295.1
Romania	71183	17160	189932	3569.0	764.6	9804.6
Slovakia	9429	7328	11286	1745.9	1359.2	2082.8
Slovenia	2551	1432	4003	1250.9	713.5	1941.3
Spain	19243	16007	24813	426.4	342.4	530.2
Sweden	4249	3568	4979	451.1	395.1	540.0
UK	17520	15506	20299	280.1	243.4	328.4
All origins	29332	19382	42566	1069.2	627.8	1744.9

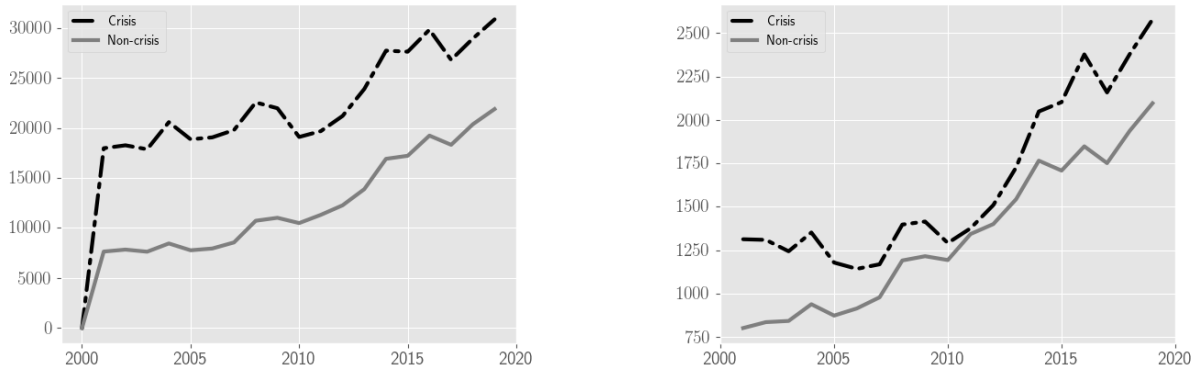
Notes: 20 yearly observations between 2000 and 2019. Total migration y_{jt}^{out} in total number of migrants. Migration rate $\tilde{y}_{jt}^{out}$ in migrants per one million inhabitants.



(a) Total Migration y_{jt}^{in} (b) Migration Rate $\tilde{y}_{jt}^{in}$

Figure C.1: Time Series Graphs Inflows

Note: Crisis countries are those for which $C_j = 1$ and non-crisis countries those with $C_j = 0$. All graphs show the yearly averages for all countries in each group.



(a) Total Migration y_{jt}^{out} (b) Migration Rate $\tilde{y}_{jt}^{out}$

Figure C.2: Time Series Graphs Outflows

Note: Crisis countries are those for which $C_j = 1$ and non-crisis countries those with $C_j = 0$. All graphs show the yearly averages for all countries in each group.

Table C.4: Comparison of Crisis and Non-Crisis Groups - Inflows and Outflows

	Average y_{jt}^{in}	Average $\tilde{y}_{jt}^{in}$	Average y_{jt}^{out}	Average $\tilde{y}_{jt}^{out}$
Crisis	29599.4	2216.1	22491.8	1615.5
Non-Crisis	15229.5	1654.2	12349.3	1300.0

Notes: All columns defined as average over both countries and time. Columns 1 and 2 measure inflows y_{jt}^{in} and $\tilde{y}_{jt}^{in}$, columns 3 and 4 outflows y_{jt}^{out}, $\tilde{y}_{jt}^{out}$ respectively.

Table C.5: Parametric Event Studies Inflows and Outflows with Clustered S.E.

	y_{jt}^{in}	$\tilde{y}_{jt}^{in}$	y_{jt}^{out}	$\tilde{y}_{jt}^{out}$
	(1)	(2)	(3)	(4)
Panel A. Parametric Event Study 1				
Post event	5253.2	726.1	-1540.5	-39.2
	(0.49)	(0.77)	(-0.22)	(-0.07)
Constant	10431.0**	1436.8***	14259.1***	1865.2***
	(2.09)	(5.01)	(4.09)	(9.77)
Panel B. Parametric Event Study 2				
Post event	-506.1	-91.2	-3641.0	-118.5
	(-0.17)	(-0.25)	(-0.98)	(-0.62)
Post event × years elapsed	4471.5	543.0	866.7	85.8
	(1.30)	(1.61)	(0.58)	(0.91)
Post event × (years elapsed)^2	-516.6	-46.4*	-93.8	-3.51
	(-1.69)	(-2.03)	(-1.07)	(-0.62)
Pre event × years elapsed	65.4	-55.4	158.7	-37.6
	(0.10)	(-1.22)	(0.23)	(-1.25)
Constant	10550.9*	1335.4***	14550.0***	1796.2***
	(2.02)	(4.15)	(4.06)	(8.59)
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓
Countries	27	27	27	27
Observations	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. Column (1) displays results for total inflows y_{jt}^{in} and column (2) for the inflow rate $\tilde{y}_{jt}^{in}$. Columns (3) and (4) do the same but for total outflows y_{jt}^{out} and the outflow rate $\tilde{y}_{jt}^{out}$. In all columns standard errors are clustered at the country level.

Table C.6: Nonparametric Event Studies Net Migration

	y_{jt}^{net}		$\tilde{y}_{jt}^{net}$	
	(1)	(2)	(3)	(4)
T_{-9}	1775.8	1775.8	180.9	180.9
	(0.40)	(0.75)	(0.44)	(1.01)
T_{-8}	1932.5	1932.5	147.1	147.1
	(0.44)	(0.76)	(0.36)	(0.98)
T_{-7}	1124.9	1124.9	94.8	94.8
	(0.25)	(0.47)	(0.23)	(0.70)
T_{-6}	1313.7	1313.7	140.8	140.8
	(0.30)	(0.54)	(0.34)	(0.93)
T_{-5}	1487.1	1487.1	65.7	65.7
	(0.34)	(0.40)	(0.16)	(0.40)
T_{-4}	4059.3	4059.3	54.0	54.0
	(0.92)	(0.76)	(0.13)	(0.27)
T_{-3}	4158.9	4158.9	82.2	82.2
	(0.94)	(0.82)	(0.20)	(0.42)
T_{-2}	2266.3	2266.3	85.3	85.3
	(0.51)	(0.72)	(0.21)	(0.58)
T_{-1}	-385.3	-385.3	-4.21	-4.21
	(-0.09)	(-0.82)	(-0.01)	(-0.04)
T_1	2639.6	2639.6	110.3	110.3
	(0.60)	(1.18)	(0.27)	(0.98)
T_2	8301.6*	8301.6	372.1	372.1
	(1.88)	(1.35)	(0.90)	(1.44)
T_3	10928.6**	10928.6	617.2	617.2*
	(2.47)	(1.60)	(1.50)	(1.96)
T_4	12889.3***	12889.3*	917.6**	917.6**
	(2.92)	(1.82)	(2.23)	(2.50)
T_5	12537.5***	12537.5*	1308.6***	1308.6*
	(2.84)	(1.74)	(3.18)	(1.96)
T_6	12554.2***	12554.2	1424.2***	1424.2
	(2.84)	(1.55)	(3.46)	(1.59)
T_7	8004.9*	8004.9	1117.8***	1117.8
	(1.81)	(1.63)	(2.71)	(1.46)
T_8	7512.0*	7512.0	1040.2**	1040.2
	(1.70)	(1.35)	(2.52)	(1.43)
T_9	6098.8	6098.8	895.9**	895.9
	(1.38)	(1.28)	(2.17)	(1.37)
T_{10}	3162.7	3162.7	588.8	588.8
	(0.72)	(1.01)	(1.43)	(1.38)
Constant	-3446.5	-3446.5**	-424.4*	-424.4***
	(-1.31)	(-2.20)	(-1.73)	(-3.61)
Adjusted R^2	0.5929	0.5929	0.4921	0.4921
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Clustered SE		✓		✓
Countries	27	27	27	27
Observations	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (3) show the results of the nonparametric event studies for total net migration y_{jt}^{net} and the net migration rate $\tilde{y}_{jt}^{net}$ respectively. Columns (2) and (4) do the same but with standard errors clustered at the country level.

