Essays on Trade, Technology, and Banking

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Economics

have examined a dissertation entitled

"Essays on Trade, Technology, and Banking"

presented by Jeffrey Wang

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature Pol Antras
Pol Antras (Apr 20, 2022 10:08 EDT)

Typed name: Prof. Pol Antras

Signature Gordon Hanson

Typed name: Prof. Gordon Hanson

Signature Elhanan Helpman
Elhanan Helpman (Apr 22, 2022 11:19 EDT)

Typed name: Prof. Elhanan Helpman

Signature Marc Melitz
Marc Melitz (Apr 19, 2022 13:34 EDT)

Typed name: Prof. Marc Melitz

Date: April 8, 2022

Essays on Trade, Technology, and Banking

A dissertation presented

by

Jeffrey Wang

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2022

Dissertation Advisor:
Professor Pol Antras

Author:
Jeffrey Wang

Essays on Trade, Technology, and Banking

Abstract

This dissertation comprises three essays on technology, international trade, and banking. The first two essays investigate industrial robots and their economic impact. In chapter 1, I construct a novel dataset that measures firm-level robot adoption among US manufacturing firms. I summarize a series of stylized facts related to robot-adopting firms in the US, and present descriptive evidence for the relationship between robots and labor market outcomes. I also develop a potential instrumental variable for robot adoption, based on immigrant shocks to the labor supply of robot-complementary workers. In chapter 2, I rely on the constructed dataset to analyze the effect of robot adoption on imports and offshoring, using a mix of theory and empirics. I find that the rise of robots in the US led to an increase in offshoring, both at the firm level and in aggregate. However, the increase in imports and the resulting efficiency gains come at the cost of domestic manufacturing workers, and particularly assembly workers. In chapter 3, coauthored with David Hao Zhang, we turn attention to small businesses in the US and their access to government support distributed through banks, such as the Paycheck Protection Program (PPP). We document geographical and racial disparities in the density of financial institutions across US commuting zones, and show that these disparities translate into differences in take-up of the PPP.

Contents

List of Tables

List of Figures

Acknowledgments

This dissertation would not have been possible without the guidance and encouragement I received from my advisors, Pol Antràs, Elhanan Helpman, Gordon Hanson, and Marc Melitz. In particular, I am indebted to Pol for his lasting support, which helped me develop and polish the first two chapters of this dissertation. I am also grateful to Elhanan, Gordon, and Marc for their advice and feedback which immensely improved this dissertation. In addition, I thank Teresa Fort for granting me access to the Census datasets used in the first two chapters, and for providing guidance as I navigated the data.

I also benefited greatly from discussions with additional faculty members and colleagues. For the first two chapters, I thank Isaiah Andrews, Adrien Bilal, Sahil Chinoy, Xiang Ding, Jiacheng Feng, Teresa Fort, Xavier Jaravel, Savannah Noray, Benji Niswonger, Rodimiro Rodrigo, Katherine Stapleton, and David Hao Zhang for helpful comments. Jim Davis at the US Census Bureau provided support with the disclosure process. For the third chapter, co-authored with David Hao Zhang, we thank Isaiah Andrews, Ed Glaeser, Claire Shi, Adi Sunderam, and Emily Williams for helpful comments.

My classmates and friends were constant sources of joy and support during the time I spent in Cambridge. I thank Jiacheng Feng and David Hao Zhang for going through all the ups and downs with me over the past six years. I also thank Lydia Cox, Yunan Ji, Gabriel Levin-Konigsberg, Jiahua Liu, Tianwang Liu, Hillary Stein, Wentao Xiong, and Allen Zhang for their inspiration and support at numerous points throughout my PhD experience. I have not discussed contents of this dissertation with my friends at HAcappella, but their company is a major reason why I persisted through the many challenges in graduate school.

Finally, I would not have completed this journey without the unconditional love from my parents. I dedicate this dissertation to them.

To my parents

Introduction

This dissertation consists of three essays related to international trade, technology, and banking. In the first essay, I start with the observation that the recent rise of scholarly interest in the economic impact of industrial robots has prompted a sharp increase in demand for datasets measuring robot use at the firm level. However, such data is lacking especially in the US. Using US Census microdata, I construct a novel dataset that measures firm-level robot adoption through robot imports. My methodology captures 86% of industrial robots used by US manufacturing firms, and uncovers three sets of descriptive facts: (1) robot adoption is lumpy; (2) robot investment is costly; (3) robot adoption exhibits selection on size and productivity. I also show that an exogenous shock to the labor supply of robot-complementary workers increases the probability of robot adoption at the firm level. Using the constructed shock as an instrument for robot adoption, I find that at the firm level, robots lead to an increase in employment and wages, and a within-firm reallocation in wage bill away from manufacturing workers.

In the second essay, I turn to the influence of robots on a particular economic outcome. What is the impact of automation on trade? This essay studies the impact of industrial robots on US manufacturing firms, with a focus on trade and offshoring. I construct a parsimonious model of two-stage production that incorporates automation and offshoring in both production of intermediate inputs (upstream production) and assembly (downstream production). I then utilize the IV strategy developed in the first essay to examine the impact of robot adoption on firm-level outcomes. Using a detailed administrative dataset of US manufacturing firms, I find that at the firm level, robot adoption has a positive effect on imports of both intermediate inputs and final outputs, but the effect is significantly larger for input imports. Robot adoption also leads to an increase in sales and productivity. These empirical findings are consistent with predictions of the model and the hypothesis that robots are mainly used in assembly. Quantitative estimation of the model confirms the comparative advantage of robots in downstream production. Following a positive shock to the productivity of robots, robot adopters increase their sales, imports, and expenditure on domestic inputs but not domestic assembly, while non-adopters contract in all dimensions due to within-industry competition. In aggregate, the rise of industrial robots in the US between 1992 and 2012 is associated with a 15% increase in imports and a 14% decrease in domestic manufacturing employment, with significant heterogeneity across different types of imports and workers. I use the estimated model to evaluate costs and benefits of robot taxes and subsidies.

In the third essay, I focus on the geographical and racial disparities in the location of financial institutions and what these disparities imply for small businesses. Many government support programs for small businesses are designed to pass through banks and credit unions. However, this poses barriers for minority communities that are less connected to financial institutions for obtaining this support. Using the latest program for supporting small businesses, the Paycheck Protection Program (PPP), as an example, we show that there was a large disparity in both the presence and density of PPP enrolled lenders by racial composition of the neighborhood. This difference is both due to a lower number of lenders in those neighborhoods in general, and by the fact that the lenders that do operate there are small credit unions without a previous relationship with the Small Business Administration. More heavily Black neighborhoods have significantly lower take-up of PPP loans, particularly in lower population (more rural) areas where this disparity is most salient. Through an instrumental variables analysis, we show that the intensive margin of access to enrolled lenders can explain about 32% of the racial disparity in take up within the relevant areas. Our results suggest that government programs that provide "support through banks” can have undesirable distributional implications.

Chapter 1

Industrial Robots in US Manufacturing: New Data and Facts

1.1 Introduction

Recent years have seen a rapid rise in interest among researchers in industrial robots and their economic effects. While earlier work in this literature focuses on the relationship between robots and economic outcomes at a broader level (e.g., country, industry), in the past few years, researchers have devoted an increasing amount of attention to the firm-level effects of robots, and more generally, automation. The availability of high-quality data that measures robot adoption at the firm level is evidently important for such an effort. However, there is a lack of data providing comprehensive information on firm-level robot adoption. In particular, most existing datasets on robot use are based on either survey data covering a small sample or data at the industry level. This poses difficulties for a thorough analysis of how robot adoption shapes firm-level incentives and outcomes.

In this paper, I address this issue by constructing a novel dataset recording firm-level robot adoption among US manufacturing firms. Although there does not exist a dataset that directly reports robot use at the firm level for the years prior to 2016, I take advantage of the fact that most industrial robots used in the US are produced abroad, and measure robot adoption through robot imports. This allows me to infer robot adoption to an extent that covers the majority of industrial robots used by US manufacturing firms.

I begin by describing how I construct the dataset. My methodology is based on highly disaggregated US Census data containing firm-level information on sales, employment, and trade for the universe of US manufacturing firms. After identifying the universe of robot importers, I exclude wholesalers and robot producers potentially purchasing robots as intermediate inputs. To validate my approach, I show that robot imports (after the cleaning steps) account for the majority of industrial robots used in the US. Cross-industry patterns of robot adoption identified using my methodology are also consistent with widely used industry-level data on robots in the US.

Using this methodology, I identify 1600 robot-adopting firms in US manufacturing. I present descriptive facts that highlight a few features of robot adoption and robot adopters in the US. First, robot investment is lumpy: the majority of robot adopters purchase robots in a single year and from a single country. Second, robot investment is costly. On average, each robot adopter purchases industrial robots at a value of $2 million. As a result, by 2012, only 0.5% of US manufacturing have adopted robots. Lastly, robot adopters are significantly larger than non-adopters along multiple dimensions. In 2012, the 0.5% of robot adopters account for around 10% of total sales and employment in the manufacturing sector. Cross-sectional regressions also confirm that robot adopters have higher sales, employment, and productivity.

I then document evolution of firm outcomes around the identified events of robot adoption. Using event study regressions in the style of Stock and Watson (2015), I show that robot adoption is associated with a within-firm increase in sales, employment, average wage, and productivity. However, firms do not only expand along these dimensions following robot adoption; they expand prior to adoption as well, and this pre-trend is statistically significant in the cases of sales and employment. In other words, firms that eventually adopt robots already tend to grow faster than non-adopters before adoption occurs. This finding points to the possibility of selection into robot adoption based on size and productivity. A set of cross-sectional regressions comparing firms that adopt robots in the future and those that do not also supports this hypothesis.

The results discussed above suggest that robot adoption is far from a random event. In order to further explore factors that determine which firms adopt robots and develop a potential instrument for robot adoption, I consider the labor supply of robot-complementary workers. Even though industrial robots are by definition automated and do not require human operation, human workers of certain occupations are helpful for the process of installing, maintaining, and repairing robots. I thus hypothesize that an increase in the availability of such workers at a geographical level leads to a higher likelihood of local firms adopting robots. I test this hypothesis in two steps. First, I identify occupations that are complementary with robots using occupational information provided by the O*NET database. Second, given that occupational composition is correlated with other characteristics of local labor markets, I use imputed immigrant inflows to proxy for the change in the labor supply of robot-complementary workers, and construct a shift-share shock following the logic of Altonji and Card (1991). Specifically, I impute the number of immigrants in robot-complementary occupations that arrived in a commuting zone (CZ) since a baseline year, 1990, based on settlement patterns of immigrants and occupational composition in 1990. This results in a shift-share shock that is correlated with the labor supply of robot complementary workers and plausibly exogenous to changes in local economic conditions that may affect robot adoption through alternative channels.

Panel regressions confirm that the constructed shock is a statistically significant predictor of robot adoption at the firm level. In addition, to support exogeneity of this shock and test it against the exclusion restriction in an instrumental variables setting, I show that it does not predict adoption of non-robot machinery, and is uncorrelated with outcomes of firms that do not adopt robots. Using this shock as an instrument in a 2SLS model, I find that robot adoption leads to an increase in overall employment and wage at the firm level, but not for manufacturing workers. This result points to the possibility of within-firm reallocation for robot-adopting firms.

This paper is related to several strands of literature. First and foremost, it complements a growing number of datasets that measure robot adoption at the firm level. The most widely used dataset on industrial robots is IFR (2018), which estimates the total number of robots delivered and used from 1993 to 2016 at the country-industry and application level. Empirical papers have utilized this data source to construct measures of robot adoption and robot exposure at the level of country, industry, or geographical areas within a country (Graetz and Michaels (2018); Artuc *et al.* (2019); Acemoglu and Restrepo (2020)). More recently, scholars have attempted to construct robot adoption variables at the firm level in various countries (Koch *et al.* (2019), Spain; Acemoglu *et al.* (2020b), Bonfiglioli *et al.* (2020), Aghion *et al.* (2020), France; Humlum (2019), Denmark). The majority of datasets used in these papers are based on survey data covering a non-comprehensive sample of firms in the respective countries. I contribute to this literature by constructing a dataset that indicates firm-level robot adoption status for the universe of manufacturing firms in the US. As far as my knowledge goes, this is the first such dataset that records robot use for an extended period of time in the US.

Second, I contribute to an emerging literature documenting stylized facts about robot adoption. At an aggregate level, I use a new data source based on customs data to show an increase in robot use and cross-industry patterns that are similar to those documented in Acemoglu and Restrepo (2020). At the firm level, I find that in the US, robot adopters are larger and more productive than non-adopters and experience persistent growth in employment, sales, and productivity, both *before* and *after* adoption. These patterns are consistent with those in Spain, France, and Denmark as documented in Koch *et al.* (2019), Acemoglu *et al.* (2020b), and Humlum (2019). In addition, similar to Humlum (2019), I provide evidence that robot adoption in the US is a lumpy and costly investment.

Finally, I provide novel results on the relationship between availability of robot-complementary workers and the likelihood of robot adoption. In contrast with an existing literature seeking to classify occupations that are particularly *vulnerable* to being displaced by automation technologies (Webb (2019), Acemoglu *et al.* (2020a)), I identify occupations that are *helpful* in the process of installing, integrating, and working with technologies. My finding of a positive relationship between exogenous shocks to the labor supply of these robot-complementary workers and robot adoption among local firms suggests that it is important to consider heterogeneous effects of robots on different types of workers.

The rest of this paper proceeds as follows. section 2.3 constructs the dataset. section 1.3 presents descriptive facts related to robot adoption in US manufacturing and results from event study regressions. section 1.4 constructs an exogenous shock to the labor supply of robot-complementary workers and presents results based on the constructed variable. section 2.7 concludes.

1.2 Data Construction

In this section, I describe the sources and methodology I use to construct a dataset on robot adoption among US manufacturing firms.

1.2.1 Data Sources

The main dataset is constructed from three data sources on US manufacturing firms, all provided by the U.S. Census Bureau. I obtain firm-level trade data from the Longitudinal Firm Trade Transaction Database (LFTTD), and information on employment, payroll, and industry classification[1] from the Longitudinal Business Database (LBD), both recorded on a yearly basis. I identify manufacturing establishments as those with NAICS codes starting with the number "3", and identify manufacturing firms as firms containing at

[1] To ensure that industry classification is consistent over time, I use the system developed by Fort and Klimek (2018) to map all NAICS codes to a unique 2012 NAICS code.

least one manufacturing establishments[2]. In addition, the dataset incorporates data from the Economic Censuses (EC), which provide sales and expenditures of the universe of US establishments in 5-year intervals, on years ending in "2" or "'7" also known as Census years. For manufacturing establishments, I also utilize input-output information provided by material trailers (MT) and product trailers (PT) of the Census of Manufactures (CMF). In certain cases, to extend EC data to a yearly basis, I conduct linear interpolation on the log of sales and input expenditures observed in the EC to infer the same variables in non-Census years. The end product is a unbalanced panel dataset of the universe of US manufacturing firms between 1992 and 2016, containing more than 3 million observations.

1.2.2 Methodology

There does not exist a dataset that reports firm-level robot adoption in the US before 2016[3]. However, taking advantage of the fact that most producers of industrial robots are headquartered outside the United States (Leigh and Kraft, 2018), I identify firm-level adoption of industrial robots using robot imports.

The Harmonized System (HS) in the US assigns a specific set of product codes corresponding to industrial robots in imported goods. I identify these codes using archive files for the US Harmonic Tariff Schedule (HTS) from 1992 to 2016, which are accessed online[4]. In every year, there are two codes corresponding to two separate categories of industrial robots. The first category is described as "*industrial robots, not specified elsewhere*"; the second category is described as "*other lifting, handling, loading or unloading machinery: industrial robots*". The specific codes vary over time, and are summarized below:

[2]Following Ding *et al.* (2020), I include non-manufacturing establishments of manufacturing firms to account for sales of manufacturing output channeling through, say, wholesale establishments.

[3]A few recent US Census datasets provide firm-level information on use of robots. For example, the 2018 Annual Capital Expenditures Survey (ACES) contains data on firm-level capital expenditures for both industrial and service robotics; the 2018 Annual Business Survey (ABS) contains an explicit question that asks respondents to report use of technologies such as robotics and cloud computing (Zolas *et al.*, 2021); the 2018 Annual Survey of Manufactures (ASM) contains three new questions on the use of robotics (Buffington *et al.*, 2018).

[4]URL: https://usitc.gov/tata/hts/archive/index.htm

- Category 1: HS code 8479899040 from 1992 to 1994; HS code 8479899540 in 1995; 6-digit HS code 847950 from 1996 to 2016;
- Category 2: 8428900010 from 1992 to 2005 (except in 1998, when it was 8428908015), 8428900120 from 2006 to 2011, 8428900220 from 2012 to 2016.

In order to identify robot adopters, I start by identifying robot importers, i.e., firms that import industrial robots based on the codes listed above. Given that not all robot importers necessarily utilize robots in their production process, I take several additional approaches to eliminate potential carry-along (wholesale) traders, and potential robot producers purchasing robots as final output or as intermediate inputs from abroad.

First, for each robot importer, I compute the total value of robot imports and exports between 1992 and 2016. As defined in Bernard *et al.* (2019), carry along traders are firms exporting products that they do not manufacture. However, in the case of capital goods such as industrial robots, a firm may simultaneously engage in carry along trade *and* adoption of robots in its own production. Therefore, I only eliminate firms whose total value of robot exports exceed robot imports. The rationale is that these firms do not retain any of their purchased robots within the US, and thus certainly do not utilize robots in production[5].

Second, I use material trailers (MT) and product trailers (PT) of the Census of Manufactures (CMF) to identify firms that produce robots or use robots as intermediate inputs. The MT and PT contain detailed information on input purchases and sales by product for a selected sample of firms[6]. MT and PT follow separate product classification systems based on SIC (in 1992) or NAICS (in 1997 and onwards). In every Census year between 1992 and 2012, there exists a unique product code referring to "industrial robots", and a unique material code referring to "Industrial robots purchased for fabrication with welding

[5] In addition to reselling robots abroad, firms may also resell robots to other domestic firms. Therefore, even if a firm's robot exports are lower than its robot imports, it might still only engage in transit trade and not utilize robots in production. However, in the data, I do not observe domestic sales by product for the majority of firms.

[6] The material trailer and product trailer data are available only for a subset of manufacturing establishments, including multi-unit firms with at least 250 employees, and a randomly selected sample of single-unit firms.

equipment". The codes are:

- Year 2012: Product code 3339997109, appears in forms 33319, 33331, 33373
- Year 2007: Product code 3339997109, appears in forms 33319, 33331, 33373
- Year 2002: Product code 33399981092, appears in forms 33319, 33331, 33373
- Year 1997: Product code 35699099, appears in forms 3505, 3506, 3529, 3538, 3539
- Year 1992: Product code 35699099, appears in forms 3505, 3506, 3514, 3529

- Year 2012: Material Code 33399903, appears in forms 33326, 33331
- Year 2007: Material Code 33399903, appears in forms 33326
- Year 2002: Material Code 333999035, appears in forms 33326
- Year 1997: Material Code 3569712, appears in forms 3529
- Year 1992: Material Code 3569712, appears in forms 3529

Based on these codes, I identify establishments that potentially produce robots and/or use robots as intermediate inputs. I then identify all firms owning such establishments and eliminate them from the list of robot adopters.

Lastly, since product-level input and output information is not available for all firms or all years, I exclude a broader set of potential robot producers identified using establishment-level information on NAICS classification. Specifically, I exclude firms whose establishments fall into the following NAICS categories: 333999 (All Other Miscellaneous General Purpose Machinery Manufacturing), 3335 (Metalworking Machinery Manufacturing, which includes molding, cutting, and forming machinery), 33392 (Material Handling Equipment Manufacturing), 333992 (Welding and Soldering Equipment Manufacturing).

One potential issue associated with the methodology outlined above is that it does not allow us to observe industrial robots purchased from domestic producers. In light of this issue, using information from MT and PT, I identify 400 robot-producing firms in the United States. Although I do not directly observe their total production of robots, I infer their domestic sales through their robot exports and export intensity. According to customs data, from 1992 to 2016, these robot producers exported a total of 19,900 industrial robots, amounting to a total value of $6 billion. Robot producers are highly export-oriented, with an overall export intensity of 35%. Based on these statistics, I infer that these firms sold 37,000 industrial robots to firms in the US over the same time period. Given that the total number of industrial robots imported by US manufacturing firms during the same time period is 192,000, I conclude that domestic robot manufacturers only account for 16% of industrial robots used in the US. It is thus reasonable to assume that identifying robot adoption using robot imports yields a representative sample.

To further validate my approach of using robot imports to measure robot adoption, I compare the aggregate volume of robot adoption measured by my methodology to the same numbers from industry-level data provided by the International Federation of Robotics (IFR), which is one of the leading data sources on robotics used in the literature (Graetz and Michaels, 2018; Acemoglu and Restrepo, 2020). IFR estimates that a total of 223,000 industrial robots are operating in the United States by 2016, while my methodology indicates that a total of 192,000 industrial robots have been adopted by US manufacturing firms through imports between 1992 and 2016. In other words, robot imports roughly account for 86% of the total number of robots used in US manufacturing. I thus conclude that my methodology does capture the majority of robots used by US manufacturing firms.

1.3 Descriptive Statistics on Robot Adoption

In this section, I present a series of descriptive facts and preliminary analysis related to adoption of industrial robots in US manufacturing. In subsection 1.3.1, I document both micro and macro-level patterns of robot adoption. In subsection 2.4.1, I use cross-sectional regressions to compare robot-adopting firms to non-adopters along several dimensions. In subsection 2.4.2, I investigate evolution of firm-level outcomes for robot adopters over time, both before and after robot adoption.

1.3.1 Basic Facts

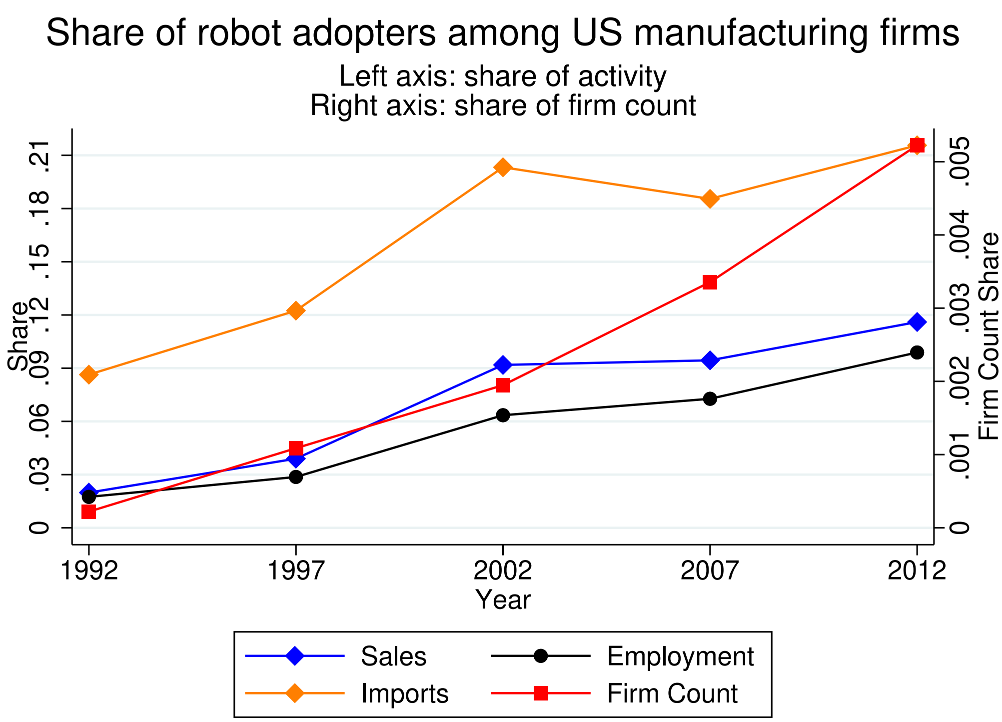
Using the methodology outlined in section 2.3, I identify around 1600 firms which have adopted robots between 1992 and 2016. Key features of the associated adoption events are displayed in Table 1.1. The average robot-adopting firm purchases 120.2 robots at a cost of $1.2 million dollars. In comparison, sales of the median US manufacturing firm in 2012 is $1.11 million. This indicates that robot investment is costly, and offers a potential explanation for the relatively small number of robot-adopting firms. In addition, robot adoption in the US is largely a one-time investment: 81% of firms purchase from one country only, 63% of firms purchase in one year only, and the peak year of investment on average accounts for 86% of total robot purchases[7]. This observed lumpiness of investment in industrial robots indicates that robot investment is pre-dominantly a one-time decision.

After evaluating micro-level patterns of robot adoption, let us turn our attention to macro-level trends. How important are robot adopters in the US economy? Figure 1.1 shows the evolution of the share of robot-adopting firms in the total number of firms, total sales, total employment, and total imports, between 1992 and 2012. The share of robot adopters has grown significantly over time in all dimensions. By 2012, about 0.5% of US manufacturing firms have adopted robots. While this may seem small in number, robot adopters are disproportionately large and play an important role in the economy: they account for close to 10% of employment and sales and around 20% of total imports. These patterns suggest that understanding what happens to robot adopters is quantitatively important for aggregate outcomes.

Finally, I investigate patterns of robot adoption across manufacturing industries. Fig-

[7] These patterns are similar to findings in Humlum (2019) for Denmark.

Figure 1.1: *Share of robot adopters among US manufacturing firms, 1992-2012*



Notes: this figure plots the change in the share of robot adopters among US manufacturing firms between 1992 and 2012 in 5-year intervals. The vertical axis corresponds to the share in sales, employment, and imports; the horizontal axis corresponds to the share in firm count.

Table 1.1: *Key characteristics of robot adoption events.* $N = 1600$

Variable	Mean	SD	Ones
Robot Value ($1000)	1235	5579	
Robot Quantity	120.2	1336	
Number of Years	2.166	2.596	1000
Number of Countries	1.364	1.318	1300
Peak Year Value Share	0.855	0.224	

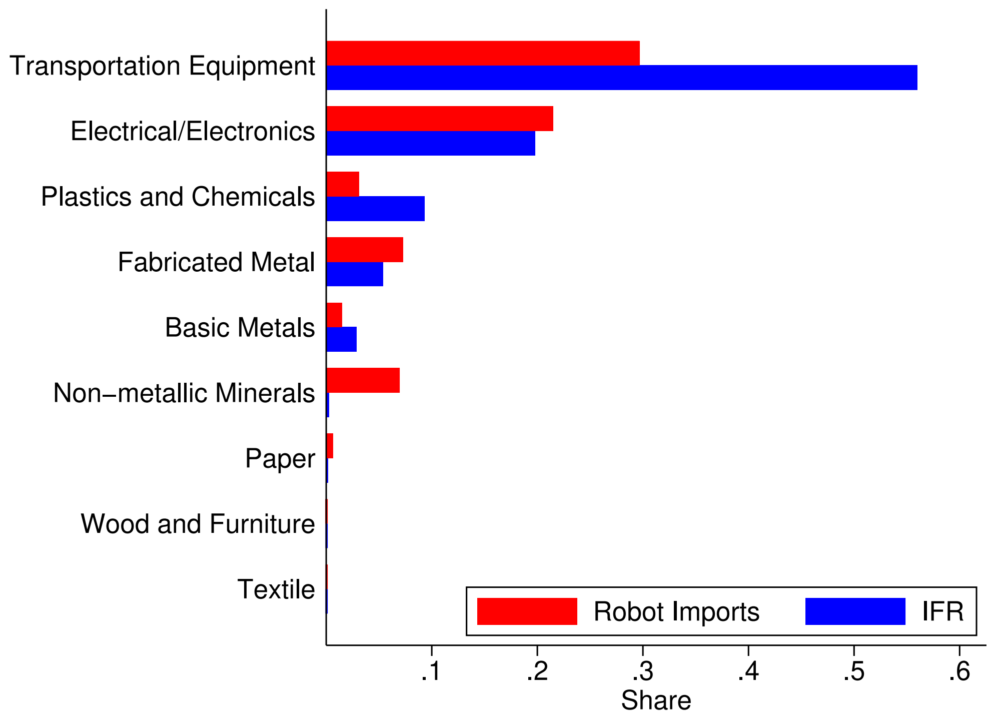
Notes: This table contains descriptive statistics for all robot adoption events. Each robot adoption event corresponds to one robot-adopting firm. For each variable, I show the mean and the standard deviation (SD) across all observations. For a robot-adopting firm, each variable is defined as follows. Robot value is the total value of robots imported by a firm; robot quantity is the total quantity of robots imported by a firm; number of years indicates the number of years in which a firm imports robots; number of countries indicates the number of countries from which a firm imports robots; peak year value share is the value share of robots purchased by a firm in the "peak year", defined as the year with the largest investment in robots. All firm counts are rounded for disclosure avoidance.

ure A.1 displays the share of robots purchased in every manufacturing industry[8], according to my methodology using robot imports and according to publicly available data provided by the IFR[9]. The IFR data indicates a much higher share of industrial robots used by the transportation equipment industry, whereas import data point to a non-negligible share of robots being used by the the non-metallic minerals industry. However, in general, the ranking of industries by robot use is consistent between the two data sources. Both datasets indicate that the transportation equipment industry and the electrical/electronics industry are by far the top two robot-using industries and account for the majority of industrial robots used by US manufacturing firms.

[8]Industries are defined in a way that is orughly consistent with 3-digit NAICS industries.

[9]For robot imports, I calculate the total number of robots purchased by firms in each manufacturing industry between 1992 and 2016. The IFR data does not provide industry breakdown before 2004, so I take the number of robots in the operational stock in 2016.

Figure 1.2: *Share of industrial robots used, by industry*



Notes: This figure plots the share of all robot-using industries among all manufacturing industries in terms of the number of robots used. For robot imports, I calculate the total number of robots imported by firms in each industry between 1992 and 2016, restricting to robot-adopting firms. For IFR, I calculate the total number of robots operating in each industry as recorded by IFR data.

1.3.2 Cross-sectional comparison

A central question related to robot adoption is: how are robot adopters and non-adopters different? I start answering this question by gauging the cross-sectional differences between robot adopters and non-adopters through the following specification:

$$Y_{it} = \alpha + \beta \text{Robot Adopter}_{it} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \varepsilon_{it}, \tag{1.1}$$

where Y_{it} is the dependent variable for firm i in year t, and $\gamma_{sc}, \gamma_{st}, \gamma_{ct}$ are industry-CZ, industry-year, and CZ-year fixed effects. Since some firms operate in multiple commuting zones and/or multiple industries, I take the industry and CZ, respectively, in which the firm has the highest employment in 1992. Fixed effects are included to control for industry and CZ-specific trends and characteristics. The coefficient β thus captures the premium of robot adopters over non-adopters in variable Y.

Table 1.2: *Premium of Robot Adopters*

	(1) Log(Sales)	(2) Log(Employment)	(3) Log(Wage)	(4) Log(Labor Prod.)
Robot Adoption	2.871***	2.316***	0.222***	0.144**
	(0.255)	(0.231)	(0.0354)	(0.0588)
Industry-Year FEs	✓	✓	✓	✓
CZ-Year FEs	✓	✓	✓	✓
Industry-CX FEs	✓	✓	✓	✓
Observations	770000	770000	770000	770000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of cross-sectional regressions following the specification in Equation 2.11. The sample used is all census-year observations between 1992 and 2012 for the universe of US manufacturing firms. All regressions include industry-year, CZ-year, and industry-CZ fixed effects. All observation counts are rounded for disclosure avoidance.

Results under Equation 2.11 are displayed in Table 2.1. Columns (1) and (2) show that robot adopters are significantly larger than non-adopters: they are nearly four times as large in sales, and more than three times as large in employment. Robot adopters also tend to pay higher wages on average, and are more productive in terms of labor productivity, as confirmed by columns (3) and (4). Therefore, it is clear that at a cross-sectional level, robot adoption is associated with a premium in size, wage, and productivity.

However, an immediate question that follows is whether the premium of robot adopters documented in Table 2.1 demonstrates meaningful firm-level changes as a consequence of robot adoption, or simply reflects ex-ante differences between adopters and non-adopters. Could adopters already be different from non-adopters *prior to* adopting robots? To shed light on this question, I restrict the sample to firms in 1992 that have not yet adopted robots, and run the following selection regression:

$$Y_{i,1992} = \alpha + \beta \text{Future Adopter}_{i,1992} + \gamma_{sc} + \varepsilon_i, \tag{1.2}$$

where Future $\text{Adopter}_{i,1992}$ is equal to one if firm i has not yet adopted robots in 1992, but ends up adopting robots in a subsequent year. The coefficient β thus captures the *ex-ante* difference between adopters and non-adopters in firm-level variable Y_i.

