The London School of Economics and Political Science

Essays on Outsourcing, Automation and Economic Justice

Daniel Chandler

A thesis submitted to the Department of Economics of the London School of Economics and Political Science
for the degree of Doctor of Philosophy

London, June 2025

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare that my thesis consists of approximately 47,900 words.

Statement of co-authored work

I confirm that Chapter 2 was jointly co-authored with Aniket Baksy and Peter Lambert. All three authors contributed to the data construction, empirical design, and analysis of the results. I drafted the majority of the text.

Acknowledgements

I am deeply grateful to my supervisor, Tim Besley, for his insight and encouragement, and for providing me with a truly inspiring model of an economist and intellectual. I would also like to thank my advisor John Van Reenen; Stephen Machin, who supervised my MRes theses; and Nava Ashraf, whose feedback and encouragement have been invaluable. I have benefited hugely from working with my two co-authors Aniket Baksy and Peter Lambert. Special thanks go to my family and especially my partner Simon, for his patience, love, and unwavering support.

I would like to gratefully acknowledge the funding I received from the ESRC through a Doctoral Studentship; a grant from the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at LSE for purchasing the dataset used in my second chapter; and finally the Open Society Foundations through their generous support of the LSE Programme on Cohesive Captialism.

Abstract

This thesis examines how domestic outsourcing and automation shape labour market outcomes, and how the philosophy of John Rawls can inform normative economics and economic policy. It is structured around three distinct essays.

The first essay provides the first empirical analysis of the impact of domestic outsourcing on workers in the UK, using matched worker-firm data and focusing on cleaners and security guards. I estimate wage penalties of 2.8% and 5.6% respectively, and show that contractor firms pay systematically lower AKM wage premiums than in-house employers. These differences reflect both lower rents per worker and, for security workers, weaker rent-sharing—highlighting the role of firm-level practices in driving wage inequality.

The second essay uses a novel proprietary survey of UK manufacturing sites between 2005 and 2023 to examine the employment effects of CNC machine tools and industrial robots. I show that both technologies are associated with significant increases in employment at adopting plants – between 6% and 9% after four years – compared to a control group of non-adopting plants. I also find positive spillovers on employment among industry competitors and a positive aggregate impact on industry-level employment, challenging the view that automation necessarily displaces workers.

The final essay argues that the standard economic interpretation of John Rawls—as an advocate of redistribution justified by a maximin social welfare function—misrepresents the spirit and substance of his theory. I argue that Rawls offers a richer conception of justice, grounded in reciprocity and focused on access to economic power and the social bases of self-respect. I outline the core features of a Rawlsian normative economics—plural in values, grounded in measurable primary goods, and psychologically realistic; and propose a policy agenda centred on predistribution and meaningful work.

Together, these essays contribute new evidence on how firms and technologies shape labour market inequality, and offer a rich non-welfarist framework for rethinking the goals and design of economic policy.

Contents

List of Figures

List of Tables

Chapter 1

Where You Work or What You Do? Domestic Outsourcing and the Wages of Low-Skilled Service Workers in the UK

ABSTRACT

Recent decades have seen a growing tendency across advanced economies for firms to "outsource" parts of their production process to specialised third-party companies. This paper uses matched worker-firm data to provide the first empirical analysis of the impact of domestic outsourcing on workers in the UK, focusing on two of the occupations most widely associated with this change: cleaners and security guards. First, using my preferred (fixed effects) specification, I estimate a wage penalty of 2.8% for cleaners and 5.6% for security workers. Second, I show that contractor companies pay systematically lower AKM wage premiums than companies that employ cleaners and security workers in-house. Finally, I argue that these low wage premiums reflect the fact that contractor firms have significantly lower rents per worker and, for security workers, a lower propensity to share rents with their employees.

Keywords: domestic outsourcing; inequality; wage determination; rent-sharing.

1.1 Introduction

Recent decades have seen a growing tendency across advanced economies for firms to "outsource" parts of their production process to specialised third-party contractors (Zwysen, 2024). While the process of *international* outsourcing (or "offshoring") has been widely studied by trade economists, until recently relatively little attention has been paid to *domestic* outsourcing – the purchase of business services from firms within the same country – even though it accounts for the lion's share of total trade in intermediate inputs.[1] This paper uses matched worker-firm data to provide the first empirical analysis of the impact of domestic outsourcing on workers in the UK, focusing on two of the occupations most widely associated with this change: cleaners and security guards. [2]

In the text-book model of the competitive labour market, domestic outsourcing should have no impact on wages. In this framework, wages are determined by the aggregate supply and demand for workers with a particular set of skills; and since outsourcing simply reallocates workers from one firm to another, it should not affect their wages. However, there is now a large literature documenting the important role that firms play in setting wages (Card et al., 2018). If anything, which company you work for is becoming more important over time: recent decades have seen a large increase in differences in average pay between firms, which in turn has been an important force increasing overall earnings inequality (Card et al., 2018; Song et al., 2019).[3] In a world where firms frequently share rents with their employees in the form of wage premiums, profitable companies may be able to cut costs by outsourcing non-core activities to contractors who often operate in fiercely competitive markets, and pay lower wages.

There are two key challenges in assessing the impact of domestic outsourcing on wages, both of which can be overcome using matched worker-firm panel data, as in this study. The first is simply to identify both in-house and

[1] Around 85% of UK intermediate service inputs are purchased domestically. See Abramovsky and Griffith (2006), and *UK Supply and Use Tables: 2022 Edition*. London: Office for National Statistics, 2025.

[2] Abramovsky and Griffith (2006) look at the effect of ICT on UK firms' choices over whether to produce in-house or to outsource services, and over the location of the activity (domestic or foreign). They do not examine the impact of these choices on wages.

[3] Song et al. (2019) find that two-thirds of the rise in earnings inequality in the USA between 1979 and 2013 was due to the growing dispersion of average earnings between firms.

outsourced workers within a given occupation. Following Dube and Kaplan (2010), I define outsourced workers using a combination of individual occupation and firm industry codes: outsourced cleaners are cleaners whose employer is classified as a "cleaning services" firm, while in-house cleaners are defined as cleaners employed in other industries. The second challenge is to isolate the impact of working for a specialised contractor from other differences in skills and productivity between in-house and outsourced workers. This is possible thanks to the availability of panel data on workers, allowing me to follow workers who move from being in-house to outsourced (and vice versa) while remaining within the same narrow occupational category.

This paper makes four contributions. First, I document an increase in the share of both cleaners and security workers who are outsourced: the proportion of cleaners employed by cleaning services firms increased from 18% in the late 1990s to around 32% in 2015, while the share of security guards employed by security services firms grew from 36% to 50% over the same period.

Second, I show that outsourced cleaners and security workers experience a wage penalty relative to their in-house counterparts that cannot be attributed to differences in worker-level characteristics. In my preferred specification with worker-level fixed effects, I estimate an average wage penalty of 2.8% for cleaners and 5.6% for security workers. I show that the overall wage penalty for cleaners and security workers can be broken down into a part that reflects the loss of public sector wage premiums, and a further penalty associated with working for a private contractor compared to being in-house within another private sector firm.

These wage penalties are prima facie evidence of firm-level differences in pay. My third contribution is to examine this directly by estimating firm-specific wage premiums using the framework developed by Abowd et al. (1999) (henceforth AKM). I show that the average wage premium in cleaning services firms is 11% lower than for firms that employ cleaners in-house, and 14% lower for firms that employ outsourced security workers. While the standard AKM framework assumes that companies pay the same premium to all their workers, I test this assumption by estimating wage premiums separately using just cleaners and security workers. I find that the wage premium paid to these workers is lower than the premium paid to employees as a whole, suggesting that companies are able to offer less favourable terms to certain groups of workers even when they are employed in-house.

Finally, I match worker data on wages and outsourcing status to firm-level financial data to shed light on why contractors pay systematically lower wages. One hypothesis is that these companies are less inclined to share profits with their workers reflecting, for example, lower unionisation rates. Alternatively, contractor companies may simply have fewer profits to share, since they sell generic services for which competition is fierce and typically based on price. Using Gross Value Added per employee as a proxy for firm-level rents, I find that both explanations have some role to play: contractor firms have significantly lower rents per worker than companies that employ cleaners and security workers in-house; while security contractors also have lower estimated rent-sharing elasticities.

1.1.1 Conceptual framework

The starting point for this paper is that – in contrast to the textbook model of the competitive labour market – firms play an important role in wage-setting. This claim is supported by a large empirical literature which shows that workers with similar observable characteristics, and indeed even the same individuals, often have very different earnings depending on where they work (Card et al., 2018; Song et al., 2019). This in turn gives rise to the possibility that domestic outsourcing, which reallocates workers from one type of firm to another, may affect wages.

The idea that firms matter for wages is consistent with several related theoretical frameworks in which labour markets are characterised by frictions such as search costs, heterogeneous preferences and imperfect information, which prevent perfect competition and give rise to rents in the employment relationship (Manning, 2011). Different models offer competing accounts of how those rents are allocated, including bargaining models, where wages are the outcome of explicit or implicit negotiations between workers (or unions) and firms over the division of rents into profits and wage mark-ups (Booth, 1994); monopsony models, in which firms face an upward-sloping labour supply curve — for example, due to limited worker mobility — which gives them wage-setting power and allows them to pay wages below the marginal product of labour (Manning, 2003); and efficiency wage and incentive pay models, where companies voluntarily pay above-market wages to elicit greater effort, reduce shirking, or lower turnover (Akerlof and Yellen, 1990; Shapiro and Stiglitz, 1984). While these models rest on different mechanisms and assumptions, they all imply that wages reflect firm-level characteristics, and specifically a positive relationship between wages and firm-level rents (Bell et al., 2024).

With this in mind, we can distinguish two broad mechanisms through which outsourcing might affect pay. First, contractor firms may be less likely to share rents with their workers. This could reflect weaker collective bargaining institutions, different wage-setting norms, or greater monopsony power. Second, contractor firms may simply generate less surplus to begin with, since they often operate in highly competitive, low-margin sectors, providing standardised services at scale with limited pricing power (Weil, 2014). Bilal and Lhuillier (2021) develop a structural model designed to capture both mechanisms: contractor firms are less productive, and hence have lower rent levels; and they have greater monopsony power, and therefore a lower propensity to share rents with workers. Their empirical results confirm that both forces are at play. I approach the same question in a more direct, reduced-form way, finding that differences in rent levels appear to be the main explanation for lower wages among outsourced workers.

While my focus is on the effects of outsourcing, it is worth briefly reflecting on why outsourcing has become more prevalent. The decision to outsource is ultimately a "make-or-buy" choice: firms internalise production when contracting is costly, and outsource when markets can provide the same service more efficiently (Williamson, 1985; Abraham and Taylor, 1996). There is some evidence that the recent increase in outsourcing reflects the diffusion of information and communication technologies (ICT), which make it easier to coordinate and monitor tasks performed outside the firm (Bergeaud et al., 2024); as well as growing competitive pressure to reduce costs, coupled with the diffusion of business strategies focused on core competencies (Weil, 2014). Although I do not attempt to adjudicate among these explanations, they provide important context for understanding the recent increase in domestic outsourcing.

1.1.2 Relation to existing literature

This paper contributes to a small but growing empirical literature that documents the rise of domestic outsourcing across a range of advanced economies, and estimates the impact on the wages of low-skilled service workers.

Early papers on this topic include Abraham (1990), which found that low-skilled outsourced workers tend to have lower wages than their in-house counterparts; and Segal and Sullivan (1997), who found that temp agency workers have lower wages, fewer benefits, and lower rates of unionisation. Dube and Kaplan (2010) is the first paper to estimate wage penalties using panel data, focusing on cleaners and security guards in the US. In the specification closest to this paper, focusing on within-occupation switchers, they estimate a wage penalty of around 6.5% for cleaners and 11.5% for security guards. Other recent papers have followed a broadly similar approach in different contexts: Goldschmidt and Schmieder (2017) estimate an average wage penalty of 9% for food, cleaners, security guards and logistics (FCSL) workers in Germany; Bergeaud et al. (2024) find a 4% wage penalty for a similar group of occupations in France; Bilal and Lhuillier (2021), also using French data, estimate that contractor firms providing FCSL services pay 12% lower AKM wage premiums on average; while Drenik et al. (2023) estimate a 14% wage penalty for temp agency workers in Argentina. While most studies focus on wages, Goldschmidt and Schmieder (2017) find that outsourced workers also have higher rates of job separation than otherwise similar in-house employees; while Bilal and Lhuillier (2021) find that contractor firms have higher rates of separation in general, suggesting that outsourcing is associated with lower job quality more broadly.

As in this study, most papers have used fixed effects designs to estimate the average difference in wages between in-house and outsourced workers. Goldschmidt and Schmieder (2017) combine this with an event-study following workers who experience an "on-site" outsourcing event, where a large employer spins out a group of workers providing a particular service such as cleaning to a legally separate business in roughly the same location.[4] This design allows for a cleaner identification of the impact of outsourcing, since in addition to following the same individuals over time, these individuals are (presumably) doing almost exactly the same jobs. However, they may not be representative of the wider impact of domestic outsourcing, since most people working for contractor companies will have joined those companies directly rather than as the result of an on-site outsourcing event. In practice,

[4]Deibler (2022) uses the same German events on workers who remain at the lead firm, and finds that remaining workers see an increase in wages, and a lower probability of separation, albeit only in companies covered by a collective bargaining agreement.

Goldschmidt and Schmieder find that the wages of outsourced workers fall by about 10% in the ten years following an on-site outsourcing event, relative to a matched control group who were not outsourced — strikingly similar to the 9% wage penalty which they estimate using the approach adopted in this paper.[5]

My contribution to this literature is threefold. First, I confirm the central finding that outsourced workers suffer a wage penalty in the UK context; and that this wage penalty is associated with a loss of firm-specific (AKM) wage premiums. While previous studies have focused exclusively on wages, I show that outsourced workers also receive substantially lower pension contributions, meaning other studies are likely to have underestimated overall negative impact on compensation. Second, I decompose the aggregate wage penalty into a component that reflects the loss of public sector wage premiums, and a component associated with working for a contractor firm compared to other private sector employers, something not previously studied. Finally, mine is the first paper to directly examine the relationship between the outsourcing wage penalty and firm-level rents.

This paper also contributes to a wider literature on the growing importance of firms in setting wages, and how this can help explain recent aggregate trends in inequality. Studies in countries including the USA (Barth et al., 2016; Song et al., 2019) and Germany (Card et al., 2018) have shown that an increasing share of wage inequality reflects differences between firms rather than within them; alongside an increase in assortative matching, where low wage workers are increasingly concentrated into low-wage firms. In the UK context, Loecker et al. (2024) document rising dispersion in firm productivity, wages, and markups, especially in the upper tail of the distribution. The rise of domestic outsourcing is likely to be a key driver of increasing occupational segregation and assortative matching (Handwerker, 2023), and hence may help explain rising between-firm inequality.

Sitting behind these results about firms and inequality are recent developments in the estimation of firm-specific (AKM) wage premiums and rent-sharing elasticities (Kline, 2024), which have shaped the empirical approach in

[5] Daruich et al. (2024) also follow individual workers who are contracted out, using Italian data that records whether job separations are the result of outsourcing. They estimate a relatively small impact on wages, alongside a much larger impact on employment: outsourced workers are 22 percentage points less likely to be employed in the year after being outsourced, and this negative effect remains significant at almost 8 percentage points five years later.

this paper and which I will come back to in Section 6. Early papers using this framework tended to focus on summarising the overall importance of firms in explaining wage inequality, using the canonical variance decomposition of log wages into parts attributable to firm effects, worker effects and the sorting of high wage workers into high wage firms. More recently, a growing number of papers have sought to explore the relationship between estimated firm fixed effects and firm-level characteristics including productivity, firm size and workplace amenities (Card et al., 2016; Bloom et al., 2018; Sorkin, 2018). These studies have identified positive hockey-stick shaped relationship between firm wage premiums and log value added per worker in a range of countries including Portugal (Card et al., 2016), Germany (Bruns, 2019), France (Coudin et al., 2018) and Canada (Li et al., 2023). I contribute to this literature by showing that contractor firms pay systematically lower wage premiums, and that this likely reflects the fact that these firms have low value added per worker.

1.2 Data

1.2.1 Worker-level data

This paper uses the Annual Survey of Hours and Earnings (ASHE) between 2002 and 2015.[6] ASHE is a longitudinal 1% sample of UK employees based on the final two digits of an individual's National Insurance (NI) number (self-employed individuals are not included). Compliance with ASHE is a legal requirement. Since information is provided by employers rather than employees ASHE suffers from less measurement error than the major alternative worker-level survey for the UK, the Labour Force Survey.

Relevant variables available in ASHE include: various measures of earnings, pension contributions, hours of work, whether an individual is on a temporary contract, detailed industry and occupation classifications, region of work, whether an individual works in the public sector, whether they are covered by a collective agreement and a unique identifier for each firm. I use hourly wages (excluding overtime) as my measure of earnings throughout this

[6] Office for National Statistics. (2025). *Annual Survey of Hours and Earnings, 1997-2024: Secure Access*. [data collection]. 26th Edition. UK Data Service. SN: 6689, DOI: http://doi.org/10.5255/UKDA-SN-6689-25. Note that although earnings data is available from 1997, firm-level identifiers are only available from 2002 and hence I restrict all my analysis to this sub-sample.

paper, adjusted to 2016 prices using the Consumer Prices Index. The main limitation of ASHE is that other than age and sex, it has no further information about individual characteristics.

1.2.2 Firm-level data

I match ASHE to the Annual Respondents Database X (ARDx), which is a representative annual survey of UK firms in the non-financial business sector (i.e. excluding the financial sector and public sector organisations), covering the period 1998 to 2014.[7] Although the ARDx is compiled from a number of underlying datasets that have changed over time, the variables have been harmonised over the entire period.

The ARDx is a stratified random sample of UK firms. All firms with more than 250 employees are sampled each year. Firms with between 10 and 250 employees that are selected are typically surveyed for two consecutive years and then exempt for a further two years. Firms with fewer than 10 employees are only sampled for a single year at a time, and then not resampled for at least three years. The implication of this sampling design is that workers in ASHE are more likely to find a match in the ARDx if they work in larger firms.[8]

The ARDx contains both employment and financial information. Employment information is limited but includes the overall number of employees and how this breaks down by full- and part-time status, and by gender. The financial variables are much more detailed. Of particular importance for my analysis are measures of Gross Value Added which I use as a proxy for firm rents.

1.2.3 Identifying outsourced workers

Conceptually, I define an outsourced worker as someone who is employed by a "contractor" (businesses services) firm to provide intermediate services to a "lead" company; and when, at least in principle, that person could provide the

[7] Office for National Statistics (ONS). (2024). *Annual Respondents Database X, 1997-2020: Secure Access.* [data collection]. 5th Edition. Office for National Statistics, [original data producer(s)]. Office for National Statistics. SN: 7989, DOI: http://doi.org/10.5255/UKDA-SN-7989-5.

[8] The unit of observation in the ARDx is the reporting unit (RU). An RU is the smallest collection of establishments (known as local units, or LUs) able to provide the financial information required by the underlying surveys. In practice, around 95% of RUs contain only one LU i.e. they are single-site businesses. A business or enterprise can in principle include a number of RUs, but in practice for all but the largest businesses, the RU is the same as the enterprise.

same service as a direct employee of the lead firm (Dube and Kaplan, 2010). In practice, I identify outsourced workers using a combination of industry and occupation codes, an approach first developed by Abraham (1990), and now widely used in the literature.

The first step is to identify all workers in a particular occupation, say cleaners, using detailed occupation codes. Outsourced cleaners are then defined as cleaners employed by firms whose primary activity is "cleaning services" according to the company's industry code. Cleaners who are employed by firms in other industries (e.g. retail) are defined as "in-house" workers.[9] A similar procedure allows us to identify security workers who work for firms specialising in the provision of business security services. I exclude workers employed by temp agencies form my analysis on the basis that I want to isolate the impact of outsourcing from the impact of engaging workers on a temporary basis. A full list of the industry and occupation codes which are used to identify in-house and outsourced cleaners and security workers is provided in Table A.4 of the appendix.[10]

In the analysis that follows, I distinguish between workers in the public and private sectors. This distinction relates to the identity of an individual's *employer*, rather than the identity of the company or sector where they ultimately perform their work. Unfortunately, my data contains no information about where outsourced workers actually provide their services. This means, for example, that it is not possible to compare the wages of in-house public sector workers with the wages of outsourced workers who provide contracted services specifically to public sector organisations. Note also that while in-house workers may be employed by public or private sector companies, outsourced workers are, by definition, in the private sector, since firms which sell cleaning and/or security services are exclusively private sector firms.

[9]Conceptually, my definition of outsourcing excludes the provision of cleaning and other services to final consumers. However, the occupation and industry codes for cleaning and security services do not distinguish individuals or companies that provide services to other firms from those that work for private individuals or households. In other words, my measure of outsourced cleaners may include some individuals who supply services to private households rather than to other businesses. In practice this is unlikely to be a major concern: the vast majority of security services are presumably provided to companies; and although many cleaners work for private households, most domestic cleaners are self-employed and therefore excluded from ASHE by construction.

[10]The period covered in this study (2002-2015) spans a number of changes in occupation and industry codes. Fortunately, these changes in classification did not have a substantial effect on any of the occupations or industries included in this study.

1.3 Descriptive statistics

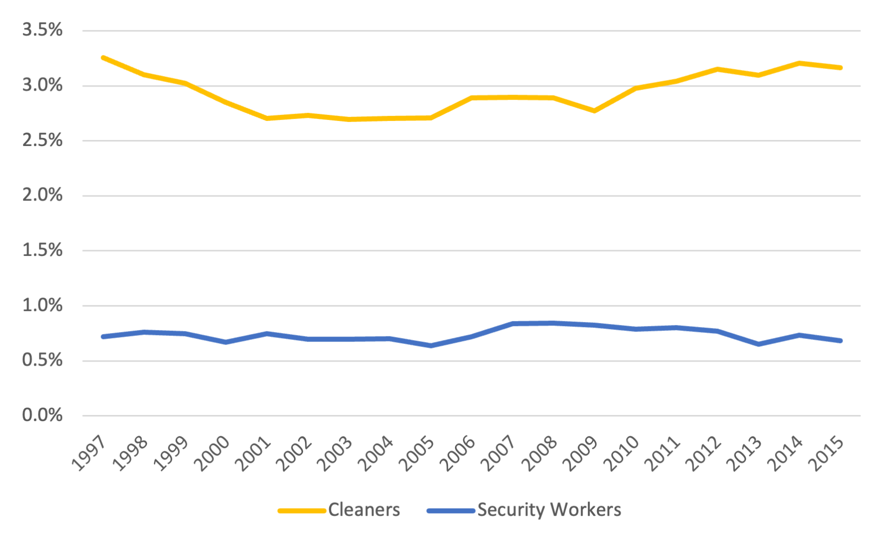
In 2015 cleaners accounted for around 3% of UK employees, while security workers accounted for around 0.75%. Figure 1 shows that the size of these occupation as a percentage of the overall workforce has been broadly stable since 1997, while Figure 2 shows that the proportion of cleaners and security guards who are outsourced has grown substantially: the share of cleaners who are outsourced has grown from 18% in the late 1990s to around 32% in 2015, while the share of outsourced security guards has grown from around 36% to 50% (though this measure is noisier, likely reflecting the smaller sample size for security workers).

Table 1.1 compares the characteristics of outsourced versus in-house cleaners and security guards, pooled over the period from 2002 to 2015 with financial values uprated to 2016 prices.[11] As expected, outsourced workers have lower hourly wages than their in-house counterparts: the raw wage penalty is 5.4% for outsourced cleaners and 22.5% for outsourced security workers. In-house and outsourced workers differ across a range of characteristics which are correlated with wages. In-house workers tend to be older, are more likely to have been in the same job for more than a year, work in larger firms, and are more likely to have their wages set by reference to a collective agreement – factors that may help explain their higher wages. A large minority of in-house cleaners and security workers (43% and 37% respectively) are public sector employees, who are generally paid significantly higher wages (see Table A2). In house cleaners are more likely to work full-time than their outsourced counterparts, though the reverse is (marginally) true for outsourced security workers. In-house workers are more likely to be female, which in general is correlated with lower wages. Interestingly, in-house cleaners and security guards are more likely to be on a temporary contracts, suggesting that outsourcing is not necessarily associated with more insecure or contingent employment.[12]

[11] As explained in Section 3.1, firm identifiers, which are necessary for estimating firm wage-premiums, are only available from 2002. I restrict the descriptive statistics to the same sample.

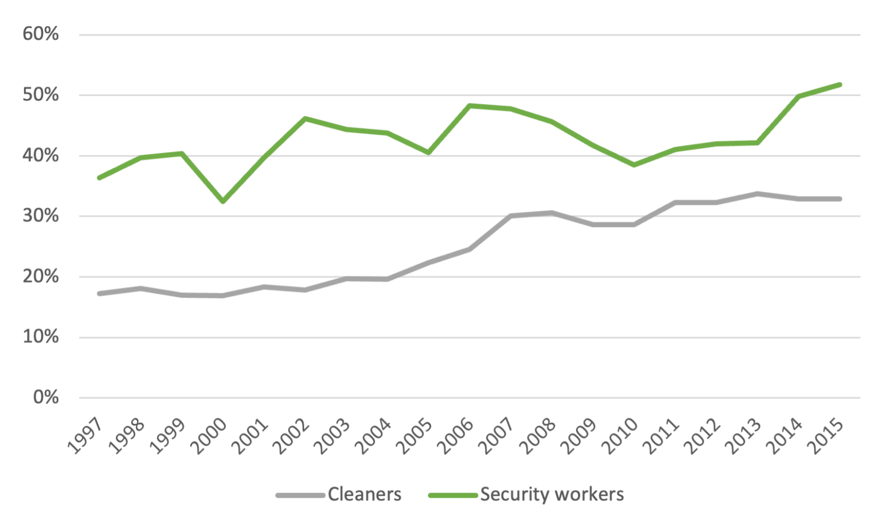
[12] Recent interest in domestic outsourcing is connected to a wider literature on the rise of "alternative" employment relationships and the "gig economy" (Polivka, 1996; Katz and Krueger, 2016; Mas and Pallais, 2016). It is important here to distinguish between two dimensions of alternative work: outsourcing and contingent work. Outsourcing encompasses arrangements where workers are employees of one firm (or self-employed), but provide services to another firm. Contingent work refers to arrangements where workers are on short-term, fixed-term or flexible employment contracts. While outsourcing and contingent work may go together, this is not necessarily the case.

Figure 1.1: Cleaners and Security Workers as Share of UK Employees



Source: ASHE.
Notes: Cleaners and security workers defined using occupation codes, as described in Appendix Table A1.

Figure 1.2: Share of Cleaners and Security Workers who are Outsourced



Source: ASHE.
Notes: Cleaners and security workers defined using occupation codes, as described in Appendix Table A1.

Table 1.1: Characteristics of In-House and Outsourced Workers

	Cleaners			Security Workers		
	In-house	Outsourced	Diff.	In-house	Outsourced	Diff.
Compensation						
Mean hourly earnings (£2016)	8.00	7.57	-0.43***	11.31	8.76	-2.54***
Fraction with employer pension contribution	0.19	0.03	-0.16***	0.39	0.07	-0.32***
Mean weekly employer pension contrib (£2016)	9.17	0.97	-8.19***	37.84	1.65	-36.20***
Individual characteristics						
Mean age	45.13	42.81	-2.32***	43.06	42.48	-0.58**
Fraction female	0.76	0.62	-0.13***	0.21	0.08	-0.13***
Fraction full time	0.31	0.28	-0.03***	0.86	0.86	0.01
Fraction in same job >1 year	0.80	0.68	-0.12***	0.83	0.76	-0.07***
Fraction temporary	0.06	0.03	-0.03***	0.05	0.03	-0.02***
Firm characteristics						
Mean firm size (employees)	10,778	6,965	-3,814***	26,063	7,272	-18,791***
Fraction public sector†	0.43	–	–	0.37	–	–
Fraction collective agreement	0.56	0.18	-0.38***	0.63	0.25	-0.38***
Observations	37,393	19,041		8,269	8,249	

Source: ASHE, 1997-2015.
Notes: * Significant at 5% level, ** at 1% level, *** at 0.1% level (two-tailed t-test).
†As explained in Section 3.2, outsourced workers are private sector employees by definition.

1.4 The impact of outsourcing on wages and compensation

1.4.1 Empirical strategy

Although outsourced cleaners and security workers are clearly paid less than their in-house counterparts, this might simply reflect differences in worker-level skills and other characteristics: in-house security guards might, for example, have more experience or training than those working for contractor companies. To isolate the impact of outsourcing on wages, I use models that exploit within-worker variation, comparing the wages of the same individuals when employed in-house and when employed by contractor firms. I do this by estimating the following regression:

$$\ln(w_{it}) = \alpha_i + x_{it}^{'}\beta + \theta_t + \gamma \text{Outsourced}_{it} + [\partial \text{Public}_{it}] + \varepsilon_{it} \tag{1.1}$$

where $\ln(w_{it})$ is the log hourly wages of individual i at time t, $x_{it}^{'}$ is a vector of time-varying individual characteristics, α_i is an individual fixed effect, θ_t is a time fixed effect capturing macroeconomic conditions that affect wages for all workers in a given period, and ε_{it} is an error term capturing other factors associated with wages. $Outsourced_{it}$ is a dummy variable indicating whether an individual is outsourced (employed by a business services firm) in period t, and $Public_{it}$ indicates whether an individual is employed by a public sector organisation. I estimate this equation separately for cleaners and security workers, since impact of being outsourced is likely to differ from one occupation to the next.

The coefficient on the outsourcing dummy (γ) is the primary object of interest. In specifications without the public sector dummy, γ can be interpreted as the average percentage wage penalty associated with being an outsourced worker in each occupation, relative to all in-house employees. Including a public sector dummy allow us to distinguish the impact of working for an outsourcing (business services) firm, relative to other private companies, from the impact of being employed in the public versus private sector. Since outsourced workers are (by definition) private sector employees, ∂ can be interpreted as the average percentage wage premium associated with working in the public sector, relative to other in-house workers. When this dummy is included, γ captures the wage penalty associated with being an outsourced worker compared to in-house workers doing the same job in the private sector.

I begin by estimating a pooled OLS specification to establish a baseline, before turning to my preferred fixed effects (within) estimator, which controls for unobserved individual heterogeneity. I include controls for sex, age, region of work, year, temporary contract and full-time status as standard; but not for firm-level characteristics such as firm size or collective pay agreements, since differences along these dimensions may be outcomes of outsourcing, the effects of which I want to capture when estimating the overall outsourcing wage penalty. Given the panel nature of my data, I cluster standard errors at the individual level.

The key advantage of the fixed effects estimator is that it controls for time-invariant worker characteristics, whether observed or unobserved — including factors such as education, ability, and work ethic — which might otherwise bias OLS estimates if they are correlated with outsourcing status. Identification using this estimator comes from cleaners and security workers who change their outsourcing status while remaining in the same occupation either side of the transition. The estimated coefficients capture the average within-worker change in wages associated with moving between in-house employment and employment at a contractor firm.

One potential concern is that in-house and outsourced workers might perform systematically different tasks, in which case wage differences could partly reflect differences in job content rather than the pure effect of outsourcing. A key advantage of focusing on cleaners and security workers is that they represent narrow occupational categories with relatively little variation in tasks across jobs, meaning we can be reasonably confident that we are estimating differences in pay for workers doing otherwise similar jobs.

1.4.2 Results: wages

The first two columns of Table 1.2 display the results of estimating equation 1.1 using pooled OLS. The estimated wage penalty is 6.4% for cleaners and 24.2% for security workers. Including a public sector dummy (column (2)) reveals a substantial public sector wage premium: wages for in-house cleaners in the public sector are 10.5% higher wages that for in-house cleaners in the private sector; while the equivalent public sector premium for security workers is

17.6%. In this specification, the coefficient on the outsourcing dummy can be interpreted at the wage penalty associated with being outsourced among private sector workers. For cleaners this penalty is 2.5%, suggesting that about half of the overall wage penalty of 6.4% reflects the loss of public sector wage premiums. For security workers, the wage penalty from working for a contractor firm compared to being in-house in the private sector is 18.3%.

Table 1.2: Effect of Outsourcing on Log Hourly Wages

	Pooled OLS		Within	
	(1)	(2)	(3)	(4)
Panel A: Cleaners				
Outsourced	-0.064***	-0.025***	-0.028***	-0.009*
	(0.003)	(0.004)	(0.004)	(0.005)
Public sector		0.105***		0.072***
		(0.004)		(0.006)
Observations	54,959	54,959	54,959	54,959
Panel B: Security Workers				
Outsourced	-0.242***	-0.183***	-0.056***	-0.051***
	(0.008)	(0.010)	(0.007)	(0.007)
Public sector		0.176***		0.055**
		(0.013)		(0.014)
Observations	16,157	16,157	16,157	16,157

Notes: Standard errors in parentheses, clustered at the individual level. * Significant at 5% level, ** at 1% level, *** at 0.1% level. In all models, the dependent variable is log hourly wages (excluding overtime) in 2016 prices. All specifications control for sex, age bands, region of work, year, temporary contract, and full-time status.

Columns (3) and (4) display the results when controlling for individual fixed effects. For cleaners the overall estimated wage penalty falls from 6.4% to 2.8%, much of which is accounted for by the loss of public sector wage premiums. For security workers, the estimated wage penalty falls to 5.6%, suggesting that unobserved individual characteristics account for most of the observed wage gap for security workers. While security workers in the public sector also receive a wage premium, the loss of these premiums only accounts for a small part of the overall wage penalty associated with working for an outsourcing company.

1.4.3 Results: pensions and total compensation

To get a full picture of the impact of being an outsourced worker on financial compensation, we must look not just at wages but at pension contributions. This is particularly important since excluding non-core workers from company-wide benefits like pensions is widely reported to be one of the key drivers of outsourcing (Weil, 2014); and yet, to the best of my knowledge, other recent studies of domestic outsourcing have ignored this question.

