Centre for Economic Performance

Discussion Paper ISSN 2042-2695

No.1868
September 2022

The impact of public transportation and commuting on urban labour markets: evidence from the new survey of London life and labour, 1929-32

Andrew Seltzer
Jonathan Wadsworth

The London School of Economics and Political Science

Economic and Social Research Council

Abstract
This paper examines the consequences of the commuter transport revolution on working-class labour markets in London, circa 1930. Using GIS-based data constructed from the New Survey of London Life and Labour, we examine the extent of commuting and estimate the earnings returns to commuting. We show that commuting was an important feature for most working-class Londoners in the early-twentieth century. Using a variety of identifying procedures to address the endogeneity of distance commuted, we estimate a likely causal return of between 1.5 to 3.5 percent of earnings for each additional kilometre travelled. We also show that commuting was an important contributor to improvements in quality of life in the early-twentieth century.

Key words: commuting, public transport, earnings, London
JEL: N34; N74; N94; J01; L91

This paper was produced as part of the Centre's Trade Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We wish to thank Jessica Bean for discussions early in the project. We also wish to thank Simon Abernathy, Mark Edele, Rob Gillezeau, Dan Hamermesh, Tim Leunig, Nigel O'Leary, Bruce McClung, and Jens Wrona for useful comments and suggestions. We have also benefitted from comments from participants at seminars at Royal Holloway, University of Melbourne, University of Swansea, Lund University, the Economic History Society Annual Conference, Frontier Research in Economic History Conference, the Economic History Association Annual Conference, the World Congress of Cliometrics, the Australasian and Asian Society of Labour Economists, Labour Markets and Living Standards in Britain Conference, Exploring the Breadth of Labour Economics Conference, the World Economic History Congress, and the Urban History Group.
Andrew Seltzer, Royal Holloway University of London and IZA. Jonathan Wadsworth, Royal Holloway University of London, CReAM, IZA and Centre for Economic Performance, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1. Introduction

The industrialisation of major urban centres in the United Kingdom during the late-eighteenth and early-nineteenth centuries led to unprecedented clustering of industry and urban population growth. Previous work on early industrialisation has been dominated by the question of whether the significant economies of scale and agglomeration generated by industrialisation was sufficient to compensate for negative crowding externalities (Lindert and Williamson 1983; Komlos 1998). A more recent literature addresses the role of public transport in expanding urban areas and alleviating the problem of crowding (Heblich, Redding and Sturm 2020; Leunig 2006). During the early phases of the Industrial Revolution, the absence of high-speed, low-cost modes of transport meant that virtually all workers lived near their place of work. However, improved transport infrastructure, starting with the railways in the mid-nineteenth century, allowed workers to commute and thus broke the geographic link between residence and workplace.

This paper further examines the consequences of the commuter transport revolution, focussing on the labour market prospects of working-class Londoners in the early-twentieth century. From the mid-nineteenth through to the early-twentieth century, London probably had the largest and most connected transport system in the world. Virtually the entire modern rail network of Greater London had been built by the end of the nineteenth century. However, scholars emphasize that early rail commuting was mostly limited to the middle class and wealthy (Crafts and Leunig 2005; Leunig 2006; Dyos 1982; Polasky 2010). Contemporary and later accounts agree that working-class Londoners generally lived very near their workplace into the early-twentieth century (Booth 1902; Ponsonby and Ruck 1930; Polasky 2010). Widespread working-class commuting in London began in the early-twentieth century with the construction of the bus, tram, and underground networks. Unlike railroads, these transport networks crossed the central areas, making it far easier to commute within the inner urban area.

Over the first three decades of the twentieth century, public transport became increasingly accessible due to infrastructure development and technological changes which led to increases in coverage, speed, and reliability and also to reductions in the cost of travel. As we show below, by 1930 virtually all residents living within 15 kilometres of the City of London lived within a few hundred meters of public transport and faced a modest-distance commute cost only a small share of typical working-class earnings.