Table 1.3: *Premium of Future Adopters*

	(1) Log(Sales)	(2) Log(Employment)	(3) Log(Wage)	(4) Log(Labor Prod.)
Future Adoption	2.807^{***}	2.270^{***}	0.219^{***}	0.157^{**}
	(0.246)	(0.204)	(0.0372)	(0.0617)
Industry-Year FEs	✓	✓	✓	✓
CZ-Year FEs	✓	✓	✓	✓
Industry-CZ FEs	✓	✓	✓	✓
Observations	770000	770000	770000	770000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of cross-sectional regressions following the specification in Equation 2.12. The sample used is the universe of manufacturing firms in 1992 that have not adopted robots. All regressions include industry-year, CZ-year, and industry-CZ fixed effects. All observation counts are rounded for disclosure avoidance.

Table 2.2 reveals that, even before robot adoption happens, future adopters already exhibit advantages in various dimensions over non-adopters. Future adopters are larger, employ more workers, pay higher wages on average, and display higher productivity. The degree of advantage is similar across the two sets of regressions. These results suggest that larger and more productive firms select into robot adoption. In other words, robot adoption is not a random event: it is likely an active choice made by firms, and firms near the right end of the size distribution are more likely to take advantage of the opportunity.

Table 1.4: *Cross-sectional Regressions, Labor Share Variables*

Panel A: Premium Regressions				
	(1)	(2)	(3)	(4)
	Labor Share of Sales	Manuf Share	Service Share	MPRO Share
Robot Adoption	0.195***	-0.0480***	0.0314***	0.0290***
	(0.0344)	(0.0333)	(0.0118)	(0.00985)
Log(Sales)	-0.120***	-0.0333***	0.0132***	0.00985***
	(0.0232)	(0.00194)	(0.00123)	(0.000793)
Industry-Year FEs	✓	✓	✓	✓
CZ-Year FEs	✓	✓	✓	✓
Industry-CX FEs	✓	✓	✓	✓
Observations	770000	770000	770000	770000
Panel B: Selection Regressions				
	(1)	(2)	(3)	(4)
	Labor Share of Sales	Manuf Share	Service Share	MPRO Share
Future Adoption	0.207***	-0.0445***	0.0292**	0.0350***
	(0.0415)	(0.0196)	(0.0120)	(0.0115)
Log(Sales)	-0.106***	-0.0324***	0.0127***	0.000912***
	(0.0149)	(0.00198)	(0.00131)	(0.000733)
Industry-Year FEs	✓	✓	✓	✓
CZ-Year FEs	✓	✓	✓	✓
Industry-CZ FEs	✓	✓	✓	✓
Observations	770000	770000	770000	770000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of cross-sectional regressions following the specification in Equation 2.11. The sample used is all census-year observations between 1992 and 2012 for the universe of US manufacturing firms. All regressions include industry-year, CZ-year, and industry-CZ fixed effects. All observation counts are rounded for disclosure avoidance.

Recently, there has emerged a debate on whether the rise of industrial robots (and more broadly, automation) displaces certain workers and threatens the future of work. We can further examine the relationship between robots and labor market outcomes at the firm level by looking at a few additional variables. In Table 1.4, I run the specifications in Equation 2.11 and Table 2.2 for several variables related to the labor share as well as within-firm worker composition. Panel A and panel B presents results from premium regressions and selection regressions, respectively.

First, I examine cross-sectional differences in the labor share of sales across adopters and non-adopters. One common view argues that robots displace workers, and thus robot-adopting firms should have lower labor share. However, column (1) of Panel A suggests that, after controlling for sales, robot adopters in fact have a higher labor share compared to non-adopters. Column (1) of Panel B further reveals that this difference is driven by a selection mechanism: firms that end up adopting robots in the future tend to have significantly higher labor share prior to adoption. One potential explanation for this phenomenon is that, holding size constant, firms whose current mode of production is more labor-intensive are more likely to select into robot adoption because it is more lucrative for them to ramp up automation.

Second, in order to investigate whether robot adopters hire different *types* of workers than non-adopters, I also analyze the share of manufacturing, service, and MPRO (management, professional, and technical) workers in total labor compensation. Column (2) shows that robot adopters spend less on manufacturing workers, both before and after robot adoption. Columns (3) and (4) show that this difference is in part due to robot adopters spending more on service workers, and in particular, workers engaged in professional, management, and technical services. In other words, robot-adopting manufacturing firms choose a different mode of production than non-adopting ones: they tend to invest less in manufacturing workers and more in non-manufacturing workers. In that sense, robot adoption may accelerate the structural change within US manufacturing firms as documented in Ding *et al.* (2020).

1.3.3 Event Studies

Results in Table 2.1 and Table 2.2 demonstrate that robot adopters are larger and more productive than non-adopters, but these advantages already exist prior to adoption. Therefore, we cannot yet associate robot adoption with a positive effect on firm-level outcomes. The natural starting point to evaluate whether robots have a meaningful effect on firm-level variables are panel regressions with firm fixed effects. It would be reasonable to run the following baseline specification:

$$Y_{it} = \alpha + \beta \text{Robot Adopter}_{it} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \gamma_{qt} + \gamma_i + \varepsilon_{it}. \tag{1.3}$$

Selection into robot adoption implies that robot adoption is correlated with other firm characteristics (e.g. productivity) that potentially affect firm-level outcomes but cannot be completely observed. This presents a major barrier to identifying the causal effect of robots. Therefore, relative to the cross-sectional specification in Equation 2.11, I introduce firm fixed effects denoted by γ_i to control for (time-invariant) firm-specific characteristics. In addition, I include quantile-year fixed effects γ_{qt}, where quantile of a firm is defined as its employment quartile in the initial year 1992. This fixed effect is intended to control for heterogeneous trends across firms with different initial firm size, thereby capturing time-varying fixed effects that are correlated with robot adoption but not directly caused by robots. Hereafter, I refer to these fixed effects as "baseline trend controls". I restrict the sample to firms that already existed in 1992, i.e. firms for which baseline trend controls are defined.

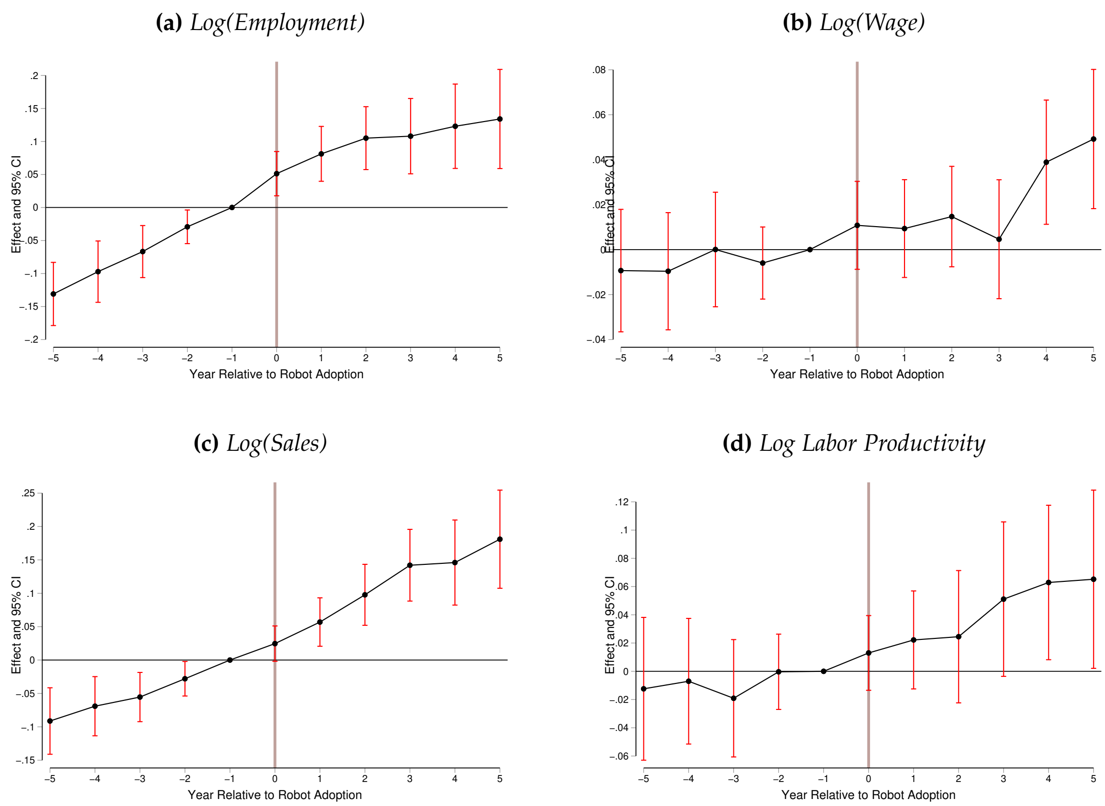
To more carefully evaluate how firm-level outcomes change both before and after robot adoption, I run a set of event study OLS regressions using the following distributed lead-lag model (Stock and Watson, 2015):

$$Y_{it} = \alpha + \sum_{k=-5}^{k=5} \delta_k \text{Robot Adopter}_{i,t-k} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \gamma_{qt} + \gamma_i + \varepsilon_{it}. \tag{1.4}$$

where Robot Adopter$_{i,t-k}$ indicates whether a firm has adopted robots in year $t-k$. Through the lead-lag coefficients, δ_k, this specification sheds light on the dynamic response of the firm-level outcome Y_{it} to robot adoption in the five years before and after adoption. I set year $t-1$ as the base level, so all coefficient estimates for δ_k should be interpreted relative to the year before robot adoption. In addition, I control for industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as quantile-year fixed effects. This exercise thus captures evolution of Y_{it} for a robot adopter over time, relative to firms in the same industry, commuting zone, and size quantile in the initial year 1992.

Results of the event study regressions are shown in Figure 2.1. The pre-adoption trends for wage and labor productivity are not distinguishable from zero. However, there are visible increasing pre-trends for employment and sales, suggesting that robot adopters are already expanding in those dimensions relative to non-adopters prior to adoption. Robot adoption is associated with an *ex-post* increase in all variables, though the increase in wage and labor productivity only becomes significant after four years.

Figure 1.3: *Event Study Regressions*



(a) *Log(Employment)* **(b)** *Log(Wage)* **(c)** *Log(Sales)* **(d)** *Log Labor Productivity*

Notes: This figures plots results of event study regressions under the specification in Equation 2.14 for four dependent variables: log(employment), log(wage), log(sales), and log of labor productivity, where labor productivity is defined as total sales divided by total employment. I set the baseline as the year prior to robot adoption, so all coefficients should be interpreted relative to the baseline year. For each regression, I show the point estimates for the 5 years before robot adoption and 5 years after robot adoption, as well as corresponding 95% confidence intervals depicted as error bars. All regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as baseline trend controls, Standard errors are clustered at the firm level.

There are two main takeaways from these results. First, robot adoption is associated with an increase in size and productivity, but in many cases, this increase occurs both before and after adoption. Therefore, it is premature to draw conclusions about the causal effect of robots on firm-level outcomes based on OLS regressions alone; alternative strategies are needed. Second, regarding the relationship between robots and workers, the event study regression for employment suggests that firms hire more workers after they adopt robots. Even if the pre-trend is taken into account, we do not observe a discernible reversal in trend around the time of robot adoption. In light of the ongoing discussion on whether robots (and more broadly, automation) would threaten the future of work, my findings provide little evidence in support of these concerns.

1.4 Robot-complementary workers and robot adoption

section 1.3 reveals that robot adoption is not a random event. subsection 2.4.1 shows that firms that are larger and more productive are more likely to adopt robots; subsection 2.4.2 indicates that robot adopters also grow faster than non-adopters prior to robot adoption. However, productivity is only one of the possible determinants of robot adoption. Are there other factors that determine which firms select into robot adoption? Beyond that, is it possible to construct or identify an exogenous shock to robot adoption? These questions are important for two reasons. First, given that robots potentially induce productivity gains at the firm level, policymakers may find it beneficial to incentivize robot adoption. Understanding the barriers to robot adoption and shocks that help firms overcome those barriers would provide valuable guidance for policies. Second, the endogeneity and pre-trend issues associated with OLS regressions means that an alternative empirical strategy is needed to evaluate the causal effect of robot adoption. One possibility is the instrumental variable strategy, and an exogenous shock to robot adoption can be potentially used for that purpose.

In this section, I investigate one potential factor that affects the process of robot adoption: the availability of robot-complementary workers. Although robots are by definition autonomous, certain human workers are still indispensable to the processes of installing and integrating robots. For example, a recent survey of the robotics industry has revealed that lack of integrators, lack of experience with automation, and cost of training [workers] are among the top challenges faced by potential users of industrial robots (McKinsey (2018)). Conversations with an expert in the robotics industry also underline the importance of human workers in robot adoption[10]. These anecdotal evidence suggests that while robots displace workers in certain occupations, they may well be complementary with other occupations. An increase in the availability of such robot-complementary workers may thus incentivize adoption among local firms.

Based on this idea, in subsection 1.4.1, I classify a list of robot-complementary occupations and develop a measure of a plausibly exogenous shock to the labor supply of workers in those occupations at the US commuting zone level. Specifically, I rely on the tendency of immigrants to settle into existing enclaves to construct a shock based on immigrant inflows of robot-complementary workers. In subsection 1.4.2, I show that the constructed shock predicts robot adoption at the firm level, and use a series of robustness checks and falsification exercises to support its exogeneity. I also use the shock as an instrument for robot adoption to estimate causal effects of robot adoption on labor market outcomes at the firm level.

1.4.1 Methodology

I identify robot-complementary occupations using the O*NET database, which is a regularly updated database maintained by the U.S. Department of Labor, containing detailed information on characteristics of occupations across the United States. The database provides an online tool, the O*NET connector, that outputs a list of matched occupations along with match scores out of 100 for a given set of keywords. The tool uses a textual algorithm that searches for the inputted keywords within the database of occupational titles, descriptions,

[10]I thank Jeff Burnstein, President at the Association for Advancing Automation, for his insights into the robotics industry.

tasks, and detailed work activities.

Using the tool, I determine occupations that match the keyword "robots". The list of occupations and their corresponding match scores are displayed in Table A.1. The match score indicates the extent to which the occupation is related to the keyword "robots". As the scores suggest, the occupations that are the most closely related to robots are, unsurprisingly, robotics technicians and robotics engineers. The top end of the list also includes other reasonable occupations, such as CNC tool operators and welding, soldering, and brazing machine operators. The list does contain several occupations that are to some extent peripheral to the process of robot adoption for manufacturing firms, e.g., biological technicians and construction laborers. However, note that those occupations generally

Table 1.5: *List of robot-complementary occupations identified by the O*NET connector*

Occupation Name	Match Score
Robotics Technicians	100
Robotics Engineers	100
Computer Numerically Controlled Tool Operators	65
Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	60
Electro-Mechanical and Mechatronics Technologists and Technicians	58
Electrical and Electronics Repairers, Commercial and Industrial Equipment	25
Commercial and Industrial Designers	25
Welders, Cutters, Solderers, and Brazers	25
Multiple Machine Tool Setters, Operators, and Tenders, Metal and Plastic	12
Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	7
Computer and Information Research Scientists	3
Mechanical Engineers	3
Biological Technicians	3
Cytogenetic Technologists	3
Pharmacy Technicians	3
Construction Laborers	3
Millwrights	3
Electromechanical Equipment Assemblers	3

I calculate the labor supply of robot-complementary workers using publicly available samples of the Decennial Censuses and American Community Surveys, provided by IPUMS (Ruggles *et al.*, 2020). All datasets used are random samples of the corresponding Census and ACS datasets. In order to reflect the relative relevance of different occupations for robot adoption, to each occupation I assign a weight that is proportional to the match score described above. Specifically, the weight is given by the match score divided by 100. Therefore, one robotics engineer in the sample is treated as one robot complementary worker, whereas one mechanical engineer is treated as 0.03 robot complementary workers. Following David *et al.* (2013), I use US commuting zones (Tolbert and Sizer, 1996) as the geographical unit of analysis for local labor markets.

While change in the labor supply of relevant workers may indeed have predictive power for robot adoption among local firms, it is not necessarily exogenous. For example, this change might be correlated with other regional economic trends due to policies and industry composition. Reverse causality may also be in play: robotics engineers may choose to relocate to certain areas precisely *in response to* bursts in robot adoption. Therefore, for our purpose, it is necessary to take advantage of exogenous variation in local labor supply that affects the availability of robot-complementary workers.

Based on this idea, I proxy for the increase in availability of robot-complementary workers using immigrant inflows. Specifically, I construct a shift-share instrument in the style of Altonji and Card (1991)[11] to capture exogenous variation in immigrant labor supply of robot-complementary workers. For each commuting zone c and year t, I define an immigrant shock $z_{c,t}$ as

$$z_{c,t} = \sum_{g} \underbrace{\Delta L^{g}_{1990,t}}_{\text{Shift}} \cdot \underbrace{\frac{L^{g}_{c,1990} - R^{g}_{c,1990}}{L^{g}_{1990} - R^{g}_{1990}}}_{\text{Share 1}} \cdot \underbrace{\frac{R^{g}_{1990} - R^{g}_{c,1990}}{L^{g}_{1990} - L^{g}_{c,1990}}}_{\text{Share 2}} \cdot \underbrace{\frac{1}{R_{c,1990}}}_{\text{Share 3}}, \tag{1.5}$$

where L denotes all workers and R denotes robot-complementary workers. This shock is the sum of a shift-share variable over country origin groups indexed by g. For each country

[11] Other papers that have constructed immigration shocks in a similar vein include Lewis (2011) and Hanson (2021).

origin group g, the individual components of the shift-share variable are defined as follows. The "shift" component represents the total number of immigrants from country origin group g that arrived between 1990 and year t. Share 1 represents the share of immigrants from g living in commuting zone c in year 1990, excluding robot-complementary workers in the calculation. Multiplying the shift component by this share apportions immigrants that arrived after 1990 to different commuting zones based on initial settlement patterns of immigrants from group g in 1990. Using past rather than realized settlement patterns makes sure that the constructed immigrant inflows are driven by generic attraction of each commuting zone to immigrants from g, not post-1990 regional trends that affect outcomes of local firms through channels other than robot adoption. By leaving out robot-complementary workers, I avoid picking up attractiveness of commuting zones specifically to those workers.

Share 2 represents the share of immigrant workers from g who work in robot complementary occupations in 1990. By excluding workers in commuting zone c from the calculation, I eliminate any effects resulting from specific occupational composition of commuting zones. Multiplying together the shift component, share 1, and share 2 delivers the imputed number of robot-complementary workers from country origin group g that arrived in commuting zone c between year 1990 and year t, based on patterns of geographical settlement and occupation composition in the baseline year, 1990. Finally, multiplying by share 3 normalizes the variable by the total number of robot-complementary workers in commuting zone c in 1990, reflecting the fact that the size of gains from an increase in the labor supply of robot-complementary workers depend on the number of such workers already residing in each commuting zone.

Finally, I compute firm-level exposure to these immigrant shocks. For each firm i, I aggregate the commuting zone-year level variable $z_{c,t}$ over the commuting zones in which the firm operates in 1992, $\mathcal{C}^i_{1992}$, to obtain the firm-level shock Z^i_t defined as:

$$Z^i_t = \log\left(\sum_{c\in\mathcal{C}^i_{1992}} w^i_{c,1992} z_{c,t-1}\right). \tag{1.6}$$

I lag the immigrant shock $z_{c,t}$ defined in Equation 2.15 by one year to allow time for the increase in labor supply to have an effect on robot adoption. When calculating the weighted sum of $z_{c,t}$ over commuting zones in $\mathcal{C}^i_{1992}$, I define the weights $w^i_{c,1992}$ to be the employment share of firm i in commuting zone c. The implicit assumption behind this definition is that the effect of local labor supply in commuting zone c on robot adoption is proportional to the share of firm i's activity in c. Lastly, I apply the logarithmic transformation to correct right skewness in the constructed weighted sum. In light of my regression analysis, If I control for time-varying fixed effects associated with each firm's predominant commuting zone, predictive power of the instrument relies on secondary commuting zones in which a firm operates.

For the constructed shock to be exogenous, it should be correlated with firm-level variables only through the channel of robot adoption. To the extent that full changes in the labor supply of robot-complementary workers are potentially endogenous, immigrant shocks are more likely to be exogenous due to a few reasons. First, while full changes might be prone to reverse causality, i.e. robot adopters hiring robotics engineers, etc., from other areas, immigrant shocks are not prone to this issue because they are based on immigration patterns in an initial year, 1990. Second, whereas full changes may be correlated with or even driven by changes in local economic conditions, my immigrant shock does not capture a significant part of those changes, since the constructed shock (1) builds upon the tendency of immigrants to be close with others from similar countries of origin rather than economic conditions of local labor markets, and (2) isolates shocks to a very specific slice of local labor supply. In the next section, I present a more detailed discussion of the potential endogeneity issues associated with the shock, and conduct robustness checks addressing those issues.

1.4.2 Results

I first test whether the constructed shock indeed has an effect on the probability of robot adoption. Specifically, I run the following specification:

$$\text{Robot Adopter}_{it} = \alpha + \beta \text{Shock}_{it} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \gamma_{qt} + \gamma_i + \varepsilon_{it}, \tag{1.7}$$

where Robot Adopter_{it} is a binary variable indicating whether firm i has already adopted robots in year t, and Shock_{it} is the firm-level immigrant shock Z_t^i defined in the previous section. Following the baseline specification in Equation 2.13, I control for a full set of fixed effects to isolate the effect of the shock on robot adoption.

Results under this specification are displayed in Table 1.6. The constructed shock is a strong predictor of robot adoption at the firm-level, regardless of how restrictive the fixed effects are. In column (1), when all relevant fixed effects are included, the estimated coefficient on the shock is still statistically significant at the 0.1% level with an F-statistic of 35.34. The estimated coefficient implies that a 10% larger shock to local robot-complementary workers due to immigrant inflows is associated with a 0.7% increase in the probability of robot adoption. Comparing column (1) with columns (2) through (4), while the F-statistic is consistently larger when fewer fixed effects are included, the magnitude of the coefficients is similar across different specifications. This indicates that controlling for additional factors – industry-CZ fixed effects and baseline trend controls – does take away some explanatory power of the constructed shock, but does not affect its predictive power in a significant way.

The constructed shock can be potentially used as a proxy for robot adoption in a 2SLS specification. In that case, this specification in Equation 1.7 can be interpreted as the first-stage regression of a 2SLS model. One key criterion for an ideal instrument is relevance. Given that standard errors are clustered, the appropriate indicator of relevance in this case would be the Kleibergen-Paap F-statistic, which is equal to the F-statistic of the first stage regressions in my case since there is only a single endogenous variable (robot adoption) (Kleibergen and Paap (2006)). The F-statistic is greater than a standard threshold of 10, which suggests that the constructed instrument is relevant as a proxy for robot adoption.

Table 1.6: *Immigrant Shock and Robot Adoption*

	Dependent Variable: Robot Adoption			
	(1)	(2)	(3)	(4)
Immigrant Shock	0.0742***	0.0845***	0.0848***	0.0734***
	(0.0125)	(0.0118)	(0.0123	(0.0118)
Industry-Year FEs	✓	✓	✓	✓
CZ-Year FEs	✓	✓	✓	✓
Firm FEs	✓	✓	✓	✓
Industry-CZ FEs	✓		✓	
Baseline Trend Controls	✓			✓
Observations	3360000	3360000	3360000	3360000
First-stage F-statistic	35.34	51.18	46.05	38.94

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of first-stage IV regressions, where the dependent variable is robot adoption, and the instrument is the constructed firm-level instrument discussed above. The sample used is all firm-year observations for US manufacturing firms between 1992 and 2016. Regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, and baseline trend controls as indicated by the checkmarks. Standard errors are clustered at the firm level. All observation counts are rounded for disclosure avoidance.

An ideal instrumental variable should also satisfy the exclusion restriction, i.e. it should affect dependent variables in the second stage only through the endogenous variable. In our context, the constructed shock should affect firm-level outcomes only through robot adoption. In the previous section, I discussed why the constructed immigrant shock is less prone to endogeneity issues than full changes in the labor supply of robot-complementary workers. However, even the constructed shock is potentially subject to concerns surrounding the exclusion restriction. To address those concerns, I implement a series of exercises detailed below.

To start off, one might be concerned that an increase in labor supply of robot-complementary workers may help firms adopt other types of non-robot machinery. If that were the case, then the estimated effects of robot adoption using my proposed IV strategy would be confounded by effects of the firm adopting other types of machinery.

I address this concern in a robustness check displayed in column (1) of Table 1.7.

Specifically, I identify three types of machinery, corresponding to three 6-digit HS codes[12], that performs similar tasks as robots but are not classified as industrial robots. I construct a binary variable corresponding to whether a firm has adopted machinery in the reference group, and regress that variable on the immigrant shock in a specification analogous to Equation 1.7. The fact that the shock does not predict adoption of machinery in this reference group at a statistically significant level suggests that we have managed to capture changes in the supply of local workers who are specifically complementary with robots, but not machinery in general.

More broadly, if the shock was correlated with other changes in the local workforce or economic conditions, and these changes affect firm-level outcomes through channels other than robot adoption, then the exclusion restriction would be violated. To alleviate those concerns, I conduct falsification tests where I directly regress firm-level outcomes on the shock, and allow the coefficient to differ between firms that adopt robots at some point between 1992 and 2016 and firms that never do. If the shock satisfies the exclusion restriction, then it should have no effects on non-robot-adopters. In columns (2) and (3), I confirm that the immigrant shock is associated with an increase in sales and employment only for robot-adopting firms. These results are consistent with the assumption that the shock affects firm-level outcomes only through the channel of robot adoption.

In Table 1.8, I use the constructed shock to study the firm-level impact of robot adoption on labor market outcomes. Specifically, I use the immigrant shock as an instrument for robot adoption in a 2SLS setting. In columns (1) and (2), the dependent variable is the log of employment and wage of workers employed by a firm across *all* sectors. The elasticity of overall employment and wage to robot adoption is positive, with the latter being statistically significant. However, when we restrict attention to workers in manufacturing plants only, the response of employment and wage is negative (though imprecisely estimated). In other

[12] The codes are 851521 (Machines and apparatus for resistance welding of metal: Fully or partly automatic), 847710 (injection-molding machines), and 851529 (Electric Machines And Apparatus For Resistance Welding Of Metal, Other Than Fully Or Partly Automatic).

Table 1.7: *Robustness checks: testing the exclusion restriction*

	(1) Reference Adoption	(2) Sales	(3) Employment
IV	0.0254	-0.137	-0.0348
	(0.0524)	(0.12)	(0.0861)
IV × Robot Ever		0.207***	0.270***
		(0.0159)	(0.0156)
Full FEs	✓	✓	✓
Trend Controls	✓	✓	✓
Observations	3360000	1540000	3360000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of robustness checks aimed at testing the exclusion restriction associated with the IV strategy used in Table 2.6, Table 2.5, and Table 1.8. In column (1), I regress a "reference adoption" variable on the instrument, Reference adoption is defined as whether a firm has adopted machinery in a reference group, consisting of three types of machinery performing similar tasks as robots (but are not identified as robots). More details on how this variable is constructed can be found in the appendix. In columns (2) through (3), I regress log(sales) and log(employment) on the constructed instrument, and the instrument interacted with a variable "robot ever", defined as a firm-level indicator of whether the firm adopts robots at any point between 1992 and 2016. The sample used is all firm-year observations for US manufacturing firms between 1992 and 2016. All regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as baseline trend controls, Standard errors are clustered at the firm level. All observation counts are rounded for disclosure avoidance.

words, robot adoption leads to an increase in overall employment at the firm level, but only favoring non-manufacturing workers.

These results are largely consistent with existing findings on the impact of automation on firm-level labor market outcomes in other advanced economies. For example, Humlum (2019) finds that adoption of industrial robots in Denmark has a positive effect on firm wage bill, but the effect is negative for production workers; Aghion *et al.* (2020) finds a positive effect of automation on firm employment in France; Koch *et al.* (2019) finds that robots lead to an increase in firm-level employment, and a decrease in the share of manufacturing employment. All of these findings point in the same general direction as Table 1.8: robots have a positive effect on overall employment at the firm level, but not necessarily for manufacturing workers. Unfortunately, the dataset does not provide firm-specific information on occupations of the workers, so it is not possible to further distinguish between the effects on, e.g., low and high skill workers within manufacturing.

Table 1.8: *IV Regressions for Labor Market Outcomes, Logged*

	(1) Employment	(2) Wage	(3) Manuf. Emp.	(4) Manuf. Wage
Panel A: IV Regressions				
Robot Adoption	0.738	1.885***	−1.571	−0.774
	(0.937)	(0.539)	(1.132)	(1.159)
Full FEs	✓	✓	✓	✓
Baseline Trend Controls	✓	✓	✓	✓
Observations	3360000	3360000	3360000	3360000
First-stage F-stat	35.34	35.34	35.34	35.34
Panel B: OLS Regressions				
	(1) Employment	(2) Wage	(3) Manuf. Emp.	(4) Manuf. Wage
Robot Adoption	0.360***	0.0718***	0.319***	0.0506***
	(0.0263)	(0.0105)	(0.0293)	(0.0106)
Full FEs	✓	✓	✓	✓
Baseline Trend Controls	✓	✓	✓	✓
Observations	3360000	3360000	3360000	3360000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Panel A of this table displays results of IV regressions at the firm level, where the dependent variables are logged labor market outcomes, and robot adoption is proxied by the firm-level instrument. Panel B displays results of OLS regressions for the same specifications. Employment and wage indicate corresponding variables across all industries a firm operates in; manufacturing employment and manufacturing wage indicate the same variables within manufacturing plants only. The sample used is all firm-year observations for US manufacturing firms between 1992 and 2016. All regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as baseline trend controls, Standard errors are clustered at the firm level. Observation counts are rounded for disclosure avoidance. "First-stage F-statistic" indicates the Kleibergen-Paap F-statistic for the IV regressions, which conincide with the F-statistic of the first-stage regression under the case of clustered standard errors and a single endogenous variable. All observation counts are rounded for disclosure avoidance.

1.5 Conclusion

In this paper, I construct a novel dataset documenting robot adoption among US manufacturing firms. Given the lack of data sources directly accomplishing this task, I rely on customs data and use robot imports to measure robot adoption. Using this methodology, I identify 1600 robot-adopting manufacturing firms in the US and capture the majority of robots used by US manufacturing firms between 1992 and 2016. The constructed dataset allows me to present a series of descriptive facts and patterns. At a macro level, the fraction of robot-adopting firms in the US grew by more than ten-fold between 1992 and 2012, and by 2012, the 0.5% of robot adopters account for around 10% of total sales and employment in the manufacturing sector. At a micro level, robot adoption appears to be a largely one-time investment that is relatively costly. Cross-sectional regressions reveal that robot adopters are larger and more productive than non-adopters, and tend to pay higher wages and spend proportionally more on service workers rather than manufacturing workers. According to event study exercises, robot adoption is associated with an increase in sales, employment, and productivity, both before and after robot adoption. Finally, I construct a plausibly exogenous shock to robot adoption based on immigrant inflows of robot-complementary workers. I show that this shock is a statistically significant predictor of robot adoption at the firm level. I also provide evidence that it satisfies the exclusion restriction, and thus can be used as an instrument for robot adoption in a 2SLS setting. Applying this methodology to labor market outcomes, I find that robot adoption triggers within-firm reallocation of wage bill from manufacturing workers to non-manufacturing workers.

My results point to several interesting features of robot adoption and robot adopting-firms. First, investment in industrial robots is lumpy and costly. This suggests that it is reasonable to think of robot adoption as a binary choice involving a significant fixed cost, both empirically and theoretically. Second, robot adoption exhibits selection on size and productivity. This implies that robot adoption is not a random event and that larger and more productive firms are more likely to reap the benefits from robots. Lastly, robot adoption seems to trigger reallocation of labor expenditure from manufacturing workers to non-manufacturing workers. This reallocation happens through two channels. Most importantly, the IV regressions indicate that robot-adopting firms decrease their share of wage bill devoted to manufacturing workers. In addition, this reallocation is amplified by a composition effect, in the sense that robot adopters already spend less on manufacturing workers prior to adoption and robot adoption expands their share in the economy.