The descriptive statistics in Table 1.1 reveal a stark gap in pension contributions between in-house and outsourced workers. While 19% of in-house cleaners and 39% of in-house security workers receive some pension contribution from their employer, this falls to just 3% and 7% among their outsourced counterparts. Among those who do receive a pension contribution, the level of weekly pension contributions is much lower (almost negligible) among outsourced workers: the average weekly pension contribution is £9.17 for in-house cleaners and £37.84 for in-house security workers, compared to just £0.97 and £1.65 per week for outsourced cleaners and security workers, respectively. Unsurprisingly, Table A3 in the appendix shows that these differences are robust to controlling for various worker characteristics: even once we control for various worker and job-level characteristics, the odds of an outsourced cleaner receiving an employer pension contribution are just over 10% of those for an in-house cleaner; and even lower once we control for individual fixed effects. The headline figures are driven in large part by the loss of public sector pension contributions, but outsourced cleaners also have lower pension contributions than their in-house counterparts in other private sector companies. The picture is broadly the same for security workers.

To bring this together with the wage analysis, Table 3 estimates the impact of being an outsourced worker on a measure of total hourly compensation including both wages and pension contributions. As expected, the wage penalty estimated in the previous section significantly under-estimates the overall compensation penalty. The headline compensation penalty estimated using OLS is 9% for cleaners and 31% for security workers, compared to 6% and 24% for wages only. Turning to using my preferred fixed effects specification, the estimated compensation penalty is 4.2% for cleaners and 7.2% for security workers, compared to 2.8% and 5.6% for wages alone.

Table 1.3: Effect of Outsourcing on Log Hourly Total Compensation

	Pooled OLS		Within	
	(1)	(2)	(3)	(4)
Panel A: Cleaners				
Outsourced	-0.091***	-0.028***	-0.042***	-0.019**
	(0.004)	(0.004)	(0.005)	(0.006)
Public sector		0.168***		0.090***
		(0.005)		(0.008)
Observations	40,073	40,073	40,073	40,073
Panel B: Security Workers				
Outsourced	-0.310***	-0.216***	-0.072***	-0.068***
	(0.010)	(0.012)	(0.009)	(0.009)
Public sector		0.263***		0.053**
		(0.015)		(0.016)
Observations	12,054	12,054	12,054	12,054

Notes: Standard errors in parentheses, clustered at the individual level. * Significant at 5% level, ** at 1% level, *** at 0.1% level. In all models, the dependent variable is log hourly total compensation (excluding overtime) in 2016 prices. All specifications control for sex, age bands, region of work, year, temporary contract, and full-time status.

1.5 Outsourcing and firm wage premiums

The previous section established that outsourced cleaners and security guards earn significantly less than their in-house counterparts, even after controlling for individual characteristics. This implies that the wage penalty associated with outsourcing is not primarily driven by differences in worker quality or skill, but rather by differences in the wage setting practices of firms that employ outsourced workers. In this section, I examine this hypothesis directly by estimating firm-specific wage premiums using the widely used framework proposed by Abowd et al. (1999) (see Kline (2024) for an up-to-date survey of the relevant literature). If contractor firms pay systematically lower premiums than in-house employers, this would help explain the observed outsourcing wage penalty.

1.5.1 Estimating wage premiums: the AKM model

The AKM framework uses panel data on worker mobility to decompose log wages into components attributable to worker and firm characteristics. This allows one to estimate wage differentials across firms that cannot be explained by observable attributes or fixed worker effects. Specifically, I estimate the following regression, where $\ln(w_{it})$ denotes the log wage of individual i in period t, α_i is an individual fixed effect, and θ_t captures time effects:

$$\ln\left(w_{it}\right) = \alpha_i + x_{it}^{'}\beta + \theta_t + \psi_{J(i,t)} + \varepsilon_{it} \tag{1.2}$$

The term $\psi_{J(i,t)}$ represents a firm fixed effect: the average wage premium (or penalty) associated with firm j, relative to the average firm.[13] These premiums can reflect factors such as productivity, rent-sharing, unobserved amenities, or differences in wage-setting practices (Kline 2024).

The key identifying assumption for this framework is that worker mobility between firms is exogenous. That is, while high-ability workers may sort into high-wage firms—a phenomenon that is well-documented—the timing and direction of their moves should not be systematically related to unobserved shocks or trends affecting wages. This assumption allows for unbiased estimation of firm effects despite the existence of sorting, but it would be violated, for example, if workers tend to move in response to firm-specific, transitory wage shocks not captured by $\psi_{J(i,t)}$ (in other words, towards companies that are temporarily increasing their wages, or away from those that are reducing them); or if firms selectively offer higher wages to new hires. In addition to exogenous mobility, the AKM estimator assumes that worker and firm effects are additively separable—meaning that there are no match-specific effects. In other words, while some workers are more productive than others, and some firms pay higher wages than others, the model does not allow for certain workers to earn systematically more (or less) at particular firms due to a special synergy

[13] Firm fixed effects are normalized to have a weighted mean of zero, allowing each coefficient to be interpreted as a premium relative to the average firm.

or penalty in that pairing.

In practice, recent papers using large administrative datasets suggest that violations of these assumptions are typically small, and tend not to lead to substantial bias in estimated firm fixed effects (Card et al., 2013, 2016; Song et al., 2019). This finding is corroborated in a comprehensive review of the literature by Kline (2024), who also provides a detailed verification of the key AKM restrictions using a benchmark dataset, finding that the model captures 84-88% of the variation in observed wage changes for workers who switch firms, depending on the density of the mobility network.

Recent work has also shown that estimates of the *variance* of firm wage effects can be upward biased in sparsely connected mobility networks, a problem known as "limited mobility bias" (Kline, 2024). This arises because firm effects are estimated with error, and when worker flows between firms are limited, it becomes difficult to separate true wage differences from statistical noise—artificially inflating the observed spread of estimated effects. This bias is an important issue for studies aiming to quantify the contribution of firm effects to overall wage inequality using the popular variance decomposition method proposed by Abowd et al. (1999). But it is less of a concern here, because my analysis compares average firm wage premiums across groups (contractors vs in-house employers), and these group mean comparisons remain unbiased under the standard AKM assumptions.[14]

1.5.2 Results: wage premiums

I estimate equation 1.2 using the multi-dimensional fixed effects estimator developed by Correia (2016), and save the estimated firm wage premium for each firm in the largest connected set of firms linked by worker mobility. When the model is estimated using the entire ASHE sample, the largest connected set covers 38% of firms and about 80% of person-year observations (see Table A.2 in the appendix for details).[15]

[14] Limited mobility can still reduce the precision of estimated firm effects—especially in smaller datasets such as ASHE. Recent methods such as cross-fitting (Kline et al., 2020) and clustering approaches (Bonhomme et al., 2019) help mitigate this problem but typically require denser mobility networks than are available in the UK ASHE data.

[15] The proportion of observations included in this largest connected group is slightly lower than in other studies: Card et al. (2016) and Goldschmidt and Schmieder (2017) have around 90% of person-year observations. This likely reflects the relatively small sample size of the ASHE data.

To compare wage premiums between in-house and contractor companies, I estimate the following regression, using the estimated AKM wage premiums as the dependent variable:

$$\widehat{\psi_{i,J(i,t)}} = \alpha_i + x_{it}^{'}\beta + \theta_t + \gamma\,\text{Outsourced}_{it} + [\partial\,\text{Public}_{it}] + \varepsilon_{it} \tag{1.3}$$

The coefficient on the outsourcing dummy can be interpreted as the mean difference in estimated firm wage premiums for lead and contractor companies, with firms weighted according to their employment share. This two-step procedure—estimating firm fixed effects and then regressing them on firm covariates—is a widely used method in papers seeking to open the black box of AKM wage premiums; and under the standard AKM exogeneity assumptions provides unbiased estimates of the coefficients in the second-step regression (Kline, 2024).[16]

Column (1) of Table 1.4 shows that, when firm wage premiums are estimated using the full sample of workers, cleaners employed by contractor firms are associated with wage premiums that are on average 8.4 percentage points lower than those paid by firms employing cleaners in-house. The corresponding figure for security workers is even larger, at 15.2 percentage points. These results confirm that a significant share of the outsourcing wage penalty reflects differences in the wage-setting practices of contractor versus in-house employers.

Column (2) decomposes these differences into two components: a *public sector premium*, and a *contractor penalty* relative to other private sector in-house employers. For cleaners, the estimated firm wage premium at public sector employers is 12.8 percentage points higher than at private sector in-house employers. Contractor firms in the private sector, by contrast, are associated with

[16]As Kline (2024, 51) explains, the two-step approach involves computing firm fixed effects and then regressing them on firm covariates. Under strict exogeneity, the firm effects are unbiased. Since the projection coefficients in the second-step regression are just a linear combination of the estimated firm effects, they inherit this property and as a result "the two-step estimator provides robust estimates of the projection regardless of the dependence between worker and firm effects." However, as Kline et al. (2020) have highlighted, standard errors from two-step regressions may be understated because they ignore estimation error in firm fixed effects. They propose an alternative method for obtaining robust standard errors based on a cross-fitting procedure, in which firm effects are re-estimated while systematically leaving out worker-firm matches. This method is not feasible given the sparse nature of the UK ASHE data. In practice the differences between the naïve and cross-fit based standard errors in the Kline (2024) benchmark are relatively small, and most empirical papers continue to use unadjusted "naïve" standard errors.

firm wage premiums that are 3.4 percentage points lower. The same broad pattern holds for security workers: public sector firms offer significantly higher wage premiums, while private contractor firms pay markedly lower premiums than other private sector employers.

Table 1.4: Effect of Outsourcing on Estimated Firm Wage Premiums

	AKM (All Workers)		AKM (Cleaners / Security Only)	
	(1)	(2)	(3)	(4)
Panel A: Cleaners				
Outsourced	-0.084***	-0.034***	-0.045***	-0.018***
	(0.003)	(0.004)	(0.003)	(0.003)
Public sector		0.128***		0.068***
		(0.004)		(0.004)
Observations	50,515	50,515	46,146	46,146
Panel B: Security Workers				
Outsourced	-0.152***	-0.113***	-0.100***	-0.096***
	(0.006)	(0.007)	(0.006)	(0.007)
Public sector		0.112***		0.010
		(0.009)		(0.010)
Observations	15,333	15,333	14,180	14,180

Notes: Standard errors in parentheses, clustered at the individual level. * Significant at 5% level, ** at 1% level, *** at 0.1% level. The dependent variable is the estimated AKM firm fixed effect (wage premium) of the firm where each worker is employed. Columns (1)–(2) use AKM estimates based on all workers in the ASHE sample; Columns (3)–(4) use estimates based only on a combined sample of cleaners and security workers. The former assumes a common wage premium; the latter allows for variation by occupation. All specifications control for sex, age bands, region of work, year, temporary contract, and full-time status.

One limitation of the standard AKM model is that it assumes that firms pay the same average percentage premium to all their employees. In our context, this implies that in-house cleaners and security workers receive the same premium as other employees in the same company. If this is the case then, other things equal, we would expect the estimated wage penalty for outsourced workers in our fixed effects specification to roughly match the average differences in AKM wage premiums between in-house and contractor companies. However, the differences in wage premiums in column (1) are significantly larger than the wage penalties estimated in the previous section. One possible explanation is that, in practice, firms can pay different wage premiums to different groups of workers. For example, Card et al. (2016), using German data, find that women typically receive lower wage premiums than men; while Lee (2016), using ASHE data in the UK, finds a very weak correlation between firm wage premiums for high- and low-skilled workers within the same firms. If companies can pay in-house cleaners and security guards lower wage premiums than their other employees, then the firm wage premiums estimated using all workers will overstate the premiums received by in-house cleaners and security workers.

I test this hypothesis by estimating AKM wage premiums twice for the same firms: once using a sample comprised of cleaning and security workers only, and again using all other workers at the same firms. If firm wage premiums are the same across workers within the same firm, we would expect the correlation between these estimates of firm-specific wage premiums to be close to 1. Goldschmidt and Schmieder (2017) perform exactly this analysis using German wage data on food, cleaning, security and logistics (FCSL) workers. A regression of firm premiums estimated using FCSL workers on those estimated using all non-FCSL workers finds a coefficient of 0.68, rising to 0.77 when they correct for measurement error using a split-sample methodology.[17] In other words, their findings suggest that if an establishment pays around 1% higher wages to non-FCSL workers than other wages, it will pay about 0.8% higher wages to FCSL workers too.

I repeat this analysis, though the results should be treated with caution since the sample sizes for the UK are much smaller than those available for Germany. When the sample is restricted to cleaning and security workers the largest con-

[17] Goldschmidt and Schmieder (2017) propose a split-sample instrumental variables approach to correct for possible downward bias due to measurement error in the firm wage premiums estimated using the sample excluding FCSL workers. They randomly divide the sample of non-FCSL workers into two equally sized groups (the division is on the worker level, so all observations for the same individual are in one group), estimate firm fixed effects separately on each sub-sample, and use the resulting fixed effects to construct an instrument. They find that this increases the regression coefficient from 0.68 to 0.77.

nected set includes 40% of firms who employ at least one cleaning or security worker over the period, and 72% of person-year observations for cleaning and security workers (see Tables A4 and A5). A regression of the firm wage premium estimated using cleaning and security workers on the wage premium estimated using all other workers yields a coefficient of 0.14, which is statistically significant at the 0.1% level. Deploying the same split sample IV method used by Goldschmidt and Schmieder (2017) to correct for possible measurement error increases the estimated regression coefficient to 0.17.

In other words, small sample size notwithstanding, it seems that in our context in-house firms are able to pay cleaners and security workers substantially lower premiums than other employees. This helps explain why the difference in wage premiums using the standard AKM model is so much larger than the observed wage penalty for cleaners and security workers. It may also help explain why the estimated wage penalties in the UK are lower than estimates for the same occupations in the US (Dube and Kaplan, 2010) and Germany (Goldschmidt and Schmieder, 2017), since in-house employers appear to be able to exclude cleaners and security guards from wider rent-sharing wage policies even without outsourcing them, driving their wages towards the level paid to outsourced workers.

Columns (3) and (4) of Table 1.4 report differences in estimated firm wage premiums based on a subsample restricted to cleaners and security workers. For cleaners, the estimated difference in average firm wage premiums between in-house and outsourced employers falls to 4.5%, compared to 8.4% when premiums are estimated using the full sample. For security workers, the difference in estimated premiums declines from 15.2% to 10.0%. These estimates are notably closer to the outsourcing wage penalties obtained using the fixed effects specification in the previous section. This convergence is not coincidental: when the AKM model is estimated using only cleaners and security workers, it relies on the same kind of within-occupation mobility as the fixed effects approach, making the two methods conceptually very similar. That the estimated firm wage premiums using these occupational subsamples align more closely with the fixed effects results reinforces the interpretation that the observed outsourcing wage penalty reflects systematic differences in average wage-setting between contractor and in-house firms.

1.6 Outsourcing and firm rents

The evidence presented so far shows that outsourced cleaners and security workers are paid less than their in-house counterparts, and that this outsourcing wage penalty can in large part be attributed to the fact that contractor firms pay systematically lower wage premiums.

But why are outsourcing firms "low-wage" firms in this sense? In this section I use firm-level data to explore whether low wage premiums at contractor companies reflect low rent levels or low rent-sharing. Both explanations have some plausibility, and they are not mutually exclusive. On the one hand, since cleaning and security service firms offer rather generic services, they may operate in competitive product markets that leave little room for firms to capture rents. On the other hand, outsourcing firms may be less likely to share rents with their employees because, for example, they are less likely to be unionised and because, having a more homogenous workforce, they may not have the same pressure to compress wages between low and high skilled workers within a given firm.

1.6.1 Data matching

As I mentioned in Section 3, ASHE is a 1% sample the employed workforce. The ARDx is an annual census of firms with more than 250 workers, and a random sample of smaller firms. It is also restricted to the non-financial business sector, so workers employed in the public and financial sectors are automatically excluded. I find a firm-level match for 42% of cleaner observations and 60% of security worker observations. The lower matching rate for cleaners compared to security workers reflects the fact that cleaners are more likely to work in smaller companies, and to be employed in the public sector.

Before proceeding to the analysis of firm rents, it is important to consider the implications of matching worker-level records to firm-level data from the ARDx. This matching process disproportionately excludes workers in smaller and more fragmented companies, and may introduce selection effects. What matters most for the validity of my estimates, however, is not whether the matched sample is representative of the full population, but whether any selection bias operates *symmetrically* across in-house and outsourced workers. For example, even if the matched sample includes only relatively high-paying firms, this would not bias my estimates of the outsourcing wage penalty provided that in-house and outsourced workers are affected in the same way. In other words, I am interested in differences between in-house and outsourced workers, and as long as those differences are preserved in the matched sample, my estimates remain informative.

To assess whether this kind of symmetric selection holds in practice, Table 1.5 compares outsourcing wage penalties for private sector cleaners and security workers estimated in both the full and matched samples.[18] The OLS estimates are very similar across the two samples: for cleaners, the wage penalty falls slightly from 2.7% in the full sample to 1.6% in the matched sample, and for security guards the results are virtually identical at 18%. This suggests that the matched sample remains reasonably representative in terms of the overall distribution of contractor and in-house firms, at least as far as cross-sectional wage differences are concerned.

[18] Since the matched sample is restricted to private sector employees, the relevant comparison is between private sector workers in the full sample and private sector workers in the matched sample.

Table 1.5: Estimated Wage Penalty for Matched and Non-Matched Samples (Private Sector Only)

	OLS		Within	
	(1)	(2)	(3)	(4)
	Full	Matched	Full	Matched
Panel A: Cleaners				
Outsourced	-0.027***	-0.016**	-0.013*	0.019
	(0.004)	(0.005)	(0.007)	(0.012)
Observations	39,271	19,543	39,271	19,543
Panel B: Security Workers				
Outsourced	-0.179***	-0.180***	-0.042***	-0.030*
	(0.010)	(0.012)	(0.011)	(0.014)
Observations	13,153	8,099	13,153	8,099

Notes: Standard errors in parentheses, clustered at the individual level. * Significant at 5% level, ** at 1% level, *** at 0.1% level. In all models, the dependent variable is log hourly wages (excluding overtime) in 2016 prices. All specifications control for sex, age bands, region of work, year, temporary contract, and full-time status.

A more pronounced difference arises in the fixed effects specifications, especially for cleaners (unsurprisingly, given the lower matching rate). In the full sample, the fixed effects estimate suggests a small outsourcing wage penalty of –1.3%, while in the matched sample it flips to a wage premium of +1.9%, albeit not statistically significant. Since the fixed effects estimator relies exclusively on workers who transition between in-house and contractor firms, this suggests that the set of such transitions is altered by the matching process. In particular, the matched sample likely over-represents transitions to relatively higher-paying contractor firms—either because these firms are more likely to be large enough to appear in the ARDx, or because the movers captured in the matched sample are positively selected on unobserved traits.

While this asymmetry warrants caution in interpreting the estimates from the matched sample, it does not undermine the broader analysis that follows. If anything, it strengthens it. As I show in the next section, even this positively selected subset of contractor firms – larger and potentially more formalised –

have substantially lower rents per worker and, in the case of security services, are less likely to share those rents with employees. This suggests that weak rent-sharing is a structural feature of outsourcing firms, not one confined to a residual group of low-quality employers.

1.6.2 Rent levels

In this section I provide some descriptive evidence about differences in rent levels between contractor companies and companies that employ cleaners and security workers in-house. As is standard in the literature, I use gross value added (GVA) per worker as a proxy for firm rents. GVA is calculated as sales minus intermediate inputs, and reflects the total value created by the firm before payments to labour and capital.

Table 1.6 displays some key summary statistics about the distribution of firm GVA per worker for in-house and outsourced workers. It confirms that contractor firms have much lower levels of GVA per worker on average: around £16,000 in companies that employ outsourced cleaners and £25,000 in those that employ outsourced security workers, compared to roughly £42,000 for those that employ in-house cleaners and £99,000 for those employing security workers in-house.

Table 1.6: Firm GVA per Worker, Summary Statistics

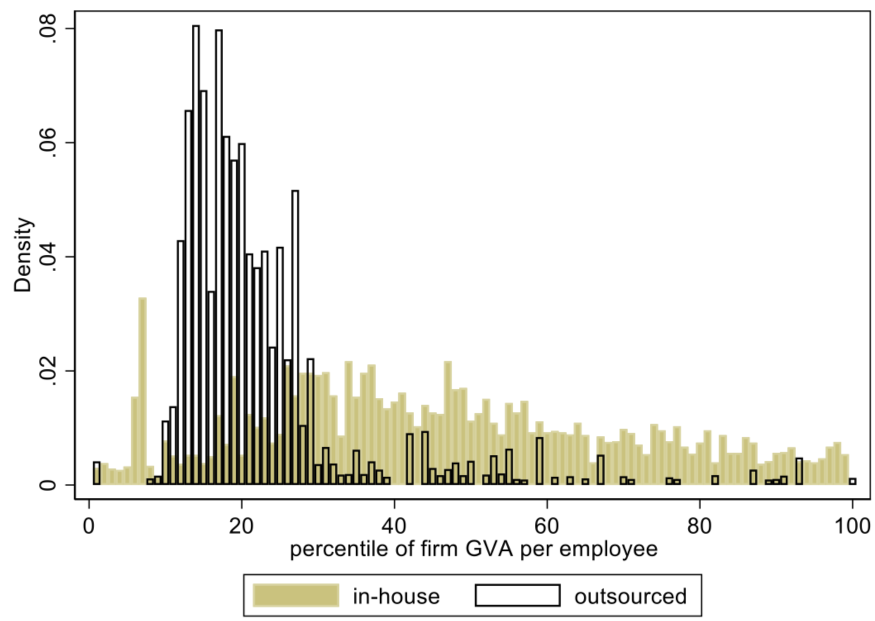
	Mean	Median	p25	p75	sd
Cleaners					
In-house	41.52	29.63	18.74	45.36	69.68
Outsourced	16.41	12.12	8.42	16.65	51.21
All	27.55	16.44	10.32	31.31	61.38
Security Workers					
In-house	98.73	39.91	22.10	132.01	162.22
Outsourced	25.30	24.44	20.87	28.15	10.47
All	52.71	25.36	21.12	36.97	105.60

Notes: Summary statistics for GVA per worker in the matched ASHE-ARDx sample. GVA measured in £'000s using 2016 prices.

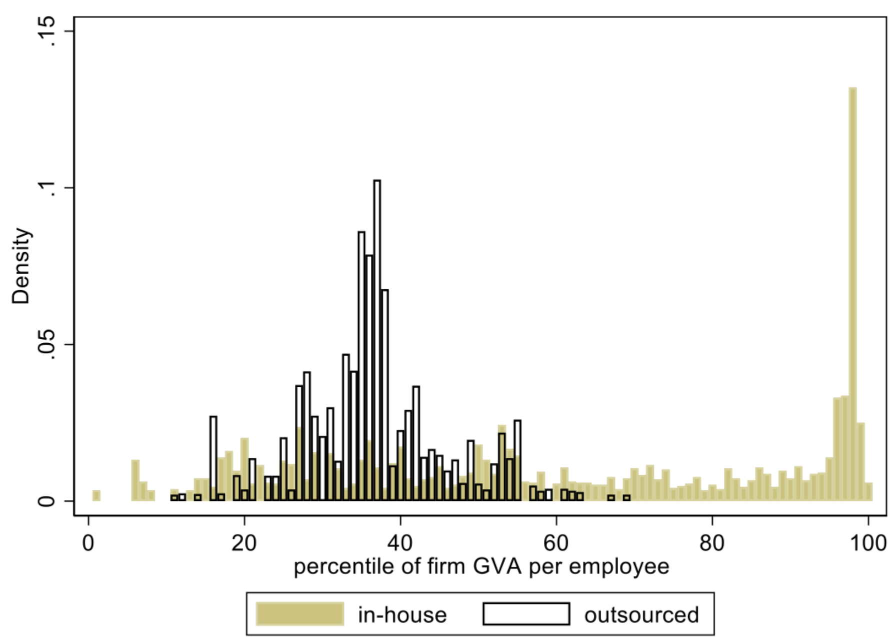
Figure 1.3 provides a more complete picture by plotting the distribution of outsourced and in-house cleaners and security workers by percentiles of firm GVA per employee. These show very clearly that outsourced workers are concentrated in low GVA firms, whereas in-house workers are employed across the full range of firms. This is especially pronounced for outsourced cleaners, nearly all of whom are employed in the bottom quartile of firms; while nearly all outsourced security workers work for firms in the bottom half of GVA per worker.

Figure 1.3: Distribution of Cleaners and Security Workers by percentile of firm GVA per employee

(a) Cleaners



(b) Security Workers



Notes: The x-axis is the percentile of firm GVA per employee in the year of observation. Note that quantiles with fewer than 10 observations have been excluded.

1.6.3 Rent sharing

It is clear that contractor firms typically have much lower rents per worker than companies that employ cleaners and security workers in-house. In this final section, we look at whether they are also less likely to share these rents with their employees.[19]

To do so, I estimate rent-sharing elasticities for both types of company using the following equation, which allows individual log wages to depend on firm-level rents and their interaction with outsourcing status:

$$\ln\left(w_{it}\right) = \alpha_i + x_{it}^{'}\beta + \theta_t + \lambda\pi_{J(i,t)} + \kappa\left(\pi_{J(i,t)} \cdot \text{Outsourced}_{it}\right) + \varepsilon_{it} \tag{1.4}$$

Here, $\pi_{J(i,t)}$ denotes quasi-rents per worker at firm $J(i,t)$, and λ captures the extent to which firms share those rents with workers.[20] The interaction term allows me to test whether this relationship differs between in-house and outsourced employment.

Conceptually, quasi-rents are defined as value added net of the competitive cost of labour and capital, i.e. the surplus available for distribution.[21] Since this is not directly observable, I use a three-year rolling average of GVA per worker as a proxy for underlying firm profitability, where the rolling average helps smooth out short-term fluctuations.[22] To address concerns about measurement error in GVA arising from noise in firm-level sales and input data, I also instrument GVA per worker using a rolling average of log turnover per employee. This widely used approach aims to reduce attenuation bias and recover a less noisy estimate of the relationship between firm rents and wages (Card et al., 2018; Van Reenen, 1996).

Estimating rent-sharing elasticities using cross-sectional data is subject to well-known biases. In particular, because higher-ability workers are more likely to match to higher-rent firms, and worker ability is often imperfectly

[19]See Figure A1 for a visual representation of the relationship between log wages and GVA per worker.

[20]This specification is closely related to the AKM framework introduced earlier in the paper, where the firm fixed effect can be interpreted as a linear function of firm quasi-rents i.e. $\psi_{J(i,t)} = \lambda\pi_{J(i,t)}$. See Card et al. (2018) for a similar formulation.

[21]See Card et al. (2018) for a derivation of the relationship between GVA and quasi-rents.

[22]My results are broadly robust to using a number of measures of GVA, including contemporaneous GVA, GVA smoothed over previous three years, and a rolling average over 3 and 5 years.

observed, cross-sectional estimates of rent sharing tend to be biased upwards (Card et al., 2018; Kline, 2024). To address this concern, I estimate rent sharing elasticities using worker fixed effects models, which control for time-invariant unobserved worker characteristics. However, while such models correct for sorting bias, they are more vulnerable to measurement error in the rent proxy, which tends to attenuate estimates of the rent-sharing elasticity toward zero (Kline et al., 2020); and they cannot fully eliminate potential bias arising from firm-specific shocks to rents that are correlated with worker mobility patterns, as would be the case if a sudden drop in profits encourages high-ability workers to move (Kline, 2024).

Recent work has moved towards using plausibly exogenous shocks to firm rents, such as the excess stock market returns associated with patent grants, in order to strengthen causal identification of rent-sharing elasticities (Kline et al., 2019; Bell et al., 2024). Since it is not possible to implement such an approach with my data, the absolute magnitude of the estimated rent-sharing elasticities below should be treated with caution. However, my focus is not on absolute magnitudes but on whether rent sharing elasticities differ across organisational forms – specifically, between contractor and in-house firms. Provided that the various biases discussed above operate similarly across the two groups – a plausible assumption especially for measurement concerns, since noise in the rent proxy is unlikely to be systematically related to organisational form – my approach to estimating differences in rent sharing between contractor and in-house firms should remain informative.

Table 1.7 displays the results of estimating equation 1.4 using both pooled OLS and the within estimator. In the OLS specification, I find an estimated elasticity of wages with respect to value added per worker of around 3.5% for cleaners and 14% for security workers, both of which increase slightly when GVA per worker is instrumented using figures for turnover. As expected, the estimated elasticities are significantly lower when using a fixed effects specification: around 1.5% for cleaners and 4% for security workers in my preferred IV specification. These estimates are low compared to other studies in the literature, which typically report elasticities in the range of 5 to 15% (Card et al., 2018). These low magnitudes likely reflect the fact that these elasticities are estimated specifically for cleaners and security workers, which as we saw in the previous section appear to receive lower wage premiums compared to other employees in the same companies, irrespective of whether they are in-house or outsourced. Indeed, as Table A7 shows, when I estimate rent-sharing elasticities for the entire matched ASHE-ARDx sample, the most plausible estimate (including worker fixed effects and using turnover as an instrument for GVA) is 7%, squarely within the standard range.

Table 1.7: Estimated Rent-Sharing Elasticities for In-House and Outsourced Workers

	Pooled OLS		Fixed Effects	
	(1)	(2)	(3)	(4)
Panel A: Cleaners				
Log GVA/emp	0.035***	0.039***	0.014***	0.014*
	(0.007)	(0.008)	(0.004)	(0.007)
Log GVA/emp × Out	-0.001	-0.002	0.006	0.005
	(0.005)	(0.005)	(0.003)	(0.003)
IV	No	Yes	No	Yes
Observations	14,545	14,545	14,545	14,545
Panel B: Security Workers				
Log GVA/emp	0.139***	0.171***	0.014	0.040**
	(0.016)	(0.014)	(0.008)	(0.013)
Log GVA/emp × Out	-0.030***	-0.022**	-0.013**	-0.015***
	(0.007)	(0.007)	(0.004)	(0.004)
IV	No	Yes	No	Yes
Observations	6,281	6,281	6,281	6,281

Notes: Standard errors in parentheses, clustered at the individual level. Dependent variable is log hourly earnings. Regressors include log GVA per employee (positive values only, smoothed over three years) and its interaction with outsourcing status. All models include controls for sex, age bands, region of work, year, temporary contract, and full-time status. Columns (2) and (4) use log turnover per employee as an instrument for GVA. * Significant at 5% level, ** at 1% level, *** at 0.1% level.

My interest here, however, is not so much the overall level of rent-sharing, but whether there is evidence of *differential* rent-sharing between contractors and companies that employ cleaners and security workers in-house, something captured by the interaction term in Table 1.7. For cleaners, I find no evidence of differential rent-sharing: the coefficient on the interaction of GVA per worker and the outsourcing dummy is very small and not significant. For security workers, however, this coefficient is negative and significant across all four specifications: in other words, contractor firms which employ security workers are less likely to share profits with their employees than firms that employ security workers in-house. The fact that outsourced security workers work at firms that have both lower levels of rents per worker *and* a lower tendency to share those rents, may help explain why the estimated penalty associated with being an outsourced security worker is so much larger than for outsourced cleaners.

1.7 Conclusion

The past fifteen years have seen a substantial increase in outsourcing among cleaners and security workers in the UK. I have shown that outsourced workers are paid less than their in-house counterparts and that these differences cannot be fully accounted for by differences in individual characteristics. Even once we control for individual fixed effects there remains an unexplained wage penalty of 2.8% for cleaners and 5.6% for security workers. This gap combines two conceptually distinct effects: first, since outsourced workers are exclusively private sector employees, they lose out on public sector wage premiums available to in-house workers in the public sector; second, outsourced workers suffer a further penalty associated with being employed by a contractor company rather than being an in-house employee in the private sector.

My analysis of estimated firm wage premiums using the AKM model confirms that the observed wage penalty associated with being outsourced reflect differences in firm- rather than worker-level characteristics. I find that estimated firm wage premiums are 4-9% lower among outsourced cleaners relative to their in-house counterparts, and 10-15% lower for outsourced security workers compared to those employed in-house, depending on the sample used to estimate those firm wage premiums. Finally, my analysis of matched worker-firm data suggests that contractor firms pay lower wages both because they have lower rents (as proxied by GVA per worker) and, for security workers, because they are less likely to share those rents with their employees.

While my findings are broadly consistent with earlier research on outsourcing-related wage penalties, this paper sheds new light on the role of firm rents in explaining those gaps. It also raises several important questions for future research. One is whether the effects of outsourcing differ across a broader set of occupations — particularly among higher-skilled workers, where the implications may be quite different due to greater bargaining power, task complexity, or access to alternative employment options. Another is to better understand why contractor firms exhibit lower profitability and (in some cases) weaker rent-sharing: is this primarily a reflection of competitive pressures, differences in productivity, or institutional factors such as union presence and bargaining coverage? It would also be valuable to examine the longer-term consequences of outsourcing for job quality and career progression.

Appendix

Table A1: Industry and Occupation Codes

Occupation Codes	**Industry Codes**
Cleaner	
SOC1990	*SIC 1992/2003*
956 Window cleaners	70.32 Management of real estate on a fee or contract basis
957 Road sweepers	74.70 Industrial cleaning
958 Cleaners, domestics	*SIC 2007*
SOC2000	81.10 Combined facilities support activities
9132 Industrial cleaning process occupations	81.21 General cleaning of buildings
9231 Window cleaners	81.22 Other building and industrial cleaning activities
9232 Road sweepers	81.29 Other cleaning activities
9233 Cleaners, domestics	
9235 Refuse and salvage occupations	
9239 Elementary cleaning occupations n.e.c.	
SOC2010	
9132 Industrial cleaning process occupations	
9231 Window cleaners	
9232 Street cleaners	
9233 Cleaners and domestics	
9235 Refuse and salvage occupations	
9236 Vehicle valeters and cleaners	
9239 Elementary cleaning occupations n.e.c.	
Security	
SOC1990	*SIC 1992 and 2003*
615 Security guards and related occupations	74.60 Investigation and security activities
619 Other security and protective service occupations n.e.c.	*SIC 2007*
SOC2000	80 Security and investigation activities
9241 Security guards and related occupations	80.1 Private security activities
9249 Elementary security occupations n.e.c.	80.2 Security systems service activities
SOC2010	80.3 Investigation activities
9241 Security guards and related occupations	
9249 Elementary security occupations n.e.c.	

Notes: SOC codes refer to occupational classifications used in the UK Standard Occupational Classification frameworks (1990, 2000, 2010). SIC codes refer to Standard Industrial Classification codes (1992, 2003, and 2007 versions). n.e.c. = not elsewhere classified.

