Summer 8-4-2020

Empirical Essays on Automation and Its Implications

Soeun Park
Southern Methodist University, soeunp@smu.edu

EMPIRICAL ESSAYS ON AUTOMATION
AND ITS IMPLICATIONS

Approved by:

Dr. Daniel Millimet
Professor of Economics

Dr. Tom Fomby
Professor of Economics

Dr. James Lake
Associate Professor of Economics

Dr. Wenhua Di
Senior Economist of the Federal Reserve
Bank, Dallas

EMPIRICAL ESSAYS ON AUTOMATION
AND ITS IMPLICATIONS

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Economics

by

Soeun Park

B.A., Economics, Sungkyunkwan University, Seoul, Korea
Master of Economics, Sungkyunkwan University, Seoul, Korea

August 4, 2020

ACKNOWLEDGEMENTS

This work could not have been accomplished without the wisdom and the support of my wonderful advisor, Prof. Millimet. I am thankful to Daniel Millimet, James Lake, Tom Fomby, and Elira Kuka for their valuable advice. I am also grateful to attendees in the Brown Bag seminar within the Department of Economics at SMU for helpful comments. The Department of Economics at SMU is a great place to learn academic knowledge as well as gives me valuable experiences in my life. I sincerely appreciate all the support and encouragement from my advisor and other professors. I also thank my family and my best friends for fully supporting me.

Park, Soeun B.A., Economics, Sungkyunkwan University, Seoul, Korea
Master of Economics, Sungkyunkwan University, Seoul, Korea

Empirical Essays on Automation
And Its Implications

Advisor: Dr. Daniel Millimet
Doctor of Philosophy degree conferred August 4, 2020
Dissertation completed July 22, 2020

This dissertation consists of three empirical studies in Economics. The first two chapters mainly focus on the effects of recent technological developments and its implications, by exploiting variation in exposure to industrial robot adoption across commuting zones. In chapter one, I empirically investigate whether low-income families are still supported by Earned Income Tax Credit (EITC) under automation. The analysis is inspired by the fact that welfare programs for low-income families in the U.S. have increasingly emphasized in-work aid over out-of-work aid since the 1990s, whereas recent technological progress such as robots and Artificial Intelligence (AI) is replacing human labor, not just routine-intensive jobs, which consequently implies that the U.S. safety net provides only modest benefits to non-workers. Based on the instrumental variable (IV) estimation, I find that there is no statistically significant difference in EITC usage across commuting zones and that EITC helps single parents out of poverty. Overall, there is no evidence that current EITC is ineffective in local labor markets experiencing growth in automation. In chapter two, I investigate the distributional effects of automation on income in the U.S. in the period 1990-2015. Applying the quantile regression framework for a group-level treatment in Chetverikov et. al. (2016), I examine the effect of each region's industrial robot adoption on the within-region income distribution. The results show that robots are much worse for low-income people as they experience a decrease in both hourly and annual earnings, in which the latter implies reduced hours worked due to robots. Also, automation makes the adverse effect of females on income stronger at the bottom of the distribution, and there is some evidence that it increases the positive effect of educational attainment on income above the middle of the distribution. Overall, the results suggest that industrial robots are more likely to harm less-skilled and low-wage workers. The last chapter empirically examines what determines the distribution of federal economic development funding in the U.S. at the county level. Since there are 425 counties in persistent poverty, I also explore whether there is any difference in the funding distribution between persistently poor counties and non-poor counties. The result presents that neither economic need nor political concerns affect the funding distribution on persistently poverty counties. Also, I find that there are no structural differences in the funding distribution.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

This dissertation is dedicated to Kyoungsook Lee, my dear mother and my best supporter forever. May she rest in peace.

CHAPTER 1

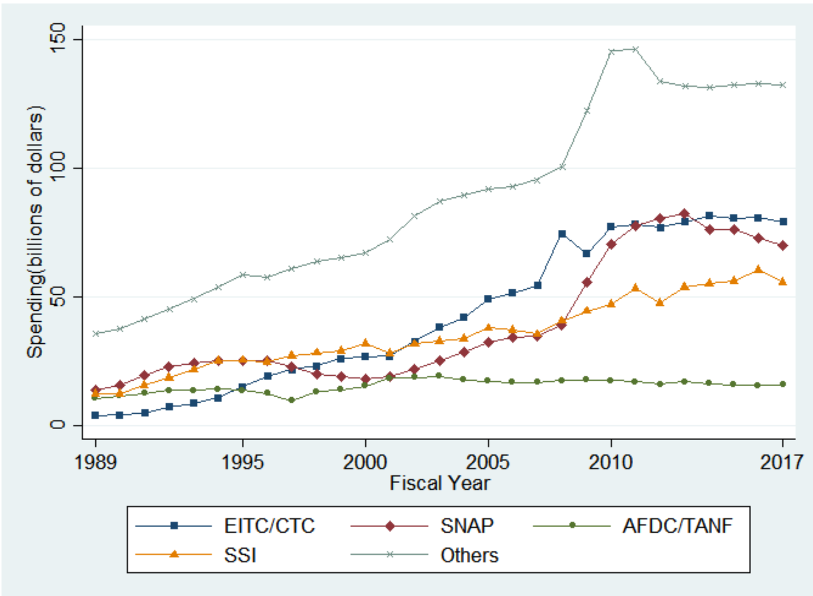
How Well Does EITC Help Low-Income Households in the Age of Automation?

1.1 Introduction

Google DeepMind developed AlphaGo, a computer program based on Machine Learning (ML), and it defeated Sedol Lee, one of the top professional Go players, 4-1 in March 2016. Since the board game Go has long been considered challenging work in the field of Artificial Intelligence (AI), the result was quite shocking and has made people fear how far AI and robot technology can develop and what they can do in the future.

As active research about the effect of recent technological development on labor markets is ongoing, frequently cited Frey and Osborne (2017) estimate that about 47% of total US employment is at high risk due to computerization in the next few decades due to recent advances in technology such as ML, AI, and Mobile Robotics (MR). This figure may be an overestimate for several reasons. First, they consider only technological capabilities which may exceed actual adoption of new technologies. Second, Arntz *et al*. (2016) point out that workers in the same occupation can actually do different tasks and that automation usually takes place at specific tasks rather than the occupation level, which show that only 9% of jobs in the U.S. are automatable based on the task-based approach. Third, the introduction of new technology can have indirect or equilibrium effects on the labor market, mitigating its negative direct impact, for example by increasing employment through improved productivity and the creation of new jobs. Regardless of the exact job loss prediction, in light of our historical experiences such as the first Industrial Revolution, one certainty is that some individuals will suffer from unemployment and wage adjustments due to technological developments, even if such developments are beneficial to society as a whole. More importantly, both studies suggest that automation is likely to be greater in jobs occupied by low-skilled and low-income workers.

While technology has progressed, making a much broader range of tasks or jobs automatable, the U.S. safety net for low-income individuals has evolved since the 1990s to require that they work to be eligible for many benefits. Earned Income Tax Credit (EITC), providing cash assistance to working families with low-income, was introduced in 1975 and has substantially expanded, especially during the 1990s; whereas the 1996 federal welfare reform changed the traditional cash programs by imposing work requirements. This leads to the growing importance of EITC, which now has become the primary safety net program for families with low- and moderate-income in the U.S. As shown in Figure 1.1, EITC has the largest expenditure as a single program, although the expenditure of Supplemental Nutrition Assistance Program (SNAP) has sharply increased due to work requirement waivers after the economic crisis in 2008.



Notes: Data, except for Aid to Families with Dependent Children (AFDC), come from https://www.usgovernmentspending.com. The specific spending categories for each program are based on the information at http://federalsafetynet.com. The EITC and the Child Tax Credit (CTC) includes only the refundable portion of the credit. Others include the following federal welfare programs: Housing Assistance from the Department of Housing and Urban Development, Child Nutrition such as school lunch, Head Start, Job Training, WIC (Women, Infant, and Children), Child Care, LIHEAP (Low Income Home Energy Assistance Program). The federal spending of AFDC is available at https://aspe.hhs.gov/system/files/pdf/167036/4spending.pdf.

Figure 1.1 The spending of safety net programs

Note that there are other cash transfer programs with work requirements that are less important than EITC, as measured by expenditure size. For example, Aid to Families with Dependent

Children (AFDC), one of the main safety net programs in the U.S. before 1996, did not require employment for eligibility for cash benefits, and in fact, its benefits fall with labor income. The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA) replaced AFDC with Temporary Assistance for Needy Families (TANF) that has strong work requirements. The other main safety net program is SNAP, previously known as Food Stamps. The 1996 PRWORA imposed work requirements on Able-Bodied Adults Without Dependents (ABAWD) to receive the SNAP benefits. However, the specific work requirements vary with programs. To be eligible for EITC, you must have earnings, but AFDC/TANF as well as SNAP accept work-related activities such as participation in job training. So, their work requirements are not as binding as EITC, and they still provide benefits to non-workers. Hence, in this paper, I focus on EITC since it is the most important safety net program and has a more explicit and stronger work requirement: positive earnings for a year that clearly reflect employment and can be easily recognized in public data.

The purpose of safety net programs with work requirements is generally to achieve income redistribution by providing income assistance through encouraging work; in particular, there is a substantial amount of literature about the positive effects of EITC on employment and poverty (Nichols and Rothstein, 2015). However, EITC provides no benefit to non-workers. If recent technological advances displace workers by automating a broad range of tasks, then EITC would not provide any assistance to displaced workers if they could not find other jobs. In this respect, there is a lively discussion recently about Universal Basic Income (UBI) that provides a certain amount of money for all citizens regardless of their income or employment status.[1] But, UBI without any eligibility criteria would require much more spending beyond the whole current safety net programs, as pointed out in Hoynes and Rothstein (2019), and we do not have strong evidence that recent technology will cause massive (long-term) unemployment yet.

Hence, my research starts with questioning whether automation overall pushes low- and moderate-income families outside the existing in-work benefits. Automation can make people ineligible for in-work benefits by either pushing them into (long-term) unemployment or by raising

[1] See Hoynes and Rothstein (2019) for further detail. The paper defines a UBI and discuss the potential role of UBI, by comparing it with the existing safety nets - welfare programs including TANF and SNAP, disability programs, Social Security Retirement, and in-work tax credit - and by providing possible effects of UBI on labor supply and human capital accumulation based on public policy related literature. The paper argues that generous UBI without eligibility would be much more expensive than the whole expenditure of current safety nets, which means that the government should substantially increase tax revenue.

earnings and pushing them beyond the eligibility range. Also, automation may push some people into existing in-work benefits if they lose their previous job but are able to find new low-wage employment that makes them eligible for EITC benefits, or if they simply have their wages reduced at their current job due to a reduction in hours as a result of fewer tasks to perform. To analyze it accurately, we need to be able to discern workers who are dismissed by automation and track their employment status (and welfare benefits) over time, which is not possible with the current public data. Thus, in this paper, I investigate this question at the local aggregate level indirectly by examining whether there is any difference across locations in the EITC usage, in terms of recipiency rates and average benefits, due to the extent of automation. Note that the interpretation of change in average benefits can be much more complicated because it is affected by changes in the number of EITC recipients as well as by changes in the earnings distribution of EITC filers, hence I focus more on recipiency rates. In addition, I also examine whether EITC still can reduce the poverty rate under automation.

Since a taxpayer should have positive earnings below a certain level to be eligible for the EITC, job loss or changing jobs due to automation will affect the EITC usage through changes in family earnings. However, it is not clear a priori how the EITC usage will change under automation. First, automation may displace some tasks and workers, thereby making them unemployed or reducing their (aggregate) working hours and earnings due to short-term unemployment, change into lower-paying jobs, or decrease in wages. So, the EITC recipiency rate at the local aggregate level may decrease due to increased unemployment, or it may increase if automation moves more families into the eligible range of earnings. Second, automation may improve the productivity of some workers if it is complementary to human labor or increases demand for workers in non-automated tasks due to the reduction in production cost even if it substitutes for human labor. Overall, it can increase the earnings of some workers due to the increase in wages or by providing opportunities to get higher-paying jobs, and thus it may decrease the EITC recipiency rate by moving some families beyond the eligible region of earnings. Third, automation may create new tasks and jobs so that it can increase employment, aggregate working hours, or earnings, but the change in the EITC recipiency rate depends on for whom newly created tasks/jobs are more favorable (ex. low-skilled or high-skilled people). Lastly, all of these effects of automation may differ across locations depending on the share of routine (or automatable) tasks/jobs.

To sum up, the effects of automation on the labor market are complex, potentially moving family earnings and EITC recipiency rates in different directions. Therefore, overall changes in the EITC usage depend on the strength of different effects. On the one hand, the EITC usage could increase in locations with higher automation if workers displaced by automation keep working by finding another (less-paid) job. In this case, we can say that EITC still supports low- and moderate-income families. On the other hand, the EITC recipiency rate could be lower in regions with higher automation if displaced workers cannot find a job or if improvements in productivity and creations of new tasks/jobs paying relatively high wages are stronger than the other effects so that there are more families with earnings outside of eligible criteria. While the former case may suggest the decreased ability of EITC supporting low-income families, the latter case does not.

To empirically examine the question of whether EITC still can support low-income families under technology-induced shocks to labor markets, automation is approximated by the adoption of industrial robots following Acemoglu and Restrepo (2017). To be specific, my empirical strategy exploits variation in exposure to industrial robot adoption across local labor markets, defined as commuting zones, to estimate the relationship between the extent of automation and both the EITC beneficiary and benefits. For the analysis, I use industrial robot data from the International Federation of Robotics (IFR) and Census micro-data from the IPUMS.

The results suggest that automation overall does not affect EITC usage across commuting zones, but that automation could have heterogeneous effects across commuting zones depending on the initial employment share in routine jobs. For single filers with children, the evidence is suggestive of a positive effect on recipiency rate in commuting zones with high routine share and no effect in commuting zones with low routine share. The increase in recipiency rate in commuting zones with high routine share seems to happen due to the movement of single filers with children into the lower part of the distribution of earnings. Considering that their earnings are more concentrated below the overall median earnings, they could be relatively more exposed to the displacement effect. The estimates also suggest the increase in the shares of those with earnings slightly above the EITC-eligible range, which implies that the expansion of EITC eligibility could reach more families with automation, given that their earnings above the EITC-eligible region are still lower than the overall median family household income.

For married filers with children, automation significantly decreases the average benefit in commuting zones with low routine share. But, it seems to come from a decrease in their recipiency rate in commuting zones, especially with low routine share, due to their movement into the higher part of the distribution of earnings, although the estimates are imprecise. In addition, the decline in the share of them being able to receive the maximum benefit of EITC could partly explain the decrease in their average benefits. Overall, considering their higher potential earnings and two potential earners, they are likely to be more exposed to other positive effects of automation.

Lastly, I focus on the ability of EITC to reduce poverty under automation by using changes in the ratio of tax-filers below certain levels of official poverty thresholds as outcome variables. The results show that EITC still can help single parents rise out of poverty under robot adoption, which especially concentrates on single filers with children whose income is either under 50 percent or under 200 percent of official poverty thresholds.

What policy should be set up for the age of automation is a quite broad question, but when we especially focus on low- and middle-income families affected by labor-replacing technology, the expansion of EITC or the increase in minimum wage level are often suggested as possible policy tools.[2] To the best of my knowledge, this paper is the first empirical study investigating how automation is moving families within the earnings distribution and how that affects EITC usage to understand how EITC supports low-income families under automation.[3] When considering that automation can negatively affect labor demand, my work contributes to the literature on EITC under economic downturns (Bitler *et al*., 2017; Jones, 2015). Also, my work contributes to the literature on the discussion about impacts of automation and its policy implications (Goos, 2018; Hoynes and Rothstein, 2019; Lordan and Neumark, 2018).

The rest of the paper proceeds as follows. Section 1.2 outlines the EITC, the impacts of recent labor-replacing technology, and the relevant policy discussion. Section 1.3 describes the empirical specification and data. The results are presented in Section 1.4, and I conclude in Section 1.5.

[2] There are a couple of studies about the effect of automation on low-skilled employment by using minimum wage changes. They show mixed results about the ability of minimum wage policy to mitigate the negative impact of automation.

[3] Although I particularly focus on automation in this paper, it can be generally interpreted as one source of systemic negative shocks to employment.

1.2 Background and related literature

1.2.1 Brief background about Earned Income Tax Credit (EITC)

The transformation from out-of-work aid to in-work aid is one of the most important changes that the U.S. safety net programs have undergone since the 1990s. The Earned Income Tax Credit (EITC), a kind of negative income tax, was first introduced in 1975 and its benefit levels and criteria have changed several times, though the most dramatic modification was made by the Omnibus Budget Reconciliation Act of 1993: the expansion of credit amount and the credit differences by the number of children. Now, it has become the most important cash transfer program of federal welfare programs, and the total amount of the refundable portion of the credits reached about $58.8 billion in 2015. Figure 1.1 shows the expenditure of major federal safety net programs: Negative Income Tax (the EITC and Child Tax Credit), Supplemental Nutritional Assistance Program, Temporary Assistance for Needy Families, Supplemental Security Income, and Others.[4] The expenditure on the EITC has been increasing and accounts for about 16.6 percent of the total spending on major federal safety net programs in 2015.

Today many states have their own EITCs that add to federal EITC. As of 2018, twenty-nine states and the District of Columbia have enacted their own EITCs. Of them, 23 states and DC have refundable EITCs, while 6 states have non-refundable EITCs.[5] More importantly, nearly all of them follow federal EITC eligibility rules, and also they determine the credits as a specific percentage of the federal credit (although it differs across states). Given state EITCs built on federal EITC, the number of people covered by EITC hardly changes when adding state EITCs. Thus, I focus on only federal EITC in this paper.

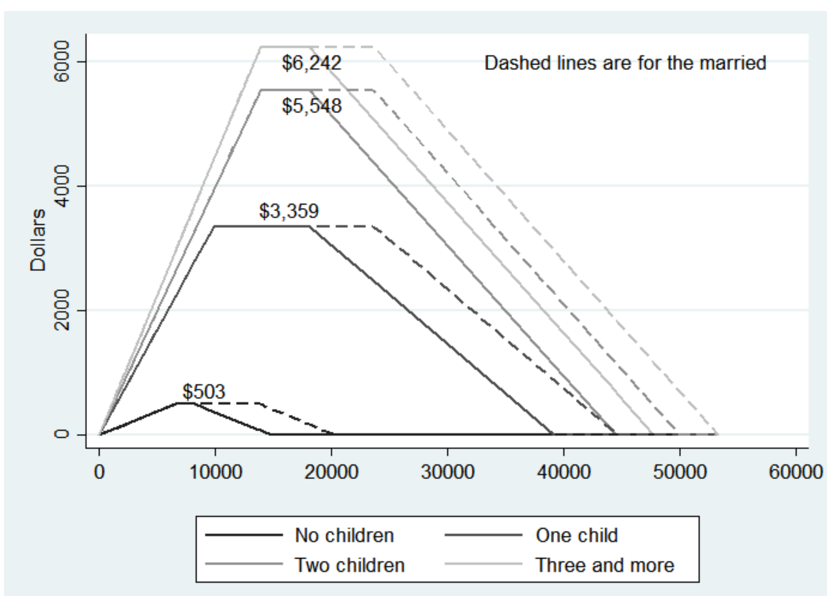
To be eligible for the EITC, taxpayers must work (as employees or self-employed), and their earnings should lie in a certain range. The earnings range for EITC is divided into three regions: i) phase-in, ii) plateau, and iii) phase-out. In the phase-in region, the credit constantly increases at

[4] Others include the following federal welfare programs: Housing Assistance from the Department of Housing and Urban Development, Child Nutrition such as school lunch, Head Start, Job Training, WIC (Women, Infant, and Children), Child Care, LIHEAP (Low Income Home Energy Assistance Program).

[5] Puerto Rico is not eligible for federal EITC and has a separate schedule with a maximum credit of between $300 and $2,000 based on family size. States with refundable EITCs are WA, OR, CA, MT, CO, NM, NE, KS, MN, IA, LA, WI, IL, MI, IN, ME, VT, NY, MA, RI, CT, NJ, MD, and DC. States with non-refundable EITCs are OK, OH, DE, VA, SC, and HI. For more details about the state EITC, see https://www.cbpp.org/research/state-budget-and-tax/states-can-adopt-or-expand-earned-income-tax-credits-to-build-a

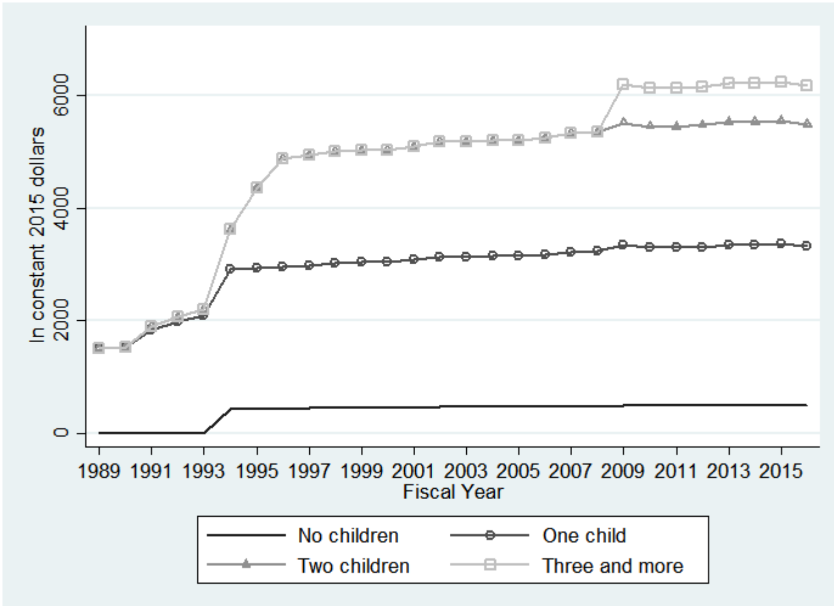
the phase-in rates, ranging from 7.65% to 45% depending on the number of children. For example, you receive 40 cents for an additional dollar of earnings if you have two children. In the plateau region, the credit remains constant at the maximum amount. In the phase-out region, the credit keeps decreasing at the phase-out rates of between 7.65% and 21.06%. So, if you have two children, your credit reduces by 20 cents for an extra dollar of earnings until the phase-out ends, where the credit equals to zero.

The thresholds for each region of earnings vary with marital status and the number of children. Since the eligible earnings ranges widen by the filing status and the number of children, EITC can support not only low-income working families but also a part of middle-income families, especially families with children. Figure 1.2 shows that, when you have two children, the maximum earnings eligible for the EITC goes to $44,454 for a single tax filer or $49,974 for a married tax filer in 2015, but the eligible income ranges for tax filers without children are quite limited. The credit difference by the number of children is more apparent in Figure 1.3, which plots the maximum credit by the number of children over time and all values are in 2015 US$. Even though it shows a large expansion before and after 1996 due to OBRA93, the credit for childless filers is still substantially small.



Notes: The figure is illustrated by author based on EITC parameters which are available at www.taxpolicycenter.org.

Figure 1.2 The EITC by the number of children and the marital status in 2015



Notes: All values are in 2015 US dollars using the Personal Consumption Expenditures (PCE) price index.

Figure 1.3 The maximum credit by the number of children

The extensive literature studying EITC shows its positive employment effect, especially on single mothers, and its success in helping low-income families out of poverty (Nichols and Rothstein, 2015; Hoynes and Rothstein 2016).[6] My work may be more related to a few studies about EITC under economic downturns when considering the negative impact of automation on employment. While high unemployment rates can arise from negative aggregate supply or demand shocks, automation is one of the sources causing negative shocks on labor demand directly, though it can also derive other positive impacts on employment as discussed below. Bitler *et al*. (2017) and Jones (2015) explore whether the EITC can respond to economic downturns by using the state unemployment rate as the measure of the economic downturn. Their results suggest that the stabilizing effect of the EITC against high unemployment rates is concentrated on married couples with children who are capable of cushioning negative effects from the recession, whereas there is no significant effect on single parents who are the majority of EITC recipients.

[6] Note that the most recent paper, Kleven (2019), provides that the employment effects of EITC in the 1990s are closely related with exposure to welfare reform (ex. state waivers and the replacement of AFDC) aided by favorable economic conditions.

1.2.2 Related literature

Let me briefly discuss the impacts of automation on employment and wages, which are indirectly reflected in the EITC participation and benefit because EITC requires earnings for its eligibility. There are several theoretical hypotheses about the effects of technological progress affecting labor demand, but the routine-biased technological change model of Acemoglu and Autor (2011), based on the task assignment framework, has been the most frequently referred one in recent years.[7] In their task framework, factors producing each task are perfect substitutes for each other, although tasks are imperfect substitutes in the production of final goods. But, each factor has a comparative advantage in different tasks with single crossing assumption,[8] which leads to the equilibrium that each factor is assigned to a different task--for example, low-skilled workers to the least complex tasks and high-skilled workers to the most complex tasks. When the digital technology has a comparative advantage over middle-skilled workers, it replaces middle-skilled workers by performing routine tasks and leads to the expansion of low and high skill tasks, which implies that some of the middle-skilled workers are allocated to tasks previously done by low- and high-skilled workers.[9] Autor and Dorn (2013) provide empirical evidence that commuting zones historically more specialized in routine-intensive occupations experienced more rapid growth in low-skill service occupations. Additionally, the technology displacing middle-skilled workers

[7] In the model, a task is defined as "*a unit of work activity producing (intermediate) output" and a skill is "a worker's endowment of capabilities for performing various tasks*." The explicit distinction between tasks and skills allows us to model the recent technology that can perform tasks previously done by workers with certain skills. While skills are applied to tasks in the task-based model, skills directly produce the final output in the Skill-Biased Technological Change (SBTC) hypothesis. The SBTC, where technology is described in a factor-augmenting form, shows that the technological change in favor of the skilled labor leads to an increase in the relative demand for skilled labor and the skill premium. However, it cannot explain the prevailing empirical phenomenon: employment polarization in the earnings distribution and wage stagnation for less-skilled labor.

[8] It means that high-skilled workers are better than middle-skilled workers, and middle-skilled workers are better than low-skilled workers as tasks are getting more complex.

[9] In Acemoglu and Autor (2011), tasks are defined in only one dimension, complexity, which seems to be defined from the human perspective. Unlike this, Feng and Graetz (2015) define tasks in two dimensions, training requirements and complexity from an engineering perspective. So, there are two types of tasks, training intensive and innate ability tasks, and each task is differentiated by complexity. In their task-based model, the firm will automate the task with more training requirements when two tasks have equal engineering complexity. By using comparative advantage properties, they show that low and high skill workers are relatively protected from technologies that facilitate automation, which tend to cause job polarization. Middle skill workers replaced by machines experience downward pressure on their wages. Also, they show that wage inequality goes up at the top, but falls at the bottom of the distribution.

decreases their wages relative to both low and high skill workers, which is related to wage polarization empirically observed.[10]

Previous studies examining the effect of digital technology like the expansion of computer usage combined with ICT (Information and Communication Technology) usually pay attention to its substitutability for routine tasks because they are easily programmable and codifiable (Auto *et al.*, 2003), and routine-task-intensive jobs are known to be generally laid in the middle of the wage distribution. But, the recent stream of technological advances like ML and AI enables us to automate tasks without fully specified instructions, so that its replaceability will not be limited to routine tasks.[11] For example, Google has been developing a driverless car, but driving a car is considered as a manual-task-intensive job that has a relatively lower wage than routine jobs. Also, paralegals and medical diagnosticians, who have a relatively high wage and used to be regarded as having strong complementarity with computers, are facing automation with developments of big data and ML.

The task-based framework modeling automation has been more generally elaborated by Acemoglu and Restrepo (2018a, 2018b, 2018c, 2019). In their model, automation has the displacement effect and the productivity effect: the former decreases labor demand in tasks previously performed by human labor, the latter raises labor demand in non-automated tasks through productivity increase. However, new technology, in general, has historically created new types of tasks (or jobs), which Acemoglu and Restrepo (2018a, 2018c, 2019) emphasize as a more powerful force to directly countervail the displacement effect by assuming that newly created tasks are more complex so that humans have a comparative advantage in these tasks.[12] Their recent empirical analysis decomposing the wage bill shows that the acceleration of displacement and the

[10] However, the impact on the relative wage of high-skilled workers compared to the low-skilled is uncertain. It increases if middle-skilled workers replaced by machines have a stronger comparative advantage in low skill tasks than high skill ones, it decreases otherwise.

[11] According to Taddy (2018), classical AI depends on "hand-specified logic rules" to solve problems, which requires that we must have full lists of all possible cases and know how to translate problems into structured data scheme. However, new AI driven by ML can input related information and learn how human acts to solve the problems.

[12] On the other hand, Bessen (2018) emphasizes the role of demand to explain how new labor-saving technology affects employment. There is the inverted U pattern of employment in the manufacturing industry during the 20th century, which can be explained by the change in price (or income) elasticity of demand. The author attributes the employment increase following the adoption of new technology to sufficient elasticity of demand. If the price elasticity of demand is greater than one or if the new technology applies to largely unmet needs, the price drop due to productivity improvement increases the demand enough to offset the negative impact on employment. Note that his argument is based on a premise that technology does not completely replace humans.

deceleration of creating new tasks contribute to the stagnation of labor demand for the last three decades, especially since 2000.

The empirical studies to date about the effects of automation, measured by industrial robots, on employment and wages have shown mixed results (Barbieri *et al*. 2019). Graetz and Michaels (2018) find the positive impact of industrial robots at the country-industry level on value-added and labor productivity, as measured by value-added per hour worked, as well as they show the negative effect on aggregated hours worked for low-skilled and middle-skilled workers. While the results in Acemoglu and Restrepo (2017) focusing on U.S. local labor markets present the negative effects of industrial robots on both employment and wages, Chiacchio *et al*. (2018) show the significant displacement effect on employment rate for six European countries, particularly of middle-educated workers and young cohort, but no impact on wage growth. Besides, Dauth *et al*. (2017) present that, in Germany, robots change the composition of aggregate employment: job losses in the manufacturing sector are fully offset by increased jobs in the service sector.

What seems obvious from the theoretical and empirical analyses is that the ongoing technological progress hurts some workers, primarily in low/middle-skilled or manufacturing jobs, even though its impact on aggregate employment is not clear. Hence, we need to contemplate in what directions our policy should respond to automation. When especially paying attention to people harmed by the technology,[13] McAfee and Brynjolfsson (2016) claim that policy in the age of automation still should encourage work by emphasizing the value of work beyond just making money, so they advocate the expansion of EITC or similar wage subsidies rather than giving cash assistance regardless of need. Korinek and Stiglitz (2017) also suggest wage subsidies and EITC for wage declines due to decreases in demand for specific types of labor that are replaced by machines. In respect to compensating for the wage declines, increasing the minimum wage can be helpful. For example, Downey (2016) focuses on the feature that automation occurs partially through deskilling--new technology simplifies tasks so that they can be performed by the technology and less-skilled workers--and presents that the minimum wage increase slows down the adoption of routine-replacing technologies. However, other empirical studies exploiting the minimum wage variation show that it negatively affects the employment of low-skilled (or low-wage) workers in routine-intensive jobs (Aaronson and Phelan, 2017; Lordan and Neumark, 2018).

[13] On the other hand, when considering that the current tax system is in favor of capital rather than labor, research about taxing robots is also ongoing. See Abbott and Bogenschneider (2018), Guerreiro *et al*. (2017), and Thuemmel (2018).

1.3 Empirics

My empirical analysis is conducted at the local labor market level, defined as Commuting Zones (CZs), which are groups of counties with strong commuting ties. CZs were developed by Tolbert and Killian (1987) and Tolbert and Sizer (1996) to provide geographical units representing local labor markets, based on residence-to-work commuting data from the 1980 and 1990 Census. A problem when applying the concept of CZs to public microdata is that the microdata report only areas that have at least 100,000 residents due to data confidentiality laws. Dorn (2009) suggests a way to identify CZs using Public Use Microdata Areas (PUMAs) which is the most disaggregated geographic unit provided in the Integrated Public Use Microdata Series (IPUMS) Census.[14] Here, I use Dorn's crosswalk files to map each PUMA to each 1990 CZs.[15] Though the 1990 Census data identifies 741 CZs, I focus on 722 CZs that cover the entire region of the U.S., except for Alaska and Hawaii.

In this section, I start with explaining the empirical specification and then describe how to construct main outcome variables about EITC usage, the explanatory variable, the exposure to industrial robots suggested by Acemoglu and Restrepo (2017), and datasets used in the empirical analysis.

1.3.1 Empirical model

My research aims to examine whether the EITC can still support low- and moderate-income families under automation. For this, I estimate the effect of automation on EITC usage by using local variation in exposure to industrial robot adoption which stems from spatial variation in the location of industries across commuting zones. The baseline specification is as follows:

$$\Delta y_{gcs,t_1-t_0} = \beta_1 EIR^{US}_{c,t_1-t_0} + \beta_2\left(EIR^{US}_{c,t_1-t_0} \cdot share\ in\ routine\ jobs_{c,t_0}\right) + \boldsymbol{X}_{c,t_0}\boldsymbol{\Gamma} + \delta_g + \theta_s + \varepsilon_{gcs} \quad (1.1)$$

[14] Since a PUMA code can represent multiple counties due to data confidentiality laws, it also can correspond to multiple CZs. Briefly, Dorn (2009) calculates a probability that a household who lives in PUMA code i also resides in CZ j, which is used for adjusting the original personal weight of IPUMS census data. See the appendix of Dorn (2009) for more details.

[15] The files are provided at http://www.ddorn.net/data.htm.

where subscripts represent tax filing cells g (filing status [single, married] × number of children [0, 1, 2+]),[16] commuting zone c, and state s. The outcome variable is defined as the change between years t_0 and t_1, and two types of variables primarily used are as follows: i) EITC recipiency rate ($y_{gcs,t}$) = $\left(\frac{EITC\ filers}{Total\ filers}\right)_t$, and ii) the average amount of credits ($y_{gcs,t}$) =$\left(\frac{Total\ amount\ of\ EITC}{Total\ filers}\right)_t$. EIR^{US}_{c,t_1-t_0} is the exposure to robots in commuting zone c in the U.S. between years t_0 and t_1, which is instrumented with EIR^{EU}_{c,t_1-t_0} based on the use of industrial robots in European countries. $\boldsymbol{X}_{c,t_0}$ includes demographic and economic characteristics in commuting zone c in the year t_0: the ratio of the working-age population, the ratio of female population, the ratio of the population with college and more education, the ratio of non-white population, the ratio of manufacturing employment, and the exposure to Chinese imports.

The employment share of routine jobs, frequently used as the measure for the Routine-Biased Technological Change (RBTC), is constructed based on the Routine Task Intensity in occupations. In the RBTC hypothesis, digital technologies have comparative advantages over medium-skilled workers in producing these tasks, which ends up displacing medium skilled labor. So, that measure can reflect the possibility of a broad range of automation, especially including other machines and computer software which do not fit into the definition of industrial robots, whereas the industrial robots can be considered as directly representing a specific type of automation.[17] Also, one might hypothesize that in commuting zones with initially high employment share in routine jobs, more workers are at risk of being replaced by industrial robots so that, among other impacts by automation, the displacement effect could be more pronounced in those regions. Therefore, the initial employment share of routine jobs is included in $\boldsymbol{X}_{c,t_0}$, and I add its interaction term with the exposure to robots to take into account the possibility that the effects of exposure to robots on the EITC usage may be heterogeneous across the levels of employment share in routine jobs in commuting zones.

[16] The classification of the tax filing group follows Bitler *et al*. (2017), which shows the possibility of heterogeneous effects of automation on EITC usage by tax filing status.

[17] The results from simple OLS regressions in Table B.1 indicate that two different measures affect the EITC usage in a different way.

The above empirical models are estimated, separately, for three groups: childless tax filers, single tax filers with children, and married tax filers with children. Here δ_g are group-specific intercepts to capture variations of credits by marital status and the number of children. So, for childless tax filers, δ_g equals to one if he/she is married, whereas for the other tax filers with children, δ_g is one if he/she has two or more children. θ_s represents the state fixed effects,[18] which also captures CZ-invariant state attributes over the time period including the effects of federal EITC policy changes. Lastly, both outcome variables have the total number of potential tax filers as the denominator, so I run the regressions by using the relevant number of potential tax filers in the baseline year as the weight, and the standard errors are clustered at the state level to allow for arbitrary spatial correlation at the state level.

As shown in Figure 1.2, the EITC amounts change with earnings in the opposite direction: your total credit will increase with an extra dollar of earnings as long as your earnings fall within the phase-in region, but if your earnings lie in the phase-out region, the total credit that you receive will decrease with an extra dollar of earnings. Of course, your credit will be constant at the maximum level with an extra dollar earned if your earnings are still in the flat region. This feature implies that, for example, the decrease in the EITC expenditure could happen when the EITC recipients are more distributed in the phase-in and the phase-out regions than in the plateau region. Hence, I examine the change in the share of EITC recipients in each eligible earnings region, rather than just focusing on the overall change in the recipiency rate.

More importantly, you can move out of the EITC eligibility because of either unemployment or earnings growth, which means that looking just at the overall recipiency rate and average benefit could miss the important point. To understand how automation is moving families around in the earnings distribution, I thus categorize the range of earnings into six groups: i) phase-in, ii) flat, iii) phase-out, iv) near phase-out, v) above the-near-phase-out, and vi) zero earnings. By calculating the ratio of each group of filers relative to the total filers, I use its changes over time as outcome variables. The earnings criteria for the first three groups follow the eligibility rule of EITC,[19] and I define the near-phase-out region from the end of the phase-out to 25 percent or 50

[18] A commuting zone can correspond to more than one state. If so, this commuting zone is assigned to a state which has the largest share of the population within the commuting zone. Visit http://www.ddorn.net/data.htm.

[19] To be eligible for the EITC, both your earnings and adjusted gross income (AGI) should belong to a specific range of income. Since the AGI is not available in the Census/ACS, I depend on only earnings.

percent above the end earnings of the phase-out region. The above the-near-phase-out region includes all observations with positive family earnings above the end of the near-phase-out.[20] In the zero earnings region, there are two types of families: potential non-filers whose earnings as well as the total income are zero, and filers who have positive total income with zero earnings.[21] In the real-world, both types may not file tax returns due to zero earnings. However, given that I construct tax-filing unit data from individual/household level data, they need to be considered as one type at the tax-filing level. Hence, the potential number of total filers, which is the denominator, includes them. Note that earnings criteria for the childless filers are not available in 1990 because they were not eligible for EITC at that time. Thus, for the childless filers in 1990, the ratio of filers in each group is set to zero.

In the analysis, the main coefficients of interest are β_1 and β_2. Since I have the interaction term, β_1 only represents the effect of the exposure to industrial robots on the change in EITC usage if the share of employment in routine jobs is zero. So, the marginal effect of the exposure to robots should be regarded as ($\beta_1 + \beta_2 \cdot share\ in\ routine\ jobs_{c,t_0}$), which depends on the level of employment share in routine jobs. If the sum of coefficients is positive at the mean value of the share in routine jobs, it implies that the higher exposure to industrial robots tends to increase the ratio of EITC recipients in the commuting zone which has the average level of employment share in routine-intensive jobs.

1.3.2 Measuring key variables

A. Outcome variables: EITC recipients and benefits

The main outcome variables are the changes in the EITC recipiency rate and average EITC benefit at the commuting zone level. To construct these variables, I first need to measure EITC eligibility and benefits for each tax filer unit and then to aggregate the data by the commuting zone. Here, I use the 1990 Census and the 2015 American Community Survey (ACS) datasets from IPUMS.[22] Since the Census and the ACS data do not contain information on the EITC usage, I

[20] The detailed criteria are given in Table B.2 in the Appendix.

[21] The latter means that they have other non-labor income such as capital gains.

[22] Note that the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (ACS) from IPUMS has not only more detailed information on income but also the values of the EITC, though tax-related variables come from the Census Bureau's tax model, not from the direct questioning. However, the ASEC (or CPS) does not contain the PUMAs, and the smallest geographical unit is county so that I am not able to identify CZs of more than half of surveyed households. Thus, I use the Census

obtain it through TAXSIM which is a publicly available microsimulation program provided by NBER.[23] But, the problem is that the IPUMS Census and ACS collect information in the household and the individual level which are not exactly matched with tax filing units. Thus, I need to convert the sample unit of those data into the family unit before I utilize TAXSIM. Hence, I briefly explain how I define the tax unit, and further details are given in the Appendix A.

I start with determining whether an individual is a qualifying child based on the tax instruction of the Internal Revenue Service (IRS): a child is an individual who is under age 19, or an individual who is a full-time student under age 24, or an individual who is permanently and totally disabled.[24] I also assign qualifying relatives to individuals based on the IRS instruction, and then define a "family" based on group identifier variables and assign the head of each family by using relationship indicator variables. After assigning an individual as a spouse of the head according to his or her relationship to the head of the family, I finally define the "tax filing unit" by assuming that every head is a tax filer who claims his/her qualifying children and relatives as dependents. There are four types of filing statuses depending on the head's marital status, the presence of a spouse, and qualifying dependents: single, head of household, married filing jointly, and married filing separately.

Tax filers whose age is 15 or less are excluded in the analysis, but tax filers whose total family income is less than or equal to zero with zero earnings are included as non-filers. If the total (family) income of the tax unit is positive even though the sum of earnings of all family members is zero, this unit is considered as (potential) tax filers. The estimated number of tax filers through the Census and the ACS are about 98 percent of the official statistics of the IRS in terms of total filers, but the number of married filing jointly is overestimated (See Table A.2). Now, I use this tax filing unit data as input for TAXSIM, but some income input for TAXSIM is assumed to be zero because

---

and the ACS data to construct outcome variables at the CZ level. Besides, their sample sizes are much larger than the ASEC. The decennial Census samples for 1990 include 5 percent of the U.S. population, and the ACS samples include 1 percent of the U.S. population.

[23] TAXSIM is the NBER's FORTRAN program for calculating liabilities under US Federal and State income tax laws from individual data. For more information, see Feenberg and Coutts (1993) and visit http://users.nber.org/~taxsim/taxsim27/

[24] According to the IRS instruction, a person is permanently and totally disabled if he/she cannot engage in any gainful activity because of a physical and mental condition, and a doctor has determined that the condition has lasted or can be expected to continue for at least a year. Because any variables cannot directly specify the disabled, I use three variables to indirectly measure disability. So, individuals are assigned as the disabled if they have both independent living difficulty (DIFFMOB) and self-care difficulty (DIFFCARE) as well as they are not in the labor force.

income variables of the Census and the ACS datasets do not exactly correspond to them (See Table A.1). Also, TAXSIM does not allow negative values for wage variables, so the negative self-employment income in the Census and the ACS is set to zero by assuming it as zero net earnings. The results through TAXSIM are reported in Table A.3, which shows that the simulated total amount of EITC benefits is about 75~79% of the official statistics. The simulation result is similar to Meyer (2010) which compares the distributions of EITC from two datasets, the IRS and CPS ASEC, and he suggests possible reasons for the discrepancy: i) IRS payments to ineligible recipients, ii) too low sample weight for EITC recipients in the CPS, and iii) underreporting of earnings in the CPS.