Table C.7: Nonparametric Event Studies Inflows

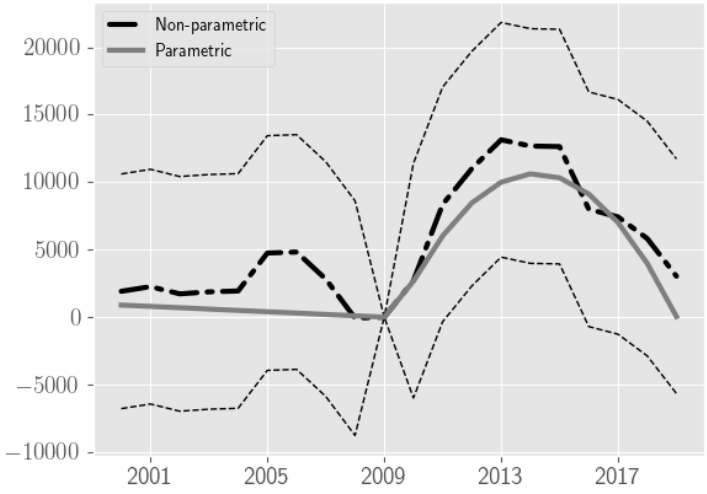
	y_{jt}^{in}		$\tilde{y}_{jt}^{in}$	
	(1)	(2)	(3)	(4)
T_{-9}	655.8	655.8	407.3	407.3
	(0.06)	(0.15)	(0.52)	(1.06)
T_{-8}	1313.7	1313.7	458.6	458.6
	(0.13)	(0.35)	(0.58)	(1.29)
T_{-7}	599.7	599.7	368.7	368.7
	(0.06)	(0.17)	(0.47)	(1.14)
T_{-6}	602.5	602.5	342.0	342.0
	(0.06)	(0.19)	(0.43)	(1.13)
T_{-5}	2663.9	2663.9	278.9	278.9
	(0.26)	(0.81)	(0.35)	(0.92)
T_{-4}	4203.1	4203.1	159.5	159.5
	(0.41)	(0.89)	(0.20)	(0.50)
T_{-3}	4319.6	4319.6	110.4	110.4
	(0.42)	(0.85)	(0.14)	(0.33)
T_{-2}	2549.8	2549.8	75.9	75.9
	(0.25)	(0.77)	(0.10)	(0.42)
T_{-1}	492.6	492.6	2.78	2.78
	(0.05)	(0.41)	(0.00)	(0.02)
T_1	298.2	298.2	8.76	8.76
	(0.03)	(0.19)	(0.01)	(0.05)
T_2	5721.4	5721.4	205.8	205.8
	(0.56)	(1.07)	(0.26)	(0.56)
T_3	8911.9	8911.9	527.1	527.1
	(0.88)	(1.25)	(0.67)	(1.09)
T_4	11953.3	11953.3	900.9	900.9
	(1.17)	(1.39)	(1.14)	(1.65)
T_5	12401.1	12401.1	1392.1*	1392.1
	(1.22)	(1.06)	(1.76)	(1.43)
T_6	12008.3	12008.3	1618.6**	1618.6
	(1.18)	(0.92)	(2.05)	(1.29)
T_7	7643.2	7643.2	1447.2*	1447.2
	(0.75)	(0.63)	(1.83)	(1.15)
T_8	5054.2	5054.2	1247.5	1247.5
	(0.50)	(0.42)	(1.58)	(1.03)
T_9	3715.2	3715.2	1137.6	1137.6
	(0.36)	(0.28)	(1.44)	(0.90)
T_{10}	2226.4	2226.4	979.9	979.9
	(0.22)	(0.18)	(1.24)	(0.86)
Constant	10872.7*	10872.7**	1360.7***	1360.7***
	(1.79)	(2.14)	(2.89)	(4.33)
Adjusted R^2	0.7584	0.7584	0.6432	0.6432
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Clustered SE		✓		✓
Countries	27	27	27	27
Observations	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (3) show the results of the nonparametric event studies for total inflows y_{jt}^{in} and the inflow rate $\tilde{y}_{jt}^{in}$ respectively. Columns (2) and (4) do the same but with standard errors clustered at the country level.

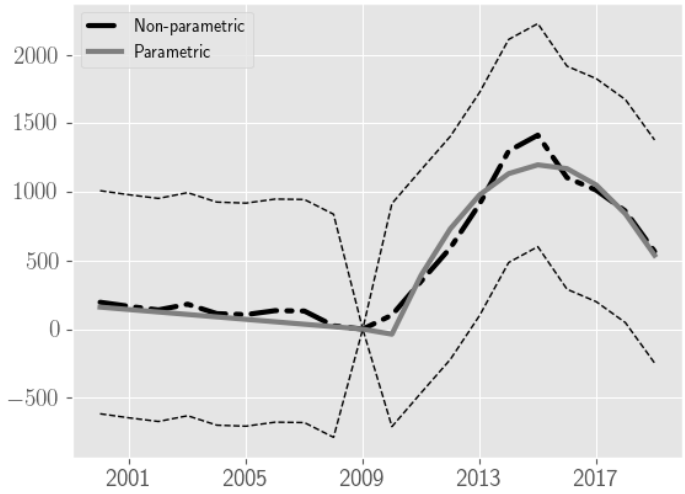
Table C.8: Nonparametric Event Studies Outflows

	y_{jt}^{out}		$\tilde{y}_{jt}^{out}$	
	(1)	(2)	(3)	(4)
T_{-9}	-1120.1	-1120.1	226.4	226.4
	(-0.16)	(-0.21)	(0.48)	(0.83)
T_{-8}	-618.8	-618.8	311.4	311.4
	(-0.09)	(-0.13)	(0.66)	(1.18)
T_{-7}	-525.2	-525.2	273.8	273.8
	(-0.07)	(-0.11)	(0.58)	(1.10)
T_{-6}	-711.2	-711.2	201.3	201.3
	(-0.10)	(-0.17)	(0.42)	(0.97)
T_{-5}	1176.8	1176.8	213.2	213.2
	(0.17)	(0.44)	(0.45)	(1.13)
T_{-4}	143.8	143.8	105.5	105.5
	(0.02)	(0.06)	(0.22)	(0.61)
T_{-3}	160.7	160.7	28.2	28.2
	(0.02)	(0.08)	(0.06)	(0.15)
T_{-2}	283.5	283.5	-9.48	-9.48
	(0.04)	(0.17)	(-0.02)	(-0.06)
T_{-1}	877.9	877.9	6.98	6.98
	(0.13)	(0.77)	(0.01)	(0.09)
T_1	-2341.3	-2341.3	-101.5	-101.5
	(-0.33)	(-1.22)	(-0.21)	(-0.87)
T_2	-2580.2	-2580.2	-166.3	-166.3
	(-0.37)	(-1.15)	(-0.35)	(-0.86)
T_3	-2016.7	-2016.7	-90.1	-90.1
	(-0.29)	(-0.78)	(-0.19)	(-0.35)
T_4	-936.0	-936.0	-16.7	-16.7
	(-0.13)	(-0.27)	(-0.04)	(-0.05)
T_5	-136.4	-136.4	83.6	83.6
	(-0.02)	(-0.03)	(0.18)	(0.20)
T_6	-545.9	-545.9	194.4	194.4
	(-0.08)	(-0.09)	(0.41)	(0.41)
T_7	-361.7	-361.7	329.4	329.4
	(-0.05)	(-0.04)	(0.69)	(0.51)
T_8	-2457.8	-2457.8	207.2	207.2
	(-0.35)	(-0.32)	(0.44)	(0.35)
T_9	-2380.0	-2380.0	241.8	241.8
	(-0.34)	(-0.25)	(0.51)	(0.33)
T_{10}	-1981.0	-1981.0	283.2	283.2
	(-0.28)	(-0.19)	(0.60)	(0.34)
Constant	14701.9***	14701.9***	1828.3***	1828.3***
	(3.51)	(4.17)	(6.46)	(8.93)
Adjusted R^2	0.7788	0.7788	0.6852	0.6852
Year FE	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Clustered SE		✓		✓
Countries	27	27	27	27
Observations	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Columns (1) and (3) show the results of the nonparametric event studies for total outflows y_{jt}^{out} and the outflow rate $\tilde{y}_{jt}^{out}$ respectively. Columns (2) and (4) do the same but with standard errors clustered at the country level.