Near-universal access to low-cost public transport may have had important consequences for the efficiency of working-class labour markets. Urban and labour economists have emphasized that commuting can increase the efficiency of labour markets by allowing more workers to meet more firms, subject to congestion externalities and the costs of firm location, housing, and commuting (Hamilton 1982, White 1988). Competitive labour market theory suggests that identical workers would be paid identical wages regardless of commute – or that unobserved worker attributes underlie any difference in earnings. But it also allows for the possibility that employers could pay higher wages to compensate for attracting workers into a central business district or a remote location (Gibbons and Machin 2006). Search theory emphasizes that productivity often depends on the specific match between workers and firms (Mortensen and Pissarides 1994; Rogerson, et al. 2005). Monopsony theory suggests that travel costs create a wedge between net wages (wages minus commuting costs) earned at local and distant employment (Bhaskar and To 1999; Bhaskar, et al. 2002; Manning 2003a; Manning 2003b). Employers can pay less than the market wage and still retain local workers because they face a cost of switching to a more remote employer.

Each of these theories suggest that high-speed, low-cost public transport effects earnings. Lower transport costs reduce any compensating wage payment needed to attract workers to a central business district or a remote location and thus may reduce wages. Lower transport costs also allow workers to search over and commute to more potential employers and thus increases the expected match-specific productivity and wages. Lower transport costs can also increase wages by reducing the wedge between net wages at local and remote employers and thus reduce monopsony power.

We explore the consequences of London's public transport networks on working-class labour markets, around 1930. We construct a unique GIS data set using the *New Survey of London Life and Labour* (henceforth *New Survey* or NSLLL). The NSLLL was conducted between 1928 and 1932 and surveyed approximately two percent of working-class households residing in the 29 London Metropolitan Boroughs and nine adjacent Municipal and County Boroughs. The data contain a range of personal, housing, and employment-related characteristics.

Crucially for our purposes, the data contains two indicators of commuting: 1) weekly expenditures on work-related travel and 2) places of residence and work. We generate GIS coordinates for residences and workplaces, assigning a single centroid to each unique street address or place name. We also generate GIS coordinates for the entire rail, Underground, tram, and bus network of Greater London. We use this GIS data to estimate crow-flies distances between residence, workplace, public transport, and two central points – the Bank of England and Charing Cross Station, the commercial and geographic centres of Greater London, respectively. The estimated distances measure commuting distance, access to public transport, and home and workplace centrality. The GIS data are then used to examine working-class commuting patterns and the returns to commuting. This is the first study, as far as we are aware, to estimate the returns to commuting distance for the period before the second world war, when high-speed, low-cost public transport was first introduced.

Whereas Booth (1902) and Llwelyn-Smith (1930) argue that in the 1890s most working-class employees worked "on the spot", by the 1930s commuting moderate distances had become the norm. Over 70 percent of workers in the NSLLL sample had a one-way commute of at least one crow-flies kilometre. Commuting followed the expected geographic pattern for a modern metropolis: there were net flows from outer boroughs to the centre, although there were also many individuals who worked locally or reverse commuted. The wealthy central boroughs, particularly the ancient centres of Westminster and the City of London, received the largest net commuting inflows.

We use these distances as explanatory variables in Mincer-type regressions on labour force status and earnings in order to estimate returns to commuting. Naïve OLS regression estimates of the effect of earnings on commuting distance are likely to suffer from bias caused by reverse causality, as workers had a degree of choice over both where to work and where to live. We address this endogeneity in three ways: first, by restricting our sample to individuals for whom residence was plausibly exogenous, second, by introducing fixed effects that control for unobserved heterogeneity at the household level and, third, by instrumenting distance commuted with the distance between birthplace and the centre of London. Our results show that the probability of employment was higher for individuals residing closer to the centre but was not affected by proximity to public transport. We also find that a one-kilometre increase in distance commuted increased earnings by about 1.5 to 3.5 percent. These results are robust across a variety of specifications. For the majority of workers in the sample, this monetary return was greater than the monetary cost of travel. For many, this difference was substantial.