After combining the Census/ACS with the simulated federal EITC data at the tax filer level, I define the EITC recipiency rate as the number of EITC filers (tax filers with positive federal EITC) per tax filer at the CZ level, and the average amount of EITC as the amount of federal EITC per tax filer (or per EITC filer) at the CZ level,[25] where I use the personal weight of the individual specified as the head of a family when I aggregate these variables by the commuting zone.

Technically, the EITC recipiency rate defined above is more like the EITC eligibility rate because TAXSIM calculates the credits only based on eligibility rules, which does not imply actual take-up of the EITC. Recent studies estimating the EITC take-up rate suggest that about 80 percent of eligible taxpayers participated in the program (Jones, 2014; Plueger, 2009). However, this is not an issue because my research question focuses on the ability of EITC to support workers who could be hurt by automation. Even if some eligible workers could not utilize EITC, it does not change the conclusion regarding the overall ability of EITC to support workers

Table 1.1 shows descriptive statistics of variables used in the regression analyses at the commuting zone level. The main outcome variables are defined as the change between 1990 and 2015 so that they can be positive or negative. Briefly looking at outcome variables, they all have positive values on average, which implies that more taxpayers became eligible for the EITC and received more credits in 2015 compared to 1990. However, based on the minimum values, some commuting zones have negative numbers, which means that taxpayers in these CZs were less eligible and paid fewer credits in 2015 even though the EITC has been more generous over time.

[25] As mentioned before, I do not consider the state EITCs since they have been built on federal EITC; nearly all of them provide just additional credits by following the same eligibility rules as federal EITC.

I should note that, unlike other types of tax-filers, childless taxpayers can not have negative values in their outcome variables because they were not eligible for the EITC in 1990.

	Mean	S.D.	Min	Max
Change in ratio of EITC filers to total filers from 1990 to 2015 (in p.p.)				
Childless filers	6.676	3.098	0.441	16.161
Single filers w/ children	1.370	10.895	-36.009	39.791
Married filers w/ children	5.028	8.953	-20.746	31.977
Change in EITC amount per tax filer from 1990 to 2015 (2015 US$)				
Childless filers	19.900	9.144	1.175	49.794
Single filers w/ children	1204.341	617.930	-100.879	3243.089
Married filers w/ children	576.723	393.469	-141.459	2004.210
Change in EITC amount per EITC filer from 1990 to 2015 (2015 US$)				
Childless filers	301.908	47.551	60.765	458.493
Single filers w/ children	1957.065	785.979	-14.932	4206.648
Married filers w/ children	1925.490	751.609	119.364	4232.837
Exposure to industrial robots				
US from 2004 to 2015	0.946	0.557	0.224	4.026
30th percentile of EU from 1993 to 2015	1.030	0.616	0.235	5.107
Mean of EU from 1993 to 2015	1.687	0.972	0.447	8.277
Other controls				
Ratio of working-age population in 1990	0.616	0.032	0.526	0.704
Ratio of female population in 1990	0.511	0.010	0.474	0.537
Ratio of population with college or more in 1990	0.286	0.062	0.137	0.468
Ratio of non-white population in 1990	0.130	0.119	0.007	0.628
Ratio of manufacturing employment in 1990	0.177	0.087	0.032	0.470
Exposure to China imports from 1991 to 2015	0.024	0.018	0.004	0.141
Share of employment in routine-intensive jobs in 1990	0.283	0.034	0.200	0.377

Note: All values are defined at the commuting zone level, so the number of observations is 722. The EITC-related variables are calculated from simulation results through TAXISM by using the 1990 Census and the 2015 ACS. See text for the detailed information on variable definitions and data sources.

Table 1.1 Summary Statistics

	All	Childless		Single w/ children		Married w/ children	
		Single	Married	One child	2+	One child	2+
p10	10,000	7,800	14,600	6,000	5,600	25,000	25,000
p25	21,000	16,000	35,000	14,000	13,000	46,000	45,800
p50	42,500	30,000	69,000	26,500	25,000	80,000	80,000
p75	81,500	50,000	114,000	45,000	42,000	124,000	129,000
p90	137,000	80,000	177,000	72,000	70,000	187,000	200,000
p95	189,000	103,000	245,000	95,000	92,000	250,000	289,000
p99	429,000	200,000	483,000	175,000	180,000	496,000	537,300
Mean	65,273	40,740	90,586	35,998	34,573	100,799	106,189

Notes: All values are calculated by the 2015 ACS. Earnings are the sum of wage and positive self-employment income at the tax filing unit. Tax-filers who have zero earnings are excluded, and I use the personal weight of each tax-filer.

Table 1.2 Earnings percentiles by tax filing group in 2015

B. Explanatory variable: Exposure to industrial robot adoption

Acemoglu and Restrepo (2017) build a theoretical model in which robots can technologically perform a range of tasks $[0, M_i]$ in industry i. They show that the total equilibrium impact of robots is the sum of displacement and productivity effect, where each effect can be expressed as a function of technological changes in robot adoption. More specifically, they derive the following expression by assuming that the amount of automatable tasks, M_i, is close to zero, $M_i \approx 0$:

$$\sum_{i \in I} l_{ci} \frac{dM_i}{(1-M_i)} \approx \sum_{i \in I} l_{ci} \frac{dR_i}{L_i} = exposure\ to\ robots_c \quad (1.2)$$

where l_{ci} is the baseline employment share in industry i in commuting zone c, L_i is baseline employment in industry i, and dR_i is changes in robot usage in industry i.

Regarding this primary explanatory variable, I utilize the industrial robot data from the International Federation of Robotics (IFR). It provides consolidated worldwide robot statistics by collecting data from robot suppliers as well as several national robot associations, and categorizes robots into two types: industrial and service robots.[26] The IFR defines the industrial robot as "*an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or*

[26] A service robot is defined by the IFR as "*a robot that performs useful tasks for humans or equipment excluding industrial automation application*," and the classification of robots is based on their intended application.

more zexs, which can be either fixed in place or mobile for use in industrial automation applications." While the definition of the industrial robot excludes other types of technologies having the potential to complement or substitute for human labor--for example, self-checkout machines and especially software like AI--it provides an internationally comparable measurement of one of the latest technologies (Acemoglu and Restrepo, 2017).

The IFR provides two kinds of information on industrial robots: the annual sales and the operational stock of industrial robots. The IFR estimates the latter by assuming that each robot serves for 12 years on average and it is immediately removed after 12 years. I use the latter dataset for the period 1993-2015 at the country-industry level because it seems a more suitable way to measure how much we are exposed to robots, considering their life span. According to IFR (2017), industrial robot sales have been growing by about 12 percent per year since 2011 and reached 294,312 units in 2016. About 74 percent of the sales were delivered to five major countries, China, South Korea, Japan, the United States, and Germany, and about 61 percent of the sales were shipped in two industries, automotive and electrical/electronics industry. The operational stock of industrial robots was 1,828,000 units in 2016.

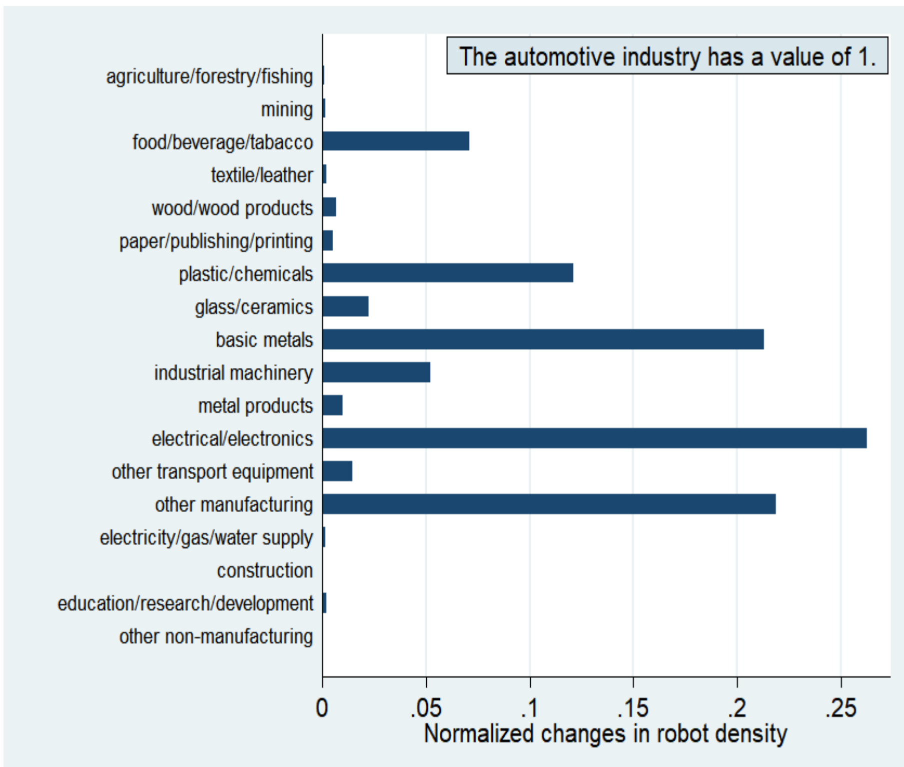
Following Acemoglu and Restrepo (2017), I construct the exposure to industrial robots (EIR_c) in the U.S. for each commuting zone c as follows:

$$EIR_c^{US} = \sum_{i \in I} l_{ci,1990} \left(\frac{IR_{i,2015}^{US}}{L_{i,1990}^{US}} - \frac{IR_{i,2004}^{US}}{L_{i,1990}^{US}} \right) \tag{1.3}$$

where $l_{ci,1990}$ is the 1990 employment share of industry i in commuting zone c, which is calculated from the 1990 Census, $IR_{i,t}$ is the operational stock of industrial robots in industry i and year t from the IFR data, and $L_{i,1990}^{US}$ is the number of workers in industry i in 1990 from the EU KLEMS data. Here, $\frac{IR_{i,t}}{L_{i,1990}^{US}}$ is the measure of industrial robot density, the operational stock of industrial robots per thousand workers, in industry i and year t. Note that I use 1990 as the baseline year because it is closer to the theoretical assumption. The industries include nineteen sectors, 6 non-manufacturing and 13 manufacturing sectors,[27] which is the classification of the IFR data.

[27] Six non-manufacturing industries are as follows: agriculture, hunting, forestry, and fishing; mining and quarrying; electricity, gas and water supply; construction; education, research and development; and other non-manufacturing industries. The manufacturing industry is categorized into 13 sectors: food products, beverages, and tobacco products; textiles, leather, and wearing apparel; wood and wood products; paper, paper products, and printing; plastic and chemical products; glass, ceramics, stone, and

They contain only private sectors, so the following industries are excluded when calculating the employment shares at the commuting zone level: public administration and defense, private households, and extra-territorial organizations and bodies. Figure 1.4 shows changes in robot density in the U.S. between 2004 and 2015, which is normalized by the change of the automotive industry because it has the largest increase over the period. Although the changes in robot density of other industries except the automotive industry are relatively small, comparatively large increases among them are found in the electrical/electronics, the basic metal, and the other manufacturing industry.



Notes: The industrial robot data comes from the IFR and the information on the number of workers in 1990 is from the EU KLEMS data. The robot density is defined as the operational stock of industrial robots per thousand workers. Changes in robot density by industry are normalized by the change of automotive industry which has the largest change over time. Thus, the normalized value of automotive industry is 1.

Figure 1.4 The change in robot density by industry in the U.S., 2004-2015

mineral products; basic metals; metal products; industrial machinery; electrical/electronics; automotive; other transport equipment; all other manufacturing sectors.

Since the EITC usage depends on employment, any shocks influencing the labor demand in commuting zone c may affect the decision of firms in that area to use robots in their production process. To deal with the possibility of endogeneity problems, I use robot usage in other nine European countries to construct an instrument, which is suggested in Acemoglu and Restrepo (2017) and defined as follows:

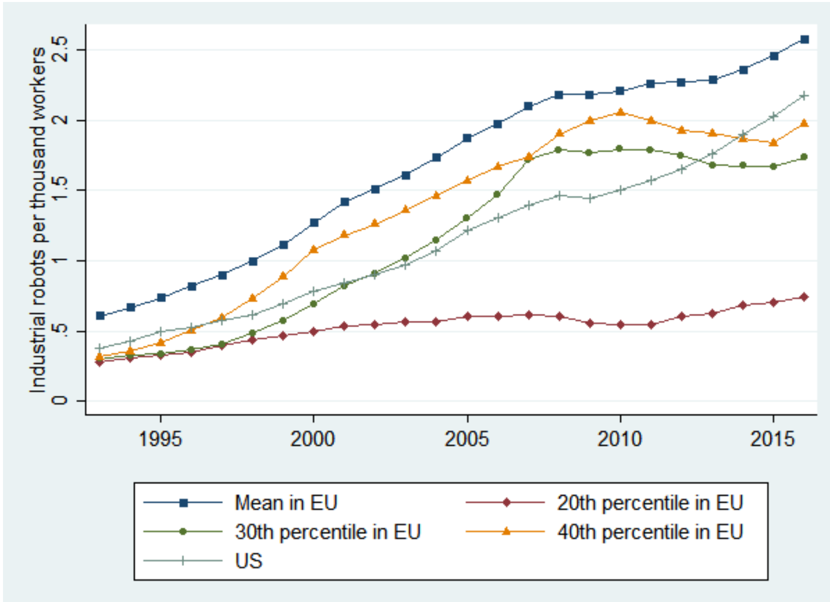
$$EIR_c^{EU} = \sum_{i \in I} l_{ci,1980} \left(p_{30}\left(\frac{IR_{i,2015}^{EU}}{L_{i,1990}^{EU}}\right) - p_{30}\left(\frac{IR_{i,1993}^{EU}}{L_{i,1990}^{EU}}\right) \right) \quad (1.4)$$

where $l_{ci,1980}$ is the 1980 employment share of industry i in commuting zone c in the U.S., which comes from the 1980 Census, the industrial robot density in industry i, $\frac{IR_{i,t}}{L_{i,1990}}$ is calculated for each European country based on the IFR and EU KLEMS data, and $p_{30}\left(\frac{IR_{i,t}}{L_{i,1990}}\right)$ denotes the 30th percentile of industrial robot density among nine European countries: Denmark, Finland, France, Germany, Italy, Norway, Spain, Sweden, and the United Kingdom. When constructing this instrument variable, I use the 1980 employment distribution of CZs, which help us mitigate bias due to contemporaneous changes in employment by expected robot adoption and focus more on the historical difference in specialized industries across CZs.[28] Since the industry-level robot data of the U.S. starts from 2004, I convert EIR_c^{US} to a 22-year equivalent change when it is instrumented with EIR_c^{EU}. For more detailed information on the construction of exposure to industrial robots, refer to the Appendix and Acemoglu and Restrepo (2017).

Figure 1.5 plots the robot adoption per thousand workers in the US and nine EU countries, which is measured in terms of the total amount of robots and employment.[29] Figure 1.6 gives the geographical distribution of exposure to robots in the U.S. from 2004 to 2015, which shows that traditional manufacturing (northeast) and high-technology (coastal) areas are relatively more exposed by industrial robots.

[28] Acemoglu and Restrepo (2017) primarily use the 1970 employment share, $l_{ci,1970}$, but they show that the results are similar when they use the distribution of employment across industries in 1980.

[29] The total amount of industrial robots is from the IFR, which equals the sum of industrial robots in each industry. In case of employment from EU KLEMS, the total is the sum of workers in all industries except three industries: Public admin and defense, compulsory social security; Private household with employed persons; Extra-territorial organizations and bodies.



Notes: The industrial robot data comes from the IFR and the information on the number of workers in 1990 is from the EU KLEMS data. The robot density is defined as the operational stock of industrial robots per thousand workers.

Figure 1.5 The robot density by time and country

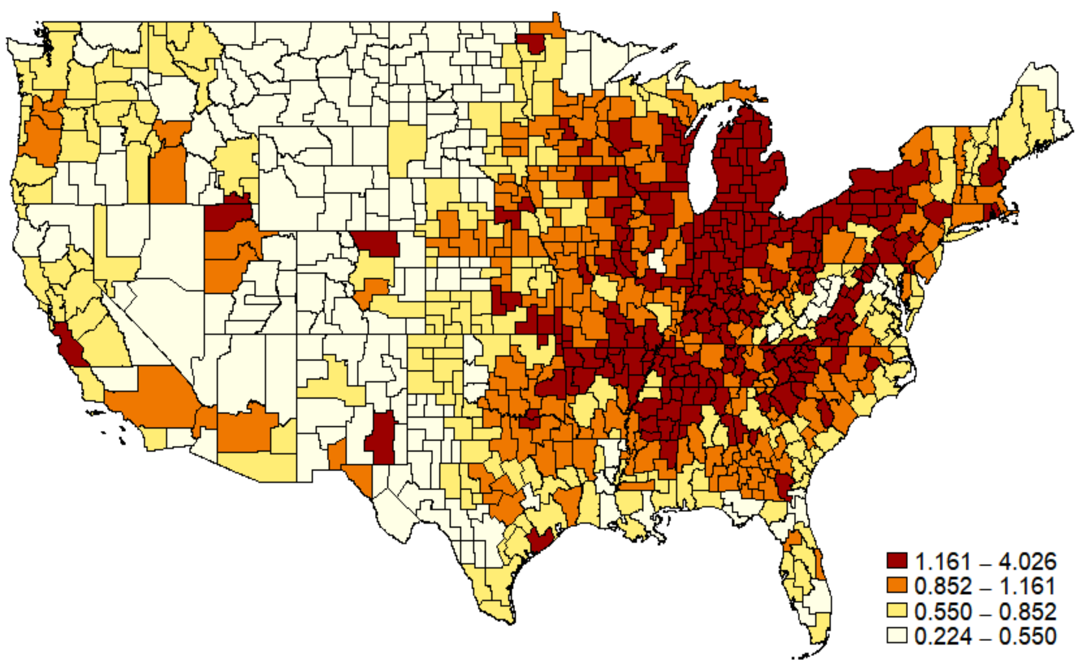


Figure 1.6 Exposure to industrial robots in the U.S. by commuting zone, 2004-2015

C. Other variables

The Routine Task Intensity (RTI) measure has been frequently exploited in literature, especially explaining job polarization, to represent the routine-based technology. So, as mentioned before, I use the share of employment in routine-intensive jobs as one of the control variables, as well as utilize its interaction with the exposure to industrial robots. The data comes from Autor and Dorn (2013), where defines it as follows:

$$Employment\ share\ in\ routine\ jobs_{c,1990} = \sum_{j \in J} \frac{L_{cj,1990} \cdot 1[RTI_j > RTI^{66}]}{L_{c,1990}} \quad (1.5)$$

where L_{cj} denotes the employment in occupation j in commuting zone c, and RTI_j measures relative routine task inputs of occupation *j* based on the U.S. Dictionary of Occupational Titles 1977.[30]

On the other hand, China's export surge has been regarded as one of the important factors affecting local labor markets in the U.S. since the 1990s. To control effects induced by changes in trade patterns, I construct an additional variable, the exposure to Chinese imports, by following previous studies' construction of the variable (Autor *et al.*, 2013; Acemoglu *et al.*, 2016). Specifically, the exposure to Chinese imports in commuting zone *c* from 1991 to 2015 is defined as follows

$$Exposure\ to\ China\ imports_c = \sum_{i \in I} l_{ci,1990} \left\{ \frac{\Delta M_i^{UC}}{Y_{i,1991} - (M_{i,1991} - X_{i,1991})} \right\} \quad (1.6)$$

where l_{ci} the employment share in industry i in commuting zone c and $\Delta M_i^{UC} \equiv M_{i,2015}^{UC} - M_{i,1991}^{UC}$ is the change in imports from China into the U.S. in industry i which is normalized by domestic absorption, approximated by the sum of domestic shipments (Y_i) and net imports ($M_i - X_i$). Trade data is available in SIC 4-digit level from World Integrated Trade Solution (WITS), and the domestic shipments data in SIC 4-digit level come from the NBER-CES Manufacturing Industry Database. Since I use the 1990 Census data to calculate the employment shares, l_{ci}, the trade data with SIC 4-digit codes are converted to the data with the Census industry codes of 1990 (Lake and

[30] $RTI_j = \ln(T_{j,1980}^R) - \ln(T_{j,1980}^M) - \ln(T_{j,1980}^A)$, where T_j^R, T_j^M, and T_j^A denote the routine, manual, and abstract task inputs in occupation j. See Autor and Dron (2013) for detailed information.

Millimet, 2018).[31] While the domestic shipments data is available only for the manufacturing industry, the trade data includes some of the agriculture, forestry, fisheries, and mining industries. For these industries, I use the mean value of domestic shipments across all 4-digit SIC manufacturing industries. All dollar amounts are adjusted to 2015 US dollars using the Personal Consumption Expenditures (PCE) price index.

Similar to the exposure to robots, there is the potential endogeneity of US exposure to China imports in the sense that any shocks affecting employment in commuting zone c also influence industry import demand. Following Acemoglu *et al*. (2016) and Lake and Millimet (2018), I construct an instrumental variable using Chinese exports to eight high income countries except for the U.S.,[32] which is based on the fact that the growth of Chinese imports since the 1990s has been mostly driven by supply-side shocks such as the increased competitiveness of Chinese manufacturing industry, the lowered trade barriers of China, and China's WTO entry (Autor at el., 2013).

$$IV\ for\ Exposure\ to\ China\ imports_c = \sum_{i \in I} l_{ci,1980} \left\{ \frac{\Delta M_i^{OC}}{Y_{i,1989} - (M_{i,1989} - X_{i,1989})} \right\} \tag{1.7}$$

where l_{ci} is the employment share in industry i in commuting zone c, and $\Delta M_i^{OC} \equiv M_{i,2015}^{OC} - M_{i,1991}^{OC}$ is the change in Chinese exports to eight non-US countries in industry i which is normalized by the U.S. domestic absorption in 1989.[33]

1.4 Results

1.4.1 Baseline results

The baseline results are presented in Table 1.3 and Table 1.4. Table 1.3 shows results from OLS estimation of the reduced form, replacing the exposure to industrial robots in the U.S. with the

[31] SIC code 3341 is matched two Census industry codes, 272 and 280. Hence, values of variables with SIC code 3341 are assigned to each Census code with the probability 0.5.

[32] The eight countries are Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland.

[33] In COMTRADE data, US exports and imports are available after 1991. So, the US exports and imports in 1989 come from https://dataweb.usitc.gov.

30th percentile of exposure to robots in nine European countries.[34] And then, Table 1.4 presents IV estimates by Two-Stage Least Squares (2SLS), where the exposures to robots and Chinese imports in the U.S. are instrumented with the relevant variables previously defined in Section 1.3.[35] For the interaction term, the product of the 30th percentile of exposure to robots in European countries and the employment share in routine jobs in 1990 is used as an instrumental variable. In both tables, the upper table (a) shows results without the interaction term between the exposure to robots and the employment share in routine jobs in 1990, and the lower table (b) presents estimates with the interaction term. The first column shows estimates for the pooled sample so that it has 722 observations. The remaining columns display estimates for three different groups, where each column has 1,444 observations: 722 CZs×2 types of marital status [single or married] for childless filers and 722 CZs×2 children groups [1 or 2+] for single/married filers with children. Panel A shows estimates for changes in the ratio of EITC filers to total (potential) tax-filer and Panel B presents results for changes in the amount of EITC per (potential) tax-filer in 2015 U.S. dollar.

Without the interaction term, both OLS and IV estimates show that the exposure to robots does not have statistically significant effects on EITC usage, whereas more taxpayers, especially tax-filers with children, are eligible for EITC in commuting zones with initially high employment share in routine-intensive jobs. When considering that, for taxpayers with children, earnings criteria for the EITC are more generous so that they include moderate-income families, the positive relation between share in routine jobs and EITC usage seems to reflect previous studies about routine-biased technology: there is a decrease in middle-wage jobs and an increase in low-wage service jobs.

The bottom table (b) in Table 1.3 and Table 1.4 presents results with the interaction based on the estimation equation (1.1). Due to the interaction term, the estimates of exposure to robots are hard to interpret so that the marginal effects of exposure to robots on EITC usage are plotted against the initial employment share of routine jobs in Figure 1.7. The left column shows its effect on the EITC recipiency rate by the tax filing group, whereas the right side presents its effect on EITC benefits per taxpayer by the tax filing group.

[34] Since the exposure to China imports may have the potential endogeneity problems, the instrument variable, measured by Chinese exports to non-US high-income countries, is used for OLS estimation.

[35] I also construct another instrument variable for the exposure to robots in the U.S. by using the mean of industrial robot usage among nine European countries, which results are reported in Table B.3.

First, figures suggest that the marginal effects of robots could vary with the employment share of routine jobs in the baseline year, even though it hardly changes the sign of the marginal effect of robots, except for the single filers with children, given that the share of employment in routine-intensive jobs in 1990 has values in the range of 0.2 to 0.377. As shown in Table 1.4-(b), most estimates on the interaction term have an opposite sign to coefficients on the exposure to robots so that it seems to weaken the marginal effect of exposure to robots as commuting zones have a higher share in routine jobs.

Second, for single filers with children, there is evidence suggestive of a positive effect on recipiency rate as the share of routine jobs in commuting zones becomes higher and no effect in commuting zones with low routine share. It may reflect the possibility that more single filers with children in commuting zones with high routine share could experience the loss of earnings due to automation. Considering that about 75% of single filers with children have earnings lower than the top of the eligible range for EITC as shown in Table 1.2, which is also lower than the overall mean earnings, they could be relatively more exposed to the displacement effect, as suggested in Acemoglu and Restrepo (2017).

Third, for the married filers with children, there is evidence suggestive of a negative effect on the recipiency rate in commuting zones with low routine share and no effect in commuting zones with high routine share. It is consistent with the hypothesis of net increases in earnings due to stronger productivity effects in commuting zones with less reliance on routine jobs. It is also consistent with the expectation that the improvement in the productivity and earnings of some workers could be completely offset by reductions in employment and/or hours for other workers performing routine tasks in commuting zones with greater reliance on routine jobs. However, note that the results could, in part, come from the ability of the married filers to cushion against income loss when considering their relatively higher potential earnings and two potential earners.

(a) Without interaction term

	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in EU	0.198	-0.190*	1.099	-0.503
1993-2015 (p30)	(0.215)	(0.112)	(0.795)	(0.671)
share in routine jobs, 1990	39.110***	-2.521	42.420*	80.819***
	(4.706)	(4.903)	(25.045)	(12.600)
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in EU	-0.223	-0.314	-19.810	-51.039
1993-2015 (p30)	(7.643)	(0.324)	(27.367)	(34.921)
share in routine jobs, 1990	830.134***	-14.322	1102.804*	1820.824***
	(203.745)	(13.913)	(573.729)	(530.690)
Observations	722	1444	1444	1444

(b) With interaction

	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in EU	0.567	-0.847	-2.692	-6.706
1993-2015 (p30)	(1.603)	(0.668)	(5.482)	(5.380)
Exposure to robots in EU	-1.080	1.922	11.097	18.218
× share in routine jobs, 1990	(4.715)	(1.973)	(15.984)	(15.597)
share in routine jobs, 1990	40.244***	-4.539	30.792	61.632***
	(5.864)	(4.717)	(35.711)	(16.566)
Joint significance (p-value)	0.623	0.122	0.338	0.381
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in EU	-93.079*	-3.867	-12.423	-471.453***
1993-2015 (p30)	(54.511)	(2.431)	(177.908)	(144.293)
Exposure to robots in EU	272.057*	10.403	-21.628	1234.723***
× share in routine jobs, 1990	(159.554)	(6.984)	(504.949)	(414.307)
share in routine jobs, 1990	544.402***	-25.245*	1125.468	520.421
	(193.372)	(14.814)	(685.844)	(554.841)
Joint significance (p-value)	0.242	0.238	0.762	0.005
Observations	722	1444	1444	1444

Note: The tables show OLS estimates of the impact of exposure to industrial robots on EITC usage. The dependent variable is the change in EITC filers to total filers ratio between 1990 and 2015 in Panel A, measured in percentage points, and the change in EITC amount per tax filer between 1990 and 2015 in Panel B. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. In both tables, automation is measured by the 30th percentile of exposure to robots in European countries. Table (a) does not include the interaction term between automation and share in routine jobs, but table (b) include it. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total filers of CZs in 1990. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table 1.3 The impact of exposure to industrial robots on EITC usage, 1990-2015 (OLS estimates)

(a) Without interaction term

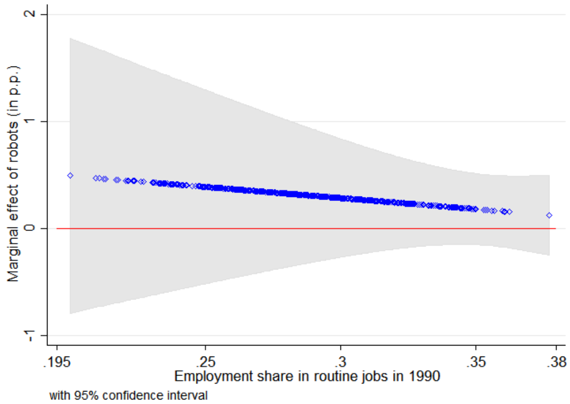
	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in US	0.186	-0.166*	0.861	-0.413
2004-2015	(0.176)	(0.097)	(0.678)	(0.570)
share in routine jobs, 1990	39.517***	-2.670	40.801*	81.173***
	(5.191)	(4.652)	(23.543)	(11.263)
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in US	2.184	-0.226	-14.571	-39.121
2004-2015	(7.114)	(0.287)	(24.854)	(31.730)
share in routine jobs, 1990	881.902***	-13.491	1151.540*	1916.332***
	(232.916)	(14.255)	(590.063)	(527.656)
Weak IV F-stat	36.471	36.806	40.472	39.414
Observations	722	1444	1444	1444

(b) With interaction

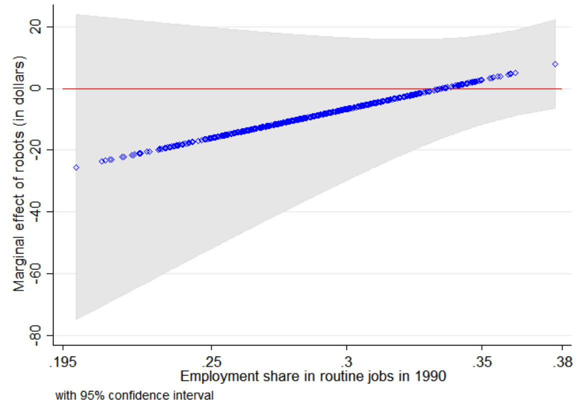
	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in US	0.905	-1.025	-1.355	-6.481
2004-2015	(1.456)	(0.687)	(4.909)	(4.760)
Exposure to robots in US	-2.070	2.472	6.381	17.507
× share in routine jobs, 1990	(4.065)	(1.883)	(13.569)	(13.252)
share in routine jobs, 1990	43.680***	-7.664	28.164	46.057*
	(9.051)	(5.532)	(41.544)	(25.292)
Joint significance (p-value)	0.534	0.153	0.348	0.377
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in US	-63.047	-3.470	-22.250	-451.113***
2004-2015	(54.327)	(2.437)	(180.948)	(166.382)
Exposure to robots in US	187.872	9.338	22.117	1188.740***
× share in routine jobs, 1990	(148.085)	(6.627)	(485.148)	(436.501)
share in routine jobs, 1990	504.016	-32.354	1107.735	-467.988
	(319.232)	(19.746)	(1041.825)	(831.998)
Joint significance (p-value)	0.338	0.362	0.820	0.024
Weak IV F-stat	20.129	20.553	21.576	21.531
Observations	722	1444	1444	1444

Note: The tables show IV estimates of the impact of exposure to industrial robots on EITC usage, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, the interaction term is also instrumented with the product of the 30th percentile of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variable is the change in EITC filers to total filers ratio between 1990 and 2015 in Panel A, measured in percentage points, and the change in EITC amount per tax filer between 1990 and 2015 in Panel B. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. Table (a) does not include the interaction term between automation and share in routine jobs, but table (b) include it. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total filers of CZs in 1990. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

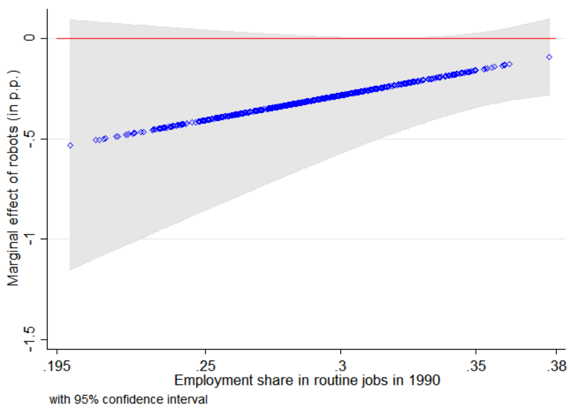
Table 1.4 The impact of exposure to industrial robots on EITC usage, 1990-2015 (IV estimates)



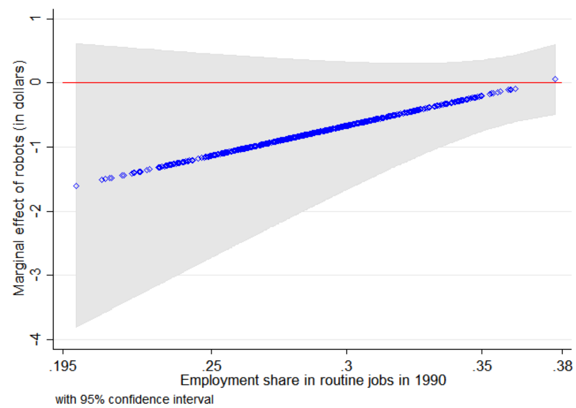
(a) All, recipiency rate



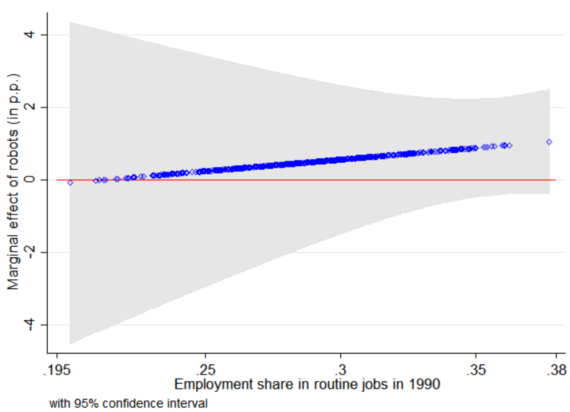
(b) All, credits per tax-filer



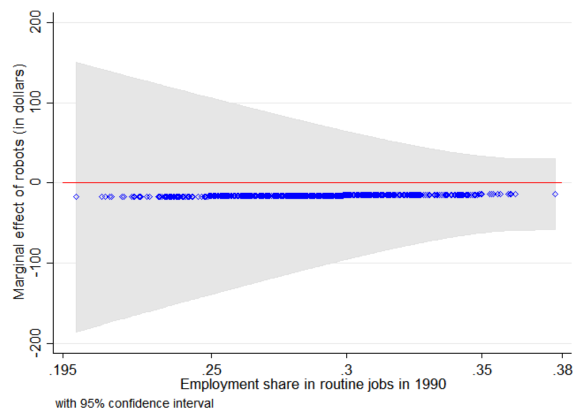
(c) Childless, recipiency rate



(d) Childless, credits per tax-filer

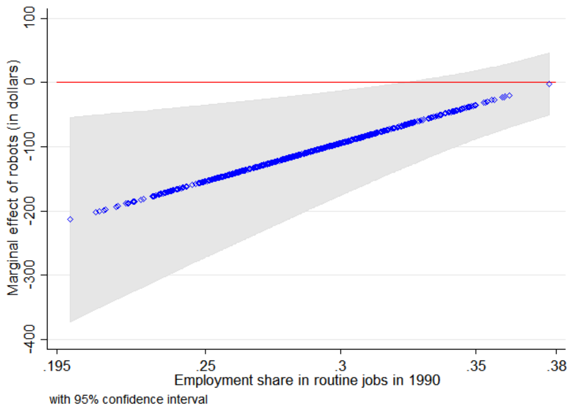


(e) Single w/ children, recipiency rate

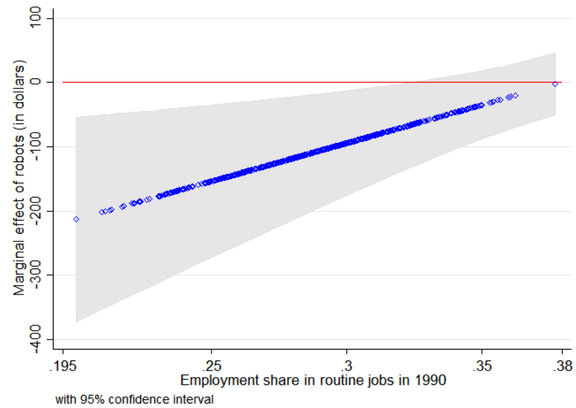


(f) Single w/ children, credits per tax-filer

(Continued)



(g) Married w/ children, recipiency rate



(h) Married w/ children, credits per tax-filer

Notes: The marginal effects and confidence intervals for each sample are calculated based on the IV estimates in Table 1.4-(b).

Figure 1.7 Marginal effects of industrial robots on EITC usage, 1990-2015

Lastly, the size of estimates on the EITC benefits per tax-filer are considerable only for the married filers with children, which seems obvious as credits are more generous with children and marriage. For the married filers with children, the effects of automation on the average EITC benefits show a pattern similar to its impacts on recipiency rate: the automation decreases the average benefits in commuting zones with low routine share, which seems to be offset as the routine share in commuting zones increases. Given that commuting zones have the average value in the share of routine jobs, Panel B of Table 1.4-(b) suggests that the EITC benefits for the married filers with children in a commuting zone where the exposure to robots equals to 1 are $115 less than ones in a commuting zone not exposed to robots.

Note that, as you can see in Table 1.4, the estimates are imprecise overall so that the above results should be viewed with caution. In addition, the lower EITC expenditure per potential tax-filer may happen since taxpayers eligible for EITC in commuting zones with more exposed to robots are distributed more in phase-in or phase-out regions. More importantly, negative effects on both the recipiency rate and the average benefit reflect results that mix reductions in EITC claims and benefits due to unemployment with reductions caused by earnings growth that push families above the eligibility cut-off. Thus, it has a shortcoming just looking at the overall recipiency rate and the average benefit.

1.4.2 Additional results

To supplement the baseline results, I turn to the extended model in this section to examine how automation is moving families around in the earnings distribution. For this, I divide the range of earnings into six regions: i) phase-in, ii) flat, iii) phase-out, iv) near phase-out, v) above the-near-phase-out, and vi) zero earnings. So, outcome variables are defined as changes in the share of filers in each region between 1990 and 2015. The first three regions follow the earnings criteria of EITC, and the near-phase-out region begins from the end of the phase-out region to 25 percent above the end earnings of the phase-out region.[36] The above the-near-phase-out region includes all families with positive earnings above the end of the near-phase-out. Note that the childless filers are not eligible for EITC in 1990, so there are no relevant earnings criteria for them. Hence, I assume that their shares in each region of earnings in 1990 are zero.

By using the same estimation equation (1.1), the IV estimates of the extended model are presented in Table 1.5. Due to the interaction term, interpretations of coefficients on the exposure to robots are not straightforward. Thus, the marginal effects of exposure to robots, which changes with the employment share of routine jobs in commuting zones, are plotted in Figure 1.8 and Figure 1.9 for the single filers with children and the married filers with children, respectively.

First, for single filers with children, Figure 1.8 shows that automation in commuting zones with high routine share seems to increase the share of those filers in the phase-in region and to decrease the share of them above the near-phase-out region. These changes seem to explain the baseline result for commuting zones with high routine share--suggestion a positive effect on the EITC recipiency rate. For commuting zones with low routine share, the baseline model--no change in the recipiency rate--seems to reflect the mixed results of an increase in flat region and a decrease in phase-out region. In commuting zones with high routine share, the relative increase of the share of single filers with children in phase-in within the eligible range for EITC might suggest the need for more income support such as a higher phase-in rate, but it requires further research about changes in the distribution of earnings of single parents under automation at the disaggregated level. Meanwhile, we can see an increase in the share of single filers with children in the near-phase-out region, which might suggest that EITC's future expansions in eligibility could support

[36] I also use a different criterion to define the near-phase-out region that begins from the end of the phase-out region to 50 percent above the earnings at the end of the phase-out region. The results are shown in Table B.5.

more single parents with automation, given that the earnings of the near-phase-out region are still lower than median household income.

Second, the results for the married filers with children are still imprecise as shown in Figure 1.9, but the estimates could suggest some movement of families through the earnings distribution, which impacts their EITC eligibility. In commuting zones with low routine share, where EITC recipiency rates and benefits appear to fall in the baseline model, as shown in Figure 1.7-(g) and Figure 1.7-(h), the largest effects of automation seem to be found in a decline of the share of families in the flat region and in a rise of the share of them above the- near-phase-out region. It implies that the decrease in overall recipiency rate of married filers with children arises from the increase in the share of families who are not eligible for EITC due to higher earnings. For commuting zones with high routine share, there is little change in the share of tax-filers across different regions of their earnings.

As a brief examination of the results in Table 1.5, the coefficients in the first three panels are about the effect of automation on the change in the share of tax-filers potentially eligible for EITC by different regions of earnings,[37] while the remaining three panels show the effect on the ratio of tax-filers by earnings regions out of eligibility for the EITC. Surprisingly, for the pooled sample and the separate filing groups, most estimates for the main effect of robots present a positive relationship between the exposure to robots and the change in the ratio of filers with earnings above the-near-phase-out. But, the coefficients on the interaction term have the opposite sign, so the main effect reduces as the employment share in routine jobs in the baseline year changes. Note that the overall effect on filers in the region of earnings above the-near-phase-out is still positive at the average level of share of routine jobs, while it turns into a negative value as the employment share of routine jobs approaches the maximum. Since the earnings of filers assigned to this region begin from a relatively lower level of median family income (up to any positive earnings above),[38] the negative sign on the interaction term may imply some extent of the effect of routine-biased technology on job polarization in the U.S., hollowing out the middle-wage jobs. The last panel shows that a commuting zone with exposure to robots is likely to have fewer filers with zero

[37] The reason I refer to tax-filers as "potentially eligible for EITC" is that they are classified by only earnings regardless of simulation results from TAXSIM.