Table A2: T-Tests of Differences in Group Means Between Private and Public Sector In-house Workers (Private – Public)

	Cleaners (1)	Security Workers (2)
Average hourly earnings (£2016)	-0.60***	-1.65***
Employer pension indicator	-0.28***	-0.37***
Weekly pension contributions (£2016)	-12.13***	-39.21***
Age at the survey reference date	-3.39***	-2.16***
Female indicator	-0.16***	-0.13***
Full-time indicator	0.17***	-0.05***
Same job for more than one year	-0.09***	-0.10***
Temporary employment indicator	-0.01***	0.02***
Employees in the enterprise (IDBR)	-4,998.09***	2,128.17
Public sector indicator	-1.00	-1.00
Covered by collective agreement	-0.67***	-0.47***
Observations	37,393	8,269

Notes: Table shows t-tests for differences in group means between in-house workers in the private versus public sector workers. Each entry shows the difference (private – public). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A3: Effect of Outsourcing on Employer Pension Contributions

	Pension Coverage (Odds Ratios)				Log Weekly Pension			
	Logit		Fixed Effects Logit		Pooled OLS		Within	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Cleaners								
Outsourced	0.110***	0.340***	0.048***	0.207***	-0.883***	-0.298**	0.068	-0.101
	(0.008)	(0.028)	(0.009)	(0.047)	(0.089)	(0.095)	(0.071)	(0.078)
Public sector		15.08***		40.18***		1.187***		-0.347***
		(1.104)		(10.127)		(0.057)		(0.065)
Observations	37,223	37,223	13,809	13,809	7,622	7,622	7,622	7,622
Panel B: Security Workers								
Outsourced	0.054***	0.113***	0.063***	0.084***	-1.609***	-1.357***	-1.075***	-1.075***
	(0.006)	(0.013)	(0.024)	(0.033)	(0.088)	(0.096)	(0.101)	(0.102)
Public sector		11.65***		18.59***		0.539***		0.0003
		(1.690)		(12.750)		(0.053)		(0.054)
Observations	11,152	11,152	6,657	6,657	3,835	3,835	3,835	3,835

Notes: Standard errors in parentheses, clustered at the individual level. * Significant at 5% level, ** at 1% level, *** at 0.1% level. Dependent variable in columns (1)–(4) is an indicator for any employer pension contribution. Dependent variable in columns (5)–(8) is log of weekly pension contribution among those who receive one. All models include controls for sex, age bands, region, year, temporary contract and full-time status.

Table A4: Samples for Estimating Firm Wage Premiums

	Person-years	Persons	Firms	Mean wage	SD wage
Whole ASHE sample	2,207,703	363,718	176,480	2.53	0.54
Largest connected set: whole sample	1,741,730	249,253	66,620	2.55	0.54
(% of whole sample)	(79%)	(69%)	(38%)		
Largest connected set: excl. cleaners & security	1,673,990	240,952	64,630	2.57	0.54
(% of whole sample)	(76%)	(66%)	(37%)		
Largest connected set: cleaners & security only	59,011	13,210	4,578	2.09	0.27
(% of whole sample)	(3%)	(4%)	(3%)		

Notes: "Largest connected set: whole sample" is the largest connected set identified when estimating the AKM model on the full ASHE sample. The next two rows show connected sets when excluding cleaners and security workers, or including only that group.

Table A5: Characteristics of Cleaners and Security Workers in Different Samples Used to Estimate AKM Model

	Cleaners			Security Workers		
	Whole sample (1)	Largest conn. set (2)	Conn. set (cleaners & security) (3)	Whole sample (4)	Largest conn. set (5)	Conn. set (cleaners & security) (6)
Compensation						
Mean hourly earnings (£2016)	7.81	7.98	7.96	10.00	10.14	9.91
Fraction with employer pension	0.19	0.22	0.23	0.31	0.32	0.30
Mean weekly employer pension (£2016)	7.70	8.97	9.35	21.12	21.23	19.21
Individual characteristics						
Mean age	44.44	44.63	45.45	43.00	43.08	43.43
Fraction female	0.68	0.66	0.68	0.15	0.15	0.13
Fraction full time	0.33	0.35	0.33	0.86	0.88	0.90
Fraction in job >1 year	0.77	0.79	0.82	0.80	0.81	0.84
Fraction temporary	0.06	0.05	0.04	0.04	0.03	0.02
Firm characteristics						
Mean firm size (employees)	8,610	10,409	13,166	15,953	17,963	21,003
Fraction public sector	0.31	0.35	0.39	0.19	0.21	0.18
Fraction covered by collective agreement	0.46	0.49	0.52	0.46	0.47	0.45
Person-years	65,411	50,427	35,481	16,488	14,173	10,566
Persons	21,016	14,378	7,898	5,068	3,969	2,303
Firms employing cleaners/security	3,052	1,506	680	924	442	210

Notes: See notes to Table A4 for description of different sub-samples. “Firms employing cleaners/security” refers to the number of unique firms who employed at least one cleaner or security worker during the sample period.

Table A6: T-Tests of Differences in Group Means Between Private Sector Workers in the Full and Matched Samples (Matched – Not Matched)

	Cleaners	Security Workers
	(1)	(2)
Average hourly earnings (£2016)	-0.07*	0.40***
Employer pension indicator	0.06***	0.16***
Weekly pension contributions (£2016)	2.73	8.41***
Age at the survey reference date	-0.59***	0.64**
Female indicator	-0.08***	0.03***
Full-time indicator	0.10***	0.06***
Same job for more than one year	0.01	0.05***
Temporary employment indicator	-0.01***	-0.02***
Employees in the enterprise (IDBR)	10,118.95***	10,462.58***
Public sector indicator	0.00	0.00
Covered by collective agreement	0.09***	0.12***
Observations	41,167	13,611

Notes: Table shows t-tests for differences in means between private sector workers in the full and matched samples. Each entry shows the difference (matched – not matched). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

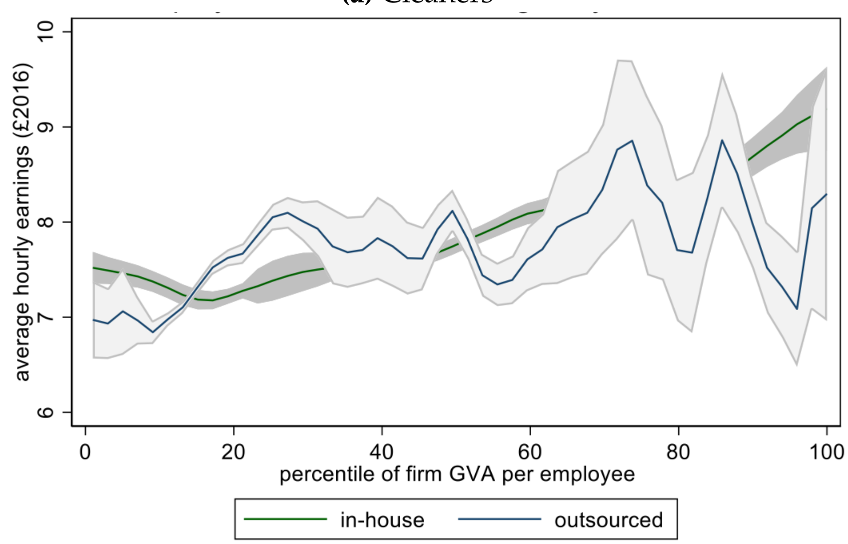
Table A7: Rent-Sharing Elasticities Estimated on the Entire Matched Sample

	OLS	IV	FE	IVFE
	(1)	(2)	(3)	(4)
Panel A: Cleaners				
Log GVA per employee	0.164***	0.238***	0.023***	0.071***
	(0.010)	(0.010)	(0.001)	(0.002)
Observations	514,434	514,434	514,434	514,434

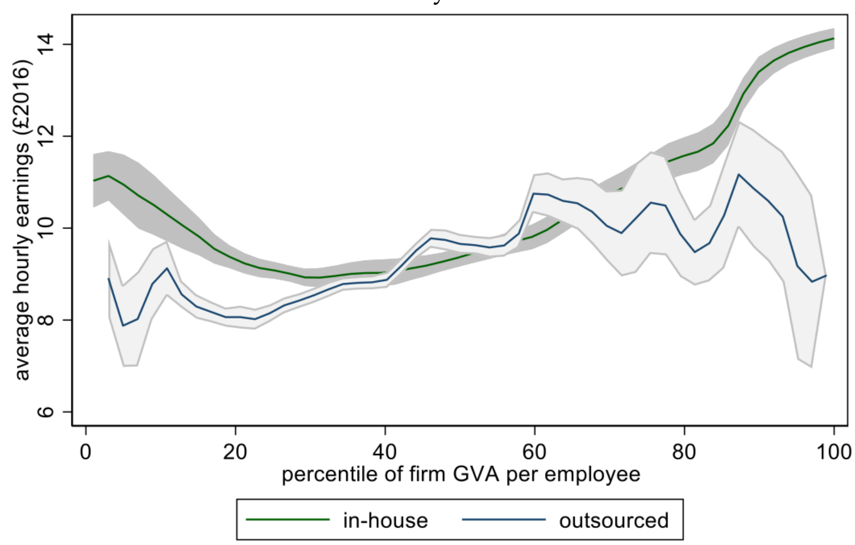
Notes: Dependent variable is log hourly wages. All specifications include controls for age bands, region, year, temporary contract, full-time status, and gender. Standard errors clustered at the individual level. * Significant at 5% level, ** at 1% level, *** at 0.1% level.

Figure A1: Fitted Polynomial of Wages and Firm GVA

(a) Cleaners



(b) Security Workers



Notes: This figure shows fitted polynomials relating log wages to firm-level gross value added (GVA) for cleaners (panel a) and security workers (panel b). These were created using the `lpoly` command, which performs a kernel-weighted local polynomial regression of `yvar` on `xvar` and displays a graph of the smoothed values with 95% confidence bands. The bandwidth for the x-value has been set to five. Estimates are based on matched private sector subsamples.

Bibliography

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Abraham, K. G. (1990). Restructuring the employment relationship: The growth of market-mediated work arrangements. In K. Abraham and R. B. McKersie (Eds.), *New Developments in the Labor Market: Toward a New Institutional Paradigm*, pp. 85–119. MIT Press.

Abraham, K. G. and S. K. Taylor (1996). Firms' use of outside contractors: Theory and evidence. *Journal of Labor Economics 14*(3), 394–424.

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

Akerlof, G. A. and J. L. Yellen (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics 105*(2), 255–283.

Barth, E., A. Bryson, J. C. Davis, and R. B. Freeman (2016). It's where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics 34*(S2), S67–S97.

Bell, B., P. Bukowski, and S. Machin (2024). The decline in rent sharing. *Journal of Labor Economics 42*(3), 683–716.

Bergeaud, A., C. Malgouyres, and P.-O. Pénin (2024). Technological change and domestic outsourcing. *Journal of Labor Economics*. Forthcoming.

Bilal, A. and H. Lhuillier (2021). Outsourcing, inequality and aggregate output. Working Paper w29348, National Bureau of Economic Research. Revised March 2025.

Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018, 1 May). The disappearing large-firm wage premium. *AEA Pap. Proc. 108*, 317–322.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework matched employer employee data". *Econometrica 87*, 699–739.

Booth, A. L. (1994). *The Economics of the Trade Union*. Cambridge University Press.

Bruns, B. (2019). Changes in workplace heterogeneity and how they widen the gender wage gap. *American Economic Journal: Applied Economics 11*(2), 74–113.

Card, D. (2011). Good jobs: The importance of who you work for. *Focus, Institute for Research on Poverty 30*(1), 23–29.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality. *Q. J. Econ. 128*(3), 967–1015.

Card, D., J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

Correia, S. (2016). A feasible estimator for linear models with multi-way fixed effects. Technical report. Working Paper.

Coudin, E., S. Maillard, and M. Tô (2018). Family, firms and the gender wage gap in france. IFS Working Paper W18/01, Institute for Fiscal Studies.

Daruich, D., M. Kuntze, P. Plotkin, and R. Saggio (2024). The consequences of domestic outsourcing on workers: New evidence from italian administrative data. Technical Report 84, Istituto Nazionale Previdenza Sociale (INPS). WorkINPS Paper No. 84.

Drenik, A., A. Camacho, and L. Fergusson (2023). Paying outsourced labor: Direct evidence from linked temp agency-worker-client data. *The Review of Economics and Statistics 105*(1), 206–216.

Dube, A. and E. Kaplan (2010). Does outsourcing reduce wages in the low-wage service occupations? evidence from janitors and guards. *Industrial & Labor Relations Review 63*(2), 287–306.

Goldschmidt, D. and J. Schmieder (2017). The rise of domestic outsourcing and the evolution of the german wage structure. *The Quarterly Journal of Economics 132*(3), 1165–1217.

Handwerker, E. W. (2023). Outsourcing, occupationally homogeneous employers, and wage inequality in the united states. *Journal of Labor Economics 41*(S1), S173–S203.

Jakubson, G. (1991). Estimation and testing of the union wage effect using panel data. *The Review of Economic Studies 58*(5), 971–991.

Katz, L. F. and A. B. Krueger (2016). The rise and nature of alternative work arrangements in the united states, 1995-2015. NBER Working Paper 22667, National Bureau of Economic Research.

Kline, P. (2024). Firm wage effects. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 5, pp. 115–181. Elsevier.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019). Who profits from patents? rent-sharing at innovative firms. *Q. J. Econ. 134*(3), 1343–1404.

Kline, P., R. Saggio, and M. Sølvsten (2020). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

Lee, I. (2016). Unobservable worker and firm effects: Evidence from the u.k. Unpublished MPhil paper, Oxford University.

Li, J., B. Dostie, and G. Simard-Duplain (2023). Firm pay policies and the gender earnings gap: The mediating role of marital and family status. *Ind. Labor Relat. Rev. 76*(1), 160–188.

Loecker, J. D., J. Eeckhout, and G. Unger (2024). Firms and inequality. *Oxford Open Economics 3*(Supplement 1), i962–i982.

Manning, A. (2003). The real thin theory: Monopsony in modern labour markets. *Labour Economics 10*(2), 105–131.

Manning, A. (2011). Imperfect competition in the labor market. In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics*, Volume 4B, pp. 973–1041. Elsevier.

Mas, A. and A. Pallais (2016). Valuing alternative work arrangements. NBER Working Paper 22708, National Bureau of Economic Research.

Polivka, A. E. (1996). Contingent and alternative work arrangements, defined. *Monthly Labor Review 119*.

Segal, L. M. and D. G. Sullivan (1997). The growth of temporary services work. *Journal of Economic Perspectives 11*(2), 117–136.

Shapiro, C. and J. Stiglitz (1984). Equilibrium unemployment as a worker discipline device. *The American Economic Review 74*(3), 433–444.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

Sorkin, I. (2018, 1 August). Ranking firms using revealed preference. *Q. J. Econ. 133*(3), 1331–1393.

Van Reenen, J. (1996). The creation and capture of rents: wages and innovation in a panel of UK companies. *The Quarterly Journal of Economics 111*(1), 195–226.

Weil, D. (2014). *The Fissured Workplace*. Harvard University Press.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press.

Zwysen, W. (2024). Working apart: Domestic outsourcing in europe. *European Journal of Industrial Relations 30*(2), 221–241.

Chapter 2

Anatomy of Automation: CNC Machines and Industrial Robots in UK Manufacturing, 2005–2023

ABSTRACT

Using a novel proprietary survey of UK manufacturing sites, we study the impact on employment of arguably the two most important industrial automation technologies of the past fifty years: computer numerical control (CNC) machine tools and industrial robots. First, we document the growing prevalence of both technologies across a wide range of industries between 2005 and 2023. Second, we use a local-projection difference-in-difference design to show that plants that adopt these technologies for the first time increase their employment by 6% to 8% after four years, compared to non-adopting plants in the same industry. Third, we find that for both technologies, automation is associated with an increase in employment among industry-competitor sites, and a positive overall impact on industry-level employment.

Keywords: automation; manufacturing; CNC machines; robots; difference-in-differences.

2.1 Introduction

How does automation affect employment? Recent advances in artificial intelligence have heightened both hopes and fears about the consequences of technological progress for workers and society more broadly. There is widespread concern that new technologies will displace workers, leading not just to temporary disruption, but to a secular decline in labour demand, putting downward pressure on wages and employment. And yet, historically, such displacement effects have typically been accompanied by improvements in productivity and the creation of new products and labour-intensive tasks, driving up the demand for labour (Autor, 2015). Which of these channels dominates is fundamentally an empirical question.

In this paper, we examine how the widespread adoption of computer numerical control (CNC) machine tools and industrial robots affected employment in the UK economy since 2005. CNC machines are programmable tooling devices that cut, drill, or shape materials (typically metals) into precisely defined components. They have been widely available since the 1980s and play a foundational role in modern manufacturing. Industrial robots, by contrast, are programmable machines designed to manipulate objects with considerable spatial flexibility. They typically operate across multiple axes using motors and sensors to guide their movements, and are commonly used for tasks such as assembly, welding, packaging, and material handling. While CNC machines are generally used to transform raw materials into components through precision machining, industrial robots are more often deployed in later stages of production to assemble or move those components. Together, these technologies represent two of the most important and widely used forms of industrial automation in recent decades.

This paper makes three main contributions. First, we introduce novel evidence on the use of these manufacturing technologies from a proprietary survey of manufacturing establishments in the United Kingdom between 2005 and 2023. This dataset, produced by the Mark Allen Group, contains information on plant-level stocks of CNC machine tools and (from 2014) indicators for the use of industrial robots on site. Crucially for our purposes, it also includes site-level information about total employment and, separately, the number of workers by broad occupational groups, including manufacturing workers.

We establish that our dataset is broadly representative of the UK manufacturing sector, closely tracking statistics obtained from the UK's statistical agency, the Office for National Statistics (ONS). We also document an increase in the use of both technologies over the period we study: the share of manufacturing production plants using CNC machines increased from about 46% in 2004 to nearly 56% in 2023, while the share of employees who work in such establishments increased from about 52% to 62%. We find that the share of plants using industrial robots has increased from about 4% in 2014 to 6% in 2023, while the share of employees working in such plants increased from around 15% to more than 25%, indicating that robot adoption has been skewed towards large plants.

Second, we use modern differences-in-differences event study methods to study the impact of CNC machine and industrial robot adoption at the extensive margin on a firm's workforce. In our preferred specification, we find that plants that adopt CNC machines for the first time increase their total employment by about 6% over the subsequent four years, compared to non-adopting plants in the same industry. Plants that adopt industrial robots for the first time also see employment increase by around 8% in the years following adoption. We find that treated plants which experience an "adoption event" exhibit broadly similar pre-treatment trends to those that do not; while the substantial cost and planning involved in adopting such technologies makes it unlikely that our results are being driven by contemporaneous shocks affecting both technology adoption and employment. Unlike much of the literature, we do not find any significant impact on the share of manufacturing workers after adoption events.

For CNC machines, we can also look at the impact of "expansion events", which occur when a plant increases the total number of CNC machines from an already positive number. We find that these events are associated with a larger increase in total plant employment than initial adoption, and a notable reduction in the share of manufacturing workers. This pattern is consistent with a learning-based view of technology adoption, in which experienced firms are better able to take advantage of new technologies, leading both to larger gains in productivity and employment, and a deeper reorganisation of their production processes (e.g. Atkeson and Kehoe, 2007).

Third, we explore the dynamics of worker reallocation across companies and industries. We first look at the impact of firm-level automation events on employment among competitors firms within the same industry. We find evidence of positive spillovers, in the sense that adoption in one plant is associated with an expansion of employment among peer firms. In order to capture the overall impact of automation on industry level employment, we adapt our main event-study specification to study industry-level automation events, defined as a relatively large annual change in the share of employees working in companies that use a given technology, compared to all year-on-year changes. Although our results vary depending on the threshold used to define an industry level event, we find positive industry level employment effects in all specifications, with varying degrees of statistical significance. These results suggest that, at least in the UK manufacturing context, the broader effects of automation on employment may be more benign than is often feared.

Related Literature. Our paper contributes to a small but growing empirical literature that uses firm-level data to look at the labour market effects of various modern technologies. The papers in this literature span a range of countries and recent time-periods, and study a variety of technologies, with a particular focus on industrial robots (for a recent overview see Aghion et al. (2023a) and Restrepo (2024)). Most use event-study designs similar to our own, though a handful combine this with a more explicitly causal approach.[1]

As in our study, this literature consistently finds that investment in new technologies is associated with an increase in total employment at focal (adopting) companies (Aghion et al., 2023a).[2] The literature is more divided when it comes to the impact on specific occupational groups. Most studies, including all those looking at industrial robots, find that investment in new technologies is associated with a reduction in employment (and in some cases wages) for "exposed" occupations, typically lower-skilled and/or manufacturing workers (Bessen, 2019; Humlum, 2021; Acemoglu and Restrepo, 2020); but other recent studies find no evidence of differential effects on different skill groups (Curtis et al., 2021; Hirvonen et al., 2023; Aghion et al., 2023c). Our results – that ex-

[1] Aghion et al. (2023c) use a shift-share IV design leveraging pre-determined supply linkages and productivity shocks. Hirvonen et al. (2023) use a quasi-experimental design comparing companies that secured access to a government subsidy programme with those that narrowly lost out.

[2] Studies which also have access to information about firm-level sales and productivity find that this increase in employment is associated with rising sales and productivity. Some of these studies also look at the impact on wages, and typically find no significant effect.

pansion but not adoption events are associated with a decline in the share of manufacturing workers – suggest that the impact of automation technologies on exposed groups of workers may increase as companies accumulate experience with a given technology.

While the studies discussed above examine the effects of technology adoption within firms, a largely separate literature focuses on the impact of automation on employment at the industry or labour market level. This literature has produced mixed results. For example, Acemoglu and Restrepo (2020) find that the addition of one industrial robot per 1,000 workers in U.S. commuting zones reduces the employment-to-population ratio by 0.4 percentage points; whereas Dauth et al. (2021), using a similar approach in Germany, find no adverse effect on regional employment. At the cross-country level, Graetz and Michaels (2018) find no aggregate employment effect of robot adoption across 17 developed economies, while Klenert et al. (2023), using similar methods for 14 European countries, find positive correlations between robot adoption and employment. Other studies, such as Webb (2020); Mann and Püttmann (2023); Kogan et al. (2023), use patent-based measures of exposure to assess occupational risk, but also reach inconsistent conclusions about aggregate employment effects. The only previous study to focus specifically on CNC adoption at the industry level is Boustan et al. (2022), who find that industries more exposed to CNC technologies experienced increases in investment, productivity, and employment, with gains for college-educated workers offsetting losses among less-educated ones.

We see our contributions to this literature as threefold.

First, we are able to exploit direct plant-level measures of two critical automation technologies: CNC machines and industrial robots. Most recent papers with access to firm-level data rely on composite measures of technology such as the total value of manufacturing capital (Aghion et al., 2023b,c) or investment in third party automation services (Bessen et al., 2023). As Aghion et al. (2023c) argue, these broad measures can help give us a sense of the impact of typical investments in manufacturing capital. But the impact of new technologies depends critically on their particular characteristics, and the degree to which they are used to displace workers rather than, say, create new products. A better understanding of the effects of modern technologies on labour markets must study the characteristics and abilities of specific technologies. Moreover, while there is now a fairly large literature looking at the impact of industrial robots, studies of CNC technology are surprisingly rare, given their huge importance for modern manufacturing. To the best of our knowledge, ours is the first paper to use firm-level data on CNC use across a wide range of industries.[3]

A further advantage of our data is that we can isolate adoption events, when a plant starts using a given technology for the first time, and in the case of CNC machines distinguish them from "expansion" events, when plants increase their use of CNC machines from an already positive baseline. Other papers, which typically use price-based measures of technology investment, cannot distinguish investment at the extensive margin – which involves the deployment of a new machine in tasks previously performed by humans, and is the canonical definition of automation, at least within the task framework – from investment at the intensive margin, which seeks to increase the productivity of capital in existing tasks. Our focus on adoption events, then, provides a cleaner empirical analogue to automation. At the same time, being able to compare the effects of adoption and expansion for CNC machines allows us to test the predictions of learning-based models of technological change (Atkeson and Kehoe, 2007).[4]

Our second key contribution is to provide a unified analysis of the implications of technology adoption at the firm and industry levels. This is clearly not possible for papers using aggregate-level data, while many papers with firm-level data have focused exclusively on firm-level outcomes (Bessen et al., 2023; Dixon et al., 2021; Koch et al., 2021). A number of papers with access to firm-level data have studied the impact of automation at one firm on "competitors" in the same industry and, like us, find negative spillover effects (Acemoglu et al., 2020; Koch et al., 2021; Aghion et al., 2023c). But few have looked at the overall impact on industry or area-level employment. An important excep-

[3] Of the two recent papers focused on CNC machines, Boustan et al. (2022) only have access to industry level data, while Bartel et al. (2007) have access to plant-level data for the US valve manufacturing sector only. Bartel et al. (2007) find that plants adopting CNC tools improve the efficiency of all stages of the production process by reducing setup times, run times, and inspection times; see increases in the skill requirements of machine operators and the adoption of new human resource practices; and shift their business strategies towards more customized products.

[4] Another advantage of measuring technology use directly is that changes in price-based measures may simply reflect shifting prices rather than changes in technology use in a given company. A downside is that we are unable to account for changes in the quality of machines across firms or over time.

tion is Aghion et al. (2023b,c), who use French firm-level data to look at the impact of investments in manufacturing capital on firm, industry and labour market outcomes, finding positive employment effects at all levels. Addressing these different levels of analysis within a single empirical setting is critical for understanding the relationships between firm-level effects, and wider general equilibrium implications.

Our third contribution is simply to provide the first UK-based study of automation in the manufacturing sector using firm-level data. Studying different countries is valuable because, as we discuss in more detail below, the impact of automation depends not simply on the technology in question, but on the wider labour market context. In other words, we cannot simply assume that the broad effects of robot adoption on aggregate employment in (say) the US or Germany will carry over to the UK. As we gather more data points from different countries and labour markets, we can better understand the contexts in which new technologies are likely to lead to positive or adverse effects for different groups.

This paper is structured as follows. Section 2.2 provides background on CNC machines and industrial robots. Section 2.3 outlines the task-based conceptual framework that guides our analysis. Section 2.4 describes our data and benchmarks it against UK administrative data. Section 2.5 outlines our empirical approach. Section 2.6 outlines our firm-level results, while Section 2.7 looks at the impact of automation on industry-level employment. Section 2.8 concludes.

2.2 Background: CNC machines and industrial robots

In this section, we provide background on the history of CNC machines and industrial robots, and their distinctive role in modern manufacturing.

At the most fundamental level, manufacturing is a process by which parts or pieces of material are cut, drilled, bent and shaped into desired shapes to produce components. These components are then assembled to produce manufactured goods. Machine tools are a broad class of machines which fulfil the first step of this process, typically through a subtractive process which removes material from the workpiece until the desired shape is achieved. Their importance for manufacturing is difficult to overstate: they are sometimes described as the "mother" machine, since, as Holland (1989b, 2) memorably put it, "every manufactured product is made by a machine tool or by a machine that was made by a machine tool".

Prior to the development of machine tools, manufacturing was the domain of skilled artisans who performed the entire range of tasks associated with the production of a final good from raw material. The late 18th and 19th century saw a paradigm change, as manufacturing embraced standardisation and specialisation, and production processes were redesigned to focus on producing large numbers of interchangeable components, which were then assembled into final products (National Research Council, 1995). As a result, manufacturing jobs were increasingly devoted to repeatedly performing the same task. Critical to the adoption of this system was the availability of special purpose machine tools built specifically for each task (Jaikumar, 2005). [5] These tools were, however, manual in their operation, controlled fully by their operators.

This started to change with the invention of numerical control (NC) machines in the 1950s and 1960s, and their widespread adoption throughout the 1970s[6]. Rapid developments in computing, including the creation of computer-aided design technologies, led to the birth of the first true computer numerical control (CNC) machines, displacing the punched cards used by NC machinery.[7] The first CNC tools designed for wide commercial application were developed in Japan in the late 1960s and their worldwide diffusion accelerated in the 1980s and 1990s[8]. Rapid improvements in both microprocessor technology and Computer-Aided Design (CAD) software in the 1980s transformed the

[5] Mechanisation yielded massive improvements in productivity, particularly when combined with the managerial and organisational changes that were taking place alongside it. Bright (1958) shows, for example, that lamps produced per operator per day rose from 160 in the 1910s and 20s, to 800 in the 1920s and 30s, and to 2,700 by the 1950s.

[6] In 1966, a report by the US National Commission on Technology, Automation and Economic Progress heralded the invention of NC machine tools as "probably the most significant development in manufacturing since the introduction of the moving assembly line" (Lynn et al., 1966)

[7] The first CNC machine was created at MIT in 1952, as part of a contract with the US Air Force to create high-precision helicopter parts. Ross (1978) describes the development of the Automatically Programmed Tool programming language, which was incorporated into a numerical control system in collaboration with the US Air Force by 1957 and demonstrated publicly in 1959.

[8] As Boustan et al. (2022) show, CNC machinery was adopted more quickly in some industries than others, reflecting the differential pace at which the underlying technology was developed for different tool types and actions, like lathing, drilling and so on.

capabilities of these machines and radically simplified the design process, leading to their widespread diffusion. CNC machine tools offer increased precision and repeatability, and significantly reduced the set-up times needed to adjust and prepare a machine between different tasks.

Industrial robots can perform many of the same tasks as CNC machine tools, such as drilling, cutting or bending. But their distinctive feature is their spatial flexibility, and their ability to move and manipulate objects in a variety of ways. Whereas CNC machine tools are mainly used in the production of parts, industrial robots are typically used for welding, sorting, painting and a variety of other tasks needed to assemble parts into final products.[9] The first industrial robots were created in the 1950s, and their use expanded during the 1960s and 1970s, especially in the automotive sector, but it is only since the 1990s that they have started to be used across a much wider range of applications including electronics, food processing, logistics, and precision engineering. As with CNC machines, industrial robots have benefited from advances in computing, as well as the development of sensors and machine vision, and recent developments in artificial intelligence continue to transform their capabilities, allowing increasing precision and flexibility.

What can we take from this brief history for thinking about the impact of the growing use of CNC machines and industrial robots in modern manufacturing? First, it's clear that both CNC machine tools and industrial robots are "automation" technologies, in the sense that they are designed to replace humans in performing specific tasks. Second, these technologies typically apply to different parts of the production process, and hence may have different implications for employment: as Boustan et al. (2022) argues, CNC machines primarily automate the work of skilled machinists with advanced motor skills, as well as lower skilled machine setters and set-up operators; while industrial robots tend to automate lower skilled jobs requiring gross motor skills connected to assembly, welding, packaging processes. Third, CNC machines are a more mature technology than industrial robots and, as we shall see in our data, are much more widely diffused across the manufacturing sector. This is relevant for thinking about learning effects, since there is likely to be a much

[9]The International Federation of Robotics defines industrial robots as "automatically controlled, reprogrammable multipurpose manipulator, programmable in three or more axes, which can be either fixed in place or fixed to a mobile platform for use in automation applications in an industrial environment".

greater stock of accumulated experience with CNC machines than for robots.

2.3 Conceptual framework

For many years, the dominant approach in the theoretical literature on automation was the canonical *factor-augmenting* framework. In this view, technological progress raises the productivity of a given factor—capital, skilled labour, or unskilled labour—uniformly across all tasks.[10] Within this framework, which dates back to Tinbergen (1974), and was developed in detail by Katz and Murphy (1992) and Goldin and Katz (2008), new technologies are generally viewed as benign in their effects on labour markets: they raise productivity, boost labour demand, and leave the labour share of income broadly unchanged, at least under realistic values for key parameters.

But as Acemoglu and Restrepo (2022a) argue, this model lacks descriptive realism. Most technologies do not improve the productivity of a factor across the board. Instead, they tend to enhance performance in specific tasks while leaving others unchanged—or, in the case of automation, replacing workers altogether in some tasks. The factor-augmenting model cannot easily capture this task-specific nature of technological change, nor the displacement of workers that is central to public and scholarly concerns about automation. Nor does it align with a growing body of empirical evidence showing that many new technologies have coincided with declining real wages and employment for exposed workers, as well as a falling labour share overall.

These limitations have led to the development of the *task-based* framework, which forms the theoretical foundation of this study. In this framework, production is understood as a set of tasks that can be performed by either human labour or capital, depending on their relative productivity and cost. Automation refers to the reallocation of specific tasks from human labour to machines. Crucially, this framework allows us to distinguish automation from two other forms of technological progress: the creation of new tasks (which can boost demand for labour) and improvements in the productivity of capital at tasks it already performs (so-called *intensive margin* progress, which does not directly displace workers). While all these forms of technological change can raise productivity, only automation—as defined here—entails a direct displacement of

[10] The following discussion draws on Acemoglu and Restrepo (2022a) and Restrepo (2024).

labour.

Automation technologies trigger two competing forces. The first is a *displacement effect*, which reduces demand for workers who previously performed the automated tasks. The second is a *productivity effect*, which makes production more efficient by lowering costs and reallocating tasks, potentially enabling firms to expand output. Whether employment rises or falls depends on the balance of these forces. Importantly, this is not something that theory alone can determine—it is an empirical question.

The task-based framework also makes clear that the impact of automation can differ significantly depending on the level of analysis. At the *firm level*, automation may reduce the number of workers needed for particular tasks but increase overall productivity. If demand for the firm's output is elastic, these productivity gains can lead to firm expansion and net employment growth. While the impact of automation on total company employment is ambiguous, according to the task-based framework it should lead to a decline in the employment share of "exposed" occupations whose tasks are directly replaced by machines.

At the *industry level*, automation can trigger a range of spillover effects that shape employment dynamics beyond the adopting firm. On the one hand, automation may lead to a reallocation of output toward more efficient producers, leading to job losses in non-adopting firms—what the literature refers to as "business stealing"—especially if industry-level demand is inelastic. If these negative spillovers are strong enough, this could lead to a decline in industry-level employment, even if automation leads to an increase in employment among focal plants. On the other hand, automation may also generate positive spillovers: by lowering prices and expanding overall demand, raising demand for complementary goods and services, or inducing learning and productivity gains through demonstration effects. At the *regional or labour market level*, aggregate employment outcomes reflect not only these firm and industry dynamics but also a wider set of general equilibrium effects including, among other things, how easily displaced workers can transition to new jobs.