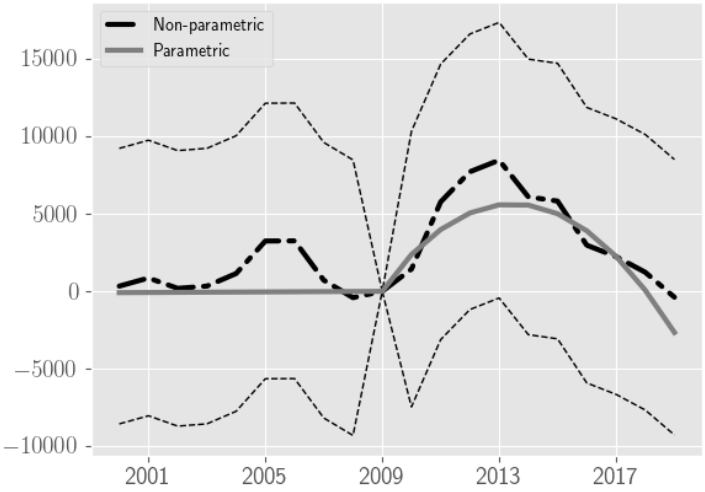
(a) Total Migration y_{jt}^{net}



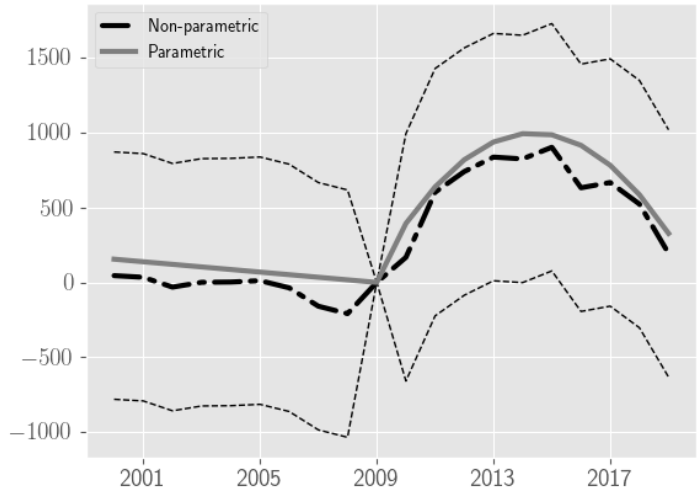
(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure C.3: Robustness 1 Event Studies

Notes: Panel (a) displays nonparametric and parametric event studies for total net flows y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification. Δ_j^u is defined as $\Delta_j^u = u_{j,2013} - u_{j,2009}$.



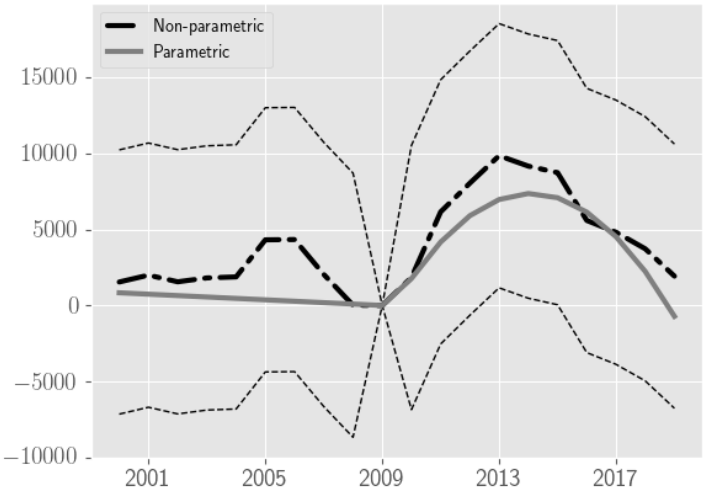
(a) Total Migration y_{jt}^{net}



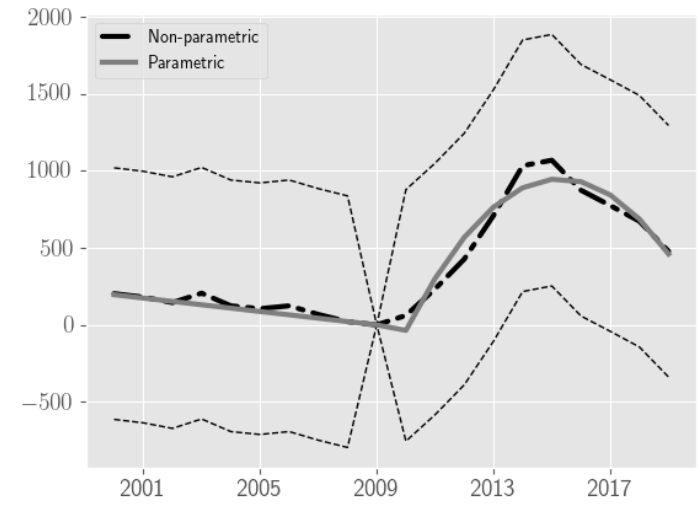
(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure C.4: Robustness 2 Event Studies

Notes: Panel (a) displays nonparametric and parametric event studies for total net flows y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification. Δ_j^u is defined as $\Delta_j^u = u_{j,2012} - u_{j,2008}$ and the threshold $\bar{\Delta} = 2.9$.



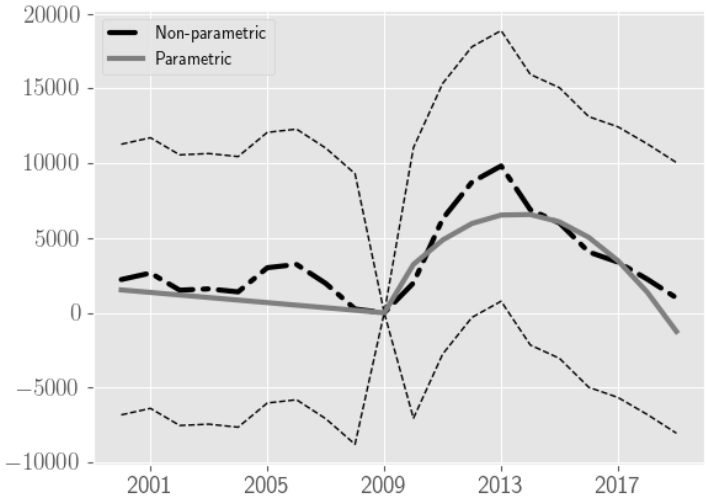
(a) Total Migration y_{jt}^{net}



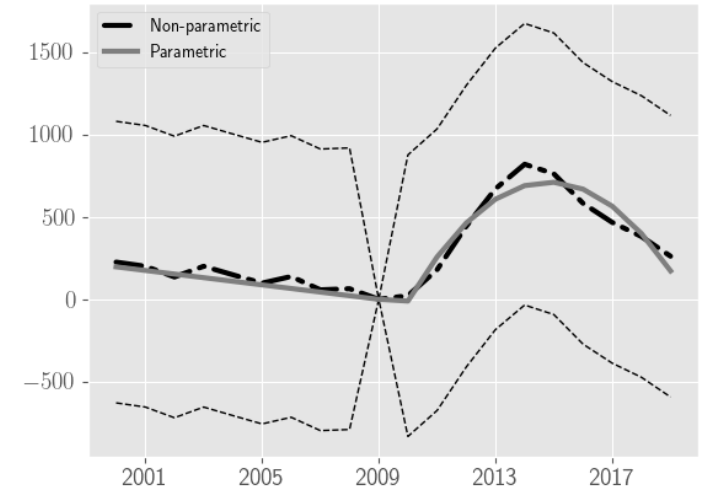
(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure C.5: Robustness 3 Event Studies

Notes: Panel (a) displays nonparametric and parametric event studies for total net flows y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification. The threshold $\bar{\Delta}$ is $\bar{\Delta} = 1$.