[38] According to the 2015 ACS, the median income of the family and the nonfamily households is $68,260 and $33,617, respectively. In the 1990 Census, the median family income (in 1989) is reported as $35,225.

earnings compared to a commuting zone not exposed to robots; however most of the estimates are imprecise.

To check whether the increase in the share of tax-filers above the-near-phase-out region in Table 1.5 is affected by migrations across commuting zones, I also construct outcome variables in a different way: the numerator is a change in the number of tax-filers in each region of earnings between 1990 and 2015, and the denominator is fixed at the total number of tax-filers in 1990. The estimates are presented in Table B.6 in the Appendix, and the marginal effects of automation for different groups of tax-filers by the routine share of commuting zones are displayed in Figure B.2 and Figure B.3. The marginal effect of automation on the change in tax-filers above the-near-phase-out region has a similar pattern, but the size of coefficients becomes larger. Although using the fixed denominator is not a perfect way to control migrations, it suggests that the previous result is not solely because families with higher earnings move into commuting zones more exposed to robots.

Lastly, to examine the ability of EITC supporting low-income families in the age of automation more directly, I explore whether EITC can alleviate poverty under the exposure to industrial robots at the commuting zone level. So, following Bitler *et al*. (2017), I calculate how many taxpayers have income below 50 percent, 100 percent, 150 percent, and 200 percent of official poverty thresholds in 1990 and 2015 at the commuting zone level,[39] and then define outcome variables as changes in the ratio of tax-filers under certain levels of income relative to total filers. Here, I use two types of income: total pre-tax income, which is the sum of family members' income of tax-filer, and total income added by the EITC. By comparing the size of estimates between two different measures, it can tell us whether the added EITC helps people out of poverty. The estimates are provided in Table 1.6, where we can see that they are statistically significant only for single filers with children, especially when focusing on the change in the ratio of filers under 50 percent and 200 percent of official poverty thresholds. For single filers with children, the ratio of tax-filers under 50 (and 200) percent of official poverty thresholds seems to decrease in commuting zones more exposed to robots, and its extents vary depending on the share of routine jobs in commuting zones. Its negative impact on the ratio of families under certain poverty thresholds is

[39] Since income variables in the 1990 Census are measured based on the previous survey year, I use the 1989 thresholds. For the official poverty thresholds, visit https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html

large in commuting zones with low routine share, and this effect seems to be diminished in commuting zones with high routine share by the displacement effect. Since the third column in Table 1.6 presents that automation reduces families under certain levels of poverty, we can say that EITC helps families get out of poverty when the absolute values of coefficients in the seventh column are larger. Figure 1.10 shows this relationship for single filers with children.

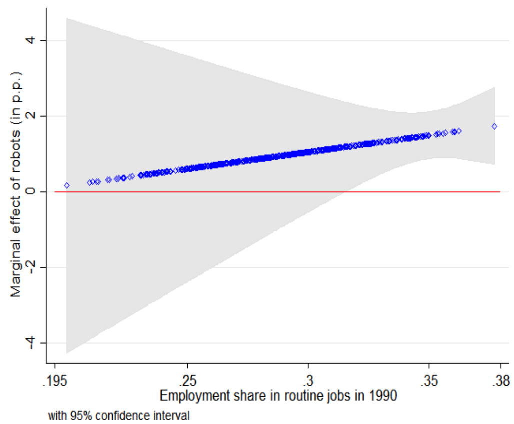
	All	Childless	Single, children	Married, children
Panel A. Change in Phase-in region				
Exposure to robots in US 2004-2015	-0.305	-1.444*	-1.608	0.890
	(0.766)	(0.777)	(5.222)	(1.207)
Exposure to robots in US × share in routine jobs, 1990	0.882	3.530	8.843	-2.798
	(2.174)	(2.175)	(14.892)	(3.365)
share in routine jobs, 1990	1.277	-20.182***	31.773	19.228**
	(4.965)	(4.677)	(32.787)	(9.074)
Joint significance (p-value)	0.920	0.015	0.000	0.579
Panel B. Change in Flat region				
Exposure to robots in US 2004-2015	-0.092	-0.701	2.098	-4.079**
	(0.599)	(0.457)	(2.463)	(1.628)
Exposure to robots in US × share in routine jobs, 1990	0.225	1.962	-6.602	11.393***
	(1.586)	(1.253)	(6.658)	(4.403)
share in routine jobs, 1990	5.698	-10.664***	22.635	7.903
	(4.182)	(3.398)	(15.602)	(9.704)
Joint significance (p-value)	0.984	0.282	0.333	0.033
Panel C. Change in Phase-out region				
Exposure to robots in US 2004-2015	0.890	0.266	-2.676	-3.307
	(1.159)	(0.933)	(4.479)	(3.113)
Exposure to robots in US × share in routine jobs, 1990	-1.955	-0.451	7.045	8.949
	(3.336)	(2.747)	(12.504)	(8.761)
share in routine jobs, 1990	19.279***	-2.267	-32.094	19.494
	(6.737)	(6.788)	(28.087)	(18.145)
Joint significance (p-value)	0.156	0.566	0.760	0.516
Panel D. Change in Near Phase-out region: from the end of phase-out to 25% above it				
Exposure to robots in US 2004-2015	0.473	-0.695	1.962	0.183
	(0.554)	(0.660)	(1.787)	(1.721)
Exposure to robots in US × share in routine jobs, 1990	-1.030	1.806	-3.926	-0.311
	(1.519)	(1.821)	(4.903)	(4.378)
share in routine jobs, 1990	7.275*	-4.910	-2.879	10.130
	(4.312)	(4.789)	(12.999)	(10.820)
Joint significance (p-value)	0.400	0.461	0.159	0.949

(continued)

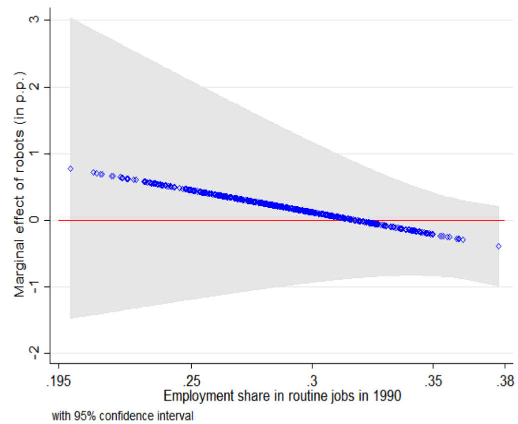
	All	Childless	Single, children	Married, children
Panel E. Change in Above Near Phase-out region				
Exposure to robots in US 2004-2015	7.929**	9.733**	6.396**	7.350
	(3.563)	(3.906)	(2.920)	(5.270)
Exposure to robots in US × share in routine jobs, 1990	-24.126**	-28.300***	-20.885***	-20.649
	(10.556)	(10.869)	(7.811)	(14.796)
share in routine jobs, 1990	76.374***	127.066***	17.400	-53.960**
	(26.093)	(26.807)	(17.854)	(27.113)
Joint significance (p-value)	0.065	0.011	0.000	0.377
Panel F. Change in Zero earnings				
Exposure to robots in US 2004-2015	-5.402**	-7.159**	-6.172	-1.037
	(2.630)	(3.461)	(4.011)	(1.478)
Exposure to robots in US × share in routine jobs, 1990	15.441**	21.453**	15.524	3.416
	(7.156)	(9.606)	(11.148)	(4.204)
share in routine jobs, 1990	-65.348***	-89.043***	-36.834	-2.794
	(17.966)	(24.551)	(31.195)	(8.066)
Joint significance (p-value)	0.067	0.023	0.188	0.253
Weak IV F-stat	20.129	20.553	21.576	21.531
Observations	722	1444	1444	1444

Note: The table shows IV estimates of the impact of exposure to industrial robots on the share of tax-filers in each region of earnings, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, the interaction term is also instrumented with the product of the 30th percentile of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variables are the change in the ratio of tax-filers in each region relative to total taxpayers from 1990 to 2015. The first three regions follow the earnings criteria for EITC. The near-phase-out region begins from the end of the phase-out region to 25 percent above the end earnings of the phase-out region. The above the-near-phase-out region includes all families with positive earnings above the end of the near-phase-out. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total tax-filers of CZs in 1990. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

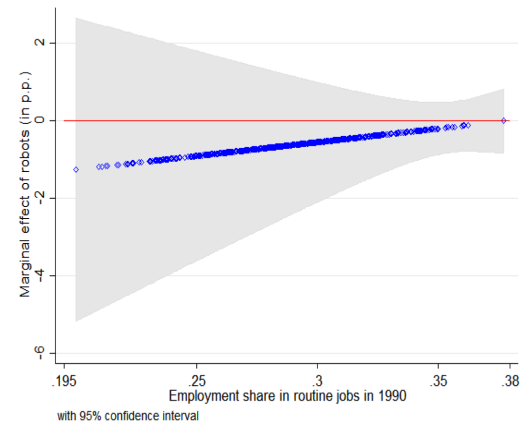
Table 1.5 The impact of exposure to industrial robots on the change in the ratio of tax-filers by the region of earnings, 1990-2015 (IV estimates)



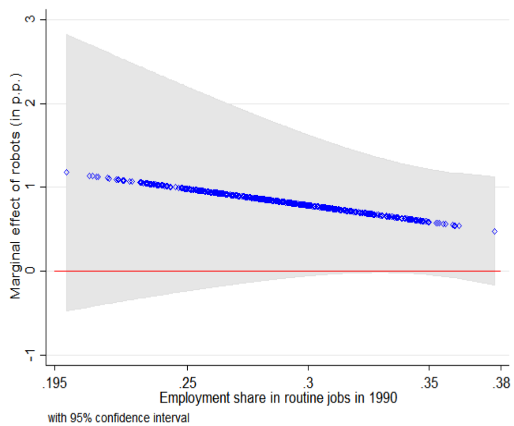
(a) Phase-in region



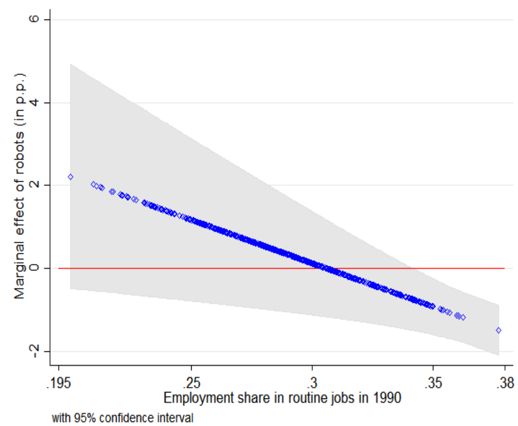
(b) Flat region



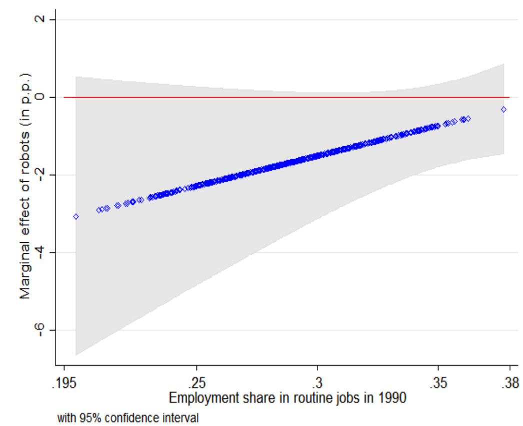
(c) Phase-out region



(d) Near phase-out region



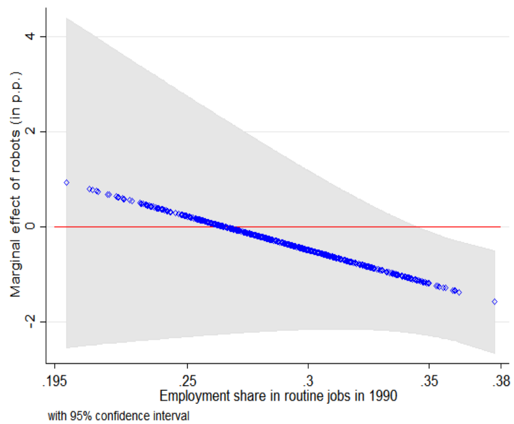
(e) Above the near phase-out region



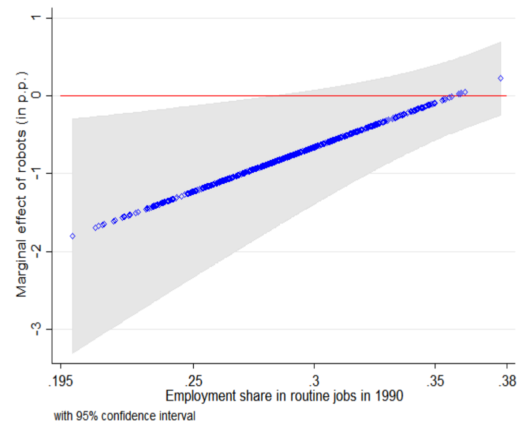
(f) Zero earnings

Notes: The marginal effects and confidence intervals for each sample are calculated based on the IV estimates in Table 1.5. The near-phase-out region is from the end of the phase-out to 25% above the end of the phase-out which is also the beginning of the above the-near-phase-out region.

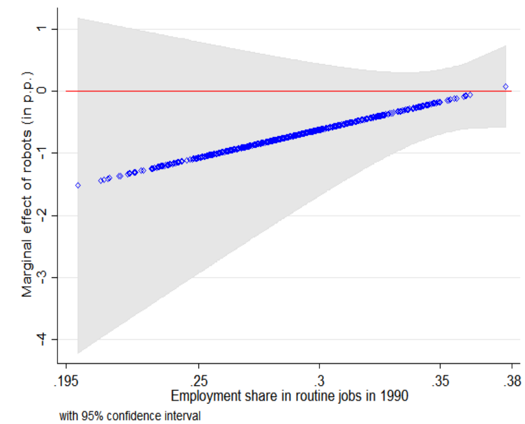
Figure 1.8 Marginal effects of industrial robots on the ratio of tax-filers by earnings region, 1990-2015, Single filers with children



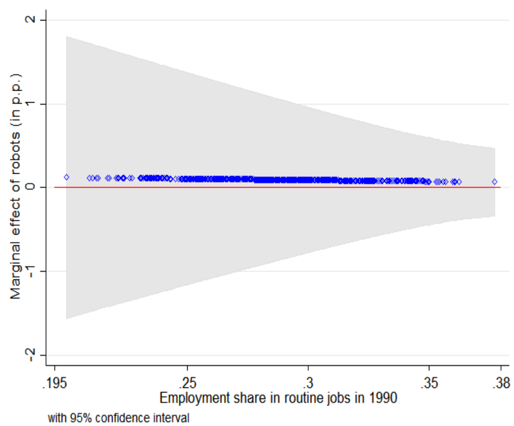
(a) Phase-in region



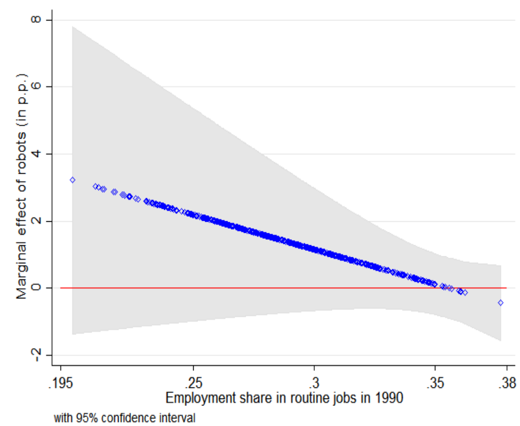
(b) Flat region



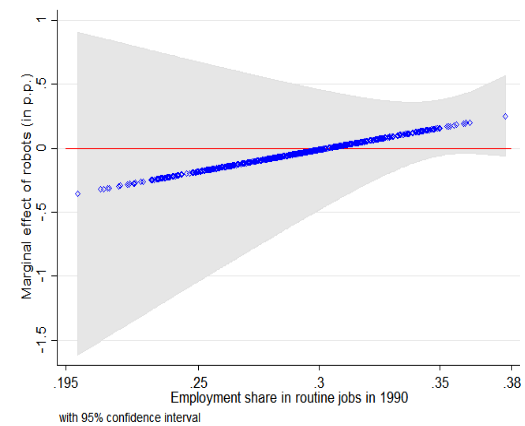
(c) Phase-out region



(d) Near phase-out region



(e) Above the near phase-out region



(f) Zero earnings

Notes: T The marginal effects and confidence intervals for each sample are calculated based on the IV estimates in Table 1.5. The near-phase-out region is from the end of the phase-out to 25% above the end of the phase-out which is also the beginning of the above the-near-phase-out region.

Figure 1.9 Marginal effects of industrial robots on the ratio of tax-filers by earnings region, 1990-2015, Married filers with children

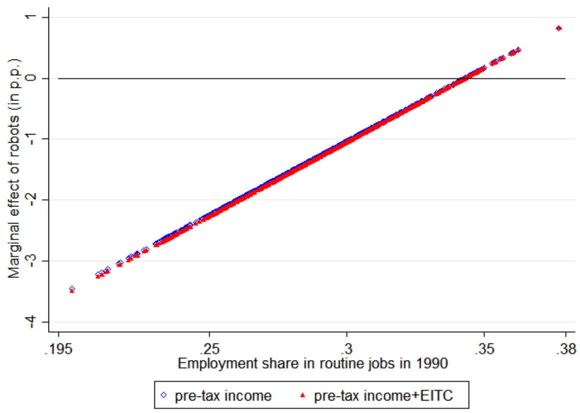
	Pre-tax income				Pre-tax income+EITC			
	All	Childless	Single, children	Married, children	All	Childless	Single, children	Married, children
Panel A. Change in ratio of tax filers under 50% of official poverty threshold, 1990-2015								
Exposure to robots in US	-1.608	-1.509*	-8.272**	-0.168	-1.632	-1.458	-8.350***	0.003
2004-2015	(1.108)	(0.906)	(3.253)	(1.472)	(1.014)	(0.912)	(3.220)	(1.120)
Exposure to robots in US	4.709	4.225*	24.104***	0.470	4.779*	4.139*	24.304***	0.016
× share in routine jobs, 1990	(3.021)	(2.417)	(8.968)	(4.031)	(2.797)	(2.440)	(9.024)	(3.061)
share in routine jobs, 1990	3.931	2.957	-21.945	10.482	3.517	2.709	-15.544	10.965
	(6.908)	(7.303)	(22.356)	(9.758)	(6.527)	(7.369)	(21.969)	(6.729)
Joint significance (p-value)	0.128	0.142	0.012	0.993	0.108	0.129	0.021	0.998
Panel B. Change in ratio of tax filers under 100% of official poverty threshold, 1990-2015								
Exposure to robots in US	-1.268	-1.181	-4.941	-2.239	-1.025	-1.313	-5.778	0.640
2004-2015	(1.601)	(1.671)	(4.029)	(3.194)	(1.589)	(1.665)	(4.330)	(2.699)
Exposure to robots in US	4.578	4.458	15.542	6.411	3.761	4.851	17.908	-1.890
× share in routine jobs, 1990	(4.500)	(4.725)	(11.031)	(9.131)	(4.445)	(4.702)	(12.232)	(7.760)
share in routine jobs, 1990	38.026***	43.429***	22.695	39.769**	34.525***	42.758***	6.234	44.697***
	(10.702)	(13.254)	(24.231)	(18.174)	(10.408)	(13.202)	(25.801)	(15.650)
Joint significance (p-value)	0.024	0.093	0.124	0.781	0.042	0.069	0.149	0.970
Panel C. Change in ratio of tax filers under 150% of official poverty threshold, 1990-2015								
Exposure to robots in US	-2.197	-2.439	-7.664*	-1.563	-1.711	-2.462	-6.812	0.667
2004-2015	(1.993)	(1.939)	(4.523)	(4.631)	(1.995)	(1.936)	(4.659)	(4.365)
Exposure to robots in US	7.829	8.822	22.758*	4.773	6.493	8.900	20.630	-1.185
× share in routine jobs, 1990	(5.652)	(5.541)	(12.466)	(12.973)	(5.663)	(5.533)	(13.043)	(12.277)
share in routine jobs, 1990	46.085***	49.914***	8.355	68.927***	47.224***	49.696***	12.282	78.202***
	(14.394)	(15.979)	(31.460)	(22.954)	(13.972)	(15.931)	(29.114)	(21.759)
Joint significance (p-value)	0.013	0.022	0.122	0.905	0.009	0.020	0.179	0.886

(continued)

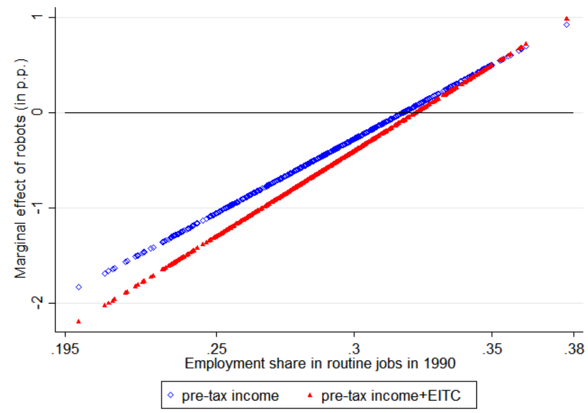
	Pre-tax income				Pre-tax income+EITC			
	All	Childless	Single, children	Married, children	All	Childless	Single, children	Married, children
Panel D. Change in ratio of tax filers under 200% of official poverty threshold, 1990-2015								
Exposure to robots in US	-2.292	-2.027	-8.013**	-5.299	-2.269	-2.037	-8.965**	-4.355
2004-2015	(2.215)	(2.231)	(3.638)	(5.103)	(2.217)	(2.234)	(3.806)	(5.097)
Exposure to robots in US	8.633	7.732	25.366***	15.809	8.522	7.755	27.505***	13.459
× share in routine jobs, 1990	(6.209)	(6.263)	(9.437)	(14.443)	(6.211)	(6.274)	(9.854)	(14.425)
share in routine jobs, 1990	53.291***	57.293***	0.322	66.906**	52.673***	57.340***	-7.245	71.485***
	(15.703)	(18.237)	(26.150)	(26.123)	(15.843)	(18.233)	(27.162)	(25.749)
Joint significance (p-value)	0.005	0.018	0.000	0.480	0.005	0.019	0.000	0.479
Weak IV F-stat	20.129	20.553	21.576	21.531	20.129	20.553	21.576	21.531
Observations	722	1444	1444	1444	722	1444	1444	1444

Notes: The table shows IV estimates of the impact of exposure to industrial robots on the share of tax-filers under official poverty thresholds, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, the interaction term is also instrumented with the product of the 30th percentile of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variables are the change in the ratio of tax-filers whose income are below official poverty thresholds to total taxpayers from 1990 to 2015. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total tax-filers of CZs in 1990. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10. **p<0.05. ***p<0.01.

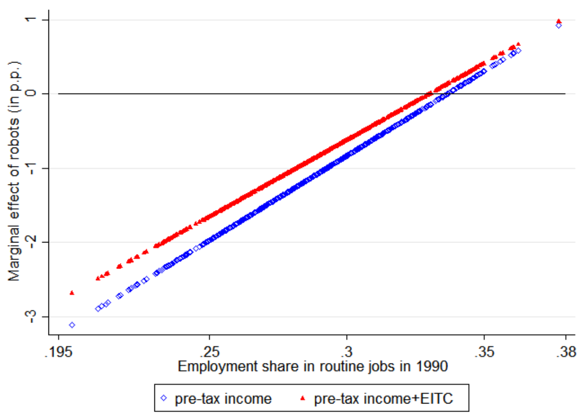
Table 1.6 The effect of exposure to industrial robots on the ratio of tax filers under official poverty thresholds (IV estimates)



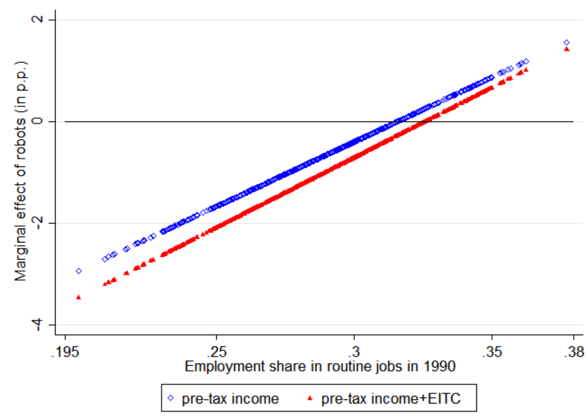
(a) Under 50% official poverty threshold



(b) Under 100% official poverty threshold



(c) Under 150% official poverty threshold



(d) Under 200% official poverty threshold

Notes: The marginal effects are calculated based on the IV estimates in Table 1.6.

Figure 1.10 Marginal effects of industrial robots on the ratio of single filers with children below official poverty thresholds, 1990-2015

1.5 Conclusion

With expectations that the recent technology such as robots and AI could have large systemic negative effects on employment through automating tasks (or jobs) previously performed by human labor, there is ongoing research about its impact on labor markets and what policy needs to mitigate negative shocks. While welfare programs in the U.S. have been transformed toward in-work aid since the 1990s, emphasizing giving work incentives to be eligible for benefits, the recent technological development leads to a discussion about Universal Basic Income that does not impose any work requirements for cash assistance. In this paper, by focusing on EITC that is now the most important and the largest cash assistance in the U.S., I examine whether the current in-work aid still can help low-income families in the sense that they still can access benefits if they can find another job after being laid off due to automation. My empirical specification is set up to investigate the above statement indirectly by questioning whether there is a difference in EITC usage across local labor markets by the extent of automation. For this, automation is measured by the exposure to industrial robots at the commuting zone level, suggested by Acemoglu and Restrepo (2017).

When using the difference in EITC usage from 1990 and 2015 as the outcome variables, the results show that there is overall no statistically significant difference in EITC recipiency rates and average benefits across commuting zones depending on the exposure to robots. However, despite their imprecision, the estimates are suggestive of heterogeneous effects of automation on EITC usage at the commuting zone level by the 1990 employment share in routine-intensive jobs. For the single filers with children, the EITC recipiency rates appear to increase in commuting zones with high routine share, which seems to come from an increase in the share of families with earnings in the phase-in region. In contrast, their recipiency rates seem to be unchanged in commuting zones with low routine share, but the results suggest the possibility of increases in the shares of families in the near phase-out and above the phase-out regions. Given that earnings in the near phase-out region are still mostly below the median household income, the expansion of EITC eligibility could support more single parents under automation. Also, the analysis presents that EITC helps the single filers with children out of poverty under automation, especially when their pre-tax income is under 50 percent or 200 percent of official poverty thresholds.

Although my work suggests that EITC has supported low- and moderate-income families under automation at the local aggregate level to date, these results may be affected by the still low adoption of industrial robots, which is growing rapidly. So, continued research will be necessary for the future. Also, given that some previous studies show negative impacts on employment, especially on low-wage workers, my work suggests that they still seem to access in-work aid through work. However, at the individual level, there could be differences in unemployment duration, earnings loss from changing jobs, working conditions, work arrangements, etc. Thus, further research needs to examine job quality for low-income workers under automation at the individual or household level.

CHAPTER 2

DISTRIBUTIONAL EFFECT OF AUTOMATION ON INCOME AND EARNINGS

2.1 Introduction

As the recent technologies such as Machine Learning (ML), Artificial Intelligence (AI), and robotics are considered as making a much broader range of tasks or jobs automatable, many studies about the effect of automation on the labor market are ongoing. While empirical studies to date have shown the mixed results about its impact on wage (or productivity), most of them focus on the mean effect of automation. However, just looking at the mean effect sometimes masks heterogeneous effects across the distribution of an outcome variable, particularly when you are interested in the change in earnings and income. For example, Autor et al. (2013) show that the increase in Chinese import competition decreases wages on average, and the distributional analysis in Chetverikov et al. (2016) presents that its negative effect is larger at the bottom of the wage distribution.

Also, the recent theoretical model about technology based on the task assignment framework assumes that each skill type of human labor for doing a task can be substituted by technology. However, the different types of labor and technology are employed in different tasks in the equilibrium since they have a comparative advantage across different tasks. For example, if technology has a comparative advantage in doing middling tasks, it replaces middle-skilled workers and affects the demand and the supply for each type of workers: a decrease in the demand for middle-skilled labor, an increase in the demand for low- and high-skilled labor, and a rise in the labor supply for low- and high-skilled tasks due to re-allocation of middle-skilled labor. Although the effect of technology on the mean wage is still important, the theoretical model at least suggests that its impact on the wage can be heterogeneous across the wage distribution when we use wage for the measure of labor productivity.

Therefore, this paper primarily aims to examine the distributional effects of automation on earnings and income. A traditional approach for estimating distributional effects using quantile regression, which enables us to estimate the impact of covariates on the conditional quantiles of the outcome variable. However, the thing is that automation, the main explanatory variable, is measured at the local labor market level, not at the individual level. It implies that the outcome variable also needs to be defined at the local labor market level to utilize the traditional quantile regression for quantile treatment effect. Considering that we are interested in income as the outcome variable, the income variable at the local labor market level are to be a kind of descriptive statistics of individual-level income variables. In this case, the traditional quantile regression examines the effect of automation just on the quantiles of the summary statistics of the outcome variable, for example, the quantiles of the `mean' income.

Given that my research goal is to study whether the effect of automation on income is different across the quantiles of income by exploiting the variation in automation across local labor markets, the traditional quantile regression is not a suitable method. One may think of using simple linear regression, defining the ratio of two specific quantiles at the local labor market level as the outcome variable such as the 90/10, the 90/50, or the 50/10 ratio. However, it still cannot provide full information about the impact of automation on income distribution, although those ratio variables are commonly used as inequality measures.

Fortunately, there is a novel econometric framework that can be applied to our research question, which is recently developed by Chetverikov et al. (2016). The model is very useful when you have a treatment at the aggregate-level (referred to "groups" which could be states, counties, etc.) and are interested in its distributional effect on outcomes observed at the micro-level (ex. individuals or firms within states). As a alternative way, you can also consider the panel data quantile regression. However, if a group-level treatment is endogenous, its effect cannot be estimated by the fixed effect approach due to within-group transformation. Thus the framework by Chetverikov et al. (2016) is a proper method for my research since I focus on identifying the effect of automation, a group-level (endogenous) treatment, on the distribution of income within each group, where groups are local labor markets.[40] The most useful thing of the framework is to allow us to control

[40] Note that this framework can be applied to many other situations, for example, a study on the effect of a school-level treatment on the within-school distribution of teachers' quality, or a study on the effect of a country-level treatment on the distribution of the firm-level productivity within each country.

individual characteristics that also could affect the outcomes when examining the effect of a group-level treatment on the outcomes. Moreover, it enables us to study whether the effects of individual characteristics on the outcomes vary with the group-level treatment. For example, given that the education attainment affects the distribution of earnings, these education effects could change with automation, such as the education effect is stronger or weaker in regions with higher automation.

While empirical studies about the impact of automation on the wage usually use hourly earnings or weekly earnings approximately calculated by several variables such as annual earnings and working hours, this paper additionally focuses on annual earnings itself as well as annual income. From the perspective of the basic economic theory, hourly earnings could be a more proper measure for the (equilibrium) wage because it represents the compensation for an additional one unit of the labor supply. So, examining the effect on hourly (or weekly) earnings can tell us roughly how automation changes the (equilibrium) wage of working population. However, automation can reduce working hours of some workers at their current job due to a decrease in tasks to perform or make some workers unemployed. Both impacts are reflected in the change in annual earnings even though some of these changes are also included in the change in hourly earnings because hourly earnings are not the direct measure from survey respondents, but are calculated from annual earnings. In addition, there are other types of income sources except for earnings, for example, the social safety net for the unemployed and low-income people. When you are interested in whether automation increases either the poor or the rich more from the perspective of income inequality, using an annual-based measure could be better.[41]

My research contributes to the literature in two ways. First, considering that the literature mainly focuses on the average effect, this paper shows the distribution effect of automation on different types of income from a perspective of empirical study. The results for income, measured for a year, suggest negative impacts of automation on the income distribution overall, and it is worse for low-income people, especially below the 30th percentile of the income distribution. However, the results for hourly earnings indicate that automation, measured by industrial robots, would decrease the wage for low-earner, but increase the wage for middle- and high-earner. This difference suggests the possibility of the adjustment of annual working hours by automation, which could occur due to short-term unemployment or decreases in working hours at employed jobs.

41 However, you need to be cautious that the income variables of the Census data do not contain all income sources that individuals can make.

Moreover, based on the flexibility of the econometric model, the analysis shows how the impact of individual-level characteristics on income is affected by automation. The results provide that i) the negative gender effect on income could be worse by automation through the adjustment of hours worked, particularly at the bottom of the distribution, ii) automation could make the positive education effect on income stronger in the middle and the top of the income distribution, due to an increase in the demand for more-skilled labor.

Secondly, from a perspective methodological application, my research suggests the importance of the consideration of individual-level characteristics for estimating the effect of a group-level treatment on income distribution. Comparing the results with Acemoglu and Restrepo (2020), the inclusion of individual-level covariates seems to reduce the negative effect of automation on low-wage earners and increase the precision of the estimates overall.

The rest of the paper proceeds as follows. Section 2.2 briefly discusses the related literature. Section 2.3 describes the econometric framework and the empirical specification, and relevant data are presented in Section 2.4. Section 2.5 provides the empirical findings, and Section 2.6 concludes.

2.2 Related literature

Research about the effect of new technology on the wage (or income) distribution has been one of the major interests in economics over the decades. Since the 1970s, many developed countries have been witnessed increases in wage inequality, and the Skill-Biased Technological Change is one hypothesis explaining that phenomenon. It assumes that technological progress augments the productivity of skilled labor more than that of unskilled labor, which then it leads to an increase in the relative demand for skilled labor and the skill premium. However, in the history of economic development, technology may not always have advanced in favor of the skilled labor, for example, automation in the nineteenth and the early twentieth century deskilled the production process and made mass production with less-skilled labor possible.

As a more recent one, the Routine-Biased Technological Change hypothesis tries to explain the impact of digital technology, represented by computer usage combined with Information and Communication Technology. The hypothesis is based on the task assignment framework and focuses on a feature of digital technology, its comparative advantage over medium-skilled labor (Acemoglu and Autor, 2011). By replacing medium-skilled workers, it can drive job polarization in employment share. The decrease in the demand for the medium-skilled labor puts downward pressure on their wages, which could lead to a rise in the upper-tail wage inequality and a fall in the lower-tail inequality.

Overall, the history implies that the impact of technological progress on the wage distribution (and inequality) can be changed depending on the features of new technology (Acemoglu, 2002). The most recent ongoing technological advances, like ML and AI, suggest the possibility that they can displace tasks (or jobs) previously considered as requiring relatively high skill, for example, paralegals and medical diagnosticians. Acemoglu and Restrepo (2018b) build a theoretical model in which both low and high skill workers are facing automation. By focusing on the equilibrium in where automation displaces a part of both low and high skill workers, they show that the change in the wage of a particular skill group of workers directly affected by automation is ambiguous because two effects created by automation, the displacement effect and the productivity effect, affects the wages in the opposite way. The former has a negative impact on the wage by reducing demand for that labor, but the latter raises the wage by increasing demand for that labor worked in non-automated tasks due to the reduction of production cost. Moreover, the displacement effect could not be limited to directly affected skill group of workers because they could displace the other skill group of workers with whom they may compete. For instance, automation displacing high skill workers also could replace some of the low skill workers since displaced high skill workers may compete with them. However, the paper shows that this high skill automation decreases the wage inequality between low- and high-skilled workers because the direct displacement effect, the decrease in the wage of high skill workers, is greater than the indirect effect, the fall in the wage of low skill workers. On the contrary, automation replacing low skill workers increases inequality.

As objective measurements for the latest technologies being developed are quite limited, empirical studies about the recent technologies largely rely on industrial robot data that is introduced in Graetz and Michaels (2018) for the first time. Those studies to date concentrate on the average effect of robot adoption on employment and wage, which shows mixed results by country and the types of dependent variables. While Graetz and Michaels (2018) present the positive impact of robots on labor productivity, measured by value-added per hour worked, Acemoglu and Restrepo (2020) find its negative effect on hourly wage in the U.S. Besides, Chiacchio et al. (2018) show that industrial robots have no impact on wage growth in six European countries.

So, my research contributes to the literature by examining the distributional effect of robots on wage (and income), which is not revealed in the mean effect. Note that Acemoglu and Restrepo (2020) show the effect of industrial robots on the wage distribution, but they do not consider individual-level characteristics that can affect the wage distribution. Hence, my research extends it by considering individual-level covariates as well as by exploring how the distributional effect of those individual-level covariates varies with the extent of robot adoption.

2.3 Econometric framework

2.3.1 Quantile regression for group-level treatments

In this section, I start by introducing the grouped instrumental variable (IV) quantile regression approach which is recently developed by Chetverikov *et al*. (2016). The methodology is practically useful when you are interested in the effect of a endogenous group-level treatment on the within-group distribution of the micro-level outcome. In this paper, we are interested in estimating the distributional effect of automation on income, by considering that individual income is partly affected by individual characteristics and that automation could be endogenous.

Chetverikov *et al*. (2016) develop the general econometric framework that can be applicable to our research interest and suggest a computationally simple two steps to estimate the group-level treatment effect. The general version is given by

$$Q_{y_{ig}|z_{ig},x_g,w_g\varepsilon_g}(\tau) = \boldsymbol{z}'_{ig}\alpha_g(\tau) + \varepsilon_{ig}(\tau) \qquad \text{for } \tau \in (0,1) \tag{2.1}$$

$$\alpha_{g,j}(\tau) = \beta_{0,j}(\tau) + \beta_{1,j}(\tau)x_g + \boldsymbol{w}'_g\gamma(\tau) + \varepsilon_{g,j}(\tau) \qquad \text{for } \tau \in (0,1) \text{ and } j = 1, \cdots, d_z \tag{2.2}$$

where $Q_{y_{ig}|z_{ig},x_g,w_g\varepsilon_g}(\tau)$ is the τth conditional quantile of y_{ig}; $\boldsymbol{z}'_{ig}$ is a d_z-vector of observable individual-level covariates including a constant (so the first component of $\boldsymbol{z}'_{ig}$ is one) and $\alpha_g(\tau) = \left(\alpha_{g,1}(\tau), \cdots, \alpha_{g,d_z}(\tau)\right)'$ is a set of group-specific effects, where the first component, $\alpha_{g,1}(\tau)$, is the intercept of the vector; x_g denotes a group-level treatment, here referred to as automation; $\boldsymbol{w}'_g$ is a vector of observable group-level control variables; and $\varepsilon_{g,j}(\tau)$ accounts for all unobservable group-level covariates that affects the distribution of y_{ig} and are also related to x_g. By assuming the form of linear regression for the group-specific effects in equation (2.2), we can study how these group-specific effects, $\alpha_{g,j}(\tau)$, depend on the extent of automation based on the estimator $\beta_{1,j}(\tau)$ for each group-specific effect.

Note that, as mentioned in Chetverikov et al. (2016), the above general version can be simplified to one equation. However, when there are group-level unobservable variables, the standard quantile regression with too many fixed effects can substantially increase the computational time (Figueiredo and Lima, 2019). In addition, the application of the classical panel data with fixed effects to address unobservable variables eliminates time-invariant variables by the within-group transformation. In our case, it implies that x_g, individual-invariant variable, will be eliminated so that we cannot retrieve our main interest, $\beta_1(\tau)$.

More importantly, the general version allows group-specific coefficients on individual-level covariates to vary across groups and specifies the form of linear regression for each component of $\alpha_g(\tau)$ as equation (2.2), which enables us to study whether the effect of individual-level covariate j on the outcome varies with the group-level treatment, x_g. For example, if the education attainment such as college degree or more has a positive impact on income, then equation (2.2) makes us examine whether the group-level treatment changes the extent of education effect on income. The same interpretations can be applied to the other group-specific coefficients on individual-level covariates used in equation (2.1). Note that the first component, $\alpha_{g,j}(\tau)$, is the group-specific coefficient on the constant term in equation (2.1). So, it can be regarded as residual income inequality because it represents the expected group income which cannot be explained by observable individual-level characteristics.

When we are interested in estimating $\beta_1(\tau)$, the effect of group-level treatment, Chetverikov et al. (2016) show that the it can be consistently estimated through the following two steps. The first step is for estimating equation (2.1), in which we separately run quantile regressions of y_{ig} on $\boldsymbol{z}'_{ig}$ for each group and each quantile index τ. Then, we can obtain $\alpha_g(\tau) = \left(\alpha_{g,1}(\tau), \cdots, \alpha_{g,d_z}(\tau)\right)'$ where

$$\hat{\alpha}_g(\tau) = \arg\min_{a} \sum_{i=1}^{N_g} \rho_\tau\left(y_{ig} - \boldsymbol{z}'_{ig}a\right) \tag{2.3}$$

and $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$. Next, as the second step is for equation (2.2), we can obtain $\beta_1(\tau)$ by performing either OLS regressions of $\alpha_{g,j}(\tau)$ on x_g and other group-level controls (w_g) for each element j and each quantile index τ if x_g is exogenous or 2SLS regressions with instrument variables if x_g is endogenous, $\mathbb{E}\left[x_g \varepsilon_g(\tau)\right] \neq 0$. Note that Chetverikov et al. (2016) emphasize that, as a quantile extension of Hausman and Taylor (1981), a specific form of internal instrument can be constructed from some of micro-level covariates, z_{ig}, if some of them, $z_{ig,k}$, are exogenous in the sense that $\mathbb{E}\left[z_{ig,k} \varepsilon_g(\tau)\right] = 0$ for all τ. However, since we have another option for the instrument variable for automation based on the previous studies about automation, I follow the way in the literature. Additionally, the estimator $\beta_1(\tau)$ is consistent and asymptotically normal under the growth condition, $G^{2/3}\left(\log N_g / N_g\right) \to 0$ as $G \to \infty$, and other regularity assumptions (see Chetverikov et al., 2016). Their simulation results show that the bias of $\beta_1(\tau)$ falls close to zero as the number of groups (G) and the minimum number of observations per group (N_g) increase from 25 to 200, which are substantially smaller than numbers in our analysis.