The dataset constructed in this paper can be used to answer a wide range of questions related to industrial robots. For example, further analysis of robot adoption across manufacturing and non-manufacturing industries may reveal new information on how the use of robots has evolved across different areas of the economy. One can also combine data on robots with input-output information to understand how the rise of robots differently affect firms along the supply chain. In Wang (2022b), I use the dataset and the constructed immigrant shock to evaluate the impact of robot adoption on trade and offshoring among US manufacturing firms.

Although this paper is specific to the US, it is worth emphasizing that the techniques used for data construction have broader relevance. In particular, customs data covering the universe of firms are widely available in many countries, whereas few countries, if any, provide data documenting robot use with the same level of coverage. The methodology introduced in this paper can thus be applied to most countries, as long as the country is not itself a major producer of industrial robots.

Chapter 2

Robots, Trade, and Offshoring

2.1 Introduction

How do automation technologies shape the global organization of production? Recent work has uncovered labor-displacing effects of technologies including routine-based technical change (Autor and Dorn, 2013; Goos *et al.*, 2014), robotics (Acemoglu and Restrepo, 2020), and artificial intelligence (Webb, 2019; Acemoglu *et al.*, 2020a). The negative labor-market effects of automation have prompted a debate on whether automation may serve as an alternative to offshoring, and thus stifle imports and induce reshoring. Given rapid developments in automation technologies, some commentators have cast serious doubt on the future of global trade (Lewis, 2014; Rodrik, 2018), while others have remained optimistic[1] (Antràs, 2020). Empirical work on the topic delivers divergent messages as well. Some have documented automation in advanced economies substituting for imports from developing countries (Artuc *et al.*, 2019; Faber, 2020; Kugler *et al.*, 2020); on the other hand, recent evidence also points to trade-promoting effects of automation (World Bank, 2020).

[1] In addition to automation technologies, a large body of recent work has also examined potential deglobalization effects of other factors, including protectionist policies (Amiti *et al.*, 2019; Fajgelbaum *et al.*, 2020; Flaaen *et al.*, 2020; Handley *et al.*, 2020), slowdown in the expansion of global value chains (World Bank, 2020), and very recently, the pandemic (Baldwin and Evenett, 2020).

In this paper, I contribute to the debate on automation and offshoring by evaluating the economic impact of a specific automation technology: industrial robotics. Defined as "automatically controlled, reprogrammable, and multipurpose [machines]", industrial robots have played an increasingly important role in manufacturing around the world since the 1990s. In 2020, 2.7 million industrial robots operate in factories around the world (International Federation of Robotics, 2020), and this number is projected to more than double by 2030 (McKinsey, 2018). Therefore, understanding the impact of industrial robots is important for economic outlook and policy design in an ever-changing world of technology.

This paper provides theory and evidence on the impact of industrial robots on US manufacturing firms, with a focus on trade and offshoring. I start by building a flexible theoretical framework of heterogeneous firms in the style of Melitz (2003) to understand interactions between robot adoption and offshoring at the firm level. Production of final goods in the model consists of two stages: input sourcing and assembly. Similar to the canonical models of importing (Gopinath and Neiman, 2014; Halpern *et al.*, 2015; Antras *et al.*, 2017), I assume that firms may import intermediate inputs from abroad. In addition, firms may also offshore assembly and import produced outputs. This two-stage setup is motivated by two observations. First, 84% of industrial robots in the world are used for tasks predominantly involved in the assembly process[2] (IFR, 2018). Capturing this comparative advantage of robots in downstream production requires the model to incorporate multiple production stages. Second, recent work on global value chains (Baldwin and Venables, 2013; Antràs and De Gortari, 2020) points to the significance of offshoring in assembly, which is distinct from input offshoring. My model explicitly distinguishes between these two types of offshoring, and thus allows robot adoption to have heterogeneous effects on upstream imports (imports of intermediate inputs) and downstream imports (imports of outputs assembled abroad).

[2]Specifically, data provided by the International Federation of Robotics (IFR) shows that 84% of industrial robots around the world in 2016 are used for three categories of applications: assembling and disassembling, welding and soldering, and material handling. The tasks associated with these applications are primarily relevant to the assembly process.

In the model, robot adoption is modeled as a binary choice, where firms pay a single fixed cost to adopt robots in both stages of production. The firm-level impact of robots on imports consists of two conflicting effects: on the one hand, robot adoption is associated with a positive productivity effect that induces firms to expand scale of production and thus increase their imports. On the other hand, robots substitute for imports in both upstream and downstream production. Allowing firms to adjust their extensive margin of imports further complicates the interaction between robots and offshoring. To highlight key mechanisms, I focus on a benchmark version of the model that underscores complementarities between robots and imports driven by the productivity effect. The benchmark model assumes that robots are only useful in assembly, and eliminates the extensive margin for downstream imports. It delivers three empirically testable predictions: first, there is selection on productivity into robot adoption; second, robot adoption has a positive effect on sales and both upstream and downstream imports at the firm level; third, this effect is larger for upstream imports than for downstream imports.

I test these predictions using highly disaggregated US Census data on manufacturing firms containing firm-level information on sales, employment, and trade from 1992 to 2016. I construct an empirical classification of upstream and downstream imports based on firm-specific input-output information, using a methodology inspired by Feenstra and Hanson (1999) and Bernard *et al.* (2020). Taking advantage of the fact that most industrial robots in the world are produced outside of the US, I measure robot adoption using robot imports, following the methodology documented in Wang (2022a).

Cross-sectional regressions confirm that robot adopters are larger, more productive, and import more, and these advantages already exist prior to robot adoption, confirming that a selection mechanism is indeed at work. However, ex-ante differences between adopters and non-adopters raise concerns of endogeneity associated with robot adoption. These concerns are confirmed by event study regressions, which show that pre-trends are indeed present in import measures. It follows that estimating the causal impact of robots on firm-level outcomes requires an alternative methodology to the OLS, such as an instrumental variables strategy.

A wealth of recent empirical studies has utilized regional variation to identify the impact of import and technology shocks (Autor and Dorn, 2013; Acemoglu and Restrepo, 2020). Following this general idea, I develop an IV strategy that exploits variation in labor supply across US commuting zones (Tolbert and Sizer, 1996), hereafter abbreviated as CZ. In contrast with a growing literature seeking to classify occupations that are particularly *vulnerable* to being displaced by automation technologies (Webb, 2019; Acemoglu *et al.*, 2020a), my instrument is based on the novel idea that certain occupations are *helpful* in the process of installing, integrating, and working with technologies. Following Altonji and Card (1991) and my earlier work in Wang (2022a), I construct a shock based on immigrant inflows of robot-complementary workers, and use it as an instrument for robot adoption at the firm level.

Using this IV strategy, I run panel regressions to evaluate the impact of robots on firm-level outcomes. My regressions control for an extensive set of fixed effects, including a firm fixed effect, and thus isolate the effect of robot adoption *at the firm level*. In the first stage, the IV is a powerful predictor of robot adoption. In the second stage, I find that robot adoption leads to an increase in sales, labor productivity, and both upstream and downstream imports, with the effect being significantly larger for upstream imports than for downstream imports. These results highlight the robot productivity effect captured by the benchmark model, and are in line with the hypothesis that robots are disproportionately useful in assembly relative to input production.

Do these effects of robots at the firm level translate into similar implications for aggregate outcomes? To answer this question, I quantitatively estimate an extended version of the model under industry equilibrium, using cross-sectional data of US manufacturing firms in 2012. Relative to the benchmark model, the extended model assumes robots to be additionally useful in input production, and allows each firm to endogenously choose the number of imported outputs. I estimate the model using a simulated method of moments.

The calibrated model closely matches overall patterns of robot adoption and importing, as well as differences between robot adopters and non-adopters in sales, input expenditure, and imports.

Based on the equilibrium in 2012, I calibrate a (reverse) increase in robot productivity that can explain the observed increase in the share of robot-adopting firms between 1992 and 2012. Comparing equilibrium outcomes between 1992 and 2012, I first show that the model corroborates empirical findings for robot adopters at the *firm* level. Following the positive robot productivity shock, adopters increase their sales, employment, imports in both stages, as well as expenditure on domestic inputs, but decrease their spending on domestic assembly. Notably, consistent with reduced-form findings, the increase in input imports among adopters is more than three times as large as the increase in output imports.

More importantly, the estimated industry equilibrium model sheds light on how robots affect non-adopters and shape aggregate outcomes in the manufacturing sector. Expansion of robot adopters comes at the expense of non-adopters, which contract in all dimensions due to within-industry competition. In aggregate, the rise of robots between 1992 and 2012 is associated with a 20% increase in input imports, and a 4% *decrease* in output imports driven by non-adopters. Overall, these changes imply a 15% *increase* in total imports. In other words, instead of restraining imports, robots have promoted trade in the manufacturing sector. The combination of the rise of robots and the resulting expansion in offshoring contributes to a 3% decrease in consumer price index of the manufacturing sector.

Despite these efficiency gains, domestic workers have suffered: robots reduce aggregate expenditure on domestic inputs and domestic assembly by 9% and 23%, respectively. Since my industry equilibrium model holds the wage level constant, these results can be interpreted as a decrease in employment of the respective types of workers due to the rise of robots between 1992 and 2012. This suggests automation as a potentially powerful explanation for the decline in US manufacturing employment over the past two decades (Pierce and Schott, 2016; Fort *et al.*, 2018).

Both reduced-form analysis and structural estimation point to significant heterogeneity in the impact of robots across upstream and downstream imports. At both firm and industry levels, imports of intermediate inputs benefit more from robots than imports of assembled outputs. This implies that countries with a comparative advantage in input production gain more from robot adoption in the United States than countries specializing in assembly. However, are these patterns contingent on the comparative advantage of robots in downstream production? How would they change if industrial robots become as productive at making, say, tires as they currently are at assembling cars? In light of these questions, I conduct a counterfactual exercise where the productivity of robots in input production rises to the same level as their productivity in assembly. It turns out that effects of this hypothetical shock on imports are the exact opposite of changes induced by the rise of robots *so far*. In aggregate, the hypothetical shock leads to a 53% reduction in input imports and a 22% rise in output imports, adding up to a 36% decrease in overall imports. These results suggest that even though robots have promoted globalization so far, this relationship may not be sustainable in case the comparative advantage of robots in assembly relative to input production fades. In terms of heterogeneity across different types of imports, future technological change may benefit or harm countries in a different way.

Finally, the calibrated model presents a natural framework to evaluate costs and benefits of robot policies. In 2017, the European Parliament voted on a proposal to tax the use of robots (Delvaux, 2016). Even though the proposal was eventually rejected, there have been ongoing discussions of the possibility of a robot tax[3]. In addition to a tax on robot investments, some have advocated for an even stronger value-added tax, targeting specific types of robots to protect particularly vulnerable workers (Oberson, 2019). On the other hand, countries such as China have used subsidies to encourage the use of robots (Lin, 2018; Cheng *et al.*, 2019). In line with these potential policy options, I use the estimated model to consider consequences of three types of robot policies: a robot investment tax that increases

[3]Public figures who have advocated for a robot tax include Bill Gates (Quartz, 2017), Congresswoman Alexandria Ocasio-Cortez (Market Watch, 2019), and NYC mayor Bill de Blasio (de Blasio, 2019).

the fixed cost of robot adoption by 30%, a value-added tax on assembly robots that reduces their productivity by 25%, and a robot subsidy that decreases the fixed cost of adoption by 30%. Effects of the robot investment tax and the robot subsidy on aggregate outcomes are relatively small, because a change in the fixed cost of adoption only affects smaller, less productive firms at the margin of adopting, but not the larger adopters which play a more important role in determining aggregate outcomes. On the other hand, a value-added tax on assembly robots does lead to a sizeable 10% increase in aggregate employment of domestic assembly workers, but this comes at the cost of lower imports and a higher price index. These results suggest that a properly designed tax may achieve reallocation towards domestic workers, but resulting efficiency losses need to be taken into account.

This paper is most closely related to an extremely recent literature which studies the firm-level impact of automation on imports and offshoring (Stapleton and Webb, 2020; Alguacil Marí *et al.*, 2020; Cilekoglu *et al.*, 2021). I complement these papers by making three contributions. First, I develop a novel instrumental variables strategy that exploits exogenous variation in labor supply of robot-complementary workers due to immigrant inflows. Compared to existing studies whose identification strategies rely on workers particularly *vulnerable* to being displaced by robots or methods such as propensity score matching, my IV strategy utilizes exogenous variation in robot-*complementary* workers that is previously unexplored. Second, I uniquely focus on the heterogeneous effects of robots on upstream and downstream imports. My results uncover nuances in the relationship between automation and offshoring and generate important implications for how countries gain or lose from robot adoption in the US. Third, my structural estimation provides the first quantitative estimates of the impact of robots on imports under industry equilibrium. The calibrated model provides a unified characterization of how adopters and non-adopters respond to a robot technology shock, and shows that positive effects of robots on imports at the firm level translate into an aggregate increase in imports of the manufacturing sector. These aggregate results suggest trade-*inducing* effects of robots consistent with Hallward-Driemeier and Nayyar (2019) and Artuc *et al.* (2020), who find that robotization in advanced economies is associated with an increase in imports from and FDI into developing countries. In the meantime, I estimate a negative effect of robots on imports of assembled goods, indicating that countries specializing in assembly might suffer from robot adoption in the US. This finding provides a potential explanation for the negative impact of robot adoption in advanced economies on imports from developing countries such as Mexico and Colombia, as documented in Artuc *et al.* (2019), Kugler *et al.* (2020), and Faber (2020).

More broadly, my paper is related to a large group of studies on the impact of automation on productivity and workers. Relative to this strand of literature, my theoretical and empirical results provide a unified explanation of both micro and macro evidence. Consistent with the majority of firm-level studies (Humlum, 2019; Koch *et al.*, 2019; Acemoglu *et al.*, 2020b; Aghion *et al.*, 2020; Bessen *et al.*, 2020; Bonfiglioli *et al.*, 2020; Dixon *et al.*, 2020), I estimate a positive effect of robots on sales and productivity. However, in aggregate, my quantitative exercise suggests that the impact of robots on total employment of the manufacturing sector is negative. Recent studies on the impact of automation on employment at the region or industry level have pointed to a mix of effects ranging from positive (Aghion *et al.*, 2020; Mann and Püttmann, 2021) to insignificant or mildly negative (Chiacchio *et al.*, 2018; Graetz and Michaels, 2018; Dauth *et al.*, 2019; Acemoglu and Restrepo, 2020). In comparison with these papers, my results contribute a unique perspective on how automation affects *aggregate* manufacturing employment. Furthermore, I distinctively point to a specific group of workers - assembly workers - who absorb the greatest damage from a rise of robots. The contrasting micro and macro employment effects of robots can be partially explained by within-industry competition, as expansion of adopters forces non-adopters to contract their employment. The mechanism of within-industry reallocation is in line with business-stealing effects of automation documented in Acemoglu *et al.* (2020b) and Aghion *et al.* (2020). Theoretically, the framework of heterogeneous firms proves to be vital for capturing this margin of adjustment.

My theoretical model adopts the framework of input offshoring in Gopinath and Neiman (2014) and Halpern *et al.* (2015), but additionally accounts for imports of produced outputs.

My structural estimation allows firms to choose their number of imported products for both input and output imports, and provides a novel characterization of the fixed cost schedule for imported outputs. I find that in general, the fixed costs of importing outputs are larger than the fixed costs of importing inputs. I construct and estimate industry equilibrium of the model using a similar framework as Antras *et al.* (2017). Finally, my paper is also related to a rich body of literature on the relationship between technology and offshoring (Grossman and Helpman, 1995; Feinberg and Keane, 2006; Grossman and Rossi-Hansberg, 2008; Bartel *et al.*, 2014; Fort, 2017), and a broader literature on technology and organization of production within firms (Brynjolfsson and Hitt, 2000; Bresnahan *et al.*, 2002; Bartel *et al.*, 2005; Abramovsky and Griffith, 2006; Acemoglu *et al.*, 2010; Bloom *et al.*, 2014). I show that a labor-displacing technology can have a positive effect on offshoring, but this effect is nuanced across different types of imports along the production process.

The rest of the paper is organized as follows. section 2.2 presents the theoretical model. section 2.3 describes the data used for empirical analysis and presents descriptive statistics related to robot adoption. section 2.4 presents the empirical results. section 2.5 describes the methodology for a quantitative exercise and reports the results. section 2.6 analyzes three counterfactual experiments based on the calibrated model. section 2.7 concludes with a discussion of future work.

2.2 Theory

In this section, I present a theoretical model connecting robots and offshoring at the firm level. The model features heterogeneous firms in the style of Melitz (2003). To shed light on the comparative advantage of robots in downstream production as discussed in section 2.1, I assume that production of final goods consists of two stages: sourcing of inputs and assembly. The model also incorporates offshoring in both stages of production, which allows robots to separately interact with upstream and downstream imports.

The full model features a rich and complex set of interactions between robot adoption and offshoring that are difficult to analyze. Therefore, in subsection 2.2.1, I introduce a highly stylized version of the model. While this benchmark model makes a few simplifications, it represents a minimal framework that is sufficient to capture the key forces governing decisions of robot adoption, and the relationship between robots and offshoring with a focus on *complementarities*. This allows the benchmark model to deliver three sharp predictions that are later tested in section 2.3 and section 2.4. I discuss those predictions in subsection 2.2.2. The extended model is mainly useful for structural estimation, so its discussion is delayed to section 2.5.

2.2.1 Benchmark Model

Consumers

Consider a domestic country, d, which contains a measure L of consumers with CES preferences over a continuum of consumer goods produced by the manufacturing sector. Products are denoted by ω. Assume that domestic consumers spend a constant level of expenditure E on manufacturing products.

Consumer preferences over manufacturing goods are given by

$$U = \left(\int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}, \tag{2.1}$$

where ω denotes the product, and $q(\omega)$ denotes the quantity of product ω

Standard CES results imply that demand $q(\omega)$ is

$$q(\omega) = EP^{\sigma-1} p(\omega)^{-\sigma},$$

where P is the aggregate ideal price index over all products

$$P = \left(\int_{\omega \in \Omega} p(\omega)^{1-\sigma} d\omega, \right)^{\frac{1}{1-\sigma}},$$

and $p(\omega)$ is the price of the product ω. Observe that total expenditure on product ω, $R(\omega)$, is simply equal to

$$R(\omega) = p(\omega)q(\omega) = RP^{\sigma-1}p(\omega)^{1-\sigma}.$$

Under the assumption that each good is produced under constant marginal costs, we know that profits associated with ω can be written as:

$$\pi(\omega) = Bc(\omega)^{1-\sigma},$$

where the demand factor B is defined as

$$B = \frac{1}{\sigma}EP^{\sigma-1}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma}, \tag{2.2}$$

and $c(\omega)$ is the marginal cost of production for consumer good ω.

Production

Consumer goods are produced by a continuum of final good producers indexed by χ engaging in monopolistic competition. Each firm has a core productivity level φ drawn from a productivity distribution. Production of each good ω involves two stages: stage 1, in which the firm procures inputs; and stage 2, in which the firm assembles the final output.

In stage 1, production of each output requires a composite intermediate input X, which is a CES aggregate of a bundle of domestic inputs, M_d, and a bundle of foreign input goods, M_f, characterized by

$$X(\omega) = \left[M_d(\omega)^{\frac{\rho-1}{\rho}} + M_f(\omega)^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}, \tag{2.3}$$

where ρ is the elasticity of substitution between inputs. Each input bundle is itself a CES aggregate of individual intermediate inputs, expressed as

$$M_d = \left[\int_{j\in\Omega_d} m_{dj}^{\frac{\rho-1}{\rho}}\,dj\right]^{\frac{\rho}{\rho-1}}, \quad M_f = \left[\int_{k\in\Omega_f} m_{fk}^{\frac{\rho-1}{\rho}}\,dk\right]^{\frac{\rho}{\rho-1}}.$$

The implicit assumption is that the elasticity of substitution across individual inputs within domestic and foreign inputs is equal to the elasticity of substitution *between* domestic and foreign inputs. Following Gopinath and Neiman (2014), I assume that all firms have free access to the same set of domestic inputs, Ω_d; meanwhile, access to foreign inputs is subject to a fixed cost schedule $f_1(n_1)$, where $n_1 = |\Omega_d|$ is the mass of foreign inputs chosen by the firm. The presence of fixed costs in foreign sourcing is consistent with empirical evidence presented in Bernard *et al.* (2009) and Antras *et al.* (2017).

Intermediate inputs are produced by perfectly competitive input producers with constant marginal costs. For notational convenience, denote the unit cost of the domestic input bundle as c_{1d}, and the cost of a foreign input bundle as c_{1f}, such that the cost of the composite intermediate good, $c(X, \omega)$, is given by[4]

$$c(X,\omega) = \left[c_{1d}(\omega)^{1-\rho} + n_1(\tau c_{1f}(\omega))^{1-\rho}\right]^{\frac{1}{1-\rho}},$$

where $\tau \geq 1$ is an iceberg trade cost.

In stage 2, I assume that each firm owns two assembly plants: one in the domestic country d, and one in the foreign country f. In each plant, the firm produces a differentiated output. I assume that all final goods produced at the foreign assembly plant are consumed by domestic consumers only. I use the foreign assembly plant to capture offshoring of assembly. In the language of Baldwin and Venables (2013), the two assembly plants correspond to two "snakes" of global production with the assembly stage happening in two different countries.

Consider a firm χ. Production of final goods follows a Cobb Douglas technology, where α is the expenditure share on the composite input good, and $1 - \alpha$ is the expenditure share on assembly. In stage 1, production at both locations uses a mix of domestic and foreign inputs, where the mass of foreign inputs n_1 is the same across both production locations for a given firm. In other words, a single choice of n_1 determines access to foreign inputs at both the domestic and the foreign assembly plant. In stage 2, assembly is completed by workers and, potentially, robots. The good produced at the domestic plant is sold directly to domestic consumers. The good produced at the foreign plant is first imported

[4] This definition only applies to goods assembled in a domestic plant. I later introduce a foreign assembly plant for which the cost of inputs is different due to geography and presence of fixed costs.

by the firm at the marginal cost, and then sold to domestic consumers[5]. Importing of the foreign-assembled good is subject to the same iceberg trade cost τ.

I model robot adoption as a binary decision, denoted as I_R. To adopt robots, each firm needs to pay a fixed cost f_R. Under the assumption that robots are only useful in assembly, adoption leads to a reduction in the unit cost of domestic assembly by a factor $\beta_{2R} > 1$[6].

Under this setup, the firm-level maximization problem for a firm with productivity φ is

$$\max_{n_1, I_R} B\varphi^{\sigma-1} \left[\left(c_{1d}^{1-\rho} + n_1(\tau c_{1f})^{1-\rho} \right)^{\frac{\alpha}{1-\rho}} \left(\frac{c_{2d}}{1 + I_R(\beta_{2R} - 1)} \right)^{1-\alpha} \right]^{1-\sigma} + \tag{2.4}$$

$$B\varphi^{\sigma-1}\tau^{1-\sigma} \left[\left((\tau c_{1d})^{1-\rho} + n_1(c_{1f})^{1-\rho} \right)^{\frac{\alpha}{1-\rho}} (c_{2f})^{1-\alpha} \right]^{1-\sigma} - f_1(n_1) - I_R \cdot f_R. \tag{2.5}$$

The firm chooses import intensity n_1 and robot adoption I_R to maximize its profits, subject to fixed costs of sourcing and robot adoption. Given that the only source of heterogeneity across firms comes from differences in productivity, firms can be indexed by φ. Profit of firm φ is thus denoted as $\pi(\varphi)$, defined as the maximand of Equation 2.4.

One main innovation of the model is that it explicitly distinguishes between two types of offshoring, captured by two separate categories of imports. **Stage 1 imports**, or **upstream imports**, represent the value of foreign inputs purchased by the firm at the domestic assembly plant. Denoting this by $M_1(\varphi)$, upstream imports can be written as:

$$M_1(\varphi) = B\alpha(\sigma-1)\varphi^{\sigma-1} \left[\left(c_{1d}^{1-\rho} + n_1(\tau c_{1f})^{1-\rho} \right)^{\frac{\alpha}{1-\rho}} \left(\frac{c_{2d}}{1 + I_R(\beta_{2R} - 1)} \right)^{1-\alpha} \right]^{1-\sigma} \cdot \frac{n_1(\tau c_{1f})^{1-\rho}}{c_{1d}^{1-\rho} + n_1(\tau c_{1f})^{1-\rho}}. \tag{2.6}$$

[5] Although I assume the foreign assembly plant to be owned by the firm, I do not take a stance on whether foreign assembly happens within boundaries of the firm. As long as the firm is able to purchase goods assembled abroad at the marginal cost, the current setup is equivalent to the alternative where imports of assembled outputs take place through arms-length transactions.

[6] This is a reduced-form formulation of the role of robots in assembly. Alternatively, I can introduce microfoundation into the model by assuming that the unit cost of domestic assembly after robot adoption becomes $\left(c_{2d}^{1-\eta} + c_{2R}^{1-\eta} \right)^{\frac{1}{1-\eta}}$, where c_{2R} is the cost of assembly under robots, and η is the elasticity of substitution between robots and domestic workers. For the purpose of this section, summarizing the effects of robots on assembly as a factor of cost reduction does not affect any implications of the model. In section 2.5, I use the microfounded setup, which allows the model to explicitly speak to the impact of robot adoption on domestic assembly workers.

The first component represents the total value of (all) inputs purchased by the domestic assembly plant, whereas the second component reflects the share of inputs purchased from the foreign country.

On the other hand, **stage 2 imports**, or **downstream imports**, capture the value of final outputs produced at the foreign assembly plant, which are purchased by the firm and return to the domestic country before being sold to domestic consumers. Denoting this as $M_2(\varphi)$, output imports can be written as:

$$M_2(\varphi) = B\sigma\varphi^{\sigma-1}\tau^{1-\sigma}\left[\left((\tau c_{1d})^{1-\rho} + n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha}{1-\rho}}(c_{2f})^{1-\alpha}\right]^{1-\sigma}.$$

These two types of imports correspond to two different types of offshoring along the production process. Upstream imports indicate offshoring of input production, while downstream imports indicate offshoring of assembly. The sum of stage 1 and stage 2 imports give the total value of imports purchased by a firm:

$$M(\varphi) = M_1(\varphi) + M_2(\varphi). \tag{2.7}$$

Finally, I denote firm sales and input expenditure of firm φ as $R(\varphi)$ and $X(\varphi)$.

2.2.2 Implications

First, examining the maximization problem leads to the following proposition:

Proposition 1. *The solution (n_1^*, I_R^*) to the firm-level profit maximization problem is increasing in the strong set order on φ, β_{2R}.*

Proposition 1 means that the solution to the profit maximization problem is *increasing* in both productivity φ and firm productivity φ. Its proof relies on elementary arguments in monotone comparative statics. It can first be shown that the profit function is supermodular on the choice variables, n_1 and I_R, and that it exhibits increasing differences in the choice variables and exogenous parameters φ and β_{2R}. Then the proposition follows from the

Topkis monotonicity theorem. A detailed proof is included in subsection A.1.1.

This result implies that both the number of imported inputs n_1 and the decision of robot adoption I_R should exhibit selection on productivity. These implications are formalized in the following two corollaries.

Corollary 1 (Selection into importing)**.** *Given two firms φ, φ' with $\varphi \geq \varphi'$, the following must be true:*

$$n_1(\varphi) \geq n_1(\varphi').$$

Corollary 2 (Selection into robot adoption)**.** *Given two firms φ, φ' with $\varphi \geq \varphi'$, it must be that $I_R(\varphi) \geq I_R(\varphi')$.*

Corollary 2 indicates that, in spite of the *direct* changes in firm-level outcomes caused by robot adoption, robot adopters have an *ex-ante* advantage in productivity compared to non-adopters. Similarly, Corollary 1 suggests that the number of imported inputs, n_1, also exhibits selection on productivity, i.e. firms with a higher productivity would always import a larger number of inputs[7]. Combining the two corollaries, the model implies that robot adopters should also on average choose a higher n_1. More broadly, compared to non-adopters, robot adopters are predicted to purchase a higher value of both upstream and downstream imports.

Next, let us turn to the impact of robots on firm-level outcomes by considering what happens to a firm if it switches from not adopting robots to adopting robots. To isolate the effect of robots, assume that this adoption decision is driven by an increase in robot productivity, β_{2R}, or a reduction in the fixed cost of robot adoption, f_R. In the following proposition, I show that robot adoption induced by such a shock leads to an expansion in sales and in both margins of imports.

Proposition 2. *Consider a firm that switches from not adopting to adopting robots, in response to*

[7] Another way to interpret Corollary 1 is that the extensive margin of input imports, given by the number of imported inputs, increases in productivity. Note that in Antras *et al.* (2017), this would be true only under specific parametric restrictions, while in my framework the result always holds.

an increase in β_{2R} or a decrease in f_R. Expressing the proportional change in a variable Y for a firm with productivity φ as $\Delta \log Y(\varphi)$, robot adoption leads to an increase in sales, input imports, and output imports of the firm, i.e.

- $\Delta \log R(\varphi) > 0$,
- $\Delta \log M_1(\varphi) > 0$,
- $\Delta \log M_2(\varphi) > 0$.

A detailed proof is included in subsection A.1.2. To understand the intuition behind Proposition 2, let us consider the forces behind the change in upstream and downstream imports. Adopting robots leads to cost reductions in domestic assembly, and induces the firm to expand the scale of production at the domestic assembly plant. This not only leads to a direct increase in the use of inputs (and thus upstream imports) due to a scale effect, but also incentivizes the firm to choose a higher number of imported inputs n_1, which translates into an increase in upstream imports as well. On the other hand, even though robots are not directly used in foreign assembly, the increase in n_1 implies that production at the foreign assembly also uses a higher number of foreign inputs. The firm thus expands production at the foreign plant as well, which leads to an increase in downstream imports. Overall, given that a robot adopter expands production at both assembly plants, its total sales increases as well.

Formally, the proportional change stage 1 and stage 2 imports can be decomposed as follows:

$$\begin{aligned}\Delta \log M_1(\varphi) = &\underbrace{(\sigma-1)(1-\alpha)\Delta \log \beta_{2R}}_{\text{Assembly Productivity Effect}} + \underbrace{\Delta \log n_1(\varphi)}_{\text{Import Extensive Margin}} + \\ &\underbrace{\left(\frac{\sigma-1}{\rho-1}\alpha - 1\right)\Delta \log\left(c_{1d}^{1-\rho} + n_1(\varphi)(\tau c_{1f})^{1-\rho}\right)}_{\text{Input Productivity Effect}}\end{aligned} \tag{2.8}$$

$$\Delta \log M_2(\varphi) = \underbrace{\left(\frac{\sigma-1}{\rho-1}\alpha\right) \Delta \log \left((\tau c_{1d})^{1-\rho} + n_1 (c_{1f})^{1-\rho}\right)}_{\text{Input Productivity Effect}}. \tag{2.9}$$

Consistent with the previous discussion, the increase in upstream imports can be decomposed into three components: an assembly productivity effect driven by the increase in stage 2 productivity (at the domestic assembly plant) due to robot adoption; an input productivity effect that summarizes the increase in stage 1 productivity following robot adoption; and an import extensive margin effect capturing the increase in n_1. Since all individual effects are positive, robot adoption unequivocally leads to an increase in upstream imports. On the other hand, the change in downstream imports consists of a single term: the input productivity effect, which is positive.

Moreover, a comparison between the two expressions reveals that the increase in upstream imports is guaranteed to be greater than the increase in downstream imports as long as a simple condition shows. I present this prediction of the model in the final proposition:

Proposition 3 (Upstream imports v.s. downstream imports)**.** *Consider a firm ω that switches from not adopting robots to adopting. Then the proportional change in upstream imports for a firm with productivity φ is higher than that in downstream imports, i.e.* $\Delta \log M_1(\varphi) > \Delta \log M_2(\varphi)$, *if*

$$\frac{\sigma-1}{\rho-1}\alpha < 1.$$

Intuitively, the expression $\frac{\sigma-1}{\rho-1}\alpha$ governs the contribution of a higher n_1 to imports in both stages. As long as this expression is not too large, the sum of the input productivity effect and import extensive margin effect for upstream imports would be larger than the input productivity effect for downstream imports. Given that the increase in upstream imports contains an additional positive assembly productivity effect, it must be larger than the increase in stage 2 imports. Note that the condition in Proposition 3 is plausible under reasonable parametric values. For example, if I set $\sigma = \rho = 4$ as in Gopinath and Neiman (2014) and Halpern *et al.* (2015), since the share of inputs in production, α, is by definition less than 1, the condition must be satisfied.