Finally, we compare our results to evidence from the late-nineteenth century. The NSLLL data are cross-sectional and we are unaware of similar microdata from other periods, which prevents us from making direct quantitative comparisons over time. However, the *New Survey* was intended as a forty-year follow-up to the more famous *Life and Labour of the People of London* (LLPL), a household survey conducted in the 1890s. Booth (1902) provides detailed summary information from the LLPL, which can be compared to our findings. Travel to work was far less of a focus of the LLPL than the NSLLL, suggesting that it was also a less important aspect of working-class lives. Homework, the most extreme absence of commuting, was common across a wide range of industries in the late-nineteenth century. Almost all working-class employees in the late-Victorian period worked within a few hundred meters from their residence (Booth 1902; Ponsonby and Ruck 1930). A simple back-of-the-envelope calculation shows that approximately 16.5 to 27.0 percent of the total increase in the real hourly earnings of the working-class earnings between 1890 and 1930 can be attributed to the effects of increased commuting distance.

2. Historical Background: The London Metropolis and its Commuting Infrastructure

London is a bicentric metropolis with a commercial centre in the City of London and a geographic and political centre in the nearby Metropolitan Borough of Westminster. Urban settlement beyond these centres dates back centuries, but accelerated after the rapid increase in population following the Industrial Revolution. Settlement outside the centres occurred in every direction. However, areas north of the River Thames and west of the city centre tended to be wealthier than areas south and east due to the prevailing winds and river flow (Heblich, Trew, and Zylberberg 2021).

During the period of our study, London was administered under the London Government Act, 1899. Under the Act, Greater London was divided into the City of London, 29 Metropolitan Boroughs in the County of London, and a ring of outer boroughs (officially County Boroughs, Municipal Boroughs, and Urban Districts).[1] Following Ponsonby and Ruck (1930), we classify the 38 Boroughs in the *New Survey* into central, middle, and exterior rings.[2]

[1] The London Government Act, 1899 was replaced by the LGA, 1963, which abolished the County of London and restructured the Metropolitan, Municipal, County Boroughs and Urban Districts into much larger London Boroughs. Throughout this paper we refer to Boroughs and Urban Districts as of the time of the *New Survey*.

[2] The central boroughs are Bermondsey, Bethnal Green, City of London, Finsbury, Holborn, St. Marylebone, St. Pancras, Shoreditch, Southwark, Stepney, and Westminster. The middle boroughs are Battersea, Chelsea,

The *New Survey* area comprised 429.9 square kilometres, about 27 percent of the total area of Greater London. It was the most densely populated part of the metropolis, with about 5,686,000 residents in 1928, approximately 72.4 percent of the total population of the Greater London area.[3] Within the *New Survey* area, population density tended to be highest near the centre.[4] It is likely that predominantly working-class areas surveyed in the *New Survey* contained much higher population densities than other areas within the same boroughs.[5]

As with most modern cities, London was characterised by economies of agglomeration and a resulting industrial concentration. Textiles, furniture-making, and box-making were heavily concentrated in the East End; the docks were located by the rivers and major canals; banking, finance, and insurance and their associated clerical employment were concentrated centrally in the City of London. As we will show, employment of the working-classes was geographically concentrated within the *New Survey* area, with the highest densities overall and for most industries individually in and around the City of London.

The concentration of employment in central areas implied that residential areas would also have been crowded, unless workers were able to live away from their employment and commute into work. However, the industrialisation of manufacturing preceded the development of high-speed transport by about half a century. In the early-nineteenth century most workers worked nearby their residence, as travel was slow (Heblich, et al. 2020). Walking

Islington, Kensington, Lambeth, and Paddington. The exterior boroughs are Camberwell, Deptford, Fulham, Greenwich, Hackney, Hammersmith, Hampstead, Lewisham, Stoke Newington, Poplar, Wandsworth, Woolwich, and the outer boroughs. The exact location of the individual boroughs can be seen in Figure 2.