When individual-level covariates are less likely to vary with the group-level variables, then we can only focus on regression of estimated coefficients $\hat{\alpha}_{g,1}(\tau)$ on the group-level treatment and other group-level covariates through OLS or 2SLS with instruments in the second step. Also, Chetverikov et al. (2016) mention that the first step can be simplified to just calculate the τth quantile of the outcome variable y_{ig} within each group when there are no individual-level covariates. Although there are several studies about the effect of a group-level treatment on the wage distribution using the methodology of Chetverikov et al. (2016), most of them apply the simple version by focusing on the coefficient on the constant term, $\alpha_{g,1}(\tau)$. In addition, most of them select τth quantile of the outcome variable in the first step, which can drastically decrease the computational time.[42] The closest study to this paper is Acemoglu and Restrepo (2020) that show the distributional effect of the exposure to industrial robots by simply calculating τth quantile of log hourly wage in the first step, which is equivalent to estimating local labor market specific quantile regressions in equation (2.1) with just an intercept. Without controlling for any other individual-level covariates, they simply regress the τth quantile of the outcome from each labor market on the group-level variables in equation (2.2). However, Adem (2018), examining the distributional effect of Chinese imports competition in the U.K., suggests the possibility that the

[42] For the wage-related literature, see Adem (2018), Autor, D. et. al. (2019), Choi and Xu (2019), Poliquin (2020), Wang et. al. (2018).

inclusion of individual-level characteristics in the first step could lead to different conclusions since it would change statistical power in estimates. Thus, my research contributes to the related literature by methodologically controlling individual-level covariates in the first step and allowing additional flexibility of the general model (2.1) and (2.2).[43] To be specific, my approach nets out individual level covariates that may differ across local labor markets and also permits analysis of the effects of automation on the returns to these individual level covariates.

2.3.2 Empirical specification

To adopt the grouped IV quantile regression approach, developed by Chetverikov et al. (2016), to the analysis of automation, let i, c, and t denote individuals, local labor markets, and years, respectively. Given that we have repeated cross-sections of individual $i = 1, \cdots, N_{ct}$ in local labor market $c = 1, \cdots, C$, measured by commuting zones (CZ), in two different time period, t_0 and t_1 and define group g as a local-labor-market-by-year cell, the micro-level empirical model for the first step can be written as follows:

$$
\begin{aligned}
Q_{y_{ict}}(\tau) = {} & \alpha_{ct,1}(\tau) + \alpha_{ct,2}(\tau) age_{ict} + \alpha_{ct,3}(\tau) age_{ict}^2 + \\
& \alpha_{ct,4}(\tau) female_{ict} + \alpha_{ct,5}(\tau) married_{ict} + \alpha_{ct,6}(\tau) white_{ict} + \\
& \alpha_{ct,7}(\tau) edu_{ict} + \varepsilon_{ict}(\tau) \qquad \text{for } \tau \in (0,1) \qquad (2.4)
\end{aligned}
$$

where the outcome variable y_{ict} is log annual income, log annual earnings, log weekly earnings or log hourly earnings of individual i in CZ c in year t, a set of individual-level covariates, y_{ig} in equation (2.1), includes age, age-squared, and dummy variables for gender (equal to one if the individual is female), marital status (equal to one if the individual is married), race (equal to one if the individual is white), and education (equal to one if the individual has college degree or more). The constant term, $\alpha_{ct,1}(\tau)$, can be interpreted as residual income because it represents the expected CZ-year log income at each quantile which cannot be explained by observable individual

[43] Two studies use the flexibility of the general version of the model by examining whether the group-specific coefficients on individual-level covariates vary with a group-level treatment. See Figueiredo and Lima (2020) and Okay and Yamadaz (2019).

covariates. Here, I can obtain $\hat{\alpha}_{ct}(\tau) = \left(\hat{\alpha}_{ct,1}(\tau), \cdots, \hat{\alpha}_{ct,7}(\tau)\right)'$ by implementing separate quantile regressions by CZ c and year t for each quantile τ using personal weights.

After the first step, we now have the CZ-level panel data with two time periods. Since automation is measured as the exposure to industrial robots (EIR) between t_0 and t_1 as you can see later, I take the long difference specification in the second step and the empirical model for the first component of the vector $\hat{\alpha}_{ct}(\tau)$ will be

$$\Delta\hat{\alpha}_{c,1}(\tau) = \beta_0(\tau) + \beta_1(\tau) \cdot \Delta EIR_c^{US} + \boldsymbol{w}_c'\theta(\tau) + \delta_c(\tau) + \epsilon_c(\tau) \tag{2.5}$$

where $\Delta\hat{\alpha}_{c,1}(\tau) = \hat{\alpha}_{c,t_1,1}(\tau) - \hat{\alpha}_{c,t_0,1}(\tau)$ is the change in the τth quantile of residual income in CZ c between t_0 and t_1 that is obtained from the first step, ΔEIR_c^{US} indicates the exposure to industrial robots in CZ c in the U.S. from t_0 and t_1, $\boldsymbol{w_c}$ represents CZ-level covariates, and δ_c is region fixed effects at the census division level. To be specific, the CZ-level covariates include the exposure to Chinese imports between t_0 and t_1, the share of employment in routine-intensive jobs in year t_0, and CZs' demographic and economic characteristics in year t_0 that are used in Acemoglu and Restrepo (2020). The last CZ-level variables include the ratio of the female population; the ratio of the population over 65 years old; the ratios of the population with no college, some college, college or associate degree, and master or doctoral degree; the ratios of whites, blacks, Hispanics, and Asians; log population; the ratio of employment in the manufacturing industry; the ratio of light manufacturing (the textile industry and the paper, publishing, and printing industry); and the female share of manufacturing employment. Note that equation (2.5) can be specified for other components of the vector $\hat{\alpha}_{ct}(\tau)$ in the same way. For example, each estimated $\beta_1(\tau)$ can present whether gender, race, and education effects on income change with automation when I use $\Delta\hat{\alpha}_{c,4}(\tau)$, $\Delta\hat{\alpha}_{c,6}(\tau)$, and $\Delta\hat{\alpha}_{c,7}(\tau)$ as the dependent variables, respectively.

Equation (2.5) can be estimated by OLS if the measure for automation is exogenous. To address the possible endogeneity that any unobservables affecting the labor demand in CZ c may impact on the adoption of robots in that CZ, I implement 2SLS regression for each quantile τ where the instrument variable is constructed by using the usage of industrial robots in several European countries. For the estimation of equation (2.5), I use the CZ share of national population in year t_0 as the weight, and the standard errors are clustered at the state level to allow for arbitrary spatial correlation at the state level.

2.4 Data

In this paper, local labor markets are defined as Commuting Zones (CZs), which are developed by Tolbert and Killian (1987) and Tolbert and Sizer (1996). As CZs are groups of counties with strong commuting ties based on residence-to-work commuting data from the 1980 and 1990 Census, it provides geographical units representing local labor markets in an economic context. I focus on 722 CZs that cover the mainland of the U.S. excluding 19 CZs located in Alaska and Hawaii. Note that Commuting Zones are identified by Public Use Microdata Areas (PUMAs), the most disaggregated geographic variable in the Census/ACS data, and I use the crosswalk files of Dorn (2009) to map each PUMA to each 1990 CZ.[44]

A. Outcome variable: Total income and earnings

For income variables, I use the 1990 Census and the 2015 American Community Surveys (ACS) from Integrated Public Use Microdata Series (IPUMS).[45] Although the Census/ACS data has several income variables according to the type of income sources, I focus on earnings and total income at the individual level. Note that the income variables in the Census/ACS are reported as income for a year; the reference period of the Census data is the previous calendar year, and it is the past 12 months for the ACS.

First, annual earnings are defined as the sum of wage/salary income and business income. Empirical studies about the labor market often exclude the self-employed, but I include them in this study. It is because I want to put more emphasis on an aspect of earnings as one income source from economic activities rather than as wage from the perspective of labor economics. Since both variables are top-coded, each income is set to equal to 1.5 times the top-coded value for observations with top-coded values.[46] Note that business income in the Census/ACS is bottom coded since it can have a negative value. Here, I do not make any adjustments for the bottom-coded values before summing two variables, but set the minimum value of annual earnings to 1 (for the log transformation), considering negative earnings as "realized" zero earnings.

[44] For the detail, see Dorn (2009) and visit http://www.ddorn.net/data.htm.

[45] I use 5% national sample for the 1990 Census and 1% national sample for the 2015 ACS.

[46] Although there are several ways to adjust top-coded values, I follow the simplest way to multiply a fixed number as in Acemoglu and Autor (2011). Note that the specific top-code and bottom-code values are found in https://usa.ipums.org/usa/volii/top_bottom_codes.shtml

Secondly, total annual income is defined as the sum of all income sources reported in the Census/ACS. In addition to earnings, it includes i) interest, dividend, and rental income, ii) retirement income such as retirement, survivor, and disability pension income, other than Social Security, iii) Social Security income such as Social Security pensions, survivors benefits, or permanent disability insurance, as well as U.S. government Railroad Retirement insurance payments, iv) welfare income such as federal/state Supplemental Security Income (SSI) payments, Aid to Families with Dependent Children (AFDC), and General Assistance (GA), and iv) other income not specified into above sources.[47] Some of these income sources also have top-coded and bottom-coded values, so I make adjustments for the top-coded values and the minimum value of total annual income in the same way as annual earnings.

Considering that the total income covers various income sources, its use as the outcome helps us to examine the effect of automation on overall income, even though it still does not reflect all income sources.[48] Note that all income variables used in the analysis are defined at the individual level, not the household level. In the case of the total income, the analysis at the household level may be more proper, considering that government supports are usually a household (or a family) base. However, it is a little unclear what characteristic at the household level affects the household income, and the selection of covariates could be more complicated for two earners in a married household, unlike the individual-level analysis. Hence, for now, I focus on the individual level income so that the married persons are considered as separate individuals.[49]

Lastly, I calculate weekly earnings and hourly earnings since annual earnings contain two types of information, the change in wage and the change in hours worked. The weekly earnings are calculated by dividing the annual earnings by weeks worked last year (or past 12 months). The latter variable is available only in intervals in the 2015 ACS data, unlike the 1990 Census, so that I use the median value of each interval. Also, the hourly earnings are defined by dividing the weekly earnings by usual hours worked per week during the previous calendar year (or 12 months).

[47] The other income is a residual variable that is included in the total income but not applied to categorized income sources.

[48] For example, the total income does not include any realized capital gains, social insurance like unemployment benefits, and tax-system-based support like Earned Income Tax Credit.

[49] The Census/ACS data collect information on income at the individual level, so the household income is calculated as the sum of all household members' income.

Note that all dollar amounts are expressed in 2015 US dollars using the Personal Consumption Expenditures (PCE) price index. For the analysis, I focus on only individuals whose age is between 25 and 64 and exclude individuals who reside in institutional group quarters.

B. Explanatory variable: Exposure to industrial robots

Automation is measured by utilizing the industrial robot data from the International Federation of Robotics (IFR) that provides internationally comparable data for one of the recent technologies. To be specific, based on the way in the working paper version of Acemoglu and Restrepo (2020), the exposure to industrial robots in the U.S. for each CZ c between two time periods (ΔEIR_c^{US}) is constructed as follows:[50]

$$\Delta EIR_c^{US} = \sum_{i \in I} l_{ci,1990} \left(\frac{IR_{i,2015}^{US}}{L_{i,1990}^{US}} - \frac{IR_{i,2004}^{US}}{L_{i,1990}^{US}} \right) \tag{2.6}$$

where $l_{ci,1990}$ is the 1990 employment share of industry i in CZ c, which is calculated from the 1990 Census. $\left(\frac{IR_{i,t}}{L_{i,1990}}\right)$ measures industrial robot density in industry i and year t, the operational stock of industrial robots ($IR_{i,t}$) per thousand workers ($L_{i,1990}$), where $IR_{i,t}$ comes from the IFR data and $L_{i,1990}$ is from the EU KLEMS data. Based on the classification of the IFR data, the industries are categorized into nineteen sectors, 6 non-manufacturing and 13 manufacturing sectors. The six non-manufacturing industries are agriculture, hunting, forestry, and fishing; mining and quarrying; electricity, gas and water supply; construction; education, research and development; and other non-manufacturing industries. The thirteen manufacturing sectors are food products, beverages, and tobacco products; textiles, leather, and wearing apparel; wood and wood products; paper, paper products, and printing; plastic and chemical products; glass, ceramics, stone, and

[50] In their published paper, the US exposure to robots is defined as follows:

$$EIR_{c,(t_0,t_1)} = \sum_{i \in I} l_{ci,1990} \left[\frac{IR_{i,t_1}^{US} - IR_{i,t_0}^{US}}{L_{i,1990}^{US}} - g_i^{US}(t_0, t_1) \frac{IR_{i,t_0}^{US}}{L_{i,1990}^{US}} \right]$$

where $g_i^{US}(t_0, t_1)$ is the growth rate of output of industry i in the U.S. between t_0 and t_1. The term in the bracket is referred to "the adjusted penetration of robots in industry i." As mentioned in their paper, the second term in the bracket is very small number because there were few robots in the early 1990s. So, most of the variation comes from the first term.

mineral products; basic metals; metal products; industrial machinery; electrical/electronics; automotive; other transport equipment; all other manufacturing sectors.

By using robot usage in other nine European countries, the instrument variable for (ΔEIR_c^{EU}) is constructed in the following way:

$$\Delta EIR_c^{EU} = \sum_{i \in I} l_{ci,1980} \left(p_{30} \left(\frac{IR_{i,2015}^{EU}}{L_{i,1990}^{EU}} \right) - p_{30} \left(\frac{IR_{i,1993}^{EU}}{L_{i,1990}^{EU}} \right) \right) \quad (2.7)$$

where $l_{ci,1980}$ is the 1980 employment share of industry i in commuting zone c in the U.S., which comes from the 1980 Census, the industrial robot density in industry i and year t, $\left(\frac{IR_{i,t}}{L_{i,1990}} \right)$, is calculated for each European country with the IFR and EU KLEMS data, and $p_{30} \left(\frac{IR_{i,t}}{L_{i,1990}} \right)$ denotes the 30th percentile of industrial robot density among nine European countries: Denmark, Finland, France, Germany, Italy, Norway, Spain, Sweden, and the United Kingdom.[51]

C. Other variables

In the second step of the estimation, several CZ-level variables are additionally included. The first one is the measure for Chinese import competition which has been considered as an important factor affecting local labor markets in the U.S. since the 1990s. Based on the previous literature (Autor et al., 2013; Acemoglu et al., 2016), the exposure to Chinese imports in CZ c between 1991 and 2015 is constructed as follows:

$$Exposure\ to\ Chinese\ imports_c = \sum_{i \in I} l_{ci,1990} \left\{ \frac{M_{i,2015}^{UC} - M_{i,1991}^{UC}}{Y_{i,1991} - (M_{i,1991} - X_{i,1991})} \right\} \quad (2.8)$$

where $l_{ci,1990}$ is the 1990 employment share in industry i in CZ c, which is calculated from the 1990 Census, and $M_{i,t}^{UC}$ represents imports from China into the U.S. in industry i in year t. The

[51] Note that the instrument variable is also based on the way in the working paper of Acemoglu and Restrepo (2020). In their published paper, the updated IV is defined as follows:

$$Exposure\ to\ Robots_{c,(t_0,t_1)} = \sum_{i \in I} l_{ci,1970} \left\{ \frac{1}{5} \sum_{j \in EURO5} \left[\frac{IR_{i,t_1}^j - IR_{i,t_0}^j}{L_{i,1990}^j} - g_i^j(t_0,t_1) \frac{IR_{i,t_0}^j}{L_{i,1990}^j} \right] \right\}$$

where $EURO5$ countries are Denmark, Finland, France, Italy, and Sweden.

change in imports is normalized by domestic absorption, approximated by the sum of domestic shipments (Y_i) and net imports ($M_i - X_i$). Trade data come from World Integrated Trade Solution (WITS) and the domestic shipments data is from the NBER-CES Manufacturing Industry Database.[52] Note that All dollar amounts are adjusted to 2015 US dollars using the Personal Consumption Expenditures (PCE) price index.

Since any shocks on the labor market in CZ c could affect the industry import demand, I also construct the instrument variable for the Chinese import competition based on Acemoglu et al. (2016) and Lake and Millimet (2018):

$$IV\ for\ Exposure\ to\ Chinese\ imports_c = \sum_{i\in I} l_{ci,1980} \left\{\frac{M_{i,2015}^{OC} - M_{i,1991}^{OC}}{Y_{i,1989} - (M_{i,1989} - X_{i,1989})}\right\} \tag{2.9}$$

where $l_{ci,1980}$ is the 1980 employment share in industry i in CZ c, which is from the 1980 Census, and $M_{i,t}^{OC}$ indicates imports from China into eight non-US countries in industry i year t, which is normalized by the U.S. domestic absorption in 1989.[53] The eight countries include Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland.

The second one is the measure for routine-intensive jobs since a CZ with an initially higher share of the employment in routine jobs may be differently affected by the adoption of new technology. The data comes from Autor and Dorn (2013) in which the variable is defined as follows:

$$Employment\ share\ in\ routine\ jobs_{c,1990} = \sum_{j\in J} \frac{L_{cj,1990} \cdot 1[RTI_j > RTI^{66}]}{L_{c,1990}} \tag{2.10}$$

where L_{ci} indicates the employment in occupation j in CZ c, and the Routine Task Intensity RTI_j in occupation j measures relative routine task inputs of the occupation j based on the U.S. Dictionary of Occupational Titles 1977.[54]

Lastly, I add the variables mentioned in Section 2.3.2, which represents CZs' demographic and industry characteristics in 1990, and they are calculated from the 1990 Census data.

---

[52] Since the industry classification of the trade and the domestic shipment data is given at the SIC 4-digit level, the SIC 4-digit codes are converted to the 1990 Census industry code.

[53] Due to the data availability in COMTRADE data, the US exports and imports in 1989 come from https://dataweb.usitc.gov.

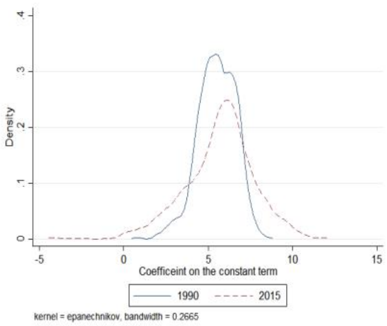
[54] See Autor and Dron (2013) for detailed information.

2.5 Results

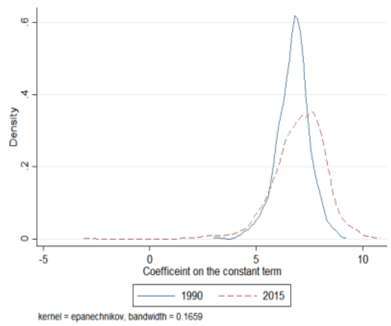
2.5.1 Baseline results

I start by briefly summarizing the estimates of equation (2.4) in the first step that is used as dependent variables in the second step estimation. After running separate quantile regressions for each CZ-year pair and each quantile, I save the estimated coefficients on the constant term and individual covariates, where each coefficient has 12,966 observations: 744 CZs×2 time periods×9 quantiles. Figure 2.1 to Figure 2.4 shows the distribution of each coefficient for each year and each quantile via Kernel Density Estimation when annual earnings are used as the outcome variable in the first step.

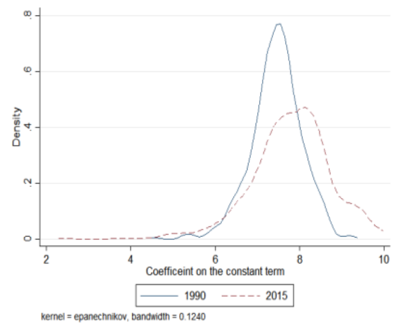
The distribution of constant term, regarded as residual earnings, in Figure 2.1 is likely to be shifted to the right as moving from a lower quantile to a higher quantile. It means that the expected CZ earnings not explained by observable individual characteristics in our specification overall tend to be larger as earnings increase. Also, when comparing two time periods, you can see that the distribution in 2015 is more dispersed, particularly above the 50th percentile of the distribution, which implies that the residual earnings differential across CZs becomes greater in 2015. Next, Figure 2.2 displays the distribution of the coefficient on female in equation (2.4), the effect of female on annual earnings. The values of the coefficient are all negative at every quantile in both years, so it indicates that women still receive lower earnings than men at each quantile in every CZs when other variables being equal. This negative effect seems to be a little stronger at the bottom quantile, but its distribution shifts to the right, closer to zero, in 2015 across all quantiles. The distribution of the coefficient on white is shown in Figure 2.3, which suggests that the distribution is quite similar over time and quantiles and that there are overall positive effects of white across all quantiles. But its contribution to annual earnings seems to be relatively smaller than other individual characteristics, given that the distribution is concentrated around zero despite heterogeneity across CZs. Lastly, Figure 2.4 presents the distribution of the coefficient on college education, which shows that having a college degree or more has an overall positive impact on annual earnings at every quantile. Although the distribution has a similar pattern over time, it looks to be shifted to the right a little in 2015, especially above the 70th percentile of the earnings distribution, which implies that the positive education effect is stronger in 2015 than in 1990.



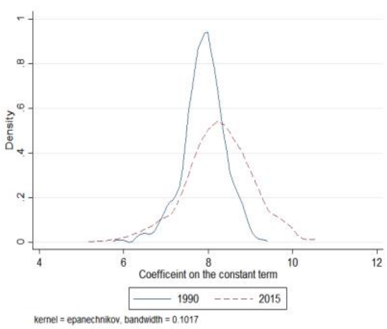
(a) 10th percentile



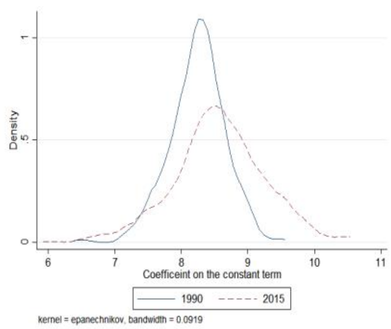
(b) 20th percentile



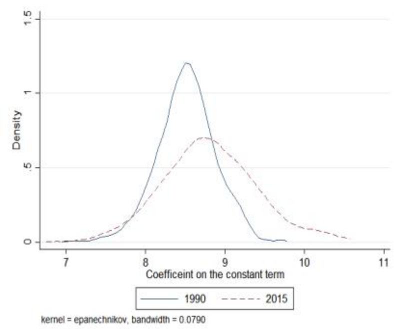
(c) 30th percentile



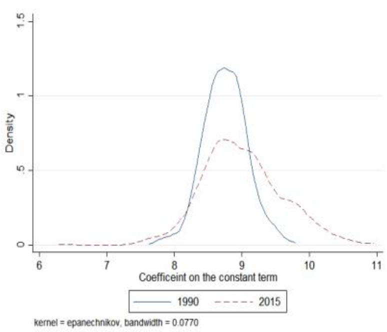
(d) 40th percentile



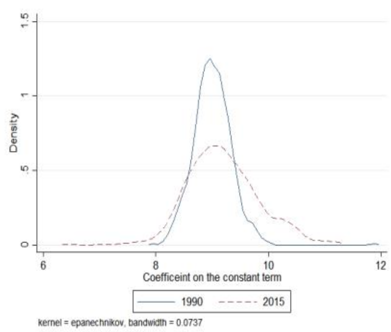
(e) 50th percentile



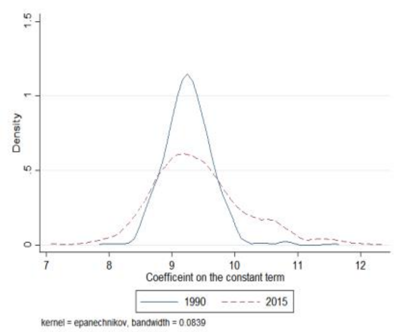
(f) 60th percentile



(g) 70th percentile



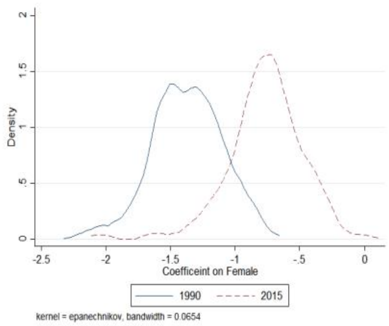
(h) 80th percentile



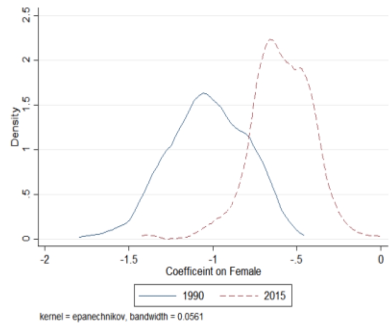
(i) 90th percentile

Notes: The graphs show the distribution of the coefficient on the constant term for each quantile, which is estimated in the first step. The number of observations for each quantile and each year is 722, and the dependent variable in the first step is annual earnings.

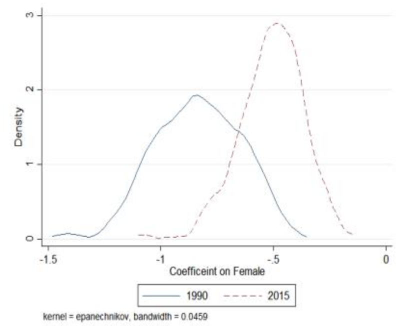
Figure 2.1 Distribution of the estimated constant term (income type: annual earnings)



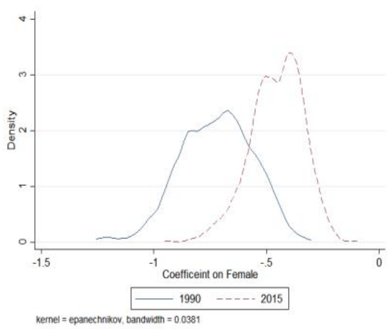
(a) 10th percentile



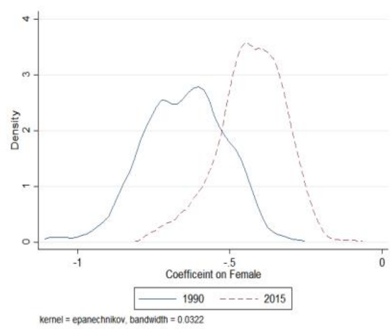
(b) 20th percentile



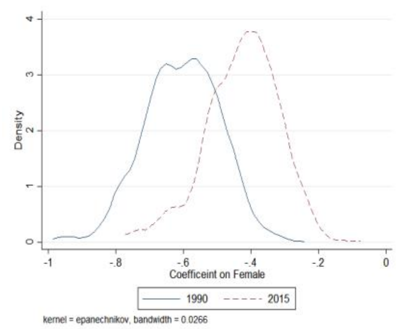
(c) 30th percentile



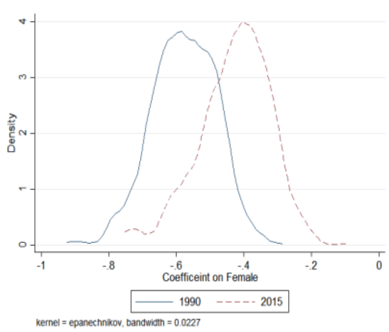
(d) 40th percentile



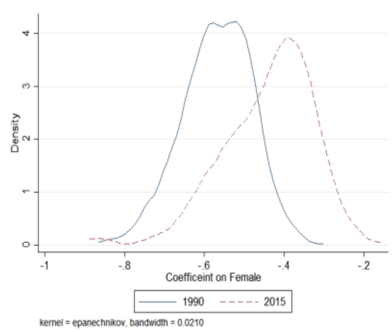
(e) 50th percentile



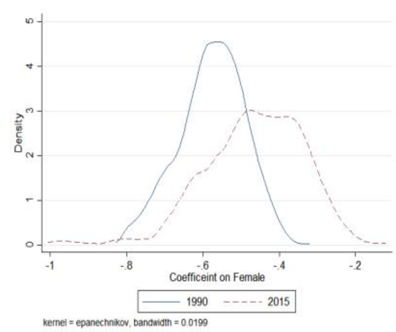
(f) 60th percentile



(g) 70th percentile



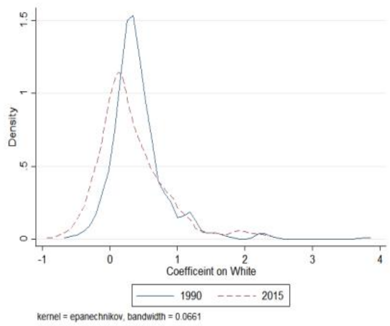
(h) 80th percentile



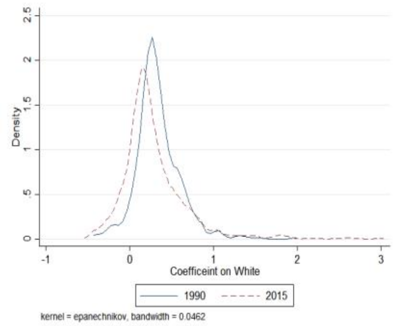
(i) 90th percentile

Notes: The graphs show the distribution of the coefficient on female for each quantile, which is estimated in the first step. The number of observations for each quantile and each year is 722, and the dependent variable in the first step is annual earnings.

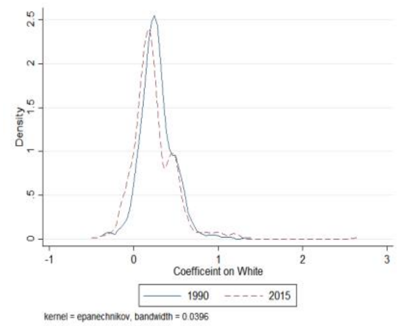
Figure 2.2 Distribution of the coefficient on female (income type: annual earnings)



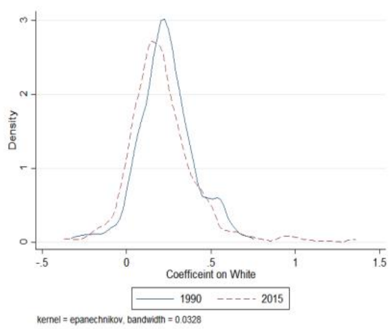
(a) 10th percentile



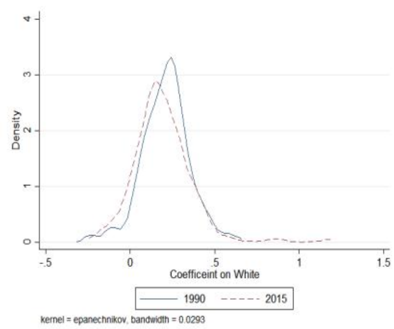
(b) 20th percentile



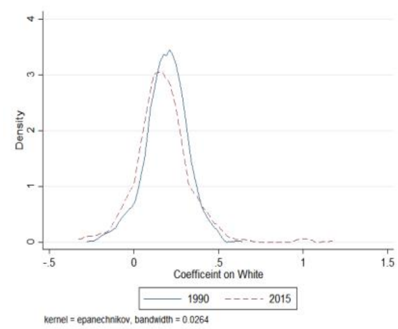
(c) 30th percentile



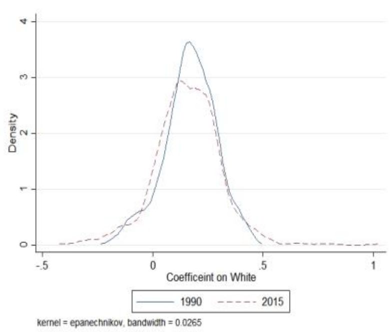
(d) 40th percentile



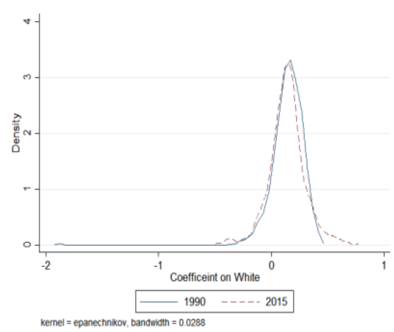
(e) 50th percentile



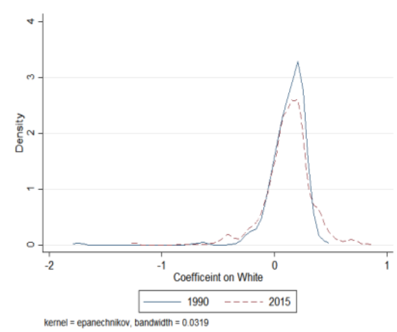
(f) 60th percentile



(g) 70th percentile



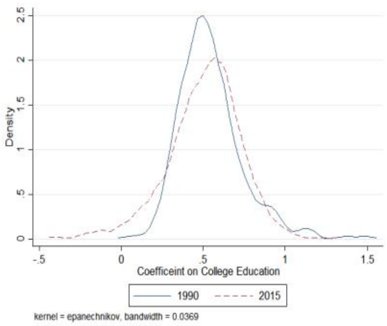
(h) 80th percentile



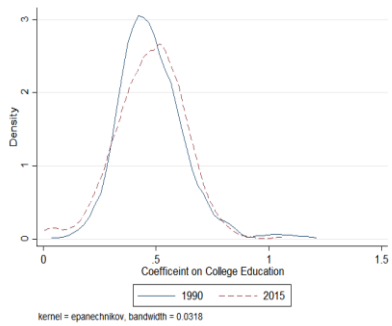
(i) 90th percentile

Notes: The graphs show the distribution of the coefficient on white for each quantile, which is estimated in the first step. The number of observations for each quantile and each year is 722, and the dependent variable in the first step is annual earnings.

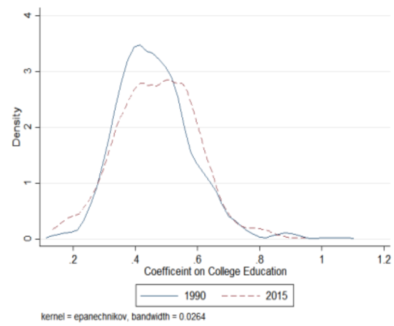
Figure 2.3 Distribution of the coefficient on white (income type: annual earnings)



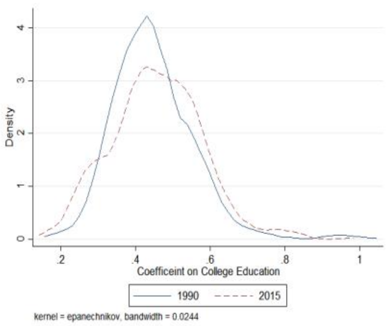
(a) 10th percentile



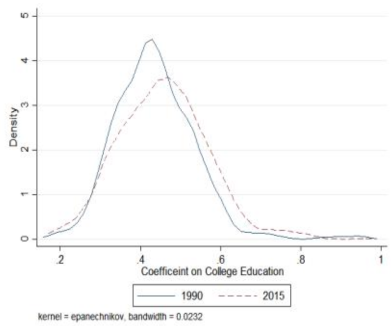
(b) 20th percentile



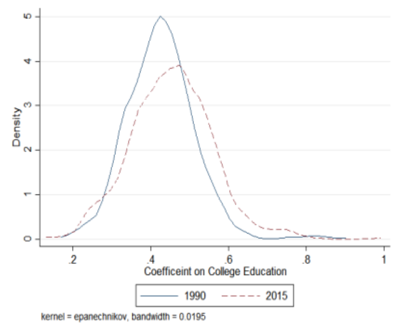
(c) 30th percentile



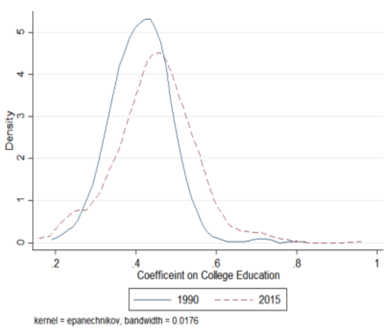
(d) 40th percentile



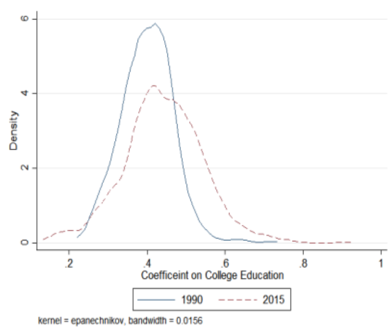
(e) 50th percentile



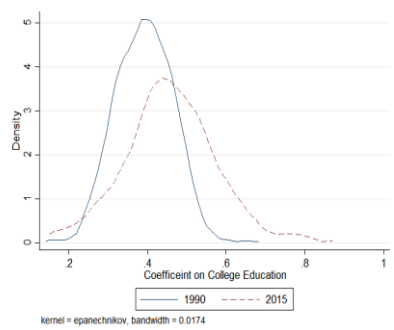
(f) 60th percentile



(g) 70th percentile



(h) 80th percentile



(i) 90th percentile

Notes: The graphs show the distribution of the coefficient on edu (having college degree or more) for each quantile, which is estimated in the first step. The number of observations for each quantile and each year is 722, and the dependent variable in the first step is annual earnings.

Figure 2.4 Distribution of the coefficient on college education (income type: annual earnings)

Among individual-level covariates included in the first step, I primarily focus on the effect of industrial robots on residual income in this section. The baseline results are presented in Figure 2.5 and Figure 2.6. The former shows the results of OLS estimation of the reduced form by quantiles, in which the exposure to industrial robots in the U.S. are replaced with the measure based on European robot adoption and the industrial composition of the CZ. And the latter presents the results of IV estimates by Two-Stage Least Squares (2SLS), where CZs' robot exposure is instrumented with the previous measure using European robot adoption.[55] The graph (a) in each Figure draws the estimates when equation (2.5) includes only the exposure to industrial robots, the graph (b) provides the results when adding region fixed effects at the census division level, the graph (c) shows the estimates when including CZ's demographic and industry characteristics in 1990 as well, and the graph (d) is the results of equation (2.5) by additionally controlling for the exposure to Chinese imports competition and the share of employment in routine-intensive jobs in 1990. These graphs imply that it could be important to control for demographic and economic characteristics of local labor markets since it could change the sign of the effect of automation, although the most estimates are not statistically significant.

The graph (d) in Figure 2.5 provides suggestive evidence that the negative impact of automation could be much worse below the 30th percentile and at the 90th percentile of the earnings distribution. Meanwhile, higher exposure to industrial robots does not affect earnings at the 10th percentile and in the middle-range of the distribution. The results indicate a possibility of reduced inequality of earnings, measured as the ratio of earnings at the 90th and 10th percentiles. But it could increase when referring to other quantiles, which implies that the distributional analysis could provide a more comprehensive understanding than focusing on representative measures of income inequality. Note that the 2SLS estimates in Figure 2.6 suggest similar results with the increase in the precision of the estimated coefficients.

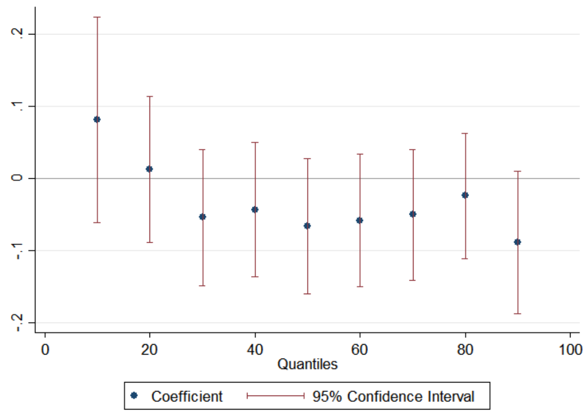
Figure 2.7 presents IV estimates via 2SLS of equation (2.5) for different types of income variables: total annual income, annual earnings, weekly earnings, and hourly earnings. Note that the graph (a) presents the results of the total annual income, which plots the estimates only above the 30th percentile of the distribution. It is because, in the analysis for the total annual income,

[55] The exposure to Chinese import competition in the U.S. is also replaced with the measure calculated by other non-U.S. countries in the OLS estimation of the reduced form. Then, in the IV estimation, the U.S. exposure is instrumented with the other countries' exposure.

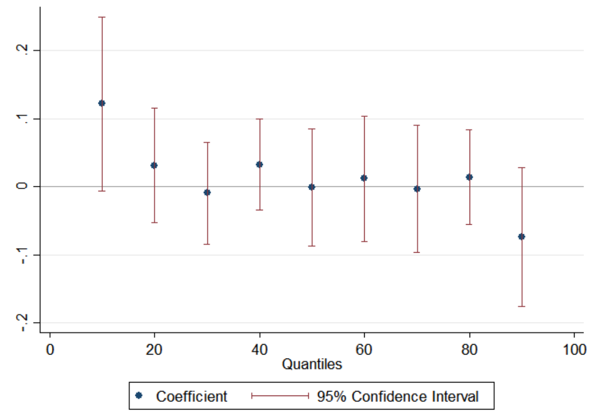
most observations in the bottom of the distribution have zero income so that quantile regressions at the 10% and 20% quantile are not identified. First, the estimates on total annual income and annual earnings present that the negative impact of automation is concentrated at the bottom of the distribution, especially on the 20th and the 30th percentiles. So, automation, measured by industrial robots, seems to be worse for low-income people and to have no effect on people with middle- and high-income, which implies that industrial robots could increase lower-tail inequality. Also, the similar results on different income measures seem to reflect that, for most people, the primary source of total income is earnings.

Second, the results on weekly earnings and hourly earnings suggest that the negative effect of industrial robots on annual earnings could be in part caused by the adjustment in working hours for a year. The graph (c) and (d) in Figure 2.7 present that robots have a negative effect on weekly and hourly earnings at the bottom of the distribution, like the results on annual earnings, and the negative impact may be weaker than one on annual earnings. Besides, the graphs also show the possibility of a positive impact on weekly and hourly earnings above the middle quantile, unlike the estimates on annual earnings, even though most estimates are statistically insignificant. Overall, the difference by different types of income variables may be suggestive evidence that robots negatively affect low-earners through the fall in wage as well as reduced working hours, where the latter could be caused by short-term unemployment or an adjustment in the working hours at current jobs.

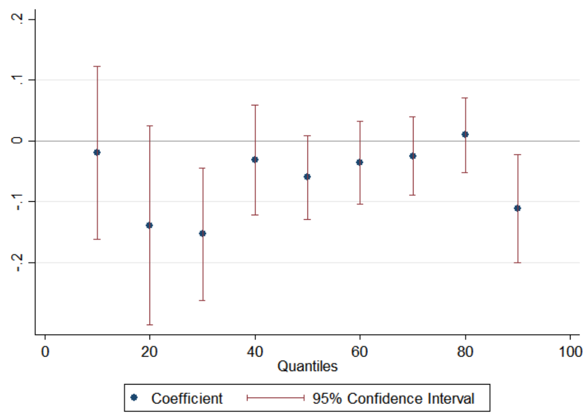
Third, the distributional effects on weekly and hourly earnings suggest that the negative mean effect of industrial robots on the wage in the literature could arise from its impact at the bottom of the wage distribution. Also, from a perspective of the theoretical model in Acemoglu and Autor (2011) and Acemoglu and Restrepo (2018b, 2020), the industrial robot, as a measure for recent technology, seems to have a comparative advantage over relatively low-skilled labor based on its negative effect at the 20th and 30th percentiles of the distribution. And, there is a possibility that it could indirectly increase the demand for relatively more skilled workers, considering its positive effect above the middle quantile of the distribution, although the estimates are not statistically significant.