To sum up, the benchmark model generates three predictions:

- **Prediction 1**: robot adoption exhibits selection on productivity, and complementarities with both upstream and downstream imports.
- **Prediction 2**: at the firm level, robot adoption leads to an increase in sales and both upstream and downstream imports.
- **Prediction 3**: following robot adoption, the increase in upstream imports is larger than the increase in downstream imports.

Given these implications, it is evident that this benchmark model is particularly useful in terms of isolating *complementarities* between robots and offshoring. This focus on complementarities is in part driven by the simplifying assumptions behind the benchmark model, which omit substitution effects between robots and offshoring. As I discuss in subsection 2.5.1, relaxing those assumptions makes the model considerably more difficult to analyze. Given that, I consider the benchmark model as a parsimonious framework that isolates the main mechanisms of robot adoption in the most transparent way possible. In section 2.5, I introduce an extended model to evaluate the role of substitution effects and whether they affect predictions of the benchmark model in a meaningful way.

2.3 Data

The theoretical framework presented in section 2.2 delivers sharp predictions regarding robot adoption and its relationship with offshoring at the firm level. Testing these predictions requires a detailed firm-level dataset connecting robots and trade. To that end, I assemble a highly disaggregated dataset of manufacturing firms using U.S. Census data. In the following, I describe my main data sources and the methodology I use to identify upstream & downstream imports and robot adoption.

My main dataset links three data sources on US manufacturing firms. I obtain firm-level trade data from the Longitudinal Firm Trade Transaction Database (LFTTD), and information on employment, payroll, and industry classification[8] from the Longitudinal Business Database (LBD), both recorded on a yearly basis. I identify manufacturing establishments as those with NAICS codes starting with the number "3", and identify manufacturing firms as firms containing at least one manufacturing establishments[9]. In addition, the dataset incorporates data from the Economic Censuses (EC), which provide sales and expenditures of the universe of US establishments in 5-year intervals, on years ending in "2" or "'7" also known as Census years. To extend EC data to a yearly basis, I conduct linear interpolation on the log of sales and input expenditures observed in the EC to infer the same variables in non-Census years. The end product is a unbalanced panel dataset of the universe of US manufacturing firms between 1992 and 2016, containing more than 3 million observations.

Unfortunately, there does not exist a dataset that reports firm-level robot adoption in the US before 2016[10]. However, taking advantage of the fact that most producers of industrial robots are headquartered outside the United States (Leigh and Kraft, 2018), I identify firm-level adoption of industrial robots using robot imports. After identifying the set of firms importing industrial robots, I rule out carry-along traders and potential robot producers that might import robots as parts. This process closely follows Wang (2022a). In that paper, I show that robot imports account for the majority of industrial robots used in the US, and that cross-industry patterns of robot adoption identified using my methodology are consistent with widely used industry-level data on robots in the US.

The benchmark model described in Section 2.2 predicts that robot adoption is associated with a greater increase in upstream imports than downstream imports. Testing this hypoth-

[8]To ensure that industry classification is consistent over time, I use the system developed by Fort and Klimek (2018) to map all NAICS codes to a unique 2012 NAICS code.

[9]Following Ding *et al.* (2020), I include non-manufacturing establishments of manufacturing firms to account for sales of manufacturing output channeling through, say, wholesale establishments.

[10]A few recent US Census datasets provide firm-level information on use of robots. For example, the 2018 Annual Capital Expenditures Survey (ACES) contains data on firm-level capital expenditures for both industrial and service robotics; the 2018 Annual Business Survey (ABS) contains an explicit question that asks respondents to report use of technologies such as robotics and cloud computing (Zolas *et al.*, 2021); the 2018 Annual Survey of Manufactures (ASM) contains three new questions on the use of robotics (Buffington *et al.*, 2018).

esis requires an empirical methodology that classifies imports according to their relative position in the production process. Corresponding to upstream imports and downstream imports in the data, I use the data to classify *input* and *output* imports at the firm level. This classification is developed using establishment-level information on inputs and outputs provided by the material trailer and product trailer files of the Census of Manufactures (CM), which record the complete set of inputs used and outputs produced by a subset of manufacturing establishments[11]. I aggregate input-output information to the NAICS level to identify potential inputs and outputs for every manufacturing firm, based on their mix of NAICS industries. I then link input-output data to the LFTTD to identify imports that are potentially inputs used by a firm, and potentially outputs produced by a firm. Since some firms possibly produce their own inputs, some imports might be simultaneously identified as *both* inputs *and* outputs for a given firm. Given my focus on upstream vs. downstream imports, I classify all identified inputs as input imports, and define output imports as products that are identified as outputs *only*, and not inputs. Both types of imports exclude industrial robots.

To better understand this classification scheme, consider an automotive firm importing two goods: (1) produced cars, and (2) car doors. Cars are flagged as an output for the firm but not an input, so they are classified as output imports for the firm. Meanwhile, car doors are flagged as both an input and an output for the firm, because the firm produces car doors in-house. However, that does not change the fact that car doors are an input in the production process of the firm. Therefore, I classify car doors as an input import.

Finally, I apply the Inverse Hyperbolic Sine (IHS) transformation to all import variables. The IHS transformation, also written as $\text{arcsinh}(\cdot)$, is defined as

$$\text{IHS}(x) = \ln\left(x + \sqrt{x^2+1}\right). \tag{2.10}$$

IHS delivers a similar interpretation as the commonly used logarithmic transformation

[11] The material trailer data are available only for a subset of manufacturing establishments, including multi-unit firms with at least 250 employees, and a randomly selected sample of single-unit firms.

in terms of elasticities[12], with the advantage that it is defined for zero-valued observations. Given that in the data, zeros are ample for the import variables, I choose the IHS transformation over the more widely used logarithmic transformation.

2.4 Empirical Results

The detailed dataset constructed in section 2.3 allows me to empirically understand robot adoption and offshoring at the firm level. My empirical analysis centers around testing the three predictions generated by the stylized model in section 2.2.

In this section, I start by providing two sets of descriptive evidence. In subsection 2.4.1, I use cross-sectional regressions to examine ex-post and ex-ante differences between robot adopters and non-adopters. These regressions provide a direct test of Prediction 1, which states that robot adoption exhibits selection on productivity and complementarities with imports. In subsection 2.4.2, I present results from event study regressions. Both sets of descriptive evidence suggest that estimating the causal impact of robots requires an alternative methodology to the OLS. To that end, I develop an instrumental variables strategy based on immigrant inflows of robot-complementary workers in subsection 2.4.3. Using the constructed IV, I use panel IV regressions to estimate the effect of robots on firm-level outcomes. subsection 2.4.4 presents results on the firm-level effects of robot adoption on sales as well as upstream and downstream imports. The results are consistent with Prediction 2 and Prediction 3, and confirm the importance of complementarities between robots and offshoring. Finally, in subsection 2.4.5, I provide a discussion of the empirical results and robustness checks.

[12]For a more detailed discussion on the IHS transformation, see Burbidge *et al.* (1988) and MacKinnon and Magee (1990).

2.4.1 Are robot adopters different from non-adopters?

To understand how robot adoption shapes firm-level outcomes, I start by gauging the cross-sectional differences between robot adopters and non-adopters. I run the following specification:

$$Y_{it} = \alpha + \beta \text{Robot Adopter}_{it} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \varepsilon_{it}, \tag{2.11}$$

where Y_{it} is the dependent variable for firm i in year t, and $\gamma_{sc}, \gamma_{st}, \gamma_{ct}$ are industry-CZ, industry-year, and CZ-year fixed effects. Since some firms operate in multiple commuting zones and/or multiple industries, I respectively take the industry and CZ in which the firm has the highest employment in 1992. Fixed effects are included to control for industry and CZ-specific trends and characteristics. The coefficient β thus captures the premium of robot adopters over non-adopters in variable Y.

Results under Equation 2.11 are displayed in Table 2.1. Columns (1) and (2) show that robot adopters are significantly larger than non-adopters: they are nearly four times as large in sales, and more than three times as large in employment. Even after controlling for sales, robot adopters import more overall, and are more likely to import, as shown in columns (3) and (4). Columns (5) and (6) confirm that the advantage in importing associated with robot adoption is present for both input and output imports.

However, the question remains as to whether the premium of robot adopters documented in Table 2.1 demonstrates meaningful firm-level changes as a consequence of robot adoption, or simply reflects ex-ante differences between adopters and non-adopters. Indeed, Prediction 1 of the theory states that robot adoption exhibits selection on productivity. Are adopters already different than non-adopters *prior to* adopting robots? To shed light on this question, I restrict the sample to firms in 1992 that have not yet adopted robots, and run the following selection regression:

$$Y_{i,1992} = \alpha + \beta \text{Future Adopter}_{i,1992} + \gamma_{sc} + \varepsilon_i, \tag{2.12}$$

where Future Adopter$_{i,1992}$ is equal to one if firm i has not yet adopted robots in 1992, but

Table 2.1: *Premium of Robot Adopters*

	(1) Log(Sales)	(2) Log(Employment)	(3) IHS(Imports)
Robot Adoption	2.871***	2.316***	3.676***
	(0.255)	(0.231)	(0.278)
Log(Sales)			0.850***
			(0.0638)
Industry-Year FEs	✓	✓	✓
CZ-Year FEs	✓	✓	✓
Industry-CX FEs	✓	✓	✓
Observations	770000	770000	770000

	(4) Import Dummy	(5) IHS(Input Imports)	(6) IHS(Output Imports)
Robot Adoption	0.130***	3.956***	3.506***
	(0.0290)	(0.220)	(0.180)
Log(Sales)	0.106***	0.722***	0.445***
	(0.00573)	(0.0579)	(0.0434)
Industry-Year FEs	✓	✓	✓
CZ-Year FEs	✓	✓	✓
Industry-CZ FEs	✓	✓	✓
Observations	770000	770000	770000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of cross-sectional regressions following the specification in Equation 2.11. The sample used is all census-year observations between 1992 and 2012 for the universe of US manufacturing firms. All regressions include industry-year, CZ-year, and industry-CZ fixed effects. All observation counts are rounded for disclosure avoidance.

ends up adopting robots in a subsequent year. The coefficient β thus captures the *ex-ante* difference between adopters and non-adopters in firm-level variable Y_i.

Table 2.2 reveals that, even before robot adoption happens, future adopters already exhibit advantages in various dimensions over non-adopters. Future adopters are larger, employ more workers, import more overall and in both upstream and downstream stages of the production process. These empirical evidence provides support for the prediction that firms with higher productivity select into robot adoption, as outlined in Prediction 1.

Table 2.2: *Premium of Future Adopters*

	(1) Log(Sales)	(2) Log(Employment)	(3) IHS(Imports)
Future Adoption	2.807***	2.270***	2.928***
	(0.246)	(0.204)	(0.415)
Log(Sales)			0.600***
			(0.0621)
Industry-Year FEs	✓	✓	✓
CZ-Year FEs	✓	✓	✓
Industry-CZ FEs	✓	✓	✓
Observations	770000	770000	770000
	(4) Import Dummy	(5) IHS(Input Imports)	(6) IHS(Output Imports)
Future Adoption	0.267***	2.928***	1.828***
	(0.0389)	(0.345)	(0.334)
Log(Sales)	0.0822***	0.464***	0.333***
	(0.00745)	(0.0533)	(0.0385)
Industry-Year FEs	✓	✓	✓
CZ-Year FEs	✓	✓	✓
Industry-CZ FEs	✓	✓	✓
Observations	770000	770000	770000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of cross-sectional regressions following the specification in Equation 2.12. The sample used is the universe of manufacturing firms in 1992 that have not adopted robots. All regressions include industry-year, CZ-year, and industry-CZ fixed effects. All observation counts are rounded for disclosure avoidance.

2.4.2 Event Studies

After verifying Prediction 1 in Section 2.4.1, I evaluate the next two predictions of the model, which state that at the firm level, robot adoption leads to an increase in sales and both upstream and downstream imports, and a greater increase in upstream imports. To evaluate the within-firm impact of robot adoption, I use panel regressions with the following baseline specification:

$$Y_{it} = \alpha + \beta \text{Robot Adopter}_{it} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \gamma_{qt} + \gamma_i + \varepsilon_{it}. \tag{2.13}$$

Selection into robot adoption implies that robot adoption is correlated with other firm characteristics (e.g. productivity) that potentially affect firm-level outcomes but cannot be completely observed. This presents a major barrier to identifying the causal effect of robots. Therefore, relative to the cross-sectional specification in Equation 2.11, I introduce firm fixed effects denoted by γ_i to control for (time-invariant) firm-specific characteristics. In addition, I include quantile-year fixed effects γ_{qt}, where quantile of a firm is defined as its employment quartile in the initial year 1992. This fixed effect is intended to control for heterogeneous trends across firms with different initial firm size, thereby capturing time-varying fixed effects that are correlated with robot adoption but not directly caused by robots. Hereafter, I refer to these fixed effects as "baseline trend controls". I restrict the same to firms that already existed in 1992, i.e. firms for which baseline trend controls are defined.

The coefficient β in Equation 2.13 is intended to identify the within-firm change in the dependent variable caused by robot adoption. However, even after including firm fixed effects and baseline trend controls, one might be concerned about giving β a causal interpretation. First, baseline trend controls only account for trends that are correlated with baseline employment, but not trends related to other firm-level variables. Even if I included a version of the baseline trend controls for *every* observable of the firm, there might still exist unobserved firm characteristics that lead to heterogeneous trends across adopters and non-adopters, thus breaking the assumption of parallel trends that is necessary for identification. For example, if adopters grow faster than non-adopters even without robot adoption, then the OLS specification in Equation 2.13 would overestimate the true value β. Furthermore, robot adoption might be induced by expectation about the future that is not manifested in firm-level outcomes. Depending on whether this expectation is positive (e.g. firms adopt in expectation of future growth in productivity) or negative (e.g. firms adopt to weather a potential negative shock), there would be an upward or downward bias associated with the OLS estimates.

A simple way to take a first look at how robots interact with firm-level outcomes and test the assumption of parallel trends are event study regressions. I use the following distributed lead-lag model (Stock and Watson, 2015):

$$Y_{it} = \alpha + \sum_{k=-5}^{k=5} \delta_k \text{Robot Adopter}_{i,t-k} + \gamma_{sc} + \gamma_{st} + \gamma_{ct} + \gamma_{qt} + \gamma_i + \varepsilon_{it}. \tag{2.14}$$

where Robot Adopter$_{i,t-k}$ indicates whether a firm has adopted robots in year $t-k$. Through the lead-lag coefficients, δ_k, this specification sheds light on the dynamic response of the firm-level outcome Y_{it} to robot adoption in the five years before and after adoption. I set year $t-1$ as the base level, so all coefficient estimates for δ_k can be interpreted relative to the year before robot adoption.

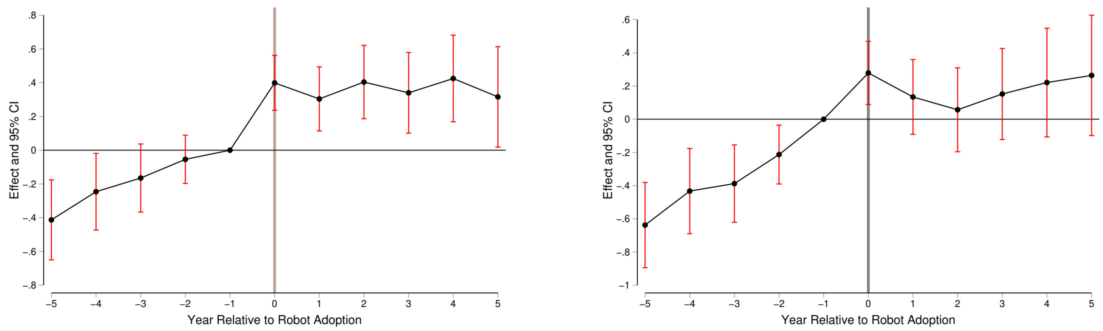
Results of the event study regressions are shown in Figure 2.1. There are visible increasing pre-trends for input imports and output imports, suggesting that robot adopters are already expanding in imports relative to non-adopters prior to adoption. Robot adoption is associated with an *ex-post* increase in both variables.

It is worth pointing out that results from these event studies are consistent with the theoretical predictions: robot adoption is associated with an increase input imports, and to some extent output imports. This increase appears to be more pronounced for input imports than for output imports: compared to year prior to robot adoption, input imports in subsequent years are consistently higher at a 5% level of statistical significance, while the effect on output imports is indistinguishable from zero in most years. The point estimates for input imports are also on average greater than those for output imports (0.4 v.s. 0.2). Importantly, the estimates for input and output imports do not reflect a mechanical increase driven by imports of industrial robots themselves, because both types of imports exclude robots.

Nevertheless, given previously discussed endogeneity concerns and evidence of divergent pre-trends across adopters and non-adopters, an alternative empirical methodology is needed to pin down empirical support for the theoretical predictions. The most obvious solution is to develop an instrumental variable exploiting exogenous variation in robot adoption, which I describe in the following section.

Figure 2.1: *Event Study Regressions*

(a) *Input Imports* **(b)** *Output Imports*



Notes: This figures plots results of event study regressions under the specification in Equation 2.14 for IHS(input imports) and IHS(output imports). I set the baseline as the year prior to robot adoption, so all coefficients should be interpreted relative to the baseline year. For each regression, I show the point estimates for the 5 years before robot adoption and 5 years after robot adoption, as well as corresponding 95% confidence intervals depicted as error bars. All regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as baseline trend controls, Standard errors are clustered at the firm level.

2.4.3 IV Strategy

The ideal instrumental variable should be *relevant* and satisfy the *exclusion restriction*. In other words, the IV should predict robot adoption with sufficient statistical power, but should also be correlated with firm-level variables only through the channel of robot adoption. More specifically, in light of endogeneity concerns mentioned above, the ideal IV should be (1) uncorrelated with firm-level trends and (2) unanticipated by the firm. Based on these criterion, I construct an instrumental variable to capture plausibly exogenous variation in robot adoption.

Following Wang (2022a), I adopt an IV strategy that relies on immigrant inflows of robot-complementary workers. Although robots are by definition autonomous, certain human workers are still indispensable to the processes of installing and integrating robots. To this end, I identify robot-complementary occupations using the O*NET database. Specifically, I use the O*NET connector to determine occupations that match the keyword "robots". The six occupations with the highest match scores are shown in Table 2.3.

Following David *et al.* (2013), I use US commuting zones (Tolbert and Sizer, 1996) as the

Table 2.3: *List of robot-complementary occupations identified by the O*NET connector, Top 6 by match score*

Occupation Name	Match Score
Robotics Technicians	100
Robotics Engineers	83
Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	56
Computer Numerically Controlled Tool Operators	56
Electro-Mechanical and Mechatronics Technologists and Technicians	48
Electrical and Electronics Repairers, Commercial and Industrial Equipment	20

geographical unit of analysis for local labor markets. At this point, it is tempting to directly use the change in the share of robot-complementary workers at the commuting zone level as the instrumental variable predicting robot adoption. However, to the extent that this change might be correlated with other regional economic trends (due to policies, industry composition, etc.), I proxy for the increase in availability of robot-complementary workers using immigrant inflows.

Specifically, I construct a shift-share instrument in the style of Altonji and Card (1991). To capture exogenous variation in immigrant labor supply of robot-complementary workers. For each commuting zone c and year t, I define an immigrant shock $z_{c,t}$ as

$$z_{c,t} = \sum_{g} \underbrace{\Delta L^{g}_{1990,t}}_{\text{Shift}} \cdot \underbrace{\frac{L^{g}_{c,1990} - R^{g}_{c,1990}}{L^{g}_{1990} - R^{g}_{1990}}}_{\text{Share 1}} \cdot \underbrace{\frac{R^{g}_{1990} - R^{g}_{c,1990}}{L^{g}_{1990} - L^{g}_{c,1990}}}_{\text{Share 2}} \cdot \underbrace{\frac{1}{R_{c,1990}}}_{\text{Share 3}}, \tag{2.15}$$

where L denotes all workers and R denotes robot-complementary workers. This shock is the sum of a shift-share variable over country origin groups indexed by g. For each country origin group g, the individual components of the shift-share variable are defined as follows. The "shift" component represents the total number of immigrants from country origin group g that arrived between 1990 and year t. Share 1 represents the share of immigrants from g living in commuting zone c in year 1990, excluding robot-complementary workers in the calculation. Share 2 represents the share of immigrant workers from g who work in robot complementary occupations in 1990. Multiplying together the shift component, share 1, and share 2 delivers the imputed number of robot-complementary workers from country origin group g that arrived in commuting zone c between year 1990 and year t, based on patterns of geographical settlement and occupation composition in the baseline year, 1990. Finally, multiplying by share 3 normalizes the variable by the total number of robot-complementary workers in commuting zone c in 1990. The final product thus captures the change in share of robot-complementary workers due to immigrant inflows based on 1990 settlement patterns. I calculate this variable using publicly available samples of the Decennial Censuses and American Community Surveys, provided by IPUMS (Ruggles *et al.*, 2020).

Then, for each firm i, I aggregate the commuting zone-year level variable $z_{c,t}$ over the commuting zones in which the firm operates in 1992, $\mathcal{C}^i_{1992}$, to obtain the firm-level instrument Z^i_t defined as:

$$Z^i_t = \log\left(\sum_{c \in \mathcal{C}^i_{1992}} w^i_{c,1992} z_{c,t-1}\right). \tag{2.16}$$

I lag the immigrant shock $z_{c,t}$ defined in Equation 2.15 by one year to allow time for the increase in labor supply to have an effect on robot adoption. When calculating the weighted sum of $z_{c,t}$ over commuting zones in $\mathcal{C}^i_{1992}$, I define the weights $w^i_{c,1992}$ to be the employment share of firm i in commuting zone c. Lastly, I apply the logarithmic transformation to correct right skewness in the constructed weighted sum.

2.4.4 Results

I first present evidence for relevance of the constructed IV in first-stage regressions. Column (1) presents results under the baseline specification. The instrument is a strong predictor of robot adoption on the firm-level (p-value < 0.001) with an F-stat of 35.34. The estimated coefficient implies that a 10% larger shock to local robot-complementary workers due to immigrant inflows is associated with a 0.7% increase in the probability of robot adoption. Given that standard errors are clustered, the appropriate indicator of first-stage relevance would be the Kleibergen-Paap F-statistic, which is equal to the F-statistic of the first stage regressions in my case since there is only a single endogenous variable (robot adoption) (Kleibergen and Paap, 2006). The F-statistic is greater than a standard threshold of 10, which suggests that the constructed instrument is relevant as a proxy for robot adoption[13]. In columns (2) - (4), I estimate variants of the first-stage regression controlling for less restrictive fixed effects. While the first-stage F-statistic is consistently larger, the magnitude of the coefficients is similar to column (1). This indicates that controlling for additional factors - industry-CZ fixed effects and baseline trend controls - does take away some explanatory power of the instrument, but does not affect relevance of the instrument in a significant way.

Table 2.4: *First-stage IV Regressions*

	Dependent Variable: Robot Adoption			
	(1)	(2)	(3)	(4)
Instrument	0.0742^{***}	0.0845^{***}	0.0848^{***}	0.0734^{***}
	(0.0125)	(0.0118)	(0.0123	(0.0118)
Industry-Year FEs	✓	✓	✓	✓
CZ-Year FEs	✓	✓	✓	✓
Firm FEs	✓	✓	✓	✓
Industry-CZ FEs	✓		✓	
Baseline Trend Controls	✓			✓
Observations	3360000	3360000	3360000	3360000
First-stage F-statistic	35.34	51.18	46.05	38.94

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of first-stage IV regressions, where the dependent variable is robot adoption, and the instrument is the constructed firm-level instrument discussed above. The sample used is all firm-year observations for US manufacturing firms between 1992 and 2016. Regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, and baseline trend controls as indicated by the checkmarks. Standard errors are clustered at the firm level. All observation counts are rounded for disclosure avoidance.

With the instrument in hand, I use 2SLS versions of the specification in Equation 2.13 to test predictions of the model regarding effects of robot adoption on firm-level outcomes. For all regressions, I supplement IV results with OLS results.

First, the model predicts that robot adoption leads to an increase in sales. Column (1)

[13] Recent work such as Lee *et al.* (2021) argues that the threshold of 10 might not be sufficient as criterion for trusting t-ratio inferences in 2SLS regressions. In subsection 2.4.5, I provide a more detailed discussion on what these new results imply for my key IV estimates.

of Table 2.5 confirms that prediction. In addition, robot adoption also causes a significant increase in exports and labor productivity. These results confirm that firms adopting robots enjoy a significant boost in productivity, and subsequently expand their scale of operations. OLS regressions in columns (4)-(6) similarly indicate a positive effect.

Table 2.5: *Regressions for sales, exports, and productivity*

	IV			OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
	Sales	Exports	Labor Prod.	Sales	Exports	Labor Prod.
Robot Adoption	1.918**	4.802***	1.764***	0.264***	0.518***	0.104**
	(0.967)	(2.332)	(0.422)	(0.0306)	(0.0667)	(0.0186)
Full FEs	✓	✓	✓	✓	✓	✓
Baseline Trend Controls	✓	✓	✓	✓	✓	✓
Observations	1540000	3360000	1540000	1540000	3360000	1540000
First-stage F-stat	24.04	35.34	24.04			

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of IV regressions, where the dependent variables are Log(sales), IHS(exports), and Log(Labor Productivity), and robot adoption is proxied by the firm-level instrument. Labor productivity is defined as total sales divided by total employment. The sample used is all firm-year observations for US manufacturing firms between 1992 and 2016. All regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as baseline trend controls, Standard errors are clustered at the firm level. Observation counts are rounded for disclosure avoidance. "First-stage F-statistic" indicates the Kleibergen-Paap F-statistic for the IV regressions, which conincide with the F-statistic of the first-stage regression under the case of clustered standard errors and a single endogenous variable. All observation counts are rounded for disclosure avoidance.

The model also predicts that robot adoption leads to an increase in both input imports and output imports. This is confirmed in Table 2.6. IV regressions in columns (1) and (2) indicate that the impact of robot adoption on both types of imports is positive and significant. Furthermore, the comparison of the effect between input and output imports is also consistent with the theory, which predicts that upstream imports should increase more than downstream imports. This result provides support for the notion that robots have a comparative advantage in downstream production.

Table 2.6: *Regressions for Input & Output Imports*

	IV		OLS	
	(1)	(2)	(3)	(4)
	Input Imp.	Output Imp.	Input Imp.	Output Imp.
Robot Adoption	3.907***	0.840***	1.022***	1.143***
	(0.877)	(0.281)	(0.110)	(0.111)
Full FEs	✓	✓	✓	✓
Baseline Trend Controls	✓	✓	✓	✓
Observations	3360000	3360000	3360000	3360000

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This table displays results of IV regressions, where the dependent variables are IHS(Input Imp) and IHS(Output Imp), and robot adoption is proxied by the firm-level instrument. Input imports are defined as the imports of intermediate inputs; output imports are defined as the imports of final outputs. The sample used is all firm-year observations for US manufacturing firms between 1992 and 2016. All regressions include industry-year, CZ-year, industry-CZ, and firm fixed effects, as well as baseline trend controls, Standard errors are clustered at the firm level. Observation counts are rounded for disclosure avoidance. "First-stage F-statistic" indicates the Kleibergen-Paap F-statistic for the IV regressions, which coincide with the F-statistic of the first-stage regression under the case of clustered standard errors and a single endogenous variable. All observation counts are rounded for disclosure avoidance.

2.4.5 Discussion and Robustness Checks

One obvious feature of the empirical results presented in subsection 2.4.4 is that IV estimates are larger than OLS estimates for almost all variables, except manufacturing employment and wage. I provide two possible reasons why this is the case. First, as discussed earlier, while existence of pre-trends might lead to an upward bias in the OLS estimates, there might also be a downward bias associated with OLS if firms adopt robots in expectation of *negative* shocks in the future. The difference between IV and OLS estimates might in part capture this downward bias. Second, from an econometric point of view, IV and OLS regressions are intended to estimate different objects. OLS regressions estimate the average treatment effect (ATE) for the population of firms, whereas IV regressions estimate the local average treatment effect (LATE) for *compliers*, defined as firms which switch from not adopting to adopting in response to the labor supply shock (Imbens and Angrist, 1994). Given that the sample of compliers is likely small, they might have particularly large responses relative to the average firm due to idiosyncratic reasons, hence driving up the magnitude of the IV estimates. This is especially plausible because as shown in Table 2.1, robot adopters are indeed much larger than non-adopters in various dimensions.

While current practice mostly relies on a threshold of 10 for first-stage F statistic to assess weak instruments, Lee *et al.* (2021) has argued that the true threshold for t-ratio inferences in 2SLS regressions to be valid is 104.7, which is substantially higher than 10. Therefore, when the first-stage F statistic is below 104.7, critical values for, e.g. a 5 percent t-test for second stages estimates, should be adjusted accordingly. Following their methodology, I compute the appropriate critical values for a 5-percent test based on my first-stage F statistic. Contrasting with the standard value of 1.96, I obtain critical values ranging between 2.29 and 2.48.

This procedure does not change any of the point estimates in Table 2.5 and Table 2.6, but it does affect statistical significance of the estimated coefficients. Specifically, the positive effect of robot adoption on sales and exports is no longer significant at the 5 percent level. Meanwhile, statistical significance for the effect of robots on labor productivity, input imports, and output imports still remains. I thus conclude that accounting for the potential impact of a weak instrument does not pose serious threats to the qualitative implications of the empirical analysis presented in subsection 2.4.4, even though it does undermine preciseness of the quantitative estimates.

In addition to the main results presented in subsection 2.4.4, I also re-run IV regressions for a set of alternative import definitions. My current definition of input and output imports excludes all types of machinery (HS 84,85), to alleviate concerns that robot adopters might also be more inclined to use other types of capital. Qualitative implications of my results remain unchanged if I only exclude imports of products in the same 4-digit HS categories as industrial robots, or if I include all machinery imports.

2.5 Quantitative Exercise

section 2.4 presents detailed empirical estimates of the effects of robot adoption at the firm level. However, it is not obvious whether and how these micro effects translate into changes in aggregate outcomes. In this section, I tackle these questions using a quantitative exercise based on cross-sectional data of U.S. manufacturing firms in 2012.

In subsection 2.5.1, I start by setting up an extended version of the model introduced in subsection 2.2.1. Compared to the benchmark model, I allow robots to be useful in both input production and assembly, and incorporate richer substitution patterns between robots and offshoring. In subsection 2.5.2, I describe the methodology I use to calibrate the model. Specifically, I set a few parameters of the model externally, and use model-based regressions and a simulated method of moments (SMM) to estimate the remaining parameters. I discuss the parametric estimates and their implications in subsection 2.5.3.

2.5.1 Extended Model

When building the benchmark model in Section 2.2.1, I make a few simplifying assumptions. Most notably, robots are assumed to be useful in assembly only, and not at all in input production. I also choose to shut down the extensive margin of imports in stage 2 by assuming that every firm produces a single good at the foreign assembly plant.

In this section, for the purpose of the quantitative exercise, I present an extended version of the model with two major amendments. First, the extended model allows robots to be used in (domestic) input production. Specifically, I assume that the cost of input production under robots is c_{1R}, and that paying a single fixed cost of robot adoption f_R grants a firm with access to robots in both stages of production. Second, the extended model incorporates an additional extensive margin of adjustment for downstream imports. Instead of assuming that each firm produces a single good at the foreign assembly plant, I allow each firm to endogenously choose the number of products to produce at the foreign assembly plant,

n_2, subject to a fixed cost schedule $f_2(n_2)$. Note that the fixed cost of sourcing in stage 2 is distinct from the fixed cost schedule in stage 1 introduced earlier, $f_1(n_1)$.