These insights help clarify the motivation for this study. First, the task-based framework explains why the impact of automation on employment is theoretically ambiguous and must be resolved through empirical analysis. Second, it highlights that the effects of automation can diverge across firms, industries, and regions, making it essential to track these dynamics at multiple levels of aggregation. And third, it underlines the importance of focusing on specific technologies and distinguishing automation from other forms of technological change. A key advantage of our study is that we are able to track the adoption of two of the most important automation technologies of recent decades. Moreover, we directly observe "adoption" events—the first time a firm introduces a given technology—which provides a clean empirical analogue to the concept of automation at the extensive margin, as opposed to other forms of technological investment.

While the task-based framework offers a powerful conceptual tool for thinking about how automation reshapes the allocation of tasks between labour and capital, it typically abstracts from the dynamic processes through which firms adjust to new technologies.[11] In particular, it overlooks the fact that adopting and integrating automation may itself be a gradual process, involving costly experimentation, reorganization, and learning. Learning-based models, such as those developed by Atkeson and Kehoe (2007), emphasize that initial adoption is only the first step in a lengthy "learning-by-doing" process, during which firms accumulate the complementary human and organizational capital needed to fully exploit a new technology. As a result, the benefits of automation may unfold gradually, with later investments yielding larger effects on productivity and employment than initial adoption.[12] As we will argue in Section 2.6, this provides a compelling account of why "expansion" events – when companies increase their use of CNC machines from an already positive baseline – may yield stronger effects than first-time adoption.

2.4 Data: The MAG manufacturing survey

A central challenge in automation research has been gaining access to firm-level data on technology adoption. In this paper, we have access to such data

[11] While the learning dynamics discussed below are not modelled explicitly in the canonical task-based framework, they are not incompatible with it.

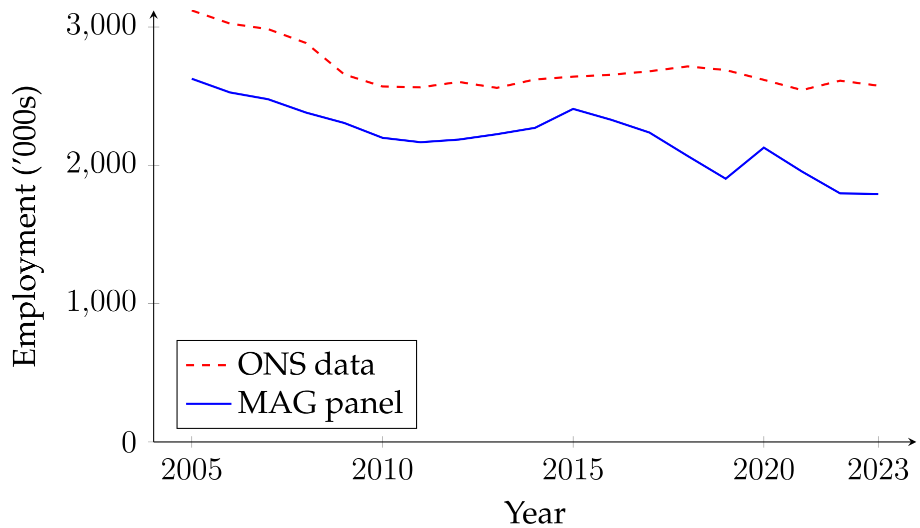
[12] As Atkeson and Kehoe (2007) show, learning-based models can help explain key facts about previous periods of rapid technological change, including the long lag between technological innovations and improvements in productivity, the slow diffusion of new technologies, and continued investment even in mature technologies, such as CNC machines. In their calibrated models internal innovations among existing users can exceed those at the technology frontier. For a discussion of similar ideas in relation to artificial intelligence see Brynjolfsson et al. (2019).

through the Mark Allen Group (MAG) Manufacturing Survey, a private census of UK manufacturing plants that has been updated continuously for over 30 years, and which we have access to from 2003 to 2023.[13] MAG is a media and information company that, among other things, publishes more than one hundred industry-focused publications across a wide range of sectors. It sells access to its manufacturing survey to clients who use it to target advertising and for sales prospecting by post, phone, and email. This exerts a strong market discipline on data quality, since clients rely on the accuracy of the information for commercial targeting, and errors would quickly be discovered, undermining MAG's reputation.

The MAG survey is conducted through a series of telephone interviews, and seeks to provide a comprehensive picture of manufacturing sites in the UK. As we can see in Figure 2.1, the MAG data largely mirrors the aggregate decline in total manufacturing employment that is observed in official national statistics over the past twenty years. Table B1 shows that the MAG survey captures around 80% of total annual manufacturing employment on average from 2005 to 2023, with a sharp decline in coverage during the Covid-19 pandemic; while Table B2 shows that most of the gap in coverage comes from missing plants with fewer than 10 employees. Although the raw MAG data is not perfectly representative of the national manufacturing sector, this is not a serious issue since our analysis is focused on within-plant changes in technology and employment.

[13]The survey was created and run by Findlay Media Limited until 2014 when it was acquired by Mark Allen Data Services (MADS), a subsidiary of the Mark Allen Group. For further information about MAG, see https://www.markallengroup.com/

Figure 2.1: Employment trends in MAG vs ONS



Notes: this table compares total employment (in '000s) in the MAG survey with official data from the quarterly Labour Force Survey, produced by the Office for National Statistics. As per Table B1, the MAG survey captures 80% of total ONS manufacturing employment on average.

For each surveyed site, the data includes the company name (note that companies can include multiple establishments), whether the company is part of a group, the postcode within which the site is located, detailed industry codes, the total number of employees at a given site, and the number of employees in four key sub-groups: manufacturing production, factory services, engineering design and electronic design. The survey also includes the self-described "primary product" manufactured at the site, as well as information about key supplier relationships, including whether a plant makes its own products and/or performs subcontract work for others, and whether they supply the aerospace, automotive or defence sectors.

The unique feature of this dataset is that it contains rich plant-level information about the use of key technologies. Of particular interest is data on the number of machine tools, broken down into Computer Numerical Controlled (CNC) versus non-CNC machine tools; and a binary variable recording whether a plant makes use of industrial robots.[14] MAG has collected data about

[14]MAG also collect more detailed information about these technologies, which we do not currently have access to. For example, they gather data about the number of CNC and non-CNC tools used for cutting vs forming, as well as a breakdown of whether a plant uses machine tools for various specific activities including drilling, grinding, milling, bending and/or pressing. They also collect a more detailed breakdown of industrial robots into

CNC machine tools since 2003, while data on industrial robots was added in 2012. However, we only use data on technology use from 2005 and 2014 respectively, since most sites are interviewed at least once every three years, allowing us to construct a reasonably accurate baseline of machine use.[15] In the subsequent analysis we limit our sample to manufacturing production sites, defined as sites that use some kind of mechanical tool. Dropping sites that never report positive tool usage in any period reduces our sample by about 40%. These non-tool using sites are likely to represent a combination of corporate offices, sales rooms, wholesalers and intermediary firms that are connected to manufacturing and hence of interest to MAG's customers, but not relevant to our analysis.[16] We also impute missing values of key employment and tool use variables using a carry-forward imputation.

2.4.1 Levels and trends in CNC and industrial robot use

Table 2.1 provides some descriptive statistics about our dataset, focusing on the manufacturing production sites that are the focus of the subsequent analysis. Our data contains 368,914 site-year observations, covering 26,974 unique sites. Of these sites, just under half use CNC machine tools at some point, while just over 6% use industrial robots. Consistent with the wider literature, we find that plants that use these advanced technologies have more employees than those that do not, though the difference is much larger for industrial robots than for CNC machines: plants that have at least one CNC machine employ 79 people on average, compared to 76 across all plants; whereas plants with robots have 234 employees on average.

four categories: Automated Handling or Storage Systems; Assembly/Welding Robots; Painting/Finishing Robots; Collaborated Robots. In recent years, they have started collecting data about the use of 3D Printing Machines and Plastics Machines.

[15]When a new question is introduced, every site is recorded as either having 0 or a positive number of tools. As a result, we cannot directly distinguish plants that have no tools from those that have not yet been sampled since the question was introduced. Since most plants are sampled at least every three years, we only use technology data from the third year after a question has been introduced.

[16]In practice, whether or not we include these sites does not make a significant difference to our results.

Table 2.1: Descriptive Statistics of UK Manufacturing Plant Survey

	Selected Years			All Years by Technology Use		
Variable	2005	2014	2023	All	CNC Users	Robot Users
Sample Size						
Number of Observations	20,087	17,923	13,133	368,914	229,697	38,079
Number of Sites	20,087	17,923	13,133	26,974	15,341	1,996
Number of Firms	19,549	17,516	12,934	26,180	14,970	1,947
Employment						
Mean	81.1	69.9	76.9	75.7	79	236.7
Total	1,629,658	1,253,437	1,010,213	27,913,290	18,147,636	9,013,907
p5	2	2	3	2	3	6
p25	7	7	8	7	8	30
Median	20	20	20	20	20	82
p75	60	50	55	55	55	230
p95	330	269	300	300	300	828
Technology Use						
Employment with CNC (%)	52	54.7	60.4	54	83	66.4
Employment with Robots (%)	-	14	27.1	22.8	26.5	63.1
Plants with CNC (%)	46.6	50.3	55.9	49.3	79.1	61
Plants with Robots (%)	-	3.9	8.3	6.2	7.2	51.4
Total Machine Tools	348,574	254,847	189,055	5,683,642	4,773,935	1,138,851
Total CNC Machine Tools	83,684	80,976	69,525	1,652,840	1,652,840	424,355
Tools per Worker	0.2	0.2	0.2	0.2	0.3	0.1
CNC Tools per Worker	0.1	0.1	0.1	0.1	0.1	0

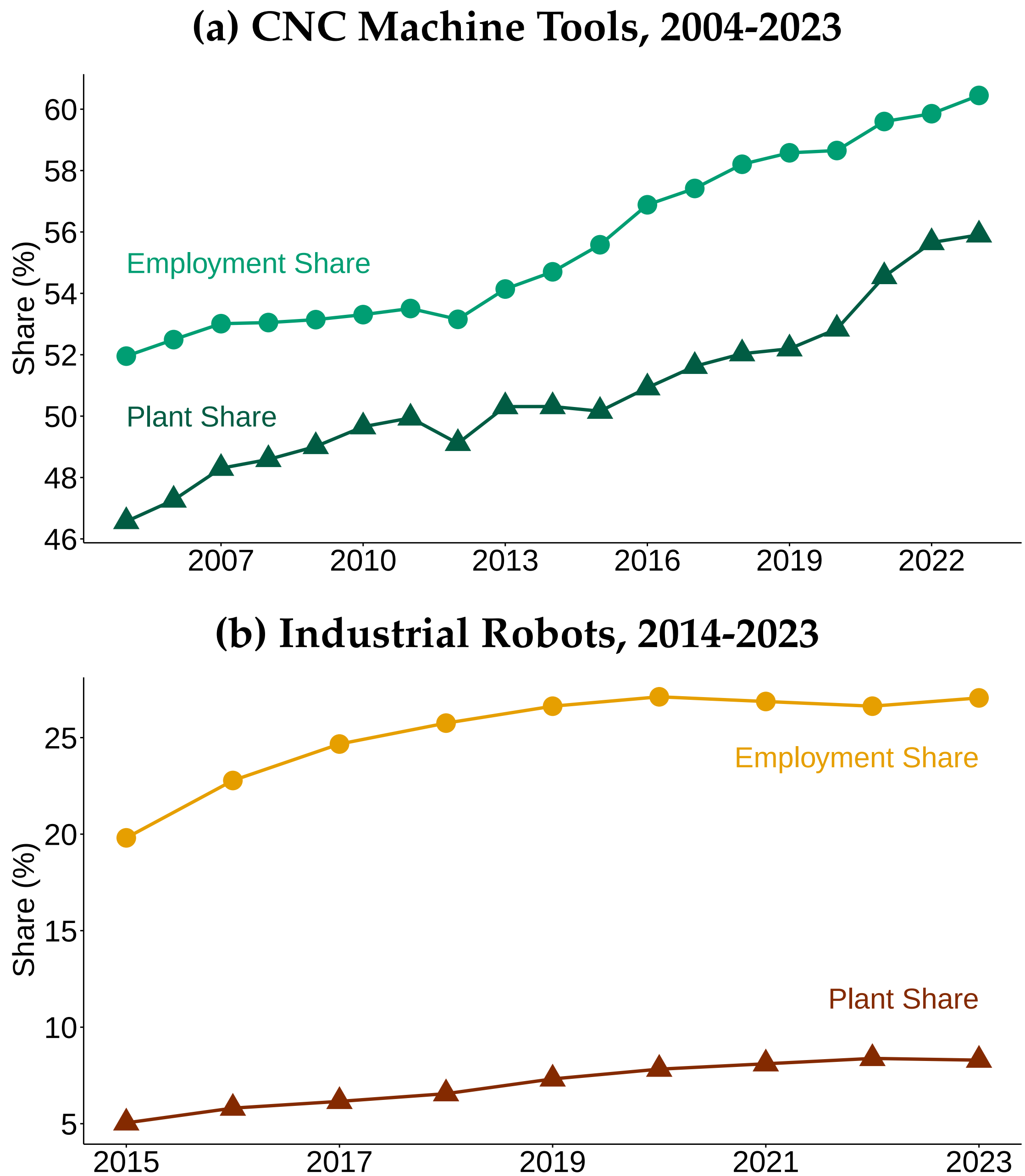
Notes: This table presents summary statistics for 'manufacturing production' plants in the MAG survey, as defined in the text. Reliable information about robot use is only available from 2015. 'CNC Users' are sites that have ever used CNC machines, 'Robot Users' are sites that have ever used robots.

Although CNC machine tools were already a relatively mature technology by the early 2000s when our data begins, the past 20 years have seen a continued increase across industries. Figure 2.2a shows that the share of manufacturing production plants using at least one CNC machine increased from about 47% in 2005 to nearly 56% in 2023, while the share of employees who work in such establishments increased from about 52% to 60%. The share of plants using industrial robots has increased from barely 4% in 2014 to 8% in 2023, while the share of employees working in plants that use industrial robots jumped from around 14% to more than 27% over the same period.

As we can see in Figure 2.3, these technologies are used more heavily in some parts of manufacturing than others. CNC machines are most prevalent in metal-heavy sectors like mechanical engineering, motor vehicles and metal manufacturing, where a clear majority of workers are employed in plants with at least one CNC machine. Industrial robots are heavily used in many of the same metal-heavy sectors, but also in other sectors where CNC machines play little role. In the "food, drink and tobacco" sector, for example, almost 43% of employees work at sites with industrial robots, compared to just 2.5% for CNC machine tools. These differences likely reflect the different capabilities of these technologies: while CNC machines are designed to cut and shape metal, industrial robots have a much wider range of possible applications. As we can see in Figure 2.4, most manufacturing sectors have seen an increase in the use of both technologies during this period, albeit at different rates.

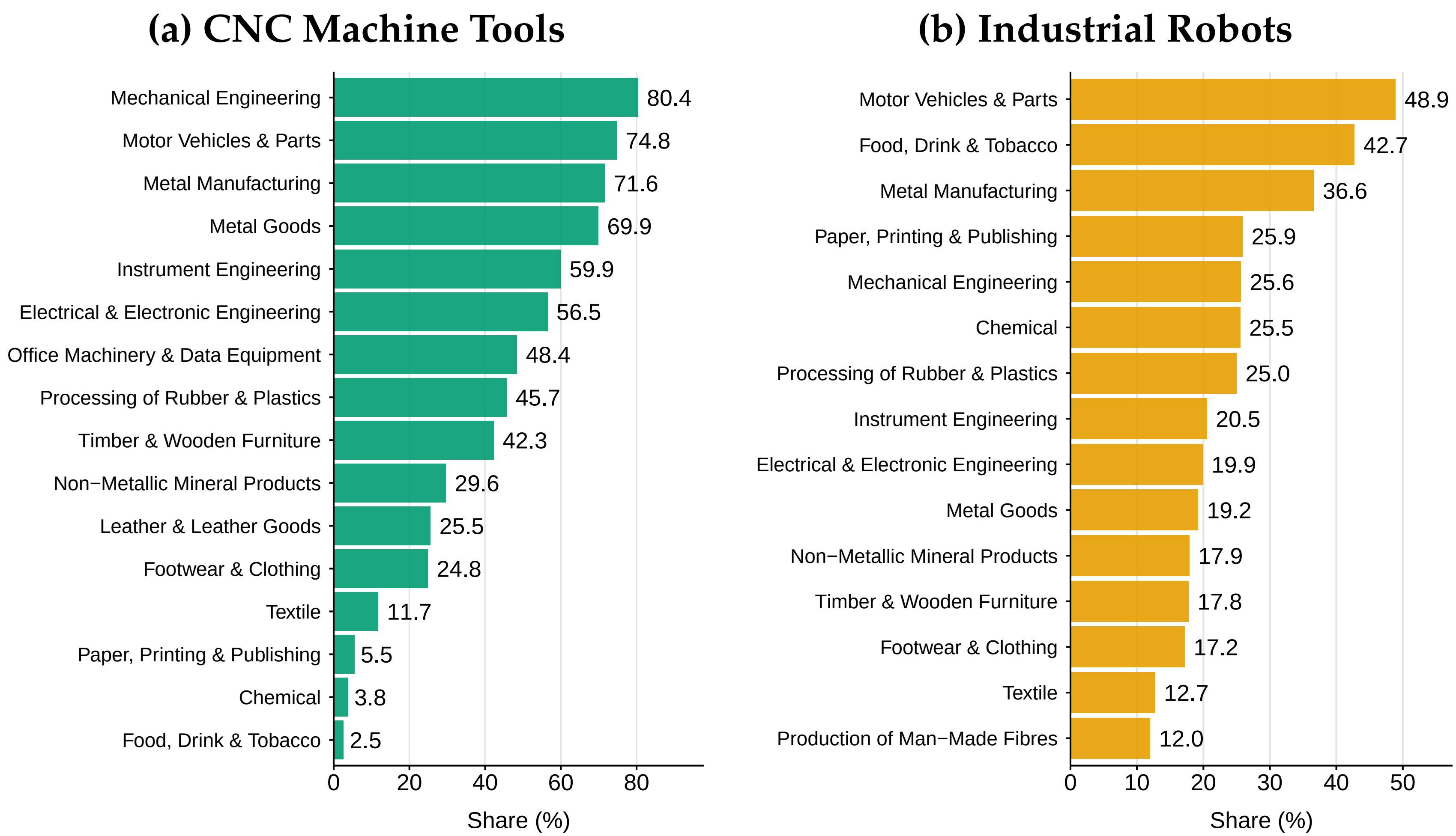
Our empirical strategy, which we discuss in the following section, takes advantage of the growing diffusion of these technologies over the past twenty years. Variation in the pace of technology adoption across industries highlights the importance of controlling for industry-level employment trends that might otherwise confound our estimates.

Figure 2.2: Automation Technology Diffusion in UK Manufacturing



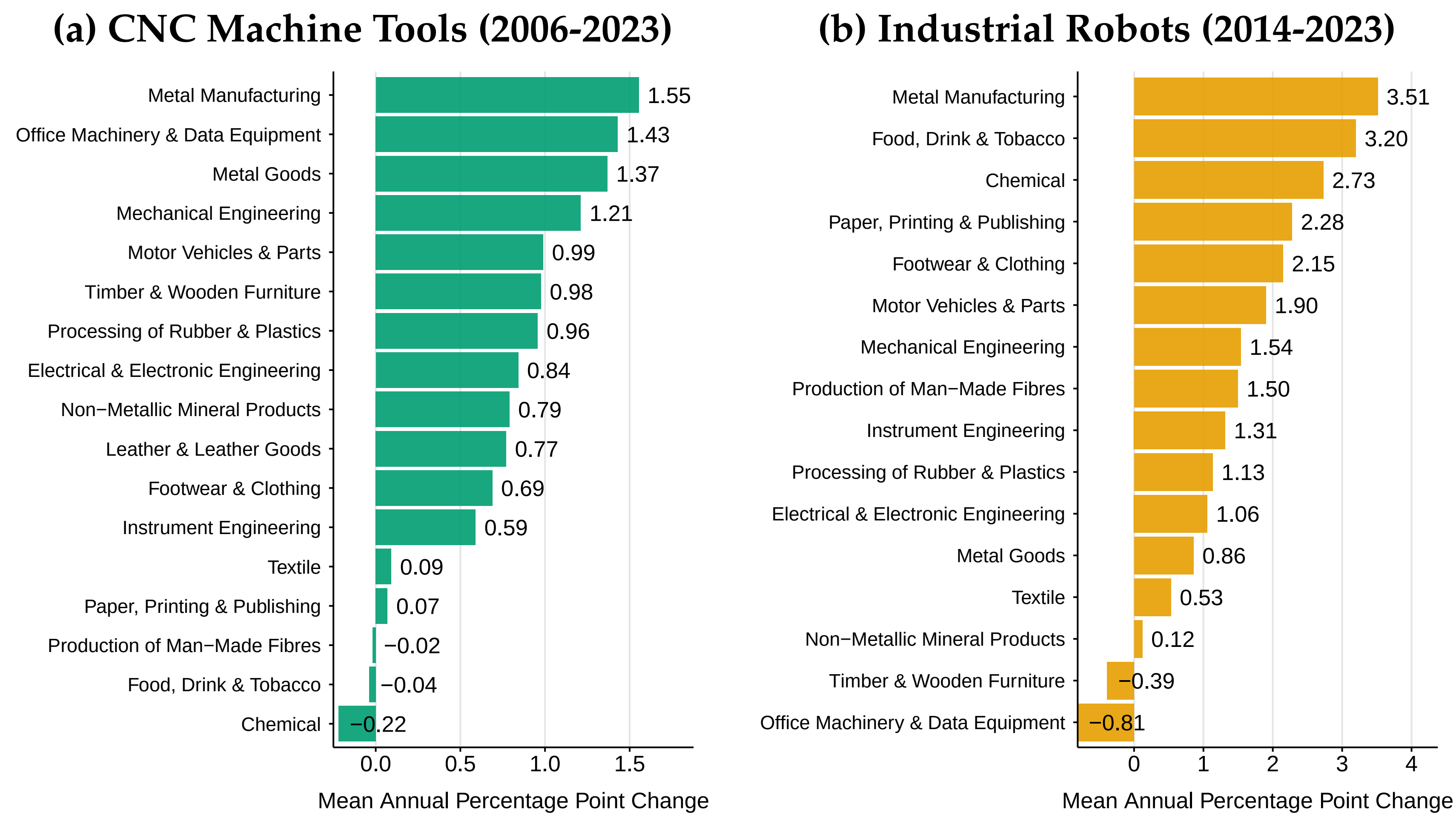
Notes: "Plant Share" is the share of plants with at least one CNC machine or industrial robot, based on the MAG survey. "Employment Share" is the share of employees at such plants. The sample is limited to "manufacturing production" plants in the MAG survey, as defined in the text. Reliable information about robot use is only available from 2015.

Figure 2.3: Share of Workers at Plants with Automation Technology in 2023, by Industry



Notes: share of employment in plants with at least one CNC machine or industrial robot. Shows top-20 manufacturing industry divisions (2-digit SIC codes). The sample is limited to "manufacturing production" plants in the MAG survey, as defined in the text.

Figure 2.4: Annualized Change in Share of Workers at Plants with Automation Technology, by Industry



Notes: annualised change in the share of employment in plants with at least one CNC machine or industrial robot. Shows top-20 manufacturing industry divisions (2-digit SIC codes). The sample is limited to "manufacturing production" plants in the MAG survey, as defined in the text.

2.5 Empirical approach

In this section we describe our empirical strategy for estimating the impact of the growing use of CNC machines and industrial robots on plant-level employment. We exploit the rich panel structure of our dataset using an event-study design, comparing the evolution of employment at focal plants after an "automation event" with a control group of non-automating sites. We do this using the Local Projection Difference-in-Differences (LP-DiD) estimator proposed by Dube et al. (2023), which addresses some of the well-known issues with standard two way fixed effects estimators in settings like ours with staggered treatment timing and heterogeneous treatment effects.

2.5.1 Automation events

We study the effect of two distinct types of automation event: **adoption events**, which occur when a firm acquires a given technology for the first time, and **expansion events**, which occur when a firm increases its stock of that technology having already adopted it in an earlier year.

Formally, we define an adoption event using a binary treatment variable $D^{adopt}_{i,t,A} \in \{0,1\}$, which switches from 0 to 1 in the first year that plant i uses automation technology $A \in$ {CNC Machines, Industrial Robots}:

$$D^{adopt}_{i,t,A} = \begin{cases} 1 & \text{if technology } A \text{ is used at any date } s \leq t \\ 0 & \text{otherwise} \end{cases}$$

An adoption event is then defined as the first difference of this variable, $\Delta D^{adopt}_{i,t,A}$, indicating the year a plant moves from non-use to use of the technology. As is standard, we model treatment status as absorbing: once adopted, $D^{adopt}_{i,t,A}$ remains 1 in all subsequent periods.

We define expansion events similarly, using a binary variable $D^{expand}_{i,t,A} \in \{0,1\}$, which switches from 0 to 1 in the first year that a plant increases its number of CNC machines, having already adopted the technology in an earlier year:

$$D^{expand}_{i,t,A} = \begin{cases} 1 & \text{if } Tech_{i,s,A} > Tech_{i,s-1,A} > 0 \text{ for any } s \leq t \\ 0 & \text{otherwise} \end{cases}$$

As with adoption events, we model the expansion indicator as absorbing. Since we observe machine counts only for CNC technologies, we are not able to define expansion events for industrial robots.

Adoption and expansion events capture distinct margins of technological change, both of which are central to understanding the labour market effects of automation. Adoption events provide a clean empirical analogue to automation as defined in the task-based framework discussed in Section 2.3, which conceptualizes automation as a reallocation of tasks from labour to capital. Our definition of adoption—capturing the first time a plant uses a given automation technology—corresponds closely to this notion of a discrete qualitative shift in the allocation of tasks from humans to machines. By contrast, expansion events may reflect either further automation at the extensive margin i.e. the replacement of humans in additional tasks, or intensive margin investments i.e. the scaling up of already automated tasks. While this limits their usefulness for identifying canonical automation effects, expansion events are of independent interest because they provide an opportunity to investigate learning effects.

As shown in Table B3, our dataset includes a substantial number of both types of events: we observe 4,355 CNC adoption events, 1,826 robot adoption events and 8,240 CNC expansion events.

2.5.2 The Local Projection Difference-in-Differences estimator

Since different plants experience automation events at different points in time, our setting is characterized by staggered treatment timing. Moreover, treatment effects are likely to vary with the number of periods since a treatment event (the "horizon") as plants adjust their production processes, organizational behaviour and employment. We also expect heterogeneity in treatment effects across plants even over the same horizon, reflecting unobserved time-varying factors such as the quality of managerial practices and introduction of new products by downstream or upstream producers.

There is now a large literature on the challenges of estimating event-studies in such a context, and in particular the problems associated with the standard Two Way Fixed Effects (TWFE) framework. These estimators can produce misleading results because they include all available non-treated units in the control group, meaning previously treated units whose outcomes may still be affected by prior treatment are included as controls for currently treated ones.

If the effect of treatment grows or shrinks over time, this can distort the comparisons. For example, treated units might appear to do worse than the control group simply because the "controls" are benefiting from prior treatment. These "forbidden comparisons" can lead not just to biased estimates, but to ones that have the wrong sign entirely.[17]

While several estimators have been proposed to address these issues, we use the Local Projection Difference-in-Differences (LP-DiD) estimator developed by Dube et al. (2023). As the name suggests, this estimator uses the local projection approach to estimating dynamic treatment effects, which relies on estimating separate regressions for each time horizon relative to treatment. At the same time, it modifies this approach to avoid forbidden comparisons, requiring that the units included in the control group are either never-treated units or have been treated sufficiently far in the past that their current outcomes are no longer influenced by the treatment. LP-DiD has a number of advantages compared to other similar estimators: it is computationally efficient and fast to implement; its identification assumptions are transparent and easy to understand; and it is highly flexible, making it easy to vary weighting schemes, choose alternative pre-treatment base periods, and pool treatment effects across different horizons. As Dube et al. (2023) emphasize, many other recent DiD estimators can be replicated as special cases of LP-DiD by adjusting these parameters.

Our preferred specification involves estimating the following regression:

$$y_{i,s,t+h} - \overline{y}_{i,s,t-4:t-1} = \beta_h \, \Delta D_{i,t,A} + \theta_{s,t} + \gamma \, \Delta y_{i,s,t-1} + e^h_{i,s,t}, \quad \forall \, h \in \{-4, -3, \ldots, 4\} \tag{2.1}$$

while restricting the sample to observations that are either:

$$\begin{cases} \text{newly treated} & \Delta D_{it} = 1, \\ \text{clean control i.e. not yet treated at } t+h & D_{i,t+h} = 0 \end{cases}$$

In the above, $y_{i,s,t+h}$ is total employment for plant i in industry s, h years after the treatment period t, and $\overline{y}_{i,s,t-4:t-1}$ is the average outcome over the pre-

[17]For a detailed discussion of these issues and potential solutions see de Chaisemartin and D'Haultfœuille (2024, 2022); Roth et al. (2023). A variety of alternative estimators that do not suffer from these difficulties have been proposed, and in Section 2.6 we show that our core results are robust to four key alternatives.

treatment periods $t-4$ to $t-1$. $\Delta D_{i,t,A}$ captures a generic automation "event", defined as a change in the underlying treatment indicator from $t-1$ to t (as above, where needed, we distinguish adoption and expansion events more explicitly as $D_{i,t,A}^{\text{adopt}}$ and $D_{i,t,A}^{\text{expand}}$). Our coefficient of interest is β_h which measures the treatment effect h periods after the treatment, and we study a time horizon from four years prior to four years after a given event. Although we estimate a variety of specifications to test the robustness of our results, in our main specification we include year-by-industry fixed effects $\theta_{s,t}$ as well as a lag of the differenced outcome variable $\Delta y_{i,s,t-1}$. As discussed below, these help mitigate a number of possible threats to identification, and to strengthen the plausibility of giving a causal interpretation to our results.

With these controls in place, we can interpret our β^h coefficients as the estimated difference in outcomes between plants that experience an automation event and those that do not, relative to pre-treatment trends and other plants in the same year and industry.

2.5.3 Identification

As with all difference-in-differences estimators, there are two critical identification assumptions

1. **Conditional parallel trends:** absent treatment, outcomes for treated and control units would have followed similar trends, conditional on covariates.

2. **Conditional no anticipation:** units do not change their employment behaviour in anticipation of treatment.

Under these assumptions, the β_h coefficient from equation (2.1) consistently estimates a weighted average across all treated cohorts of the Average Treatment Effect on the Treated h periods after an automation event. If these conditions hold, we would expect our estimated pre-treatment coefficients to be flat and close to zero. While this is not a sufficient condition for identification, it provides a useful first check on the plausibility of the assumptions.

There are several possible threats to identification, which we seek to mitigate. The fundamental concern is that companies which adopt new technologies might, for a variety of reasons, already be on a faster growth trajectory, thus invalidating the key parallel trends assumption.[18] It is possible, for example, that companies which adopt CNC machines or industrial robots are disproportionately concentrated in industries that are on more rapid growth trajectories. We include a full set of industry-year fixed effects in our main specification to account not just for secular industry-level trends that may be correlated with technology adoption, but for time-varying industry-level shocks, like a sudden spike in global demand for cars, that might simultaneously raise the probability of technology adoption and affect plant employment.

Even within a given industry, technology adopters might differ in unobserved ways that also affect their employment trajectory. They might, for example, have more capable or better informed managers who are simultaneously more likely to invest in new technology and pursue a range of other strategies that drive faster employment growth. Alternatively, firms in faster growing market segments may be more likely to both expand their employment and adopt new technologies, simply to keep up with rising demand. These concerns help to motivate our inclusion of a lag of the differenced outcome variable, which allows us to account for *firm-level* differences in pre-treatment trajectories that might otherwise bias our results (we effectively partial out these dynamics to better isolate the causal effect of treatment). Dube et al. (2023) recommend this approach in contexts like ours where the outcome variable is likely to be subject to momentum or drift, following a well-established tradition in panel data models.

As we shall see, conditional on these controls, we find that treated and control units exhibit broadly parallel pre-treatment trends. But even if pre-trends appear similar, bias may still arise if treatment coincides with unobserved shocks. For instance, it is possible that a spike in demand leads companies to immediately install new machines; or that new managers arrive in a plant and instantly install new technologies, while also overhauling other workplace practices that affect employment levels. However, such contemporaneous shocks are unlikely to explain our results, because decisions to install CNC machines and industrial robots typically involve substantial planning, capital investment approval processes, and installation lead times that can span

[18] See, for example, Acemoglu et al. (2023), who find that US companies that adopt new technologies including AI and robotics were already larger and growing faster than other similar companies that did not; and Bessen et al. (2023), who documents that Dutch companies that experience an automation event have a higher average growth rate than those which do not.

multiple quarters or even years.[19]

While our discussion so far has focused on threats to the parallel trends assumption, we also consider the possibility of bias arising from anticipatory adjustments to employment prior to treatment. As we have just commented, automation events are likely to be the result of planning over multiple years, which raises the possibility that plants might hire new or different kinds of workers in anticipation of the arrival of new technologies, leading to biased results. Such anticipation effects would also be mitigated by including a lag of the differenced outcome variable, and even without including this lag, we do not find evidence of significant anticipation effects.

In the absence of a valid instrument or other (quasi-)experimental estimation strategy, we cannot justify a strictly causal interpretation of our results, and identifying such a strategy is an important priority for future research. Even so, it is worth noting that Aghion et al. (2023c, 19-20), who combine an event-study design similar to our own with a shift-share IV strategy leveraging pre-determined supply linkages and productivity shocks, find that OLS-based estimates remain positive but of smaller magnitude. In other words, their results suggest that if anything, a strategy such as ours might under-estimate the positive effects of automation at the firm level. In any case, while we cannot rule out all sources of endogeneity, the absence of differential pre-trends in most specifications, the careful use of industry-time controls, and the inclusion of firm-level dynamics all help to strengthen the plausibility of a causal interpretation.