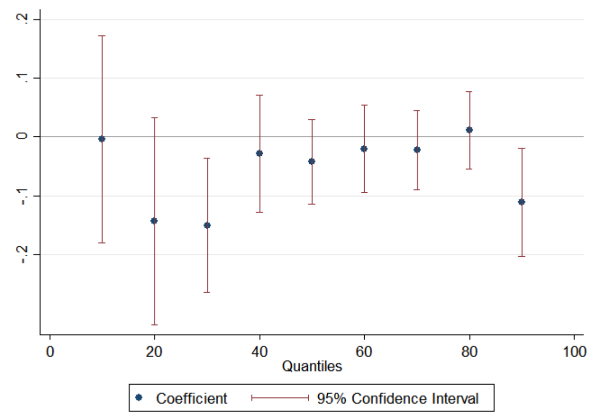
(a) With no controls



(b) Adding census division



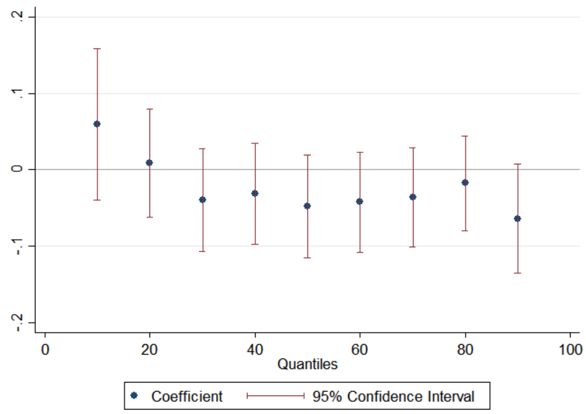
(c) Adding demographic and industry characteristics



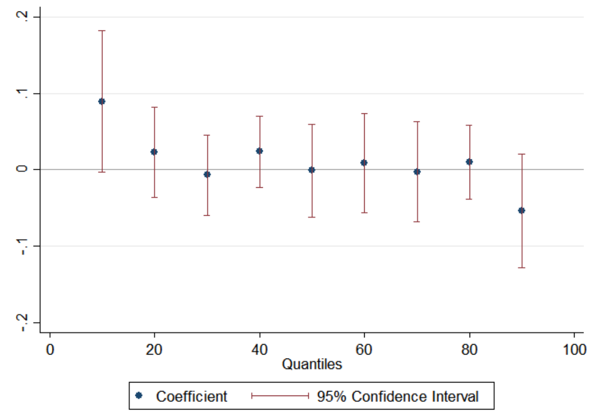
(d) Adding Chinese imports and routineness

Notes: The figures show OLS estimates of $\beta_1(\tau)$ in the second step, which indicates the effect of automation on the change in residual log annual earnings at each quantile. Here, automation is measured by the 30th percentile of exposure to robots in European countries. Figure (a) shows the results without any other group-level variables. Figure (b) is the results when adding only a dummy variable for Census division. Figure (c) presents the results when additionally including demographic and industry characteristics of CZs in 1990, and Figure (d) shows the results of the full specification by additionally controlling for Chinese import competition (using its IV) and the share of employment in routine-intensive jobs in 1990 (from Autor and Dorn, 2013). All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

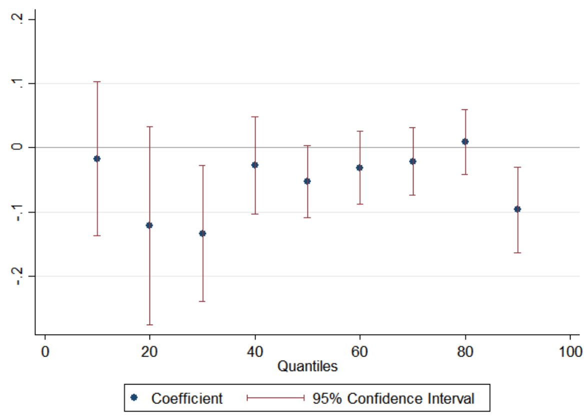
Figure 2.5 The effect of exposure to industrial robots on the distribution of annual earnings (OLS estimates)



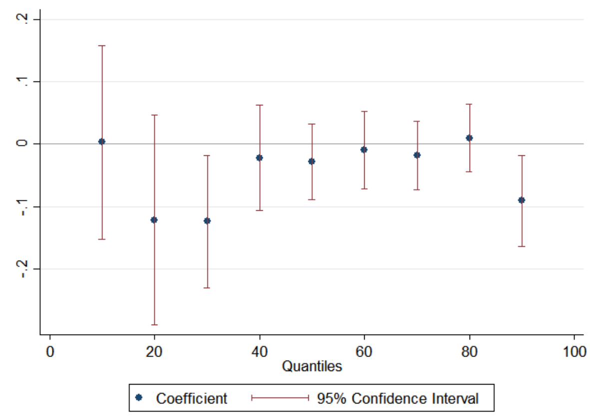
(a) With no controls



(b) Adding census division



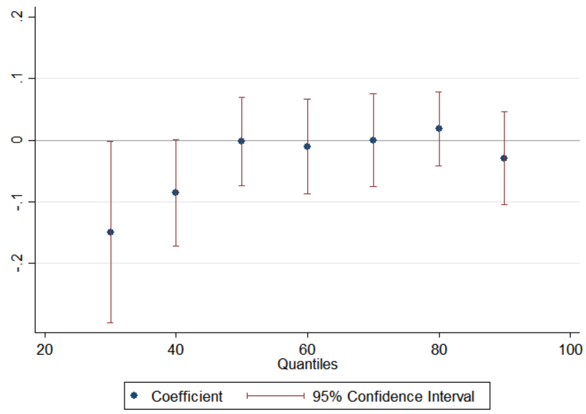
(c) Adding demographic and industry characteristics



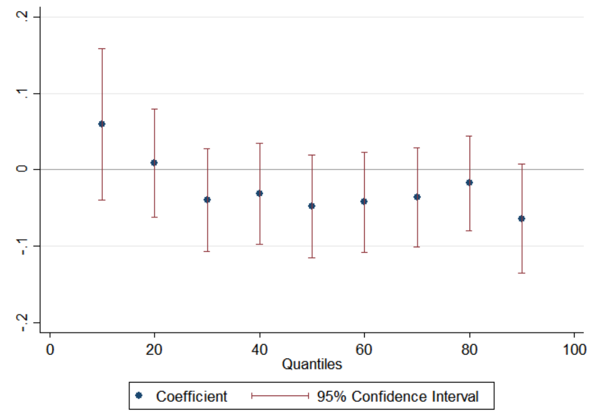
(d) Adding Chinese imports and routineness

Notes: The figures show 2SLS estimates of $\beta_1(\tau)$ in the second step, which indicates the effect of automation on the change in residual log annual earnings at each quantile. Here, the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. Figure (a) shows the results without any other group-level variables. Figure (b) is the results when adding only a dummy variable for Census division. Figure (c) presents the results when additionally including demographic and industry characteristics of CZs in 1990, and Figure (d) shows the results of the full specification by additionally controlling for Chinese import competition (using its IV) and the share of employment in routine-intensive jobs in 1990 (from Autor and Dorn, 2013). All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

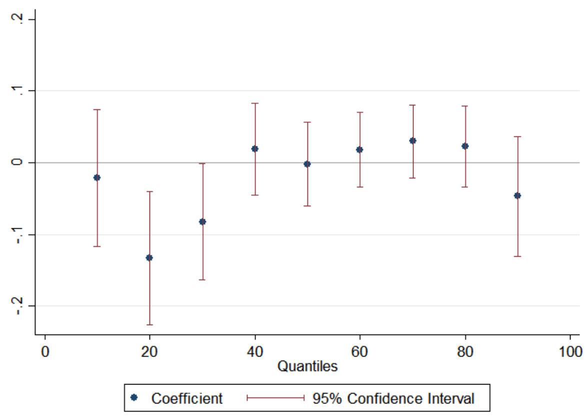
Figure 2.6 The effect of exposure to industrial robots on the distribution of annual earnings (2SLS estimates)



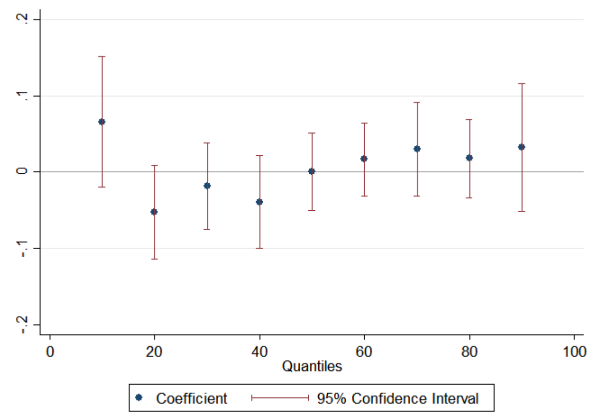
(a) Change in the log of total annual income



(b) Change in the log of annual earnings



(c) Change in the log of weekly earnings



(d) Change in the log of hourly earnings

Notes: The figures show 2SLS estimates of $\beta_1(\tau)$ in equation (2.5), where the estimates indicate the effect of automation on the change in residual log income at each quantile for different types of income measures. Here, the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. Note that figure (a) shows the estimates above the 30th percentile of the distribution since most of the observations have zero income at the bottom distribution when we focus on the total annual income. All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

Figure 2.7 The effect of exposure to industrial robots on the income distribution by different income variables (2SLS estimates)

Note that Acemoglu and Restrepo (2020) also examine the distributional effect of the industrial robots based on the unconditional wage at each quantile. For the comparison with the results in this paper, Figure B.4 in the Appendix shows the OLS estimates of the reduced form for different types of income variables and Figure B.5 presents the IV estimates, where the updated measure for the exposure to industrial robots, provided in Acemoglu and Restrepo (2020), are used.[56] Although the updated version increases the precision of the estimates, Figure B.4 and Figure B.5 present the distributional effect of automation in a similar way. More importantly, the graph (d) in both figures suggests the importance of the inclusion of individual-level covariates in the first step since controlling for individual characteristics reduces the negative effect of automation on low-wage earners as well as increases the precision of the estimates across quantiles.

2.5.2 Additional results

So far, we focus on the distributional effect of automation on residual income after controlling individual characteristics that could affect the income level. In addition to this, by utilizing the flexibility of the general model (2.1) and (2.2), I can explore whether the effects of individual-level covariates on the income distribution vary with the extent of automation. For example, let assume that having a college degree or more has a positive impact on earnings across all quantiles. Then, this positive impact on earnings will be stronger in a local labor market with relatively higher automation if it increases the labor demand for more educated people or their productivity, and so on.

Figure 2.8 to Figure 2.10 presents the 2SLS estimates of equation (2.5), where the dependent variables are the change in the coefficient on individual-level covariates are obtained from quantile regressions of different income variables. To be specific, Figure 2.8 shows the results when the dependent variable is the coefficients on female ($\Delta\hat{\alpha}_{c,4}(\tau)$), and each graph in Figure 2.8 display how the gender effect on the different types of income measures changes depending on the exposure to industrial robots. As shown in Figure 2.2, the negative values of the coefficient on female become smaller across quantiles over time. However, the graph (a), (b), and (c) in Figure 2.8 show that automation negatively affects the change in the gender effect on income, especially at the lower quantiles of the income distribution. In other words, automation makes the negative effect of gender on income stronger at the bottom of the income distribution. Note that there is a negative effect on the return to being female for annual earnings, but not for hourly earnings, even

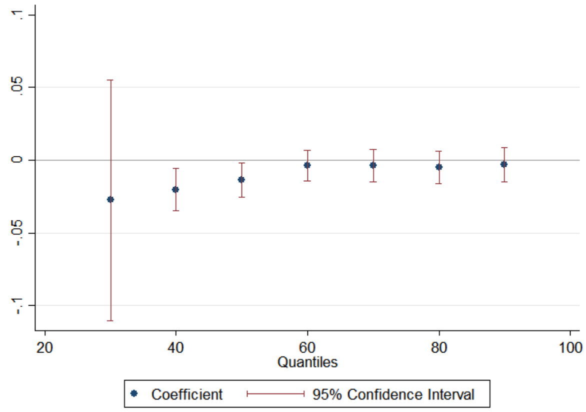
[56] Note that as mention in Section 2.4, the updated measure uses industrial robot adoption of five European countries that have a closer pattern to the U.S. and includes the industry's growth rate multiplied by the robot density in the initial period. Since the added term contributes little to the variation in the exposure to industrial robots, I here stick to the previous version. The data on the updated measure is available at https://economics.mit.edu/faculty/acemoglu/data.

at the lower quantile shown in Figure 2.8-(d). It suggests that the adverse effect of automation on females operates primarily through hours worked. Also, this result, in part, may come from the fact that women in married households usually tend to be secondary earners who are more likely to change hours worked in response to economic shocks.

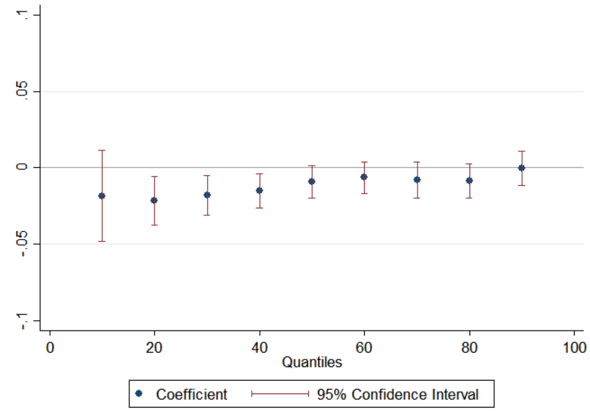
Next, Figure 2.9 plots the estimates when the dependent variable is the coefficients on white ($\Delta\hat{\alpha}_{c,6}(\tau)$), which shows that most estimates are statistically insignificant. Since Figure 2.3 suggests that whites' income is a little higher when other things being equal and Figure 2.7 also shows that the effect of automation on residual earnings is worse at the bottom of the distribution, you may have suspected that automation may have more harmed minority workers. Due to the lack of statistical power, Figure 2.9 indicate that the positive effect on the return to being white seems not to be affected by the extent of automation. However, small positive estimates--close to zero but not zero--suggest a possibility that minority workers in the middle-range of the distribution could be adversely affected by industrial robots.

Lastly, the results from the coefficients on education ($\Delta\hat{\alpha}_{c,7}(\tau)$) are presented in Figure 2.10. The coefficient on education estimates the effect of having a college degree or more on income, and it has positive values at all quantiles of the income distribution shown in Figure 2.4. Moreover, this positive effect of educational attainment seems to be a little larger above the 50th percentile of the income distribution over time. Figure 2.10 shows how automation affects the effect of educational attainment across the income distribution. Overall, all graphs in Figure 2.10 provide suggestive evidence that automation makes the effect of educational attainment on income stronger above the middle of the income distribution. It may reflect that automation, measured by industrial robots, is more likely to increase the college wage premium by raising the demand for more educated (or skilled) people, such as persons who can install, program, and maintain robots.[57]

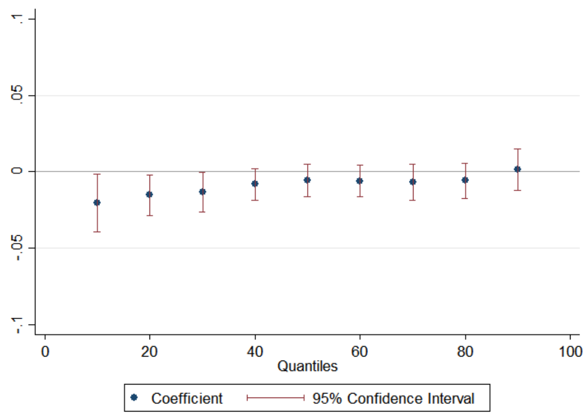
[57] Leigh and Kraft (2017) mentions that the adoption of robot technology needs "*both the science involved in creating robots and the competencies involved in optimizing their use*."



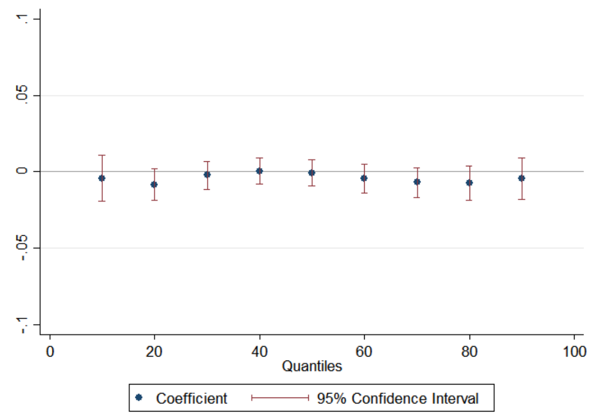
(a) Income variable in the 1st step:
log of total annual income



(b) Income variable in the 1st step:
log of annual earnings



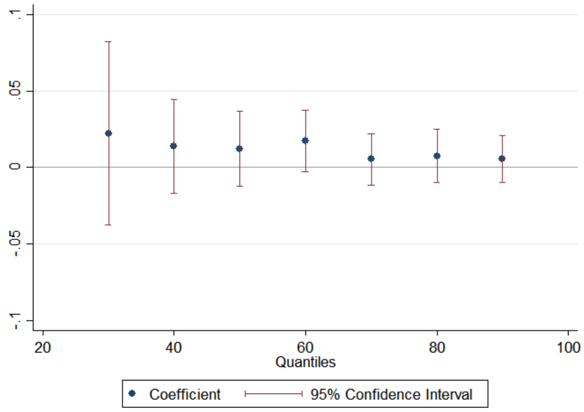
(c) Income variable in the 1st step:
log of weekly earnings



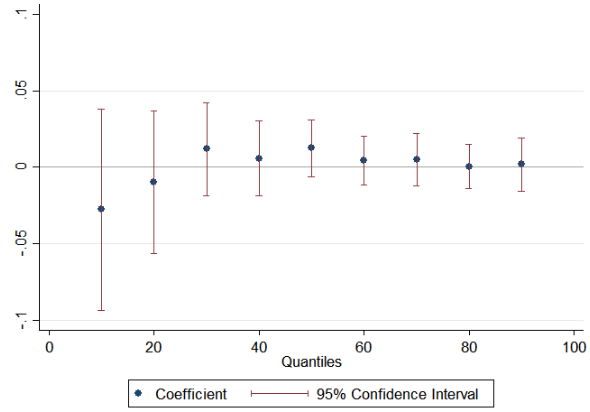
(d) Income variable in the 1st step:
log of hourly earnings

Notes: The figures show 2SLS estimates of $\beta_1(\tau)$ in equation (2.5), where the estimates indicate the impact of automation on the change in the gender effect on income (the coefficient on the female in the first step) at each quantile for different types of income measures. Here, the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. Note that figure (a) shows the estimates above the 30th percentile of the distribution since most of the observations have zero income at the bottom distribution when we focus on the total annual income. All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

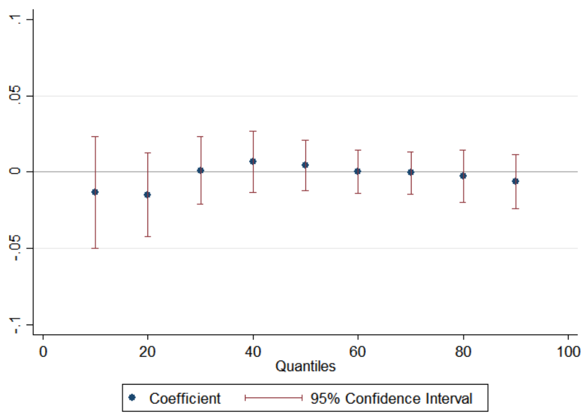
Figure 2.8 The effect of exposure to industrial robots on the coefficient on female across the income distribution by different income variables (2SLS estimates)



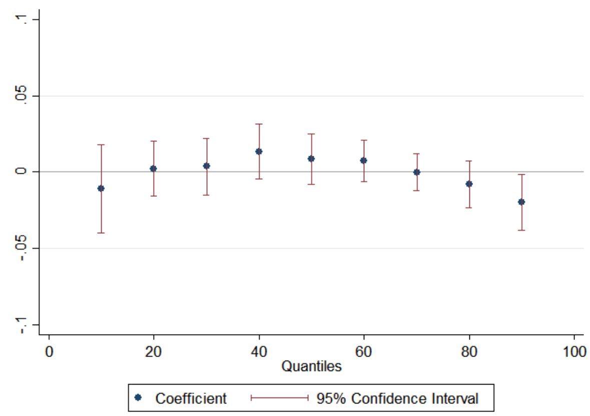
(a) Income variable in the 1st step: log of total annual income



(b) Income variable in the 1st step: log of annual earnings



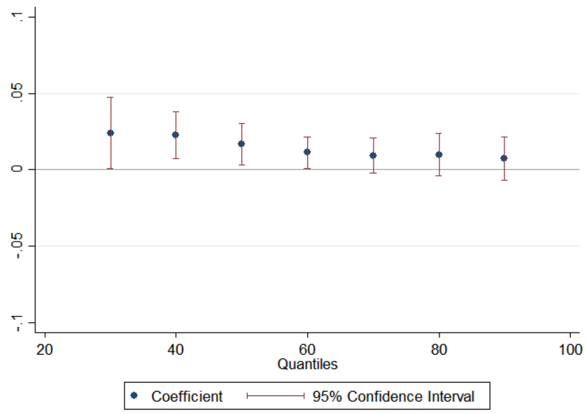
(c) Income variable in the 1st step: log of weekly earnings



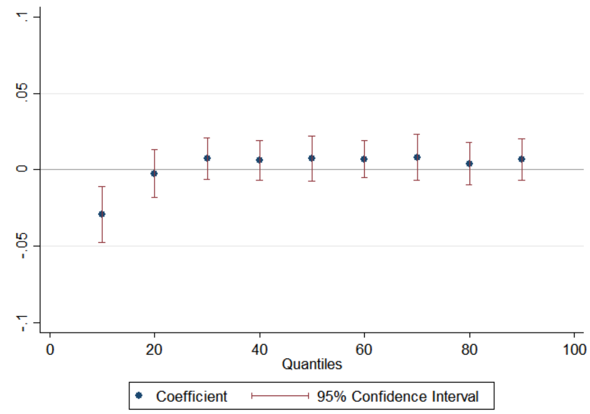
(d) Income variable in the 1st step: log of hourly earnings

Notes: The figures show 2SLS estimates of $\beta_1(\tau)$ in equation (2.5), where the estimates indicate the impact of automation on the change in the race effect on income (the coefficient on the white in the first step) at each quantile for different types of income measures. Here, the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. Note that figure (a) shows the estimates above the 30th percentile of the distribution since most of the observations have zero income at the bottom distribution when we focus on the total annual income. All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

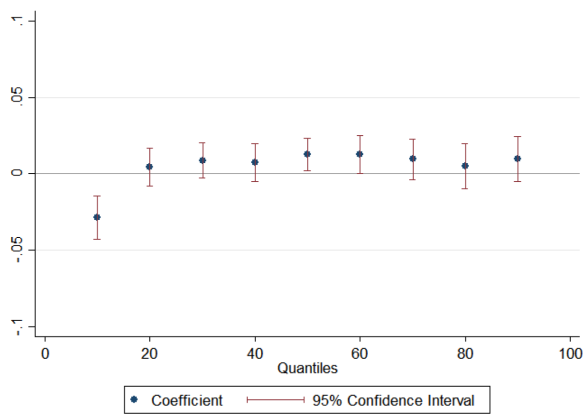
Figure 2.9 The effect of exposure to industrial robots on the coefficient on white across the income distribution by different income variables (2SLS estimates)



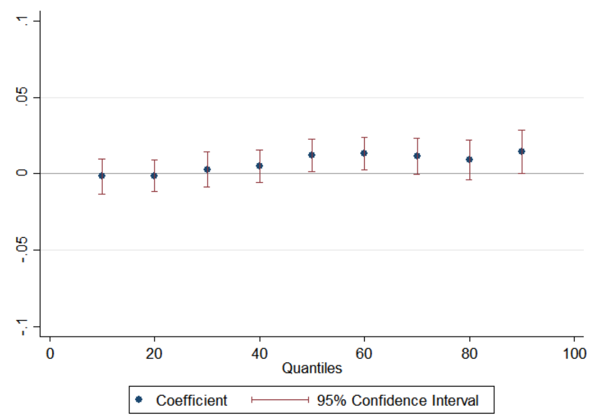
(a) Income variable in the 1st step:
log of total annual income



(b) Income variable in the 1st step:
log of annual earnings



(c) Income variable in the 1st step:
log of weekly earnings



(d) Income variable in the 1st step:
log of hourly earnings

Notes: The figures show 2SLS estimates of $\beta_1(\tau)$ in equation (2.5), where the estimates indicate the impact of automation on the change in the education effect on income (the coefficient on the college education in the first step) at each quantile for different types of income measures. Here, the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. Note that figure (a) shows the estimates above the 30th percentile of the distribution since most of the observations have zero income at the bottom distribution when we focus on the total annual income. All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

Figure 2.10 The effect of exposure to industrial robots on the coefficient on college education across the income distribution by different income variables (2SLS estimates)

2.6 Conclusion

Our historical experience and theoretical models in the literature suggest that the effect of new technologies on wage distribution can be affected by the feature of the technology, such as how it interacts with different types of labor. By using the industrial robot data from the IFR, as a measure for one of the recent technologies, and utilizing the quantile regression approach for a group-level treatment, developed by Chetverikov et. al. (2016), this paper contributes the literature that primarily focuses on the average effect of robots on wages. In addition to the distributional effect of robots on residual income, I also examine whether the effect of individual-level characteristics on income changes with the variation in the exposure to industrial robots, by exploiting the flexibility of the econometric model.

In the analysis, I use different income variables to study the effect of robots on wages as well as overall income, since automation can affect annual income or annual earnings through the change in hours worked and other income sources such as government supports. The results on annual earnings (and income) suggest that robots hurt low-income people below the 30th percentile of the distribution and overall no effect on people above the middle distribution. Also, the result on hourly earnings indicates that robots are much worse for low-earners since they could go through the decrease in wages as well as in hours worked, unlike middle- and high-earners. This suggests that industrial robots are likely to be the technology more affecting relatively low-skilled labor.

The additional results on the impact of individual-level characteristics present that robots negatively affect the return to being female for annual earnings, but not for hourly earnings. It could reflect that women are more likely to adjust hours worked responding to shocks. Also, the analysis shows that industrial robots could increase the college wage premium in the middle and the top of the distribution since it could raise the demand for more educated persons to program and manage robots.

Note that my results suggest the importance of the inclusion of individual-level characteristics, comparing the analysis based on the unconditional wage in Acemoglu and Restrepo (2020). But, my work has the vulnerability of automation measure, since the updated measure in Acemoglu and Restrepo (2020) seems to increase the precision of the estimates. Also, even though I examine the effect on automation on annual income, each quantile is based on the individual-level income. Given that the financial support from the government is usually based on the household (or family) level, the distributional effect of automation on annual income needs to be addressed at the household-level in further research.

CHAPTER 3

PERSISTENT POVERTY COUNTIES AND THE DISTRIBUTION OF ECONOMIC DEVELOPMENT FUNDING

3.1 Introduction

After the New Deal,[58] series of programs of the administration of President Franklin D. Roosevelt, poverty became a national issue again in the mid-1960s. President Lyndon B. Johnson declared a war on poverty with the Great Society initiative that aimed at the elimination of poverty. For example, the Economic Opportunity Act of 1964 created a lot of programs to fight against poverty such as the Job Crops, Head Start, and so on. Besides, the administration of President Johnson passed Appalachian Regional Development Act of 1965, the Food Stamp Act of 1964, Medicare Act of 1965, etc. The census data showed that the U.S. poverty rate declined from 19.5 percent in 1963 to 12.6 percent in 1970,[59] which is still lower than 13.5 percent, the poverty rate in 2015.

Since a huge drop in the poverty rate during the 1960s, it has been changed cyclically by increasing in recessions and decreasing during booms: the poverty rate is in the range of 11.1 to 15.2 percent between 1971 and 2015. In some different way this implies that the poverty rate does not fall below a certain level for three decades and that there are people living in chronic poverty. In addition, we can see that, geographically, several areas have had relatively high poverty rates for over thirty years. These so called persistent poverty counties are defined by the U.S. Department of Agriculture (USDA), which means a county is considered as a persistent poverty

[58] The main purpose of the program is referred to as “3Rs”, Relief, Recovery, and Reform: relief of economic distress for the unemployed and poor, recovery from the great recession, and reform of the financial system to prevent a crisis from happening again.

[59] Hence, poverty rate means the ratio of the number of people whose income falls below the federal poverty line to total population at the county level. Note that the official definition of poverty in the U.S. is based on family income, which means if the total income of a family is lower than the federal poverty line of that family size, then every person in that family is considered to be in poverty.

county if its poverty rates are greater than or equal to 20 percent in each decennial censuses since 1970. Most of these counties are concentrated in the southeast region, more specifically Appalachia, the Mississippi Delta, Rio Grande Valley, and the southern black belt area.

After the great recession in 2008, congressman Clyburn suggested the 10-20-30 rule to help persistent poverty counties, which was inserted in Title 1, Section 105 of the American Recovery and Reinvestment Act of 2009[60]:

> *SEC. 105. Of the amounts appropriated in this title to the "Rural Housing Service, Rural Community Facilities Program Account", the "Rural Business-Cooperative Service, Rural Business Program Account", and the "Rural Utilities Service, Rural Water and Waste Disposal Program Account", at least 10 percent shall be allocated for assistance in persistent poverty counties: Provided, That for the purposes of this section, the term "persistent poverty counties" means any county that has had 20 percent or more of its population living in poverty over the past 30 years, as measured by the 1980, 1990, and 2000 decennial censuses.*

In other words, this rule requires at least 10 percent of federal funding for rural development to be directed to counties that have experienced chronic poverty. He believes that these counties still have insufficient infrastructure so that they can't overcome poverty even though they have the will to do. As the representative of South Carolina 6th District, he mentions in his essay that nearly half of counties in his congressional district are in persistent poverty and persons living in these area want him to recruit business and industries to relieve the lack of opportunities. However, he says that it is very difficult because of the lack of basic infrastructure in these areas, for example "portable water, sufficient sewage, adequate roads, and appropriate bridges (Clyburn 2014, p.9)". Actually, Islam *et al*. (2015) examines reasons why certain counties have been persistently poor and shows that about 80 percent of the overall income difference between persistently poor and non-poor counties can be explained by lower levels of production factors. That is, some counties

[60] Congressman Clyburn proposed the 10-20-30 Act in 2014 that applied to other programs in addition to the above three accounts, but it was not passed. After that, several bills including a provision of the 10-20-30 rule had been introduced by the 114th Congress. Although these bills were not enacted into law, it shows Congress's interest in poverty issues. These bills are as follows: H.R. 1360 (America's FOCUS Act of 2015), H.R. 5393 (Commerce, Justice, Science, and Related Agencies Appropriations Act, 2017), H.R. 5054 (Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Appropriations Act, 2017), H.R. 5538 (Department of the Interior, Environment, and Related Agencies Appropriations Act, 2017), and S. 3067 and H.R. 5485 (Financial Services and General Government Appropriations Act, 2017).

are in chronic poverty because they have lower production endowments rather than using the factors less efficiently.

The ARRA of 2009 appropriated a total of 1.66 billion for the three accounts: 130 million for rural community facilities program account, 150 million for rural business program account, and 1.38 billion for rural water and waste disposal program account. As mentioned in Dalaker (2017), the 10-20-30 rule did not increase spending for three rural development accounts, but it seems mainly to affect a way to invest the existing funds. However, it is not clear whether or not these federal funding was actually allocated according to economic conditions within persistently poor counties. In fact, I find that, based on our data, persistent poverty counties already have received a little more than 10 percent of funding from these three accounts before the 10-20-30 rule. Besides, almost 80 percent of three accounts have been allocated to non-persistent poverty counties. So, I try to check whether the funding from these three account is assigned by economic conditions within persistent poverty counties. On the other hand, according to GAO report, Congress appropriated $6.2 billion in fiscal year 2010 for community and economic development programs as part of developing and helping economically distressed areas (GAO, 2012). This report studies whether federal economic development grant funding was distributed in accordance with local economic condition by focusing on cities and non-metro counties, separately. It shows that per capita grant funding distribution for population in poverty (or unemployed population) in both cities and non-metro counties was not consistent with either poverty rates or unemployment rates.

To sum up, this paper examines determinants of distribution of federal funding at the county level by focusing on several economic development programs that are more closely related to improve regional infrastructure. The choice of economic development programs used in this study is based on 80 programs suggested by GAO (2011). The selected programs are operated by two agencies, the U.S. department of agriculture and the U.S. department of housing and urban development. According to GAO (2012), the amount of fund appropriated to HUD in the fiscal year 2010 consists of about 67 percent of total appropriation on 80 economic development programs. Also, the amount of fund allocated to USDA accounts for about 18 percent. So, I would like to concentrate on programs operated by these two agencies to find out what factors determine the distribution of federal funds for economic development at the county level. More specifically, by comparing determinants of federal funding between persistent poverty counties and non-persistent poverty counties, I try to examine whether or not factors related to economic need play a role in the distribution of funding and whether or not there is any kind of structural difference in the allocation of funding.

3.2 Previous literature

GAO (2012) examined the distribution of economic development grant funding separated by cities or non-metro (rural) counties. It used federal economic development grant funds that was obligated to each region in fiscal year 2010, and compared grant funding per person in poverty (or per unemployed person) by poverty rates (or unemployment rates). This report showed that the distribution of grant funding was not consistent with either poverty rates or unemployment rates in both regions. For example, rural counties with the lowest poverty rates received more money than counties with higher poverty rates. Except for counties with the highest unemployment rates, rural counties with lower unemployment rates also received more funding than counties with relatively high unemployment rates. However, this report just provided the analysis with descriptive statistics.

Before moving to next topic, Dewar (1998) suggested that political aspects should be considered as one explanation for ineffectiveness of economic development programs. She argued that funding of economic development program was not distributed in ways that can achieve its economic development outcome by reviewing a single program, Minnesota's economic recovery fund which purpose was to stimulate economic development and increase employment opportunities. In her previous papers, she showed that the distribution of funds weakly correlated with the degree of economic distress, but more thriving counties received more funds when comparing more distressed counties. As the political economy perspective, she argued that an economic development program could be operated to consolidate re-election and to seek budgetary security. Since the changing political environment made administrators be more responsive to elected officials' desires and administrators of the program could not know how long their program would last, they would like to run the program to make sure its survival and to guarantee its budget.

Actually, there are a lot of studies showing that political concerns play an important role in determining the distribution of federal spending as well as economic conditions do. With regard to political variables, previous literature showed that politicians can affect federal fiscal decisions for their constituents (Chubb, 1985), that politicians can treat electively competitive areas differently to get favorable results in election or to compensate for their constituents' support, that areas with high voter turnout may receive more funds because of a reelection strategy of politicians (Fleck, 1999), and that party affiliation also can have influence in federal grant distribution (Levitt and Snyder, 1995).

More specifically, Levitt and Snyder (1995) examined the effect of political parties on the distribution of spending on federal assistance programs for the period of 1984 to 1990 by using district level data. They mentioned three possible models about the effect of political parties on federal funding allocation. First, the standard view is that distributive politics is decentralized so that districts whose representatives have the position in committee chairs might get more money, second, as the strong party model, it says that parties can act as a unitary organization to maximize reelection goals, and third model considers parties as groups of politicians who can cooperate for some common goals, but it has limitation due to their conflict preferences, which implies that the majority party in Congress do or don't have influence on the distribution of spending. After controlling several demographic and socio-economic variables, they found that funding increased as the number of Democratic voters in a district rose, but the party affiliation of a district's representative had no significant effect on the distribution of federal funds.

As a more recent study, Berry *et al*. (2010) argued that previous literature on distributive politics understated the effect of the president on the distribution of federal outlays by heavily focusing on the roles of committee and majority parties. They suggested that the president could affect appropriation process both ex ante and ex post because not only can the president exert formal proposal authority over the budget, but also the president can control agency administrators who distribute federal funds. Specifically, each agency should submit detailed reports to the Office of Management and Budget (OMB) who can return them with amendment or correct them itself to adjust discrepancies in favor of the president's policy priorities. Also, the president can reprogram or transfer funds with Congress's approval so that he can affect the allocation of spending ex post. To reflect various political economic concerns, they used dummy variables indicating whether the representative's party was the same with the president's party and the majority party, an indicator for the first terms of representatives, and the president's vote margin, etc. Their empirical analysis was implemented at both congressional district and county levels for the period of 1984 to 2007. In the case of county-level analysis, they used counties with only one congressional district. The results showed that both districts and counties whose representative was in the same party with the president received more funds. Also, the more narrowly the representative was elected, the more federal funding districts or counties received.

Deemer (2015) examined the determinants of federal spending distribution by concentrating on TANF (Temporary Assistance to Needy Families) and place-based federal program outlays, separately. She argued that place-based development programs could be an important strategy to develop the economy (also see Markusen and Galsmeier, 2008), but it also could exacerbate geographical inequalities in development. This is because she found that place-based federal investment had a positive impact on income level of underdeveloped areas (Isserman and Rephann, 1995; Mencken, 2000; Mencken and Tolbert, 2005), but other studies using descriptive analyses showed that place-based federal funding was not distributed evenly across counties. Therefore, she tried to figure out the relationship of distribution of place-based federal spending with socioeconomic, political, institutional, and spatial characteristics at the county level.[61] The result from OLS with the state-fixed effects showed that republican party strength, measured by counties' share of total votes cast for the Republican presidential candidate, had a statistically significant and negative effect on the average funding per capita during 2003-05, and county interparty competition, measured by the voting margin, had also a negative effect. Bureaucratic capacity, measured by the ratio of federal and local government employment to total county workforce, had a statistically significant and positive impact on the mean of federal funding. On the other hand, both median family income and unemployment rate were not significant under a specification including the past expenditures.

On the other hand, the literature studying the political determinants of federal spending at the state level focused on the role of U.S. senate because each state has two senators, but the number of representatives is assigned on the basis of population. Therefore, per capita representation in the house can be adjusted, whereas per capita representation in the senate varies with population, which means that states with smaller population are overrepresented in the Senate. Atlas *et al*. (1995) introduced this effect of overrepresentation by small states by showing that states with higher per capita senate representation received more per capita net spending. After Atlas *et al*. (1995), Hoover and Pecorino (2005) reexamined this effect by using more recent and

[61] This paper was written from a sociological point of view, so hence she said that "the political economy approach situates class, racial, and electoral struggles for power at the heart of social forces driving the formation of social policy while the institutional approach asserts that institutions independently shape social policy as well as structure and constrain political-economic forces" (Deemer, 2015, p.26).

disaggregated data. They confirmed that there was a positive relationship between per capita senate representation and federal spending, and that this effect was the strongest for procurement spending when they divided total expenditures into five categories such as procurement, grants, wages, retirement, and other. However, Douglas and Reed (2016), which replicated the analysis of Hoover and Pecorino (2005), showed that this small state effect disappeared when they (i) used cluster robust standard errors, and (ii) included a variable indicating population growth.[62]

Unlike the above literature, Hall (2010) analyzed that rurality or economic need in local areas affected the distribution of federal economic development grand funds by focusing on 120 counties of Kentucky for the period of 1993 to 2003. While previous studies concentrated the roles of politics and local administrative capacity to explain the inequity of distribution of federal grants, he focused on the roles of rurality and local economic need after controlling government capacity and political variables. He considered "capacity" as administrative and financial capacity which made it much easier to write grant application and to provide matching funds. To represent these kind of capacities, he used local administrative employment, local government revenues, etc. With regard to politics, he used fraction of county voters registered Republican, a dummy variable indicating whether at least one county representative is in the same party of the U.S. House, a dummy variable representing whether a county has multiple congressional districts, and so on. According to his empirical analysis, rurality increased grant funding per capita, whereas economic need did not necessarily increase grant funding per capita. Hence, economic need was measured by KEPI (Kentucky Economic Progress Index) which indicated a county's economic progress relative to the whole state, and it had a statistically significant and positive coefficient. This result implied that federal grant funding was not targeted to much needier counties. He argued that it reflected either that place-based and people-based targeting might be used simultaneously when considering the allocation of funding or that grant funding might be targeted to rapidly developing counties to get the highest returns on investment.

[62] See Larcinese *et al.* (2013).

3.3 Data and descriptive statistics

3.3.1 Persistent poverty counties

In this paper, I define a persistent poverty county by following the definition of USDA. In other words, I determine a county as a persistent poverty county when its poverty rates are 20 percent or more in each of the 1970, 1980, 1990, and 2000 decennial censuses. Although the 10-20-30 rule in ARRA defined persistent poverty counties by using three decennial censuses since 1980, almost 97 percent of these poverty counties (13 out of 425) also had poverty rates higher than 20 percent in the 1970 decennial census data. In addition, The Decennial Census of Population and Housing was the only source of poverty data at county level before the mid-1990s, and it stops collecting data on income after the Census 2000 so that it can't be used to calculate poverty rates since 2000. There are two data sets available to compute poverty rates in the time period after 2000: The American Community Survey (ACS) and the Small Area Income and Poverty Estimates (SAIPE). While the former includes other topics besides poverty, the latter has only information on poverty. Since the number of persistent poverty counties can change depending on different data, I use only decennial censuses to determine whether a county is persistently poor or not. In fact, about 96 percent of persistent poverty counties based on the period of 1980 to 2000 (16 out of 425) also had poverty rates greater than 20 percent in the 2010 SAIPE (the Small Area Income and Poverty Estimates).