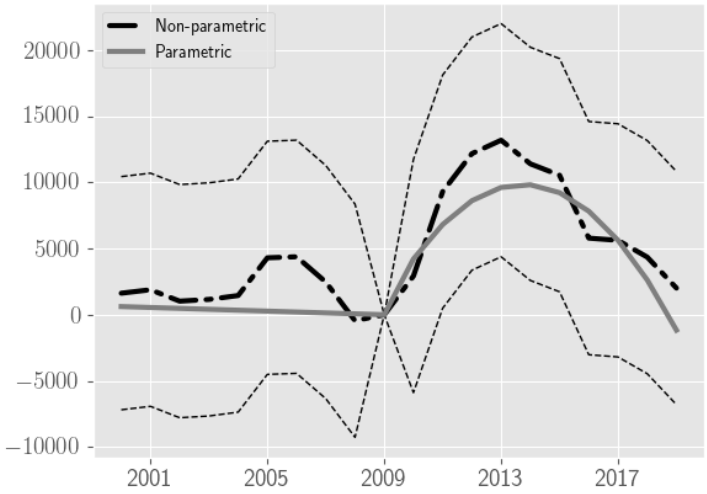
(a) Total Migration y_{jt}^{net}



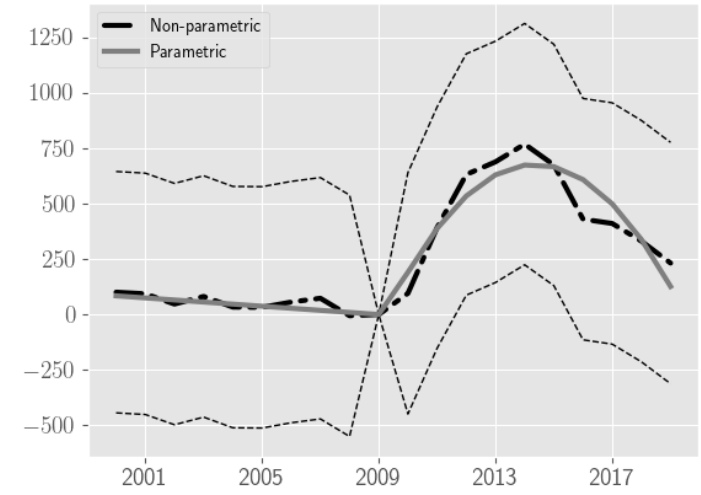
(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure C.6: Robustness 4 Event Studies

Notes: Panel (a) displays nonparametric and parametric event studies for total net flows y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification. The threshold $\bar{\Delta}$ is $\bar{\Delta} = 0$.



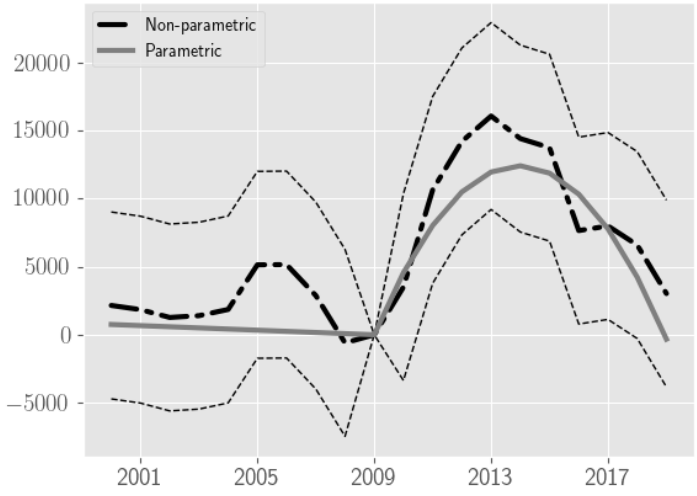
(a) Total Migration y_{jt}^{net}



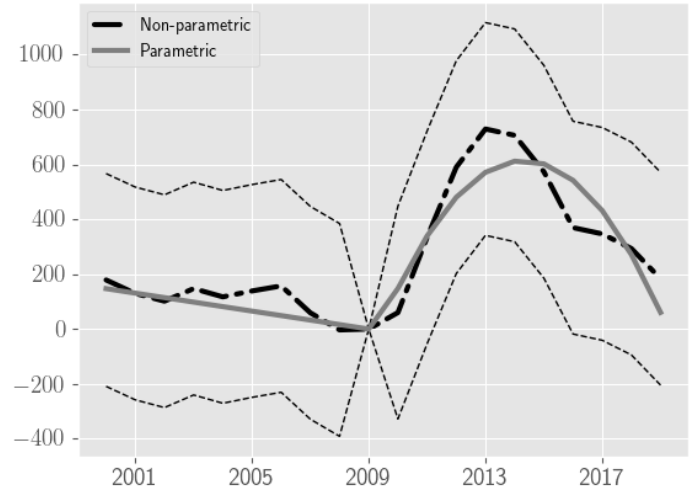
(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure C.7: Robustness 5 Event Studies

Notes: Panel (a) displays nonparametric and parametric event studies for total net flows y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification. Croatia is excluded from the data set.



(a) Total Migration y_{jt}^{net}



(b) Migration Rate $\tilde{y}_{jt}^{net}$

Figure C.8: Robustness 6 Event Studies

Notes: Panel (a) displays nonparametric and parametric event studies for total net flows y_{jt}^{net} and panel (b) for the net migration rate $\tilde{y}_{jt}^{net}$. The black dash-dotted lines present the nonparametric estimates, the thin dashed lines their 95% confidence intervals and the grey solid line the second parametric specification. Croatia, Romania and Bulgaria are excluded from the data set.

Table C.9: Robustness Event Studies with Total Migration y_{jt}^{net}

	y_{jt}^{net}					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Parametric Event Study 1						
Post event	6279.3***	3168.0**	4043.9***	3258.6**	5944.9***	7676.9***
	(4.50)	(2.23)	(2.91)	(2.26)	(4.20)	(6.79)
Constant	-3529.1	-3869.5	-3834.6	-3782.3	-3507.0	-2010.6
	(-1.51)	(-1.61)	(-1.61)	(-1.57)	(-1.51)	(-1.11)
Panel B. Parametric Event Study 2						
Post event	2670.8	2373.9	1771.7	3239.0	4215.2	4568.3*
	(0.83)	(0.72)	(0.55)	(0.97)	(1.29)	(1.78)
Post event × years elapsed	3810.0***	1882.8	2743.9**	1900.4	2999.1**	3969.6***
	(3.03)	(1.46)	(2.18)	(1.45)	(2.35)	(3.96)
Post event × (years elapsed)^2	-455.8***	-271.3**	-335.6**	-266.5*	-399.6***	-501.5***
	(-3.38)	(-1.97)	(-2.49)	(-1.90)	(-2.93)	(-4.68)
Pre event × years elapsed	-99.5	9.61	-92.8	-170.3	-68.4	-84.2
	(-0.29)	(0.03)	(-0.27)	(-0.48)	(-0.20)	(-0.31)
Constant	-3711.5	-3843.9	-4035.8	-4264.8	-3625.4	-2152.6
	(-1.54)	(-1.50)	(-1.63)	(-1.64)	(-1.52)	(-1.18)
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Countries	27	27	27	27	26	24
Observations	540	540	540	540	520	480

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. In all columns the dependent variable is total net migration y_{jt}^{net}. The first two columns vary the definition of Δ_j^u with column (1) $\Delta_j^u = u_{j,2013} - u_{j,2009}$ and in column (2) $\Delta_j^u = u_{j,2012} - u_{j,2009}$. The next columns vary the threshold $\bar{\Delta}$ where in column (3) $\bar{\Delta} = 1$ and in column (4) $\bar{\Delta} = 0$. The final two columns vary the countries included, with column (5) excluding Croatia and column (6) excluding Croatia, Romania and Bulgaria.

Table C.10: Robustness Event Studies with Migration Rate $\tilde{y}_{jt}^{net}$ and Clustered S.E.

	$\tilde{y}_{jt}^{net}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Parametric Event Study 1						
Post event	701.9	643.8*	515.3	332.9	415.0	315.7
	(1.51)	(1.71)	(1.22)	(0.87)	(1.29)	(1.66)
Constant	-395.9***	-348.2***	-414.9***	-434.1***	-405.4***	-329.4***
	(-3.95)	(-2.87)	(-3.83)	(-3.33)	(-4.06)	(-4.25)
Panel B. Parametric Event Study 2						
Post event	-36.9	394.3	-35.4	-10.7	190.8	147.9
	(-0.14)	(1.64)	(-0.14)	(-0.04)	(0.90)	(0.66)
Post event × years elapsed	475.7*	276.0	372.1	299.8	223.8*	216.4***
	(1.72)	(1.26)	(1.50)	(1.39)	(1.97)	(2.90)
Post event × (years elapsed)^2	-45.7*	-31.5	-35.2	-31.0	-25.7**	-25.1***
	(-2.01)	(-1.69)	(-1.68)	(-1.69)	(-2.24)	(-2.86)
Pre event × years elapsed	-17.8	-17.3	-21.8	-22.0	-9.38	-16.4
	(-0.84)	(-0.70)	(-1.01)	(-0.83)	(-0.46)	(-0.88)
Constant	-428.7***	-394.2**	-462.0***	-496.3***	-421.6***	-357.1***
	(-3.59)	(-2.45)	(-3.44)	(-2.82)	(-3.52)	(-3.55)
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓
Countries	27	27	27	27	26	24
Observations	540	540	540	540	520	480

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. In all columns the dependent variable is the net migration rate $\tilde{y}_{jt}^{net}$. The first two columns vary the definition of Δ_j^u with column (1) $\Delta_j^u = u_{j,2013} - u_{j,2009}$ and in column (2) $\Delta_j^u = u_{j,2012} - u_{j,2009}$. The next columns vary the threshold $\bar{\Delta}$ where in column (3) $\bar{\Delta} = 1$ and in column (4) $\bar{\Delta} = 0$. The final two columns vary the countries included, with column (5) excluding Croatia and column (6) excluding Croatia, Romania and Bulgaria.