The firm-level maximization problem under the extended framework can be characterized as

$$\max_{n_1,n_2,I_R} B\varphi^{\sigma-1}\left[\left(c_{1d}^{1-\rho}+I_R\cdot c_{1R}^{1-\rho}+n_1(\tau c_{1f})^{1-\rho}\right)^{\frac{\alpha}{1-\rho}}\left(\frac{c_{2d}}{1+I_R(\beta_{2R}-1)}\right)^{1-\alpha}\right]^{1-\sigma}+$$
$$n_2 B\varphi^{\sigma-1}\tau^{1-\sigma}\left[\left((\tau c_{1d})^{1-\rho}+I_R\cdot c_{1R}^{1-\rho}+n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha}{1-\rho}}\left(c_{2f}\right)^{1-\alpha}\right]^{1-\sigma}-f_1(n_1)-f_2(n_2)-I_R\cdot f_R. \tag{2.17}$$

Compared to the benchmark model, the extended model features a more complex set of interactions between robots and offshoring at the firm level. Given that robots are used in input production as well, robot adoption now has an additional negative effect on upstream imports due to a substitution effect. Since firms can additionally adjust the number of imported outputs, the change in downstream imports also incorporates a new channel of adjustment, which feeds back to affect upstream imports. These newly added forces makes it considerably more difficult to extract clean predictions from the extended model relative to the benchmark model.

I thus refrain from an in-depth analysis of the extended model. However, it is more feasible to evaluate complex mechanisms of the extended model using a quantitative approach. A quantitative estimation of the model not only speaks to firm-level interactions between robot adoption and offshoring, but also allows me to evaluate the effects of robots on non-adopters and aggregate outcomes, and conduct counterfactual exercises.

For the purpose of structural estimation, I close this discussion by considering industry equilibrium of the model. Assume that consumers in the domestic country spend a constant share η of income on the manufacturing sector, and the remaining share $1-\eta$ of income on a perfectly competitive non-manufacturing sector that uses labor as the only input. Furthermore, assume that $1-\eta$ is large enough to pin down the wage level. Given that wage is set as the numeraire, we can thus exogenously pin down the wage level to be equal to 1.

Within the manufacturing sector, assume that all potential entrants need to pay an entry cost f_e to obtain a productivity draw from the distribution $G(\varphi)$. Since productivity draws are realized after firms pay the entry cost, industry equilibrium is governed by the free entry condition, which says that expected profits should be equal to the entry cost. It can be written as:

$$\left[\int_{\tilde{\varphi}}^{\infty} \pi(\varphi) dG(\varphi)\right] = f_e \tag{2.18}$$

Note that only firms with productivity $\varphi > \tilde{\varphi}$ will obtain positive profits and thus remain in production.

The free entry condition in Equation 2.18 pins down a unique level of market demand B. Given the level of B, we can calculate the revenue and profits for any given productivity level φ, and thus obtain the cutoff productivity $\tilde{\varphi}$ as well. Finally, to close the model, we use the fact that expenditure on the manufacturing sector is a constant share η of consumer income to determine the equilibrium mass of manufacturing firms:

$$N = \frac{\eta L}{\int_{\tilde{\varphi}}^{\infty} R(\varphi) dG(\varphi)} \tag{2.19}$$

2.5.2 Calibration Procedure

Structural estimation is based upon the firm-level maximization problem in the extended model, presented in Equation 2.17. Relative to the theory, I make a few amendments to ease the estimation process. First, I assume that the iceberg trade cost τ is equal to one[14]. Second, I set the domestic costs in both stages to 1, i.e. $c_{1d} = c_{2d} = 1$. This normalization is harmless because the true magnitude of these costs will be absorbed into B.

[14] For estimation purposes, this is equivalent to assuming that there are three countries - domestic, foreign input producer, and foreign assembly location - and the trade cost between any two countries is equal to τ.

Based on these assumptions, the maximization problem can be rewritten as

$$\max_{n_1,n_2,I_R} B\varphi^{\sigma-1}\left[\left(1+I_R\cdot c_{1R}^{1-\rho}+n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha}{1-\rho}}\left(\frac{1}{1+I_R(\beta_{2R}-1)}\right)^{1-\alpha}\right]^{1-\sigma}+ \tag{2.20}$$

$$n_2 B\varphi^{\sigma-1}\tau^{1-\sigma}\left[\left(1+I_R\cdot c_{1R}^{1-\rho}+n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha}{1-\rho}}\left(c_{2f}\right)^{1-\alpha}\right]^{1-\sigma}-f_1(n_1)-f_2(n_2)-I_R\cdot f_R. \tag{2.21}$$

Mapping the model to data, I measure the number of imported inputs, n_1, and the number of imported outputs, n_2, by counting the number of 4-digit HS products imported by a firm within each category. Following Melitz and Redding (2015), I assume the productivity distribution $G(\varphi)$ follows a Pareto distribution with shape parameter κ. I also parameterize the fixed cost schedule by assuming that they take on the same functional forms as the fixed costs of importing foreign inputs in Gopinath and Neiman (2014):

$$f_1(n_1)=\beta_{1m}n_1^{\beta_{1p}},\quad f_2(n_2)=\beta_{2m}n_1^{\beta_{2p}}. \tag{2.22}$$

Under this setup, the model contains the following parameters: ρ, α, σ, κ, c_{1f}, c_{2f}, B, c_{1R}, β_{2R}, f_R, β_{1m}, β_{1p}, β_{2m}, β_{2p}. I estimate these parameters in three steps.

In the first step, I externally set four parameters: the elasticity of substitution across inputs in the composite input good, ρ; the elasticity of substitution across outputs in consumer demand, σ; the share of inputs in the Cobb-Douglas production function faced by all firms, α; and the shape of the Pareto distribution of productivity draws, κ.

The baseline values for these parameters are displayed in Table 2.7. The value of elasticity of substitution between consumer goods, $\sigma=4$, is similar to what Antras *et al.* (2017) estimates for a cross-section of US manufacturing firms in 2007 and falls within the range of estimates in Broda and Weinstein (2006)[15]. The value of elasticity of substitution between inputs, $\rho=4$, is identical to the value used in Gopinath and Neiman (2014) and

[15] Broda and Weinstein (2006) estimates estimate a mean elasticity of 4 and a median of 2.2 at the SITC-3 level.

estimated in Halpern *et al.* (2015). The share of inputs in production, $\alpha = 2/3$, is calculated from the input share of value-added in manufacturing using the 2012 BEA input-output tables for the US. Lastly, I set $\kappa = 4.25$ following Melitz and Redding (2015).

Table 2.7: *External Parameters*

Parameter	Description	Source	Value
σ	Elasticity of substitution across output varieties	Standard	4
ρ	Elasticity of substitution between domestic and foreign inputs	Standard	4
α	Share of inputs in production	2012 BEA I-O Table	2/3
κ	Pareto shape of productivity distribution	Melitz & Redding (2015)	4.25

In the second step, I estimate the cost of foreign inputs c_{1f} and cost of assembly at the foreign plant c_{2f}. To that end, I exploit the relationship between these costs, the number of imported inputs and outputs, and the observed share of input imports and output imports. To understand the logic of identification, consider the cost of foreign inputs, c_{1f}. Although this variable is not directly observed, we know that it determines the value of foreign inputs purchased. Holding the number of imported inputs n_1 fixed, we can directly relate c_{1f} with the total value of input imports, which is observed in the data. An analogous idea applies to output imports. Formally, it can be established from the model that for non-robot adopters,

$$S_1 = \frac{M_{1f}}{M_{1d} + M_{1f}} = \frac{n_1 c_{1f}^{1-\rho}}{1 + n_1 c_{1f}^{1-\rho}}, \tag{2.23}$$

$$S_2 = \frac{M_{2f}}{\text{Domestic Sales} + \text{Foreign Sales}} = \frac{n_2 \left(c_{2f}^{(1-\alpha)(1-\sigma)} \right) \frac{\sigma-1}{\sigma}}{1 + n_2 \left(c_{2f}^{(1-\alpha)(1-\sigma)} \right)}, \tag{2.24}$$

where S_1 and S_2 are input imports normalized by total input expenditure and output imports normalized by total sales, respectively. These two equations can be rearranged to obtain:

$$\log(c_{1f}^{1-\rho}) = \log(S_1) - \log(1 - S_1) - \log(n_1), \tag{2.25}$$

$$\log \left(c_{2f}^{(1-\alpha)(1-\sigma)} \right) = \log(S_2) - \log \left(\frac{\sigma - 1}{\sigma} - S_2 \right) - \log(n_2). \tag{2.26}$$

Given that we observe S_1, S_2 as well as n_1, n_2 for each firm, I estimate the left hand side parameters by taking the sample mean of the right hand side. I find that importing one foreign input reduces composite cost of inputs by 0.3%, whereas offshoring assembly of one output is on average associated with an increase in sales equivalent to 0.3% of domestic sales.

I estimate the remaining parameters $\Theta = [B, c_{1R}, \beta_{2R}, f_R, \beta_{1m}, \beta_{1p}, \beta_{2m}, \beta_{2p}]$ using a simulated method of moments (SMM), where I simulate a large number of firms according to the model and match moments between the simulated firms and the data. Descriptions of the parameters and moment(s) used to target identification of each parameter are listed in Table 2.8. The vector of moments $m(\Theta)$ consists of six groups of moments, which I describe in detail below.

The first moment is the number of firms with sales below median sales in the data, \$1.11 million. The second moment is the share of robot adopters, which helps identify the fixed cost of robot adoption as well as productivity of robots. Directly estimating the productivity of robots is difficult, because the data does not allow us to observe the value of inputs or outputs produced by robots. Therefore, I use the log premium in sales and input expenditures as the third and fourth moments to infer c_{1R} and β_{2R}. Due to selection on productivity, robot adopters will have an ex-ante advantage in size over non-adopters, which is determined by the productivity distribution. However, any additional difference in sales would be driven by how productive robots are. Meanwhile, the effect of robots on input-usage is two-folded: on the one hand, robots increase productivity of adopters and allow them to expand operations; on the other hand, robots substitute for inputs exerting a negative effect on the observed input expenditures. Therefore, the more productive robots are in stage 1, the wider the gap between the input premium and sales premium.

Finally, to estimate the fixed cost schedule for input (output) importing, I use the share of firms importing 0, 1, 2, 3, 4, 5, 6-10, and 10+ inputs (outputs). Note that, although numbers of imported inputs and outputs are discrete in the data, in the simulation I assume firms to choose a continuous n_1 and n_2 to ease computational burden. Therefore, I calculate the final two sets of moments in the simulated data by classifying all firms choosing $n_1 < 1$ as importing 0 inputs, and all firms choosing $n_2 < 1$ as importing 0 inputs, and so on for the rest.

Table 2.8: *SMM Parameters*

Parameter	Description	Source of Identification
B	Demand factor	Median sales among all manuf firms
c_{1R}	Robot capability in input production	Input premium of robot adopters
β_{2R}	Robot capability in assembly	Sales premium of robot adopters
f_R	fixed cost of robot adoption	Share of robot adopters
β_{1m}, β_{1p}	Fixed cost schedule for input imports	Distribution of # of imported inputs
β_{2m}, β_{2p}	Fixed cost schedule for output imports	Distribution of # of imported outputs

Practically, the estimation algorithm is carried out as follows. First, I obtain 1200 draws of core productivity φ from the Pareto distribution with shape parameter κ. I use a stratified random sampling technique to oversample higher productivity draws. Given that productivity is the only source of heterogeneity across firms, these 1200 productivity draws constitute the entire simulated sample of firms. Second, I make an initial guess for the vector of parameters $\hat{\Theta}$. Based on the parameter guess, for every firm, I solve the firm-level maximization problem where each firm chooses n_1, n_2, and I_R to maximize profits. Aggregating firm-level outcomes based on their optimal choices, I calculate the set of moments $m(\hat{\Theta})$ conditional on the parameter guess. Finally, I use a bounded Nelder-Mead simplex search method to update and calibrate a value of $\hat{\Theta}$ that minimizes the objective function, determined by the distance between moments of the simulated data and moments calculated from the data:

$$\Theta^* = \arg\min_{\Theta}[m - m(\hat{\Theta})]^T \mathbf{W}[m - m(\hat{\Theta})] \tag{2.27}$$

where m is the vector of moments calculated from the data and $\mathbf{W}$ is a weighting matrix. In the baseline estimation, I use an identity matrix as the weighting matrix, so all moments are weighted equally[16].

2.5.3 SMM Results

The estimated parameters are displayed in Table 2.9. Standard errors have been calculated using 25 bootstrapped samples and are pending data release. The estimated values of c_{1R} and β_{2R} imply that, holding all else equal, for the median firm, adopting robots leads to a 10.2% reduction in the cost of producing intermediates, and a 33.7% reduction in the cost of assembly. These cost reductions in stage 1 and 2 translate into a 6.9% and 12.8% reduction in total costs, and a 24.0% and 50.9% increase in total sales, respectively. In other words, slightly less than two thirds of gains in overall productivity from robot adoption can be attributed to the use of robots in assembly, and the productivity of robots in assembly is more than three times as large as their productivity in input production. This comparison confirms our hypothesis that robots have a comparative advantage in downstream production.

According to the estimated value of f_R, robot adoption is associated with a fixed cost of \$8.2 million. This is substantially higher than the average cost of purchasing robots observed in the data (\$1.1 million), suggesting that in addition to the sheer cost of purchasing the machinery, there are significant costs associated with the process of installing and integrating robots. The estimated fixed cost is almost prohibitively high for the average firm: to give a perspective, it is more than six times larger than annual sales of the median firm, which provides an explanation for why levels of robot adoption remain low despite potential productivity gains. However, the fixed cost is equal to only 8.8% of the sales of the median *robot adopter*.

The estimated values of $\beta_{1m}, \beta_{1p}, \beta_{2m}, \beta_{2p}$ indicate that the fixed cost of input importing and output importing can be written as $f_1(n_1) = 0.0019n_1^{2.051}$ and $f_2(n_2) = 0.0011n_2^{2.631}$. The function for fixed cost of input importing implies similar ranges of fixed costs as in

[16] In an undisclosed analysis, I use the optimal weighting matrix under 2-step GMM, i.e. the estimated variance-covariance matrix of the moments, and results do not change in a meaningful way.

Gopinath and Neiman (2014) and Halpern *et al.* (2015).The estimates also suggest that the fixed costs of foreign assembly are generally higher than those of importing inputs: for example, importing 10 inputs requires a fixed cost of $214,000, whereas importing 10 outputs requires a fixed cost of $469,000.

Table 2.9: *Estimated Parameters*

B	c_{1R}	β_{2R}	f_R	β_{1m}	β_{1p}	β_{2m}	β_{2p}
0.172	1.381	1.514	8.212	0.0019	2.051	0.0011	2.631
(0.006)	(0.105)	(0.123)	(0.926)	(0.0003)	(0.184)	(0.0002)	(0.207)

Notes: This table reports estimates for the vector of parameters Θ. Estimation uses the universe of manufacturing firms in 2012. Standard errors have been computed using results from 25 bootstrap samples and are pending data release.

In terms of model fit, the simulated model closely matches moments targeted in estimation. Table 2.10 shows that sales of the median firm are extremely close between the data and the model ($1.11 million v.s. $1.13 million). The model also matches the share of robot adopters (0.39% v.s. 0.4%), and the log premium of robot adopters in sales and input expenditure. Regarding the distribution of input and output imports across firms, Figure 2.2 illustrates that the model replicates the general distribution in the data, although it slightly overestimates the share of firms importing 0 or 1 inputs and underestimates the share of firms importing 6 to 10, or more than 10 inputs.

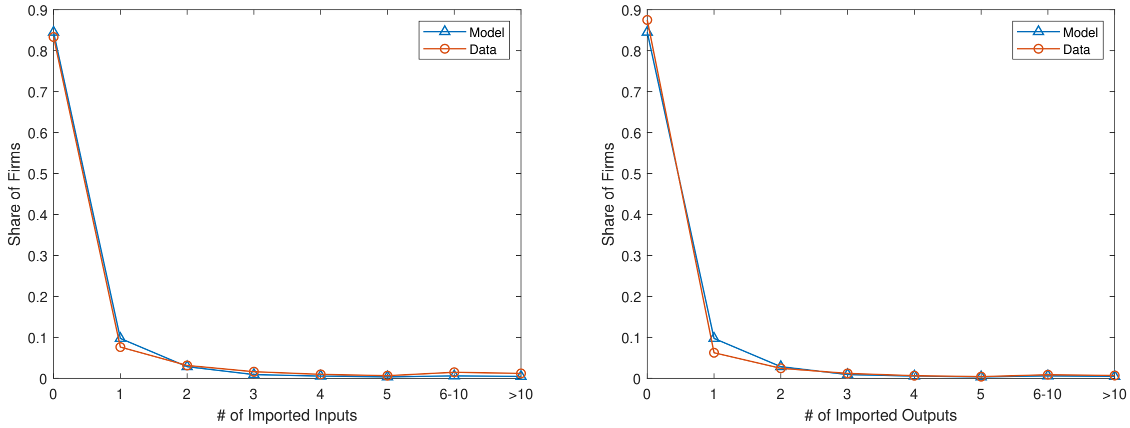
Table 2.10: *Moments: Model vs. Data*

Moment	Data	Model
Median Sales ($ Millions)	1.11	1.13
Share of Robot Adopters	0.0039	0.0040
Sales Premium of Robot Adopters	4.745	4.753
Inputs Premium of Robot Adopters	4.477	4.473

Finally, the simulated model also performs reasonably well in matching untargeted moments. In both the data and the model, robot adopters account for around 20% of total sales. In aggregate, input imports in the simulated model are four times as large as output imports, whereas they differ by a factor of 4.5 in the data.

Figure 2.2: *Distribution of imported inputs and imported outputs, model vs. data*

(a) *Imported inputs* **(b)** *Imported outputs*



2.6 Counterfactual Exercises

In this section, I utilize the model to conduct three counterfactual experiments. In the first counterfactual, I take advantage of the heterogeneous firm framework to analyze the impact of the rise of robots between 1992 and 2012 on adopters and non-adopters, as well as its effect on aggregate outcomes. In the second counterfactual, I hypothesize an increase in the productivity of robots in input production and examine how it affects upstream and downstream imports differently. In the third and final counterfactual, I evaluate costs and benefits of three possible robot policies: a robot investment tax, a value-added tax on assembly robots, and a robot subsidy.

2.6.1 Counterfactual 1: Rise of robots between 1992 and 2012

In this counterfactual, I examine the rise of robots between 1992 and 2012 and its implications. Specifically, starting from the calibrated model in 2012, I conduct a reverse counterfactual to calibrate the *decrease* in productivity of robots in input production and assembly relative to the baseline in 2012 such that the fraction of firms adopting robots matches the observed fraction in 1992. I then compute firm-level and aggregate outcomes in the equilibrium prior to the increase in robot productivity. This exercise allows me to examine how the rise of robots between 1992 and 2012 shapes firm-level decisions as well as expenditure on imports and domestic workers in the entire manufacturing sector.

I assume that the proportional increase in the productivity of robots is identical across input production and assembly. Under this assumption, in order to explain the degree of change in the share of robot adopters between 1992 and 2012, robot productivity needs to have increased by 55.7%. More details on how I calibrate this change is included in subsection A.3.1.

Given the equilibria in 1992 and 2012, we can compare outcomes across 1992 and 2012 to examine the effect of the rise of robots. I separate all firms into three groups: always-adopters, defined as firms which adopt robots in both 1992 and 2012; new adopters, which switch from not adopting robots in 1992 to adopting in 2012; and non-adopters, which adopts in neither year. I evaluate the change in aggregate outcomes on the industry level, as well as the change in the same outcomes for these three groups separately.

The calibrated changes are shown in Table 2.11. The first three columns indicate changes in corresponding outcomes within the three groups of firms, while the fourth column indicates change in aggregate outcomes. First, in light of the empirical results estimating the effect of robot adoption on firms that switch from not adopting to adopting, let us focus on new adopters. Consistent with the reduced-form findings in Table 2.5 and Table 2.6, the counterfactual exercise suggests that new adopters increase their sales and imports in both stages of production. Notably, the increase in input imports is more than three times as large as the increase in output imports, again supporting the notion that robots have a comparative in downstream production. Furthermore, new adopters also increase their expenditure on domestic inputs. Since my model fixes the wage level, this can be interpreted as new adopters employing a higher number of domestic workers in the input sector. In contrast, new adopters decrease their employment in domestic assembly by a substantial proportion.

For always-adopters, all outcomes move in the same direction as for new adopters, except the magnitude is consistently smaller. This is because always-adopters already adopt robots in 1992, so the size of the productivity gains is smaller for them than for new adopters. Meanwhile, non-adopters contract in all dimensions due to within-industry competition, driven by the increase in productivity among the two groups of adopters. In the model, these competitive forces are driven by a decrease in the demand factor B.

In aggregate, since I assume consumer expenditure on the manufacturing sector to be constant, total sales do not change. My exercise thus sets aside the effect of robots on aggregate productivity and focuses on the change in expenditure on imports and domestic workers as *a share of total output*. The rise of robots between 1992 and 2012 is associated with a 20.2% increase in input imports and a 3.7% decrease in output imports, which implies a 14.6% increase in total imports. Given that non-adopters contract both types of imports to a similar extent, the difference between upstream and downstream imports is mainly driven by the two groups of adopters, which increase their input imports substantially more than output imports. This suggests that, contrary to the common perception, the rise in robots between 1992 and 2012 has in fact *promoted* importing in the US manufacturing sector, though the effect exhibits significant heterogeneity across different *types* of imports in the production process. Given that countries have tremendously different patterns of specialization in input production v.s. assembly, this heterogeneity implies that as sourcing locations, countries gain or lose from robot adoption in the U.S. to vastly different extents. For example, robots may reduce imports from countries with a strong comparative advantage in assembly, despite the overall *increase* for all countries. My framework thus reconciles studies suggesting a positive effect of robots on aggregate trade (Hallward-Driemeier and Nayyar, 2019; Artuc *et al.*, 2020) with studies finding a negative impact of robot adoption in the U.S. on imports from developing countries including Mexico and Colombia (Artuc *et al.*, 2019; Kugler *et al.*, 2020; Faber, 2020).

Turning attention to domestic outcomes, I find that, as a result of the rise in robots, the consumer price index decreased by 3.4%. While this reflects an increase in output of the manufacturing sector, it is not necessarily an accurate measure of the welfare effects of robots, because wages are fixed in our industry equilibrium model. These efficiency gains are also overshadowed by the fact that domestic employment in manufacturing decreased by 13.7%. This decrease is concentrated in assembly workers (22.5%) but also impacts workers engaging in input production (9%). In comparison, Fort *et al.* (2018) estimates that US manufacturing employment decreased by 25% between 1992 and 2012. This suggests that the rise in automation technologies such as industrial robotics likely played an important role in the decline of manufacturing employment. Moreover, consistent with their discussion, effects of technology and trade on domestic workers are not separate from each other. Instead, as my model suggests, a technology shock like the rise in robots may promote trade, which amplifies the negative effect on domestic workers.

Table 2.11: *Impact of the rise in robots, 1992-2012*

	Always-adopters	New adopters	Non-adopters	Overall
Share of Firms (%)	0.022	0.378	99.6	1
Total Sales	0.388	0.647	-0.099	0
Total Imports	0.209	0.446	-0.174	0.146
Input Imports	0.250	0.505	-0.180	0.202
Output Imports	0.026	0.158	-0.156	-0.037
Domestic Employment	-0.119	-0.060	-0.098	-0.137
Domestic Inputs	0.215	0.237	-0.097	-0.090
Domestic Assembly	-0.503	-0.521	-0.099	-0.225
Consumer Price index				-0.034

Notes: "Always adopters" are firms that adopt robots in both 1992 and 2012 according to the calibrated model. "New adopters" are firms that do not adopt robots in 1992 but switch to adopting robots in 2012. "Non-adopters" are firms that do not adopt robots in either year.

An important feature of my model is that it incorporates decisions of offshoring and robot adoption in a unified framework. This feature introduces an additional channel of adjustment - the choice of imported inputs and outputs - through which robots affect aggregate outcomes such as imports. I explore the role of these extensive-margin import adjustments in subsection A.3.3. I find that around a third of the aggregate increase in imports is driven by firm-level adjustments at the extensive margin. However, fixing sourcing strategies does not have a quantitatively significant impact on the changes in employment or the price index.

2.6.2 Counterfactual 2: Rise in robot productivity in inputs

Counterfactual 1 reveals that, at least between 1992 and 2012, robots have led to an increase in total imports of US manufacturing firms. However, this is contingent on the fact that robots are particularly capable in assembly relative to input production. Given the rapid developments in robot technologies, one might wonder what happens if robots become equally productive in stage 1 as in stage 2. Would this reverse the trade-inducing effects of robots, especially for input imports?

In counterfactual 2, I explore this possibility by studying the consequences of a hypothetical situation: an isolated increase in productivity of input production, given by a decrease in the cost of input production under robots, i.e. c_{1R}. I calibrate a decrease in c_{1R} such that robot adoption leads to a proportionally equal reduction in the unit cost of inputs as the reduction in the unit cost of assembly. I find that a 54.1% decrease in c_{1R} is required.

The impact of such a rise in productivity of robots in inputs is summarized in Table 2.12. Similar to counterfactual 1, which assumes an leveled increase in robot technology across production stages, the directed technical change in counterfactual 2 also leads to an increase in sales of adopters, and a decrease in the consumer price index of the manufacturing sector. However, effects on imports are vastly different. Robots becoming productive in input production induces a significant increase in output imports among adopters and in aggregate, but greatly undermines input imports due to the substitution effect. Aggregating these two effects, the value of total imports decrease by 34.3%. In other words, the positive relationship between robots and offshoring may not be sustained if the technology of industrial robots grows in a different direction. In light of counterfactual 1, which suggests that countries specializing in assembly may particularly lose out from robot adoption in the U.S., counterfactual 2 demonstrates that this cross-country comparison is contingent on technological characteristics of industrial robotics.

Regarding domestic workers, a rise of robots in input production is again associated with a decline in manufacturing employment, but this time workers engaging in input production suffer more.

Table 2.12: *A hypothetical rise in robot productivity in input production*

	Always-adopters	New adopters	Non-adopters	Overall
Share of Firms (%)	0.400	0.700	98.9	1
Total Sales	0.273	1.661	-0.205	0
Total Imports	-0.327	0.021	-0.328	-0.343
Input Imports	-0.487	-0.333	-0.356	-0.506
Output Imports	0.349	1.48	-0.309	0.235
Domestic Inputs	-0.371	-0.188	-0.202	-0.286
Domestic Assembly	0.269	-0.232	-0.204	-0.202
Price index				-0.073

Notes: "Always adopters" are firms that adopt robots both before and after the increase in robot productivity. "New adopters" are firms that do not adopt robots in the baseline model but switch to adopting robots after the increase in robot productivity. "Non-adopters" are firms that do not adopt robots in either scenario.

2.6.3 Counterfactual 3: Robot tax vs. robot subsidies

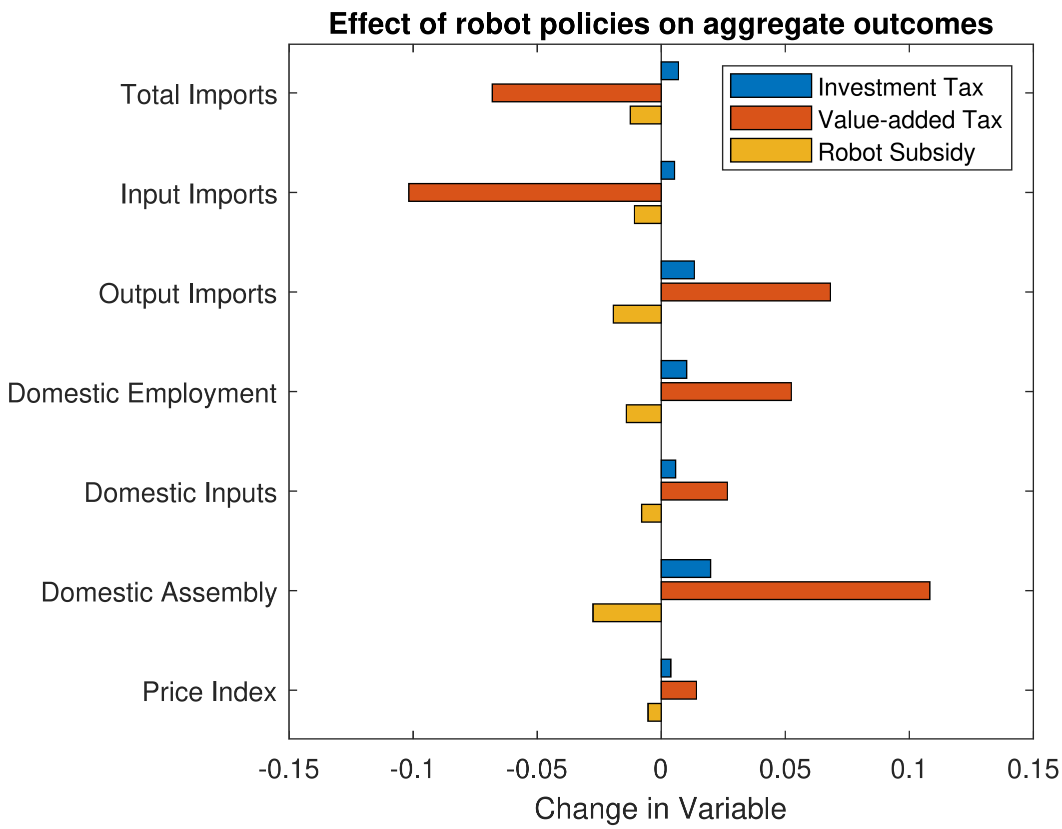
In the final counterfactual, I turn to evaluate several possible robot policies. In 2017, concerns of the negative effects of robots on workers prompted the European Parliament to vote on a proposal for a robot tax. Although the proposal was eventually rejected, discussions over this possibility has continued gaining attention. For example, Acemoglu *et al.* (2020c) argues that the U.S. tax code systematically favors capital, and advocates for an automation tax. Practical options for a robot tax include a standard tax on robotic investments, and a value-added tax on specific types of robots to help workers particularly vulnerable to being displaced by automation. On the other hand, countries like China have taken the opposite approach of subsidizing robots in an effort to boost productivity and increase competitiveness of the manufacturing sector (Lin, 2018; Cheng *et al.*, 2019).

In light of these divergent policy possibilities, I use the model to consider three types of robot policies. The first policy is a robot investment tax, which increases the fixed cost of adopting robots by 30%. This is the standard idea of a robot tax that makes robots more costly as an investment. The magnitude of 30% is close to the optimal level of capital tax proposed by Acemoglu *et al.* (2020c), 26.65%. The second policy is a 25% value-added tax on assembly robots. This policy essentially decreases the productivity of robots used in assembly, in the hope of benefiting domestic assembly workers. I set a 25% tax to match the effective tax rate on workers with top 1% income (McClelland and Airi, 2021). The third policy is a subsidy on robot investments that decreases the fixed cost of robot adoption. In practice, this can be implemented through, for instance, an investment tax credit. I set the size of subsidy to be 30%, symmetric with the robot investment tax.

For convenience, I refer to these policies as robot tax 1, robot tax 2, and robot subsidy. Figure 2.3 displays the change in various aggregate outcomes following each of the three policies. The first major takeaway is that policies targeting the fixed costs of robot adoption (robot tax 1 and robot subsidy) have relatively small effects on aggregate outcomes. A 30% tax on investments in robots only results in in 1% increase in domestic employment of manufacturing workers, and the effects on other outcomes are similarly inconsequential. This is the case because a tax on fixed costs only in a direct way affects firms at the margin of adopting, which are the smallest and least productive firms *among robot adopters*. Even though the tax induces these firms to stop adopting robots, the larger, more productive robot adopters, which play a more important role in the economy, continue adopting, which explains why the aggregate effects of such a tax are small. A similar logic applies for robot subsidies.

On the other hand, a 25% value-added tax on assembly robots seems to have a much more substantial effect. It in part reverses the effects of robots on imports, reducing total imports by 6.8%. Most importantly, it leads to an increase in domestic employment, particularly in assembly (10.8%). While this tax does accomplish the goal of helping assembly workers, this achievement comes at the cost of a 1.4% higher consumer price index. Therefore, policymakers face a trade-off between efficiency gains associated with robots and welfare of manufacturing workers.

Figure 2.3: *Model-implied change in aggregate outcomes following robot policies*



Notes: this figure plots the change in seven aggregate outcomes following three types of robot policies. Robot tax 1 is a robot investment tax that increases the fixed cost of adoption by 30%. Robot tax 2 is a 25% value-added tax on assembly robots. Robot subsidy decreases the fixed cost of adoption by 30%.