Another issue is that the number of persistent counties can be different depending on whether poverty rates are rounded to zero or one decimal places. If we round off to the nearest whole number, counties which poverty rates are greater than or equal to 19.5 percent will be included in persistent poverty counties. In the latter case, persistent poverty counties will consist of from counties with 19.99 percent poverty rates. In this paper, I focus on the former one, and Table 3.1 shows how the number of persistent poverty counties change when we use different numbers as the criterion for 20 percent. The total number of counties is 3,101 that excludes counties in Alaska, Hawaii, and the District of Columbia, and 13.7 percent of total counties correspond to persistent poverty counties.

	rounded to zero decimal place	rounded to one decimal place
The 1970, 1980, 1990, and 2000 decennial censuses	425	382

Table 3.1 The number of persistent poverty counties

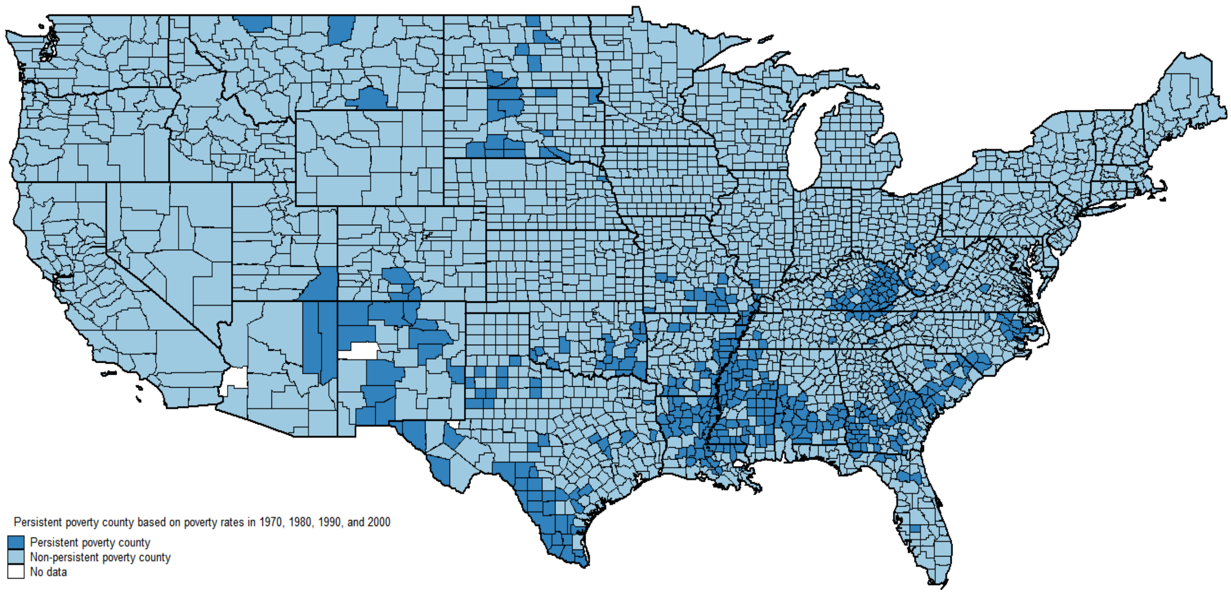


Figure 3.1 Persistent poverty counties during 1970-2000

Before moving to next variables, let's look at characteristics of persistent poverty counties by using the 2000 decennial census (Table 3.2). With regard to demographic and racial characteristics, persistent poverty counties tend to have small population size, but more Hispanic or Latino persons. Also, there is the difference in education attainment between persistently poor counties and non-poor counties. The fraction of people with less than high school is much higher in persistently poor counties. According to employment variables, less people participate in the labor market and are employed in persistent poverty counties. In contrast, the unemployment rate and the fraction of people employed in state or local government are higher in persistently poor counties. The last set of variables is related to household characteristics, which shows that persistent poor counties tend to have female-headed household more, lower median household income and mean earnings. Besides, except for retirement income, households living in persistently poor counties receive more income support: the fractions of households with social security income, supplemental security income, and public assistance income are higher.

	Non-persistent poverty		Persistent poverty	
	Mean	SD	Mean	SD
Demographics/race				
Total population	99,115	313,397	32,174***	85,631
Fraction female	0.504	0.018	0.509***	0.024
Fraction under age 17	0.252	0.03	0.271***	0.039
Fraction age 18 to 64	0.598	0.04	0.590***	0.039
Fraction over age 65	0.150	0.042	0.139***	0.034
Fraction Hispanic or Latino	0.056	0.095	0.100***	0.219
Fraction White (neither Hispanic nor Latino)	0.852	0.145	0.591***	0.251
Education attainment (25 years and over)				
Fraction less than high school	0.208	0.073	0.346***	0.073
Fraction high school graduate	0.349	0.067	0.333***	0.051
Fraction some college or associate degree	0.270	0.053	0.205***	0.045
Fraction Bachelor's degree or higher	0.173	0.079	0.116***	0.049
Employment				
Labor force participation	62.2	6.3	52.9***	5.7
Employment rate	58.6	6.7	48.1***	5.7
Unemployment rate	5.3	2.2	8.9***	3.3
Fraction in the 1st industries [a]	0.071	0.079	0.083***	0.058
Fraction in manufacturing	0.159	0.091	0.163	0.090
Fraction government workers	0.160	0.054	0.208***	0.066
Household characteristics				
Fraction married	0.568	0.057	0.509***	0.070
Fraction female headed	0.096	0.030	0.160***	0.054
Median household income	36,818	8,358	25,342***	3,537
Fraction with earnings	0.779	0.056	0.718***	0.056
Mean earnings	45,436	9,735	36,408***	4,113
Fraction with social security income [b]	0.304	0.064	0.319***	0.055
Mean social security income	10,941	876	9,300***	754
Fraction with supplemental security income	0.044	0.018	0.096***	0.033
Mean supplemental security income	6,086	827	5,605***	464
Fraction with public assistance income [c]	0.03	0.014	0.057***	0.029
Mean public assistance income	2,316	814	2,078***	602
Fraction with retirement income	0.171	0.044	0.153***	0.034
Mean retirement income	15,696	4,023	14,672***	4,146
Population density per square mile	254.7	1,682.4	142.9	1,567.2
Fraction non-metro counties (1993)	0.700		0.934***	
Total sum of population	258,259,662		13,185,394	
Total sum of Population in poverty	30,114,260		3,478,465	
Poverty rate	11.7		26.4	
Total number of counties	2,676		425	

a. 1st industries include Agriculture, forestry, fishing and hunting, and mining.
b. It does not include Medicare reimbursements
c. Public assistance income includes general assistance and TANF.
Note: All variables come from the 2000 decennial census data.

Table 3.2 The characteristics of persistent poverty counties

3.3.2 Economic development funding

A. The background of economic development programs

GAO (2011) provides a list of selected economic development programs that are operated under four agencies, Department of Commerce (USDC), Department of Agriculture (USDA), Department of Housing and Urban Development (USHUD), and Small Business Administration (SBA). In this paper, I concentrate on programs operated by USDA and USHUD because (i) three accounts of USDA rural development are explicitly mentioned in ARRA of 2009 to make federal funds spent to persistent poverty counties on the basis of total sum of appropriations from them, and (ii) Congressman Clyburn proposed the 10-20-30 Act of 2014 including several programs, and one of them is a program under USHUD, especially the Office of Community Planning and Development[63]. Besides, GAO (2011) provides appropriation information of the selected 80 programs in the fiscal year 2010, which shows that the total appropriation of 14 programs of USHUD consists of about 67 percent of the appropriation of 80 programs.

■ Three rural development accounts of USDA

Many programs for rural economic development have been implemented by different federal agencies, but half of those are administered by USDA (Cowan, 2016). There are three agencies for rural development in USDA: The Rural Housing Service (RHS), the Rural Business-Cooperative Service (RBS), and the Rural Utilities Service (RUS). The 10-20-30 provision in ARRA requires that at least 10 percent of the amounts appropriated in three rural development accounts of USDA will be allocated to persistent poverty counties. Hence, three rural development accounts are as follows: Rural Community Facilities Program Account of Rural Housing Service, Rural Business

[63] The 10-20-30 Act of 2014 required that at least 10 percent of appropriations from each of the following accounts shall be allocated to persistent poverty counties in each of fiscal years 2015 through 2025:

(1) “Department of Agriculture, Rural Development Programs”.
(2) “Department of Commerce, Economic Development Administration, Economic Development Assistance Programs”.
(3) “Department of Commerce, National Institute of Standards and Technology, Construction”.
(4) “Department of Education, Fund for the Improvement of Education”.
(5) “Department of Education, Fund for the Improvement of Postsecondary Education”.
(6) “Department of Labor, Employment and Training Administration, Training and Employment Services”.
(7) “Department of Health and Human Services, Health Resources and Services Administration”.
(8) “Department of Housing and Urban Development, Economic Development Initiative”.
(9) “Department of Justice, Office of Justice Programs”.
(10) “Environmental Protection Agency, State and Tribal Assistance Grants, Water and Wastewater”.
(11) “Department of Transportation, Federal Highway Administration, Transportation Community and System Preservation”.
(12) “Department of the Treasury, Community Development Financial Institutions”.

Program Account of Rural Business-Cooperative Service, and Rural Water and Waste Disposal Program Account of Rural Utilities Service.

Let's look at the basic information on three accounts[64]. First, RHS implements various programs for individual homeownership and rental housing. In addition to these programs, the Rural Community Facilities Program Account, which was designed by the 10-20-30 rule, has been used for "essential community facilities" in rural areas with a population of 20,000 or less. Hence, essential facilities include fire stations, community centers, child care centers, medical clinics, and so on. Secondly, one of the major concerns regarding rural development policies has been related to employment. Before the mid-1950s, these policies aimed at jobs in primary industries such as agriculture, mining, fishing, etc., and after that, manufacturing jobs were the main target of employment policy in rural areas. However, these days, the manufacturing sector in rural areas is competing with developing countries so that rural areas are trying to improve the manufacturing sector with high skilled labor. So, RBS operates many programs to help local entrepreneurs. More specifically, the Business and Industry loans are available for business acquisition, construction, expansion, etc., the main purpose of the Rural Business Opportunity grants is to provide training and technical assistance, and the Rural Business Enterprise Grants are more focused on small and emerging business. Thirdly, RUS mostly aimed at providing or constructing essential rural infrastructure like electrification, telecommunications, and water infrastructure. In addition, recent infrastructure includes facilities for health service delivery and broadband telecommunications. Most programs under the Rural Water and Waste Disposal Program Account are made for construction and improvements of rural water systems (e.g. storage, treatment, purification, and so on).

■ Other development programs of USDA except for three accounts

The list of other rural development programs of USDA is summarized in Table 3.3. Hence, I exclude other programs of RHS because most of them are not explicitly related with economic development. For example, there are three programs of RHS, Very Low to Moderate Income Housing Loans, Very Low-Income Housing Repair Loans and Grants, Rural Rental Assistance Payments, which consist of almost 85 percent of other programs operated by the office of RHS.

[64] These contests are based on Cowan (2016).

The first one provides loans to help households get decent housing in rural areas, the second one also gives loans or grants to homeowners to help them repair their properties or remove health and safety hazards, and the last one is implemented in the form of direct payments to reduce the rents paid by low income families.

■ Other development programs of USHUD

The U.S. Department of Housing and Urban Development mainly concentrate on creating "strong, sustainable communities and quality affordable homes for all" [65]. Especially, the office of Community Planning and Development operates the program as well-known as Community Development Block Grant Programs (CDBG). It is basically in the form of formula grant and provides funds to states, cities, and counties to help them become urban communities by giving decent living environment and increasing economic opportunities for low and moderate income people. The block grant is a kind of hybrid grant because it should be spent within certain categories (hence, community or economic development), but has local discretion to distribute funds within categories (Hall, 2010).

In Table 3.3, I mark whether the type of grant is project or formula when programs include grants. The important difference among different grant programs is the degree of targeting versus distributional equity. Formula grants are supposed to be distributed on the basis of predetermined characteristics such as population so that it puts more weight on distributional equity. On the other hand, categorical or project grants are usually likely to be targeted to specific areas based on income, poverty, or other variables so that it aims at reducing the gap between rich and poor areas. Project grants are often competitive in receiving money and have stringent requirements about how the grants are allocated, but block grants, one type of formula grants, tend to give state or local government control of the use of grants (Hall, 2010). According to Schneider and Ji (1990), while Democrats prefer targeting grants, Republicans favor distributional equity and less federal discretion.

[65] https://portal.hud.gov/hudportal/HUD?src=/about/mission

Agency		Office	Program name	CFDA code	Grant type
USDA	Three accounts	RHS	Community facilities loans and grants	10.766 (10.780)	P
		RBS	Business and industry loans	10.768 (10.782)	-
			Rural business enterprise grants	10.769 (10.783)	P
			Rural business opportunity grants	10.773	P
		RUS	Water and waste disposal system for rural communities	10.760 (10.781)	P
			Technical assistance and training grants	10.761	P
			Solid water management grants	10.762	P
			Emergency community water assistance grants	10.763	P
			Water and waste disposal loan and grants (Section 306C)	10.770	P
			State bulk fuel revolving fund grants	10.857	P
			Assistance to high energy cost rural communities	10.859	P
			Household water well system grant program	10.862	P
	Other Programs	RBS	Intermediary relending program	10.767	-
			Rural cooperative development grants	10.771	P
			Empowerment zones program	10.772	P
			Renewable energy systems and energy efficiency improvement program	10.775	P
			Rural economic development loans and grants	10.854	P
			Bio-refinery assistance	10.865	-
			Rural energy for America program recovery	10.868	-
			Rural micro-entrepreneur assistance program	10.870	P
		RUS	Broadband initiatives program – recovery	10.787	P*
			Rural electrification loans and loan guarantees	10.850	-
			Rural telephone loans and loan guarantees	10.851	-
			Rural telephone bank loans	10.852	-
			Distant learning and telemedicine loans and grants	10.855	P
			Public television station digital transition grant program	10.861	P*
			Community connect grant program	10.863	P
			Grant program to establish a fund for financing water and wastewater projects	10.864	P
			Rural broadband access loans and loan guarantees	10.886	-

(continued)

Agency	Office	Program name	CFDA code	Grant type
USHUD	Community Planning and Development	Community development block grants(CDBG)/Entitlement grants	14.218 (14.253)	F
		CDBG/Small cities program	14.219	F
		CDBG/Secretary's discretionary fund-insular area	14.225 (14.254)	P
		CDBG/Secretary's discretionary fund/community development tech assist	14.227	P
		CDBG/State's program	14.228 (14.255)	F
		CDBG/secretary fund-special projects	14.232	P*
		CDBG/Economic development initiative	14.246	P
		EDI Special Projects	14.251	P

Note: CFDA (Catalog of Federal Domestic Assistance) is the basic reference of federal programs and helps researchers to identify each program. The CFDA number in parentheses means that it has the same program name, but is implemented as a part of Recovery Act.

Table 3.3 The selected economic development programs (loans/grants)

B. Data on economic development funding

Data on economic development funding summarized in Table 3.3 comes from the Consolidated Federal Funds Report (CFFR). I use data for the period of 1993 to 2010, but 11 years of data between 1993 and 2003 are collected from "the National Archives Catalog,[66] whereas data for the remaining period from 2004 to 2010 is available on the website of USDA's Economic Research Service (ERS)[67]. Note that ERS produces the federal funds data by applying its own screening method to calculate the amount of funding that each county receives accurately. It's because some funds include substantial subcontracting so that these kinds of funds would be redistributed to surrounding areas. Thus, ERS generate a variable which name is "disposition code" and funds with zero disposition code can be considered accurate at the county level. According to the documentation of ERS, even if it reported at the county level, the data with disposition code 1 are not considered accurate at the county level because 25 percent or more of the funds flowed to state capitols[68]. However, the problem is that ERS provides only 7 years of data. Since all programs in Table 3.3 have zero or one disposition code and the ratio of observations with one disposition code

[66] https://catalog.archives.gov/search?q=*:*&f.parentNaId=626196&f.level=fileUnit&sort=naIdSort%20asc

[67] https://www.ers.usda.gov/data-products/federal-funds

[68] See https://www.ers.usda.gov/data-products/federal-funds/documentation/

is less than 2 percent except for 2009, I decide to include observations with disposition code one. Then, based on the CFDA code of each program I construct the data set of economic development funds for the period of 1993 to 2003 by using CFFR data files.

The CFFR report, not raw data, provides dollar amounts under several categories,[69] retirement and disability, other direct payments, grants, procurement contracts, and salaries and wages at the county level. According to categories, CFFR 2010 report mentions that direct payments and salaries and wages reflect actual expenditures or outlays, whereas grants and procurement contracts represent “obligated funds”. In some case, the amount of federal funding has a negative value, which reflects de-obligations of funding that had been previously awarded. The types of funding that I can examine in this paper are loans and grants, and these funding are less closely related to actual spending. Thus, I replace the negative values of de-obligations with the positive values to reflect obligations in the first place (Deemer, 2015).

Table 3.4 shows the mean of total funding on economic development programs. First, the fraction of counties who don’t receive money have decreased in both persistent poverty counties and non-persistent poverty counties. Per capita funding is on average greater in persistently poor counties, which may happen because of relatively small population size in these counties. So, I also calculate the average amount of funding per person in poverty, which shows that persistently poor counties receive less money. Similarly, this is probably in part because percent of people in poverty is higher in these counties. The last column shows what percent of total funding is distributed to persistent poverty counties, which is less than 10 percent. But, as you can see in Appendix, it mainly stems from funding of HUD. In case of funding allocated to USDA, about 10 percent has been distributed to persistently poor counties in most years.

Figure 3.2 through Figure 3.4 represents geographical distribution of total funding in each year. White colors indicate counties who did not get money. Blue colors represent how much counties received funds, which is comprised of four ranges minimum to the first quartile, the first quartile to median, median to the third quartile, and the third quartile to maximum in each year. Black dots indicate persistent poverty counties. As time passed, more counties have received more funds, and this seems to occur regardless of whether a county is persistently poor.

Year	Non-persistent poverty counties	Persistent poverty counties

[69] According to the report, interest paid on federal government debt, travel expenses when not provided under procurement contract, and international payments and foreign aid were excluded. See U.S. Census Bureau (2011).

	Fraction not receiving	Funding per capita	Funding per persons in poverty	Fraction not receiving	Funding per capita	Funding per person in poverty	Percent of funding spent to PPC	
							total	3 accounts
1993	0.497	46	343	0.569	61	249	8.5	15.6
1994	0.485	44	326	0.539	42	150	6.5	13.6
1995	0.467	51	390	0.494	57	206	7.5	14.0
1996	0.478	50	389	0.520	59	211	7.1	15.3
1997	0.445	47	376	0.405	76	301	8.4	15.4
1998	0.414	68	524	0.447	63	265	7.4	14.0
1999	0.405	54	482	0.421	88	410	9.0	15.3
2000	0.423	78	735	0.424	83	381	8.7	14.1
2001	0.336	98	850	0.308	140	704	8.6	11.7
2002	0.322	112	944	0.325	115	522	7.7	11.7
2003	0.334	101	919	0.334	146	706	11.9	14.4
2004	0.352	109	942	0.384	107	494	6.6	12.7
2005	0.380	116	939	0.391	95	381	7.5	14.8
2006	0.315	90	724	0.313	70	287	2.4	10.6
2007	0.316	84	690	0.289	117	512	5.0	11.0
2008	0.304	92	741	0.315	82	339	6.2	9.4
2009	0.268	164	1339	0.280	155	635	6.1	12.2
2010	0.204	230	1793	0.224	258	1010	8.2	11.7

Table 3.4 Descriptive statistics of total amount of selected economic development programs

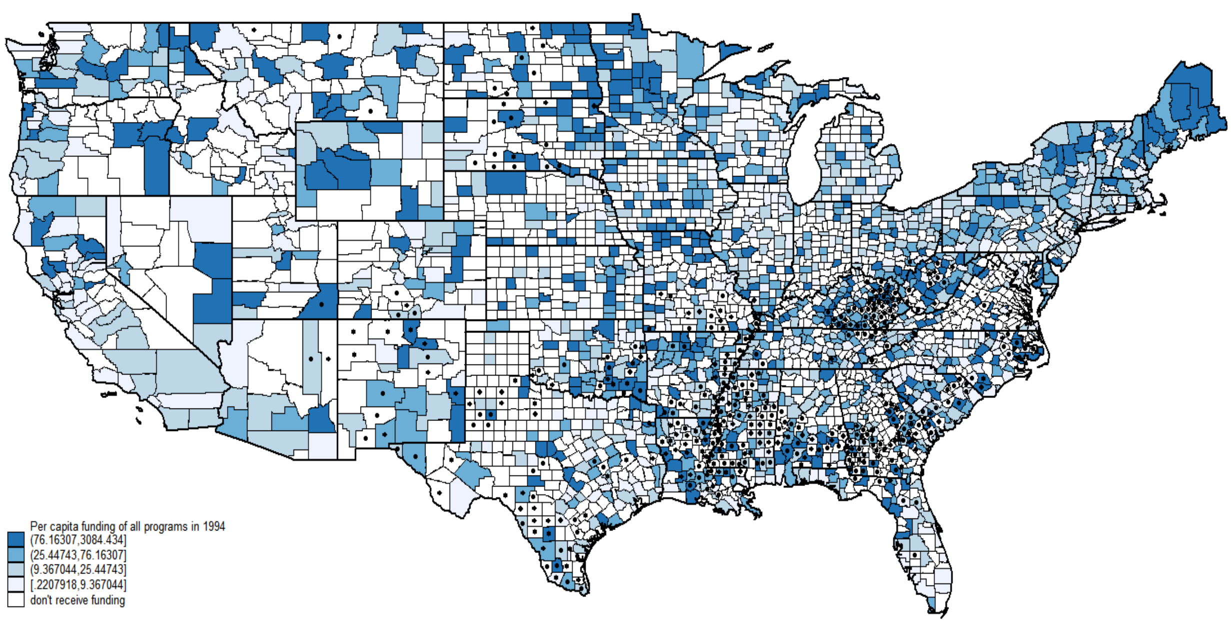


Figure 3.2 Per capita funding of selected all programs in 1994

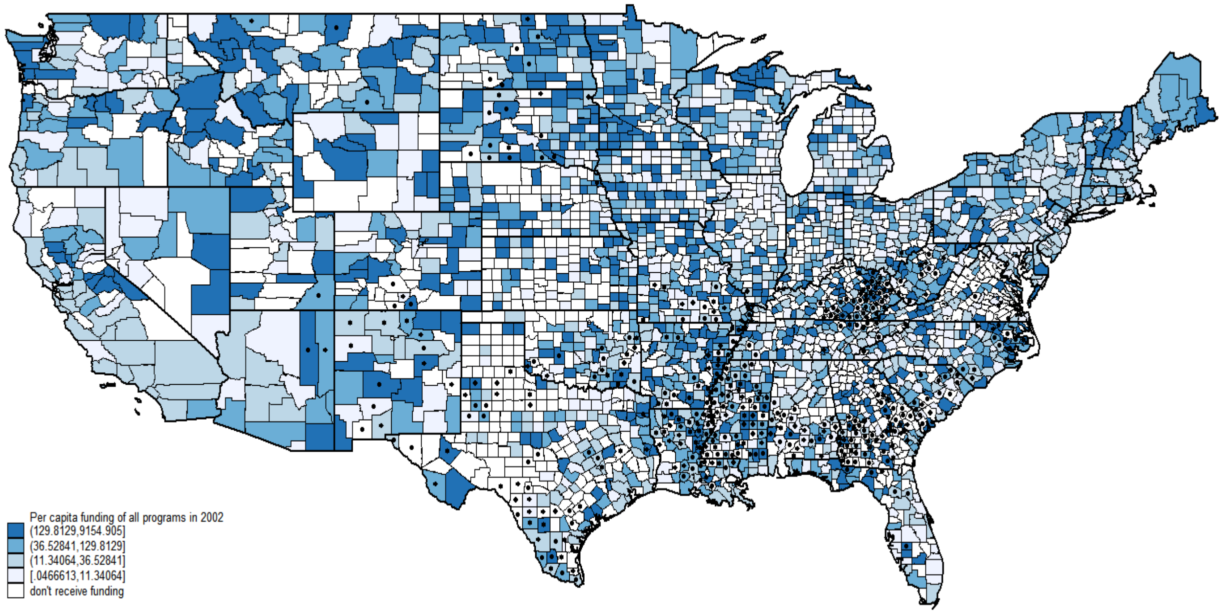


Figure 3.3 Per capita funding of selected all programs in 2002

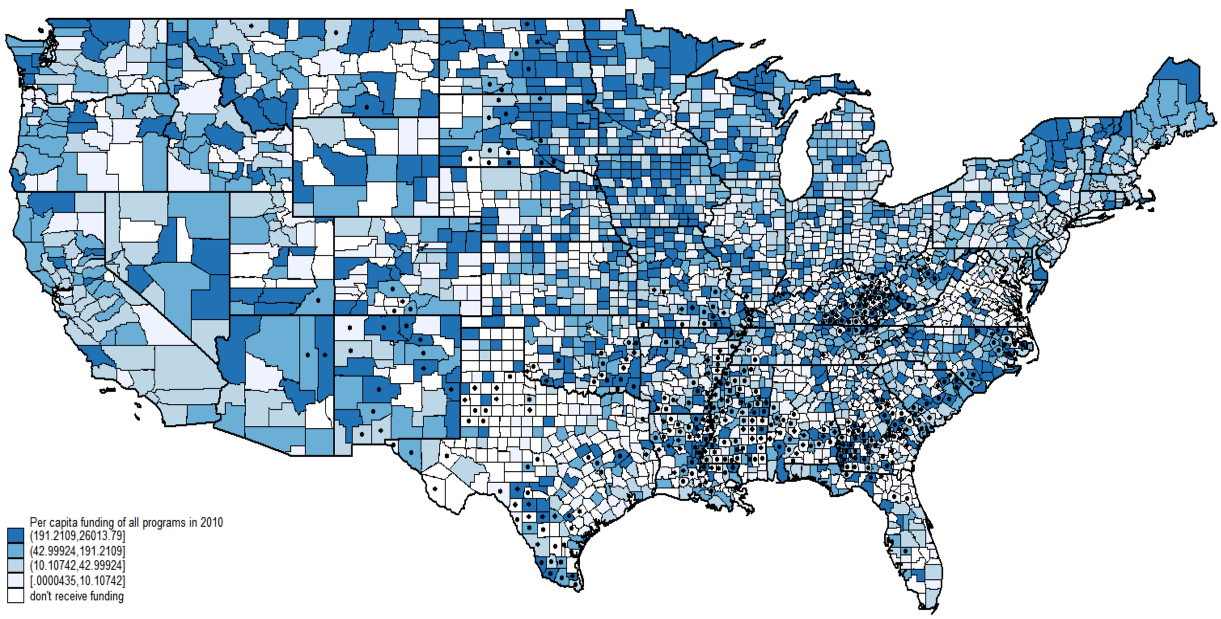


Figure 3.4 Per capita funding of selected all programs in 2010

3.3.3 Other variables

A. Variables representing economic conditions

This paper focuses on persistent poverty counties that are likely to have less population than non-persistent poverty counties. The decennial census survey collecting lots of characteristics of all counties in U.S. was stopped after 2000, and the American Community Survey provides 5-year estimates for all counties. The 5-year estimates of ACS are all period estimates that represent data over a 60-mont period so that the user guide recommend not to use overlapped period of data. Also, the ACS 5-year estimates are available from a relatively recent period, 2006-2010. Thus, in this paper, I use several basic variables representing economic conditions provided by Bureau of Economic Analysis (BEA). More specifically, the following variables are available: (i) per capita income, (ii) per capita income support that includes Supplemental security income (SSI) benefits, Earned Income Tax Credit (EITC), Supplemental Nutrition Assistance Program (SNAP), and other income maintenance benefits such as Temporary Assistance for Needy Families (TANF), the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), etc., (iii) each ratio of farm, manufacturing, and state and local government employment.

Other variable, unemployment rates, comes from Local Area Unemployment Statistics of Bureau of Labor Statistics (BLS). Finally, both poverty rates and median household income for the period of 1993 to 2010 are collected from the Small Area Income and Poverty Estimates (SAIPE). Note that the SAIPE does not provide county-level data in 1994 and 1996. So, for these years I use previous year data, which means that poverty related variables of 1994 (1996) have the same figures with data of 1993 (1995).

Table 3.5 shows descriptive statistics of first and last years of the whole sample periods. As you can expect, both per capita income and median household income are lower in persistent poverty counties, whereas unemployment rate and poverty rate are higher in persistently poor counties. Percent of farm employment, measured by the number of jobs, is higher in these counties in both years.

	Non-persistent poverty		Persistent poverty	
	Mean	SD	Mean	SD
1993				
Per capita income	18,106	3,686	13,759***	1,956
Per capita income support	278	123	610***	188
Percent of farm emp. (jobs)	10.3	9.9	12.5***	9.3
Percent of manufacturing emp. (jobs)	14.3	10.1	15.8**	11.0
Unemployment rate	6.5	3.0	9.6***	4.0
Poverty rate	14.4	4.8	29.0***	6.2
Median household income	28,996	6,407	19,536***	3,037
2010				
Per capita income	34,795	8,363	27,072***	4,194
Per capita income support	735	248	1,337***	378
Percent of farm emp. (jobs)	7.7	7.8	9.3***	7.2
Percent of manufacturing emp. (jobs)	9.0	6.5	8.3*	6.9
Unemployment rate	9.0	3.0	11.6***	3.2
Poverty rate	15.2	4.6	27.0***	5.6
Median household income	44,826	10,233	31,488***	3,806

Table 3.5 Descriptive statistics about economic conditions: 1993 and 2010

B. Political economy variables

With regard to political variables, I use several variables that can reflect elections and capacity of local government. Data on the president election results at the county level is collected from Atlas of U.S. President Elections. I construct data set with total votes cast, votes cast for the Republican presidential candidate, and votes cast for the Democrats presidential candidate from the following election years: 1992, 1996, 2000, 2004, 2008. Then, I calculate (i) voter participation rates which are divided by total population from BEA data,[70] (ii) percent of votes cast for Republican candidate to represent the degree of support for Republican, and (iii) voting margin between two parties to reflect competitive elections. The bureaucratic capacity is measured by the ratio of state and local employment to total employment, which are available on BEA data set. Note that employment of BEA data is measured by the number of jobs.

Since the U.S. presidential elections are held every four years, the above elections variables have the same values for 4 years after the election year. More important thing is that the elected

[70] Atlas of U.S. President Elections does not provide information on voting age population.

president actually exerts his power over the budget in the fiscal year after next. For example, the U.S. presidential election of 1992 was held on November, 1992. Bill Clinton who was elected from this election started his job from January 20, 1993. Then, budget proposals for the fiscal year 1994 usually was submitted on February or March, 1993. Since the funding of economic development programs is based on each fiscal year, I match the values of election variables in 1992 to those between 1994 and 1997, and so on. That is, election variables from 2002 to 2004 have the same figures with ones in 2000.

Table 3.6 represents descriptive statistics of political economy variables in 1992, 2000, and 2008. These years are chosen because the presidential party was changed as a result of elections. Overall, voter participation rates are lower in persistent poverty counties although they have increased over time. Also, you can see that percent of votes cast for Republican candidate has been increased over time, but in all three years, non-persistent poverty counties tend to support more Republican when comparing to persistent poverty counties. Percent of government employment is relatively high in persistently poor counties.

	Non-persistent poverty		Persistent poverty	
	Mean	SD	Mean	SD
Election year: 1992				
Voter participation rate	43.1	7.5	37.6***	7.1
Percent of votes cast for Rep.	40.2	8.2	37.0***	10.3
Voting margin (absolute value)	13.2	10.4	20.2***	14.4
Percent of government emp.	14.9	5.7	19.5***	7.5
Election year: 2000				
Voter participation rate	40.5	7.3	34.9***	6.6
Percent of votes cast for Rep.	58.0	11.4	50.4***	13.2
Voting margin (absolute value)	24.0	17.3	21.4***	15.9
Percent of government emp.	15.3	5.8	20.8***	8.4
Election year: 2008				
Voter participation rate	45.2	7.5	40.6***	8.4
Percent of votes cast for Rep.	56.0	14.1	52.0***	16.2
Voting margin (absolute value)	25.5	18.1	27.6**	17.8
Percent of government emp.	15.6	5.9	20.8***	8.1

Table 3.6 Descriptive statistics of political economy variables

3.4 Empirical analysis and results

Previous literature used different estimation strategy, that is, some studies implemented cross-sectional regressions, whereas other studies employed spatial regression models or panel data analysis with fixed effects. Deemer (2015) used cross-section data and run regressions by using spatial regression model and OLS with state-fixed effects. She showed that results from two types of regressions were very similar to each other.

Based on previous literature and purposes and eligibility of economic development programs, I specify the following model to examine determinants of funding distribution:

$$y_{ct} = (1 - d_c)\boldsymbol{X_{ct}\beta_1} + d_c\boldsymbol{X_{ct}\beta_2} + \alpha_c + \mu_t + d_c\mu_t + \varepsilon_{ct} \tag{3.1}$$

where subscript c denotes counties and t denotes time. Hence, α_c denotes county-fixed effects, μ_t is year dummies to control changes in federal funding over time, d_c denotes a dummy variable indicating whether a county is persistently poor or not, $\boldsymbol{X_{ct}}$ includes economic need variables and political economy variables, and y_{ct} is per capita funding or funding per person in poverty. Since d_c is a dummy variable representing persistent poverty counties, estimated $\boldsymbol{\beta_2}$ reflects how county-level characteristics affect the funding allocation in persistent poverty counties. So, by testing a hypothesis $\boldsymbol{\beta_1} = \boldsymbol{\beta_2}$, we can see whether there is a structural change in the distribution of funds between two types of counties.

Table 3.7 through Table 3.9 shows results when the dependent variable is per capita funding. First, Table 3.7 shows that signs of estimated coefficients change depending on whether time effects are included or not. In any case, political economy variables are not significant, so the distribution of funding looks like to be affected by economic conditions. However, it shows that non-persistent poverty counties with more distressed economic conditions receive less funds, and the funding distributed to persistent poverty counties is not determined by economic conditions when including year effects. Except for unemployment rate, there are no structural differences in federal funding distribution.

Table 3.8 shows regression results by agencies and Table 3.9 summarizes results according to types of grant by excluding loans. As you can see it in Table 3.3, only HUD operates programs which funding type is formula grants. In addition, almost 90 percent of HUD funding addressed in this paper comes from CDBG/Entitlement grants and CDBG/State's program so that results between Table 3.8 and Table 3.9 are quite similar to each other. More specifically, with regard to USDA funding or project grants, funds distributed to persistently poor counties are not affected by economic need when I include time effects. Among political variables, percent of government employment reflecting capacity is only significant and it has a positive impact on funding distribution. In case of HUD funding or formula grants, economic need variables cannot explain determinants of funding distribution. Of political variables, voting margin has a significant and positive effect, which is opposed to the general hypothesis that interparty competition is related to more generous social policies (Deemer, 2015). This result may imply that more political cooperation would be required to receive more economic development funds.

In appendix, Table B.7 through Table B.9 provides results that come from the same specification with former analysis, but now the dependent variable is defined as federal funding per person in poverty. Overall, the size of coefficients become larger than previous analysis, but the signs of significant coefficients are mostly the same. Table B.7 shows that non-persistent poverty counties still receive less funding as their unemployment rates and poverty rates are higher. Although the degree of accordance with poverty rate is much worse in non-persistently poor counties, the distribution of funding to persistent poverty counties is also not consistent with poverty rates. Besides, distribution of funding per person in poverty by agencies or the type of grants represent similar results. That is, funding operated by USDA or project grants are allocated according to economic need conditions even if they are negatively related to each other. On the other hand, distribution of funding operated by HUD or formula grants is just affected by political variables, especially voting margin.

Variables	Non-persistent poverty counties		Persistent poverty counties		Structural break	
Poverty rate (%)	-3.3377**	-2.7693*	-5.2973***	-2.186		
	(1.554)	(1.5612)	(1.5414)	(2.0698)		
Median household income ($)	0.0020***	-0.0029***	0.0022	0.0027		
	(0.0006)	(0.0009)	(0.0022)	(0.004)		
Per capita income support ($)	0.1647***	-0.0953*	0.1169***	0.0049		
	(0.0342)	(0.0499)	(0.0344)	(0.0449)		
Unemployment rate	4.0662***	-5.2918***	4.6530***	-0.1916		**
	(1.458)	(1.9977)	(1.5946)	(1.5075)		
Percent of farm emp.	-6.9652**	-4.9697*	-2.9618	-2.7507		
	(3.1835)	(2.8726)	(2.7052)	(2.7523)		
Percent of manufacturing emp.	1.9745**	2.3969***	1.5758	2.0001*		
	(0.9651)	(0.8995)	(1.1678)	(1.1664)		
Percent of government emp.	1.2729	0.2092	3.5981*	3.3704		
	(2.4151)	(2.3958)	(2.069)	(2.1442)		
Voter participation rate (%)	-1.3721	-0.5779	-0.9879	-0.2678		
	(1.0879)	(1.584)	(1.5672)	(1.8308)		
Percent of vote for Rep.	-0.6879	-1.7356	-0.4212	-0.7052		
	(0.9976)	(1.6169)	(0.8904)	(1.2651)		
Voting margin (%)	0.6043	0.5757	0.2203	0.1367		
	(0.402)	(0.4349)	(0.3293)	(0.3317)		
Constant	49.0931	316.6497*	21.6567	14.8972		
	(82.0919)	(163.8367)	(110.3989)	(153.9118)		
Year fixed effect	No	Yes	No	Yes	No	Yes
Num. of obs.	41,554		6,373			
Num. of groups	2,600		422			

* p<.1; ** p<.05; *** p<.01

Table 3.7 Determinants of per capita funding on economic development programs

Variables	USDA					HUD				
	Non-persistent poverty counties		Persistent poverty counties		Structural break	Non-persistent poverty counties		Persistent poverty counties		Structural break
Poverty rate (%)	-3.9493***	-2.9863**	-5.1549***	-2.1701		0.7384	0.2495	-0.0035	0.0199	
	(1.4424)	(1.3985)	(1.5414)	(2.0682)		(0.5833)	(0.6947)	(0.0566)	(0.0412)	
Median household income ($)	0.0018***	-0.0026***	0.0021	0.0026		0.0002*	-0.0003	0.0001	0.0002	
	(0.0006)	(0.0009)	(0.0022)	(0.004)		(0.0001)	(0.0003)	(0.0001)	(0.0002)	
Per capita income support ($)	0.1579***	-0.0989**	0.1170***	0.0046		0.0091	0.0046	0.0011	0.0018	
	(0.0303)	(0.0468)	(0.0344)	(0.0449)		(0.0156)	(0.0164)	(0.0015)	(0.002)	
Unemployment rate	4.7470***	-5.4465***	4.6926***	-0.2044	**	-0.7146	0.1788	-0.0039	0.0256	
	(1.2756)	(1.7808)	(1.593)	(1.5064)		(0.7042)	(0.8921)	(0.0362)	(0.0341)	
Percent of farm emp.	-7.6562**	-6.1249**	-2.9433	-2.7764		0.6549	1.1244*	0.0461	0.0501	
	(3.158)	(2.7977)	(2.7085)	(2.7603)		(0.4088)	(0.6631)	(0.0663)	(0.0616)	
Percent of manufacturing emp.	2.5714***	2.6219***	1.6219	1.9407*		-0.5284	-0.2176	0.038	0.0543***	
	(0.762)	(0.8044)	(1.1681)	(1.1668)		(0.591)	(0.4)	(0.028)	(0.0209)	
Percent of government emp.	3.9655**	2.8883*	3.4731*	3.2976		-2.737	-2.6913	0.056	0.0476	
	(1.5938)	(1.6128)	(2.0733)	(2.1486)		(1.8144)	(1.7689)	(0.0679)	(0.0736)	
Voter participation rate (%)	-1.7385*	-0.4806	-1.0098	-0.2886		0.3753	-0.1358	0.0432	-0.0076	
	(0.8968)	(1.4796)	(1.5686)	(1.8344)		(0.6166)	(0.5692)	(0.0424)	(0.0642)	
Percent of vote for Rep.	-0.6061	-1.1474	-0.4042	-0.6569		-0.1120*	-0.5879*	-0.0326**	-0.0704***	
	(0.996)	(1.5799)	(0.8896)	(1.2657)		(0.0622)	(0.3535)	(0.0143)	(0.018)	
Voting margin (%)	0.5198	0.4568	0.2194	0.1363		0.0874*	0.1174*	0.0117**	0.0147***	
	(0.399)	(0.4301)	(0.3281)	(0.3308)		(0.051)	(0.0687)	(0.0056)	(0.0055)	
Constant	28.7643	250.2547	19.1853	18.3924		18.2205	66.6883	-5.6025	-4.6683	
	(81.2865)	(158.8414)	(110.5084)	(154.0977)		(11.2425)	(41.5868)	(4.8167)	(4.1014)	
Year fixed effect	No	Yes	No	Yes		No	Yes	No	Yes	
Num. of obs.	41,554		6,373			41,554		6,373		
Num. of groups	2,600		422			2,600		422		

* p<.1; ** p<.05; *** p<.01

Table 3.8 Determinants of per capita funding on economic development programs by agency

Variables	Project grant					Formula grant				
	Non-persistent poverty counties		Persistent poverty counties		Structural break	Non-persistent poverty counties		Persistent poverty counties		Structural break
Poverty rate (%)	-0.6914***	-0.7108**	-1.0792**	-0.7465		0.7297	0.2537	-0.0417	-0.0153	
	(0.2657)	(0.2886)	(0.4972)	(0.5751)		(0.5828)	(0.6938)	(0.0263)	(0.0147)	
Median household income ($)	0.0003	-0.0006**	0.0004	0.0006		0.0002*	-0.0003	0	0	*
	(0.0002)	(0.0003)	(0.0005)	(0.0007)		(0.0001)	(0.0003)	(0)	(0)	
Per capita income support ($)	0.0506***	-0.0047	0.0367***	0.003		0.0078	0.0041	-0.0004	-0.0007	
	(0.0109)	(0.0094)	(0.0099)	(0.0121)		(0.0156)	(0.0164)	(0.0004)	(0.0005)	
Unemployment rate	0.6806**	-1.0377**	0.6929	-0.7022	**	-0.6882	0.1928	0.0348*	0.0304*	
	(0.2957)	(0.4669)	(0.5389)	(0.4511)		(0.7034)	(0.891)	(0.0191)	(0.0177)	
Percent of farm emp.	-2.1415*	-1.8283*	-0.3805	-0.5143		0.6221	1.0911*	-0.0102	0.0035	
	(1.2105)	(0.9787)	(0.7368)	(0.778)		(0.4071)	(0.6619)	(0.0137)	(0.0121)	
Percent of manufacturing emp.	0.3711	0.3068*	0.7530**	0.6973*		-0.5282	-0.2166	-0.0097	0.0051	
	(0.2303)	(0.1611)	(0.3226)	(0.3691)		(0.5906)	(0.3996)	(0.0103)	(0.0068)	
Percent of government emp.	1.3870***	1.1119***	1.2164**	1.2321**		-2.7276	-2.6825	0.0113	0.003	
	(0.3683)	(0.369)	(0.5265)	(0.5418)		(1.8128)	(1.7675)	(0.0102)	(0.009)	
Voter participation rate (%)	-0.2901	-0.6664	-0.2001	-0.3895		0.3563	-0.1352	-0.0046	0.003	
	(0.2462)	(0.4538)	(0.4663)	(0.5355)		(0.6164)	(0.5691)	(0.0355)	(0.0394)	
Percent of vote for Rep.	-0.6369	-0.7498	-0.2294	-0.4274		-0.1061*	-0.5817*	0.0033	-0.0072	*
	(0.4828)	(0.7724)	(0.2063)	(0.3239)		(0.0618)	(0.3532)	(0.0087)	(0.0088)	
Voting margin (%)	0.1876	0.1491	-0.0001	-0.0297		0.0860*	0.1172*	0.0036*	0.0045**	
	(0.1717)	(0.185)	(0.1242)	(0.1233)		(0.0508)	(0.0686)	(0.0019)	(0.0018)	
Constant	16.8041	95.8146	-13.1208	26.0548		20.0186*	66.4483	2.2429	1.3788	
	(30.9353)	(67.2732)	(24.6937)	(32.7432)		(11.2518)	(41.5721)	(2.2422)	(1.8963)	
Year fixed effect	No	Yes	No	Yes		No	Yes	No	Yes	
Num. of obs.	41,554		6,373			41,554		6,373		
Num. of groups	2,600		422			2,600		422		

* p<.1; ** p<.05; *** p<.01

Table 3.9 Determinants of per capita funding on economic development programs by the type of grant

3.5 Conclusion

Although the United States is one of the most developed counties, there are a lot of counties in chronic poverty. Based on the definition of persistent poverty county suggested by USDA, there are 425 persistent poverty counties which poverty rates are continuously 20 percent or more since 1970. The 10-20-30 rule for persistently poor counties was inserted into ARRA by Congressman Clyburn, who argues that these persistent poor counties are still in the lack of basic infrastructure. When considering existence of chronically poor counties, federal assistance for economic development still remains important. In this paper, I mainly focus on place-based targeting funds of economic development programs because “places do matter in wealth and income creation beyond the sum of the firms, workers, and owners of resources within them (Markusen & Glasmeier, 2008, p.83)”.