Table C.11: Robustness Event Studies with Total Migration y_{jt}^{net} and Clustered S.E.

	y_{jt}^{net}					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Parametric Event Study 1						
Post event	6279.3	3168.0	4043.9	3258.6	5944.9	7676.9**
	(1.43)	(0.64)	(0.90)	(0.62)	(1.32)	(2.13)
Constant	-3529.1**	-3869.5*	-3834.6**	-3782.3	-3507.0**	-2010.6***
	(-2.28)	(-1.79)	(-2.20)	(-1.60)	(-2.26)	(-3.67)
Panel B. Parametric Event Study 2						
Post event	2670.8	2373.9	1771.7	3239.0*	4215.2	4568.3
	(1.02)	(1.16)	(0.76)	(1.71)	(1.70)	(1.68)
Post event × years elapsed	3810.0*	1882.8	2743.9	1900.4	2999.1	3969.6**
	(1.81)	(0.85)	(1.31)	(0.82)	(1.44)	(2.14)
Post event × (years elapsed)^2	-455.8*	-271.3	-335.6	-266.5	-399.6	-501.5*
	(-1.91)	(-1.17)	(-1.45)	(-1.11)	(-1.59)	(-2.01)
Pre event × years elapsed	-99.5	9.61	-92.8	-170.3	-68.4	-84.2
	(-0.54)	(0.06)	(-0.54)	(-1.01)	(-0.35)	(-0.47)
Constant	-3711.5**	-3843.9	-4035.8**	-4264.8	-3625.4**	-2152.6***
	(-2.15)	(-1.56)	(-2.06)	(-1.58)	(-2.10)	(-2.97)
Year FE	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓
Clustered SE	✓	✓	✓	✓	✓	✓
Countries	27	27	27	27	26	24
Observations	540	540	540	540	520	480

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A presents results for the parametric event study 1 as in Equation 3.3 and panel B for the parametric event study 2 as in Equation 3.4. In all columns the dependent variable is the net migration rate y_{jt}^{net}. The first two columns vary the definition of Δ_j^u with column (1) $\Delta_j^u = u_{j,2013} - u_{j,2009}$ and in column (2) $\Delta_j^u = u_{j,2012} - u_{j,2009}$. The next columns vary the threshold $\bar{\Delta}$ where in column (3) $\bar{\Delta} = 1$ and in column (4) $\bar{\Delta} = 0$. The final two columns vary the countries included, with column (5) excluding Croatia and column (6) excluding Croatia, Romania and Bulgaria.

Table C.12: Nonparametric Event Studies Robustness 1 and 2

	y_{jt}^{net}		$\tilde{y}_{jt}^{net}$		y_{jt}^{net}		$\tilde{y}_{jt}^{net}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T_{-9}	1913.1	1913.1	196.1	196.1	334.0	334.0	45.4	45.4
	(0.43)	(0.81)	(0.47)	(1.10)	(0.07)	(0.18)	(0.11)	(0.26)
T_{-8}	2257.8	2257.8	166.3	166.3	859.1	859.1	34.4	34.4
	(0.51)	(0.89)	(0.40)	(1.11)	(0.19)	(0.41)	(0.08)	(0.23)
T_{-7}	1724.6	1724.6	140.0	140.0	191.1	191.1	-31.4	-31.4
	(0.39)	(0.72)	(0.34)	(1.05)	(0.04)	(0.11)	(-0.07)	(-0.26)
T_{-6}	1871.0	1871.0	181.8	181.8	336.7	336.7	0.71	0.71
	(0.42)	(0.77)	(0.44)	(1.21)	(0.07)	(0.18)	(0.00)	(0.00)
T_{-5}	1938.7	1938.7	113.0	113.0	1148.0	1148.0	2.63	2.63
	(0.44)	(0.52)	(0.27)	(0.69)	(0.25)	(0.43)	(0.01)	(0.02)
T_{-4}	4749.3	4749.3	105.7	105.7	3246.5	3246.5	11.8	11.8
	(1.07)	(0.90)	(0.26)	(0.52)	(0.72)	(0.85)	(0.03)	(0.08)
T_{-3}	4824.1	4824.1	135.0	135.0	3256.6	3256.6	-36.2	-36.2
	(1.09)	(0.96)	(0.33)	(0.69)	(0.72)	(0.89)	(-0.09)	(-0.23)
T_{-2}	2815.0	2815.0	132.0	132.0	715.6	715.6	-159.1	-159.1
	(0.64)	(0.90)	(0.32)	(0.92)	(0.16)	(0.31)	(-0.38)	(-1.02)
T_{-1}	-73.4	-73.4	24.6	24.6	-410.9	-410.9	-208.7	-208.7
	(-0.02)	(-0.15)	(0.06)	(0.22)	(-0.09)	(-0.85)	(-0.50)	(-1.63)
T_1	2703.2	2703.2	102.4	102.4	1430.5	1430.5	167.8	167.8*
	(0.61)	(1.21)	(0.25)	(0.91)	(0.32)	(0.75)	(0.40)	(1.80)
T_2	8359.6*	8359.6	348.9	348.9	5772.3	5772.3	603.2	603.2***
	(1.89)	(1.36)	(0.84)	(1.34)	(1.28)	(1.21)	(1.43)	(2.92)
T_3	11010.6**	11010.6	591.9	591.9*	7713.5*	7713.5	741.1*	741.1***
	(2.49)	(1.62)	(1.43)	(1.85)	(1.71)	(1.37)	(1.76)	(2.83)
T_4	13141.9***	13141.9*	911.5**	911.5**	8458.9*	8458.9	837.8**	837.8**
	(2.97)	(1.87)	(2.20)	(2.48)	(1.87)	(1.41)	(1.99)	(2.70)
T_5	12678.9***	12678.9*	1299.1***	1299.1*	6092.5	6092.5	825.6*	825.6
	(2.87)	(1.77)	(3.14)	(1.94)	(1.35)	(0.83)	(1.96)	(1.48)
T_6	12636.4***	12636.4	1414.6***	1414.6	5822.2	5822.2	903.5**	903.5
	(2.86)	(1.56)	(3.42)	(1.58)	(1.29)	(0.69)	(2.15)	(1.26)
T_7	7994.2*	7994.2	1104.3***	1104.3	2976.4	2976.4	632.7	632.7
	(1.81)	(1.62)	(2.67)	(1.44)	(0.66)	(0.59)	(1.50)	(1.08)
T_8	7430.2*	7430.2	1013.5**	1013.5	2240.5	2240.5	668.0	668.0
	(1.68)	(1.33)	(2.45)	(1.39)	(0.50)	(0.37)	(1.59)	(1.18)
T_9	5823.9	5823.9	860.8**	860.8	1230.7	1230.7	524.0	524.0
	(1.32)	(1.21)	(2.08)	(1.31)	(0.27)	(0.23)	(1.25)	(1.02)
T_{10}	3033.9	3033.9	566.2	566.2	-380.8	-380.8	193.8	193.8
	(0.69)	(0.96)	(1.37)	(1.32)	(-0.08)	(-0.12)	(0.46)	(0.57)
Constant	-3411.4	-3411.4**	-427.2*	-427.2***	-3494.0	-3494.0	-395.2	-395.2**
	(-1.29)	(-2.17)	(-1.73)	(-3.64)	(-1.16)	(-1.57)	(-1.41)	(-2.53)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE		✓		✓		✓		✓
Countries	27	27	27	27	27	27	27	27
Observations	540	540	540	540	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This table presents detailed results for the nonparametric event studies of robustness tests one and two. Columns (1) and (2) as well as (5) and (6) have total net migration y_{jt}^{net} as the dependent variable, columns (3), (4), (7) and (8) the net migration rate $\tilde{y}_{jt}^{net}$. Columns (1) - (4) are for the first robustness exercise, i.e. Δ_j^u is defined as $\Delta_j^u = u_{j,2013} - u_{j,2009}$ and columns (5) - (8) are for the second robustness exercise in which $\Delta_j^u = u_{j,2012} - u_{j,2008}$. Columns (2), (4), (6) and (8) cluster standard errors by country level.