2.7 Conclusion

This paper investigates the impact of industrial robots on US manufacturing firms. Centered on the hypothesis that robots have a comparative advantage in assembly relative to input production, I construct a stylized model to highlight *complementarities* between robot adoption and offshoring at the firm level and *heterogeneity* in the effect of robots across upstream and downstream imports. My empirical analysis based on a novel instrumental variables strategy confirms the importance of both forces. At the firm level, robots lead to an increase of both imports of intermediate inputs and imports of assembled outputs, with the former being significantly larger. A quantitative exercise shows that these firm-level changes translate into similar movements in aggregate outcomes. In the manufacturing sector, the rise of robots between 1992 and 2012 is associated with a significant increase in input imports, and a modest decrease in output imports, which implies an increase in overall imports. These results suggest that, contrary to the popular notion that robots lead to reshoring and a decline in trade, they have been trade-inducing for US manufacturing firms. The rise of robots also leads to a decrease in the consumer price index of the manufacturing sector. However, these efficiency gains come at the cost of domestic manufacturing workers, especially assembly workers. The calibrated model indicates that the rise of robots can explain a substantial fraction of the decline in US manufacturing employment over a similar time period. These results highlight a trade-off between efficiency gains and welfare of domestic workers from the perspective of robot policies.

Aside from results on the impact of robot adoption, an additional contribution of this paper is to show that exogenous changes in labor supply of robot-complementary workers form a statistically significant predictor of robot adoption. This result complements existing work on identifying occupations exposed to technology by exhibiting the importance of occupations that are *integral* to technology adoption. I believe that the methodology I used to construct the labor supply shock based on immigrant inflows can be extended to study questions related to technology in broader settings.

Chapter 3

The Cost of Banking Deserts: Racial Disparities in Access to PPP Lenders and their Implications[1]

3.1 Introduction

Many government lending programs for small businesses, such as the Small Business Administration (SBA) 7a and 504 programs, are primarily intermediated by banks and credit unions. The main avenue of government support for small businesses during the Covid-19 crisis, the Payment Protection Program (PPP), was also designed to be intermediated by third party lenders who are tasked for taking in loan applications and loan forgiveness applications.[2] This channel of support has its advantages: banking institutions may have better connections with their local communities and more resources to quickly distribute funds than government agencies, which explains the widespread use of third party lenders

[1] Co-authored with David Hao Zhang

[2] PPP loans are structured as loans but are forgivable if the business used at least 60% of the loan for eligible payroll costs over a span of 24 weeks, among other conditions. Therefore, it is more akin to a grant for maintaining employment than a traditional loan.

as distributors of government support.

Nevertheless, we show that over-reliance on established financial institutions can have undesirable distributional implications. In particular, it leaves behind minority communities which have less access to these financial institutions. Taking the PPP program as an example, we show that Zip codes with a greater proportion of Black population have worse access to enrolled lenders. A 10 percentage point increase in the share of Black population in a Zip code correlates with a 1.0% decrease in the likelihood of having any enrolled lender in their Zip code, and, conditional on having at least one, a 4.1% decrease in the number of branches of enrolled lenders in their Zip code. This is both because there are fewer lenders in Black neighborhoods in general, and because the lenders that *are* there are more likely to be small credit unions with no previous relationships with the SBA and are less likely to enroll in the program.

These differences in geographical access to financial institutions, which we find to be particularly severe in lower population (more rural) areas, correlate with the lower take-up of PPP loans in those areas. Furthermore, borrowers in Black neighborhoods that do receive the government support tend to use Fintech lenders and travel further. Therefore, our descriptive results suggest that the unequal distribution of financial institutions across geographical access may be a driver of differential pass-through of government support. As an analogy to food deserts, which are well-defined geographical areas with limited access to healthy and affordable foods[3], we study the disparate distribution of financial institutions in the form of "government support banking deserts," which we characterize as geographical areas with limited access to financial institutions that have sufficient scale and experience to intermediate government support programs for small businesses.

To assess the causal effect of the presence of financial institutions in a neighborhood, we use an instrumental variables approach. More specifically, we use the failure of small community banks, which tend to be acquired by neighboring larger banks, to instrument

[3] Food deserts are defined by the USDA: https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/. Academic studies of food deserts include Walker *et al.* (2010) and Allcott *et al.* (2019).

for the entry of a larger bank which is more likely to enroll in the PPP program. We find that neighborhoods where their community bank failed, all else equal, ended up with more enrolled lenders and higher take-up of PPP loans. The failures of commercial banks in the US are primarily driven by exposure to commercial real estate loans as well as residential mortgage backed securities (RMBS), and not small business loans, as examined in Antoniades (2020). More importantly for the validity of our estimates, to the extent that community bank failures are correlated with an unobserved deterioration of local economic conditions, it would only make our estimates more conservative since areas with worse economic conditions have lower take-up of PPP because businesses in those areas tend to close altogether. We find that Zip codes exposed to past local bank failures have more enrolled branches, with a 10% increase in the number of enrolled branches in a Zip code corresponding to a 1.2% increase in the take-up of PPP loans. Our IV analysis shows that access to financial institutions explains 32% of the difference in PPP take-up among the lower population (more rural) areas.

To be clear, we view the causal effect of geographical proximity to enrolled financial institutions on PPP take-up as capturing more than the physical costs of travelling further to reach an enrolled lender, much like how standard models of trade costs captures more than physical transportation costs.[4] In particular, enrolled financial institutions tend to engage in advertising and promotion of the PPP program both among their existing small business clients and more broadly, which may increase awareness and decrease misconceptions about the program's requirements. Information frictions are an important driver of PPP take-up among small businesses, as shown in Humphries *et al.* (2020). The impact of the banking deserts we study, then, may be due to the alleviation of information frictions in addition to the physical costs of access.

Our paper is related to the literature on the unequal access to bank branches and banking deserts. Earlier studies have documented the existence of areas with few to no

[4] For example, social connectedness has been found to be an important explanation for trade costs increasing with distance, as shown in Bailey *et al.* (2020).

bank branches which are more concentrated in poorer areas, though the correlation between minority presence and banking deserts depends on the definition of "minority" used. In particular, Morgan *et al.* (2016) finds that banking deserts are negatively correlated with income but are less prominent in majority-minority areas, whereas Kashian *et al.* (2018) finds that African Americans are more likely to live in banking deserts, a result corroborated by more geographically isolated studies including Hegerty (2016) and Miller (2015). We contribute to this literature by showing that the disparity in access to lenders is increased if we expand the definition of "banking deserts" to count only lenders that enroll in government support programs for small businesses. In other words, not only do African American neighborhoods have fewer banks and credit unions, the banks and credit unions that do operate there are less likely to be able and willing to serve as an intermediary for government programs for small businesses. In addition, we demonstrate a tangible consequence of unequal access, in the form of lower take-up of government support.

We also contribute to the literature on PPP loans. Granja *et al.* (2020) finds that, due to selective decisions by large lenders on where to approve Round 1 PPP loans, Round 1 PPP loans flowed to areas that are less hit by the crisis. We look at disparities in geographical access to banking rather than selective intermediation by large banks, and focus on PPP enrollment by the end of Round 2 when the problem of selective intermediation is less severe.[5] More importantly, we focus on how differences in the presence of enrolled lenders in a neighborhood cause disparities in access by race, which is a problem distinct from that of larger lenders selectively choosing which business' application to process first. We also use an instrumental variables approach to to assess the causal effects of enrolled branches on take-up. Howell *et al.* (2021) examines the types of lenders used by minority-owned businesses, though not geographical differences in access or their implications.

[5]Indeed, journalists have reported that PPP take-up by Black congressional districts was lower in Round 1 but entirely caught up by the end of Round 2, suggesting that selective intermediation by large banks may be a temporary problem: https://www.bloomberg.com/graphics/2020-ppp-racial-disparity/. The reason we find a persistent racial disparity in take-up at the end of Round 2 is by looking at lower population areas where differences in geographical access to enrolled banks is the greatest, and by conducting analyses at the Zip code level which more closely approximates racial distributions of neighborhoods.

The literature finds mixed effects of PPP takeup on employment, depending on the types of business examined. Granja *et al.* (2020) and Chetty *et al.* (2020) find that the effect of PPP loans on employment is small in aggregate. On the other hand, Bartik *et al.* (2020c) shows that PPP loans did have a large employment effect for a sample of smaller businesses. Since the businesses in the rural, minority areas where the disparities we identify are largest are more likely to be small, our results may have employment implications.

Finally, our paper contributes to the broader literature on business behavior during Covid-19. Bartik *et al.* (2020a) finds that many small businesses are financially fragile, and many businesses (about 43%) temporarily closed during the crisis. Wang *et al.* (2021) finds that bankruptcy takeup is lower during Covid-19 than earlier downturns. In terms of re-opening, Balla-Elliott *et al.* (2020) finds that Covid-19 demand expectations explain a large part of re-opening decisions. Bartik *et al.* (2020b) finds that better educated and higher paid industries are more likely to switch to remote work.

The rest of this paper is organized as follows. Section 3.2 describes our data and summary statistics. Section 3.3 shows our reduced-form descriptive results on the geographical distribution of enrolled lenders and its correlation with the racial composition of neighborhoods. Section 3.4 presents our instrumental variables results. Section 3.5 concludes.

3.2 Data

We use the Federal Deposit Insurance Corporation's Summary of Deposits (SOD) data to get information on the branch locations of banks and the National Credit Union Administration (NCUA)'s Credit Union and Corporate Call Report Data to get information on the branch locations of credit unions. We merge data on financial institutions with PPP data through August 08, 2020 and data from the 7(a) and 502 loan programs released by the SBA.[6] The

[6]We used the March 2, 2021 version of the SBA PPP data in this paper.

data includes information about the names and Zip codes of the businesses that took out the loans as well as the names of the financial institutions. We use a fuzzy matching procedure to match the names of the financial institutions in the PPP data with the corresponding entities in the SOD and NCUA Call report data, using state as tie-breakers. Furthermore, we identify FinTech lenders using a list adapted from Howell *et al.* (2021).

We then combine the data on PPP loans with Zip code Business Patterns (ZBP) data in 2018 from the US Census Bureau to compute the take-up rate of PPP loans by Zip code. The ZBP includes data on the size distribution of businesses on a Zip code level. Take-up rate is calculated for each Zip code i using the following equation[7]:

$$\text{PPP Take-up Rate}_i = \frac{\text{\# of PPP loans}_i}{\text{\# of small businesses}_i}. \tag{3.1}$$

To make the data comparable, we take the following approaches. First, since the ZBP excludes businesses from certain industries as well as self-employed individuals, we remove PPP loan records for businesses with corresponding features.[8] Second, we calculate the zip-level number of small businesses as the number of businesses in each Zip code with less than 500 employees[9]. In addition to the take-up rate, we also use the ZBP to calculate average employment size of small businesses and share of businesses in each NAICS-2 industry, both on a Zip code level.

We obtain Zip-level demographic and geographical information from various data

[7]For a small number of Zip codes, we calculate a take-up rate greater than one. This most likely arises from 2018 ZBP data being outdated compared to the PPP data, which concerns small businesses in 2020. We set the take-up rate of these Zip codes to one.

[8]Specifically, we exclude PPP loans taken by business in the following sectors: crop and animal production (NAICS 111,112), rail transportation (NAICS 482), Postal Service (NAICS 491), pension, health, welfare, and vacation funds (NAICS 525110, 525120, 525190), trusts, estates, and agency accounts (NAICS 525920), office of notaries (NAICS 541120), private households (NAICS 814), and public administration (NAICS 92). We also exclude loans taken by businesses classified as "Self-employed Individuals".

[9]According to the eligibility conditions for PPP loans, posted on the SBA website `www.sba.gov/funding-programs/loans/coronavirus-relief-options/paycheck-protection-program`, small businesses can apply if their have less than 500 employees, or if they meet the SBA industry size standard if more than 500. Since the size standard varies across industries defined at a detailed level, our measure might underestimate the number of small businesses. We correct this bias by controlling for 2-digit NAICS industry shares of small businesses in each Zip code in our analysis.

sources. Racial composition is based on 5-year estimates of the American Community Survey (ACS) from 2014 to 2018 on a census tract level, which we then collapse to a Zip code level using a crosswalk file provided by the Office of Policy Development and Research of U.S. Department of Housing and Urban Development (HUD).[10] Population density for Zip codes is calculated using the 2010 Decennial Census and 2013 U.S. Gazetteer files. Distance between Zip codes is provided by the NBER Zip Code Distance Database. Commuting zones are defined over counties and cover the entire US, and we obtain their definitions and their population from the US Department of Agriculture Economics Research Service. Furthermore, we separate commuting zone populations into three categories:

$$CZ\_cat = \begin{cases} 1 \text{ , if commuting zone population} & \leq 515,013, \\ 2 \text{ , if commuting zone population} & > 515,013 \text{ and } \leq 1,604,457, \\ 3 \text{ , if commuting zone population} & > 1,604,457, \end{cases} \tag{3.2}$$

where the cut-offs correspond to the 50th and 75th percentiles of commuting zone population by Zip code. We picked this cut-off because more than half of US Zip codes are rural,[11] such that a Zip code in an area with a lower population are more rural and more "remote" in the sense of having less people within their commuting area. As an example, Williamstown, MA, a college town located in a rural area, has a Zip code with CZ_cat = 1. Amherst, MA, which is also in a small college town but is in the more populated Five Colleges area, is located in a Zip code with CZ_cat = 2. Cambridge, MA has a CZ_cat = 3.

Finally, we use the Federal Deposit Insurance Corporation (FDIC)'s Failed Bank List to construct our instrument. We take the entire list of bank failures before 2020, which includes the period from October 1, 2000 to December 31, 2019. We then restrict the sample to small banks that have less than or equal to 10 branches before they failed, which are banks that

[10] We focus our analysis at the zip code level because the welfare of businesses in minority neighborhoods may be more relevant to minorities than the owners' race per se. We also conduct a robustness check of our lender type results using the owners' race identified from their names using the algorithm of Ambrose *et al.* (2020).

[11] Source: http://proximityone.com/zip_urban_rural.htm

are less to have a relationship with SBA and are less likely to enroll in the PPP program. The historical failure of these small local banks, conditioned on the 2000 number of small bank branches, are correlated more enrolled branches and more take-up of PPP loans in the Zip code and uncorrelated with the unemployment claim rate in February 2020. Summary statistics for our variables are shown in Table 3.1.

Table 3.1: *Summary statistics of our variables*

	Mean	10th	Median	90th	*N*
Bank Presence					
Has Branch	0.69	0	1	1	28,969
Log(Branch)	1.10	0	1.10	2.40	19,951
Has Enrolled	0.64	0	1	1	28,969
Log(Enrolled)	0.98	0	0.69	2.20	18,520
Small Bank #	1.15	0	1	3	28,969
has_failed_branch	0.04	0	0	0	28,969
Demographics					
black_ratio	0.09	0.00	0.03	0.28	28,864
Log(Establishments)	4.25	1.79	4.22	6.71	28,969
Log(Population)	8.17	6.04	8.26	10.38	28,864
Population Density	4.81	2.08	4.43	8.16	28,175
PPP					
Takeup	0.60	0.33	0.60	1	28,969
Remote	0.12	0	0	1	4,696,969
log(Miles)	0.76	0	0	3.04	3,808,316

Note: this table shows the summary statistics of our Zip code level and PPP loan level samples.

3.3 Descriptive Analysis

3.3.1 Access to lenders

As a first step, we investigate if there is differential access to financial institutions associated with racial composition of Zip codes. The primary specification is

$$Y_{ic} = \alpha + \beta \text{Black Ratio}_i + \mathbf{X}_i\delta + \gamma_c + \varepsilon_{ic} \tag{3.3}$$

where Y_{ic} is the outcome variable of Zip code i in country c; Black Ratio$_i$ is the share of Black population in Zip code i; $\mathbf{X}_i$ is a vector of Zip code characteristics; and γ_c is a county-level fixed effect. For the outcome variable Y_{ic}, we use a dummy variable of whether the Zip code contains any lender, as well as log of the number of branches (of any lender). We thus examine racial disparity in both extensive and intensive margins of access to financial institutions. To control for other factors correlated with racial composition that might affect access (e.g. industrial composition and size), we include the following Zip code-level variables in the control vector $\mathbf{X}_i$: log number of establishments and squared, log population and squared, population density, and share of small businesses within each 2-digit NAICS industry. Standard errors are clustered on the Commuting Zone level.

In addition, we also examine heterogeneity in racial disparity across neighborhoods in commuting zones with different populations. As described earlier, we categorize Zip codes into three groups by the population of the respective commuting zones they are located in. "Low population" areas are Zip codes with a total 2010 population of under 515,013 in their commuting zone, representing 50% of all Zip codes. "Medium population" areas are Zip codes with a total population between 515,014 and 1,604,456 in their commuting zone, representing 25% of all Zip codes. Finally, "high population" areas are Zip codes with a population over 1,604,457 in their commuting zone, representing 25% of all Zip codes. For simplicity, these three areas are denoted as CZ group 1, 2, and 3, or "Low population CZ", "Medium population CZ", and "High population CZ", respectively. We separately estimate the racial gap in access for these three areas by interacting the coefficient on Black Ratio with dummies for the different areas in the following specification:

$$Y_{ic} = \alpha + \sum_k \beta_k \{\text{CZ group} = k\}_i \times \text{Black Ratio}_i + \mathbf{X}_i\delta + \gamma_c + \varepsilon_{ic} \tag{3.4}$$

where $\{\text{CZ group} = k\}_i$ is equal to one if Zip code i is located in one of the group k commuting zones.

Results under Equation (3.3) and Equation (3.4) are shown in Table 3.2. Column (1) shows that Zip codes with a 10% higher proportion of Black population is 0.77% less likely to contain any lender. Column (3) repeats the analysis in (1) using log(number of lenders within Zip code) as the dependent variable, conditional on the number of lenders being greater than zero, and shows that the gap is even wider in the intensive margin: a 10% higher Black population share is associated with a 2.1% lower number of branches, conditional on the Zip code having at least one lender. Decomposing across Zip codes by population of their commuting zones in Column (2) demonstrates that this racial disparity in access is statistically significant for Zip codes in both low density and medium density commuting zones, and is the strongest for low density areas. The racial gap in the intensive margin is statistically significant for all three CZ groups, as shown in Column (4) with the most significant gap being in the high density commuting zones.

To confirm that racial disparity in access is relevant in the context of PPP, Table 3.3 investigates whether Black Zip codes have worse access specifically to lenders that are enrolled in the PPP program. Columns (1) and (3) show that the difference in access is larger (more negative) in magnitude than before, suggesting that Black Zip codes have a even more pronounced disadvantage in access to PPP lenders rather than lenders in general. Decomposing the racial gap across areas with different population densities in columns (2) and (4) shows that the disparity in access to enrolled lenders are greater for the lower population commuting zones (CZ_cat=1) in both extensive and intensive margins.

Comparing coefficients between Table 3.2 and 3.3 shows that the racial disparity is worse when focusing on access to lenders enrolled in the PPP program rather than lenders in general. This indicates that the *kind* of lenders in the area matters in addition to the presence of *any* lender. By construction, it must be that lenders in more heavily Black neighborhoods are less likely to enroll. Indeed, Figure 3.1, which is a bin-scatter plot with the percent of Black residents in a Zip code on the x-axis and the fraction of bank and credit union branches that enrolled in PPP in a Zip code on the y-axis, shows that the fraction of lenders branches enrolled in PPP in a Zip code is negatively correlated with its percent of Black residents.

Table 3.2: *Access to Lenders in Zip Code*

	(1) Has Branch	(2) Has Branch	(3) Log(Branch)	(4) Log(Branch)
black_ratio	-0.0774***		-0.210***	
	(0.0215)		(0.0455)	
CZ_cat=1 × black_ratio		-0.219***		-0.185**
		(0.0507)		(0.0792)
CZ_cat=2 × black_ratio		-0.0864***		-0.176***
		(0.0332)		(0.0610)
CZ_cat=3 × black_ratio		-0.0212		-0.249***
		(0.0325)		(0.0756)
Constant	0.0293	0.0273	0.259***	0.287***
	(0.0365)	(0.0363)	(0.0509)	(0.0523)
2-digit NAICS controls	Yes	Yes	Yes	Yes
Log(Establishments) controls	Yes	Yes	Yes	Yes
Log(Population) controls	Yes	Yes	Yes	Yes
Population density	Yes	Yes	Yes	Yes
Observations	28136	27249	19237	18449
heightCounty FE	Yes	Yes	Yes	Yes

Robust standard errors clustered by county Zone in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows OLS regressions of the presence of lenders (banks and credit union branches) in a Zip code. Column (1) is a regression of "Has Branch" on the ratio of Black population in the Zip code, black_ratio, with controls for the shares of 2-digit NAICS industry presence, log number of small establishments in the Zip code and squared, log population and squared, population density, as well as county fixed effects. Column (2) is the same regression but with black_ratio interacted with our three categories of commuting zone population. Columns (3) and (4) are the same regressions but on the intensive margin of the log number of branches in a Zip code, conditional on it being greater than zero.

To verify that the relationship between the percent of residents that are Black and the probability of lender enrollment holds after we include county fixed effects and our control variables, we run branch level regressions of enrollment on lender and Zip code level controls with standard errors clustered by lender and county and results shown in Table 3.4. Column (1) of Table 3.4 shows that lender enrollment is decreasing in the percent of Black residents in a Zip code. Column (2) shows that the coefficient relationship between

Table 3.3: *Access to Enrolled Lenders in Zip Code*

	(1) Enrolled	(2) Enrolled	(3) Log(Enrolled)	(4) Log(Enrolled)
black_ratio	-0.103***		-0.407***	
	(0.0225)		(0.0456)	
CZ_cat=1 × black_ratio		-0.259***		-0.477***
		(0.0492)		(0.0837)
CZ_cat=2 × black_ratio		-0.100***		-0.392***
		(0.0383)		(0.0644)
CZ_cat=3 × black_ratio		-0.0471		-0.412***
		(0.0322)		(0.0711)
Constant	-0.0360	-0.0277	0.457***	0.507***
	(0.0339)	(0.0351)	(0.0622)	(0.0621)
2-digit NAICS controls	Yes	Yes	Yes	Yes
Log(Establishments) controls	Yes	Yes	Yes	Yes
Log(Population) controls	Yes	Yes	Yes	Yes
Population density	Yes	Yes	Yes	Yes
Observations	28136	27249	17993	17232
heightCounty FE	Yes	Yes	Yes	Yes

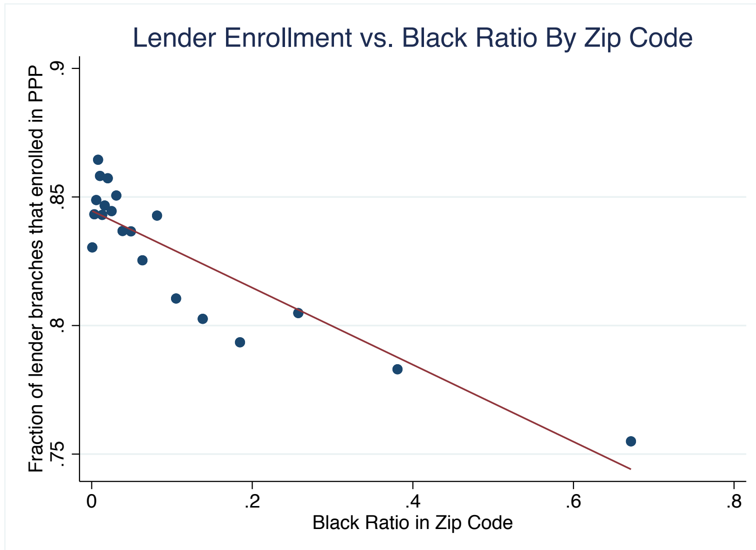
Robust standard errors clustered by county in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows OLS regressions of the presence of lenders (banks and credit union branches) enrolled in the PPP program in a Zip code. Column (1) is a regression of "Has Enrolled" on the ratio of Black population in the Zip code, black_ratio, with controls for the shares of 2-digit NAICS industry presence, log number of small establishments in the Zip code and squared, log population and squared, population density, as well as county fixed effects. Column (2) is the same regression but with black_ratio interacted with our three categories of commuting zone population. Columns (3) and (4) are the same regressions but on the intensive margin of the log number of enrolled lender branches in a Zip code, conditional on it being greater than zero.

enrollment and the percent of Black residents is slightly stronger in the lower density commuting zones (CZ_cat=1 and CZ_cat=2). Finally, Column (3) of Table 3.4 shows that the relationship between lender enrollment and the Black ratio is mostly explained by three additional variables and their interactions: (i) a dummy variable for whether the lender is a credit union, (ii) a dummy variable for whether the lender is small (less than 10 branches), and (iii) a dummy variable for whether the lender is enrolled in either SBA 7(a)

Figure 3.1: *Fraction of Branches in Zip Code Enrolled in PPP vs Ratio of Residents that are Black*



Note: This figure shows a bin-scatter plot of the fraction of bank and credit union branches in a Zip code that enrolled in the PPP program on the y-axis and the ratio of residents in the Zip code that are Black on the x-axis.

or 504 programs. While a small racial disparity remains, these three variables reduces the coefficient of black_ratio on enrollment by around three-quarters.

3.3.2 PPP take-up by racial composition of establishment neighborhood

Given the suggestive evidence that more heavily Black neighborhoods have worse access to lenders in general as well as PPP lenders in particular, we then examine if this racial gap translates to a racial disparity in access to support through the PPP program. As a first step, the bin scatter plot in Figure 3.2 shows a clearly negative correlation between PPP take-up rate and the share of Black population on a Zip code level. To perform a more careful descriptive analysis, we again invoke the specifications in equations (3.3) and (3.4), this time using take-up rate as the dependent variable. Our regressions thus estimate the neighborhood-level relationship between the share of Black population and the fraction of eligible businesses that end up receiving support through PPP.

Table 3.5 displays our results on the PPP take-up by racial composition of the establish-

Table 3.4: *Probability of Enrollment by Branch*

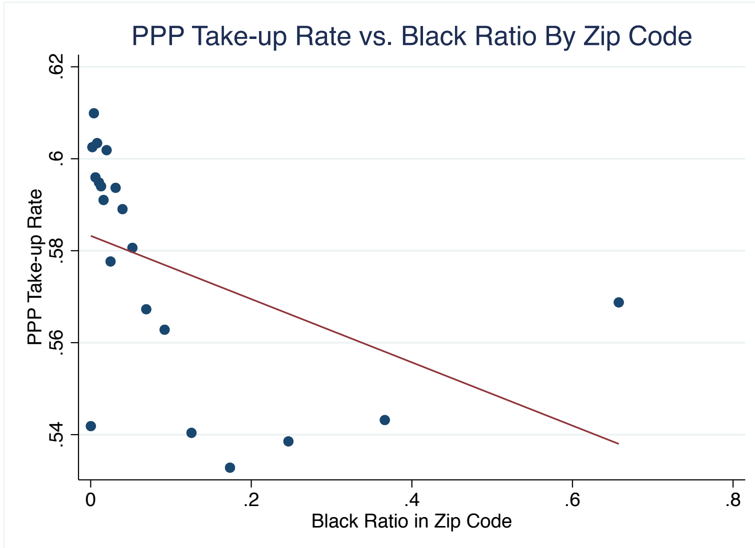
	(1) Enroll	(2) Enroll	(3) Enroll	(4) Enroll
black_ratio	-0.178***	-0.0411***		
	(0.0293)	(0.0120)		
CZ_cat=1 × black_ratio			-0.204***	-0.0778***
			(0.0444)	(0.0272)
CZ_cat=2 × black_ratio			-0.204***	-0.0371
			(0.0394)	(0.0231)
CZ_cat=3 × black_ratio			-0.156***	-0.0382***
			(0.0333)	(0.0139)
credit_union		-0.271***		-0.269***
		(0.0479)		(0.0488)
enrolled_sba		0.218***		0.210***
		(0.0322)		(0.0317)
small_lender		-0.0458**		-0.0461**
		(0.0213)		(0.0212)
credit_unionXsmall_lender		-0.249***		-0.255***
		(0.0404)		(0.0415)
credit_unionXenrolled_sba		0.274***		0.274***
		(0.0450)		(0.0457)
Constant	0.799***	0.794***	0.819***	0.804***
	(0.0513)	(0.0451)	(0.0512)	(0.0458)
2-digit NAICS controls	Yes	Yes	Yes	Yes
Log(estab) and squared	Yes	Yes	Yes	Yes
Log(Population) and squared	Yes	Yes	Yes	Yes
Population density	Yes	Yes	Yes	Yes
Observations	105047	105047	98721	98721
heightCounty FE	Yes	Yes	Yes	Yes

Robust standard errors clustered by lender and county in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows the results of OLS regressions on branch enrollment.

ment neighborhoods. In columns (1) and (2), we first look at the correlation between share of Black population and take-up rate without including any of the control variables. We find that on average, Zip codes with a 10% higher proportion of Black population have a

Figure 3.2: *Fraction of small establishments in Zip code getting PPP loans vs. share of Black population in Zip code, binscatter plot*



Note: This figure shows a bin-scatter plot of the fraction of small establishments in a Zip code that took out a PPP loan on the y-axis and the ratio of residents in the Zip code that are Black on the x-axis.

1.3% lower take-up rate of PPP loans. Separating this effect across Commuting Zones with different population densities in column (2) shows that the racial disparity is almost entirely restricted to the low population and medium population Commuting Zones. Controlling for neighborhood-level characteristics in columns (3) and (4) makes the coefficients somewhat smaller in magnitude, but does not take away statistical significance. To put our estimates in perspective, a small business in a Zip code with 100% Black population would be 8.4% less likely to obtain a PPP loan than its counterpart in a Zip code with no Black population, and 25.0% less likely than its counterpart if it is located in a low population Commuting Zone.

To summarize, controlling for county fixed effects and local characteristics, we find that small businesses Zip codes with a larger share of Black population are less likely to take up PPP loans. This effect is the most severe within low-density Commuting Zones, and also statistically significant within medium-density Commuting Zones. These results suggest that racial disparity in pass-through of government support does exist, and the relatively more rural areas suffer in particular.

Table 3.5: *Fraction of small establishments in a Zip code getting loans, by categories of commuting zone population*

	(1) Takeup Rate	(2) Takeup Rate	(3) Takeup Rate	(4) Takeup Rate
black_ratio	-0.117***		-0.0844***	
	(0.0360)		(0.0294)	
CZ_cat=1 × black_ratio		-0.374***		-0.250***
		(0.0327)		(0.0300)
CZ_cat=2 × black_ratio		-0.204***		-0.134***
		(0.0292)		(0.0231)
CZ_cat=3 × black_ratio		-0.00607		-0.0228
		(0.0557)		(0.0468)
Constant	0.613***	0.621***	0.199***	0.207***
	(0.00334)	(0.00216)	(0.0258)	(0.0251)
2-digit NAICS controls	No	No	Yes	Yes
Log(Establishments) controls	No	No	Yes	Yes
Log(Population) controls	No	No	Yes	Yes
Population density	No	No	Yes	Yes
Observations	28864	27928	28136	27249
heightCounty FE	Yes	Yes	Yes	Yes

Robust standard errors clustered by county in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table shows OLS regressions on the PPP take-up rate in a Zip code. Column (1) is a regression of PPP take-up on the ratio of residents in a Zip code that are Black ("black_ratio") controlling only for county fixed effects. Column (2) is the same regression but interacted with categories of commuting zone population. Column (3) adds additional controls for shares of 2-digit NAICS industries in a Zip code, log number of small establishments in a Zip code, log population in a Zip code, and the population density. Column (4) is the same regression as Column (3) but interacted with our three categories of commuting zone density.

3.3.3 Type of PPP Lender Used

In Section 3.1, we provided evidence that Black neighborhoods have fewer lenders in their Zip codes, and this applies to both lenders in general and lenders enrolled in the PPP program. However, one might still wonder if this necessarily implies racial disparity in access overall. For example, if a Black neighborhood has a comparable number of PPP lenders in *nearby* Zip codes, then it could have similar levels access to lenders despite the lack of financial institutions within the neighborhood itself.

To verify if the racial gap in access extends beyond the neighborhood level, we look at the type of lenders that PPP borrowers use and how they differ by zip codes. First, we define a set of lenders as "FinTech", based in part on a list from Howell *et al.* (2021). Second, we estimate the distance of each business to its lender by computing the distance between Zip code of the business and Zip code of the closest branch of the lender. Based on this distance measure, we define a loan as "remote" if the business is more than 100 miles away from the lender. Small businesses get their loans from a remote lender may have applied through online platforms or a remote application process. Third, we examine the distance between the small business and their lender PPP lender, conditional on it being non-FinTech and non-remote.