Previous studies about federal funding distribution used the whole amount of federal funds or just focused on a single program. Looking at a single program can be problematic because there may be the zero-inflated bias due to the small number of awards. Thus, this paper especially concentrates on economic development programs to examine determinants of funding distribution at the county level by comparing results between persistent poor counties and non-poor counties. I start to examine from funding on rural development programs of USDA and expand it to include several programs of HUD.

Regression results show that (i) non-persistent poverty counties with more distressed economic conditions such high poverty rates and high unemployment rates receive less per capita funding, (ii) overall, per capita funding distributed to persistent poverty counties is not affected by economic need or political concerns, (iii) distribution of both rural development funding operated by USDA and project grant are likely to be influenced by economic need, whereas both CDBG operated by HUD and formula grant are not, (iv) voting margin affects the distribution of CDBG (and formula grant) positively, which implies political cooperation may be required to get funding, and (v) structural differences in funding distribution between persistently poor counties and non-poor counties are hardly observed when including time effects.

When considering that economic development programs selected in this paper broadly have been aimed at developing distressed areas and helping low and moderate income people, the above results make me question the validity of place-based targeting. We may need better targeting criteria as Markusen and Glasmeier (2008) say, “Better targeting and performance standards should be implemented and changes in crude place-based eligibility explored”.

APPENDIX A

DATA APPENDIX

A.1 Simulating the amount of EITC using Census and ACS (American Community Survey) data

In this paper, I use the decennial Censuses data for 1990 and 2000 and the ACS data for the periods after 2000 to calculate the Earned Income Tax Credit (EITC) through TAXSIM provided by NBER.[71] The Census and ACS data have less income-related variables than the CPS data, but they contain the Public Use Microdata Area (PUMA) which is a geographic variable at the most disaggregated level so that I can assign commuting zones to each household by using crosswalk files of David Dorn.[72] The thing is that the Census and ACS data are collected in the household and the individual level, which both are not directly matched with tax filing units. So firstly, I tried to convert the household and individual level data into the family level data by using several variables such as the relationship with the head of household, marital status, age, etc. Then, I estimated the amount of EITC through TAXSIM.

Now I explain in detail how I define the tax units. First, I determine qualifying children and relatives based on the tax instruction of IRS, which is one of the main factors used to calculate the EITC. Actually, according to the instruction, for a primary tax filer to claim someone as qualifying children or relatives, they should be related to the primary tax filer like sons, daughters, siblings, etc.[73] However, when I use the Census and ACS survey data, I cannot make sure who is actually

[71] TAXSIM is the NBER's FORTRAN program for calculating liabilities under US Federal and State income tax laws from individual data. For more information, see http://users.nber.org/~taxsim/taxsim27/

[72] Visit the following website: http://www.ddorn.net/data.htm

[73] In addition, according to the IRS instruction for taxpayers, for qualifying dependents to be claimed, they also should be U.S. citizen, U.S. national, U.S. resident alien, or a resident of Canada or Mexico. However, there are not proper variables in the Census and ACS data to apply the citizen test. There is a variable indicating the citizenship status, but it says that about 95% of respondents have citizenship. So, I assume that all observations meet the requirement of the citizenship test.

claimed for tax filing even though the data has a variable showing the relationship with the head of household or the reference person. For example, a person whose age is 16 and relationship with the head of household is non-relatives cannot be considered as a qualifying child if I use the relationship condition, but there is still a chance for him or her to be claimed as a qualifying child of someone who is not observed in the survey data. Since the number of qualifying children is very important for the amount of EITC, I do not use the relationship restriction to include every possible qualifying child. Note that I used the IRS instruction of the year before each survey year until 2000 because the Census data for 1990 and 2000 collected income information during the previous calendar year. The ACS data for the periods after 2000 surveys income information during 12 months prior to the survey date, so I use the IRS instruction of the ACS survey year. More specifically, the following individuals are assumed as qualifying children and relatives:

- A qualifying child is an individual who is not the head of household (subfamily) or the spouse of the head, and
  - he or she is under age 19, or
  - he or she is under age 24 and in school, or
  - he or she has both independent living difficulty (DIFFMOB) and self-care difficulty (DIFFCARE), and he or she is not in labor force.[74]
- A qualifying relative is an individual who is not assigned as a qualifying child and is not the head of household (subfamily) or the spouse of the head, and
  - his or her gross income is less than a certain amount ($2,000 in 1990, $2,750 in 2000, $3,650 in 2010, and $4,000 in 2015).[75]

Second, I define a "family" based on several group identifier variables in the dataset, and then assign the head of a family by using relationship variables. Hence, the group identifiers in the Census and ACS data do not work perfectly so that there are families who do not have the head after I assign it. Also, there are qualifying children who are not linked to (the head of) a family. In one part, it happens because I define qualifying children without considering the relationship with

[74] The third condition is for including "permanently and total disabled" individuals as a candidate for qualifying children. DIFFMOB indicates that the respondent is difficult or impossible to perform basic activities outside the home alone for at least 6 months. DIFFCARE indicates that respondents have difficulty for at least 6 months to take care of their own personal needs, such as bathing, dressing, or getting around inside the home.

[75] Hence, the gross income is calculated by subtracting the sum of two income variables, INCWELFR (public assistance income) and INCSUPP (supplementary security income), from total personal income (INCTOT).

head of household, but, in other part, it occurs due to the inaccuracy of group variables. Therefore, I make a few assumptions to reflect and adjust these issues.

- A same-sex married couple is not treated as a married couple. Each individual of the couple is considered the head of each family unit.
- If a qualifying child defined in the first step does not have a mother or a father within a household, this child is linked to the head of the household.[76]
- If a qualifying child defined in the first step has a parent but does not have the head because of the inaccuracy of group identifiers, this child is assigned to his or her parent and that parent is considered the head of the family.
- I exclude observations who are defined as qualifying children but do not have any other persons in the same household.

Third, I assign a spouse of the head by using the relationship with the reference person, and I finally define tax filing units by assuming that each head is a tax filer who represents his or her family and claims his or her qualifying children and relatives as dependents. The determination of filing status depends on the head's marital status, whether the head has a spouse or not, and whether the head has qualifying dependents or not. Note that "No spouse" or "No dependents" mean that the head does not have a spouse or dependents within the *surveyed* same unit.

- Single Filing Status: The head does not have the spouse or dependents, and the head's marital status is not "married, spouse present" or "married, spouse absent."
- Head of Household: The head whose marital status is not "married, spouse present" does not have the spouse but dependents.[77]
- Married Filing Jointly: The head whose marital status is "married, spouse present" has the spouse.
- Married Filing Separately: The head whose marital status is "married, spouse absent" does not have the spouse or dependents.

[76] When I check whether a qualifying child without a head has a mother or a father, I use variables that indicate a personal number of the child's mother or father. Notice that these "mom (pop) location" variables are constructed after the survey finished so that they even embrace a link between children and their possible parents.

[77] Since a same-sex married couple is not considered a married couple in this analysis, one of the couple is treated as "Head of household" if he or she has dependents.

Before moving on to the next step, there are remaining observations whose filing status is not yet assigned. They are neither qualifying children nor qualifying relatives as defined above so that they cannot be linked to a primary taxpayer (family head).[78] For example, they can be a child who lives with his or her parents but is not qualifying children or qualifying relatives, parents who live with their children but are not qualifying relatives, an unmarried partner or one of the same-sex married couple who does not have any qualifying children, and so on. I basically assume that they are single filers, but include only observations who do not have their own children or qualifying children within the surveyed household because I want to at least rule out the possibility that they can file tax returns with dependents. For example, if a household consists of the head, two qualifying children, and the unmarried partner, each adult, the head and unmarried partner, could claim one child as a dependent separately.

Given the tax filing units defined above, now I need income information for each tax filer to simulate the EITC amounts through TAXSIM. Table A.1 shows which variables of the Census/ACS datasets are used for TAXSIM inputs. Some TAXSIM inputs are set to zero because there are no corresponding variables in the Census/ACS data. In TAXSIM, pwages and swages include wage and salary income as well as self-employment income that can be a negative value in the Census/ACS. Since TAXISM does not allow negative numbers for wage variables, I assume that negative business and farm income means "no net earnings from self-employment," and change negative values to zero.

Note that TAXSIM requires the number of dependents to reflect personal exemptions and the number of children under the certain age to estimate some relevant credits. According to the introductory page of TAXSIM, the number of children under age 18 seems to be used for the calculation of EITC, but I use the number of qualifying children as defined in this paper because it follows the requirements for the EITC even though it is not measured exactly due to the data availability.

Table A.2 shows the number of potential tax filers, which is population estimates weighted by the person weight of the family head. The numbers in square brackets are the official number of tax filers by filing status from IRS. The estimated numbers are about 90 percent of official statistics in terms of total filers, but the married jointly filers are relatively overestimated. Table A.3 reports

[78] Since I use only the income criterion when defining qualifying relatives, some of the remaining observations are categorized as qualifying relatives. However, they are non-relatives of the head according to the relationship variable.

results through TAXSIM, the number of tax filers with positive federal earned income credits and the sum of federal earned income credits by the number of qualifying children, which shows that the total number of EITC filers and the total amount of credits are estimated less than the official statistics. Meyer (2010) compares the distributions of EITC by the filing status and the number of qualifying children in two different datasets: IRS and CPS ASEC. It shows that both the total number of EITC recipients and the total benefits calculated from the CPS are less than two-thirds of those from the IRS. The author suggests possible reasons for this discrepancy: i) IRS payments to ineligible recipients, ii) too low sample weight for EITC recipients in the CPS, and iii) underreporting of earnings in the CPS. In addition, the numbers of EITC filers with one qualifying child are relatively underestimated when compared to official statistics, whereas the numbers of EITC filers with three or more qualifying children are overestimated which are not reported separately in Table A.3. The figures from the CPS ASEC of Meyer (2010) have a similar pattern: the CPS captures less amount of EITC of filers with one qualifying child, compared to filers with two or more qualifying children. This result may occur due to the three reasons mentioned above, or it might imply the strategic behavior of EITC filers with more than two children by claiming and splitting eligible children.

TAXSIM		Census/ACS	
Inputs	Definition	Variables	Definition
pwages	wage and salary income of primary taxpayer (include self-employment)	incwage, incbus	wage and salary income, business and farm income
swages	wage and salary income of spouse (include self-employment)	incwage, incbus	wage and salary income, business and farm income
dividends	dividend income	-	set to zero
intrec	interest received	-	set to zero
stcg	short term capital gains and losses	-	set to zero
ltcg	long term capital gains and losses	-	set to zero
otherprop	other property income, including unearned partnership and S-corp income, rent, non-qualified dividends, capital gains distributions on form 1040, other income or loss not otherwise enumerated here	incinvst	interest, dividend, and rental income
nonprop	other non-property income such as alimony, fellowships, state income tax refunds (itemizers only) adjustments and items such as alimony paid, Keogh and IRA contributions, foreign, income exclusion, NOLs	-	set to zero
pensions	taxable pensions and IRA distributions	incretir	retirement income
gssi	gross social security benefits	incss	social security income
ui	unemployment compensation received	-	set to zero
transfers	other non-taxable transfer income such as welfare, workers comp, veterans benefits	incwelfr, incsupp	welfare (public assistance) income, supplementary security income
rentpaid	rent paid	-	set to zero
proptax	real estate taxes paid	-	set to zero
otheritem	Other Itemized deductions that are a preference for the Alternative Minimum Tax	-	set to zero
childcare	child care expenses	-	set to zero
mortgage	Deductions not included in "otheritem" and not a preference for the AMT	-	set to zero

Notes: Definitions of all TAXSIM inputs are available at https://users.nber.org/~taxsim/taxsim27.

Table A.1 TAXSIM inputs corresponding income-related variables from the Census/ACS

Filing Status	1990	2000	2010	2015
Single	40,945,293	48,308,239	58,105,871	64,158,258
	NA	[56,927,117]	[64,846,356]	[71,086,947]
Head of household	11,916,518	15,426,213	19,264,278	19,590,580
	NA	[17,781,482]	[21,916,717]	[22,134,303]
Married, jointly	52,270,664	56,262,140	57,265,226	58,471,498
	NA	[49,980,900]	[53,526,090]	[54,294,820]
Married, separately	1,151,280	1,803,227	2,007,645	2,483,756
	NA	[2,385,646]	[2,532,292]	[2,977,192]
Non-filer	2,037,417	3,101,752	3,853,861	4,249,706
Total of Potential filers	108,321,172	124,901,571	140,496,881	148,953,798
	NA	[127,075,145]	[142,892,051]	[150,493,263]

Notes: Data come from the decennial Censuses and the ACS, and they are converted from the household level to the family level data to calculate the number of tax filers. To get population estimation, I used the personal weight of the reference person in a family. The "Non-filer" is the samples who have both zero (or negative) earnings and zero (or negative) total family income. Samples that primary taxpayers are under age 16 are excluded. The numbers in square brackets are the official statistics from the IRS website.

Table A.2 The number of potential tax filers by filing status

	TAXSIM Results				Official Statistics	
	Number		Amount		Number	Amount
All filers						
1990	9,707,848	(83%)	5,092,376,519	(77%)	11,696,000	6,595,000,000
2000	17,006,068	(88%)	24,078,072,092	(75%)	19,258,715	31,901,107,000
2010	24,367,166	(89%)	47,013,929,006	(79%)	27,367,756	59,562,029,000
2015	24,183,844	(89%)	52,217,223,339	(77%)	27,305,404	67,783,979,000
Filers with no qualifying children						
1990	-	-	-	-	-	-
2000	3,857,894	(120%)	696,130,128	(108%)	3,222,299	644,529,000
2010	7,129,271	(107%)	1,877,094,444	(107%)	6,647,462	1,752,786,000
2015	7,325,571	(108%)	2,166,107,788	(112%)	6,756,859	1,932,666,000
Filers with one qualifying child						
1990	4,675,931	-	2,475,062,136	-	NA	NA
2000	5,635,819	(72%)	7,848,965,491	(65%)	7,802,846	12,005,739,000
2010	7,422,605	(74%)	14,212,121,960	(68%)	10,000,746	21,014,164,000
2015	7,390,815	(73%)	16,034,960,383	(66%)	10,090,090	24,426,268,000
Filers with two or more qualifying children						
1990	2,986,546	-	1,549,126,910	-	NA	NA
2000	7,512,355	(91%)	15,532,976,472	(81%)	8,233,571	19,250,839,000
2010	9,815,290	(92%)	30,924,712,602	(84%)	10,719,546	36,795,083,000
2015	9,467,458	(91%)	34,016,155,169	(82%)	10,458,452	41,425,045,000

Notes: TAXSIM results are simulated values obtained through TAXSIM with the Census/ACS data. The official statistics are from www.taxpolicycenter.org, which does not provide statistics of EITC distribution by the number of qualifying children in 1990 (Fiscal Year 1989).

Table A.3 Simulated number of EITC filers and amounts of EITC

A.2 Constructing the `exposure to robots' variable

The industrial robot data comes from the International Federation of Robotics (IFR) who collects data provided by almost all industrial robot suppliers all over the world. The IFR calculates the operational stock of robots based on the annual sales of robots by assuming that their service life is 12 years on average. The industrial robot data are available for the period 1993-2016, but during the same period, the industry-level data are only available for 9 European countries: Denmark, Finland, France, Germany, Italy, Norway, Spain, Sweden, and the United Kingdom.[79] In the US, the industry-level data starts from 2004.[80] For those ten countries, I use information on the operational stock of robots in 19 industries, as in Acemoglu and Restrepo (2017): 6 non-manufacturing sectors and 13 manufacturing sectors.[81] However, as mentioned in Acemoglu and Restrepo (2017), some robots still remain unspecified, and the number of unspecified robots are not small in some countries and some time periods even though its total share of the total robots has declined over time, 19.2% in 2011 to 12.8% in 2016. I allocate these unspecified robots to 19 industries, using the proportion of each industry to the total by country. In addition, Denmark's robot data are not classified by industry from 1993 to 1995, so I impute missing values by deflating the 1996 robot stocks by industry using the growth rate of the total stock of robots of Denmark between each missing year and 1996.

The measure of robot density, the number of industrial robots per thousand workers $\left(\frac{R_i}{L_i}\right)$, for ten countries utilizes employment information of EU KLEMS data[82] released in March 2008.[83]

[79] According to the IFR's report, the industrial classification has followed the International Standard Industrial Classification of All Economic Activities (ISIC) revision 4 since 2010.

[80] Until 2010, the IFR reports only the overall stock of robots for North America (the United States, Canada, and Mexico). However, based on reports after 2010, the operational stock of robots in the US accounts for more than 90 percent on average.

[81] Six non-manufacturing industries are as follows: agriculture, hunting, forestry, and fishing; mining and quarrying; electricity, gas and water supply; construction; education, research and development; and other non-manufacturing industries. The manufacturing industry is categorized into 13 sectors: food products, beverages, and tobacco products; textiles, leather, and wearing apparel; wood and wood products; paper, paper products, and printing; plastic and chemical products; glass, ceramics, stone, and mineral products; basic metals; metal products; industrial machinery; electrical/electronics; automotive; other transport equipment; all other manufacturing sectors.

[82] Since the EU KLEMS data do not include information on Norway, I use the mean value of employment in three Scandinavian countries (Denmark, Finland, and Sweden).

[83] The EU KELMS data released after 2012 follows ISIC Rev. 4 industry classification. But it is available from 1995, so for the periods before 1995 it includes only the estimates. Besides, the EU KLEMS data classified by ISIC Rev.4 have 34 industry categories, whereas the data under ISIC Rev.3 have 72 categories. More importantly, the 1990's data on two countries, Sweden and the US, are not available under the ISIC Rev 4.

According to the IFR's report, the industrial classification has followed the International Standard Industrial Classification of All Economic Activities (ISIC) revision 4 since 2010,[84] whereas the 2008 KLEMS data is classified by the ISIC Rev. 3. So, I reassign the industry codes on the basis of the release note of EU KELMS 2012 to make the 2008 data roughly follow the ISIC Rev. 4. Also, I use the number of US equivalent workers when calculating the number of industrial robots per thousand workers for European countries as mentioned in Acemoglu and Restrepo (2017). It is because the same number of workers may not imply the same labor intensity between countries unless they have the same total hours worked.[85]

Lastly, I calculate employment shares in industry i in commuting zone c, $l_{ci} = \frac{L_{ci}}{L_c}$, using the U.S. decennial censuses from 1980 and 1990. A small problem is that the U.S. Census Bureau has its own industrial classification. So, I assign the industry codes of ISIC Rev. 3 to each Census industrial categories and then make it follow the ISIC Rev. 4 in the same way as the EU KLEMS data. The detailed procedures are given in the Appendix A.3. Note that, when I calculate the total employment by CZs from the Census or by countries from the KLEMS, the following industries are excluded in this analysis: public administration and defense, private households, and extra-territorial organizations and bodies.

A.3 The adjustment of industry classification from Census to ISIC Rev. 4

The detailed procedures to make the Census industry codes roughly correspond to the ISIC Rev. 4 are as follows: (i) Assign 1987 SIC 4-digit codes to Census 1990 industry codes (ind1990) using the 1990 Census Sample Codebook.[86] Because one Census code often corresponds to multiple SIC codes, I calculate the employment share of each SIC code within a Census code using 1990 County Business Pattern (CBP) data and use it as the weight mapping from the Census industry to SIC 4-digit industry. Assuming that the employment shares by SIC 4-digit are not different across commuting zones, I use the information only at the national level. Note that some industries in the 1990 CBP data are actually reported at 3-digit level, not 4-digit level. Hence, if an industry is

---

[84] However, the industry categories of the IFR do not correspond completely with the ISIC Rev. 4.

[85] I first calculate the hours worked per worker by industry in the US, then divide the total hours worked in each EU country by the working hours of one US worker.

[86] See https://usa.ipums.org/usa/volii/codebooks.shtml

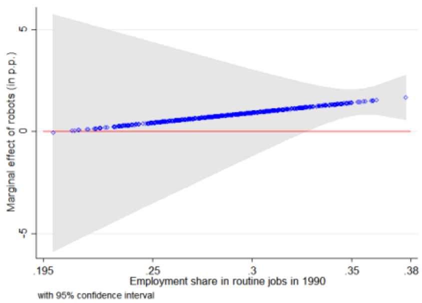
reported only at the 3-digit level and the 3-digit code has only one 4-digit code, the 3-digit code is considered equivalent to the 4-digit code. If the CBP data reports an industry only at the 3-digit level and the 3-digit code is disaggregated into multiple 4-digit codes, I assign equal probabilities to each 4-digit code in the 3-digit code. (ii) Based on the industry concordance from 1987 SIC 4-digit to ISIC Rev 3. 4-digit,[87] I can assign 2-digit codes of ISIC Rev. 3 to each Census industry with the final weight indicating the share of each Census industry that maps to a given 2-digit ISIC Rev. 3 code. But, according to the industry concordance, one SIC 4-digit code is often mapped to multiple ISIC Rev. 3 4-digit codes. Since I could not figure out the share of each ISIC 4-digit code corresponding to a given SIC 4-digit code, I assign equal probabilities to each ISIC 4-digit code. iii) The ISIC Rev. 3 codes assigned to every Census industry code are eventually mapped to the ISIC Rev. 4 codes by following the release note of EU KELMS 2012.

[87] For the industry concordance, visit https://www.macalester.edu/research/economics/page/haveman/Trade.Resources/tradeconcordances.html

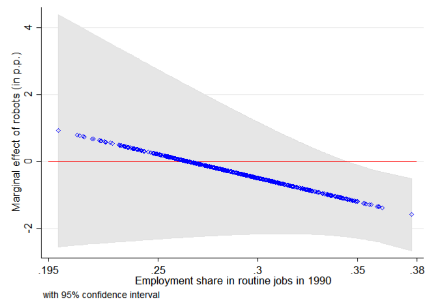
APPENDIX B

SUPPLEMENTARY MATERIAL

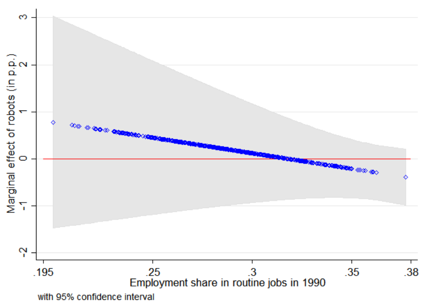
B.1 Figures and Tables for Chapter 1



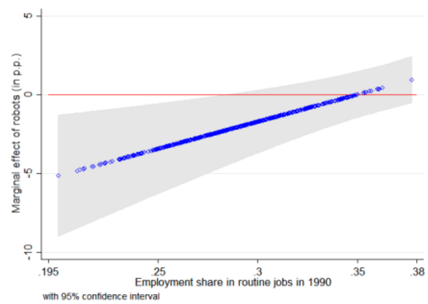
(a) Phase-in, single w/ children



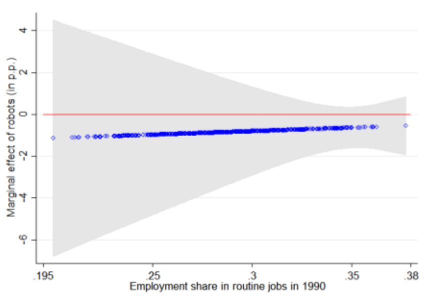
(b) Phase-in, married w/ children



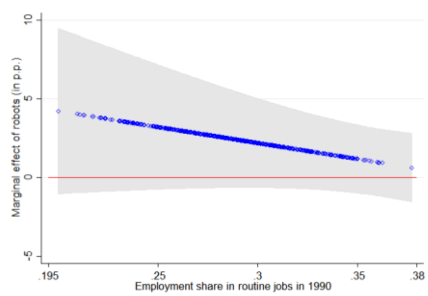
(c) Flat, single w/ children



(d) Flat, married w/ children



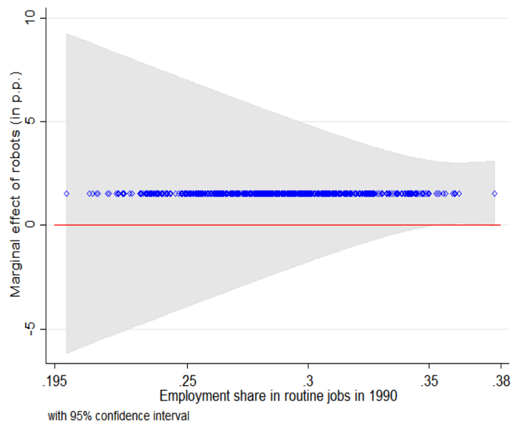
(e) Phase-out, single w/ children



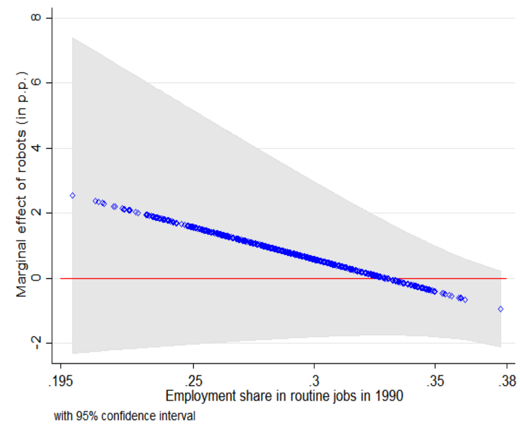
(f) Phase-out, married w/ children

Notes: The marginal effects and confidence intervals for each sample are calculated based on the IV estimates in Table B.4.

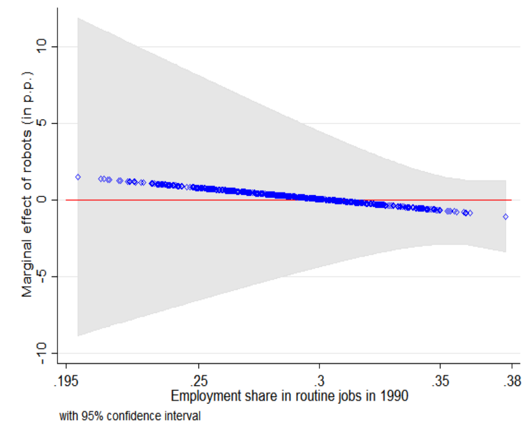
Figure B.1 Marginal effects of industrial robots on the share of EITC filers within the eligible region, 1990-2015



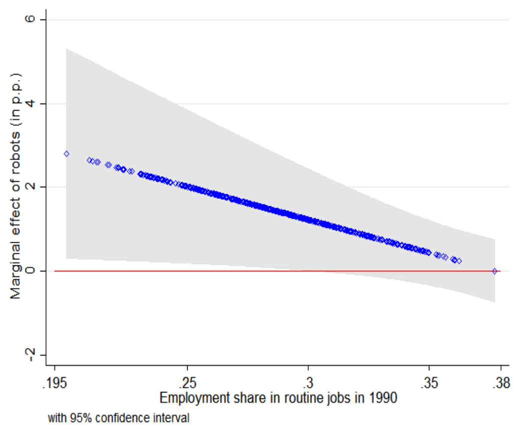
(a) Phase-in region



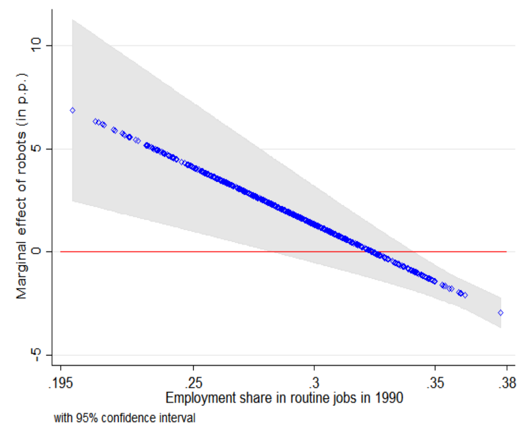
(b) Flat region



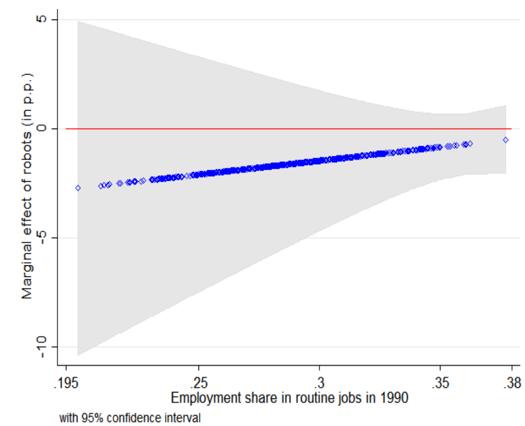
(c) Phase-out region



(d) Near phase-out region



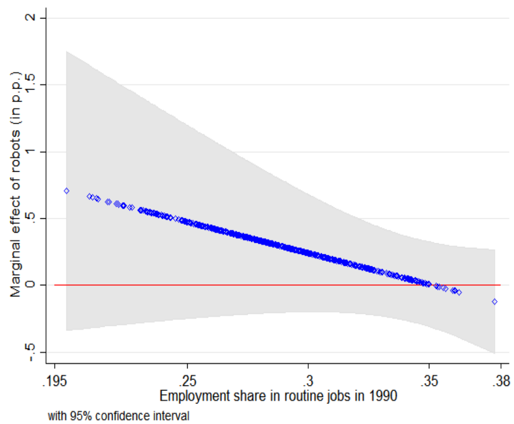
(e) Above the near phase-out region



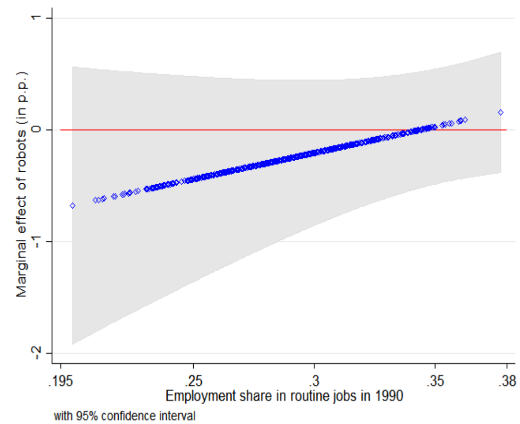
(f) Zero earnings

Notes: The marginal effects and confidence intervals for each sample are calculated based on the IV estimates in Table B.6. The near-phase-out region is from the end of the phase-out to 25% above the end of the phase-out which is also the beginning of the above the-near-phase-out region.

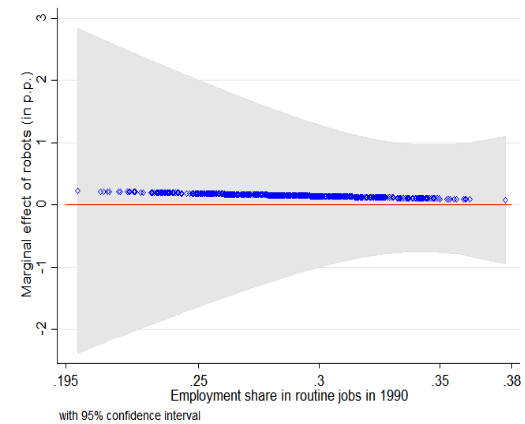
Figure B.2 Marginal effects of industrial robots on the ratio of tax-filers by earnings region, 1990-2015, Single filers with children (fixed denominator)



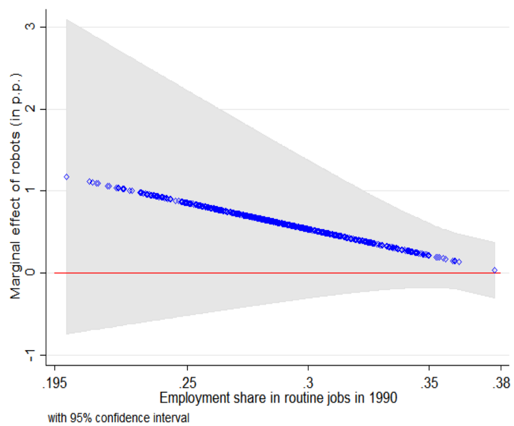
(a) Phase-in region



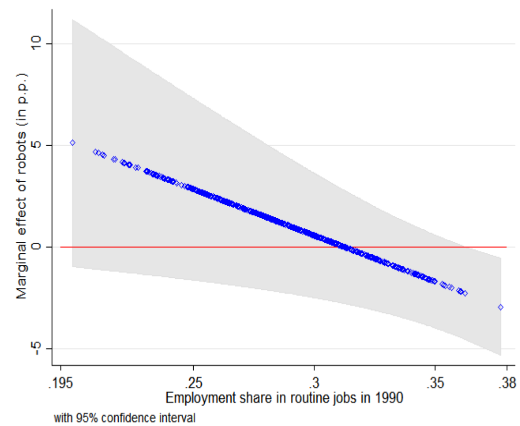
(b) Flat region



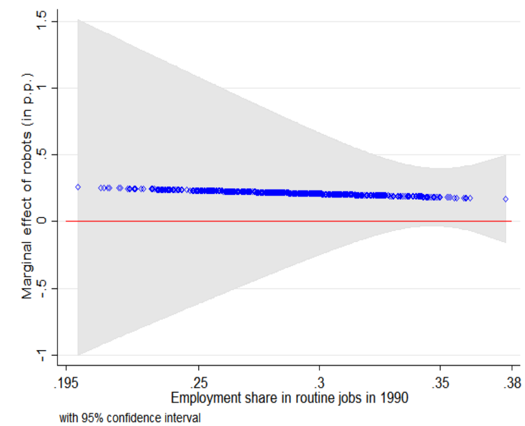
(c) Phase-out region



(d) Near phase-out region



(e) Above the near phase-out region



(f) Zero earnings

Notes: The marginal effects and confidence intervals for each sample are calculated based on the IV estimates in Table B.6. The near-phase-out region is from the end of the phase-out to 25% above the end of the phase-out which is also the beginning of the above the-near-phase-out region.

Figure B.3 Marginal effects of industrial robots on the ratio of tax-filers by earnings region, 1990-2015, Married filers with children (fixed denominator) *(a) The impact of routine-biased technology*

	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Share of employment	38.412***	-2.740	41.361*	79.796***
in routine jobs, 1990	(4.381)	(4.523)	(24.032)	(11.045)
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Share of employment	842.258***	-14.677	1106.443*	1883.188***
in routine jobs, 1990	(195.814)	(12.966)	(579.458)	(514.996)
Observations	722	1444	1444	1444

(b) The impact of exposure to industrial robots

	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in EU	0.017	-0.241*	1.177	-0.879
1993-2015 (p30)	(0.227)	(0.120)	(0.834)	(0.675)
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in EU	-4.874	-0.495	-29.071	-59.624*
1993-2015 (p30)	(7.797)	(0.349)	(26.293)	(33.816)
Observations	722	1444	1444	1444

Note: The tables show the OLS estimates of the impact of each measure of technology, the share of employment in routine jobs in 1990 and the exposure to industrial robots, on EITC usage. In the table (a), regressions do not include the latter measure of technology, whereas the first measure is excluded in the regressions of table (b). The dependent variable is the change in EITC filers to total filers ratio between 1990 and 2015 in Panel A, measured in percentage points, and the change in EITC amount per tax filer between 1990 and 2015 in Panel B. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, and the exposure to Chinese imports. Also, all regressions, except those for the pooled sample, include group-specific intercepts and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total filers of CZs in 1990. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table B.1 The impact of each measure of technology on EITC usage, 1990-2015

	Num. of qualifying children	1990 (FY1989)	2015	
		Single/Married	Single	Married
Flat	0	NA	6,580	
	1	6,500	9,880	
	2	6,500	13,870	
	3+	6,500	13,870	
Phase-out	0	NA	8,240	13,760
	1	10,240	18,110	23,630
	2	10,240	18,110	23,630
	3+	10,240	18,110	23,630
Near Phase-out	0	NA	14,820	20,340
	1	19,340	39,131	44,651
	2	19,340	44,454	49,974
	3+	19,340	47,747	53,267
Above the Near Phase-out (25% above the end of phase-out)	0	NA	18,525	25,425
	1	24,175	48,914	55,814
	2	24,175	55,568	62,468
	3+	24,175	59,684	66,584
Above the Near Phase-out (50% above the end of phase-out)	0	NA	22,230	30,510
	1	29,010	58,697	66,977
	2	29,010	66,681	74,961
	3+	29,010	71,621	79,901

Note: The table shows the criteria for each region of earnings. Note that each number is also the ending earnings of the previous level of earnings region. The near-phase-out region begins from the end of phase-out earnings, and the above the-near-phase-out region starts from 25% or 50% above the end of phase-out earnings.