Table C.13: Nonparametric Event Studies Robustness 3 and 4

	y_{jt}^{net}		$\tilde{y}_{jt}^{net}$		y_{jt}^{net}		$\tilde{y}_{jt}^{net}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T_{-9}	1545.8	1545.8	203.0	203.0	2225.1	2225.1	226.3	226.3
	(0.35)	(0.74)	(0.49)	(1.18)	(0.48)	(1.29)	(0.52)	(1.19)
T_{-8}	1997.9	1997.9	180.6	180.6	2661.1	2661.1	201.2	201.2
	(0.45)	(0.88)	(0.44)	(1.24)	(0.58)	(1.31)	(0.46)	(1.21)
T_{-7}	1558.6	1558.6	144.6	144.6	1514.4	1514.4	135.9	135.9
	(0.35)	(0.76)	(0.35)	(1.14)	(0.33)	(0.89)	(0.31)	(0.98)
T_{-6}	1814.2	1814.2	205.7	205.7	1603.1	1603.1	201.0	201.0
	(0.41)	(0.86)	(0.50)	(1.42)	(0.35)	(0.90)	(0.46)	(1.25)
T_{-5}	1881.8	1881.8	123.8	123.8	1399.4	1399.4	149.2	149.2
	(0.42)	(0.59)	(0.30)	(0.82)	(0.30)	(0.54)	(0.34)	(0.98)
T_{-4}	4325.1	4325.1	105.2	105.2	3019.7	3019.7	98.4	98.4
	(0.98)	(0.95)	(0.25)	(0.58)	(0.66)	(0.84)	(0.23)	(0.58)
T_{-3}	4343.7	4343.7	123.9	123.9	3235.4	3235.4	138.8	138.8
	(0.98)	(1.01)	(0.30)	(0.70)	(0.70)	(0.94)	(0.32)	(0.81)
T_{-2}	2044.5	2044.5	68.0	68.0	1963.8	1963.8	57.7	57.7
	(0.46)	(0.76)	(0.16)	(0.44)	(0.43)	(0.90)	(0.13)	(0.29)
T_{-1}	17.0	17.0	21.0	21.0	255.9	255.9	64.5	64.5
	(0.00)	(0.04)	(0.05)	(0.17)	(0.06)	(0.54)	(0.15)	(0.38)
T_1	1840.3	1840.3	61.8	61.8	1997.6	1997.6	22.1	22.1
	(0.42)	(0.90)	(0.15)	(0.59)	(0.43)	(1.04)	(0.05)	(0.23)
T_2	6169.0	6169.0	233.3	233.3	6290.4	6290.4	180.3	180.3
	(1.39)	(1.13)	(0.56)	(0.93)	(1.37)	(1.36)	(0.41)	(0.73)
T_3	8061.8*	8061.8	427.0	427.0	8743.0*	8743.0	444.9	444.9
	(1.82)	(1.29)	(1.03)	(1.37)	(1.90)	(1.58)	(1.02)	(1.57)
T_4	9868.1**	9868.1	709.4*	709.4*	9826.4**	9826.4	672.3	672.3**
	(2.23)	(1.51)	(1.71)	(1.99)	(2.13)	(1.66)	(1.54)	(2.12)
T_5	9181.4**	9181.4	1032.2**	1032.2	6896.5	6896.5	821.8*	821.8
	(2.07)	(1.30)	(2.49)	(1.68)	(1.50)	(0.90)	(1.89)	(1.49)
T_6	8753.4**	8753.4	1067.9**	1067.9	6002.8	6002.8	763.8*	763.8
	(1.98)	(1.09)	(2.57)	(1.32)	(1.30)	(0.68)	(1.75)	(1.08)
T_7	5601.2	5601.2	873.2**	873.2	4069.2	4069.2	584.0	584.0
	(1.26)	(1.15)	(2.10)	(1.29)	(0.88)	(0.77)	(1.34)	(1.03)
T_8	4826.6	4826.6	774.3*	774.3	3388.2	3388.2	468.0	468.0
	(1.09)	(0.86)	(1.87)	(1.19)	(0.74)	(0.53)	(1.07)	(0.83)
T_9	3734.6	3734.6	672.5	672.5	2251.8	2251.8	382.2	382.2
	(0.84)	(0.77)	(1.62)	(1.15)	(0.49)	(0.40)	(0.88)	(0.75)
T_{10}	1931.5	1931.5	477.6	477.6	997.5	997.5	261.9	261.9
	(0.44)	(0.63)	(1.15)	(1.25)	(0.22)	(0.29)	(0.60)	(0.77)
Constant	-3638.7	-3638.7*	-456.0*	-456.0***	-4057.7	-4057.7	-496.5*	-496.5***
	(-1.30)	(-2.04)	(-1.74)	(-3.47)	(-1.30)	(-1.70)	(-1.69)	(-2.89)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE		✓		✓		✓		✓
Countries	27	27	27	27	27	27	27	27
Observations	540	540	540	540	540	540	540	540

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This table presents detailed results for the nonparametric event studies of robustness tests three and four. Columns (1) and (2) as well as (5) and (6) have total net migration y_{jt}^{net} as the dependent variable, columns (3), (4), (7) and (8) the net migration rate $\tilde{y}_{jt}^{net}$. Columns (1) - (4) are for the third robustness exercise, i.e. with threshold $\bar{\Delta} = 1$ and columns (5) - (8) are for the fourth robustness exercise in which $\bar{\Delta} = 0$. Columns (2), (4), (6) and (8) cluster standard errors by country level.

Table C.14: Nonparametric Event Studies Robustness 5 and 6

	y_{jt}^{net}		$\tilde{y}_{jt}^{net}$		y_{jt}^{net}		$\tilde{y}_{jt}^{net}$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
T_{-9}	1626.8	1626.8	100.8	100.8	2155.6	2155.6	178.9	178.9
	(0.36)	(0.63)	(0.36)	(0.60)	(0.62)	(0.80)	(0.91)	(1.20)
T_{-8}	1886.9	1886.9	93.5	93.5	1848.0	1848.0	129.7	129.7
	(0.42)	(0.68)	(0.34)	(0.64)	(0.53)	(0.64)	(0.66)	(0.97)
T_{-7}	1026.1	1026.1	47.2	47.2	1272.7	1272.7	101.5	101.5
	(0.23)	(0.39)	(0.17)	(0.35)	(0.36)	(0.46)	(0.52)	(0.86)
T_{-6}	1156.8	1156.8	81.7	81.7	1403.1	1403.1	147.5	147.5
	(0.26)	(0.43)	(0.29)	(0.56)	(0.40)	(0.50)	(0.75)	(1.17)
T_{-5}	1449.2	1449.2	33.4	33.4	1860.4	1860.4	117.3	117.3
	(0.32)	(0.35)	(0.12)	(0.19)	(0.53)	(0.42)	(0.60)	(0.79)
T_{-4}	4300.8	4300.8	32.3	32.3	5139.8	5139.8	138.8	138.8
	(0.96)	(0.74)	(0.12)	(0.15)	(1.47)	(0.82)	(0.70)	(0.77)
T_{-3}	4378.2	4378.2	56.1	56.1	5159.6	5159.6	157.2	157.2
	(0.98)	(0.79)	(0.20)	(0.27)	(1.48)	(0.87)	(0.80)	(0.89)
T_{-2}	2467.0	2467.0	73.4	73.4	2884.9	2884.9	59.2	59.2
	(0.55)	(0.72)	(0.26)	(0.49)	(0.83)	(0.76)	(0.30)	(0.37)
T_{-1}	-451.8	-451.8	-5.74	-5.74	-594.7	-594.7	-3.09	-3.09
	(-0.10)	(-0.92)	(-0.02)	(-0.05)	(-0.17)	(-1.15)	(-0.02)	(-0.03)
T_1	2924.2	2924.2	95.6	95.6	3508.1	3508.1	60.1	60.1
	(0.65)	(1.22)	(0.34)	(0.80)	(1.00)	(1.39)	(0.30)	(0.57)
T_2	9300.9**	9300.9	392.7	392.7	10648.3***	10648.3	333.9*	333.9
	(2.08)	(1.40)	(1.42)	(1.42)	(3.05)	(1.47)	(1.69)	(1.22)
T_3	12169.0***	12169.0	632.7**	632.7*	14213.1***	14213.1*	589.7***	589.7*
	(2.72)	(1.67)	(2.28)	(1.87)	(4.07)	(1.81)	(2.99)	(1.76)
T_4	13184.1***	13184.1*	689.8**	689.8**	16076.7***	16076.7*	728.7***	728.7**
	(2.94)	(1.71)	(2.49)	(2.30)	(4.60)	(1.97)	(3.70)	(2.38)
T_5	11397.5**	11397.5	769.9***	769.9*	14421.3***	14421.3*	705.6***	705.6***
	(2.54)	(1.50)	(2.77)	(1.95)	(4.13)	(2.02)	(3.58)	(2.81)
T_6	10539.3**	10539.3	675.8**	675.8	13751.6***	13751.6*	573.3***	573.3*
	(2.35)	(1.26)	(2.44)	(1.42)	(3.94)	(1.82)	(2.91)	(1.95)
T_7	5786.9	5786.9	430.9	430.9	7657.8**	7657.8*	0	369.4
	(1.29)	(1.25)	(1.55)	(1.39)	(2.19)	(1.90)	(.)	(1.66)
T_8	5618.1	5618.1	411.4	411.4	7996.3**	7996.3*	346.7*	346.7
	(1.25)	(1.02)	(1.48)	(1.18)	(2.29)	(1.75)	(1.76)	(1.59)
T_9	4353.7	4353.7	331.5	331.5	6594.2*	6594.2*	293.3	293.3
	(0.97)	(0.94)	(1.19)	(1.05)	(1.89)	(1.80)	(1.49)	(1.38)
T_{10}	2014.8	2014.8	232.5	232.5	3031.0	3031.0	183.3	183.3
	(0.45)	(0.65)	(0.84)	(1.01)	(0.87)	(1.10)	(0.93)	(1.00)
Constant	-3446.5	-3446.5**	-424.4***	-424.4***	-2026.6	-2026.6***	-358.0***	-358.0***
	(-1.33)	(-2.19)	(-2.64)	(-3.60)	(-1.03)	(-3.87)	(-3.22)	(-3.52)
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Clustered SE		✓		✓		✓		✓
Countries	27	27	27	27	27	27	26	24
Observations	540	540	540	540	540	540	520	480