2.6 Firm-level results

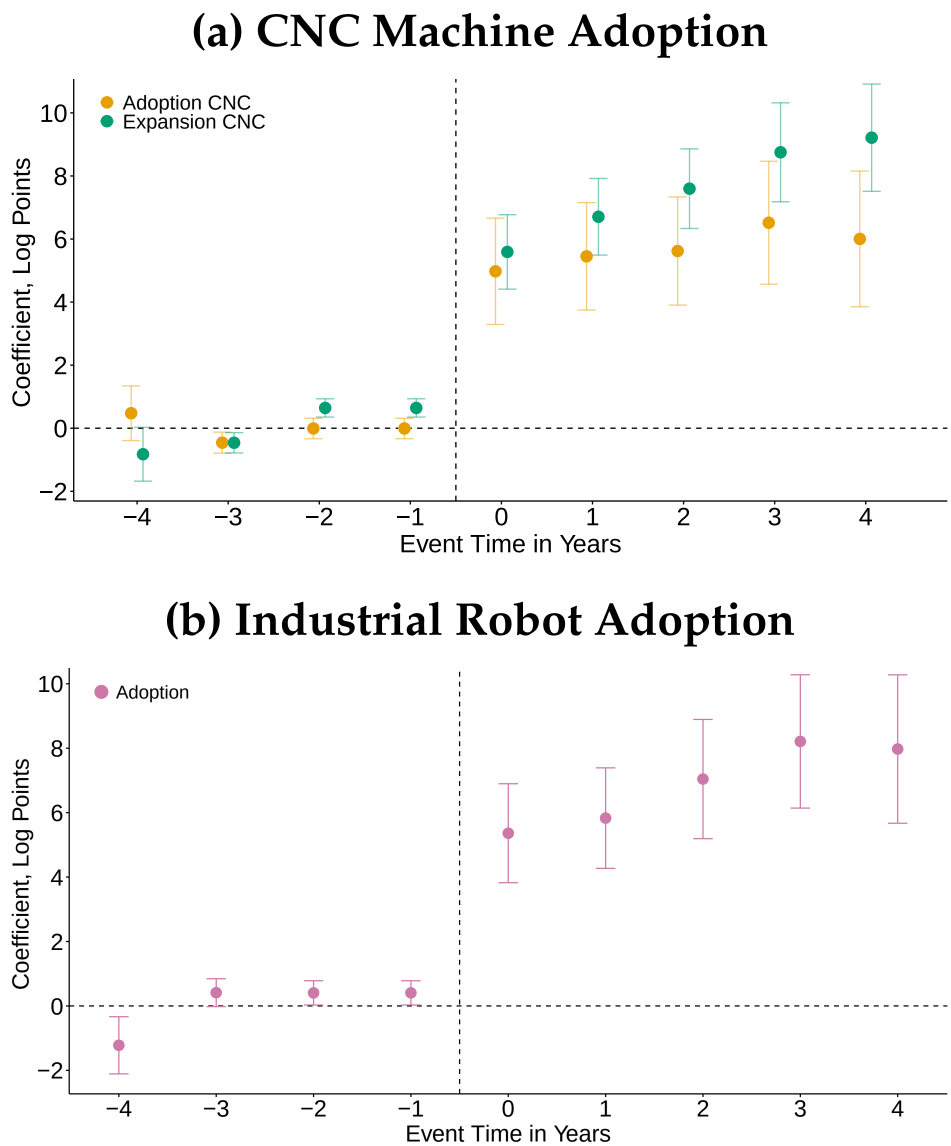
Figure 2.5 presents our event-study estimates of plant-level employment around automation events. For CNC adoption, treated plants experience an immediate 5% increase in total employment relative to controls, growing modestly to around 6% four years after adoption. Industrial robot adopters show a similar pattern: a 5% jump in employment in the year of treatment that rises to

[19] The high cost of the organizational changes required to implement a successful adoption of ICT are emphasized in Brynjolfsson and Hitt (2000); we expect that the changes required to implement a new machine tooling system are similar. For industrial robots, Humlum (2021) estimates that the total cost (equipment and non-equipment) of robot adoption for the typical firm can be up to 25% of the firm's sales, with the equipment component estimated at around 13%.

roughly 8% after four years. In other words, for both technologies, firm-level productivity effects appear to dominate displacement effects, leading to a positive overall impact on employment. Importantly, our event studies support the two critical identification assumptions: treated and control firms exhibit broadly parallel pre-treatment trends, conditional on our industry and lagged dependent variable controls; and there is no evidence of significant anticipation effects.

Figure 2.5 also displays the results for CNC expansion events, which occur when a plant increases the total number of CNC machines from an already positive number. Although there is modest evidence of a positive pre-treatment trend, CNC expansion events are associated with an even larger increase in total plant employment of about 6% in the year following an expansion event, rising to 9% after four years. In contrast to adoption events, the increase in employment following an expansion event does not plateau after four years, indicating that such events may not only increase the level of employment, relative to controls, but also shift firms onto a different growth trend.

Figure 2.5: Plant-level Event Study, Total Employment



Notes: These figures show LP-DiD estimates of employment effects following automation events, as defined in the text. Panel (a) uses $N = 368,914$ observations from 2005-2023, with 4,355 adoption events and 8,240 expansion events. The post-treatment pooled coefficient is 5.058^{***} (adoption) and 6.821^{***} (expansion). Panel (b) examines industrial robot adoption using $N = 229,731$ observations from 2010-2023, of which $1,826$ experienced an adoption event. The post-treatment pooled coefficient is 6.98^{***}. Both designs use 6-digit Industry Codes $\times$ Year fixed effects and a lag of the differenced outcome variable in period $t-1$ as controls, as per equation 2.1, with standard errors clustered at the plant-level. Vertical bars represent 95% confidence intervals.

Although the positive difference between expansion and adoption events is modest, it is consistent with the idea that plants undergo a learning process after adopting new technologies. As discussed in Section 2.3, in learning-based models it takes time for firms to accumulate the human, organizational, and process-specific capital needed to make the most of a given technology, and to undertake the complex work of re-shaping production processes and cultivating new supplier and customer relationships Atkeson and Kehoe (2007). Once a plant has climbed the steepest section of this learning curve, further investments — such as adding more CNC machines — may deliver disproportionately larger returns, since the firm already knows how to integrate new capacity into its routines and can redeploy it immediately into higher-value tasks.

2.6.1 Alternative specifications and robustness

Figure 2.6 compares our baseline technology adoption results from the previous section to four alternative specifications of the LP-DiD estimator. As explained in Section 2.5, our primary concern is to control for employment-related factors that are correlated with treatment (technology adoption), and hence satisfy the conditional parallel trends assumption.

To recap, our baseline specification includes year by industry fixed effects, as well as a lag of the differenced outcome variable. The latter mechanically flattens the estimated pre-treatment coefficients in $t-1$ and $t-2$. To see how these controls affect our results we look first at the simplest reasonable specification with year fixed effects only (in green). In contrast to our baseline, we find that treated units have a slightly positive pre-treatment employment trend; and that this pre-trend is more pronounced for sites that adopt CNC machines than industrial robots. Nevertheless, we continue to observe a clear jump in employment compared to untreated units in the year of treatment, followed by a gradual further increase over the subsequent four years.

In our third specification (in purple) we add a full set of year by industry controls, to allow for the possibility that technology adoption is correlated with broader industry-level employment trends, but once again omit the lag of the differenced outcome variable. Compared to our specification with year fixed effects only, this further reduces the observed pre-treatment trends for CNC adopters (though not for robots), suggesting that treated units are slightly more likely to be in faster-growing industries; and for both CNC and robots we continue to observe a substantial and sustained jump in employment after treatment, compared to untreated units. In order to demonstrate that our results are not being driven by the lag of the differenced outcome variable, in Table B6 we report the four year pooled post-treatment coefficient with no lag and with one, two and three lags. Varying the number of lags makes little difference to our results.

Our fourth specification (in blue) adopts a slightly different approach: instead of using the lag of the differenced outcome variable to control for differential trends immediately prior to treatment, we control for linear plant-specific time trends in employment across the whole span of our data.[20] In this specification, the modified parallel trends assumption is that, after removing site-specific trends, treated and control units would have been parallel in deviations from those trends; and we can interpret the estimated LP-DiD coefficients as the average effect of treatment relative to each plant's own linear employment trend. The estimated pre-treatment coefficients in this model are largely flat and close to zero; while the estimated treatment effects continue to be significant, albeit slightly smaller in magnitude than in our baseline model, and (thanks to the addition of many extra parameters) less precise.

Finally, we include a specification (in red) that repeats our baseline specification, but restricts the control group to "future adopters", in other words plants that eventually experience an adoption event for the technology in question. We do this by dropping "never-treated" units from our sample. This approach is inspired by Bessen et al. (2023), who argues that sites which adopt advanced technologies are likely to differ in unobserved ways from sites that never adopt. By comparing adopters with future adopters, rather than with never adopters, this specification exploits the *timing* rather than the incidence of adoption events, and is more likely to compare like with like. The downside is that the size of control group is significantly reduced. Again, this specification supports our headline findings: we find no evidence of differential pre-treatment trends between current and future adopters; and we find that

[20] This is accomplished by explicitly adding a unit fixed effect to the standard LP-DiD specification. As the authors of the relevant command explain, since unit fixed effects are already filtered out by the differencing of the outcome in LP-DiD, "adding unit fixed effects to the LP-DiD specification is equivalent to including unit-specific linear time trends". See http://fmwww.bc.edu/repec/bocode/l/lpdid.sthlp

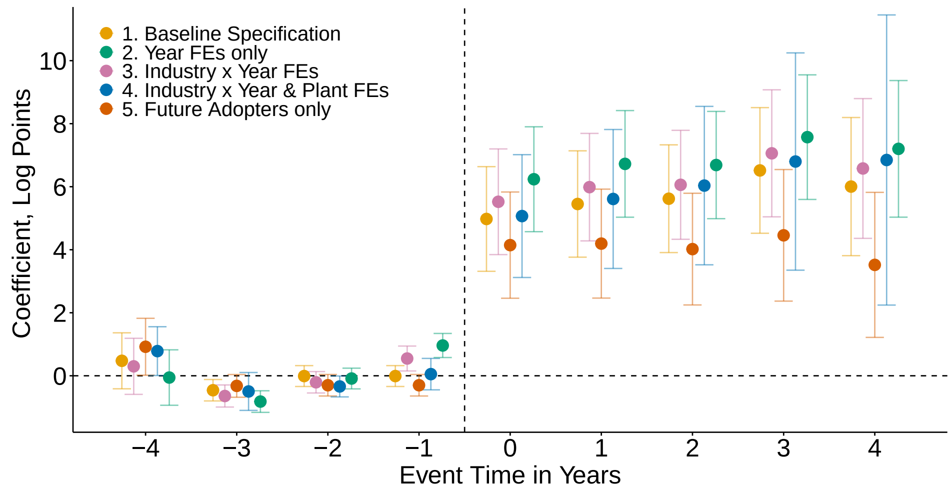
automation events are associated with a statistically significant and sustained increase in employment, albeit slightly lower than in our baseline.

In addition to these alternative specifications, we conduct a placebo test using our baseline specification. Specifically, we conduct fifty separate trials in which we randomly assign technology adoption events across the complete sample, preserving the temporal distribution of events throughout. In contrast to the alternative specifications above — which test the robustness of the conditional parallel-trends assumption by varying our controls — this exercise performs a simple randomization-inference check: the aim is to confirm that our results capture a genuine treatment effect rather than a mechanical artifact of the LP-DiD estimator or a chance fluctuation in the data. We report our results in Figure 2.7. On the right is the estimated impact of technology adoption after four years, as reported in Figure 2.5, where the standard error is depicted as normally distributed in blue shading. On the left (in green shading), we report the distribution of the four-year post-treatment coefficient across our fifty placebo trials. Our results are reassuring: for both CNC machines and robots, the distribution of placebo coefficients is centred around zero, and barely overlaps with the estimates from our baseline specification.

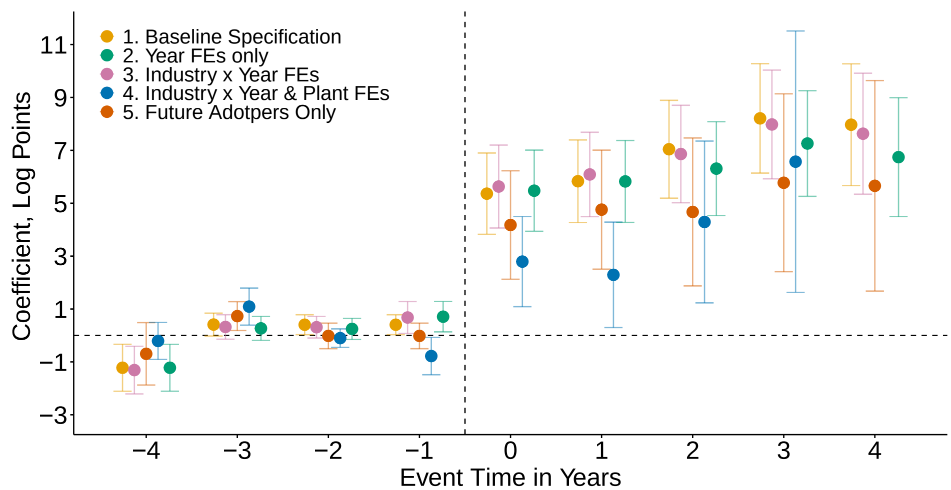
To complete our robustness tests, in Table B4 we show the pooled post-treatment coefficient obtained from estimating our core specification using four alternative estimators put forward in the recent literature on difference-in-differences estimation in the context of staggered treatments and heterogeneous treatment effects. While different estimators yield slightly different point estimates, the results are remarkably similar across these specifications. Finally in B5 we show that the significance of our core results is also robust to a range of alternative clustering strategies.

Figure 2.6: Alternative Plant-level Event Study Specifications

(a) CNC Machine Adoption



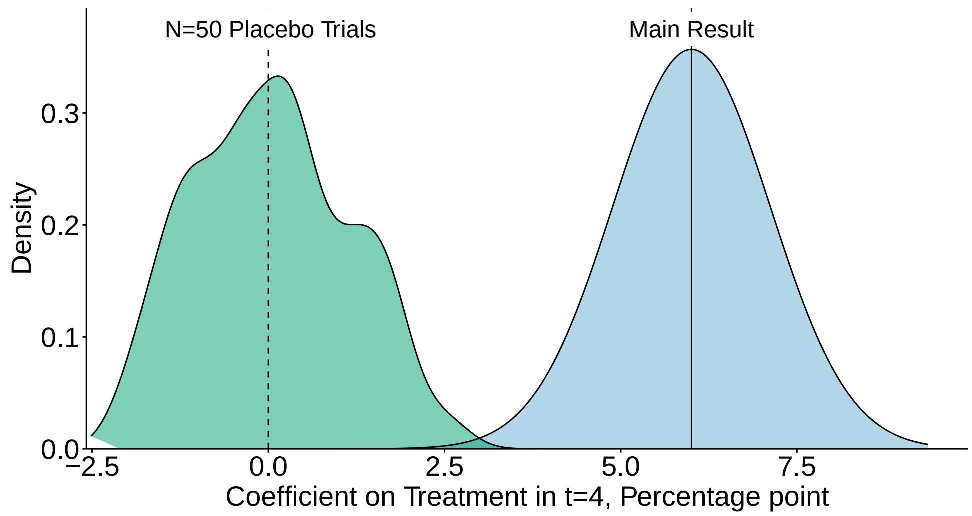
(b) Industrial Robot Adoption



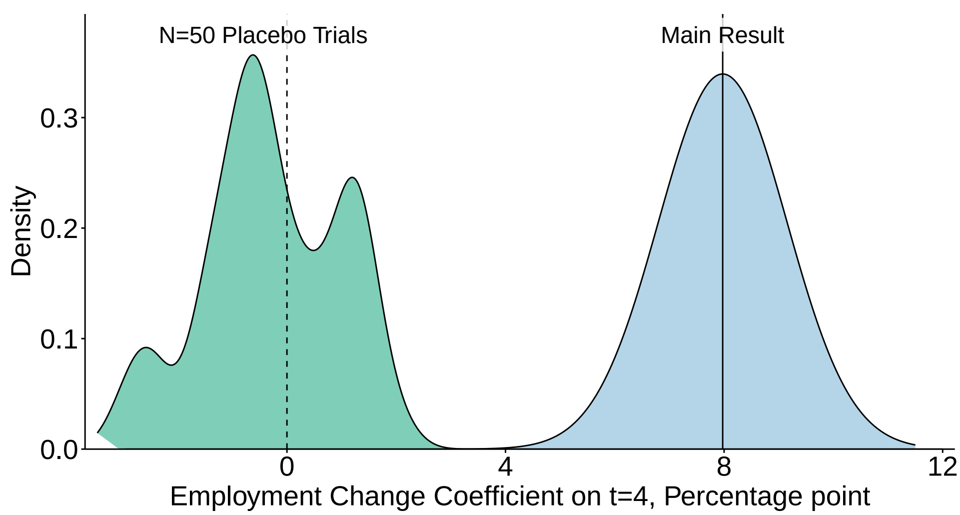
Notes: These figures compare five alternative LP-DiD specifications, as described in the text. The baseline specification is that from Figure (2.5), which includes Industry by Year fixed effects and a lag of the differenced outcome variable. Specification 2. includes Year fixed effects only; 3. includes Industry by Year fixed effects only; 4. includes Industry by Year fixed effects plus a linear site-specific trend; and 5. runs the baseline specification on the subsample of units who experience an adoption event at some stage i.e. it removes never-treated units from the control group, leaving only 'future treated' ones.

Figure 2.7: Placebo Test of Baseline Event-study Results

(a) CNC Machine Adoption



(b) Industrial Robot Adoption



Notes: This figure reports the coefficient estimate from $t = 4$ found in Figure (2.5), where the standard error is depicted as normally distributed in blue shading. Additionally, the left-most distribution in both panels shows a kernel-density plot of the same coefficient estimate from $N = 50$ placebo trials, where technology adoption events are randomised across the complete sample, preserving the temporal distribution throughout. The mean/median of the placebo-treatment coefficient estimates in panel (a) are $0.04/0.03$ and in panel (b) are $-0.28/0.43$.

2.6.2 Impact on employment composition

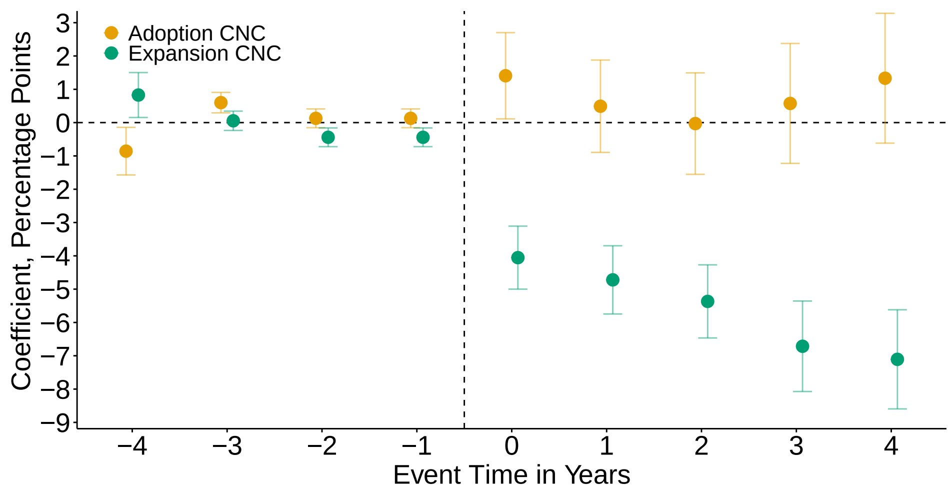
One of the key predictions of the task-based framework is that automation will reduce the share of employment among workers whose tasks are displaced. Our data single out employment of "manufacturing workers", who are most likely to be exposed to automation as a result of the installation of CNC machines and industrial robots.

In Figure 2.8 we report the results of our baseline LP-DiD specification, but replacing the outcome variable with the share of manufacturing employees at a given site. For adoption events, our results do not support the prediction of the task model. For CNC machines, there is no discernible effect of adoption events on the share of manufacturing employees. For industrial robots there is some evidence of a slight decline, but this follows a significant negative pre-trend, and in any case the individual post-treatment coefficients are not significantly different from zero.

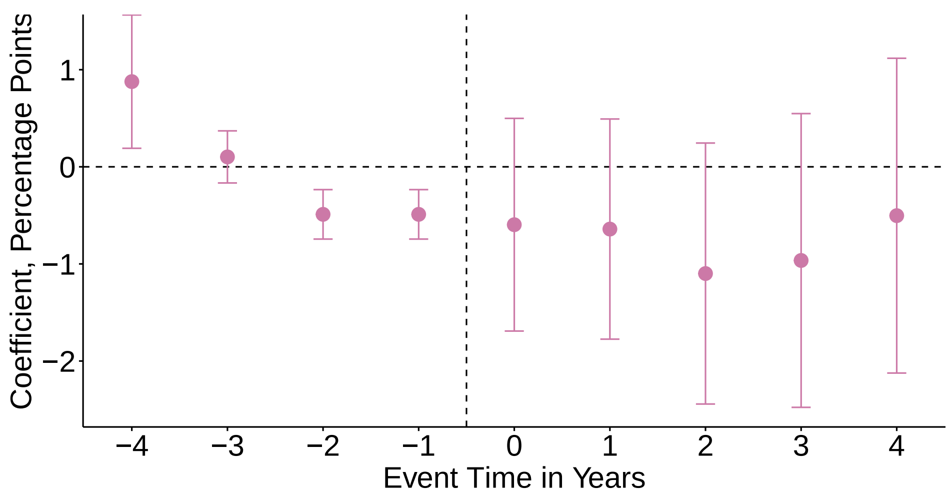
However, we find that CNC expansion events are associated with a large and statistically significant reduction in the share of manufacturing workers of around 8 percentage points after six years. This is further suggestive evidence that the distinction between adoption and expansion events is economically significant. These results are also consistent with the idea that plants which adopt new technologies undergo a learning process. On this view, it takes time for firms to work out how to reorganise their production processes in ways that make the most of a given new technology. If this is the case, we would expect first adopters to use technologies in ways that are more easily adapted to existing work processes, and more likely to complement existing workers. However, as firms accumulate more know-how about a new technology, they are likely to see opportunities to redesign processes in more fundamental ways that might automate away existing tasks. Investigating this mechanism further, both theoretically and empirically, is an important question we leave for future research.

Figure 2.8: Plant-level Event Study, Share of Manufacturing Employment

(a) CNC Machine Adoption



(b) Industrial Robot Adoption



Note: These figures show LP-DiD estimates of employment-composition effects following automation events, as defined in the text. The outcome variable is the share of manufacturing workers to higher skilled designers/engineers. Panel (a) uses $N = 368,914$ observations from 2005-2023, with 4,355 'adoption' events and 8,240 expansion events. The post-treatment pooled coefficient is 1.541^{*} (adoption) and -5.854^{***} (expansion). Panel (b) examines industrial robot adoption using $N = 229,731$ observations from 2010-2023, of which $1,826$ experienced an adoption event. The post-treatment pooled coefficient is -0.611. Both designs use 6-digit Industry Codes $\times$ Year fixed effects and a lag of the differenced outcome variable in period $t-1$ as controls, as per equation 2.1, with standard errors clustered at the plant-level. Vertical bars represent 95% confidence intervals.

2.7 Industry-level results

As we saw in our discussion of the task-based framework in Section 2.3, even if automation events are associated with an increase in employment at focal plants, this need not translate into an increase in industry-level employment. This depends on how automation at focal plants affects the employment of their competitors. On the one hand, it's possible that as automating firms become more productive, we see a reallocation of production away from less productive firms. A number of empirical papers have documented evidence of such "business stealing" effects, where automating companies expand at the expense of their competitors (Aghion et al., 2023c; Acemoglu et al., 2020; Koch et al., 2021). If this is the case, the eventual impact at the industry level will still depend on whether the expansion among automating firms outweighs the contraction of their competitors. While both Acemoglu et al. (2020) and Aghion et al. (2023c) find evidence of business stealing, they find opposite overall effects on industry employment. On the other hand, automation can also lead to positive spillover effects on competitor firms. If demand is elastic, lower prices as a result of automation may expand total output in the industry, for example, especially if capacity constraints or market frictions limit the dominance of automating companies. In addition, automation may increase demand for complementary inputs or services, some of which may be supplied by other firms in the same sector, or trigger broader productivity improvements through demonstration and diffusion effects.

In this section, we first test for evidence of business stealing, before looking directly at the impact of industry-level automation events on total industry employment.

2.7.1 Impact on competitor plants

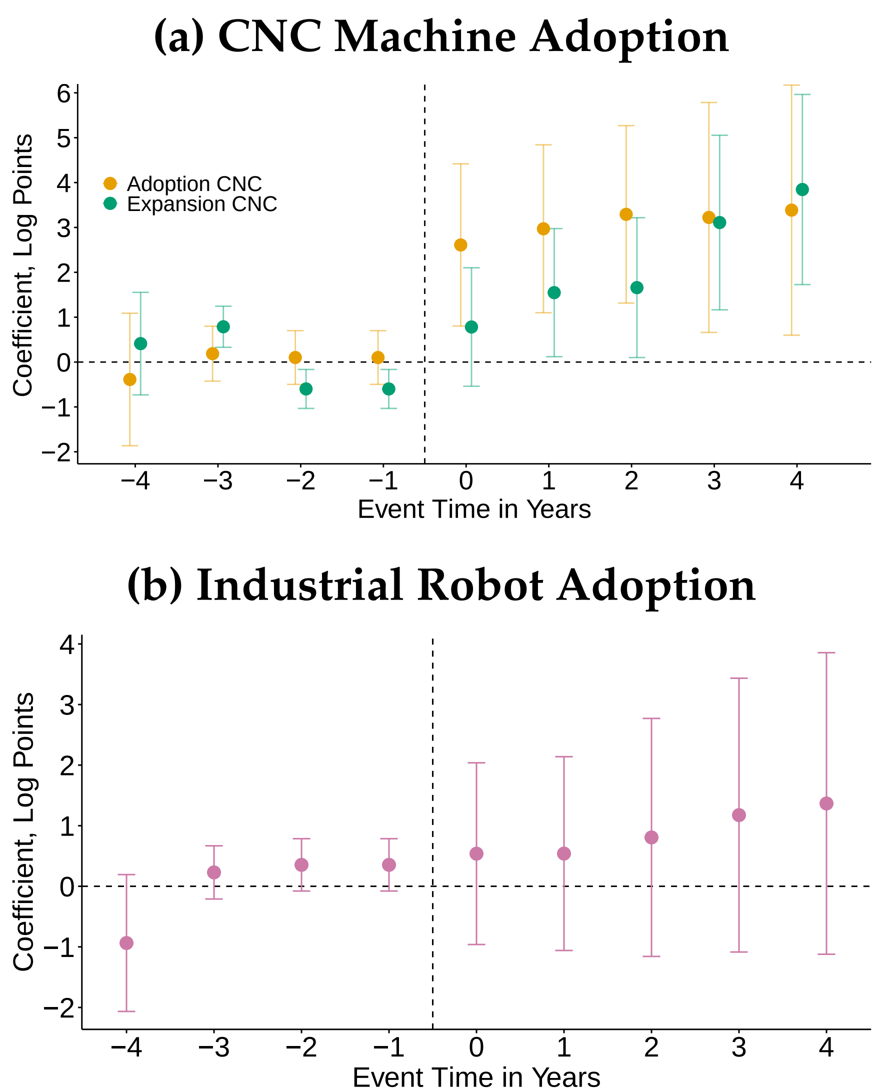
In order to look at whether automating firms steal business from their competitors, we first need to define who those competitors are. The best proxy for this in our data is whether two sites share a given industry code, which is a reasonably strong indicator that they are producing similar products and competing in the same market segments. We then follow the strategy proposed by Aghion et al. (2023c), who replace each firm's own employment outcome in their event study specification with the employment of its competitors. Specifically, we adapt our firm-level LP-DiD specification by replacing the outcome variable for each plant-year observation with the total employment among other firms in the same six-digit SIC code:

$$Y_{i,t,s}^{-i} = \sum_{\substack{j \neq i \\ s_j = s}} y_{j,t} \tag{2.2}$$

This specification captures how the employment of a treated firm's competitors evolves in response to automation events. Specifically, the estimated LP-DiD coefficients compare the change in competitors' employment for treated firms h years after treatment to the change in competitors employment' for untreated firms over other h-year periods in our dataset. If there is business stealing, we would expect treated competitors' employment to decline in the periods following a treatment event compared to the trajectory at untreated firms over similar time horizons. Conversely, if there are positive spillover effects, we would expect automation at one firm to lead to an increase in employment among its peers, relative to other firms and periods.

As we can see in Figure 2.9, we find evidence of positive spillover effects rather than business stealing. For CNC machines, automation at a focal plant is associated with a statistically significant increase in competitors' employment of around 3% after four years. This is true for both adoption and expansion events. The direction of travel appears to be the same for competitors of plants that adopt industrial robots, though the magnitude of the effect is smaller and not statistically significant.

Figure 2.9: Spillover Effects on Industry-Competitor Employment



Notes: These figures show LP-DiD estimates of the effect of focal plant automation events on industry-competitor employment, as described in the text. The outcome is the leave-out sum of employment among other sites in a given industry and year, defined in Equation 2.2. Panel (a) uses $N = 368,914$ observations from 2005-2023, with 4,355 adoption events and 8,240 expansion events. The post-treatment pooled coefficient is 3.315^{***} (adoption) and 2.416^{***} (expansion). Panel (b) examines industrial robot adoption using $N = 229,731$ observations from 2010-2023, of which $1,826$ experienced an adoption event. The post-treatment pooled coefficient is 0.636. Both designs use 6-digit Industry Codes $\times$ Year fixed effects and a lag of the differenced outcome variable, as per equation 2.1, with standard errors clustered at the plant-level. Vertical bars represent 95% confidence intervals.

A natural concern with these results is that they are simply capturing positive industry-level trends that are common to treated companies and their competitors in a given industry. So, for example, we might worry that companies that automate are more likely to be in faster growing industries, and that as a result employment among their peers is likely to grow faster than that of peers in other industries. We mitigate this concern directly by continuing to include industry and year fixed effects, and a lag of the differenced outcome variable. More importantly, if this was driving our results, it would show up in a positive pre-treatment trend among treated units, something we do not observe.[21]

2.7.2 Industry-level automation events

We now turn to look more directly at the implications of automation for industry employment, taking into account both the impact on focal firms and on their competitors. Following Aghion et al. (2023c), we adapt our LP-DiD design once again, but this time embracing a fully industry-level specification: studying the impact of industry-level automation events on total industry employment. Intuitively, we study what happens in an industry after a significant increase in the penetration of a given automation technology, taking into account the full set of direct, indirect and general equilibrium effects.

We define an industry-level automation event as a year in which we see a large increase in the share of employees working for companies that use a given technology, relative to all positive annual increases in this measure. Specifically, we first measure all positive annual changes in the share of employees working at plants using a given technology in each industry, and then define an industry automation event as an increase in this share above a given percentile threshold in the distribution of all positive year-on-year changes.[22] Focusing on the share of employees working at firms that use a given technology, rather

[21] Another potential concern is that the outcome variable for untreated units in this specification, namely total employment among all other firms in their industry, will include the employment of treated units in that industry. Since treated firms generally see a (relative) increase in employment after automation events, this will mechanically increase employment among their peer group. If anything, however, this would lead us to understate positive spillovers, since it would mechanically increase the outcome variable for untreated units when one of their peers is treated. To examine this concern we have also looked at specifications using total employment among never treated peer companies in the same industry, and find qualitatively similar results.

[22] For comparison, Aghion et al. (2023c) define industry-level investment events as an above-median annual change in the balance sheet value of industry equipment.

than simply the share of companies using that technology, means the influence of a given firm's adoption is weighted by its size. For example, an automation investment at a large multi-site employer will have a much larger effect on industry employment dynamics than adoption by a small fringe firm. Unlike at the firm-level, where adoption and expansion provide natural binary thresholds for defining events, the definition of an automation event at the industry level is inherently somewhat arbitrary. As such, we display our results for three thresholds: the top half, the top third, and the top quartile of positive year-on-year changes in the share of employees working at companies that use a given technology.

To examine the impact of these industry-level events we estimate the following adapted LP-DiD regression:

$$y_{s,t+h} - \overline{y}_{s,t-4:t-1} = \beta_h \, \Delta D_{s,t,A} + \theta_t + \gamma \, \Delta y_{s,t-1} + e^h_{s,t}, \quad \forall \, h \in \{-4, -3, \ldots, 4\} \tag{2.3}$$

while restricting the sample to observations that are either:

$$\begin{cases} \text{newly treated} & \Delta D_{s,t} = 1, \\ \text{clean control i.e. not yet treated at } t+h & D_{s,t+h} = 0 \end{cases}$$

In this specification, $y_{s,t+h}$ denotes total employment in industry s, h years after the baseline year t; $\overline{y}_{s,t-4:t-1}$ is the average of the outcome in the four years prior to treatment; $\Delta D_{s,t,A}$ is an indicator for an automation "event" in year t, defined as a large increase in the share of employees working at technology-using plants in industry s; θ_t captures year fixed effects, controlling for economy-wide shocks that might be correlated with automation events; and $\Delta y_{s,t-1}$ is the lagged first-difference of the outcome variable.[23] The coefficient of interest is β_h, which captures the dynamic effect of automation on industry-level employment h years after the event.

This industry-level LP-DiD approach allows us to trace out the evolution of employment following major increases in automation intensity, comparing industries that experience an automation event to those that remain untreated over a comparable time horizon. Of course, it's possible, for example, that

[23] Although we no longer include industry level controls, note that because LP-DiD uses differenced outcome variables it implicitly controls for industry fixed effects.

strong demand growth in a given industry drives both large automation events and an increase in employment growth relative to other industries. As before, the absence of significant pre-treatment trends is a necessary but not sufficient condition for satisfying the critical parallel trends and no anticipation conditions.

Our results are displayed in Figure 2.10. The first thing to note is that we do not observe significant pre-trends. Turning to our post-treatment coefficients, in most specifications the impact of automation events on industry-level employment is positive, and we can largely rule out negative employment effects. This is consistent with our finding that firm-level automation events increase employment at both focal firms, and at competitor firms within a given industry. Looking in more detail at the results for CNC machines, we find that the estimated effects vary substantially from one event threshold to the next. The largest and only statistically significant effect is based on a threshold set at the median of positive changes in the share of employees at CNC-using companies. When we focus on the top third or top quarter of changes, we get smaller effects, and lose significance.