Figure 3.3 shows the correlation between distance to PPP lender and the proportion of Black population on a Zip code level through two bin scatter plots. Panel (a) shows that businesses in Zip codes with a higher Black population are significantly more likely to apply for PPP through a FinTech lender. Panel (b) shows that, conditional on non-FinTech lenders, Zip codes with a higher Black share have a higher share of "remote" PPP loans on average. Panel (c) shows that, even if we only inspect the non-FinTech and non-remote loans, businesses in Zip codes with a higher ratio of Black population obtain loans from lenders that are further away. These results indicate that the disparity in access to PPP lenders is correlated with higher usage of FinTech and further away options.

To more systematically analyze the relationship between racial composition and remoteness of PPP loans, we conduct loan-level regressions using the following specifications:

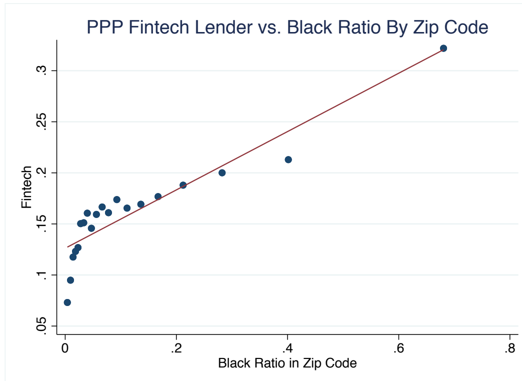
$$Y_{biz} = \alpha + \beta \text{Black Ratio}_i + \mathbf{X}_i \delta + \gamma_c + \varepsilon_{biz} \tag{3.5}$$

$$Y_{biz} = \alpha + \sum_k \beta_k \{\text{CZ group} = k\}_i \times \text{Black Ratio}_i + \mathbf{X}_i \delta + \gamma_z + \varepsilon_{biz} \tag{3.6}$$

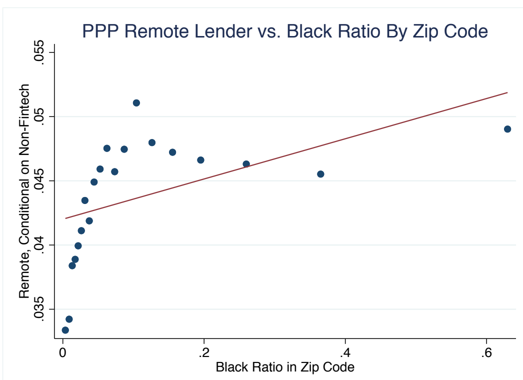
where Y_{biz} is the outcome variable for a loan taken by business b in Zip code i and county

Figure 3.3: *Type of PPP Lender used by Black Ratio in Zip Code*

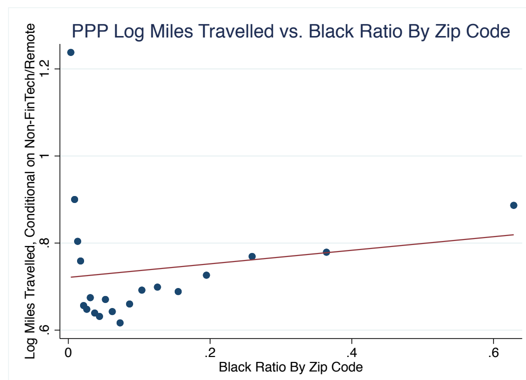
(a) *Share of PPP Loans from Fintech Lender*



(b) *Conditional on Non-FinTech, share of PPP Loans from Remote Lender*



(c) *Conditional on Non-FinTech and Non-Remote, Log Miles to PPP Lender*



Note: These figures show bin-scatter plot of the share of loans by PPP lender type on the y-axes, and the ratio of residents in the Zip code that are Black on the x-axis. Figure 3.3a shows whether the loan was taken out with a Fintech lender, Figure 3.3b shows whether the loan was taken out with a remote lender (a lender that is more than 100 miles away from the small business) conditional on non-Fintech, and Figure 3.3c shows the log miles travelled to the lender (the distance between the business and the nearest lender branch) conditional on non-Fintech and non-Remote.

z; γ_c are county fixed effects. Like before, we control for a vector of Zip code-level characteristics $\mathbf{X}_i$, and interact the coefficient on Black Ratio with dummies for different Commuting Zone population groups to investigate heterogeneity in the racial gap. For the outcome variable, we use a dummy variable equal to one if the PPP lender is remote, as well as log distance of business to its lender.

Results for regressions in equations 3.5 and 3.6 are shown in Table 3.6. Column (1) shows that a 10% higher ratio of Black population in the Zip code in which a business is located is associated with a 1.6 percentage point increase in the likelihood of the lender to be FinTech. Column (2) shows that within all Commuting Zone groups, businesses in Black Zip codes are more likely to acquire PPP loans from a FinTech lender, with the difference being highest among the high density Commuting Zones. Column (3) and (4) show that businesses in Zip codes with a higher share of Black population are also more likely to get their PPP loan from a remote lender, although the magnitude of the difference is lower than it is for FinTech and is concentrated among the high population commuting zones. As we turn to the intensive margin of distance, Column (5) reveals that business in Black Zip codes travel further overall to obtain PPP loans: businesses in a Zip code with a 10% higher Black ratio are on average 2.4% further away from their lenders. Column (6) confirms that this racial disparity is statistically significant regardless of population density of the Commuting Zone, and is the highest among the lower population commuting zones. Our FinTech and distance results suggest that FinTech penetration may be more successful at making up for the racial disparities in PPP access in more urban, higher population areas, whereas businesses from minority Zip Codes in more rural areas still travelled further to get a loan from a brick and mortar lender.

3.4 Instrumental Variables estimates

Given the descriptive evidence in Section 3.3, which suggests that more heavily Black Zip codes have (1) worse access to lenders in general and particularly PPP lenders and (2) take up PPP loans at a lower rate, we now attempt to establish a connection between access to lenders and pass-through of PPP support, and examine how this connection can explain the racial disparity in take-up. Ideally, we would like to run the following regressions:

$$\text{Takeup}_{iz} = \alpha^1 + \sum_k \gamma_k^1 \{\text{CZ group} = k\}_i \times \text{Enrollment}_i + \mathbf{X}_i\delta^1 + \gamma_c^1 + \varepsilon_{iz} \tag{3.7}$$

$$\text{Takeup}_{iz} = \alpha^0 + \sum_k \beta_k^0 \{\text{CZ group} = k\}_i \times \text{Black Ratio}_i + \mathbf{X}_i\delta^0 + \gamma_c^0 + \varepsilon_{iz} \tag{3.8}$$

Table 3.6: *Type of PPP Lender Used Regressions*

	(1) FinTech	(2) FinTech	(3) Remote	(4) Remote	(5) Log(Miles)	(6) Log(Miles)
black_ratio	0.166***		0.00872***		0.244***	
	(0.0248)		(0.00270)		(0.0252)	
CZ_cat=1 × black_ratio		0.0528***		0.00784		0.417***
		(0.00727)		(0.00507)		(0.0817)
CZ_cat=2 × black_ratio		0.100***		-0.00371		0.210***
		(0.0142)		(0.00663)		(0.0556)
CZ_cat=3 × black_ratio		0.206***		0.0134***		0.242***
		(0.0350)		(0.00276)		(0.0291)
Constant	0.184***	0.143***	0.0305***	0.0230**	3.013***	2.869***
	(0.0275)	(0.0264)	(0.0109)	(0.0108)	(0.145)	(0.144)
2-digit NAICS controls	Yes	Yes	Yes	Yes	Yes	Yes
Log(Establishments) controls	Yes	Yes	Yes	Yes	Yes	Yes
Log(Population) controls	Yes	Yes	Yes	Yes	Yes	Yes
Population density	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4342923	3880797	3896283	3528177	3724185	3377126
heightCounty FE	Yes	Yes	Yes	Yes		

Robust standard errors clustered by county in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Column (1) of this table displays the result of an OLS regression on an indicator variable for whether the loan was taken out with a Fintech lender and the ratio of Black residents in the Zip code ("black_ratio"), controlling for 2-digit NAICS shares, log number of establishments in a Zip code and squared, log population and squared, population density, and county fixed effects. Column (2) is the same regression but with black_ratio interacted with our categories of commuting zone density. Column (3) displays the results of an OLS regression on an indicator variable for whether the loan was taken out with remote lender (a lender whose nearest branch is more than 100 miles away) conditional on non-Fintech, with the same dependent variables as Column (1). Column (4) is the same regression but with black_ratio interacted with our categories of commuting zone density. Column (5) is the same regression as Column (3) but with the dependent variable changed to log 1 + number of miles to the nearest branch of the PPP lender, conditional on non-Fintech and non-remote. Column (6) is the same regression as Column (5) but with black_ratio interacted with our categories of commuting zone density.

In equation (3.7), we regress PPP take-up rate of Zip code i in Commuting Zone z on a Zip-code level enrollment variable as well as the ratio of Black population, allowing for different coefficients across Commuting Zones with different population densities. For the enrollment variable, we use indicators on both extensive and intensive margins, i.e. a dummy variable for whether the Zip code contains a PPP-enrolled lender, and log of the number of PPP-enrolled lenders. The estimated coefficient $\hat{\beta}_k^1$ measures the relationship between enrollment of local lenders in the PPP program and take-up of PPP loans for each CZ group k. Moreover, by comparing the estimated coefficients on Black Ratio across Equation (3.7) and Equation (3.8), $\hat{\beta}_k^1$ and $\hat{\beta}_k^0$, we can also gauge the extent to which local access can explain racial disparity in the pass-through of PPP support.

However, the specification in Equation 3.7 is prone to endogeneity, since PPP enrollment is likely correlated with other local factors contained in the error term ε_{iz} that might affect PPP take-up. For example, better local economic conditions would prompt banks in general (including PPP-enrolled ones) to establish more branches in the corresponding Zip codes, and simultaneously lead to higher PPP take-up rate since local businesses are more resilient to the Covid shock and less likely to shut down. This would generate an upward bias in our estimate of the association between enrollment and take-up, $\hat{\gamma}_k^1$.

Therefore, to assess the causal effect of local access to PPP-enrolled lenders on take-up, we use the failure of small community banks, defined as banks with fewer than 10 branches, to proxy for PPP enrollment of local banks. Specifically, we define a dummy variable that is equal to 1 if the Zip code contains at least one branch of a failed bank. The rationale for choosing this instrument is that failed branches tend to be acquired by larger financial institutions, which are more likely to enroll in the PPP compared to small lenders. As a result, Zip codes with branches of failed banks should have more PPP-enrolled lenders relative to those without, after controlling for other local characteristics.

Regarding external validity of the instrument, Antoniades (2020) has shown that the failure of commercial banks in the US is primarily driven by exposure to commercial real estate loans as well as residential mortgage backed securities (RMBS), and not small business loans. In other words, bank failures may be mainly driven by shocks uncorrelated with local conditions of small businesses that would affect take-up of PPP loans. Moreover, even if some bank failures are indeed driven by deteriorating local economic conditions and thus violate the exclusion restriction, that would only make our IV estimate more conservative: businesses in areas with worse economic conditions are more likely to close down during the pandemic, and thus should be *less* likely to take-up PPP loans, rather than being more likely as we find. In line with this argument, the Appendix shows that Zip codes with branches of failed banks have a similar claim rate for unemployment insurance (UI) prior to Covid-19, in February 2020, with the point estimate for the difference being 0.265% and not statistically significant. Furthermore, Zip codes with higher UI claim rate in February are strongly correlated with a lower PPP take-up rate. Therefore, if anything, using this instrument may generate a downward bias on our estimated impact of access on PPP take-up.

Using the proposed instrument, we supplement the main regression in Equation 3.7 with a first-stage regression:

$$\text{Enrollment}_{ic} = \kappa + \rho\{\text{Has Failed Branch}\}_i + \mathbf{X}_i\mu + \varphi_c + \xi_{ic} \tag{3.9}$$

where Enrollment_{ic} is an enrollment variable of Zip code i in county c, and $\{\text{Has Failed Branch}\}_i$ is a dummy variable equal to one if Zip code i contains at least one failed bank branch.

Results for the first stage IV regression are shown in Table 3.7. As previously discussed, for the enrollment variable, we use both a dummy variable for whether a Zip code contains any PPP-enrolled lender, and log number of PPP-enrolled lenders a Zip code. The results indicate that our instrument is mostly relevant in the intensive margin. More specifically, Column (1) indicates that neighborhoods with branches of failed banks are about equally likely to have an enrolled lender as neighborhoods without failed banks. One possible explanation for this is that Zip codes with failed banks *and* no PPP-enrolled banks ex-ante is a small sample, and may be simply economically unattractive for big banks to enter through acquisitions. In contrast, Column (2) shows that, conditional on a Zip code containing at least one PPP-enrolled lender, Zip codes with at least one failed bank branch have a 9.1% higher number of enrolled lenders. This confirms the internal validity of our instrument for the intensive margin of access to PPP lenders.

For the second stage of the IV regression, we first run simpler specifications that isolate

Table 3.7: *Instrumental variables first stage estimates*

	(1) Has Enrolled	(2) Log(Enrolled)
has_failed_branch	-0.00290	0.0908^{***}
	(0.00889)	(0.0151)
Constant	-0.0560^{*}	0.312^{***}
	(0.0316)	(0.0622)
2-digit NAICS controls	Yes	Yes
Log(Establishments) controls	Yes	Yes
Log(Population) controls	Yes	Yes
Population density	Yes	Yes
Small Bank # controls	Yes	Yes
Observations	27957	17397
heightCounty FE	Yes	Yes

Robust standard errors clustered by county in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: this table shows our first stage regressions of an indicator variable for whether a Zip code has an enrolled lender in Column (1), and the log number of enrolled branches (conditional on there being at least one) in Column (2). We control for 2-digit NAICS shares, log number of small establishments and squared, log population and squared, population density, and the 2000 number of small banks in the Zip code and squared, and county fixed effects.

the impact of local access to enrolled lenders on PPP take-up:

$$\text{Takeup}_{ic} = \alpha + \gamma \text{Enrollment}_i + \mathbf{X}_i \delta^1 + \gamma_c + \varepsilon_{ic} \tag{3.10}$$

$$\text{Takeup}_{ic} = \alpha + \sum_k \gamma_k \{\text{CZ group} = k\}_i \times \text{Enrollment}_i + \mathbf{X}_i \delta + \gamma_c + \varepsilon_{ic} \tag{3.11}$$

Results for Equation (3.10) and Equation (3.11) are shown in Table 3.8. Column (1) shows that doubling the number of PPP-enrolled lenders in a Zip code would lead to an increase of 11.9 percentage points in the take-up rate of PPP loans. Column (2) examines this effect separately for areas with different population densities, and finds the magnitude to vary between 11.8 and 12.6 percentage points. These results show that local access to particular lenders does matter for actual pass-through of government support through PPP loans.

Table 3.8: *Instrumental Variables Second Stage Estimates*

	(1) Takeup	(2) Takeup
Log(Enrolled)	0.119^{**}	
	(0.0505)	
CZ_cat=1 × Log(Enrolled)		0.118^{**}
		(0.0599)
CZ_cat=2 × Log(Enrolled)		0.149^{***}
		(0.0546)
CZ_cat=3 × Log(Enrolled)		0.126^{**}
		(0.0533)
2-digit NAICS controls	Yes	Yes
Log(Establishments) controls	Yes	Yes
Log(Population) controls	Yes	Yes
Population density	Yes	Yes
Small Bank # controls	Yes	Yes
Observations	17397	16638
First stage F-stat	36.09	10.39
heightCounty FE	Yes	Yes

Robust standard errors clustered by county in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: this table shows our second stage regressions of take up on the log number of enrolled branches (conditional on there being at least one) in Column (1), and the log number of enrolled branches (conditional on there being at least one) interacted with our Commuting Zone population category in Column (2). We control for 2-digit NAICS shares, log number of small establishments and squared, log population and squared, population density, the 2000 number of small banks in the Zip code and squared, and county fixed effects.

Finally, we run the regressions in Equation 3.7 and Equation 3.8 to examine the explanatory power of the intensive margin access to PPP lenders for racial disparity in PPP take-up. For the second stage regression in Equation 3.7, we instrument for enrollment using our bank failure IV. Table 3.9 displays the results. Column (1) corresponds to Equation 3.8 for the sample of Zip codes with at least 1 enrolled lender; column (2) corresponds to Equation 3.7, which includes both Black ratio and log number of PPP-enrolled lenders as regressors. Comparing across the two columns, the coefficient on Black Ratio, which measures the racial disparity in take-up, decreases for all zip code densities after controlling for the number of enrolled lenders. More specifically, the results suggest that access to enrolled lenders can explain 32.8% of the racial gap in low density areas, and 57.5% of the racial gap in medium density areas.

3.5 Conclusion

Many government support programs are designed to be intermediated by banks and credit unions. We show that one drawback of this design is that it exacerbates the distributional effects of "government support banking deserts". All else equal, neighborhoods with a greater share of Black population have fewer lenders, and the lenders that are present there are less likely to enroll in the PPP program. We show using an instrumental variables approach that the significantly lower take up rate of Black neighborhoods in lower population areas can be partially explained by this difference in access to enrolled lenders. Our results suggest that alternative implementations of government support programs, e.g. by combining "support through banks" with more outreach centers in banking deserts, or providing more incentives for credit unions in minority neighborhoods to participate in SBA programs, may improve the distributional impact of government support for small businesses.

Table 3.9: *Instrumental Variables Implication*

	(1) Takeup	(2) Takeup
CZ_cat=1 × black_ratio	-0.241***	-0.162***
	(0.0306)	(0.0406)
CZ_cat=2 × black_ratio	-0.0817***	-0.0367
	(0.0223)	(0.0285)
CZ_cat=3 × black_ratio	0.0512	0.0955
	(0.0532)	(0.0585)
CZ_cat=1 × Log(Enrolled)		0.112*
		(0.0598)
CZ_cat=2 × Log(Enrolled)		0.140**
		(0.0546)
CZ_cat=3 × Log(Enrolled)		0.123**
		(0.0540)
Constant	0.424***	
	(0.0479)	
2-digit NAICS controls	Yes	Yes
Log(Establishments) controls	Yes	Yes
Log(Population) controls	Yes	Yes
Population density	Yes	Yes
Small Bank # controls	Yes	Yes
Observations	16638	16638
First stage F-stat		10.32
heightCounty FE	Yes	Yes

Robust standard errors clustered by county in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: Column (1) of this table is an OLS regression of take-up on the ratio of the population that is Black ("black_ratio") interacted with our categorical variable for Commuting Zone density within the Zip codes with at least one enrolled lender, controlling for 2-digit NAICS shares, log number of small establishments and squared, log population and squared, population density, the 2000 number of small banks in the Zip code and squared, and county fixed effects. Column (2) is the second stage of an IV regression that adds in the log of the number of enrolled branches interacted with our categories of Commuting Zone population.

References

Abramovsky, L. and Griffith, R. (2006). Outsourcing and offshoring of business services: How important is ict? *Journal of the European Economic Association*, **4** (2-3), 594–601.

Acemoglu, D., Autor, D., Hazell, J. and Restrepo, P. (2020a). *AI and jobs: Evidence from online vacancies*. Tech. rep., National Bureau of Economic Research.

—, Griffith, R., Aghion, P. and Zilibotti, F. (2010). Vertical integration and technology: theory and evidence. *Journal of the european economic Association*, **8** (5), 989–1033.

—, Lelarge, C. and Restrepo, P. (2020b). Competing with robots: Firm-level evidence from france. In *AEA Papers and Proceedings*, vol. 110, pp. 383–88.

—, Manera, A. and Restrepo, P. (2020c). *Does the US tax code favor automation?* Tech. rep., National Bureau of Economic Research.

— and Restrepo, P. (2020). Robots and jobs: Evidence from us labor markets. *Journal of Political Economy*, **128** (6), 2188–2244.

Aghion, P., Antonin, C., Bunel, S. and Jaravel, X. (2020). What are the labor and product market effects of automation? new evidence from france.

Alguacil Marí, M. T., Lo Turco, A. and Martínez-Zarzoso, I. (2020). *What is so special about robots and trade?* Tech. rep., cege Discussion Papers.

Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I. and Schnell, M. (2019). Food Deserts and the Causes of Nutritional Inequality*. *The Quarterly Journal of Economics*, **134** (4), 1793–1844.

Altonji, J. G. and Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, Trade, and the Labor Market*, University of Chicago Press, pp. 201–234.

Ambrose, B. W., Conklin, J. N. and Lopez, L. A. (2020). Does Borrower and Broker Race Affect the Cost of Mortgage Credit? *The Review of Financial Studies*, **34** (2), 790–826.

Amiti, M., Redding, S. J. and Weinstein, D. E. (2019). The impact of the 2018 tariffs on prices and welfare. *Journal of Economic Perspectives*, **33** (4), 187–210.

Antoniades, A. (2020). Commercial bank failures during the great recession: The real (estate) story. *Available at SSRN 2325261*.

Antràs, P. (2020). *De-Globalisation? Global value chains in the post-COVID-19 age*. Tech. rep., National Bureau of Economic Research.

Antràs, P. and De Gortari, A. (2020). On the geography of global value chains. *Econometrica*, **88** (4), 1553–1598.

Antras, P., Fort, T. C. and Tintelnot, F. (2017). The margins of global sourcing: Theory and evidence from us firms. *American Economic Review*, **107** (9), 2514–64.

Artuc, E., Bastos, P. and Rijkers, B. (2020). Robots, tasks, and trade.

—, Christiaensen, L. and Winkler, H. J. (2019). *Does Automation in Rich Countries Hurt Developing Ones?: Evidence from the US and Mexico*. The World Bank.

Autor, D. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, **103** (5), 1553–97.

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R. J. and Stroebel, J. (2020). *International trade and social connectedness*. Tech. rep., National Bureau of Economic Research.

Baldwin, R. and Venables, A. J. (2013). Spiders and snakes: Offshoring and agglomeration in the global economy. *Journal of International Economics*, **90** (2), 245–254.

Baldwin, R. E. and Evenett, S. (2020). Covid-19 and trade policy: Why turning inward won't work.

Balla-Elliott, D., Cullen, Z. B., Glaeser, E. L., Luca, M. and Stanton, C. T. (2020). Business reopening decisions and demand forecasts during the covid-19 pandemic. *Working Paper*.

Bartel, A. P., Lach, S. and Sicherman, N. (2005). Outsourcing and technological change.

—, — and — (2014). Technological change and the make-or-buy decision. *The Journal of Law, Economics, & Organization*, **30** (1), 165–192.

Bartik, A. W., Bertrand, M., Cullen, Z. B., Glaeser, E. L., Luca, M. and Stanton, C. T. (2020a). How are small businesses adjusting to covid-19? early evidence from a survey. *Working Paper*.

—, Cullen, Z. B., Glaeser, E. L., Luca, M. and Stanton, C. T. (2020b). What jobs are being done at home during the covid-19 crisis? evidence from firm-level surveys. *Working Paper*.

—, —, —, —, — and Sunderam, A. (2020c). The targeting and impact of paycheck protection program loans to small businesses. *Working Paper*.

Bernard, A. B., Blanchard, E. J., Van Beveren, I. and Vandenbussche, H. (2019). Carry-along trade. *The Review of Economic Studies*, **86** (2), 526–563.

—, Fort, T. C., Smeets, V. and Warzynski, F. (2020). *Heterogeneous globalization: Offshoring and reorganization*. Tech. rep., National Bureau of Economic Research.

—, Jensen, J. B. and Schott, P. K. (2009). 14. importers, exporters, and multinationals: A portrait of firms in the us that trade goods. In *Producer Dynamics*, University of Chicago Press, pp. 513–556.

Bessen, J., Goos, M., Salomons, A. and van den Berge, W. (2020). Firm-level automation: Evidence from the netherlands. In *AEA Papers and Proceedings*, vol. 110, pp. 389–93.

Bloom, N., Garicano, L., Sadun, R. and Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, **60** (12), 2859–2885.

Bonfiglioli, A., Crinò, R., Fadinger, H. and Gancia, G. (2020). Robot imports and firm-level outcomes.

Bresnahan, T. F., Brynjolfsson, E. and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The quarterly journal of economics*, **117** (1), 339–376.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, **121** (2), 541–585.

Brynjolfsson, E. and Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic perspectives*, **14** (4), 23–48.

Buffington, C., Miranda, J., Seamans, R. *et al.* (2018). *Development of survey questions on robotics expenditures and use in US manufacturing establishments*. Tech. rep.

Burbidge, J. B., Magee, L. and Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, **83** (401), 123–127.

Cheng, H., Jia, R., Li, D. and Li, H. (2019). The rise of robots in china. *Journal of Economic Perspectives*, **33** (2), 71–88.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M. and Team, T. O. I. (2020). The economic impacts of covid-19: Evidence from a new public database built using private sector data. *Working Paper*.

Chiacchio, F., Petropoulos, G. and Pichler, D. (2018). *The impact of industrial robots on EU employment and wages: A local labour market approach*. Tech. rep., Bruegel working paper.

Cilekoglu, A., Moreno, R. and Ramos, R. (2021). The impact of robot adoption on global sourcing.

Dauth, W., Findeisen, S., Suedekum, J. and Woessner, N. (2019). The adjustment of labor markets to robots. *Journal of the European Economic Association*.

David, H., Dorn, D. and Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, **103** (6), 2121–68.

de Blasio, B. (2019). Why american workers need to be protected from automation. *Wired*.

Delvaux, M. (2016). *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics*. Tech. rep., European Parliament Committee on Legal Affairs.

Ding, X., Fort, T. C., Redding, S. J. and Schott, P. K. (2020). Structural change within versus across firms: Evidence from the united states.

Dixon, J., Hong, B. and Wu, L. (2020). *The employment consequences of robots: Firm-level evidence*. Statistics Canada.

Faber, M. (2020). Robots and reshoring: Evidence from mexican labor markets. *Journal of International Economics*, **127**, 103384.

Fajgelbaum, P. D., Goldberg, P. K., Kennedy, P. J. and Khandelwal, A. K. (2020). The return to protectionism. *The Quarterly Journal of Economics*, **135** (1), 1–55.

Feenstra, R. C. and Hanson, G. H. (1999). The impact of outsourcing and high-technology capital on wages: estimates for the united states, 1979–1990. *The quarterly journal of economics*, **114** (3), 907–940.

Feinberg, S. E. and Keane, M. P. (2006). Accounting for the growth of mnc-based trade using a structural model of us mncs. *American Economic Review*, **96** (5), 1515–1558.

Flaaen, A., Hortaçsu, A. and Tintelnot, F. (2020). The production relocation and price effects of us trade policy: the case of washing machines. *American Economic Review*, **110** (7), 2103–27.

Fort, T. and Klimek, S. (2018). *The Effects of Industry Classification Changes on US Employment Composition*. Tech. rep., US Census Bureau, Center for Economic Studies.

Fort, T. C. (2017). Technology and production fragmentation: Domestic versus foreign sourcing. *The Review of Economic Studies*, **84** (2), 650–687.

—, Pierce, J. R. and Schott, P. K. (2018). New perspectives on the decline of us manufacturing employment. *Journal of Economic Perspectives*, **32** (2), 47–72.

Goos, M., Manning, A. and Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American economic review*, **104** (8), 2509–26.

Gopinath, G. and Neiman, B. (2014). Trade adjustment and productivity in large crises. *The American Economic Review*, pp. 793–831.

Graetz, G. and Michaels, G. (2018). Robots at work. *Review of Economics and Statistics*, **100** (5), 753–768.

Granja, J., Makridis, C., Yannelis, C. and Zwick, E. (2020). Did the paycheck protection program hit the target? *Working Paper*.

Grossman, G. M. and Helpman, E. (1995). Technology and trade. *Handbook of international economics*, **3**, 1279–1337.

— and Rossi-Hansberg, E. (2008). Trading tasks: A simple theory of offshoring. *American Economic Review*, **98** (5), 1978–97.

Hallward-Driemeier, M. and Nayyar, G. (2019). *Have Robots Grounded the Flying Geese?: Evidence from Greenfield FDI in Manufacturing*. The World Bank.

Halpern, L., Koren, M. and Szeidl, A. (2015). Imported inputs and productivity. *American Economic Review*, **105** (12), 3660–3703.

Handley, K., Kamal, F. and Monarch, R. (2020). *Rising import tariffs, falling export growth: When modern supply chains meet old-style protectionism*. Tech. rep., National Bureau of Economic Research.

Hanson, G. H. (2021). *Immigration and Regional Specialization in AI*. Tech. rep., National Bureau of Economic Research.

Hegerty, S. W. (2016). Commercial bank locations and "banking deserts": A statistical analysis of milwaukee and buffalo. *The Annals of Regional Science*, **56** (1), 253–271.

Howell, S. T., Kuchler, T. and Stroebel, J. (2021). Which lenders had the highest minority share among their payment protection program (ppp) loans? *Working Paper*.

Humlum, A. (2019). Robot adoption and labor market dynamics. *Princeton University*.

Humphries, J. E., Neilson, C. A. and Ulyssea, G. (2020). Information frictions and access to the paycheck protection program. *Working Paper*.

IFR (2018). *World Robotics Report*. Tech. rep.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** (2), 467–475.

Kashian, R. D., Tao, R. and Drago, R. (2018). Bank deserts in the usa and the great recession: geography and demographics. *Journal of Economic Studies*.

Kleibergen, F. and Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, **133** (1), 97–126.

Koch, M., Manuylov, I. and Smolka, M. (2019). Robots and firms.

Kugler, A. D., Kugler, M., Ripani, L. and Rodrigo, R. (2020). *Us robots and their impacts in the tropics: Evidence from colombian labor markets*. Tech. rep., National Bureau of Economic Research.

Lee, D. S., McCrary, J., Moreira, M. J. and Porter, J. R. (2021). *Valid t-ratio Inference for IV*. Tech. rep., National Bureau of Economic Research.

Leigh, N. G. and Kraft, B. R. (2018). Emerging robotic regions in the united states: insights for regional economic evolution. *Regional Studies*, **52** (6), 804–815.

Lewis, C. (2014). Robots are starting to make offshoring less attractive. *Harvard Business Review*.

Lewis, E. (2011). Immigration, skill mix, and capital skill complementarity. *The Quarterly Journal of Economics*, **126** (2), 1029–1069.

Lin, O. (2018). Chinese robotic firms thrive, thanks to subsidies. *Asia Times*.

MacKinnon, J. G. and Magee, L. (1990). Transforming the dependent variable in regression models. *International Economic Review*, pp. 315–339.

Mann, K. and Püttmann, L. (2021). Benign effects of automation: New evidence from patent texts. *The Review of Economics and Statistics*, pp. 1–45.

Market Watch (2019). Bill gates finds an ally in washington for his idea to tax robots: Alexandria ocasio-cortez.

McClelland, R. and Airi, N. (2021). *Effective Income Tax Rates Have Fallen for The Top One Percent Since World War II*. Tech. rep., Tax Policy Center.

McKinsey (2018). *Industrial robotics – insights into the sector's future growth dynamics*. Tech. rep.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *econometrica*, **71** (6), 1695–1725.

— and Redding, S. J. (2015). New trade models, new welfare implications. *American Economic Review*, **105** (3), 1105–46.

Miller, M. J. (2015). Social finance in black geographies: A statistical analysis of locations in los angeles county. *Harvard Journal of African American Public Policy*, p. 78.

Morgan, D. P., Pinkovskiy, M. L. and Yang, B. (2016). *Banking deserts, branch closings, and soft information*. Tech. rep., Federal Reserve Bank of New York.

Oberson, X. (2019). *Taxing Robots*. Edward Elgar Publishing.

Pierce, J. R. and Schott, P. K. (2016). The surprisingly swift decline of us manufacturing employment. *American Economic Review*, **106** (7), 1632–62.

Quartz (2017). The robot that takes your job should pay taxes, says bill gates. *Interview with Bill Gates by Kevin J. Delaney. Quartz, February*, **17**.

Rodrik, D. (2018). *New technologies, global value chains, and developing economies*. Tech. rep., National Bureau of Economic Research.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J. and Sobek, M. (2020). Ipums usa: Version 10.0 [dataset]. minneapolis, mn: Ipums; 2020.

Stapleton, K. and Webb, M. (2020). Automation, trade and multinational activity: Micro evidence from spain. *Working Paper*.

Stock, J. H. and Watson, M. W. (2015). Introduction to econometrics 3rd ed.

Tolbert, C. M. and Sizer, M. (1996). *US commuting zones and labor market areas: A 1990 update*. Tech. rep.

Walker, R. E., Keane, C. R. and Burke, J. G. (2010). Disparities and access to healthy food in the united states: A review of food deserts literature. *Health Place*, **16** (5), 876 – 884.