Notes: t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

This table presents detailed results for the nonparametric event studies of robustness tests five and six. Columns (1) and (2) as well as (5) and (6) have total net migration y_{jt}^{net} as the dependent variable, columns (3), (4), (7) and (8) the net migration rate $\tilde{y}_{jt}^{net}$. Columns (1) - (4) are for the fifth robustness exercise, i.e. with Croatia excluded and columns (5) - (8) are for the sixth robustness exercise in which Croatia, Romania and Bulgaria are excluded. Columns (2), (4), (6) and (8) cluster standard errors by country level.

Bibliography

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, *80*(5), 1977–2016.

Amiti, M., Itskhoki, O., & Konings, J. (2019). International shocks, variable markups and domestic prices. *Review of Economic Studies*, *86*(6), 2356–402.

Anderson, J. E. & van Wincoop, E. (2004). Trade Costs. *Journal of Economic Literature*, *42*(3), 691–751.

Atkeson, A. & Burstein, A. (2008). Trade costs, pricing-to-market, and international relative prices. *American Economic Review*, *98*(5), 1998–2031.

Azinović, M., Gaegauf, L., & Scheidegger, S. (2019). Deep Equilibrium Nets. Mimeo, University of Zurich. `https://ssrn.com/abstract=3393482`.

Bach, F. (2017). Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, *18*(1), 629–681.

Bagwell, K. & Lee, S. H. (2020). Trade policy under monopolistic competition with firm selection. *Journal of International Economics*, *127*, 103379.

Bagwell, K. & Staiger, R. W. (2004). *The Economics of the World Trading System*. The MIT Press, Cambridge MA.

Bandeira, G., Caballé, J., & Vella, E. (2019). Fiscal Austerity and Greek Migration: A Missing Link. Working Papers 1075, Barcelona Graduate School of Economics.

Barron, A. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, *39*(3), 930–945.

Barros, P. P. & Cabral, L. (1994). Merger policy in open economies. *European Economic Review*, *38*(5), 1041–1055.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: a survey. *The Journal of Machine Learning Research*, *18*(153), 1–43.

Bellman, R. (1958). *Dynamic programming*. Princeton University Press.

Bertoli, S., Brücker, H., & Fernández-Huertas Moraga, J. (2016). The european crisis and migration to germany. *Regional Science and Urban Economics*, *60*, 61–72.

Beyer, R. C. M. & Smets, F. (2015). Labour market adjustments and migration in Europe and the United States: how different? *Economic Policy*, *30*(84), 643–682.

Blanchard, O. J. & Katz, L. F. (1992). Regional evolutions. *Brookings Papers on Economic Activity*, *1*, 1–75.

Blonigen, B. A. & Prusa, T. J. (2008). Antidumping. In E. Choi & J. Harrigan (Eds.), *Handbook of International Trade* chapter 9, (pp. 251–284). Wiley-Blackwell.

Boyd, J. P. (2001). *Chebyshev and Fourier spectral methods* (2nd ed., rev ed.). Mineola, N.Y: Dover Publications.

Boyd, J. P. & Petschek, R. (2014). The Relationships Between Chebyshev, Legendre and Jacobi Polynomials: The Generic Superiority of Chebyshev Polynomials and Three Important Exceptions. *Journal of Scientific Computing*, *59*(1), 1–27.

Brander, J. A. (1995). Chapter 27 strategic trade policy. volume 3 of *Handbook of International Economics* (pp. 1395 – 1455). Elsevier.

Brander, J. A. & Spencer, B. J. (1984). Trade warfare: Tariffs and cartels. *Journal of International Economics*, *16*(3–4), 227–242.

Breinlich, H., Nocke, V., & Schutz, N. (2019). Merger Policy in a Quantitative Model of International Trade. *The Economic Journal*, *130*(626), 393–421.

Broda, C. & Weinstein, D. (2006). Globalization and the gains from variety. *Quarterly Journal of Economics*, *121*(2), 541–85.

Brumm, J. & Scheidegger, S. (2017). Using Adaptive Sparse Grids to Solve HighDimensional Dynamic Models. *Econometrica*, *85*, 1575–1612.

Caliendo, L., Dvorkin, M., & Parro, F. (2019). Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock. *Econometrica*, *87*(3), 741–835.

Caliendo, L., Parro, F., Rossi-Hansberg, E., & Sarte, P.-D. (2017). The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy. *The Review of Economic Studies*, *85*(4), 2042–2096.

Cameron, A. & Miller, D. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*, 317–372.

Carvalho, V. M. & Gabaix, X. (2013). The Great Diversification and Its Undoing. *American Economic Review*, *103*(5), 1697–1727.

Carvalho, V. M. & Grassi, B. (2019). Large firm dynamics and the business cycle. *American Economic Review*, *109*(4), 1375–1425.

Chaney, T. (2008). Distorted Gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review*, *98*(4), 1707–1721.

Costinot, A., Rodríguez-Clare, A., & Werning, I. (2020). Micro to macro: Optimal trade policy with firm heterogeneity. *Econometrica*, *88*(6), 2739–2776.

Cover, T. & Thomas, J. (2012). *Elements of Information Theory, 2nd edition*. Wiley.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, *2*(4), 303–314.

Dao, M., Furceri, D., & Loungani, P. (2017). Regional Labor Market Adjustment in the United States: Trend and Cycle. *The Review of Economics and Statistics*, *99*(2), 243–257.

Davis, S. J., Faberman, R. J., & Haltiwanger, J. (2006). The Flow Approach to Labor Markets: New Data Sources and MicroMacro Links. *Journal of Economic Perspectives*, *20*(3), 3–26.

De Stefano, M. & Rysman, M. (2010). Competition policy as strategic trade with differentiated products. *Review of International Economics*, *18*(4), 758–771.

Decressin, J. & Fatás, A. (1995). Regional labor market dynamics in europe. *European Economic Review*, *39*(9), 1627–1655.

Demidova, S. & Rodríguez-Clare, A. (2013). The simple analytics of the melitz model in a small economy. *Journal of International Economics*, *90*(2), 266 – 272.

Den Haan, W. J., Ramey, G., & Watson, J. (2000). Job Destruction and Propagation of Shocks. *American Economic Review*, *90*(3), 482–498.

Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, *126*(3), 903–983.

di Giovanni, J. & Levchenko, A. A. (2012). Country Size, International Trade, and Aggregate Fluctuations in Granular Economies. *Journal of Political Economy*, *120*(6), 1083–1132.

di Giovanni, J., Levchenko, A. A., & Méjean, I. (2014). Firms, Destinations, and Aggregate Fluctuations. *Econometrica*, *82*(4), 1303–1340.