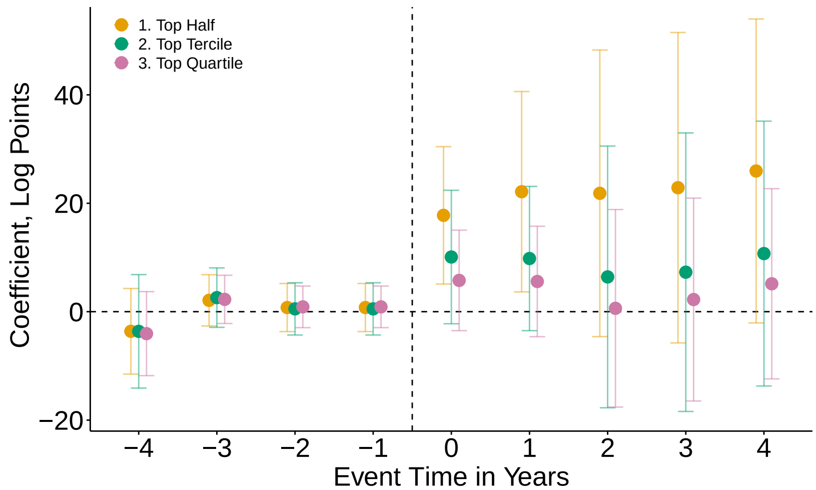
At first sight, it might seem surprising that more extreme automation events (those in the top quartile rather than the top half) appear to have a smaller impact on total industry employment. But it's important to remember that changes in these event thresholds affect the composition both of the treatment and control groups: when we shift to top-quartile events, industries that experience changes above the median but below the top-quartile, and which likely grow as a result, become part of the control group. This attenuates the estimated treatment effect, since industries in the revised control group are also likely to experience moderate automation-related growth.

Turning to our results for industrial robots, we find more consistent results across the three thresholds, with a gradual but clearly positive effect emerging four years after a major industry level automation event. As before, we estimate the largest effect when using the median threshold.

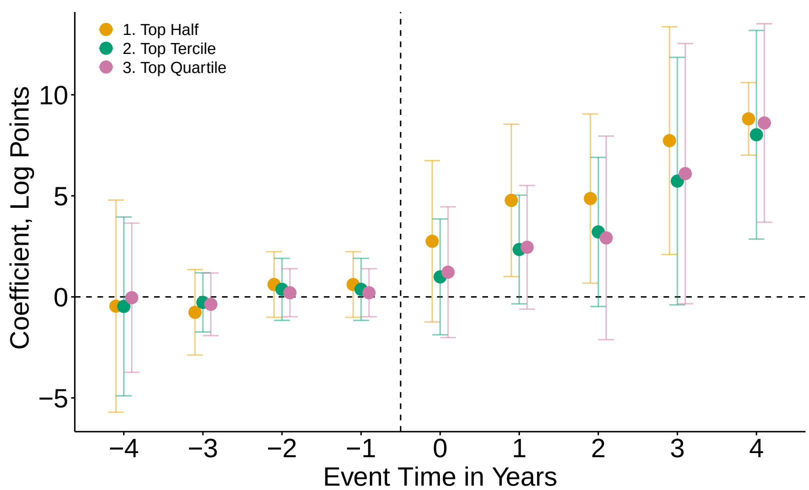
As a final step in the industry-level analysis, we look at the impact of the same set of industry-level automation events on the industry-level share of manufacturing employees. As we saw in Section 2.3, one of the clearest predictions of the task-based framework is that automation will reduce the employment share of the most exposed workers at both the firm and industry level. In practice, as we discussed in the Introduction, other empirical studies have reached mixed conclusions on this question. Our results contribute to this mixed picture. When it comes to industry-level CNC events, it is difficult to discern any clear pattern: we find a slight but not significant increase in the manufacturing share immediately after an automation event, which fades out by year four. For industrial robots, we find a gradual but (by year four) statistically significant reduction in the share of manufacturing workers across all three event threshold specifications, reversing an apparently positive pre-treatment trend.

Figure 2.10: Industry-level Event Studies, Total Employment

(a) CNC Machine Adoption



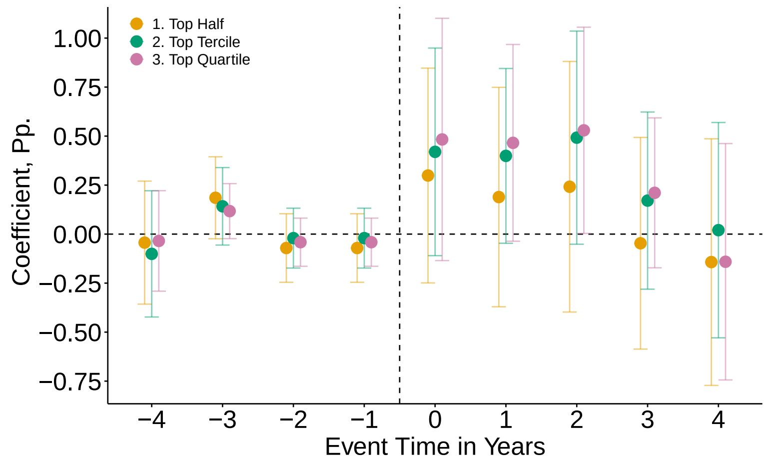
(b) Industrial Robot Adoption



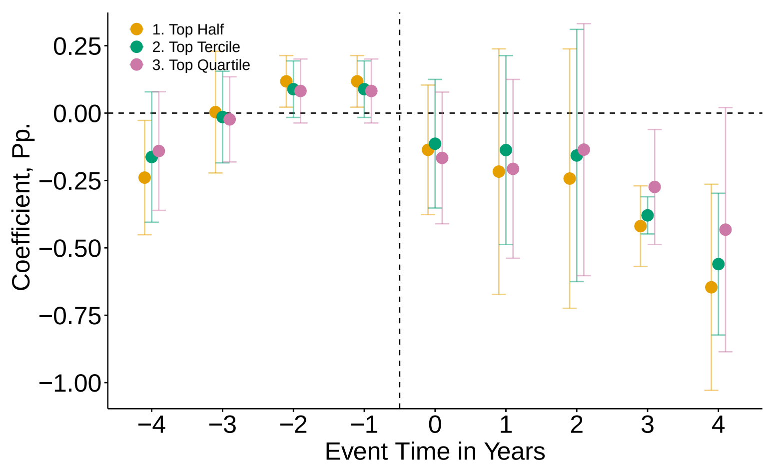
Notes: These figures show LP-DiD estimates of the effect of industry-level automation events on total industry employment, as described in the text. Industry is defined at the six digit SIC code. Events are defined as an increase in the industry share of employees working in firms with a given technology above a given percentile threshold in the distribution of all positive year-on-year changes. We use three thresholds: the top half, top third and top quarter of positive year on year changes. These correspond to an increase of 5%, 11%, 17%, respectively for CNC penetration, and 18%, 52%, and 90% for robots. All designs include year fixed effects and a lag of the differenced outcome variable, as per equation 2.3, with standard errors clustered at the industry level. Vertical bars represent 95% confidence intervals.

Figure 2.11: Industry-level Event Studies, Share of Manufacturing Employment

(a) CNC Machine Adoption



(b) Industrial Robot Adoption



Notes: These figures show LP-DiD estimates of the effect of industry-level automation events on the industry share of manufacturing employment. Industry is defined at the six digit SIC code. Events are defined as an increase in the industry share of employees working in firms with a given technology above a given percentile threshold in the distribution of all positive year-on-year changes. We use three thresholds: the top half, top third and top quarter of positive year on year changes. These correspond to an increase of 5%, 11%, 17%, respectively for CNC penetration, and 18%, 52%, and 90% for robots. All designs include year fixed effects and a lag of the differenced outcome variable, as per equation 2.3, with standard errors clustered at the industry level. Vertical bars represent 95% confidence intervals.

2.8 Conclusion

This paper provides new evidence on one of the most pressing questions in labour economics: how does automation affect employment? Using a novel dataset tracking the adoption of CNC machine tools and industrial robots across UK manufacturing plants from 2005-2023, we find that automation events are associated with positive employment effects, both at automating plants and the wider industry level.

Our key finding is that automation technologies increase employment at adopting firms by approximately 6% to 8% in the four years following adoption. This result is robust across different specifications and estimation strategies, and holds for both CNC machines and industrial robots. In other words, productivity effects clearly dominate displacement effects at the firm level, consistent with theories emphasizing automation's role in expanding output and creating complementary tasks.

While initial technology adoption shows little impact on the share of manufacturing workers, CNC expansion events are associated with a significant reduction in the manufacturing employment share of around 8%. This pattern is consistent with learning-based theories of technology adoption, suggesting that firms initially use automation in ways that complement existing workers but gradually reorganize production processes in more transformative ways as they accumulate experience with the technology.

At the industry level, we find evidence of positive spillover effects rather than business stealing, with automation events at focal firms associated with employment growth among competitors. Industry-wide automation events generally show positive or neutral effects on total employment.

Our findings are broadly consistent with other studies using firm-level data on technology, and suggest that fears of widespread job displacement from automation may be overstated. Looking forward, several important questions remain. While our findings at the firm and industry levels are predominantly positive, a complete assessment of automation's labour market impact requires examining general equilibrium effects at the labour market and national levels. Workers displaced from non-adopting firms or industries may face adjustment costs and transitions that are not captured in our plant and industry-level analysis. The mechanisms driving positive spillover effects between firms also deserve further investigation.

Finally, while our results provide valuable insights into the employment effects of CNC machines and industrial robots—two foundational automation technologies—caution is warranted in extrapolating these findings to newer technologies, such as artificial intelligence. The task-based framework suggests that different technologies may have fundamentally different employment consequences depending on their specific capabilities, the nature of the tasks they automate, and the wider market context. As automation technologies continue to evolve, understanding how the employment effects vary across different types of technological change will be crucial for anticipating future labour market dynamics.

Appendix

Table B1: Aggregate Employment in MAG Survey vs ONS

	(1)	(2)	(3)
Year	MAG Panel Data	ONS Data	MAG/ONS (%)
2005	2,626	3,120	84.17
2006	2,527	3,025	83.54
2007	2,478	2,985	83.02
2008	2,381	2,884	82.56
2009	2,306	2,657	86.79
2010	2,199	2,570	85.56
2011	2,167	2,564	84.52
2012	2,186	2,603	83.98
2013	2,225	2,560	86.91
2014	2,271	2,621	86.65
2015	2,408	2,641	91.18
2016	2,329	2,655	87.72
2017	2,237	2,680	83.47
2018	2,068	2,715	76.17
2019	1,903	2,689	70.77
2020	2,128	2,618	81.28
2021	1,955	2,544	76.85
2022	1,797	2,612	68.80
2023	1,793	2,576	69.60
Mean	2,165	2,695	80.33

Notes: This table compares total employment (in ’000s) in the MAG survey with official data for total UK manufacturing employment in the the quarterly Labour Force Survey, produced by the Office of National Statistics.

Table B2: Plant Size Distribution in MAG vs ONS UK Manufacturing Plants, 2016

Size	MAG Emp. ('000s)	ONS Emp. ('000s)	MAG/ONS (%)
1–9	41	238	17.2
10–49	308	461	66.8
50–249	668	606	110.2
250–499	396	248	159.7
500+	892	851	104.9
All	2305	2404	95.9

Notes: This table compares total employment (in '000s) in the MAG survey with official data for total UK manufacturing employment from the quarterly Labour Force Survey, produced by the Office of National Statistics, broken down into different plant-size bins. For plants with more than 50 employees, the number of plants in the MAG survey exceeds the number of plants in the ONS data. This is likely to be driven by the fact that MAG's coverage is driven by commercial requirements and not strictly limited to the manufacturing sector; and it may also reflect varying definitions of employment, such as the separate treatment of full-time and part-time workers in the MAG survey.

Table B3: Count of Automation Events, by Year

	Events		
Year	CNC Adoption Events	CNC Expansion Events	Robot Adoption Events
2005	562	705	-
2006	274	479	-
2007	422	742	-
2008	277	810	-
2009	189	442	-
2010	265	534	-
2011	193	691	-
2012	172	466	-
2013	318	561	-
2014	231	477	-
2015	134	327	242
2016	265	397	204
2017	342	511	243
2018	137	214	173
2019	87	153	162
2020	125	194	106
2021	230	347	51
2022	99	144	638
2023	33	46	7
Total	4,355	8,240	1,826

Notes: This table presents plant-level counts of 'events' in the MAG survey. Reliable information about robot use is only available from 2015.

Table B4: Effect of Automation Events on Log Employment: Alternative Estimators

Dependent Variable:	Δ Ln Employment$_t$											
Treatment Definition	CNC Adoption				CNC Expansion				Robot Adoption			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables												
Treatment Effect	5.74***	5.74***	5.06***	5.63***	7.67***	7.68***	6.82***	8.63***	6.95***	6.95***	6.98***	7.05***
	(0.828)	(0.828)	(0.932)	(0.861)	(0.549)	(0.550)	(0.664)	(0.592)	(1.05)	(1.05)	(1.03)	(1.17)
ΔLn Employment$_{t-1}$	-0.037***	-0.037***	0.703***	-0.034***	-0.035***	-0.035***	0.706***	-0.035***	-0.024***	-0.026***	0.718***	-0.022***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)	(0.009)	(0.005)
Estimator	CDLZ	CS	BJS	CC	CDLZ	CS	BJS	CC	CDLZ	CS	BJS	CC
Fixed-effects												
year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC1980-year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics												
Observations	159,262	159,262	123,011	147,825	143,218	143,218	108,379	130,261	112,053	67,868	65,284	104,908
R^2	0.050	0.050	0.230	0.053	0.051	0.051	0.232	0.056	0.041	0.047	0.227	0.043

Notes: This table reports the pooled post-treatment effect of automation events on total plant-level employment using four alternative estimators (1) CDLZ is the Stacked Estimator Cengiz et al. (2019); (2) CS uses the approach from Callaway and Sant'Anna (2021) (3) BJS uses the approach from Borusyak et al. (2024); (4) CC applies a Composition Correction to our baseline estimator to ensure stable comparison groups, as outlined in Dube et al. (2023). All specifications include Year by Industry fixed effects. Standard errors are clustered at the plant level. Significance: *** 0.01, ** 0.05, * 0.1.

Table B5: Effect of Automation Events on Log Employment: Sensitivity to Clustering Method

Dependent Variable:	Δ Ln Employment$_t$											
Treatment Definition	CNC Adoption				CNC Expansion				Robot Adoption			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables												
Treatment Effect	5.74***	5.74***	5.74***	5.74***	7.67***	7.67***	7.67***	7.67***	6.95***	6.95***	6.95***	6.95***
	(0.828)	(0.970)	(0.929)	(0.828)	(0.549)	(0.636)	(0.637)	(0.549)	(1.05)	(1.11)	(1.14)	(1.05)
ΔLn Employment$_{t-1}$	-0.037***	-0.037***	-0.037***	-0.037***	-0.035***	-0.035***	-0.035***	-0.035***	-0.024***	-0.024***	-0.024***	-0.024***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)
Clustering	Plant	Industry	Ind×Year	Robust	Plant	Industry	Ind×Year	Robust	Plant	Industry	Ind×Year	Robust
Fixed-effects												
year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC1980-year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics												
Observations	159,262	159,262	159,262	159,262	143,218	143,218	143,218	143,218	112,053	112,053	112,053	112,053
R^2	0.050	0.050	0.050	0.050	0.051	0.051	0.051	0.051	0.041	0.041	0.041	0.041

Clustered (cluster_var) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: This table reports the pooled post-treatment effect of automation events on log of total plant-level employment, using our baseline specification, but varying the clustering method. Plant = plant level clustering, as in our baseline; Industry = SIC1980 clustering; Industry×Year = SIC1980×Year clustering; Robust = heteroskedasticity-robust only. All specifications use the CDLZ estimator with Industry by Year fixed effects.

Table B6: Effect of Automation Events on Log Employment: Sensitivity to Number of Outcome Lags

Dependent Variable:	Δ Ln Employment$_t$											
Treatment Definition	CNC Adoption				CNC Expansion				Robot Adoption			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables												
Treatment Effect	5.45***	5.74***	6.16***	5.22***	7.42***	7.67***	6.97***	6.34***	6.94***	6.95***	6.97***	6.65***
	(0.764)	(0.828)	(0.887)	(0.937)	(0.498)	(0.549)	(0.621)	(0.650)	(1.05)	(1.05)	(1.05)	(1.04)
ΔLn Employment$_{t-1}$		-0.037***	-0.040***	-0.038***		-0.035***	-0.037***	-0.039***		-0.024***	-0.020***	-0.021***
		(0.005)	(0.006)	(0.005)		(0.005)	(0.005)	(0.006)		(0.006)	(0.006)	(0.007)
ΔLn Employment$_{t-2}$			-0.038***	-0.041***			-0.042***	-0.046***			-0.024***	-0.017**
			(0.006)	(0.006)			(0.006)	(0.007)			(0.007)	(0.007)
ΔLn Employment$_{t-3}$				-0.043***				-0.052***				-0.026***
				(0.007)				(0.008)				(0.009)
Number of Lags	0	1	2	3	0	1	2	3	0	1	2	3
Fixed-effects												
year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC1980-year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics												
Observations	178,806	159,262	140,622	123,011	162,605	143,218	125,042	108,379	129,636	112,053	95,202	79,141
R^2	0.049	0.050	0.051	0.051	0.050	0.051	0.052	0.052	0.041	0.041	0.044	0.047

Notes: This table reports the pooled post-treatment effect of automation events on log of total plant-level employment, using our baseline LP-DiD specification, but varying the number of lags of the differenced outcome variable. All designs include Year by 6-digit Industry fixed effects. Standard errors are clustered at plant level. Significance: *** 0.01, ** 0.05, * 0.1.

Bibliography

Acemoglu, D., G. Anderson, D. Beede, C. Buffington, E. Childress, E. Dinlersoz, L. Foster, N. Goldschlag, J. Haltiwanger, Z. Kroff, P. Restrepo, and N. Zolas (2023). Advanced technology adoption: Selection or causal effects? *AEA Pap. Proc. 113*, 210–214.

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics 4*, 1043–1171.

Acemoglu, D., C. Lelarge, and P. Restrepo (2020). Competing with robots: Firm-level evidence from france. *AEA Papers and Proceedings 110*, 383–388.

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

Acemoglu, D. and P. Restrepo (2022a). A task-based approach to inequality. *Oxford Open Economics (Deaton Review) 2024*(3).

Acemoglu, D. and P. Restrepo (2022b). Tasks, automation, and the rise in U.S. wage inequality. *Econometrica 90*(5), 1973–2016.

Adachi, D., D. Kawaguchi, and Y. U. Saito (2024). Robots and employment: Evidence from japan, 1978–2017. *J. Labor Econ.*, 591–634.

Aghion, P., C. Antonin, S. Bunel, and X. Jaravel (2023a). The effects of automation on labor demand: A survey of the recent literature. In L. Y. Ing and G. M. Grossman (Eds.), *Robots and AI: A New Economic Era*. Routledge.

Aghion, P., C. Antonin, S. Bunel, and X. Jaravel (2023b). The local labor market effects of modern manufacturing capital: Evidence from france. *AEA Papers and Proceedings 113*, 219–223.

Aghion, P., C. Antonin, S. Bunel, and X. Jaravel (2023c). Modern manufacturing capital, labor demand, and product market dynamics: Evidence from france. CEP Discussion Paper 1910, Centre for Economic Performance, London School of Economics.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and innovation: An inverted-U relationship. *Q. J. Econ. 120*(2), 701–728.

Agrawal, A., J. Gans, and A. Goldfarb (2019). *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Athey, S. and G. W. Imbens (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *J. Econom. 226*(1), 62–79.

Atkeson, A. and P. J. Kehoe (2007). Modeling the transition to a new economy: Lessons from two technological revolutions. *American Economic Review 97*(1), 64–88.

Autor, D. (2015). Why are there still so many jobs? the history and future of workplace automation. *J. Econ. Perspect. 29*(3), 3–30.

Bartel, A., C. Ichniowski, and K. Shaw (2007). How does information technology affect productivity? plant-level comparisons of product innovation, process improvement, and worker skills. *The Quarterly Journal of Economics 122*(4), 1721–1758.

Bessen, J. (2019). Automation and jobs: when technology boosts employment. *Econ. Policy 34*(100), 589–626.

Bessen, J., M. Goos, A. Salomons, and W. van den Berge (2023). What happens to workers at firms that automate? *Rev. Econ. Stat.*, 1–45.

Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event-study designs: robust and efficient estimation. *Rev. Econ. Stud.*, rdae007.

Boustan, L. P., J. Choi, and D. Clingingsmith (2022). Automation after the assembly line: Computerized machine tools, employment and productivity in the united states. Technical Report w30400, National Bureau of Economic Research.

Bright, J. R. (1958). *Automation and Management*. Boston, MA: Division of Research, Graduate School of Business Administration, Harvard University.

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

Brynjolfsson, E., D. Rock, and C. Syverson (2019). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In

A. Agrawal, J. Gans, and A. Goldfarb (Eds.), *The Economics of Artificial Intelligence*, pp. 23–60. University of Chicago Press.

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-differences with multiple time periods. *J. Econom. 225*(2), 200–230.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The effect of minimum wages on low-wage jobs. *Q. J. Econ. 134*(3), 1405–1454.

Curtis, E. M., D. G. Garrett, E. C. Ohrn, K. A. Roberts, and J. Serrato (2021). Capital investment labor demand. (Working Paper 29485).

Dauth, W., S. Findeisen, J. Suedekum, and N. Woessner (2021). The adjustment of labor markets to robots. *J. Eur. Econ. Assoc. 19*(6), 3104–3153.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *Am. Econ. Rev. 110*(9), 2964–2996.

de Chaisemartin, C. and X. D'Haultfœuille (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *Econom. J.*.

de Chaisemartin, C. and X. D'Haultfœuille (2024). Difference-in-differences estimators of intertemporal treatment effects. *Rev. Econ. Stat.*, 1–45.

Dixon, J., B. Hong, and L. Wu (2021). The robot revolution: Managerial and employment consequences for firms. *Manage. Sci. 67*(9), 5586–5605.

Dube, A., D. Girardi, O. Jorda, and A. Taylor (2023). A local projections approach to difference-in-differences. Technical report, Cambridge, MA.

Gardner, J. (2022). Two-stage differences in differences. *arXiv [econ.EM]*.

Goldin, C. and L. F. Katz (2008). *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *J. Econom. 225*(2), 254–277.

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

Gregg, A. G. (2020). Factory productivity and the concession system of incorporation in late imperial russia, 1894–1908. *American Economic Review 110*(2), 401–427.

Hirvonen, J., A. Stenhammar, and J. Tuhkuri (2023). New evidence on the effect of technology on employment and skill demand. Working Paper.

Holland, M. (1989a). *When the machine stopped : a cautionary tale from industrial America*. Harvard Business School Press.

Holland, M. (1989b). *When the machine stopped: Cautionary tale from industrial America*. Boston, MA: Harvard Business School Press.

Humlum, A. (2021). Robot adoption and labor market dynamics. mimeo.

Jaikumar, R. (2005). From filing and fitting to flexible manufacturing: A study in the evolution of process control. *Foundations and Trends® in Technology, Information and Operations Management 1*(1), 1–120.

Karshenas, M. and P. L. Stoneman (1993). Rank, stock, order, and epidemic effects in the diffusion of new process technologies: An empirical model. *Rand J. Econ. 24*(4), 503–528.

Katz, L. F. and K. M. Murphy (1992). Changes in relative wages, 1963–1987: Supply and demand factors. *Quarterly Journal of Economics 107*(1), 35–78.

Klenert, D., E. Fernández-Macías, and J.-I. Antón (2023). Do robots really destroy jobs? evidence from europe. *Econ. Ind. Democr. 44*(1), 280–316.

Koch, M., I. Manuylov, and M. Smolka (2021). Robots and firms. *Economic Journal 131*(638), 2553–2584.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2023). Technology and labor displacement: Evidence from linking patents with worker-level data. Working Paper 31846, National Bureau of Economic Research.

Lavoratori, K. and D. Castellani (2021). Too close for comfort? microgeography of agglomeration economies in the united kingdom. *J. Reg. Sci. 61*(5), 1002–1028.

Lynn, F., T. Roseberry, and V. Babich (1966). A history of recent technological innovations. In *Technology and the American Economy: Report of the National Commission on Technology, Automation, and Economic Progress*. Washington, DC: U.S. Government Printing Office.

Mann, K. and L. Püttmann (2023). Benign effects of automation: New evidence from patent texts. *Rev. Econ. Stat. 105*(3), 562–579.

National Research Council (1995). *Information Technology for Manufacturing: A Research Agenda*. Washington, DC: The National Academies Press.

Restrepo, P. (2024). Automation: Theory, evidence, and outlook. *Annual Review of Economics 16*, 1–25.

Ross, D. T. (1978). Origins of the apt language for automatically programmed tools. *ACM SIGPLAN Notices 13*(8), 61–99.

Roth, J., P. H. C. Sant'Anna, A. Bilinski, and J. Poe (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *J. Econom. 235*(2), 2218–2244.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J. Econom. 225*(2), 175–199.

Tinbergen, J. (1974). *Income Distribution: Analysis and Policies*. Amsterdam: North-Holland.

Webb, M. (2020). The impact of artificial intelligence on the labor market. Working paper, SSRN ID: 3482150.

Chapter 3

Towards a Rawlsian Economics

ABSTRACT

This chapter argues that the standard economic interpretation of John Rawls—as an advocate of redistribution justified by a maximin social welfare function—misrepresents the spirit and substance of his theory. Rather than endorsing a more egalitarian form of welfarism, Rawls offers a conception of justice grounded in reciprocity and concerned with the lifetime prospects of the least advantaged—not only for income, but also for economic power and the social bases of self-respect. I outline the core features of a Rawlsian normative economics: plural in values, grounded in measurable primary goods, and psychologically realistic in its emphasis on reciprocity; and I argue that such a framework can help shift economic thinking beyond the conventional focus on redistribution toward a broader policy agenda centred on predistribution and meaningful work.

Keywords: John Rawls; welfare economics; economic justice; predistribution; workplace democracy.

3.1 Introduction

John Rawls is widely considered to be one of, if not the most influential political philosopher of the twentieth century. The publication of his landmark book *A Theory of Justice* (TJ) in 1971 is recognised as marking the (re)birth of modern political philosophy (Kymlicka, 2002), after a period in which Peter Laslett, the respected historian of political thought, had famously declared that "For the moment, anyway, political philosophy is dead."[1] Rawls's ideas transformed the discipline, prompting what has been described as "an outpouring of philosophical literature on social, political, and economic justice unmatched in the history of thought"(Freeman, 2010). In 1974, just three years after the publication of *TJ* , Robert Nozick claimed that "Political philosophers now must either work within Rawls's theory, or explain why not" (Nozick, 1974, 183). This assessment has largely stood the test of time: nearly forty years later, the philosopher G. A. Cohen wrote that "at most two books in the history of Western political philosophy have a claim to be regarded as greater than *A Theory of Justice*: Plato's *Republic* and Hobbes's *Leviathan*" (Cohen, 2009, 11).

Rawls's primary philosophical aim was to articulate a systematic alternative to the then-dominant utilitarianism. He did this by reimagining the social contract tradition associated with thinkers such as Locke, Hobbes and Rousseau, arguing that we should organise society according to the principles that rational individuals would select behind a "veil of ignorance", without knowledge of their individual characteristics. Rawls used this "original position" thought experiment to propose two principles of justice, which he referred to collectively as "justice as fairness". The first "basic liberties" principle states that all citizens are entitled to an equal set of fundamental personal and political freedoms, including freedom of belief and expression, and the right to a democratic political system. His second "equality" principle has two parts: "fair equality of opportunity", meaning every citizen should have an equal chance to develop their talents and abilities, and to compete for positions on the basis of merit; and the "difference principle", which states that inequalities can be justified only if they serve to maximise the life chances of the least well-off,

[1] Laslett was responding to the rise of logical positivism (roughly, the idea that scientific knowledge is the only form of factual knowledge), which had led philosophers to eschew substantive questions about how we should organise society in favour of linguistic analysis of political concepts (Laslett, 1956, vii). For a detailed historical account of the reception and impact of Rawls's ideas within political philosophy, see Forrester (2019)

by creating the incentives that underpin well-functioning markets, innovation and economic prosperity.

Rawls's ideas generated an unusual degree of interest and engagement from economists. Prominent figures including Kenneth Arrow, Paul Samuelson, Anthony Atkinson and Edmund Phelps wrote lengthy responses to *TJ*, and Rawls responded in leading economic journals including the *American Economic Review* and *Quarterly Journal of Economics* (Walraevens, 2023; Guizzo and Paré-Ogg, 2023). Most discussion focused on the difference principle: while some challenged it on a philosophical level, others – led by Atkinson and Phelps – sought to incorporate it into welfare economics, and specifically the burgeoning field of optimal tax theory in the form of the "Rawlsian" or "maximin" social welfare function (SWF), which defines social welfare solely in terms of the welfare of the least well-off individual.[2] Over the 1970s and 80s, this became the standard interpretation of Rawls's ideas within economics. As a result, even as political philosophers largely abandoned utilitarianism in favour of social contract theories similar to Rawls's, the welfarist foundations of normative economics remained largely unchanged.

In this paper I will argue that the maximin interpretation of Rawls seriously misconstrues his ideas and their implications for economic policy. At a philosophical level, it presents Rawls's theory as a variety of welfarism, if not utilitarianism, when Rawls's explicit aim was to provide a systematic *alternative* to utilitarianism as the basis for public policy; and it ignores Rawls's basic liberties and opportunity principles, both of which have important implications for economic institutions. It also misinterprets the difference principle in crucial ways: whereas the maximin SWF seeks to maximise the welfare of the poorest individual at a given point in time, the difference principle is concerned with the lifetime opportunities of the lowest paid workers. This focus on shaping people's opportunities, rather than controlling outcomes, reflects the fundamental importance of reciprocity within Rawls's theory. Finally, whereas the maximin SWF and other welfarist principles are typically appled to the distribution of income, the difference principle is explicitly concerned with a richer set of "primary goods" including the distribution of positions of power and control and the "social bases of self-respect", among which access to paid work

[2] I will use the term maximin (rather than Rawlsian) SWF since, as we shall see, it only superficially resembles Rawls's theory.

has a particularly important place.

The association of Rawls with the maximin SWF has contributed to the widespread belief that, in practical terms, his theory is primarily a justification for high levels of redistribution. This interpretation made his ideas relatively easy to incorporate into the dominant policy paradigm within economics, which focuses on maximising economic efficiency by correcting market failures, and then addressing concerns about inequality through redistributive taxation. But Rawls explicitly rejected "welfare state capitalism", which he characterised as relying on the "redistribution of income to those with less at the end of each period", in favour of a "property-owning democracy" in which the state would seek to ensure "the widespread ownership of productive assets and human capital (that is, education and trained skills) at the beginning of each period". In this sense, he was an early advocate for what we would now call "predistribution".[3]

Getting Rawls right is not simply a matter of historical accuracy. The central argument of this paper is that Rawls's philosophy, properly understood, offers a rich resource both for developing a non-welfarist normative economics, and for shaping a policy agenda that looks beyond the conventional focus on redistribution. A Rawlsian normative economics would prioritise individual freedom and democracy over the pursuit of economic growth, and recognise the importance of opportunity, reciprocity, and of work as a source of self-respect – addressing well-known limitations of the still-dominant utilitarian paradigm. This in turn would justify a policy agenda focused on bringing about a more equal distribution of human and physical capital, and hence of market incomes, including through prioritising investment in the education of non-college workers; and ensuring broad access to good jobs, by embracing greater workplace democracy, among other things.

This paper draws on and contributes to various related literatures. There is now a substantial literature that seeks to clarify, elaborate and build on Rawls's ideas about economic justice, challenging the prevalent notion that he was simply an advocate for an expanded welfare state (O'Neill and Williamson, 2012; Freeman, 2013; Thomas, 2016). This literature has transformed how Rawls is understood within political philosophy, but has had little impact on how his ideas are understood and applied within economics, which is the central focus

[3]This term was coined by Hacker (2011).

of this paper. Second, this paper connects to active debates about the need to incorporate a wider set of fairness related concerns within welfare economics and optimal tax theory (Atkinson, 2009; Coyle et al., 2023; Fleurbaey and Maniquet, 2018). Rawls's philosophy clearly supports such an approach. Finally, Rawls's ideas can help clarify the moral basis for growing interest among economists and policy makers in a shift from redistribution to what has been variously termed "predistribution", "productivism" and "supply side progressivism" – in effect, a new focus on questions about work and production (Diamond and Chwalisz, 2015; Rodrik, 2023a; Democracy Journal, 2023). These terms are often used vaguely and little attention has been given to the underlying ethical arguments for such a shift. Rawls's philosophy can shed new light on why these new approaches are desirable.

3.2 Domesticating Rawls: the maximin social welfare function

The publication of Rawls's *A Theory of Justice* in 1971 sparked intense debate, not just among philosophers, but among economists. Most of the discussion focused on the difference principle, which was typically – and as we shall see, not entirely accurately – interpreted as the claim that society should seek to maximise the welfare or income of the least well-off individual.

A number of distinguished economists including Paul Samuelson, John Harsanyi and Kenneth Arrow engaged directly with Rawls's philosophical argument for the difference principle (Igersheim, 2023, 2022). Many, including Harsanyi and Arrow, broadly accepted the basic logic of the original position thought experiment – that we should organise society according to the principles that rational agents would choose if they were deprived of knowledge about their individual position in society, as if behind a "veil of ignorance".[4] But they rejected Rawls's reasoning for the difference principle, arguing that the "maximin" decision rule, whereby parties in the original position seek to maximise their prospects under the worst case (minimum) scenario, rests on a psychologically unrealistic degree of risk-aversion. They argued that rather

[4]Harsanyi had invoked a similar hypothetical choice situation in a paper published in 1955, three years before Rawls published his first article on "justice as fairness" (Harsanyi, 1955; Peter, 2009).

than focusing exclusively on the worst outcome, rational agents would seek to maximise their expected utility, and hence endorse average utilitarianism.[5] These criticisms prompted Rawls to clarify that the case for the difference principle within the original position does not rest on maximin reasoning or strong psychological assumptions about risk-aversion.[6] Instead, Rawls's emphasised the benefits of the difference principle in securing the willing support rather than the grudging acceptance of the least well off, and the benefits this would have for the "stability" of liberal democracy.