[3] *UK Census* data, reprinted in *London Statistics* and Llewellyn-Smith (1930a).

[4] The population densities per square kilometre in 1931 were 2303.2, 2127.1, 1007.0, 1323.7, and 499.3 for the central ring, middle ring, exterior ring, *New Survey* area, and Greater London, respectively (*London Statistics*).

[5] We are unaware of data on population density for levels below Metropolitan Boroughs. However, one of the initial reasons for undertaking the both the LLPL and NSLLL was a perception of widespread crowding in working-class areas (Booth 1902; Llewellyn-Smith 1930a). It is clear from the summary volumes of both surveys that working-class dwellings were small, often crowded, and located close together (Booth 1902; Llewellyn-Smith 1930b). The NSLLL data show that approximately 7.7 percent of households and 12 percent individuals lived in crowded households, defined as more than two people to a room (Llewellyn-Smith 1930b; Hatton and Bailey 1998). If crowding is defined by the so-called Manchester standard of over 2.5 individuals to a bedroom (Llewellyn-Smith 1930b), approximately 34.7 percent of households and 46.4 percent of individuals lived in crowded conditions.

averaged at most five kilometres per hour, and probably somewhat less.[6] The only faster alternative was horse-drawn hackney carriages (and horse-drawn omnibuses and trams from 1829 and 1861, respectively), which averaged 6-10 kilometres per hour, but were beyond the means of but the very wealthy (London Transport Museum 2020).

The development of the faster transport infrastructure needed for longer-distance commutes occurred from the mid-nineteenth through the early-twentieth centuries. Railways were first built in London in 1836. Trams, buses, and London Underground followed later in the nineteenth century but did not have comprehensive networks which were fast, reliable, and inexpensive until the early-twentieth century. Table 1 shows some statistics on speed, coverage, and cost of the four modes of public transport in 1900-07, 1913, and 1929. The decline in cost per mile is particularly noteworthy, because, as we will show later, by 1930 public transport was well within the means of a large majority of working-class employees. The dramatic improvements in public transport shown in Table 1 led to an increase in usage, which far outpaced the population growth over the same period.

Railways were the first available mode of high-speed transport. Most of the modern rail network was complete by the end of the nineteenth century. In 1907, the average scheduled speed for commuting trains into central London terminal stations was about 32.3 kilometres an hour (*Statistics London*, 1907).[7] The availability of fast transport led to the growth of middle-class residential suburbs. Residential movement away from the centre is particularly evident for the City of London, which experienced a dramatic decline in residential population from

[6] Five kilometres per hour is a widely cited average walking speed first proposed by the Scottish mountaineer William Naismith. However, Naismith's rule assumes level ground, no encumbrances, and no stoppages; and thus, urban walking speed was probably substantially slower. We follow Leunig (2006) and assume a walking speed of 4.0 kilometres an hour throughout this paper.

[7] *London Statistics* (1907) shows the inward speed for 20 suburban train routes with terminus at a central London station between 8:00 and 9:00 a.m. The figure of 32.3 kilometres per hour is the average across these routes, weighted by the number of trains on each route. This is similar to other estimates of rail speed. Leunig (2006) uses surviving train timetables and calculates that the scheduled rail speed for "minor journeys" (inter-city routes with many stops to provide local service) were 30.4 and 32.8 kilometres per hour in 1887 and 1910, respectively.

130,117 in 1801, to 43,882 in 1891, and 15,758 in 1931 (UK Census), although employment continued to grow over this period.[8]

Figure 1, panel A, shows the railway network circa 1930 and the borders of the County of London and the *New Survey* area. The map of the network shows two important features of rail travel. First, there was extensive coverage of the exterior of the Greater London area. Virtually all built up locations within the area were connected to the centre by rail. Second, rail commuting *across* the central areas of the City of London and Westminster was much more difficult, as the terminal stations of the network were at what was then the outskirts of the central area when the rail networks were first built in the mid-nineteenth century.[9] There were few direct connections between the terminal stations. These two characteristics meant that rail travel was typically used for longer-distance commutes from the suburbs to the centre or, less frequently, reverse commuting out from the centre to industrial suburbs.