Dixit, A. (1984). International Trade Policy for Oligopolistic Industries. *Economic Journal*, *94*(376a), 1–16.

Donald, S. & Lang, K. (2007). Inference with difference-in-differences and other panel data. *The Review of Economics and Statistics*, *89*(2), 221–233.

Dornbusch, R., Fischer, S., & Samuelson, P. A. (1977). Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods. *American Economic Review*, *67*(5), 823–839.

Duarte, V. (2018). Machine Learning for Continuous-Time Finance.

Eaton, J. & Grossman, G. M. (1986). Optimal Trade and Industrial Policy Under Oligopoly. *The Quarterly Journal of Economics*, *101*(2), 383–406.

Eaton, J., Kortum, S., & Sotelo, S. (2012). International Trade: Linking Micro and Macro. NBER Working Papers No. 17864.

Eurostat (2021a). Glossary: Labour force. Accessed on March 23, 2021.

Eurostat (2021b). Glossary: Unemployment. Accessed on March 23, 2021.

Eurostat (2021c). Population change - demographic balance and crude rates at national level. Accessed on March 23, 2021.

Eurostat (2021d). Unemployment by sex and age. Accessed on March 23, 2021.

Felbermayr, G., Jung, B., & Larch, M. (2013). Optimal tariffs, retaliation, and the welfare loss from tariff wars in the melitz model. *Journal of International Economics*, *89*(1), 13 – 25.

Fernández-Villaverde, J. & Guerrón-Quintana, P. A. (2020). Estimating dsge models: Recent advances and future challenges. Working Paper 27715, National Bureau of Economic Research.

Fernández-Villaverde, J., Hurtado, S., & Nuño, G. (2019). Financial frictions and the wealth distribution. Working Paper 26302, National Bureau of Economic Research.

Fernández-Villaverde, J., Nuño, G., Sorg-Langhans, G., & Vogler, M. (2021). The dual role of migration - insurance and volatility in the schengen area. Work in Progress.

Furlanetto, F. & Robstad, A. (2019). Immigration and the macroeconomy: Some new empirical evidence. *Review of Economic Dynamics*, *34*, 1–19.

Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics*, *1*(1), 255–294.

Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. *Econometrica*, *79*(3), 733–772.

Gaubert, C. & Itskhoki, O. (2020). Granular comparative advantage. *Journal of Political Economy*. forthcoming, `http://www.princeton.edu/~itskhoki/papers/GranularCA.pdf`.

Glorot, X. & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Aritficial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, volume 9, (pp. 249–256).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Grassi, B. (2017). Io in i–o: Size, industrial organization, and the input-output network make a firm structurally important. Manuscript, Department of Economics, Bocconi University.

Greenwood, M. J. (1997). Chapter 12 internal migration in developed countries. volume 1 of *Handbook of Population and Family Economics* (pp. 647–720). Elsevier.

Haaland, J. I. & Venables, A. J. (2016). Optimal trade policy with monopolistic competition and heterogeneous firms. *Journal of International Economics*, *102*, 85 – 95.

Hagedorn, M. & Manovskii, I. (2008). The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited. *American Economic Review*, *98*(4), 1692–1706.

Han, J., Jentzen, A., & E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, *115*(34), 8505–8510.

Hauser, D. & Seneca, M. (2019). Labor mobility in a monetary union. Bank of Canada Staff Working Paper 2019-15, Bank of Canada.

Head, K. & Ries, J. (1997). International mergers and welfare under decentralized competition policy. *Canadian Journal of Economics*, *30*(4), 1104–23.

Helpman, E. & Krugman, P. (1989). *Trade Policy and Market Structure*. The MIT Press, Cambridge MA.

Horn, H. & Levinsohn, J. (2001). Merger Policies and Trade Liberalisation. *The Economic Journal*, *111*(April), 244–276.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

House, C. L., Proebsting, C., & Tesar, L. L. (2018). Quantifying the benefits of labor mobility in a currency union. Working Paper 25347, National Bureau of Economic Research.

Huart, F. & Tchakpalla, M. (2015). Labour mobility in the euro area during the great recession. Vol 12.

Izquierdo, M., Jimeno, J. F., & Lacuesta, A. (2016). Spain: From massive immigration to vast emigration? *IZA Journal of Migration*, *5*(10), 1–20.

Kaplan, G. & Schulhofer-Wohl, S. (2017a). Understanding the Long-Run Decline in Interstate Migration. *International Economic Review*, *58*(1), 57–94.

Kaplan, G. & Schulhofer-Wohl, S. (2017b). Understanding the long-run decline in interstate migration. *International Economic Review*, *58*(1), 57–94.

Kennan, J. & Walker, J. R. (2011). The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, *79*(1), 211–251.

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, *10*(2), 1–26.

Lozej, M. (2019). Economic migration and business cycles in a small open economy with matching frictions. *Economic Modelling*, *81*, 604–620.

Lucas, R. E. (1997). Chapter 13 internal migration in developing countries. volume 1 of *Handbook of Population and Family Economics* (pp. 721–798). Elsevier.

Maffei-Faccioli, N. & Vella, E. (2020). The asymmetric unemployment response of natives and foreigners to migration shocks. Working Papers 2020008, The University of Sheffield, Department of Economics.

Maliar, L., Maliar, S., & Winant, P. (2019). Will Artificial Intelligence Replace Computational Economists Any Time Soon? CEPR Discussion Papers 14024, C.E.P.R. Discussion Papers.

McCaffrey, D. F. & Bell, R. M. (2006). Improved hypothesis testing for coefficients in generalized estimating equations with small samples of clusters. *Statistics in Medicine*, *25*(23), 4081–4098.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, *71*(6), 1695–1725.

Molloy, R., Smith, C., & Wozniak, A. (2017). Job changing and the decline in long-distance migration in the united states. *Demography*, *54*.

Molloy, R., Smith, C. L., Trezzi, R., & Wozniak, A. (2016). Understanding Declining Fluidity in the U.S. Labor Market. *Brookings Papers on Economic Activity*, *47*(1 (Spring), 183–259.

Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, *25*(3), 173–196.

Mundell, R. A. (1961). A theory of optimum currency areas. *American Economic Review*, *51*(4), 657–665.

Nakamura, E. & Steinsson, J. (2018). Identification in macroeconomics. *Journal of Economic Perspectives*, *32*(3), 59–86.

Oikonomou, M. (2020). Migration, search and skill heterogeneity. Job Market Paper.

Petrosky-Nadeau, N., Zhang, L., & Kuehn, L.-A. (2018). Endogenous Disasters. *American Economic Review*, *108*(8), 2212–2245.

Pissarides, C. (2009). The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer? *Econometrica*, *77*(5), 1339–1369.

Redding, S. J. (2010). The empirics of new economic geography. *Journal of Regional Science*, *50*(1), 297–311.

Redding, S. J. (2013). *Economic Geography: A Review of the Theoretical and Empirical Literature,*, (pp. 497–531). London: Palgrave Macmillan UK.

Redding, S. J. & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*(1), 21–58.

Richardson, M. (1999). Trade and competition policies: Concordia discors? *Oxford Economic Papers*, *51*(4), 649–664.

Rokicki, S., Cohen, J., Fink, G., Salomon, J. A., & Landrum, M. B. (2018). Inference with difference-in-differences with a small number of groups. *Medical Care*, *56*(1), 97–105.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Schneider, J. (2012). Practical measures for reducing irregular migration. Federal Office for Migration and Refugees Working Paper No. 41.

Shimer, R. (2005). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American Economic Review*, *95*(1), 25–49.

Sirignano, J. & Spiliopoulos, K. (2018). DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, *375*, 1339–1364.

Smith, C. & Thoenissen, C. (2019). Skilled migration and business cycle dynamics. *Journal of Economic Dynamics and Control*, *109*, 103781.

Statistisches Bundesamt (2021). Wanderungen zwischen deutschland und dem ausland: Deutschland, jahre, staaten der europäischen union, geschlecht. Wiesbaden, Accessed on March 22, 2021.

Sutton, R. S. & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book.

Vogel, D. & Aßner, M. (2011). Umfang, entwicklung und struktur der irregulären bevölkerung in deutschland. Mimeo. Bundesamt fur Migration und Flüchtlinge, Hamburg und Nürnberg.

Vogler, M. (2021). The migration response to the euro crisis - an empirical investigation. Working Paper.