More pertinent for this paper is that, alongside these criticisms, other prominent economists sought to incorporate Rawls's ideas, and specifically the difference principle, into the burgeoning field of optimal tax theory, newly invigorated by the publication of a seminal paper by James Mirrlees's in 1971 (the same year as the publication of *TJ*). Previously, optimal tax theory had largely treated individuals' decisions about whether and how much to work

[5]Unlike classical (or sum total) utilitarianism, which aims to maximize the total amount of happiness or well-being in society, average utilitarianism seeks to maximise the average happiness per person.

[6]Rawls's initial response, published as "Some Reasons for the Maximin Criterion" in the *American Economic Review* in 1974, appeared to accept that the choice of the difference principle in the original position rests on a psychological assumption about risk-aversion. But in later work he explicitly rejected this claim, arguing that "The widespread idea that the argument for the difference principle depends on extreme aversion to uncertainty is a mistake, although a mistake unhappily encouraged by the faults of exposition in *Theory*." (JAF 43). The mature version of Rawls's argument is set out in *Justice as Fairness: A Restatement* (Rawls, 2001). Rawls accepts that, as a general rule, rational agents faced with uncertainty should maximise expected utility. However, he argues that under certain circumstances it is rational to follow the maximin decision rule instead, choosing the option where the worst-case outcome is as good as possible. Specifically, maximin is rational when (1) there is no basis for assigning probabilities to different outcomes; (2) there is an option where the worst-case scenario is decent; and (3) some options could lead to outcomes that are completely intolerable. These circumstances apply when the parties in the original position are faced with a choice between utilitarianism and Rawls's two principles taken together ('justice as fairness'), since Rawls's principles guarantee that everyone would have basic freedoms and a decent social minimum, whereas utilitarianism leaves open the possibility of being deprived of their fundamental freedoms if this could be shown to increase social welfare. Rawls argues that while this maximin argument decisively establishes the case for the basic liberties principle, it has little role to play in the choice of the difference principle. Once we accept the basic liberties as given, maximin reasoning no longer applies because even the worst outcomes are tolerable. His argument for the difference principle rests instead on a comparison between that principle and "restricted utilitarianism", which accepts the basic liberties principles and a basic minimum, and then maximises average utility above this level. Rawls continued to defend the DP on the basis that it is the "most reasonable" conception of economic justice, but he recognised that argument for the DP over restricted utility rests on "a more delicate balance of less decisive considerations" (JAF, 95), and that the DP is but one of a family of "reasonable" conceptions of distributive justice.

as fixed, irrespective of tax rates. Within a utilitarian approach, and given the assumptions of identical utility functions and diminishing marginal utility, this had led figures like Edgeworth and Pigou to advocate 100% marginal tax rates and the equalisation of post-tax incomes. Mirrlees provided a way of incorporating behavioural responses into the analysis of optimal taxation, offering a unified framework for balancing equality and efficiency considerations in tax design. Like Rawls's difference principle, the Mirleesian approach recognised that some degree of inequality may be justified to provide work incentives, thereby increasing total income and social welfare.

In Mirrlees's framework, and in most of the subsequent literature, social welfare is defined in classical utilitarian terms as the sum of individual utilities. Utility is typically assumed to depend on the final distribution of income (and the disutility of labour), which in turn determines people's opportunities for consumption. The government then chooses the tax system which achieves the social-welfare-maximising distribution of income, after taking into account the revenue requirements of the government (assumed to be fixed), and the way in which people respond to the incentives created by taxes and transfers.[7]

Mirrlees was able to derive a small number of theoretical implications about the optimal tax system, including a demogrant (lump sum transfer) at the bottom of the ability/income distribution; a positive marginal tax rate further up the income distribution; and, under certain conditions, a zero marginal tax rate at the very top. Mirrlees also conducted numerical simulations which suggested that the optimal tax schedule would be approximately linear, in contrast to contemporaneous tax schedules which were strongly progressive; and that optimal tax rates might be much lower than those then in place, especially for high earners. He estimated, for example, that the optimal marginal tax rate on roughly the top 1% of earners was between 15% and 21%, depending on the underlying distribution of skills, at a time when the actual top marginal tax rate in the UK was 75% (Walraevens, 2023, 865).[8]

[7]In the original Mirrlees (1971) model, individuals are assumed to have a pre-existing ability which determines their hourly wages, and then face a choice over how many hours to work, taking the tax and benefit system as given. Other similar papers adopted different earnings models where, for example, people choose not their hours of work, but how many years to invest in wage-enhancing education. By assumption, the government cannot directly tax ability, or indeed other fixed characteristics associated with ability (if it could, then the optimal tax would be a lump-sum tax based on individual earning ability); instead, the government must tax people's earnings directly.

[8]These results came as a surprise to Mirrlees, who in his own words "had expected the rigorous

In response to Mirrlees's work, and motivated in part by a desire to see whether it was possible to justify more progressive tax schedules, a number of other economists – led by Tony Atkinson and Edmund Phelps – started to explore the implications of adopting a "Rawlsian", or "maximin", social welfare function (SWF), which defines social welfare solely in terms of the welfare of the least well-off individual.[9] Over the 1970s, this became a standard reference point in optimal tax theory, in large part thanks to its incorporation into Atkinson and Joseph Stiglitz's hugely influential graduate textbook *Lectures on Public Economics*. This text, following earlier papers by Atkinson, presented the maximin SWF as the egalitarian limiting case of a wider class of social welfare functions with varying degrees of inequality aversion, with utilitarianism representing the other extreme.[10] Inasmuch as economists encounter Rawls today, this is the form in which they tend to do so.

Phelps and Atkinson also led the way in exploring the implications of the maximin SWF for tax policy. As expected, compared to standard utilitarianism, the maximin SWF implies higher optimal tax rates at every income level, often by a significant margin (Atkinson and Stiglitz, 1980). Atkinson, for example, estimated that the optimal marginal tax rate at the median income was 52% under the maximin SWF, compared to just 21% under the utilitarian one (Atkinson and Stiglitz, 1980, 350). Ogura (1977) generalised this result, proving that the "the Rawlsian taxation scheme is really the heaviest [higher at each level of income] of any reasonable taxation, reasonable being defined as being justifiable by some additive social welfare function" (341). In contrast to the utilitarian case with its approximately linear structure, optimal marginal tax rates under the Rawlsian tax schedule were estimated to fall from around 50 per cent at the median to around 25 per cent for the top percentile (Atkinson and Stiglitz, 1980, 350). The late 1990s and early 2000s saw the development of a new sufficient statistics approach to optimal taxation, pioneered by Saez

analysis of income-taxation in the utilitarian manner to provide an argument for high tax rates" (Mirrlees, 1971, 207)

[9]Atkinson appears to have developed the idea of a maximin SWF independently of Rawls, though he subsequently referenced Rawls in defending the plausibility of this position (Walraevens, 2023, 862-4). Phelps, by contrast, had read early drafts of *TJ* and in his first paper on this topic, published in 1973, described Rawls's theory as "the first complete principle of social choice to command wide and serious interest since the time of sum-of-satisfactions utilitarianism" (Phelps, 1973, 332).

[10]The utilitarian social welfare function simply sums utility across individuals, and in this sense gives no direct importance to the distribution of utility across individuals.

(2001) among others, which made it easier to estimate optimal tax rates by using key parameters relating to the elasticity of labour supply. More recent studies using this approach have confirmed the general finding that optimal tax rate are higher under a Rawlsian SWF compared to a utilitarian one, albeit with a slightly different profile across the income distribution compared to earlier studies (Ghatak and Jaravel, 2020).

3.3 Beyond maximin

Although Rawls's philosophy was meant to represent a fundamental break with utilitarianism and welfarism more broadly, economists largely incorporated his ideas into that tradition. According to the maximin interpretation of his ideas, the guiding principle of public policy is to maximise the welfare the least well-off individual, where in practice welfare is often simply assumed to be a function of disposable income. In this section I will examine what this interpretation gets wrong about Rawls, and attempt to recover a richer account of his ideas. In doing so, we will start to see how they depart from the welfarist framework, with important practical implications.

3.3.1 Rawls didn't modify welfarism, he rejected it

At a philosophical level, the fundamental problem is that the maximin SWF depicts Rawls's philosophy as a form of welfarism, if not utilitarianism. Welfarism is the claim that we should rank social states solely in terms of the subjective well-being (or welfare, or utility) of individuals in those states (Sen, 1979). Utilitarianism is the most widely used variety of welfarism, and in its classical form compares social states in terms of the sum of individual utilities. Since the maximin SWF compares social states solely in terms of the welfare of the least well-off individual, it too represents a kind of welfarism.

But Rawls was strongly critical of utilitarianism, and of welfarism more broadly: in the preface to *A Theory of Justice* he described his principal aim as seeking to develop "a reasonably systematic *alternative* to utilitarianism" (TJ, xi; italics added).[11] Rawls didn't simply reject the principle of maximising util-

[11] In what follows, I refer to Rawls's three major works by their initials: TJ for *A Theory of Justice* (revised edition, 1999), PL for *Political Liberalism* (expanded edition, 2005), and JAF for *Justice as Fairness: A Restatement* (2001). Full publication details are provided in the bibliography.

ity as the basis for public policy, he rejected the fundamental logic of welfarism as an ethical framework. He recognised that it was rational for individuals to organise their life so as to maximise the satisfaction of their desires, making sacrifices and trade-offs such as consuming less now to have more income in retirement. But he argued that we cannot simply transfer this logic to the organisation of society, sacrificing the interests of some people to increase the sum of satisfaction across society as a whole. To do so, he argued, is to treat society as if it were a single agent and hence fail to "take seriously the distinction between persons" (TJ, 24). For Rawls, the aim of a democratic society is not to maximise welfare, utility, or indeed any other conception of the individual or collective "good", but to specify fair terms of cooperation among free and equal citizens, each with distinct interests and desires. The original position thought experiment is designed to reflect this conception of society as "a fair system of social cooperation", by asking us to imagine principles that every citizen could rationally accept.[12]

Of course, the structure of the original position itself does not rule out the possibility that the parties might endorse some form of utilitarian principle for regulating the basic structure of society, and much of *TJ* is devoted to making the case for why the parties would choose Rawls's principles over some form of utilitarianism. Rawls uses this thought experiment to develops a more systematic critique of utilitarianism, setting out a variety of reasons why rational parties would prefer his two principles of justice including, most importantly, that they would provide a more secure basis for protecting fundamental individual freedoms and democratic institutions.[13]

[12]Rawls described "the idea of society as a fair system of social cooperation over time" as "the most fundamental idea" in his conception of justice (JAF, 5).

[13]While utilitarian thinkers such as John Stuart Mill have defended such freedoms on the basis that they are likely to maximise utility, in the final analysis, they are always subject to the utilitarian calculus, and must be sacrificed if doing so would increase social welfare. So, for example, if an authoritarian dictatorship was able to achieve a higher level of well-being as a democratic one, a welfarist would have no reason to object. This, Rawls argued, meant that utilitarianism, and by extension welfarism, was not an appropriate conception of justice for a democratic society.

3.3.2 Rawls's freedom, opportunity and sustainability principles also have far-reaching economic implications

A more detailed discussion of Rawls's critique of utilitarianism, and indeed of welfarism and consequentialism more broadly, is beyond the scope of this essay. In the remainder of this section, I will focus instead on how the maximin SWF misrepresents the substantive of content of his conception of justice, as summarised by his two principles:

Rawls's Two Principles of Justice (PL, 5-6)

> First principle [basic liberties]: Each person has an equal claim to a fully adequate scheme of equal basic rights and liberties, which scheme is compatible with the same scheme for all; and in this scheme the equal political liberties, and only those liberties, are to be guaranteed their fair value.
>
> Second principle: Social and economic inequalities are to satisfy two conditions: first, they are to be attached to positions and offices open to all under conditions of fair equality of opportunity [fair equality of opportunity]; and second, they are to be to the greatest benefit of the least advantaged members of society [the difference principle], consistent with the just savings principle.

The most obvious problem is with the maximin SWF is that, like all welfarist principles, it places no direct importance on individual freedom, democracy or equality of opportunity. As a result, it effectively ignores Rawls's basic liberties and fair equality of opportunity principles, both of which take have a strict "lexical" priority over the difference principle. This priority means we cannot sacrifice individual freedoms, or deny some people equality of opportunity, even if doing so would increase economic growth and raise the living standards of the least well off. Of course, the maximin SWF was never *meant* to offer an interpretation of these other aspects of Rawls's theory.[14] But these other principles clearly have important implications for economic institutions and

[14] It's possible to think about the maximin SWF as being constrained by a prior commitment to equal basic liberties and fair equality of opportunity, much like the difference principle, and it seems likely that this is what most economists have in mind – simply taking for granted that policies to promote the welfare of the least well off will take place against a background of basic rights, freedoms and opportunities.

the degree of inequality that we should tolerate as a society, and in this sense the maximin SWF is, at the very least, seriously incomplete as an account a Rawlsian vision for economic justice.

Take the basic liberties principle, which insists that every citizen is entitled to an equal set of fundamental individual rights and freedoms, including both personal freedoms of speech and religion and a broad set of political freedoms that are the basis for a democratic government. Regarding the latter, Rawls insists that the state should seek to guarantee not just formal equal political liberties like the right to vote and freedom of political speech, but the "fair value" of those freedoms, in the sense that citizens' opportunities for political influence should not depend on access to financial resources.[15] As we shall see, he argued that this is likely to require direct limits on extreme concentrations of economic power and wealth. Rawls also specifies elsewhere that his first principle should be thought as being preceded by "a lexically prior principle requiring that basic needs be met", on the basis that doing so "is a necessary condition for citizens to understand and to be able fruitfully to exercise the basic rights and liberties" (JAF, 44). This overlooked 'basic needs' principle provides a rationale for a degree of redistribution, as well as certain essential public services, even before we start to think about how we might apply the more well-known difference principle.

Fair equality of opportunity – the idea that every citizen should have an equal chance to develop their innate talents and abilities, and to compete for jobs and position on the basis of merit – also has far reaching implications for social and economic policy. Indeed, despite its familiarity, this is a radical principle that would likely require measures to alleviate if not eliminate child poverty, alongside an education system that guarantees equal educational opportunities irrespective of parental income. Note also Rawls's requirement that the difference principle must be applied in a way that is "consistent with the just savings principle" (JSP) – a principle of intergenerational justice that is generally understood to include a commitment to what has come to be known as sustainable development (development that meets the needs of the present without compromising the ability of future generations to meet their needs). This would justify policies to ensure that a society's economy operates within

[15] Rawls defined the notion of fair value as follows: "Citizens similarly gifted and motivated have roughly an equal chance of influencing the government's policy and of attaining positions of authority irrespective of their economic and social class" (JAF, 46).

certain safe ecological limits.[16]

3.3.3 The difference principle is concerned with access to primary goods, not welfare

By reducing Rawls conception of justice to the difference principle, then, the maximin SWF ignores his other principles, which themselves have important implications for economic policy. But it also misrepresents the difference principle itself. For a start, that principle is concerned not with the welfare or utility of the least well off, but with their access to "primary goods" – all-purpose resources that people have reason to value, whatever their particular values or conception of the good. This focus on the distribution of primary goods – including, though as we shall see not limited to, income and wealth – as opposed to preferences or subjective well-being is another fundamental departure from the dominant welfarist framework.

Rawls offers several reasons for this shift of focus. One is the classic "expensive tastes" argument – in other words, someone whose happiness depends on fine dining and constant travel should not be entitled to more resources than someone with more modest desires. As Rawls puts it, people must "take responsibility for their ends" (PL 33-34). A more practical reason is that primary goods are directly observable, and relatively easy to compare across individuals, whereas utility and preferences are not.[17] In Rawls's terminology they better satisfy the "publicity" condition, by which principles of justice must provide a reasonably transparent and practical basis for public deliberation. The fact that primary goods are measurable, at least in principle, also offers a way out of the well-known conceptual challenges associated with inter-personal comparisons of utility, and makes Rawls's theory an attractive practical framework for public policy, something we will return to in Section 5.

[16] Although Rawls said very little about the environment, he recognized that the question of intergenerational justice included both "the question of the appropriate rate of capital saving and of the conservation of natural resources and the environment of nature" (TJ, 118-119). A number of other thinkers have used and extended his ideas to develop a more detailed Rawlsian account of environmental justice and sustainability (Manning, 1981; McKinnon, 2012; Töns, 2021).

[17] Rawls argues that the difficulties in measuring utility "are so great and the approximations are so rough that deeply conflicting opinions may seem equally plausible to different persons." (TJ, 282). In this way, welfarism introduces a degree of indeterminacy about the desirability of different policies, which in turn "is likely to increase disputes and mistrust" (JAF, 127), and may be abused by powerful groups (TJ, 282).

3.3.4 The difference principle is about maximising the productive opportunities of low-wage workers, not redistribution

A second way in which maximin misconstrues the difference principle is that it is focused on maximising the welfare of the least well off at a given point in time, whereas the difference principle is concerned with maximising their "life-prospects", defined as "differences in citizens' (reasonable) expectations of primary goods over a complete life" (JAF, 59). Rawls is also clear that the "least well off" refers not to the poorest individuals in society (typically careers, the sick and the unemployed) but to the least advantaged workers, understood as the group whose natural abilities are least favoured in a market economy at a given point in time.[18] Rawls isn't entirely clear how we should define this group, suggesting in different places the class of minimum-wage workers or even simply the bottom half of earners. In a modern knowledge economy that prizes cognitive skills and academic education, non-college educated workers might be a natural (if broad) proxy for this group.

The aim of the difference principle is to organise our economic institutions so that this group has better opportunities to earn a decent living than under any alternative economic system, rather than simply to maximise their welfare or income at any given moment. In this sense, Rawls's theory is fundamentally about opportunities rather than outcomes. This helps explain why Rawls presented FEO and the DP as two parts of a single principle of economic justice, rather than as stand-alone principles: while FEO is meant to ensure that everyone has fair access to jobs and positions, the DP is concerned with making sure that the rewards attached to those positions are fair.

From this perspective, the problem with the maximin SWF and other welfarist principles is that they seek to bring about a certain distribution of resources in society without paying any attention to the process through which these resources have been generated. Rawls referred to this way of thinking as "allocative justice", which is concerned with "how a given bundle of commodities is to be distributed, or allocated, among various individuals whose particular needs, desires, and preferences are known to us, and who have not cooperated in any way to produce those commodities." Rawls contrasts alloca-

[18]This is not to say that justice for carers, the sick and disabled doesn't matter, but simply that we need a different principle to think about this.

tive justice with "distributive justice", which is grounded in the idea of society as a "fair system of social cooperation", and hence seeks to strike a fair balance between contribution and reward (JAF, 50). For Rawls the aim is not to achieve a pre-specified distribution of resources, but to organise our economic institutions such that "when everyone follows the publicly recognized rules of cooperation, and honours the claims the rules specify, the particular distributions of goods that result are acceptable as just (or at least as not unjust) whatever these distributions turn out to be" (JAF, 50).

This emphasise on opportunities rather than outcomes is important from a moral perspective, because it leaves plenty of room for individual choice and responsibility: as Rawls put it, "What a person is entitled to depends on what he does" – how hard they work, whether they decide to study, and so on. It also has practical advantages. Critics of the difference principle have sometimes suggested that it would require constant interference to divert funds to the least well off, conjuring dystopian images of excessive government control.[19] But the difference principle is concerned with the structural inequalities that affect the life prospects of different social groups – those that result from the way we organize our social and economic institutions; rather than inequalities that inevitably arise as people make choices and go about their lives (Nagel, 2003). It requires no more interference than any conventional system of government funded by taxation.

3.3.5 The difference principle governs the distribution not just of income and wealth, but of economic power and opportunities for self-respect

Although the maximin SWF is concerned with maximising the welfare of the least well off, in practice welfare is typically equated with income, and the focus is on designing a tax system that will maximise the financial resources available to redistribute to the poor. By contrast, the difference principle is concerned not just with income and wealth, but with the distribution of two other crucial primary goods: the "powers and prerogatives of offices and positions of authority

[19] This is the essence of Robert Nozick's so-called Wilt Chamberlain argument in Nozick (1974, 160-164). For Rawls's response, see PL, 283.

and responsibility" (PPO), and the "social bases of self-respect" (SBSR).[20]

PPO refers to the way in which authority and decision-making rights are vested in different positions, whether in public bodies or private companies. Rawls recognised that, as with inequalities of income and wealth, a degree of workplace hierarchy can be justified, since this is essential for the proper functioning of large organisations, which allows society to benefit from economies of scale. But such inequalities must be justified to, and have compensating benefits for, those who end up at the bottom of the decision-making pile, through higher wages, better funded public services, and so on.[21]

Although often overlooked, Rawls stated in multiple places that the social bases of self-respect were "the most important primary good", since "Without [self-respect] nothing may seem worth doing." (TJ, 386). Whereas self-respect is a subjective state of mind, and depends in part on a person's character and relationships in ways that are properly outside the scope of public policy, the *social bases* of self-respect refer to objective and measurable features of our economic and political institutions that shape citizens opportunities for self-respect, and hence are the proper subject of a theory of justice, and indeed of public policy.

Rawls said very little about what the social bases of self-respect might be, and this continues to be a source of debate. But as Moriarty (2009) and others have argued, opportunities for meaningful work stand out as among the most important social bases of self-respect. To see why, it's worth unpacking Rawls's notion of self-respect a little further.

Rawls defined self-respect as having two aspects: "First... it includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect implies a confidence in one's ability, so far as it is within one's power, to fulfil one's intentions" (TJ, 386). The first aspect – having confidence that our goals in life are worthwhile – depends, in turn, on two things, which I will refer to as opportunities for "meaningful activities" (including meaningful work) and "social recognition". The former is meant to capture the essence of what Rawls

[20] This aspect of Rawls's theory has often been overlooked (and not just by economists) in part because it is relatively under-developed on Rawls's own writing. But amongst philosophers there is now a broad consensus that the difference principle should be interpreted as maximising lifetime access to an index of these three primary goods (Van Parijs, 2003)

[21] For a detailed discussion and development of Rawls's ideas about PPO, see Arnold (2012).

referred to as the Aristotelian Principle, which he defined as the claim that "other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity" (TJ, 374).[22] Rawls gives the example of chess versus checkers, arguing that someone who can do both will prefer chess since it is more complex, and hence ultimately more rewarding. This is meant to describe a widespread psychological tendency (rather than a moral principle), which Rawls argued could help us understand many of the things which humans do and value, and might have an evolutionary basis, since enjoying complex activities is likely to generate benefits for a particular community (TJ, 375-8). Rawls argues that we must give significant weight to this principle in the design of social institutions "otherwise human beings will find their culture and form of life dull and empty. Their vitality and zest will fail as their life becomes a tiresome routine" (TJ, 377).

There are clearly many activities that could satisfy this principle, but paid work arguably has a special importance (Moriarty, 2009). The importance of paid work within a Rawlsian framework is note grounded not in the "perfectionist" claim that such activities are essential to living a meaningful life, an idea sometimes associated with Marx and the socialist tradition (Hsieh, 2008, 76). Rather it reflects the fact that paid work is one of the most important ways in which most people fulfil their obligation to contribute to society in return for sharing in its benefits. Since citizens are effectively required to devote much of their time to work, and since work shapes their lives and personalities in such profound ways, society also has an obligation to ensure that paid work is consistent with developing a secure sense of self-respect.[23]

Having confidence that one's goals in life are worthwhile also depends, according to Rawls, on "finding our person and deeds appreciated and confirmed by others who are likewise esteemed and their association enjoyed"; in other words, on social recognition (TJ, 386). Without the recognition of our peers, Rawls argued, it is hard to maintain a strong sense that our goals are worth-

[22] As Rawls explained, "The intuitive idea here is that human beings take more pleasure in doing something as they become more proficient at it, and of two activities they do equally well, they prefer the one calling on a larger repertoire of more intricate and subtle discriminations."

[23] As Moriarty (2009, 452) puts it "people's participation in work... is often *mandatory* and *extensive*. Moreover, work can be monotonous, routine, and "deadening to human thought and sensibility" (TJ, 546). By contrast, people's participation in nonwork associations is usually *optional* and *limited*."

while, and "we cannot pursue them with pleasure or take delight in their execution". Of course, we can't expect all our fellow citizens to endorse our values and goals — in a diverse society, where people have deeply different moral and religious beliefs, this is all but impossible. What matters is that "there should be for each person at least one community of shared interests to which he belongs and where he finds his endeavours confirmed by his associates." The basic liberties play an important role here, since they are a precondition for a rich and diverse associational life, in which a variety of religious and ethical communities can flourish. Participation in paid work is also vital, since it is one of the most important ways that people contribute to society and hence earn the respect of their fellow citizens.

Most people need opportunities for meaningful work and social recognition in order to feel that their goals in life are worthwhile, or so Rawls argues. In addition, self-respect depends on having "confidence in one's ability... to fulfill one's intentions" – in other words, a sense of agency or independence. This also points us towards the importance of paid work, since being unemployed and dependent on the state for income can undermine one's sense of agency, especially given the intrusive systems of assessment and conditionality that exist in most means-tested benefit systems. By contrast, having decent opportunities for paid work can be a vital source of the agency and independence that are so essential for self- respect.

3.4 Rawls on economic institutions

As should be clear by now, rather than representing a particularly egalitarian form of welfarism, Rawls's philosophy offers a vastly richer account of economic justice that breaks with the welfarist tradition in a variety of important ways. But what are the practical and policy implications?

Rawls's most systematic comments about economic institutions are contained in *Justice as Fairness: A Restatement*, a short book based on his lecture notes that was published in 2001, a year before his death.[24] In contrast to the dominant approach within welfare economics, which tends to focus on optimising individual policy instruments (especially taxation), Rawls encourages us to think about the justice of economic institutions in the round. He organises

[24] There is also a detailed discussion in TJ, 228-242

his analysis around a comparison of five stylised "regimes": laissez-faire capitalism, welfare state capitalism, property-owning democracy, liberal socialism and state socialism with a command economy (JAF, 134-141).[25] Rawls immediately rules out state socialism – a one-party regime with a centrally planned economy – on the basis that by failing to allow a significant role for democracy, markets and freedom of occupational choice, it violates the basic liberties principle. He also swiftly rejects laissez faire capitalism – a regime that "aims for economic efficiency and growth constrained only by a rather low social minimum" – on the basis that it fails to make any serious effort to achieve either fair equality of opportunity or the difference principle (JAF, 137).

Perhaps more surprisingly, given his reputation as an advocate for an expanded welfare state, he also rejected welfare state capitalism (WSC). Rawls characterises WSC as a private property-based market economy that seeks to ensure that "none should fall below a decent minimum standard of life" (JAF, 139), an aim which it achieves through "the redistribution of income to those with less at the end of each period, so to speak". The principal problem with WSC, Rawls argues, is that it "permits very large inequalities in the ownership of real property (productive assets and natural resources) so that the control of the economy and much of political life rests in few hands" (JAF, 138). This concentration of power and wealth means WSC is unable to guarantee the "fair value of the political liberties", as required by the basic liberties principle. Moreover, unlike the difference principle, which seeks to ensure a fair balance of contribution and reward, WSC simply aims to top up people's incomes so they can meet their essential needs, and hence fails uphold a principle of reciprocity in economic relations (JAF, 138). In addition, by failing to provide people with opportunities for meaningful work, Rawls worried that WSC would give rise to "a discouraged and depressed underclass many of whose members are chronically dependent on welfare" (JAF, 140).

Having rejected WSC, Rawls argues that his principles are, at least in theory, compatible with either liberal (democratic) socialism (LS) or property-owning democracy (POD). Like WSC, both LS and POD would be constitutional liberal democracies in which markets play a central role in economic life, thereby respecting important basic liberties such as freedom of occupational choice and

[25] This discussion and some of the terminology, including POD, was influenced by the economist James Meade, and especially his book *Efficiency, Equality and the Ownership of Property*. For a history of the varied use of the term POD, see Jackson (2012).

the right to own personal property.[26] The crucial distinction between these regimes relates to the ownership of the means of production. Under WSC, "a small class... have a near monopoly of the means of production". This combination of markets with the concentration of private ownership is, for Rawls, the defining feature of capitalism, which in turn gives rise to a society in which there are two distinct classes: capitalists, who derive their income largely from profits, and workers, who get their income from labour. Both LS and POD seek to democratise ownership of the means of production, albeit in different ways, and in this sense although both preserve a central role for markets, Rawls describes them as "alternative[s] to capitalism" (JAF, 135). Under liberal socialism, ownership would be socialised: companies would be owned and controlled by the state and/or their employees, but would operate within a system of reasonably free and competitive markets. A property-owning democracy, by contrast, would maintain predominantly private ownership of the means of production, but ownership would be widely, if not quite equally, shared, and the sharp distinction between capitalists and workers would largely disappear.

3.4.1 The case for a Property-Owning Democracy

Rawls leaves open the choice between liberal socialism and property-owning democracy, arguing that this will depend on their feasibility and efficiency, and their fit with the historical circumstances and political traditions of a given society. Even so, he largely throws his weight behind property-owning democracy, and it is the contrast between POD and WSC that best reveals the distinctive content of his theory.

Before turning to the reasons why Rawls favours POD over WSC, it's worth saying a little more about the key features of a POD. As we have seen, this is a broadly market-based, private property economy, in which ownership of productive resources is widely, if not equally, shared. Rawls's conception of productive resources here is broad, encompassing both ownership of physical

[26] Rawls distinguishes between the right to own *personal* property, such as housing and clothing, which is protected by the basic liberties principle, and the wider question of the ownership of the means of production, which is not (JAF, 138). Private ownership of personal property is a basic liberty because this is essential for leading a free and independent life. Whether companies should be owned privately or by the state, or some combination of the two is, for Rawls, a question to be resolved by reference to the difference principle. In other words, we should adopt whichever regime is most able to maximise the lifetime expectations of the least well off.

productive assets, such as companies and natural resources, but also, crucially, human capital. Moreover, Rawls highlights that in contrast to WSC, which is focused on "the redistribution of income to those with less *at the end of each period*", in a POD the aim is to ensure "the widespread ownership of productive assets and human capital (that is, education and trained skills) *at the beginning of each period*" (italics added). In other words, Rawls makes the case for what has come to be known as predistribution, meaning an economic strategy that seeks to achieve a more equal distribution of market incomes rather than rather than relying on redistribution.[27] This isn't just about a more equal distribution of money, but about transforming the world of work. As Rawls put it, in a POD "the narrowing and demeaning features of the division [of labour] should be largely overcome" (JAF, 177).

To be clear, although Rawls rejects WSC, this doesn't mean a POD would eliminate income transfers altogether: some form of redistribution would almost certainly be required to support those who cannot work, and hence to meet the requirements of Rawls's basic needs principle. Neither does it mean that society would do away with other familiar elements of the welfare state, including the public provision of healthcare, education and other public services, all of which are essential for achieving fair equality of opportunity. It simply means that when it comes to raising the living standards of the lowest paid workers, as required by the difference principle, society should focus on expanding their opportunities for decent employment rather than income transfers.

Rawls makes two key arguments in favour of POD, which broadly correspond to his criticisms of WSC. The first is that, by tackling the extreme concentration of economic power and wealth, POD would be better able to protect the fair value of the political liberties. This argument implicitly rests on the (empirical) claim that we cannot fully insulate the democratic process from wider inequalities through policies like limits on donations to political parties, at least not beyond a certain level of extreme wealth and economic power. More fundamentally, Rawls argues that, unlike WSC, POD would regulate economic

[27]Rawls also throws his weight behind the idea of "Society as employer of last resort through general or local government, or other social and economic policies", arguing that "Lacking a sense of long-term security and the opportunity for meaningful work and occupation is not only destructive of citizens' self-respect but of their sense that they are members of society and not simply caught up in it. This leads to self-hatred, bitterness, and resentment." (PL, lvii).

inequalities according to a principle of reciprocity: rather than simply topping up the incomes of the poor, they would ensure that the lowest paid workers have the opportunity to contribute to economic life in return for a fair reward.

There is also a third Rawlsian argument that point to the limits of WSC, and supports the case for a POD (I say "Rawlsian" because although a number of prominent Rawls scholars have made this argument, Rawls did not explicitly do so himself).[28] At the heart of this argument is the fact that the difference principle is concerned not only with the distribution of financial resources, but with the distribution of "powers and prerogatives of offices" (PPO) and the "social bases of self-respect" (SBSR). From this perspective, the problem with WSC is that, by relying exclusively on redistribution, it largely leaves the sphere of work and production untouched: it does nothing (at least not directly) to shape power relations within the workplace, or to expand access to meaningful work, which as we have seen is among the most important social bases of self-respect. A POD, by contrast, would bring about a much more equal distribution of economic power and control, and in doing so could bring about a wider transformation in the organisation and content of work(O'Neill, 2012, 88-89).

There continues to be significant debate about the details of Rawls's argument for POD, even among those who broadly endorse Rawls's philosophical framework and principles. Edmundson (2017) argues in his book *John Rawls, Reticent Socialist*, that Rawls should have rejected POD in favour of LS, on the basis that state ownership of the "commanding heights" of the economy is the only way to guarantee the fair value of the political liberties.[29] Thomas (2016) argues in *Republic of Equals: Predistribution and Property-Owning Democracy* that Rawls should have rejected liberal socialism in favour of POD, on the basis that (among other things) mandating that all companies be either state- or worker-owned would place excessive limits on freedom of occupational choice. Others have argued that Rawls was wrong to reject welfare state capitalism, and that it is both possible and preferable to achieve his principles through the famil-

[28] As O'Neill (2012) explains, this third argument is likely to provide the strongest grounds for POD over WSC within the Rawlsian framework, since it is at least conceivable that society could achieve the fair value of the political liberties through rigorous regulation of campaign finance and high levels of taxation; and we can imagine a welfare state regime that seeks to recognise a principle of reciprocity through, for example, a combination of work-related conditionality and in-work tax credits.

[29] This is, of course, a strong empirical claim. But if true it would imply a clear preference for liberal socialism, since Rawls's first principle takes priority over the second.

iar institutions of the welfare state, such as cash transfers and public services, without a large-scale shift in property ownership (Schemmel, 2015).

A detailed engagement with this literature is beyond the scope of this paper. In any case, the framing of this debate as a choice between a discrete set of economic regimes may obscure more than it reveals. While comparisons of stylised economic regimes can be a useful heuristic for drawing out certain high-level choices about, say, the role of markets and private ownership, it is unlikely to be a useful guide for real world choices about economic institutions and policies. As O'Neill (2020) has argued, this approach falsely narrows the range of options that are available to us. Economic systems differ along multiple dimensions, including the extent to which we rely on markets and how they are regulated, the balance of private and public ownership, the level of taxation and the role of the state in providing public services. Instead of thinking about the form of our economy as a choice between rigid alternative *systems*, we need to think about how we can combine these elements in different ways.