Although rail commuting transformed the lives of the wealthy and middle classes, most scholars have argued that, with the exception of a relatively small number of relatively high earners, few working-class employees commuted by rail in the nineteenth century.[10] Ponsonby and Ruck (1930) argue that even around 1930, rail was a relatively infrequent mode of commuting for the working-class, typically used only for very long commutes (over 16 kilometres) or as a substitute for the Underground in the south and east of the metropolis.

---

[8] The City of London undertook *Day Censuses* in 1866, 1881, 1891, and 1911 to demonstrate its continued commercial importance. The *Reports of the Day Censuses* show that many non-residents entered the City every day and that employment increased substantially even as the residential population declined. The estimated daytime population of the City of London was 261,061 in 1881; 301,384 in 1891; 364,061 in 1911; and 436,721 in 1921 (*Day Census,* 1881-1911 and *UK Census,* 1921).

[9] The opening of the Metropolitan Underground Line in 1863 provided limited connections between the terminal rail stations and the central areas of London. However, coverage was limited and the cost of an additional ticket would have been beyond the means of most working-class employees in the late-nineteenth century.

[10] See Ponsonby and Ruck (1930), Dyos (1953), Polasky (2010). Even into the twentieth century, rail commuting from the outer boroughs was not practical for most working-class households. These households often had multiple earners, thereby would have required multiple fares. Working-class employees also worked very long hours and thus faced time constraints. In addition, many had sufficiently irregular employment that they could not be sure of work when they would have had to board a train (Booth 1902; Maddison 1964; Huberman and Minns 2007).

Working-class access to public transport in London improved dramatically in the early-twentieth century. New infrastructure and technology, most notably the development of the Underground network and the replacement of horse-drawn buses and trams by their motorised counterparts, reduced the cost and increased the speed and reliability of travel, making it practical to commute between inner-city locations.[11] Buses and trams were cheap, with fares starting at ½d, whereas Underground fares were typically 2d or more.[12] The Underground lines, trams, and buses were owned separately, but were regulated under a common framework established under the London Traffic Act of 1924. Public transport companies sometimes competed for passengers on the same route until their consolidation under the London Passenger Transport Act and the formation of the London Transport Board in 1933.

Although its first underground train line was opened in 1863, the core of the modern Underground network was built in the early twentieth century. The Central London Railway (the Central Line), Baker Street & Waterloo Railway (the Bakerloo Line), Piccadilly & Brompton Railway (the Piccadilly Line), and Charing Cross, Euston & Hampstead Railway (the Northern Line) opened in 1900, 1906, 1906, and 1907, respectively.[13] The Underground ran at similar speeds to mainline rail. However, unlike the rail network, the Underground network was designed to cross the inner city and made it feasible to commute between most locations in the north and west regions of the NSLLL area. A limiting factor on Underground usage was its cost, with 58 percent of journeys in 1930 costing 2d or more

[11] In addition to public transport, bicycles were an important form of transport for the working-class. Aldred (2014) argues that bicycles were relatively cheaper and much more widely used for commuting in 1930 than in 1900. It is not possible to determine the exact number of workers commuting by bicycle, as the *New Survey* recorded transport expenses rather than mode of transport. Nevertheless, commuting by bicycle is specifically mentioned for 288 individuals and it is likely that many individuals with zero or missing commuting expenditures cycled to work. Private cars and motorcycles were beyond the means of almost all workers in our sample. These modes of transport are specifically mentioned for only 12 individuals.
[12] All prices in this paper are reported in "old" pounds sterling, where one pound (£) equals 20 shillings (s) and one shilling equals 12 pence (d).
[13] By 1907 the routes that would become the Central, District, Metropolitan, Central, Bakerloo, Piccadilly, and Northern Lines were all largely completed. Although the outer termini of these lines would be extended between 1907 and 1930, no new lines were opened from 1907 until the opening of the Victoria Line in 1969.