Table B.2 The beginning earnings of each regions by filing groups

(a) Without interaction term

	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in US	0.175	-0.147	0.844	-0.308
2004-2015	(0.168)	(0.091)	(0.646)	(0.514)
share in routine jobs, 1990	39.502***	-2.640	40.793*	81.284***
	(5.169)	(4.677)	(23.555)	(11.407)
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in US	3.009	-0.204	-11.900	-32.732
2004-2015	(6.755)	(0.272)	(24.538)	(27.724)
share in routine jobs, 1990	883.002***	-13.457	1152.851*	1923.068***
	(233.542)	(14.269)	(591.487)	(536.605)
Weak IV F-stat	42.699	42.846	49.483	46.015
Observations	722	1444	1444	1444

(b) With interaction

	All	Childless	Single, children	Married, children
Panel A. Change in ratio of EITC filers to total filers, 1990-2015 (in p.p.)				
Exposure to robots in US	0.439	-0.924	-1.989	-6.472
2004-2015	(1.534)	(0.686)	(5.006)	(4.801)
Exposure to robots in US	-0.756	2.222	8.107	17.680
× share in routine jobs, 1990	(4.312)	(1.889)	(13.863)	(13.433)
share in routine jobs, 1990	41.024***	-7.135	24.727	45.783*
	(8.914)	(5.312)	(42.189)	(24.564)
Joint significance (p-value)	0.575	0.191	0.318	0.389
Panel B. Change in EITC amount per tax filer, 1990-2015 (2015 US$)				
Exposure to robots in US	-77.181	-3.749	-19.452	-448.125***
2004-2015	(51.399)	(2.462)	(164.897)	(158.450)
Exposure to robots in US	229.633	10.145	21.618	1191.318***
× share in routine jobs, 1990	(142.961)	(6.729)	(442.919)	(422.687)
share in routine jobs, 1990	420.507	-33.981*	1110.012	-469.066
	(276.262)	(18.988)	(950.340)	(776.570)
Joint significance (p-value)	0.225	0.313	0.883	0.018
Weak IV F-stat	24.631	24.886	28.029	26.349
Observations	722	1444	1444	1444

Note: The tables show IV estimates of the impact of exposure to industrial robots on EITC usage, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the mean of EU's exposure to robots, the interaction term is also instrumented with the product of the mean of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variable is the change in EITC filers to total filers ratio between 1990 and 2015 in Panel A, measured in percentage points, and the change in EITC amount per tax filer between 1990 and 2015 in Panel B. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. Table (a) does not include the interaction term between automation and share in routine jobs, but table (b) include it. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to China imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total filers of CZs in 1990. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table B.3 The impact of exposure to industrial robots on EITC usage, 1990-2015 (IV estimates using the mean of exposure to robots in other countries)

	All	Childless	Single, children	Married, children
Panel A. Change in Phase-in region				
Exposure to robots in US	0.322	-1.602	-2.030	3.750
2004-2015	(4.465)	(6.623)	(6.774)	(3.759)
Exposure to robots in US	-2.290	1.934	9.831	-14.103
× share in routine jobs, 1990	(12.945)	(19.006)	(19.122)	(10.188)
share in routine jobs, 1990	47.248*	-35.282	46.870	53.514*
	(26.410)	(40.071)	(40.338)	(28.823)
Joint significance (p-value)	0.501	0.315	0.000	0.012
Panel B. Change in Flat region				
Exposure to robots in US	-1.721	-4.034	3.841	-12.020***
2004-2015	(2.388)	(4.112)	(4.402)	(4.110)
Exposure to robots in US	4.505	13.020	-13.212	34.369***
× share in routine jobs, 1990	(6.487)	(11.282)	(12.030)	(11.125)
share in routine jobs, 1990	-1.153	-34.728	25.524	-12.664
	(16.778)	(27.925)	(23.675)	(25.781)
Joint significance (p-value)	0.761	0.270	0.043	0.007
Panel C. Change in Phase-out region				
Exposure to robots in US	1.398	5.636	-1.811	8.270
2004-2015	(5.107)	(5.439)	(6.618)	(5.552)
Exposure to robots in US	-2.215	-14.954	3.381	-20.266
× share in routine jobs, 1990	(14.786)	(15.452)	(18.809)	(14.996)
share in routine jobs, 1990	-46.095	70.010**	-72.394*	-40.850
	(31.158)	(33.387)	(40.431)	(32.742)
Joint significance (p-value)	0.565	0.479	0.447	0.282
Weak IV F-stat	24.176	20.553	23.478	26.894
Observations	722	1444	1444	1444

Note: The table shows IV estimates of the impact of exposure to industrial robots on the share of EITC filers in each eligible region, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, the interaction term is also instrumented with the product of the 30th percentile of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variable is the change in the share of EITC filers in phase-in region between 1990 and 2015 in Panel A, the change in flat region in Panel B, and the change in phase-out region in Panel C. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as EITC filers of CZs in 1990. Note that the number of EITC filers of Childless sample is zero in 1990 so I use the number of total filers in 1990 as the weight. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table B.4 The impact of automation on the change in the share of EITC filers within eligible region, 1990-2015 (IV estimates)

	All	Childless	Single, children	Married, children
Panel A. Near phase-out region: from the end of phase-out to 50% above it				
Exposure to robots in US	1.448*	-0.530	4.592*	1.551
2004-2015	(0.828)	(1.547)	(2.572)	(1.733)
Exposure to robots in US	-3.600	0.985	-12.023*	-2.285
× share in routine jobs, 1990	(2.290)	(4.307)	(6.591)	(4.785)
share in routine jobs, 1990	18.907**	-0.944	11.451	15.625
	(7.937)	(10.503)	(19.695)	(12.468)
Joint significance (p-value)	0.087	0.110	0.189	0.048
Panel B. Above near phase-out region				
Exposure to robots in US	6.953*	9.568**	3.766*	5.982
2004-2015	(3.628)	(4.035)	(2.234)	(5.098)
Exposure to robots in US	-21.557**	-27.480**	-12.788**	-18.675
× share in routine jobs, 1990	(10.753)	(11.183)	(5.990)	(14.184)
share in routine jobs, 1990	64.742**	123.100***	3.070	-59.455**
	(26.301)	(27.328)	(11.076)	(24.245)
Joint significance (p-value)	0.094	0.021	0.000	0.101
Weak IV F-stat	20.129	20.553	21.576	21.531
Observations	722	1444	1444	1444

Note: The table shows IV estimates of the impact of exposure to industrial robots on the share of EITC filers in each eligible region, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, the interaction term is also instrumented with the product of the 30th percentile of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variables are the changes in the ratio of tax-filers above the eligible region for EITC, where the near-phase-out region starts from the end of the phase-out to 50% above it. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total tax-filers of CZs in 1990. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table B.5 The impact of automation on the change in the ratio of tax-filers in earnings region above EITC eligibility, 1990-2015 (IV estimates)

	All	Childless	Single, children	Married, children
Panel A. Change in Phase-in region				
Exposure to robots in US	-0.975	-2.874	1.539	1.638
2004-2015	(1.274)	(1.802)	(8.580)	(1.221)
Exposure to robots in US	2.211	6.536	0.013	-4.658
× share in routine jobs, 1990	(3.528)	(5.051)	(23.462)	(3.513)
share in routine jobs, 1990	3.500	-29.889***	91.303**	31.599***
	(7.334)	(10.535)	(43.504)	(9.936)
Joint significance (p-value)	0.338	0.001	0.112	0.406
Panel B. Change in Flat region				
Exposure to robots in US	-0.642	-1.226	6.484	-1.616
2004-2015	(1.133)	(1.082)	(5.121)	(1.343)
Exposure to robots in US	1.542	3.090	-19.666	4.694
× share in routine jobs, 1990	(3.048)	(3.023)	(13.418)	(3.714)
share in routine jobs, 1990	11.599*	-14.406**	107.882***	37.573***
	(6.761)	(7.062)	(28.990)	(9.477)
Joint significance (p-value)	0.753	0.184	0.090	0.442
Panel C. Change in Phase-out region				
Exposure to robots in US	0.446	0.183	4.435	0.391
2004-2015	(2.462)	(2.279)	(11.603)	(3.021)
Exposure to robots in US	-1.363	-0.688	-14.568	-0.83
× share in routine jobs, 1990	(7.018)	(6.650)	(31.883)	(8.678)
share in routine jobs, 1990	44.598***	0.778	190.690***	93.229***
	(15.814)	(14.909)	(60.675)	(24.355)
Joint significance (p-value)	0.973	0.959	0.660	0.967
Panel D. Change in Near Phase-out region: from the end of phase-out to 25% above it				
Exposure to robots in US	0.681	-1.460	5.948**	2.451
2004-2015	(1.034)	(1.645)	(2.692)	(2.104)
Exposure to robots in US	-2.098	3.480	-15.731**	-6.388
× share in routine jobs, 1990	(2.839)	(4.611)	(7.254)	(5.674)
share in routine jobs, 1990	21.198**	-6.239	66.640***	49.670***
	(8.545)	(10.890)	(20.959)	(14.438)
Joint significance (p-value)	0.614	0.134	0.086	0.459

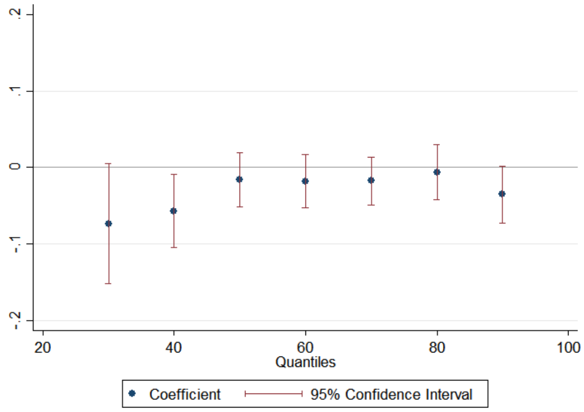
(continued)

	All	Childless	Single, children	Married, children
Panel E. Change in Above Near Phase-out region				
Exposure to robots in US 2004-2015	17.342***	15.877**	17.935***	14.163**
	(5.555)	(7.224)	(4.900)	(6.623)
Exposure to robots in US × share in routine jobs, 1990	-55.725***	-51.802***	-55.295***	-45.338**
	(15.576)	(19.460)	(13.351)	(18.230)
share in routine jobs, 1990	247.128***	270.255***	172.332***	119.272*
	(60.433)	(70.967)	(32.729)	(62.058)
Joint significance (p-value)	0.000	0.000	0.000	0.009
Panel F. Change in Zero earnings				
Exposure to robots in US 2004-2015	-8.440*	-16.802**	-5.257	0.359
	(4.991)	(8.531)	(8.579)	(1.486)
Exposure to robots in US × share in routine jobs, 1990	21.904	44.903*	12.583	-0.503
	(13.804)	(23.870)	(23.558)	(4.254)
share in routine jobs, 1990	-37.941	-130.540**	49.959	14.307
	(28.612)	(53.501)	(55.024)	(9.534)
Joint significance (p-value)	0.175	0.087	0.539	0.243
Weak IV F-stat	20.129	20.553	21.576	21.531
Observations	722	1444	1444	1444

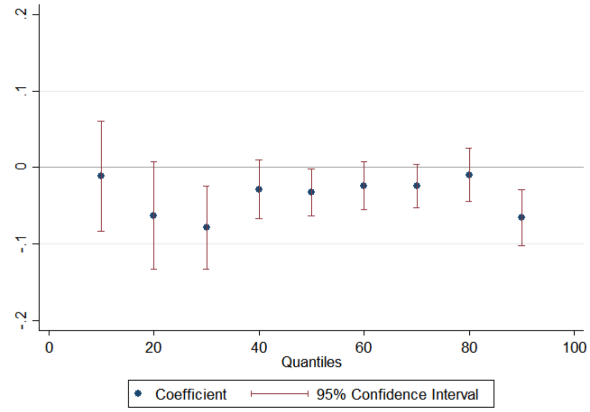
Note: The table shows IV estimates of the impact of exposure to industrial robots on the share of tax-filers in each region of earnings, where the exposure to robots in the U.S. between 2004 and 2015 is instrumented with the 30th percentile of EU's exposure to robots, the interaction term is also instrumented with the product of the 30th percentile of EU and the share in routine jobs, and the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. The dependent variables are the change in the ratio of tax-filers in each region relative to total taxpayers from 1990 to 2015, but the denominator is fixed as the total tax-filers in 1990. The first three regions follow the earnings criteria for EITC. The near-phase-out region begins from the end of the phase-out region to 25 percent above the end earnings of the phase-out region. The above the-near-phase-out region includes all families with positive earnings above the end of the near-phase-out. The models for the pooled sample have one observation per commuting zone (CZ), whereas the remaining models have two observations per CZ. Specifically, Childless sample is estimated at the marital status-CZ level, and the others are estimated at the number of children-CZ level. All regressions include a constant, demographic characteristics of CZs in 1990 (the ratio of working-age population, the ratio of female population, the ratio of population with college or more education, the ratio of non-white population), the share of employment in manufacturing, the exposure to Chinese imports, and the share of employment in routine jobs in 1990 (from Autor and Dorn, 2013). Also, all regressions include group-specific intercepts (except for the pooled sample) and state fixed effects. In addition, all regressions are weighted by relevant denominators of each outcome variable such as total tax-filers of CZs in 1990. The weak IV test reports F-statistics of Kleibergen-Paap. Standard errors are clustered by state and shown in parentheses. *p<0.10, **p<0.05, ***p<0.01.

Table B.6 The impact of exposure to industrial robots on the change in the ratio of tax-filers with the fixed denominator, 1990-2015 (IV estimates)

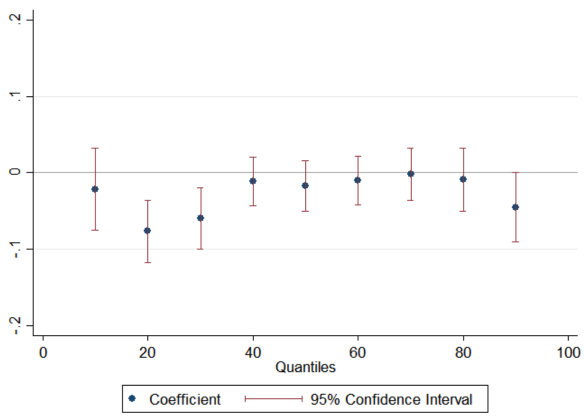
B.2 Figures for Chapter 2



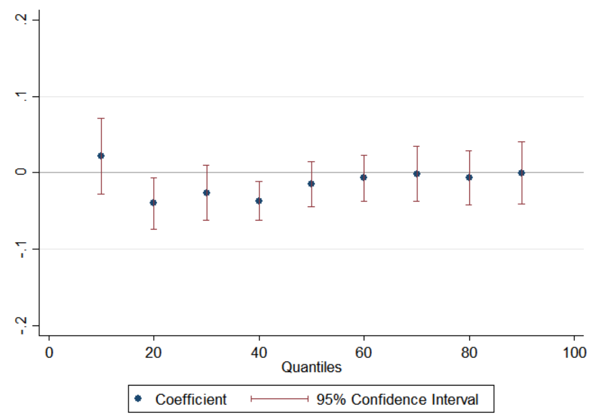
(a) Change in the log of total annual income



(b) Change in the log of annual earnings



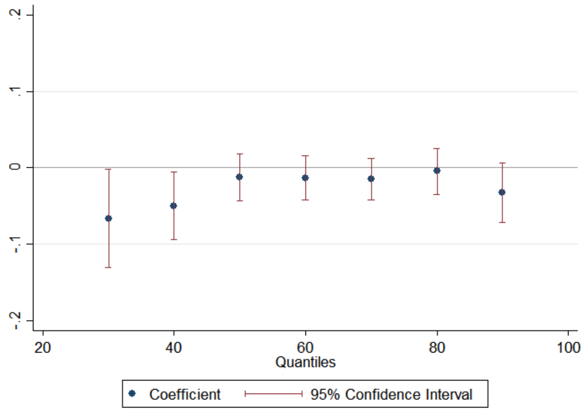
(c) Change in the log of weekly earnings



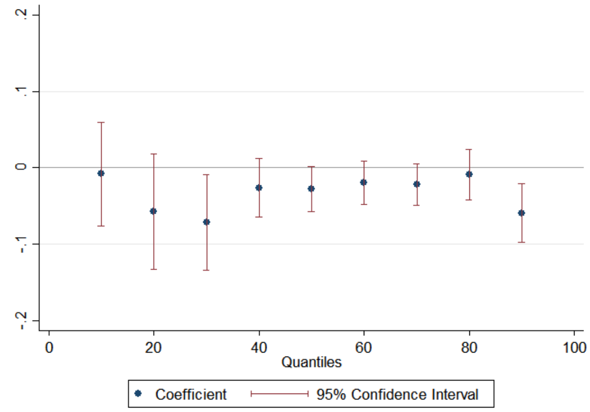
(d) Change in the log of hourly earnings

Notes: The figures show OLS estimates of $\beta_1(\tau)$ in equation (2.5), which indicates the effect of automation on the change in residual log income at each quantile for different types of income measures. Here, automation is measured by the mean of the adjusted penetration of robots in five European countries, which is available at https://economics.mit.edu/faculty/acemoglu/data. Note that figure (a) shows the estimates above the 30th percentile of the distribution since most of the observations have zero income at the bottom distribution when we focus on the total annual income. All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

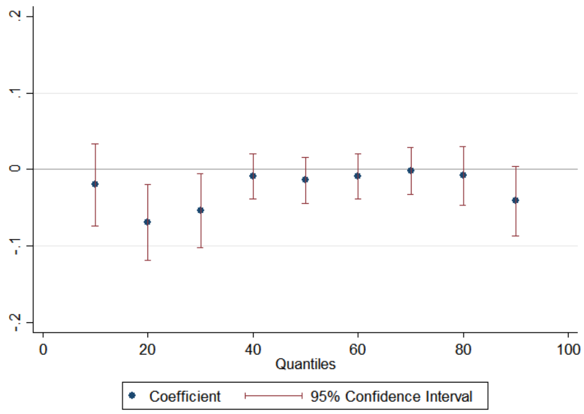
Figure B.4 The effect of exposure to industrial robots on the income distribution by different income variables (OLS estimates with the updated version for exposure to robots in Acemoglu and Restrepo (2020))



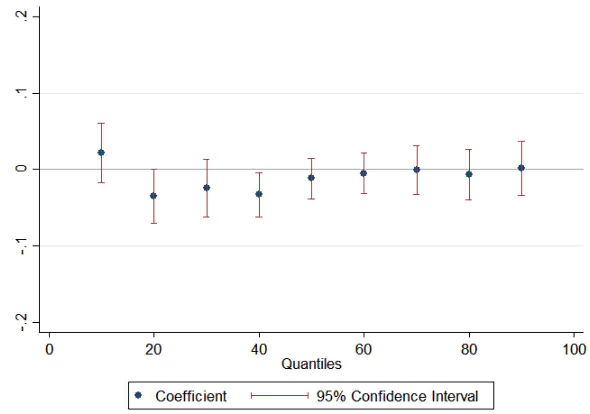
(a) Change in the log of total annual income



(b) Change in the log of annual earnings



(c) Change in the log of weekly earnings



(d) Change in the log of hourly earnings

Notes: The figures show 2SLS estimates of $\beta_1(\tau)$ in equation (2.5), where the estimates indicate the effect of automation on the change in residual log income at each quantile for different types of income measures. Here, the exposure to robots in the U.S. is instrumented with the mean of the adjusted penetration of robots in five European countries, which is available at https://economics.mit.edu/faculty/acemoglu/data. And the exposure to Chinese imports is instrumented with other eight countries' exposure to Chinese imports. Note that figure (a) shows the estimates above the 30th percentile of the distribution since most of the observations have zero income at the bottom distribution when we focus on the total annual income. All regressions are weighted by each CZ's share of population in 1990, and standard errors are clustered by state.

Figure B.5 The effect of exposure to industrial robots on the income distribution by different income variables (2SLS estimates with the updated version for exposure to robots in Acemoglu and Restrepo (2020))

B.3 Figures and Tables for Chapter 3

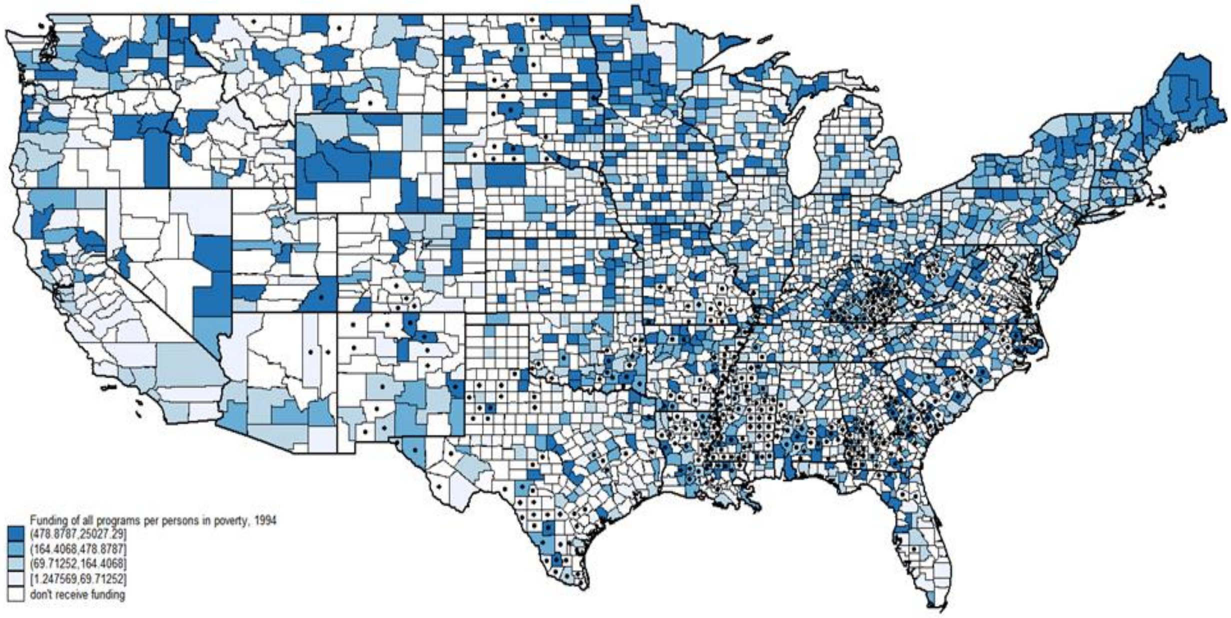


Figure B.6 Total amount of funding of all programs per person in poverty, 1994

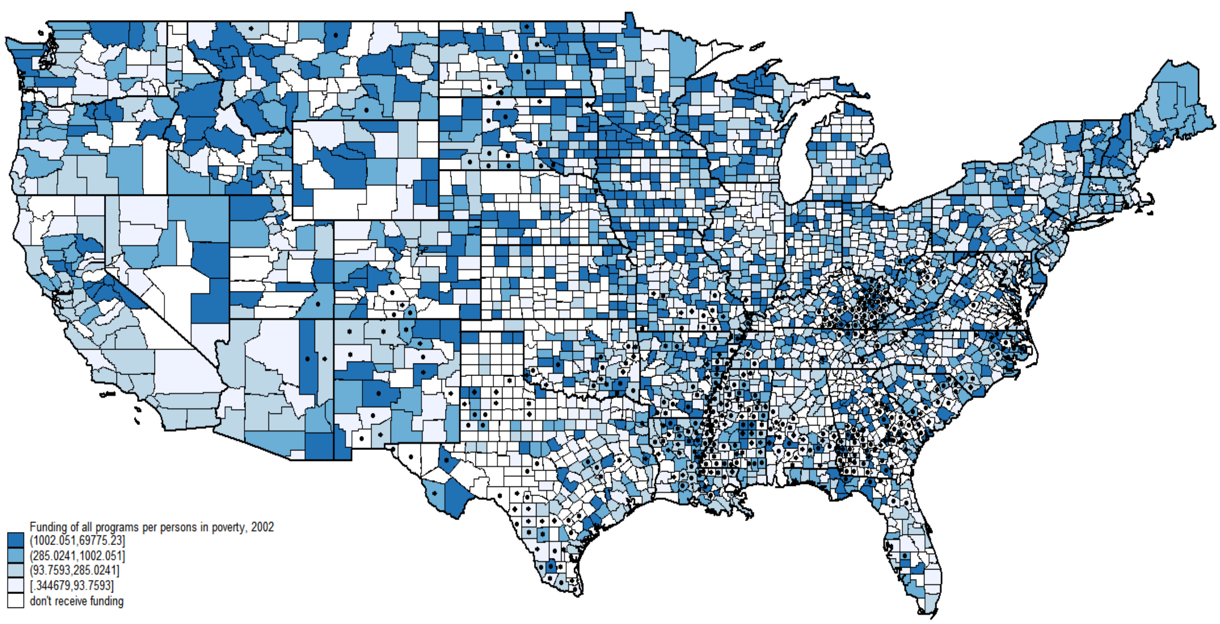


Figure B.7 Total amount of funding of all programs per person in poverty, 2002

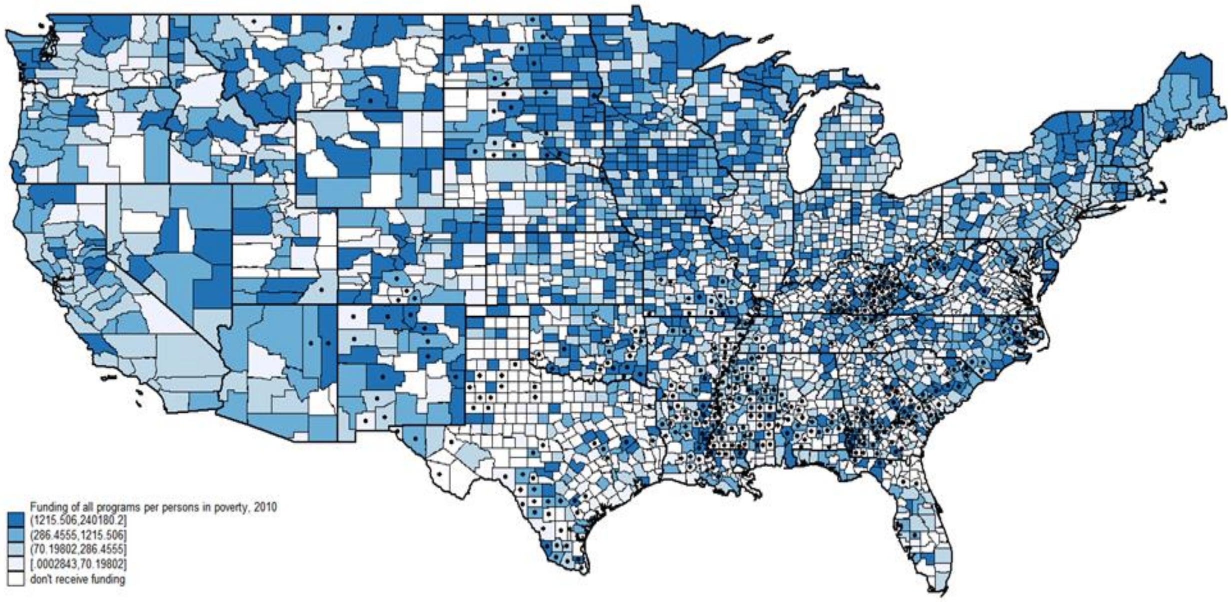


Figure B.8 Total amount of funding of all programs per person in poverty, 2010

Variables	Non-persistent poverty counties		Persistent poverty counties		Structural break	
Poverty rate (%)	-62.6899***	-56.5744***	-34.4475***	-18.5108**	**	**
	(12.4557)	(12.71)	(6.8854)	(8.8563)		
Median household income ($)	0.0255***	-0.0150*	0.0126	0.0193		*
	(0.0052)	(0.0082)	(0.0102)	(0.0186)		
Per capita income support ($)	0.9493***	-1.1919***	0.3669***	-0.0762	*	**
	(0.2802)	(0.4113)	(0.1412)	(0.1815)		
Unemployment rate	46.8584***	-34.4448**	22.5552***	2.2375	*	**
	(12.4227)	(17.0933)	(6.904)	(6.6979)		
Percent of farm emp.	-67.0652**	-50.4901**	-14.095	-12.5423	*	
	(26.4737)	(23.7995)	(11.223)	(11.2977)		
Percent of manufacturing emp.	19.7020**	23.6893***	7.637	10.2433**		
	(8.8073)	(8.2397)	(5.2762)	(5.1702)		
Percent of government emp.	8.6906	-0.4262	18.4621**	16.8390*		
	(20.4296)	(20.1734)	(9.1668)	(9.4951)		
Voter participation rate (%)	-10.7682	-3.351	-3.3514	-0.3969		
	(9.825)	(13.3691)	(7.1051)	(8.1157)		
Percent of vote for Rep.	-6.4221	-16.1804	-0.0041	-2.0128		
	(8.3721)	(13.4778)	(4.1004)	(5.7435)		
Voting margin (%)	5.7107*	5.6139	0.6433	0.43		
	(3.4456)	(3.6917)	(1.5366)	(1.5458)		
Constant	690.5085	2900**	236.6254	57.8126		
	(711.6541)	(1400)	(511.8287)	(702.5584)		
Year fixed effect	No	Yes	No	Yes	No	Yes
Num. of obs.	41,554		6,373			
Num. of groups	2,600		422			

Table B.7 Determinants of funding on economic development programs per person in poverty

Variables	USDA						HUD					
	Non-persistent poverty counties		Persistent poverty counties		Structural break		Non-persistent poverty counties		Persistent poverty counties		Structural break	
Poverty rate (%)	-62.6281***	-53.0479***	-33.5603***	-18.1795**	**	**	1.0109	-3.2704	-0.2041	-0.1821		
	(12.0462)	(11.3303)	(6.8838)	(8.848)			(3.2016)	(5.7395)	(0.3066)	(0.224)		
Median household income ($)	0.0238***	-0.0128	0.0123	0.0185			0.0018**	-0.0021	0.0007	0.001		
	(0.0051)	(0.0078)	(0.0103)	(0.0186)			(0.0009)	(0.0025)	(0.0006)	(0.0008)		
Per capita income support ($)	0.9513***	-1.1701***	0.3715***	-0.0759	**	**	0.0175	-0.0166	0.0007	0.0042		
	(0.2561)	(0.3934)	(0.1413)	(0.1816)			(0.1128)	(0.1158)	(0.0074)	(0.0085)		
Unemployment rate	50.5143***	-37.7818**	22.6433***	2.1099	**	**	-4.0593	3.3928	0.024	0.1359		
	(11.7604)	(14.9656)	(6.9)	(6.6985)			(3.9979)	(8.139)	(0.1356)	(0.138)		
Percent of farm emp.	-71.8352***	-58.8564**	-13.9938	-12.6009	**	*	4.4850*	8.1627*	0.2453	0.2226	*	*
	(26.3745)	(23.3402)	(11.2357)	(11.3248)			(2.4417)	(4.8077)	(0.2786)	(0.2422)		
Percent of manufacturing emp.	25.5880***	26.8409***	7.8751	10.0360*	**	*	-5.3357	-3.0828	0.1458	0.1800*		
	(6.8205)	(7.4118)	(5.2755)	(5.171)			(5.575)	(3.606)	(0.136)	(0.0927)		
Percent of government emp.	30.0925**	20.7857	17.9701*	16.5723*			-21.6997	-21.2504	0.2676	0.2511		
	(12.8491)	(12.9508)	(9.1898)	(9.5174)			(15.9216)	(15.4936)	(0.3519)	(0.3838)		
Voter participation rate (%)	-14.7929*	-3.9546	-3.4463	-0.4927			4.144	0.3537	0.2027	-0.0385		
	(7.986)	(12.7385)	(7.1129)	(8.1372)			(5.7297)	(4.0898)	(0.2324)	(0.3455)		
Percent of vote for Rep.	-5.8727	-11.9812	0.0333	-1.8594			-0.8076*	-4.2168	-0.1163**	-0.2698***		
	(8.362)	(13.1805)	(4.0976)	(5.7463)			(0.4367)	(2.875)	(0.0577)	(0.0824)		
Voting margin (%)	4.9296	4.5836	0.6551	0.4426			0.7939*	1.0065	0.0402*	0.0540**		
	(3.4164)	(3.6422)	(1.5309)	(1.5415)			(0.4615)	(0.6304)	(0.0235)	(0.0223)		
Constant	478.2293	2400*	217.9828	70.258			192.6853*	550.8243	-22.1091	-16.043		
	(704.2858)	(1300)	(512.6953)	(703.9815)			(100.6886)	(345.6266)	(26.203)	(20.6974)		
Year fixed effect	No	Yes	No	Yes			No	Yes	No	Yes		
Num. of obs.	41,554		6,373				41,554		6,373			
Num. of groups	2,600		422				2,600		422			

* p<.1; ** p<.05; *** p<.01

Table B.8 Determinants of funding on economic development programs per person in poverty by agency

Variables	Project grant						Formula grant					
	Non-persistent poverty counties		Persistent poverty counties		Structural break		Non-persistent poverty counties		Persistent poverty counties		Structural break	
Poverty rate (%)	-10.1426***	-10.7471***	-6.9659***	-5.2068**			0.9977	-3.1649	-0.4119**	-0.3084**		
	(2.1952)	(2.3038)	(2.192)	(2.543)			(3.1953)	(5.7302)	(0.1899)	(0.1503)		
Median household income ($)	0.0036**	-0.0046**	0.0029	0.0036		**	0.0016*	-0.0022	-0.0001	0	*	
	(0.0019)	(0.0023)	(0.0022)	(0.0031)			(0.0009)	(0.0025)	(0.0002)	(0.0002)		
Per capita income support ($)	0.3507***	-0.1102	0.1250***	-0.0329	**		0.0105	-0.0179	-0.0014	-0.0014		
	(0.0894)	(0.0718)	(0.0406)	(0.0532)			(0.1128)	(0.1157)	(0.0023)	(0.0032)		
Unemployment rate	5.8486**	-7.9917**	3.6911*	-2.227			-3.8972	3.4368	0.1272	0.122		
	(2.3865)	(3.9153)	(2.1445)	(1.9217)			(3.9894)	(8.1287)	(0.0857)	(0.0916)		
Percent of farm emp.	-18.9681*	-15.9149**	-1.9845	-2.1803			4.2316*	7.9074*	-0.0282	0.0118	*	*
	(10.0701)	(8.0781)	(3.0801)	(3.2555)			(2.42)	(4.7952)	(0.0655)	(0.0744)		
Percent of manufacturing emp.	3.5260*	3.4463***	3.5324**	3.6501**			-5.3431	-3.0813	-0.0643	-0.0106		
	(1.8576)	(1.2548)	(1.3746)	(1.6749)			(5.5718)	(3.6022)	(0.0548)	(0.0436)		
Percent of government emp.	10.1650***	7.9465***	5.5142**	5.4198**			-21.5788	-21.1403	0.0041	-0.0288		
	(2.6808)	(2.6792)	(2.17)	(2.2231)			(15.9079)	(15.4805)	(0.0687)	(0.0648)		
Voter participation rate (%)	-1.6281	-5.0683	-0.642	-1.6004			3.9332	0.3127	0.0358	0.0805		
	(2.0429)	(3.7502)	(2.111)	(2.3708)			(5.7269)	(4.0883)	(0.2131)	(0.2392)		
Percent of vote for Rep.	-5.3787	-7.1818	-0.813	-2.0931			-0.7532*	-4.1546	0.0096	-0.0167	*	
	(4.0585)	(6.4671)	(0.939)	(1.5113)			(0.4321)	(2.8725)	(0.0418)	(0.047)		
Voting margin (%)	1.474	1.2306	-0.1198	-0.2037			0.7871*	1.0085	0.0173*	0.0205**	*	
	(1.4262)	(1.536)	(0.5647)	(0.5521)			(0.4595)	(0.629)	(0.0094)	(0.0101)		
Constant	163.719	881.8327	-26.2083	158.2196			209.7872**	551.5508	15.4344	8.8782		
	(257.5091)	(562.3013)	(112.7216)	(144.0834)			(100.762)	(345.472)	(12.7949)	(11.2076)		
Year fixed effect	No	Yes	No	Yes			No	Yes	No	Yes		
Num. of obs.	41,554		6,373				41,554		6,373			
Num. of groups	2,600		422				2,600		422			

* p<.1; ** p<.05; *** p<.01

Table B.9 Determinants of funding on economic development programs per person in poverty by the type of grant

BIBLIOGRAPHY

Aaronson, D. and B.J. Phelan (2017), "Wage Shocks and the Technological Substitution of Low-wage Jobs," *The Economic Journal*, 129(617), 1-34.

Abbott, R. and B. Bogenschneider (2018). "Should Robots Pay Taxes: Tax Policy in the Age of Automation," *Harvard Law & Policy Review*, 12, 145.

Acemoglu, D. (2002), "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, 40(1), 7-72.

Acemoglu, D. and D. Autor (2011), "Skills, Tasks, and Technologies: Implications for Employment and Earnings," Handbook of Labor Economics, 4, 1043--1171.

Acemoglu, D. and P. Restrepo (2017), "Robots and Jobs: Evidence from US Labor Markets," NBER Working Paper No. 23285.

Acemoglu, D. and P. Restrepo (2018a), "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 108(6), 1488-1542.

Acemoglu, D. and P. Restrepo (2018b), "Low-Skill and High-Skill Automation," *Journal of Human Capital*, 12(2), 204-232.

Acemoglu, D. and P. Restrepo (2018c), "Artificial Intelligence, Automation and Work," NBER Wokring Paper No. 24196.

Acemoglu, D. and P. Restrepo (2019), "Automation and New Tasks: How Technology Displaces and Reinstates Labor," *Journal of Economic Perspectives*, 33(2), 3-30.

Acemoglu, D. and P. Restrepo (2020), "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 128(6).

Acemoglu, D., D. Autor, D. Dorn, G.H. Hanson, and B. Price (2016), "Import Competition and the Great US Employment Sag of the 2000s," *Journal of Labor Economics*, 34 (S1), S141-S198.

Adem (2018), "Distributional Effect of Import Shocks on the British Local Labour Markets," Working paper

Arntz, M., T. Gregory, and U. Zierahn (2016), "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis," OECD Social, Employment, and Migration Working Papers, 189.

Atlas, C.M., T.W. Gilligan, R.J. Hendershott, and M.A. Zupan (1995), “Slicing the Federal Government Net Spending Pie: Who Wins, Who Loses, and Why,” *The American Economic Review*, 85(3), 624-629.

Autor, D.H. and D. Dorn (2013), "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economics Review*, 103, 1553-1597.

Autor, D.H., D. Dorn, and G. Hanson (2013), "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103(6), 2121-2168.

Autor, D.H., D. Dorn, and G. Hanson (2019), "When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men," *American Economic Review: Insights*, 1(2): 161-178.

Autor, D.H., F. Levy, and R.J. Murnane (2003), "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118(4), 1279-1333.

Barbieri, L.M., C. Piva, C. Mussida, and M. Vivarelli (2019), "Testing the Employment Impact of Automation, Robots and AI: A Survey and Some Methodological Issues," IZA Discussion Paper No. 12612

Berry, C.R., B.C. Burden, and W.G. Howell (2010), “The President and the Distribution of Federal Spending,” *American Political Science Review*, 104(04), 783-799.

Bessen, J. (2018), "AI and Jobs: The Role of Demand," NBER Working Paper No. 24235

Bitler, M, H. Hoynes, and E. Kuka (2017), "Do In-Work Tax Credits Function as a Safety Net?" *Journal of Human Resources*, 52(2), 319-350.

Chetverikov, D., B. Larsen, and C. Palmer (2016), "IV Quantile Regression for Group-level Treatments, with an Application to the Distributional Effects of Trade," *Econometrica*, 84(2), 809-833.

Chiacchio, F., G. Petropoulos, and D. Pichler (2018), "The Impact of Industrial Robots on EU Employment and Wages: A Local Labour Market Approach," Bruegel Working Paper.

Choi, J. and M. Xu (2019), "The labor market effects of the China Syndrome: Evidence from South Korean manufacturing," *World Econ*. pp 1-49.

Chubb, J.E. (1985), “The Political Economy of Federalism,” *American Political Science Review*, 79(04), 994-1015.

Clyburn, J.E. (2014), "Developing the Will and the Way to Address Persistent Poverty in America," *Harvard Journal on Legislation*, 51, 1.

Cowan, T. (2016), "An Overview of USDA Rural Development Programs," Congressional Research Service report, RL31837.

Dalaker, J. (2017), "The 10-20-30 Rule and Persistent Poverty Counties," Congressional Research Service report, R44748.

Dauth, W., S. Findeisen, J. Südekum, and N. Woessner (2017), "German Robots-the Impact of Industrial Robots on Workers," IAB-Discussion Paper, No. 30/2017

Deemer, D.R. (2015), "Spatial Inequalities in the Fiscal Distribution of the US Welfare State (Doctoral dissertation, The Ohio State University).

Dewar, M.E. (1998), "Why State and Local Economic Development Programs Cause So Little Economic Development," *Economic Development Quarterly*, 12(1), 68-87.

Dorn, D. (2009), "Essays on Inequality, Spatial Interaction, and the Demand for Skills," Dissertation University of St. Gallen no. 3613.

Douglas, S. and W.R. Reed (2016), "A Replication of "The Political Determinants of Federal Expenditure at the State Level" (Public Choice, 2005)," *Public Finance Review*, 44(4), 549-558.

Downey, M. (2016), "Partial Automation: Routine-Biased Technical Change, Deskilling, and the Minimum Wage," Working Paper.

Feenberg, D. and E. Coutts (1993), "An Introduction to the TAXSIM Model," Journal of Policy *Analysis and management*, 12(1), 189-194.

Feng, A. and G. Graetz (2015), "Rise of the Machines: The Effect of Labor-Saving Innovations on Jobs and Wages," IZA Discussion Paper No. 8836.

Figueiredo, E. and L.R. Lima (2020), "Do Economic Integration Agreements Affect Trade Predictability? A Group Effect Analysis," *Canadian Journal of Economics*, 53(2), 637-664.

Fleck, R.K. (1999), "The Value of the Vote: A Model and Test of the Effects of Turnout on Distributive Policy," *Economic Inquiry*, 37(4), 609-623.

Frey, C. B. and M. A. Osborne (2017), "The Future of Employment: How Susceptible are Jobs to Computerisation?" *Technological Forecasting and Social Change*, 114, 254-280.

GAO (2011), "List of Selected Federal Programs that Have Similar or Overlapping Objectives, Provide Similar Services, or Are Fragmented Across Government Missions," GAO-11-474R.

GAO (2012), “The Distribution of Federal Economic Development Grants to Communities with High Rates of Poverty and Unemployment,” GAO-12-938R.

Goos, M. (2018), "The Impact of Technological Progress on Labour Markets: Policy Challenges," *Oxford Review of Economic Policy*, 34, 362-375.

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

Guerreiro, J., S. Rebelo, and P. Teles (2017), "Should Robots Be Taxed?" NBER Working Paper No. 23806.

Hall, J.L. (2010), “The Distribution of Federal Economic Development Grant Funds: A Consideration of Need and the Urban/Rural Divide,” *Economic Development Quarterly*, 24(4), 311-324.

Hausman, J. AND W. Taylor (1981), "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377-1398.

Hoover, G.A. and P. Pecorino (2005), “The Political Determinants of Federal Expenditure at the State Level,” *Public Choice*, 123(1-2), 95-113.

Hoynes, H.W. and J. Rothstein (2016), "Tax Policy Toward Low-Income Families," NBER Working Paper No. 22080.

Hoynes, H.W. and J. Rothstein (2019), "Universal Basic Income in the US and Advanced Countries," NBER Working Paper No. 25538.

Islam, T.M., J. Minier, and J.P. Ziliak (2015), “On Persistent Poverty in a Rich Country,” *Southern Economic Journal*, 81(3), 653-678.

Isserman, A. and T. Rephann (1995), “The Economic Effects of the Appalachian Regional Commission: An Empirical Assessment of 26 Years of Regional Development Planning,” *Journal of the American Planning Association*, 61(3), 345-364.

Jones, M.R. (2014), "Changes in EITC Eligibility and Participation, 2005--2009," Center for Administrative Records Research and Applications Working Paper, 04.

Jones, M.R. (2015), "The EITC over the Business Cycle: Who Benefits?" CARRA Working Paper.

Korinek, A. and J. E. Stiglitz (2017), "Artificial Intelligence and Its Implications for Income Distribution and Unemployment," NBER Working Paper No. 24174.

Lake, M. and D.L. Millimet (2018), "Good Jobs, Bad Jobs: What's Trade Got To Do With It?", Working Paper.

Larcinese, V., L. Rizzo, and C. Testa (2013), “Why Do Small States Receive More Federal Money? US Senate Representation and the Allocation of Federal Budget,” *Economics & Politics*, 25(3), 257-282.

Leigh, N.C. and B.R. Kraft (2017), "Emerging robotic regions in the United States: insights for regional economic evolution," *Regional Studies*, 52, 804-815

Levitt, S.D. and J.M. Snyder Jr (1995), “Political Parties and the Distribution of Federal Outlays,” *American Journal of Political Science*, 958-980.

Lordan, G. and D. Neumark (2018), "People Versus Machines: The Impact of Minimum Wages on Automatable Jobs," *Labour Economics*, 52, 40-53.

Markusen, A. and A. Glasmeier (2008), “Overhauling and Revitalizing Federal Economic Development Programs,” *Economic Development Quarterly*, 22(2), 83-91.

McAfee, A. and E. Brynjolfsson (2016), "Human Work in the Robotic Future: Policy for the Age of Automation," *Foreign Affairs*, 95(4), 139.

Mencken, F.C. (2000), “Federal Spending and Economic Growth in Appalachian Counties,” *Rural Sociology*, 65(1), 126-147.

Mencken, F.C. and C.M. Tolbert (2005), “Federal Public Investment Spending and Economic Development in Appalachia,” *Rural sociology*, 70(4), 514-539.

Nichols, A. and J. Rothstein (2015), "The Earned Income Tax Credit," NBER Working Paper No. 21211.

Okay, T. and K. Yamadaz (2019), "Heterogeneous Impact of the Minimum Wage: Implications for Changes in Between- and Within-group," Working paper

Plueger, D. (2009), "Earned Income Tax Credit Participation Rate for Tax Year 2005." IRS Research Bulletin.

Poliquin, C.W. (2020), "The Effect of the Internet on Wages," Working paper

Rupasingha, A. and S.J. Goetz (2007), “Social and Political Forces as Determinants of Poverty: A Spatial Analysis,” *The Journal of Socio-Economics*, 36(4), 650-671.

Schneider, M. and B.M. Ji (1990), “The Political Economy of Intergovernmental Grant Seeking: Targeting and Suburbs, 1977 and 1982,” *American Journal of Political Science*, 408-420.

Taddy, M. (2018), "The Technological Elements of Artificial Intelligence," NBER Working Paper No. 24301.

Thuemmel, U. (2018), "Optimal Taxation of Robots, CESifo Working Paper, No. 7317.

Tolbert, C.M. and M. Sizer (1996), "U.S. Commuting Zones and Labor Market Areas: A 1990 Update," Rural Economy Division, Economic Research Service, U.S. Department of Agriculture.

Tolbert, C.M. and M.S. Killian (1987), "Labor Market Areas for the United States," Agriculture and Rural Economy Division, Economic Research Service, U.S. Department of Agriculture.

U.S. Census Bureau (2011), “Consolidated Federal Funds Report for Fiscal Year 2010 – State and County Areas.”

Wang, Z., S. Wei, X. Yu, and K. Zhu (2018), "Re-examining the Effects of Trading with China on Local Labor Markets: A Supply Chain Perspective," NBER Working Paper No. 24886