3.5 Towards a Rawlsian normative economics: plural, measurable, psychologically realistic

Up to this point, my aim has been to recover a more complete and faithful account of Rawls's approach to economic questions, highlighting aspects that are overlooked by the maximin interpretation common among economists. But what can economists gain from this exercise, beyond a more accurate understanding of an important historical figure? In this section I will highlight three features of Rawls's theory that make it an attractive basis for a non-welfarist normative economics: that it would be pluralist, measurable, and grounded in a realistic moral psychology. In the next section I will argue that these ideas also point towards a new direction for economic policy.

Ever since its emergence in the nineteenth century, normative economics has been dominated by the welfarist and utilitarian tradition – hence the more familiar term welfare economics. One of the defining features of this tradition is its commitment to a unitary value: individual welfare or utility, aggregated across society. This singular focus has an important methodological advantage: it lends itself naturally to the optimisation frameworks that economists favour, allowing social choices to be modelled in formally tractable ways. Although many economists have acknowledged the ethical limitations of welfarism, including its inability to account for values such as rights, freedoms, or fairness (Sen, 1987); it is widely assumed that there is no alternative that could provide an equally systematic basis for public policy.

Before the publication of *TJ* in 1971, this assumption was shared by many philosophers too. Indeed, as we have seen, Rawls described the absence of a systematic alternative to utilitarianism as the primary motivation for developing his theory. He argued that philosophers had been left to choose between utilitarianism and what he called "intuitionism" – an approach which recognises a plurality of values, but denies that there is any coherent basis for prioritising them, and that the best we can do is appeal to our intuitions about particular cases (TJ, xvii-xviii, 30-40). Simple as it might seem, one of Rawls's most important philosophical innovations was to organise his principles into a clear structure of priority – basic liberties take precedence over equality of opportunity, which in turn takes precedence over the difference principle; and, through the original position, to provide a compelling rationale for doing so. In his review of *TJ,* the philosopher Bernard Williams singled this out as an "outstanding feature" of Rawls's approach, arguing that it made his theory "at once complete and humane enough to satisfy moral demands, and rigorously unified enough to meet the rational requirements of one who wants more than disconnected insights" (Williams, 2015, 83).

Of course, Rawls's theory still has its ambiguities. Critics have challenged the lexical priority that Rawls attaches to his principles, arguing that fully achieving fair equality of opportunity, for example, would absorb so much social resources that nothing would be left to tackle inequality; and Rawls himself argued that this lexical ordering should be thought of as "an illuminating approximation" rather than a rigid requirement (TJ, 40). Moreover, Rawls doesn't provide any systematic way to assess possible trade-offs between the index of social primary goods governed by the difference principle (income and wealth, positions of power and control, the social basis of self-respect); though he suggested that we should resolve this indexing problem by adopting the perspective of the least well off. Even so, although his principles can only provide a partial ordering of political and economic institutions, they still serve as an action-guiding basis for public deliberation, clearly favouring some institutions and policies over others.

A second attractive feature of a Rawlsian normative economics is that, by focusing on the distribution of primary goods rather than welfare or utility, it would be grounded in objectively measurable features of the world. This is essential if normative economics is to fulfil its primary function as framework for positive economics and public policy. The subjective nature of utility is widely recognised as a key weakness of the utilitarian tradition: indeed, one of the central tenets of modern welfare economics is the "impossibility" of interpersonal comparisons of utility. This in turn has led to the development of purely ordinal criteria based on (objective and measurable) revealed preferences, such as Pareto improvements, according to which we should adopt policies that make someone better off and no-one worse off. This is an attractive but largely irrelevant criteria, since Pareto improvements are so rare. Economists have developed more widely applicable alternatives like the Kaldor-Hicks criterion, whereby a policy is desirable if those who gain could hypothetically compensate those who are made worse off. But this is very hard to justify as the basis for public policy unless compensation is actually paid, at which point the Kaldor-Hicks criterion is simply a Pareto improvement by another name.

A normative economics grounded in Rawls's principles would avoid these difficulties. Rawls is clear that primary goods must be measurable, at least in principle: this is one reason why, for example, he singles out the *social bases* of self-respect – objective features of our social institutions that shape people's opportunities for self-respect, like access to work – rather than self-respect in itself, which is a subjective state of mind. This is not to say the Rawlsian approach would be free from conceptual or practical measurement challenges: it would require us, for example, to specify the social basis of self-respect in more detail; and to provide a tractable definition of the least well off. But it provides a framework that can be brought to data and inform public debate.

A third advantage of a Rawlsian normative economics is that it would be grounded in a more realistic moral psychology than that assumed by classical welfare economics. This mattered to Rawls, who argued that principles of justice must be reasonably likely to secure the willing support not just of hypothetical parties to the original position, but real-life citizens. This is, for Rawls, essential for the "stability" of a conception of justice. Traditional utilitarianism rests on the psychology of altruism: it assumes that individuals are willing to treat the welfare of all others with the same concern as their own (TJ, 164). This in turn gives rise to a peculiar tension with most positive economics, which assumes that individuals are motivated by self-interest – a tension that is typically resolved by assuming the existence of a social planner who can impose altruistically motivated utilitarian policies and institutions on an otherwise self-interested citizenry.

Rawls rejected both pure altruism and pure self-interest as empirically and normatively misguided. His own theory is explicitly built instead on the psychology of reciprocity, broadly conceived: the idea that individuals are willing to cooperate with others on fair terms when others are likewise willing to do so.[30] This avoids the unrealistic moral demands of utilitarianism, which ask citizens to sacrifice their deepest commitments whenever doing so would increase the aggregate welfare. And it provides a more plausible account of how social institutions can sustain themselves over time, by creating conditions in which cooperation is stable because it is mutually beneficial and perceived as fair.

Rawls's emphasis on the importance of reciprocity as the cornerstone of a realistic moral psychology has been largely supported by subsequent empirical research (Bowles and Gintis, 2013). Lab experiments consistently find that people frequently make personal sacrifices to uphold norms of reciprocity (usually by punishing those who they think have acted unfairly), while anthropologists have found evidence of reciprocity in the form of sharing food and other resources beyond the immediate family in all societies going back to the advent of Homo sapiens. This is not to say that empirical work in moral psychology uniquely supports Rawls's specific principles, which clearly offer a very particular interpretation of what reciprocity might entail when it comes to the design of social institutions. But there is good reason to think that a psychologically realistic normative economics would take some notion of reciprocity rather than altruism or self-interest as its starting point.

[30]Rawls thought it was vital that any theory of justice be compatible with a realistic moral psychology, and the subject is at the heart of the (overlooked) third part of *A Theory of Justice*. Here, Rawls sets out a detailed account of how citizens might come to develop a sense of justice, drawing on the work of developmental psychologist Lawrence Kohlberg (Riker, 2014).

3.6 Towards a Rawlsian economy: predistribution and good jobs

While Rawls's ideas offer a promising foundation for a non-welfarist normative economics, they also have starkly different policy implications compared to the conventional welfarist approach. This is true across a range of policy areas, but I will focus here on how his principles point towards the importance of predistribution and broad access to good jobs, since this represents a significant departure from the dominant policy paradigm in economics.[31]

The welfarist framework typically assumes that welfare is derived from consumption and leisure, which in turn is a function of disposable income; and that work is simply a source of disutility (since it reduces leisure), something people do to increase the income available for consumption. From this perspective, what matters is the final distribution of income and hence of opportunities for consumption, rather than the balance of predistribution or redistribution. Of course, the welfarist framework could, in principle, incorporate non-monetary sources of welfare that are associated with work, and there is a growing body of economic research seeking to shed light on the complex relationship between work and well-being (De Neve and Ward, 2025). But for the most part, welfare economics has had little to say about the sphere of work, other than that it should be organised efficiently to maximise the production of welfare-enhancing consumer goods and services. This broad neglect of the sphere of production in theory has largely been mirrored in economic policy, which has rested on the idea that the role of government is to maximise productive efficiency by tackling market failures and externalities, and then address concerns about inequality through redistributive taxation (though as we shall see in the next section, this is starting to change).

Simply recognising that predistribution and good jobs are important policy goals represents a significant departure from the familiar welfarist approach. Strictly speaking, this is as far as Rawls's philosophy on its own can take us – highlighting important policy goals that have been neglected. But what does this require in practice? What kinds of institutions and policies are most likely to achieve these goals? Rawls's discussion of property-owning democracy only

[31]For a discussion of the practical implications of Rawls's principles across a wide range of policy areas, see Chandler (2023).

provides the barest outline of an answer to these questions. This was intentional: Rawls recognised that these are ultimately empirical questions, and the proper domain not of political philosophy but of the social sciences, especially economics. It would be foolish to try and provide exhaustive or definitive answers in a short article. Instead, in the rest of this section I simply want to highlight two areas where a proper understanding of Rawls's principles would justify a different policy direction than the welfarist framework, even a supposedly Rawlsian one.[32]

3.6.1 A Rawlsian economics would prioritise education for non-college educated workers

Across the world, recent decades have seen a massive expansion of academic higher education, both driven and accompanied by large increases in public subsidies. But whereas huge amounts of public money have been invested in subsidising university-level education, in most countries public spending on post-secondary vocational education and training has lagged behind. This isn't universally true: some countries, like Germany, have a long tradition of excellent vocational education, and have achieved a more balanced profile of public investment. But in countries such as the UK and USA, vocational education is fragmented and of variable quality, and in many countries, students pursuing vocational routes are not eligible for the generous student loans and tuition fee waivers available for academic courses. To take just one example, a 2012 study found that average public funding per full-time student for vocational courses in the UK was around £2,150 per year, compared to almost £8,400 for university undergraduates (Wolf, 2015, 15).

The strong emphasis on academic higher education over vocational education is frequently justified on the basis that in a modern knowledge economy, the kinds of skills gained through academic higher education are the key drivers of innovation and growth. This claim is usually supported by highlighting the substantial wage premium earned by graduates compared to those

[32] One challenge in comparing Rawls's framework with the conventional welfarist approach is that almost any policy can, in principle, be justified within the welfarist framework given the right empirical assumptions about utility or efficiency. My aim in the sections that follow is not to argue that the policies I describe can *only* be justified on Rawlsian grounds, but to show that, under *similar* empirical assumptions, Rawls's principles lead to *different* conclusions than standard welfarist reasoning.

who enrol in vocational courses. There is some debate about whether investments in academic HE yield higher returns than those in vocational education, reflecting the difficulties of disentangling the causal effect of education from the selection effects that arise as people with different levels of underlying ability opt into different routes (Matthewes and Ventura, 2022). But let's suppose for the sake of argument that marginal public investments in academic higher education typically yield a larger overall return in terms of GDP. In this case, it's easy to see how a policy that prioritises academic over vocational education could be justified within a conventional welfarist framework, on the basis that society should maximise economic growth and then use the tax system to make sure the proceeds are widely shared.[33]

What would a Rawlsian perspective bring to this question? Both fair equality of opportunity and the difference principle are relevant here. FEO requires that access to education, whether academic or vocational, should be based on merit rather than individual or parental income. At the post-secondary level, where there is no expectation of universal participation, this could be achieved through a system of income-contingent loans, as is the case in the UK, Australia and a growing number of other countries; or through full public funding, as is common in Europe. But FEO alone cannot tell us how to prioritise academic versus vocational routes. After all, this principle is consistent both with a society in which public subsidies are reserved for a small academic elite, or one in which all subsidies are directed towards vocational courses, as long as access is based on merit rather than money.

To answer this question, we need the difference principle. From this perspective, our aim is not simply to maximise economic growth, but to design an economic system that will maximally benefit the class of workers whose natural capacities are likely to earn the lowest reward in a modern economy, in other words non-college educated workers. Moreover, we should seek to achieve this

[33]There has been surprisingly little discussion amongst economists about the normative basis for education policy. Indeed, across the seven volumes of the definitive Elsevier *Handbook of the Economics of Education* that have been published since 2006, there is not a single Chapter devoted to this question. The most straightforward rationales for public intervention in education within the welfarist framework are (1) to address individual credit constraints that prevent people from enrolling in education that would yield a private net benefit; and (2) to account for a wide range of positive social, political and economic externalities associated with higher levels of education. Although economists frequently refer to concepts like equality of opportunity and social mobility in justifying education policies, these have no direct basis in the standard welfarist framework.

by investing in the productive capacity and human capital of non-college educated workers directly, rather through income transfers. We could call this the "inclusive opportunity" education strategy, where the aim is to target educational subsidies to enhance the employment opportunities of lower earners. In contrast to the "growth first" strategy, which focuses on the marginal impact on economic output, the Rawlsian criteria for targeting education subsidies would be the marginal impact on the earnings of non-college workers (or however we decide to define the least well off).

This doesn't mean we should abandon public subsidies for academic higher education entirely. Such subsidies have indirect benefits for non-college educated workers, through higher tax revenues and by encouraging innovation and job creation that raises the wages and employment prospects for those without college degrees. If we removed all public subsidies from academic higher education, we would likely reduce the income and job opportunities of non-college educated workers. The optimal Rawlsian policy will need to strike a balance between these competing forces, and identifying the right balance is a complex empirical question that deserves further study.[34] But it's clear that the Rawlsian approach would justify much greater investment in vocational education and training compared to either the status quo or the standard welfarist framework, even if the estimated economic returns are lower than for academic courses. This could justify a range of specific policies, such as a lifetime free entitlement to citizens who have not achieved a basic level of education; and/or a system of individual lifetime leaning accounts that provide larger absolute subsidies for vocational rather than academic courses.[35]

3.6.2 A Rawlsian economics would embrace workplace democracy

One of the characteristic features of contemporary capitalism is the concentration of formal decision-making rights in the hands of company owners, or shareholders, typically known as the shareholder primacy model. Although

[34] Answering this question means looking not only at the additional tax revenues that could be generated from higher earnings for college educated workers; but at the spillover effects on lower income workers – a topic that is harder to study, and less well understood.

[35] Although the discussion here has focused on post-secondary education, investing in early years and school education may be a more effective way to bring about a fair distribution of human capital.

shareholders usually delegate their decision-making powers to managers, since managers ultimately report to the board of directors, and directors are appointed by shareholders, in the final analysis it is shareholders who have the power to decide how companies are run. The model is typically justified on grounds of economic efficiency (Collison et al., 2011). The efficiency case for shareholder primacy is largely based on theory of property rights and incomplete contracts developed by Hart, Grossman and Moore (Hart, 1988; Zingales, 2016). In a world where contracts are inevitably incomplete, this theory suggests that residual decision making rights should be held by shareholders since they have a clear incentive to maximise the long-run profitability of a company, which in the standard welfare economic framework will lead to a Pareto efficient outcome. Moreover, without these powers, investors may be unwilling to make often risky and irreversible financial investments in companies, since they might fear that workers would capture any residual profits, depriving them of a return on their investment. Indeed, in its canonical form this theory suggests that giving power to workers is likely to undermine investment, productivity, and ultimately social welfare.[36]

In fact, there is plenty of debate amongst economists about whether shareholder primacy is as efficient as its advocates claim. In reality, shareholders often seek short-term financial gain over long-run profitability; and in the presence of externalities, there is often a gap between profitability and social welfare. There is now a large body of theoretical and empirical work demonstrating that putting power in the hands of workers may also be good for economic efficiency. Giving workers more power can encourage them to make productivity enhancing firm-specific investments, and to share information with managers and owners, by proving assurance that workers will share in the benefits (Addison, 2009; Kruse et al., 2010; Jäger et al., 2022).

What does the Rawlsian framework add to this debate? While it is possible to make an argument for workplace democracy purely on efficiency grounds – that it would boost motivation, reduce turnover, or encourage firm-specific investment – the Rawlsian case rests on the intrinsic importance of power, recog-

[36]In an influential paper, the economists Michael Jensen and William Meckling argued that giving workers more power would lead to a process whereby they would hereby they would "begin 'eating [the firm] up"', and that this would make it hard for firms to raise new investment, causing "a significant reduction in the country's capital stock, increased unemployment reduced labor income, and an overall reduction in output and welfare" (Jensen and Meckling, 1979, 504).

nition, and the social bases of self-respect. To see this, let's suppose, for the sake of argument, that shareholder primacy tends to maximise economic output, and that we can disregard any externalities. In this case, the welfarist framework would provide strong support for the shareholder primacy model on efficiency grounds. But the Rawlsian framework would reject it. For Rawls, the potential benefits of shareholder primacy for economic efficiency have to be weighed against the loss of power for ordinary workers, and the likely negative consequences for the availability of work that offers opportunities for individual fulfilment and social recognition. From a Rawlsian perspective if a more democratic workplace is likely to expand opportunities for meaningful work then, other things equal, this would be a strong argument in its favour.

How much workplace democracy is likely to be justified according to the Rawlsian perspective? As ever, the answer depends on the evidence: how effective is workplace democracy for improving the quality of work? What are the trade-offs with investment, innovation and productivity? And how should we weight these different goals? These are not easy questions to answer. But the recent evidence suggests that, at least for countries like the US and UK, a significant move towards greater workplace democracy in the form of German-style co-determination could easily be justified. As Jäger et al. (2022) have argued "the available micro evidence [on European models of co-determination] points to zero or small positive effects of codetermination on worker and firm outcomes and leaves room for moderate positive effects on productivity, wages, and job stability." This is true even in Germany, where co-management is at its most extensive. This not to say that countries like the UK and USA could introduce German-style co-management and immediately expect to see the same effects, since the success of co-management depends on a wider set of labour market institutions such as sector-level wage bargaining, as well as a relatively cooperative culture of industrial relations.[37] But if it is possible to increase worker power without any harm to firm performance, productivity and GDP, then within the Rawlsian framework it seems clear that we should be moving in this direction.

[37] As Jäger et al. (2022, 857) argue, one of the reasons why we don't see major effects of co-management is that in practice "existing codetermination laws convey little authority to workers."

3.7 Why Rawls?

Although Rawls is often interpreted as a defender of generous redistribution, his philosophy points towards an economic regime focused on predistribution and broad access to good jobs. In fact, a growing number of prominent economists have made the case for a similar shift in direction in recent years. Daron Acemoglu, for example, has argued that "we need to abandon the fantasy of building a better society just by redistributing income" and that "Policymakers' first priority should be creating "good", high-wage jobs" (Acemoglu, 2019). According to Branko Milanovic "we should aim for an egalitarian capitalism based on approximately equal endowments of both capital and skills across the population" (Milanovic, 2019, 46). In a similar vein, Dani Rodrik has coined the term "productivism" to describe an emerging economic paradigm that "focuses less on redistribution… and more on creating economic opportunity by working on the supply side of the economy to create good, productive jobs for everyone" (Rodrik, 2023a).

If there is already such broad interest in predistribution and good jobs, why do we need Rawls? Despite the growing interest in these ideas, there has been very little discussion about their underlying justification, at least amongst economists. For the most part, economists tend to appeal to some notion of economic efficiency, allowing them to operate within the broadly welfarist framework. One of the most common arguments for a greater focus on predistribution, for example, is that it provides a way to reduce inequality without increasing taxes, avoiding the efficiency costs associated with taxation (Milanovic, 2019, 44-46).[38] In his essay on productivism, (Rodrik, 2023a) develops a more nuanced efficiency based argument for policies to promote good jobs, on the basis that doing so will generate a variety of positive social, political and economic externalities, and that a shortage of good jobs generates a range of negative ones including "exclusion, broken families, drug abuse… polarization, the rise of populism". Alongside efficiency, others have claimed that further increases in taxation are simply not politically feasible.

These arguments provide valid reasons to support a shift towards predistribution within both a welfarist and a Rawlsian framework. But they suffer from

[38] For a discussion of the variety of ways in which predistribution has been defined and defended in recent years, and an attempt to clear up a great deal of conceptual confusion around this term, see (O'Neill, 2020)

two important limitations. First, they rely on contentious empirical claims. While some pre-distributive policies like increasing minimum wages would not require any direct increase in taxes, many of the policies needed to increase the wages of lower earners, like investment in vocational education, are likely to require significant public spending; and any serious strategy to bring about more equal distribution of wealth would likely require a major increase in taxes on existing stocks of wealth. More importantly, they fail to identify the most important moral *reasons* why we should care about predistribution and the availability of good jobs, namely: the importance of work for reciprocity, in the sense of maintaining a healthy connection between contribution and reward, and of work as a crucial source of individual meaning, social recognition, and ultimately self-respect. The great strength of Rawls's philosophy is that, while it recognises the validity of considerations of efficiency and feasibility, it provides a more direct basis for these broader moral concerns about dignity and reciprocity which appear to motivate so much of the wider public and political interest in questions about the importance of work.[39]

Even if we accept the need to clarify the moral case for broad policy goals such as predistribution and good jobs, why look to Rawls? After all, Rawls's own comments about these questions are relatively sparse – so much so that he has often been interpreted as an advocate for a more conventional redistributive approach. It is largely thanks to the work of scholars such as Samuel Freeman and Martin O'Neill over the past twenty years that this aspect of his ideas is now more widely understood. This work has not only clarified Rawls's own writings, but added to them in important ways. Other recent philosophers have taken up similar topics, often in considerably more detail. In her widely cited book *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*, Elizabeth Anderson has sought to revive a long tradition of market-based liberal thinking that is critical of the excessive power of owners and managers over ordinary workers, arguing that the concentration of power in the hands of shareholders in contemporary American resembles

[39]To take two examples, Keir Starmer, the UK Prime Minister, has criticised the tax-and-spend paradigm, asking whether "even when we enjoyed the most sustained period of redistribution in British history, did it also redistribute dignity and respect to working people, or was that glue… starting to come a little unstuck?" (Starmer, 2023). In a similar vein, President Biden put "good jobs" at the centre of his economic agenda, claiming that "a job is about [a] lot more than a pay cheque. It's about your dignity. It's about respect." (The White House, 2024)

a dictatorship.[40] Michael Sandel has made a powerful case for what he calls "contributive justice", defined as "an opportunity to win the social recognition and esteem that goes with producing what others need and value" (Sandel, 2020, 206). Danielle Allen has develop an explicitly non-Rawlsian argument for greater democracy at work, arguing that this follows from a conception of justice that takes political equality as its foundational value (Allen, 2023).

Each of these thinkers – and others besides – offer valuable perspectives on why work matters, and this is not the place for a detailed comparison. The case for revisiting Rawls is twofold. First, his ideas have had more influence on academic and (to a lesser extent) public discourse than any other twentieth century philosopher, and on these questions they continue to be widely misunderstood, especially amongst economists. Second, whereas other recent thinkers have addressed questions about work in more detail, Rawls's theory provides us with a framework that can situate these concerns within a reasonably complete conception of justice that also encompasses questions about individuals rights and freedoms, equality of opportunity, and economic justice more broadly. In many (though not all) cases, more detailed discussions can be viewed as friendly elaborations or revisions to a fundamentally Rawlsian perspective.

3.8 Conclusion

In this paper, I have argued that the conventional economic interpretation of John Rawls's theory — as a justification for redistribution based on the maximin social welfare function — misrepresents both the spirit and substance of his philosophy. Far from endorsing a narrow focus on maximising the income of the least well off, Rawls provides a richer and more demanding account of economic justice. His difference principle is not about outcomes at a moment in time, but about the lifetime opportunities of the least advantaged workers, understood in terms of access not only to income and wealth, but also to positions of power and control and to the social bases of self-respect.

Taken together, these ideas point towards very different approach to achieving economic justice than that supported by conventional welfare economics:

[40] As Anderson (2017, 70-71) puts it: "The vast majority [of workers] are subject to private, authoritarian government, not through their own choice, but through laws that have handed nearly all authority to their employers."

one that prioritises predistribution over redistribution, and places the structure of work and production at the heart of public policy. Rawls's framework supports widely shared ownership of capital, stronger labour market institutions, and new forms of economic organisation that disperse power and promote meaningful work. The goal is not simply to ensure that the least well off have enough to live on, but to create an economy in which all citizens have a fair chance to develop their talents, to make a meaningful contribution, and to enjoy the respect of others.

Even so, the task of developing a truly Rawlsian economics remains at an early stage. Although there is now a rich philosophical literature exploring Rawls's ideas about economic justice, there are important normative questions that require further attention: how broadly should we define the "least advantaged"? How should we handle potential trade-offs between income and wealth and the social bases of self-respect? What exactly is the time horizon over which the difference principle is meant to apply?

The discussion in this essay also raises new questions for economists: what kinds of policies would be most effective in bringing about a more equal distribution of market earnings? What are the potential trade-offs, if any, between relying on redistribution and predistribution; and what is the optimal balance between them? What is the connection between workplace democracy and the quality of work? If workers value more democratic workplaces, why are they so rare? A serious engagement with Rawls's ideas would put these often-neglected questions at the heart of economic research.

Bibliography

Acemoglu, D. (2019). It's good jobs, stupid. Technical report, Economics for Inclusive Prosperity.

Addison, J. (2009). *The Economics of Codetermination: Lessons from the German Experience*. Springer.

Allen, D. (2023). *Justice by means of democracy*. Chicago, IL: University of Chicago Press.

Anderson, E. (2017). *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*. Princeton University Press.

Arnold, S. (2012). The difference principle at work. *The Journal of Political Philosophy 20*(1), 94–118.

Atkinson, A. B. (2009). Economics as a moral science. *Economica 76*(s1), 791–804.

Atkinson, A. B. and J. E. Stiglitz (1980). *Lectures on Public Economics*. McGraw-Hill Book Company.

Bowles, S. and H. Gintis (2013). *A Cooperative Species: Human Reciprocity and Its Evolution* (Reprint Edition ed.). Princeton University Press.

Chandler, D. (2023). *Free and Equal: What Would a Fair Society Look Like?* (1 ed.). Allen Lane.

Cohen, G. A. (2009, 30 June). *Rescuing Justice and Equality*. Harvard University Press.

Collison, D., S. Cross, J. Ferguson, D. Power, and L. Stevenson (2011). Shareholder primacy in UK corporate law: An exploration of the rationale and evidence. Technical Report Research report 125, Association of Chartered Certified Accountants.

Coyle, D., M. Fabian, E. Beinhocker, T. Besley, and M. Stevens (2023). Is it time to reboot welfare economics? overview. *Fisc. Stud. 44*(2), 109–121.

De Neve, J.-E. and G. Ward (2025). *Why workplace wellbeing matters: The science behind employee happiness and organizational performance*. Boston, MA: Harvard Business Review Press.

Democracy Journal (2023). Symposium: The new supply-side progressivism. *Democracy: A Journal of Ideas*, Issue 68 (June 2023).

Diamond, P. and C. Chwalisz (2015). *The Predistribution Agenda: Tackling Inequality and Supporting Sustainable Growth*. Bloomsbury Publishing.

Edmundson, W. A. (2017). *John Rawls: Reticent Socialist*. Cambridge University Press.

Fleurbaey, M. and F. Maniquet (2018). Optimal income taxation theory and principles of fairness. *Journal of Economic Literature 56*(3), 1029–1079.

Forrester, K. (2019). *In the Shadow of Justice: Postwar Liberalism and the Remaking of Political Philosophy*. Princeton University Press.

Freeman, S. (2010). A new theory of justice. *The New York Review of Books*.

Freeman, S. (2013). Property-owning democracy and the difference principle. *Analyse & Kritik 35*(1).

Ghatak, M. and X. Jaravel (2020). Is funding a large universal basic income feasible? a quantitative analysis of UBI with endogenous labour supply. *LSE Public Policy Review 1*(2), 11.

Guizzo, D. and C. Paré-Ogg (2023). Economics with(out) ethics? an interdisciplinary encounter between public economists and john rawls in the 1970s. *The European Journal of the History of Economic Thought*, 1–28.

Hacker, J. (2011). The institutional foundations of middle-class democracy. *Policy Network 6*(5), 33–37.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J. Polit. Econ. 63*(4), 309–321.

Hart, O. D. (1988). Incomplete contracts and the theory of the firm. *J Law Econ Organ 4*(1), 119–139.

Hsieh, N.-H. (2008). Survey article: Justice in production. *J. Polit. Philos. 16*(1), 72–100.

Igersheim, H. (2022). Rawls and the economists: The (im)possible dialogue. *Revue économique 73*(6), 1013–1038.

Igersheim, H. (2023). Samuelson against 'rawls's gratuitism': Some lessons on the misunderstandings between rawls and the economists. *The European Journal of the History of Economic Thought 30*(5), 883–905.

Jackson, B. (2012). Property-owning democracy: A short history. In M. O'neill and T. Williamson (Eds.), *Property-Owning Democracy: Rawls and Beyond*, pp. 33–52. Wiley Blackwell.

Jäger, S., S. Noy, and B. Schoefer (2022). What does codetermination do? *ILR Review 75*(4), 857–890.

Jensen, M. C. and W. H. Meckling (1979). Rights and production functions: An application to labor-managed firms and codetermination. *The Journal of Business 52*(4), 469–506.

Kruse, D. L., R. B. Freeman, and J. R. Blasi (2010). *Shared capitalism at work: Employee ownership, profit and gain sharing, and broad-based stock options*. National Bureau of Economic Research Conference Report. Chicago, IL: University of Chicago Press.

Kymlicka, W. (2002). *Contemporary Political Philosophy: An Introduction*. Oxford University Press.

Laslett, P. (Ed.) (1956). *Philosophy, Politics and Society*. B. Blackwell.

Little, D. (2014). Rawls and economics. In *A Companion to Rawls*, pp. 504–525. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Manning, R. (1981). Environmental ethics and rawls' theory of justice. *Environ. Ethics 3*(2), 155–165.

Matthewes, S. H. and G. Ventura (2022). On track to success? returns to vocational education against different alternatives. (038).

McKinnon, C. (2012). *Climate Change and Future Justice: Precaution, Compensation and Triage*. Routledge.

Milanovic, B. (2019). *Capitalism, Alone: The Future of the System That Rules the World*. Harvard University Press.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Rev. Econ. Stud. 38*(2), 175–208.

Moriarty, J. (2009). Rawls, self-respect, and the opportunity for meaningful work. *Social Theory and Practice 35*(3), 441–459.

Nagel, T. (2003). Rawls and liberalism. In S. Freeman (Ed.), *The Cambridge Companion to Rawls*, Volume 62, pp. 85. Cambridge University Press Cambridge.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.

Ogura, S. (1977). More on rawlsian optimal income taxation: A complementary note on E. S. phelps's "taxation of wage income for economic justice"*. *Q. J. Econ. 91*(2), 337–344.

O'Neill, M. (2012). Free (and fair) markets without capitalism: Political values, principles of justice, and property-owning democracy. In M. O'Neill and T. Williamson (Eds.), *Property-Owning Democracy: Rawls and Beyond*, pp. 75–100. John Wiley & Sons.

O'Neill, M. (2020). Power, predistribution, and social justice. *Philosophy 95*(1), 63–91.

O'Neill, M. and T. Williamson (Eds.) (2012, 17 January). *Property-Owning Democracy: Rawls and Beyond*. John Wiley & Sons.

O'Neill, M. (2020). Social justice and economic systems: On rawls, democratic socialism, and alternatives to capitalism. *Philosophical Topics*. Available at http://eprints.whiterose.ac.uk/165444/.

Peffer, R. G. (2014). Basic needs. In J. Mandle and D. A. Reidy (Eds.), *The Cambridge Rawls Lexicon*, pp. 50–54. Cambridge University Press.

Peter, F. (2009). Rawlsian justice. In *The Handbook of Rational and Social Choice*, pp. 433–456. Oxford University Press.

Phelps, E. S. (1973). Taxation of wage income for economic justice*. *Q. J. Econ. 87*(3), 331–354.

Rawls, J. (1974). Some reasons for the maximin criterion. *Am. Econ. Rev. 64*(2), 141–146.

Rawls, J. (2001). *Justice as Fairness: A Restatement*. Harvard University Press.

Rawls, J. (2005). *Political liberalism*. Columbia University Press.

Rawls, J. (2009). *A Theory of Justice: Revised Edition*. Harvard University Press.

Riker, W. (2014). Kohlberg, lawrence. In J. Mandle and D. A. Reidy (Eds.), *The Cambridge Rawls Lexicon*, pp. 405–406. Cambridge University Press.

Rodrik, D. (2023a). On productivism. https://drodrik.scholar.harvard.edu/sites/scholar.harvard.edu/files/dani-rodrik/files/on_productivism.pdf.

Rodrik, D. (2023b). On productivism.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Rev. Econ. Stud. 68*, 205–229.

Sandel, M. J. (2020). *The Tyranny of Merit: What's Become of the Common Good?* Penguin UK.

Schemmel, C. (2015). How (not) to criticise the welfare state. *J. Appl. Philos. 32*(4), 393–409.

Sen, A. (1979). Utilitarianism and welfarism. *The Journal of Philosophy 76*(9), 463–489.

Sen, A. (1987). *On Ethics and Economics*. Wiley.

Starmer, K. (2023). Starmer's full economy speech: 'labour will offer a new deal for the public'. https://labourlist.org/2023/12/keir-starmer-full-economy-speech-resolution-foundation-today-policy-spending-public-services/. Accessed: 2025-6-8.

The White House (2024). Remarks by president biden on his administration's historic support for unions. https://bidenwhitehouse.archives.gov/briefing-room/speeches-remarks/2024/11/01/remarks-by-president-biden-on-his-administrations-historic-support-for-unions-philadelphia-pa/. Accessed: 2025-6-8.

Thomas, A. (2016). *Republic of Equals: Predistribution and Property-Owning Democracy*. Oxford University Press.

Töns, J. (2021). *John Rawls and Environmental Justice: Implementing a Sustainable and Socially Just Future* (1 ed.). Routledge.

Van Parijs, P. (2003). Difference principles. In S. Freeman (Ed.), *The Cambridge Companion to Rawls*, pp. 200–240. Cambridge University Press.

Walraevens, B. (2023). Rawls's maximin and optimal taxation theory. *The European Journal of the History of Economic Thought 30*(5), 860–882.

Williams, B. (2015). *Essays and Reviews: 1959-2002* (Reprint edition ed.). Princeton University Press.

Wolf, A. (2015). Heading for the precipice: Can further and higher education funding policies be sustained? Technical report, The Policy Institute at King's College London.

Zingales, L. (2016). Corporate governance. In *The New Palgrave Dictionary of Economics*, pp. 1–11. London: Palgrave Macmillan UK.