(Ponsonby and Ruck 1930, p. 187). While the average fare for each mile on the Underground was lower than the bus or tram (Table 1), the Underground was more expensive for shorter journeys. Ponsonby and Ruck (1930) argue that circa 1930 the Underground was the primary mode of working-class travel for distances between 3.2 and 19.3 kilometres.

Figure 1, panel B shows the Underground network in 1930. Unlike the rail network, the Underground network was geographically concentrated, with over 80 percent of stations located north of the River Thames and west of the eastern boundary of the City of London. This concentration occurred for both geological and economic reasons. North of the River Thames, the soil near the surface is predominantly "London clay", which is comparatively easy and inexpensive to tunnel through and is largely impermeable to water (Paul 2016). In most areas south of the Thames, London clay is covered by sand and silt, which is porous and difficult to tunnel through. Even today, virtually all the deep underground rail network is located in the areas where the London clay is near the surface. The boroughs east of the City of London were poorer than those to the west, and the Underground was generally not extended to this area, regardless of geological suitability.[14] The outer parts of the metropolis, with the exception of a few wealthier areas to the north and west, were generally not serviced by the Underground because the density of traffic would not have been sufficient to justify the high initial fixed investment.

Private horse-drawn carriages (or "hackneys") have been used in London for centuries. Horse-drawn omnibuses open to the public were first run in 1829 (London Transport Museum 2020). Horse-drawn trams date back to 1860. In the early-twentieth century, electric, diesel, and petrol engines replaced horses (London Transport Museum 2020). Around 1930, buses and trams were similar in terms of vehicle design, speed, and cost, with the only major difference being that trams ran on fixed lines whereas buses could be run on any road (Ponsonby and

[14] The entire area east of the City of London contained only one Underground line (District) and 9 stations.

Ruck 1930; London Transport Museum 2020). Both were substantially slower than rail or the Underground but were also cheaper on short routes. Ponsonby and Ruck (1930) argue that workers used buses and trams interchangeably on journeys of up to 3.2 kilometres.

The bus and tram networks in 1930 are shown in Figure 1, panels C and D. Trams were contained within inner-London, with few routes extending beyond the boundaries of the *New Survey* area. Tram density was highest in areas without Underground lines. On the other hand, buses were the most widely distributed form of public transport. There were 209 bus routes within Greater London in 1931, covering virtually all built-up areas. Virtually all residents of the *New Survey* area had access to at least one bus route, and only a few households in the outer boroughs were located more than a few hundred meters from a route.

3. Data

Our primary source of data are records from the *New Survey of London Life and Labour,* a household survey of working-class residents of the 29 Metropolitan Boroughs and nine outer boroughs conducted between 1928 and 1932. Most of the original record cards have survived intact and were encoded in the 1990s by the team of Roy Bailey, Dudley Baines, Timothy Hatton, Paul Johnson, Anna Leith, and Angela Raspin. The original cards from the Municipal Boroughs of Walthamstow and Tottenham, the two northernmost boroughs in the sample, have been lost.[15] The computerized records are freely available from the UK Data Service (Johnson, et al. 1999). The computerized records contain 26,915 households, 94,137 individuals, and 49,445 income earners, about two percent of the working-class population of London.