Wang, J. (2022a). Industrial robots in us manufacturing: New data and facts.

— (2022b). Robots, trade, and offshoring: A perspective from us firms.

—, Yang, J., Iverson, B. and Kluender, R. (2021). Bankruptcy and the covid-19 crisis. *Working Paper*.

Webb, M. (2019). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.

World Bank, a. (2020). *World Development Report 2020: Trading for Developmentin the Age of Global Value*. Tech. rep., World Bank Publications.

Zolas, N., Kroff, Z., Brynjolfsson, E., McElheran, K., Beede, D. N., Buffington, C., Goldschlag, N., Foster, L. and Dinlersoz, E. (2021). *Advanced Technologies Adoption and Use by US Firms: Evidence from the Annual Business Survey*. Tech. rep., National Bureau of Economic Research.

Appendix A

Appendix to Chapter 2

A.1 Theory

A.1.1 Proof of Proposition 1

The proof proceeds in three steps. For the first step, I show that the profit function $\pi(n_1, I_R|\varphi, \beta_{2R})$ is supermodular in the choice variables, $X = (n_1, I_R)$. That would be the case if and only given values of φ and β_{2R}, for any pair $X = (n_1, I_R)$ and $X^{'} = (n_1^{'}, I_R^{'})$, it can be shown that

$$\begin{aligned}\pi(\max(n_1, n_1^{'}), \max(I_R, I_R^{'}))|\varphi, \beta_{2R}) - \pi(n_1, I_R|\varphi, \beta_{2R}) \geq \\ \pi(n_1^{'}, I_R^{'}|\varphi, \beta_{2R}) - \pi(\min(n_1, n_1^{'}), \min(I_R, I_R^{'}))|\varphi, \beta_{2R}).\end{aligned} \tag{A.1}$$

Without loss of generality, it suffices to consider a single case: $n_1 \leq n_1^{'}, I_R = 1, I_R^{'} = 0$. Note that for any given φ and β_{2R}, we can write

$$\begin{aligned}\frac{\partial \pi(n_1, 1|\varphi, \beta_{2R})}{\partial n_1} = B\varphi^{\sigma-1}\left[\left(\frac{c_{2d}}{\beta_{2R}}\right)^{1-\alpha}\right]^{1-\sigma}\left(c_{1d}^{1-\rho} + n_1(\tau c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(\tau c_{1f})^{1-\rho} + \\ B\varphi^{\sigma-1}\tau^{1-\sigma}\left[(c_{2f})^{1-\alpha}\right]^{1-\sigma}\left((\tau c_{1d})^{1-\rho} + n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(c_{1f})^{1-\rho} - f_1^{'}(n_1),\end{aligned} \tag{A.2}$$

and

$$\frac{\partial\pi(n_1,0|\varphi,\beta_{2R})}{\partial n_1} = B\varphi^{\sigma-1}\left[(c_{2d})^{1-\alpha}\right]^{1-\sigma}\left(c_{1d}^{1-\rho}+n_1(\tau c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(\tau c_{1f})^{1-\rho}+$$
$$B\varphi^{\sigma-1}\tau^{1-\sigma}\left[(c_{2f})^{1-\alpha}\right]^{1-\sigma}\left((\tau c_{1d})^{1-\rho}+n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(c_{1f})^{1-\rho}-f_1'(n_1). \tag{A.3}$$

Since β_{2R}, it can be shown that

$$\frac{\partial\pi(n_1,1|\varphi,\beta_{2R})}{\partial n_1} > \frac{\partial\pi(n_1,0|\varphi,\beta_{2R})}{\partial n_1}. \tag{A.4}$$

In words, Equation A.4 says that the marginal gains in profits from an increase in the imported number of inputs, n_1, is higher when the firm adopts robots ($I_R = 1$) than when the firm does not ($I_R = 0$). This result directly leads us to conclude that $\pi(n_1, I_R|\varphi, \beta_{2R})$ is supermodular in the choice variables, $X = (n_1, I_R)$.

For the second step, I show that the profit function $\pi(n_1, I_R|\varphi, \beta_{2R})$ exhibits increasing differences in the choice variables $X = \{n_1, I_R\}$ and parameters $\{\varphi, \beta_{2R}\}$. I proceed by examining each possible pair of choice variable and exogenous parameter. First, fixing $I_R = 1$, observe that

$$\frac{\partial^2\pi(n_1,1|\varphi,\beta_{2R})}{\partial n_1\partial\varphi} = B(\sigma-1)\varphi^{\sigma-2}\left[\left(\frac{c_{2d}}{\beta_{2R}}\right)^{1-\alpha}\right]^{1-\sigma}\left(c_{1d}^{1-\rho}+n_1(\tau c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(\tau c_{1f})^{1-\rho}+$$
$$B(\sigma-1)\varphi^{\sigma-2}\tau^{1-\sigma}\left[(c_{2f})^{1-\alpha}\right]^{1-\sigma}\left((\tau c_{1d})^{1-\rho}+n_1(c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(c_{1f})^{1-\rho}>0, \tag{A.5}$$

and

$$\frac{\partial^2\pi(n_1,1|\varphi,\beta_{2R})}{\partial n_1\partial\beta_{2R}} = B\varphi^{\sigma-1}\left[(c_{2d})^{1-\alpha}\right]^{1-\sigma}(1-\alpha)(\sigma-1)\beta_{2R}^{(1-\alpha)(\sigma-1)-1}$$
$$\left(c_{1d}^{1-\rho}+n_1(\tau c_{1f})^{1-\rho}\right)^{\frac{\alpha(1-\sigma)}{1-\rho}-1}\frac{\alpha(1-\sigma)}{1-\rho}(\tau c_{1f})^{1-\rho}>0. \tag{A.6}$$

Analogous results can be shown for cases where $I_R = 0$. Next, fixing n_1 and defining the marginal gains in profits from robot adoption as

$$A(n_1, \varphi, \beta_{2R}) = \pi((n_1, 1|\varphi, \beta_{2R}) - \pi((n_1, 0|\varphi, \beta_{2R}), \tag{A.7}$$

it follows that

$$\frac{\partial A(n_1, \varphi, \beta_{2R})}{\partial \varphi} = B(\sigma - 1)\varphi^{\sigma-2} \cdot$$
$$\left(\left[\left(c_{1d}^{1-\rho} + n_1 (\tau c_{1f})^{1-\rho} \right)^{\frac{\alpha}{1-\rho}} \left(\frac{c_{2d}}{\beta_{2R}} \right)^{1-\alpha} \right]^{1-\sigma} - \left[\left(c_{1d}^{1-\rho} + n_1 (\tau c_{1f})^{1-\rho} \right)^{\frac{\alpha}{1-\rho}} (c_{2d})^{1-\alpha} \right]^{1-\sigma} \right) > 0, \tag{A.8}$$

and

$$\frac{\partial A(n_1, \varphi, \beta_{2R})}{\partial \beta_{2R}} = B\varphi^{\sigma-1} \left[\left(c_{1d}^{1-\rho} + n_1 (\tau c_{1f})^{1-\rho} \right)^{\frac{\alpha}{1-\rho}} (c_{2d})^{1-\alpha} \right]^{1-\sigma} (1-\alpha)(\sigma-1)\beta_{2R}^{(1-\alpha)(\sigma-1)-1} > 0. \tag{A.9}$$

This completes the proof of increasing differences between the choice variables and exogenous parameters.

Finally, based on the two intermediate findings, I apply the Topkis monotonicity theorem to conclude that the optimal choice of n_1, I_R that maximizes the profit function is increasing in the strong set order on exogenous parameters φ, β_{2R}.

A.1.2 Proof of Proposition 2 and Proposition 3

Consider the decomposition of the change in upstream imports shown in Equation 2.8. The assembly productivity effect is clear positive. Meanwhile, note that the following inequality can be established for any φ:

$$\Delta \log n_1(\varphi) \geq \Delta \log \left(c_{1d}^{1-\rho} + n_1(\varphi)(\tau c_{1f})^{1-\rho} \right), \tag{A.10}$$

which implies that the sum of the import extensive margin effect and the input productivity effect in Equation 2.8 is greater than the expression shown in Equation 2.9. Given that robot adoption must induce the choice of a higher n_1, Equation 2.9 must be positive. In other words, robot adoption is associated with an increase in downstream imports. Furthermore, the discussion above implies that Equation 2.8 is greater than Equation 2.9. This not only proves that robot adoption is associated with an increase in upstream imports, but also completes the proof of Proposition 3.

A.2 Data

A.2.1 Constructing the main dataset

The main dataset used for empirical analysis in section 2.3 and section 2.4 connects three firm-level US Census datasets: the Longitudinal Business Database (LBD), the Longitudinal Foreign Trade Transactions Database (LFTTD), and the Economics Censuses (EC). I start by obtaining information on employment, payroll, and industry classification from the LBD, which contains records on the establishment level and provides a unique firm identifier for every business operating in the US between 1992 and 2016. I exclude all establishments with negative employment or payroll. In addition, I restrict my attention to manufacturing firms, defined as firms containing at least one manufacturing establishment. A manufacturing establishment is in turn defined as one whose NAICS code according to the Fort (2017) definition starts with a "3", corresponding to manufacturing industries. I thus aggregate the LBD to a firm-year-level dataset, containing manufacturing firms only.

I then incorporate information on imports and exports obtained from the LFTTD. For imports, I calculate the overall value of imports at the firm-year level, as well as input and output imports as defined in subsection A.2.5. Given that robot adoption is identified using imports, I exclude imports of industrial robots from all import definitions. In addition, in the baseline specification for upstream and downstream imports, I exclude all imports of machinery (HS 84 or 85). This final step is taken to ensure that the effect of robots on imports is not driven by idiosyncratic factors making firms more likely to adopt machinery in general.

Finally, I incorporate data on sales and input expenditure from the EC. Specifically, I take sales from all economic censuses, and calculate input expenditure as the sum of the total cost of materials and total compensation for production workers from the Census of Manufactures (CMF). Since these data are available every 5 years, I impute data for non-census years by conducting a linear interpolation on the logarithmic transformation of sales and input expenditure. This method of interpolation is based on the assumption that these variables grow at a constant rate over time.

A.2.2 Details on HS codes for robots

HS codes for robots are identified using archive files for the US Harmonic Tariff Schedule (HTS) from 1992 to 2016, which are accessed online[1]. In every year, there exists two codes corresponding to two categories of industrial robots. The first category is described as "*industrial robots, not specified elsewhere*"; the second category is described as "*other lifting, handling, loading or unloading machinery: industrial robots*". The specific codes for each category vary over time, and are summarized below:

- Category 1: HS code 8479899040 from 1992 to 1994; HS code 8479899540 in 1995; 6-digit HS code 847950 from 1996 to 2016;
- Category 2: 8428900010 from 1992 to 2005 (except in 1998, when it was 8428908015), 8428900120 from 2006 to 2011, 8428900220 from 2012 to 2016.

A.2.3 Identifying robot adopters

As discussed in the main text, to identify robot adopters, I start by identifying firms that import industrial robots based on the codes listed in Appendix A.2.2. Given that not all robot importers necessarily utilize robots in the production process, I then take several

[1] URL: https://usitc.gov/tata/hts/archive/index.htm

approaches to eliminate potential carry-along (wholesale) traders, and potential robot producers purchasing robots assembled abroad as final output or as intermediate inputs.

First, for each robot importer, I compute the total value of robot imports and exports between 1992 and 2016. As defined in Bernard *et al.* (2019), carry along traders are firms exporting products that they do not manufacture. However, in the case of capital goods such as industrial robots, a firm may simultaneously engage in carry along trade *and* adoption of robots in its own production. Therefore, I only eliminate firms whose total value of robot exports exceed robot imports. The rationale is that these firms do not retain any of their purchased robots within the US, and thus do not utilize robots in production with certainty[2].

Second, I use material trailers (MT) and product trailers (PT) of the Census of Manufactures (CMF) to identify firms that produce robots or use robots as intermediate inputs. The CMF contains data on the universe of manufacturing establishments on years ending in "2" or "7" ("Census years"). The MT and PT contain detailed information on input purchases and sales by product for a selected sample of firms. MT and PT follow separate product classification systems based on SIC (in 1992) or NAICS (in 1997 and onwards). In every Census year between 1992 and 2012, there exists a unique product code referring to "industrial robots", and a unique material code referring to "Industrial robots purchased for fabrication with welding equipment" The codes are:

- Year 2012: Product code 3339997109, appears in forms 33319, 33331, 33373
- Year 2007: Product code 3339997109, appears in forms 33319, 33331, 33373
- Year 2002: Product code 33399981092, appears in forms 33319, 33331, 33373
- Year 1997: Product code 35699099, appears in forms 3505, 3506, 3529, 3538, 3539

[2]In addition to reselling robots abroad, firms may also resell robots to other domestic firms. Therefore, even if a firm's robot exports are lower than its robot imports, it might still only engage in transit trade and not utilize robots in production. However, in the data, I do not observe domestic sales by product for the majority of firms.

- Year 1992: Product code 35699099, appears in forms 3505, 3506, 3514, 3529

- Year 2012: Material Code 33399903, appears in forms 33326, 33331
- Year 2007: Material Code 33399903, appears in forms 33326
- Year 2002: Material Code 333999035, appears in forms 33326
- Year 1997: Material Code 3569712, appears in forms 3529
- Year 1992: Material Code 3569712, appears in forms 3529

Based on these codes, I identify establishments that potentially produce robots and/or use robots as intermediate inputs. I then locate all firms containing such establishments and eliminate them from the list of robot adopters.

Lastly, since product-level input and output information is not available for all firms or all years, I exclude a broader set of potential robot producers identified using establishment-level information on NAICS classification. Specifically, I exclude firms whose establishments fall into the following NAICS categories: 333999 (All Other Miscellaneous General Purpose Machinery Manufacturing), 3335 (Metalworking Machinery Manufacturing, which includes molding, cutting, and forming machinery), 33392 (Material Handling Equipment Manufacturing), 333992 (Welding and Soldering Equipment Manufacturing).

A.2.4 Details on robot producers

Using the methodology outlined in subsection A.2.3, I identify 400 robot-producing firms in the United States. Since I do not directly observe their total production of robots, I infer domestic sales through their robot exports and export intensity. Over the time period from 1992 to 2016, these robot producers exported a total of 19,900 industrial robots, amounting to a total value of $6 billion. Robot producers are highly export-oriented, with an overall export intensity of 35%. Based on these observations, I infer that these firms sell 37,000 industrial robots to firms in the US. Given that the total number of industrial robots imported by US manufacturing firms during the same time period is 192,000, I conclude that domestic robot manufacturers only account for 16% of industrial robots used in the US. It is thus reasonable to assume that identifying robot adoption using robot imports yields a representative sample.

A.2.5 Classifying input and output imports

In the main text, I outlined a methodology to identify input and output imports at the firm level. The methodology relies on input and output information extracted from Material Trailers (MT) and Product Trailers (PT) of the Census of Manufactures (CMF), and connects inputs used and outputs produced by each firm (and firms in the same industries) with its imports observed from LFTTD. In this section, I describe this methodology in detail.

I start by identifying inputs and outputs used by each manufacturing firm between 1992 and 2016. The starting point are the MT and PT, which provide material and product codes for intermediate inputs used and outputs produced by a selected sample of manufacturing establishments. Since these information is not available for the universe of manufacturing establishments, I aggregate them to the NAICS-level in the following sense: a given product is classified as a potential input for a 6-digit NAICS industry as long as some establishment classified under the industry is recorded to use the product as an input according to the Material Trailer files. Analogously, a product is classified as a potential output for a NAICS industry as long as some establishment in that industry produces the product as an output according to the Product Trailer files. The NAICS codes used are based on the time-consistent classification developed by Fort and Klimek (2018). This process yields a list of potential inputs and outputs for each NAICS industry in Census years. To extend this information to a yearly basis, I apply the list in 1992 to years 1993-1996, the list in 1997 to years 1998-2002, and so on. The end product is a list of inputs and outputs for all manufacturing industries in all years from 1992 to 2016.

Next, I assign these industry-level information to firms based on their industry composition. In other words, a product is classified as an input for a firm if it is classified as an input for any industry the firm operates in. A similar idea applies to outputs. Note that a product can be simultaneously classified as an input and an output for a given firm. This process results in a dataset of potential inputs and outputs at the firm-year level.

Finally, I use the constructed firm-level input-output information to flag all imports purchased by US manufacturing firms between 1992 and 2016. To this end, I create a concordance between material/product codes and HS codes, using SIC and NAICS codes as the middle point. As discussed in the main text, I flag an imported product as an input if it corresponds to a potential input used by the importing firm. Meanwhile, I flag an imported product as an output if it corresponds to a potential output, and *not* a potential input.

A.2.6 Complete list of robot-complementary occupations

The full list of robot-complementary occupations identified using the O*NET connector is displayed in Table A.1.

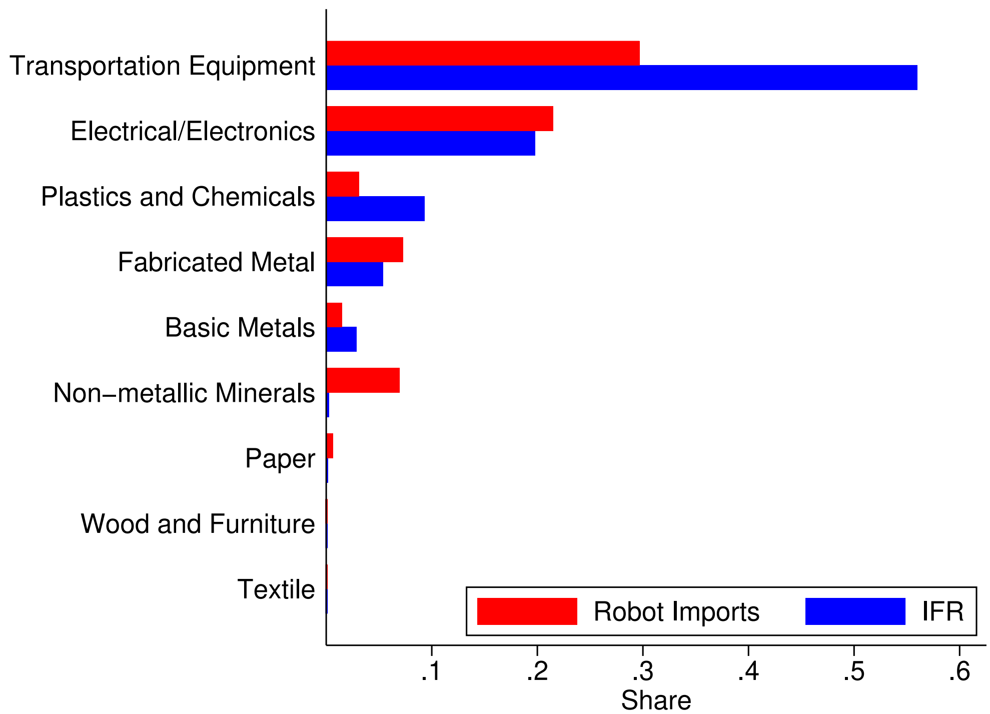
A.2.7 Constructing the immigrant shock

As discussed in the main text, the CZ-year level instrument $z_{c,t}$ in Equation 2.15 is constructed as a weighted sum over country origin groups. The idea behind these country origin groups is that immigrants in a group have a tendency to settle following geographical settlement patterns of previous immigrants from the same group. I separate all immigrants into 15 groups: Mexico, Cuba, Central America, South America, Southeastern Asia and Pacific Islands, Philippines, Caribbeans, Southwestern Europe, Eastern Europe and Russia, Africa, China, Japan and Korea, the Middle East, UK, Australia & New Zealand, Northern Europe, and Southwestern Asia.

Table A.1: *Full list of robot-complementary occupations identified by the O*NET connector*

Occupation Name	Match Score
Robotics Technicians	100
Robotics Engineers	100
Computer Numerically Controlled Tool Operators	65
Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	60
Electro-Mechanical and Mechatronics Technologists and Technicians	58
Electrical and Electronics Repairers, Commercial and Industrial Equipment	25
Commercial and Industrial Designers	25
Welders, Cutters, Solderers, and Brazers	25
Multiple Machine Tool Setters, Operators, and Tenders, Metal and Plastic	12
Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	7
Computer and Information Research Scientists	3
Mechanical Engineers	3
Biological Technicians	3
Cytogenetic Technologists	3
Pharmacy Technicians	3
Construction Laborers	3
Millwrights	3
Electromechanical Equipment Assemblers	3

Figure A.1: *Share of all robot-using industries among all manufacturing industries*



Notes: This figure plots the share of all robot-using industries among all manufacturing industries in terms of the number of robots used. For robot imports, I calculate the total number of robots imported by firms in each industry between 1992 and 2016, restricting to robot-adopting firms. For IFR, I calculate the total number of robots operating in each industry as recorded by IFR data.

A.2.8 Cross-industry comparison for all manufacturing industries

A.3 Quantitative Exercise

A.3.1 Re-establishing equilibrium for Counterfactuals

In this section, I provide details on how I re-establish equilibrium for the counterfactual exercises. For all 3 exercises, I start from the estimated equilibrium in 2012, and modify selected parameters such that free-entry condition is satisfied.

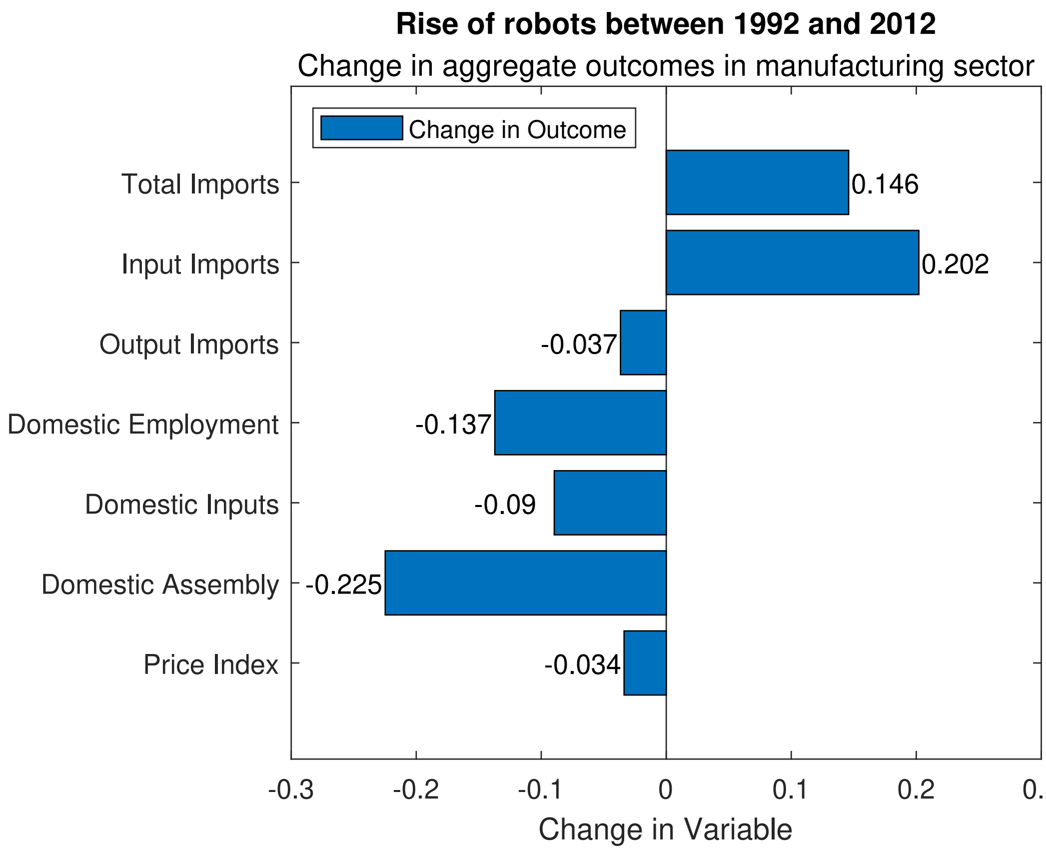
For counterfactual 1, starting from the estimated equilibrium in 2012, I assume that both c_{1R} and c_{2R}, costs of production under robots, have decreased by the same proportion β between 1992 and 2012. I then calibrate the value β as well as a new demand factor B such that (1) the share of robot adopters matches the observed share in 1992 and (2) the free entry condition in Equation 2.18 is satisfied. In this process, I assume all other parameters of the model, including the cost of entry, stay unchanged. Specifically, I calibrate the values of β and B using a bounded Nelder-Mead simplex search method. After obtaining estimates of β and B, I calculate equilibrium outcomes in 1992 and compare then with outcomes in 2012 to gauge the effect of a rise in robots. For counterfactuals 2 and 3, I also start from the estimated equilibrium in 2012, and calibrate a new demand factor B such that the free entry condition is satisfied after an increase in productivity of robots or after a robot policy is implemented.

A.3.2 Figure displaying results under counterfactual 1

A.3.3 Fixing extensive margin of imports for counterfactual 1

I consider the same increase in robot productivity as in counterfactual 1, but under the assumption that firms keep their choices of n_1 and n_2 unchanged. I then calculate the changes in aggregate outcomes under fixed extensive margins of importing, and compare

Figure A.2: *Model-implied change in aggregate outcomes following robot productivity shock: flexible vs. fixed sourcing strategy*



Notes: this figure plots the change in seven aggregate outcomes following an increase in robot productivity under two assumptions. "Flexible strategy" assumes that firms are able to freely adjust their imported number of inputs and outputs, n_1 and n_2. "Fixed strategy" assumes that firms do not adjust their imported number and outputs.

them to those under the baseline model displayed in Table 2.11.

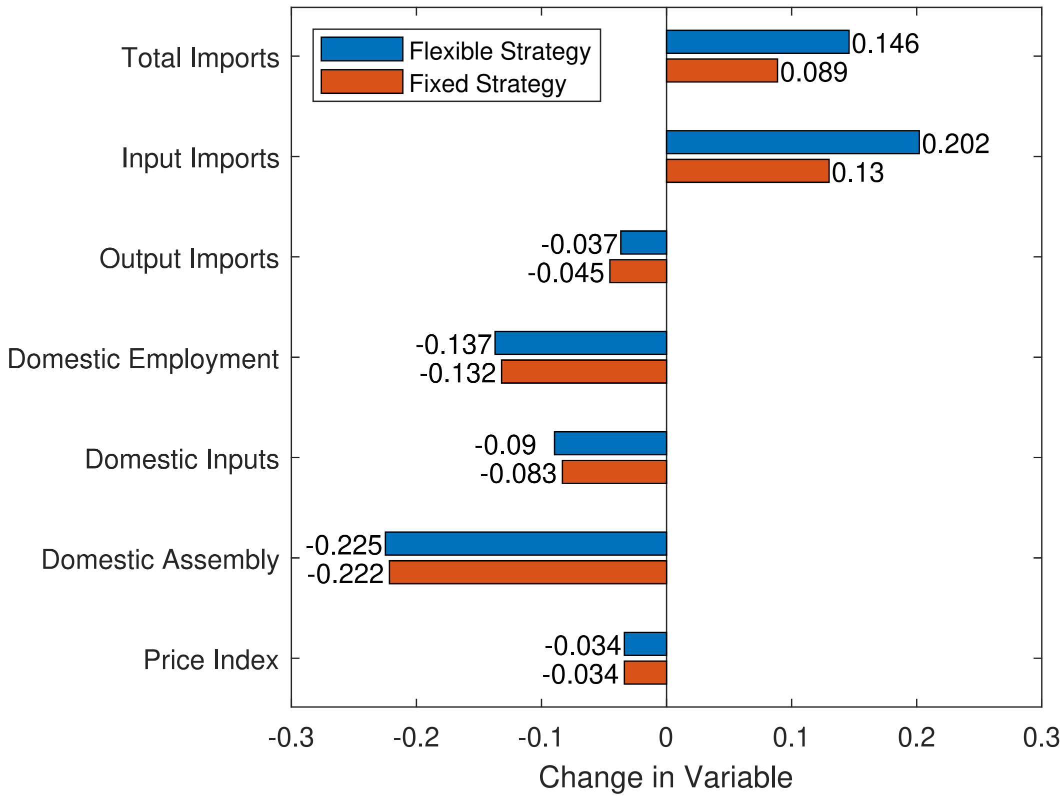
The comparison is illustrated in Figure A.3. Blue bars correspond to changes in aggregate outcomes following the calibrated robot productivity shock between 1992 and 2012, under the assumption that firms are allowed to freely adjust their imported number of inputs and outputs, n_1 and n_2. Red bars correspond to changes in aggregate outcomes following the same shock, under the assumption of "fixed strategy", i.e. firms hold on to the same choices of n_1 and n_2. Allowing firms to adjust the extensive margin of imports turns out to have a significant impact on aggregate changes in imports, particularly input imports. Under fixed strategies of importing, the rise of robots between 1992 and 2012 would only be associated with a 8.9% increase in imports of the manufacturing sector, which is 5.7 percentage points lower than the increase under flexible strategies. In other words, around a third of the aggregate increase in imports is driven by firm-level adjustments at the extensive margin. Meanwhile, fixing sourcing strategies does not have a quantitatively significant impact on the changes in employment or the price index.

A.3.4 Alternative to counterfactual 1: decrease in fixed costs

In subsection 2.6.1, I calibrate the observed increase in the share of robot-adopting firms under the assumption that it is only driven by an increase in productivity of robots. In reality, this increase is also potentially driven by a decrease in fixed costs of robot adoption. In fact, an extreme alternative would be to assume that the rise of robots is *only* due to a decrease in fixed costs.

I conduct a counterfactual exercise under that alternative assumption. Results of this exercise are presented in Table A.2. In order to explain the rise in the share of robot adopters from 0.022% in 1992 to 0.4% in 2012, the fixed costs of robot adoption in 1992 needed to be 9.5 times as large as fixed costs of robot adoption in 2012. Given this decrease in fixed costs between 1992 and 2012, overall outcomes of the manufacturing sector still respond in the same direction as in the baseline counterfactual: the robot shock induces an increase

Figure A.3: *Model-implied change in aggregate outcomes following robot productivity shock: flexible vs. fixed sourcing strategy*



Notes: this figure plots the change in seven aggregate outcomes following an increase in robot productivity under two assumptions. "Flexible strategy" assumes that firms are able to freely adjust their imported number of inputs and outputs, n_1 and n_2. "Fixed strategy" assumes that firms do not adjust their imported number and outputs.

in total imports, a decrease in domestic manufacturing employment, and a decrease in the consumer price index. However, the magnitude of the changes is consistently smaller than the ones shown in Table 2.11. This is because firms that already adopt robots in 1992 *do* benefit from an increase in the productivity of robots, but they *do not* benefit from a decrease in fixed costs.

In reality, the increase in productivity of robots and the decrease in fixed cost of adoption probably both contributed to the rise of robots. Therefore, the changes in aggregate outcomes estimated under the two sets of assumptions can be interpreted as bounds for the true impact of the rise of robots between 1992 and 2012.

Table A.2: *Impact of the rise in robots: decrease in fixed costs, 1992-2012*

	Always-adopters	New adopters	Non-adopters	Overall
Share of Firms (%)	0.022	0.378	99.6	1
Total Sales	-0.075	0.705	-0.072	0
Total Imports	-0.082	0.527	-0.127	0.038
Input Imports	-0.076	0.589	-0.132	0.067
Output Imports	-0.103	0.224	-0.114	-0.065
Domestic Employment	-0.070	-0.031	-0.071	-0.137
Domestic Inputs	-0.070	0.272	-0.070	-0.057
Domestic Assembly	-0.073	-0.504	-0.071	-0.153
Consumer Price index				-0.024

Notes: "Always adopters" are firms that adopt robots in both 1992 and 2012 according to the calibrated model. "New adopters" are firms that do not adopt robots in 1992 but switch to adopting robots in 2012. "Non-adopters" are firms that do not adopt robots in either year.

A.3.5 Table for counterfactual 3

Table A.3: *Change in aggregate outcomes under different robot policies*

	Investment Tax	Robot Income Tax	Adoption Subsidy
Total Imports	0.012	-0.025	-0.015
Input Imports	0.010	-0.044	-0.013
Output Imports	0.016	0.031	-0.019
Domestic Inputs	0.006	0.054	-0.008
Domestic Assembly	0.019	0.135	-0.027
Price index	0.006	0.022	-0.005

Notes: "Always adopters" are firms that adopt robots both before and after the increase in robot productivity. "New adopters" are firms that do not adopt robots in the baseline model but switch to adopting robots after the increase in robot productivity. "Non-adopters" are firms that do not adopt robots in either scenario.