The NSLLL was created to follow the LLPL, which influences both the sample and the questions. The LLPL surveyed residents of the 29 Metropolitan Boroughs, excluding the City

[15] The adjacent boroughs of Leyton and Hornsey comprise slightly over 3 percent of the NSLLL sample and thus it is likely that the share of records lost was fairly small.

of London, which had few working-class residents by the 1890s. The NSLLL also included the County of London (again excluding the City of London). However, by the late-1920s there had been outward movement of working-class residences, thus the NSLLL also included nine adjacent outer boroughs.[16]

The sample is limited to working-class households, defined by the head of household not working in a white-collar occupation. Most households surveyed earned well below £250, approximately the median household income in London in 1929. The mean (median) wage earnings for households with at least one income earner was £110/14s/5d (£100) and only 7.5 percent of households in the sample earned more than £200.[17]

The NSLLL was structured by individual household. Each record card contains background information about each member of the household: age, gender, place of birth (for those aged 14 and over), relation to head of household, and different sources of non-wage income. The cards also contain the following additional information for each working member of the household: earnings in the previous week and in a full-time week, hours worked in the previous week and in a full-time week, occupation, employer name, place of work, and transport expenditures.[18] A complete list of the information on the record cards is shown in Appendix I. Summary statistics for variables used in the paper are shown in Appendix I, Table A.I.1.

---

[16] These were Acton, Barking, East Ham, Hornsey, Leyton, Tottenham, Walthamstow, West Ham, and Willesden. These boroughs contained the majority of working-class residents of the Greater London area outside the County of London (Llewellyn-Smith 1930a).

[17] These figures likely overestimate annual earnings, which we estimate using the standard approach of multiplying pay in a full week by 50 weeks. Although not nearly as prevalent as in the 1890s, there still existed intermittent jobs in 1930 which provided less than 50 full-time weeks of employment per year. A rough indication of the extent of intermittency can be obtained by comparing hours worked in the previous week and hours worked in a full-time week. Approximately one percent of those reporting positive hours in a full-time week also reported zero hours in the previous week. Hours worked in the previous week is missing for another 2.5 percent. Approximately 7.0 percent of workers reported working less than full time hours in the previous week, compared to only about 1.0 percent who reported working more than full time.

[18] The cards also contain information on the dwelling: address; borough; rent paid; number of bedrooms; and whether it contains a kitchen, pantry, scullery or larder, bath, parlor, garden, yard, and allotment. We do not use these data in this paper.

For our purposes, the most important feature of the data is that it contains information about travel to work. The only direct information is expenditure on transport. However, using transport expenditure to measure commuting is problematic for our purposes. This information is missing for about 30.8 percent of workers who worked a positive number of hours in the previous week and for 46.4 percent of individuals in the data who were assigned an earner number.[19] In addition, many respondents did not supply easily quantifiable answers to the question about transport expenditures, e.g. “bicycle” or “it varies”. Moreover, it is possible that some non-commuting-related travel costs are included in the responses even though the question was clearly intended to cover commuting costs only (see Appendix I). Finally, the monetary cost of travel does not incorporate the implicit costs of workers’ time commuting and thus would not reflect the full cost of transport even absent errors in the data. For these reasons, we only use transport expenditures for robustness tests, rather than as the main indicator of commuting in our analysis.

As an alternative to travel expenditure, we measure the crow-flies distances between individuals’ residence and workplace, residence (workplace) and the geographic and commercial centres of London at Charing Cross and the Bank of England, and residence (workplace) and nearest available public transport. We first generate GIS coordinates for each relevant point of interest using Streetmap.co.uk and National Library of Scotland (2020). For each unique location in the data we generate GIS coordinates for a single centroid, typically at the centre of each street or place name. In addition to home and workplace, we have gathered GIS data for the entire public transport network within the Greater London area. We then used the GIS coordinates and the Great Circle Distance formula to construct several variables

[19] When coding transport costs, we have handled missing observations in two ways. First, we simply leave these as missing and drop the observations from the analysis. Secondly, we recode missing values to zero if the individual’s residence and workplace were less than a kilometre apart. We generally prefer the second approach, as one kilometre is a plausible walking distance and virtually no individuals in the data residing within a kilometre of their workplace reported non-zero transport costs.

measuring crow-flies distances between home, workplace, the nearest public transport, and the nearer of the two centres of London. Appendix II outlines in detail the procedures used to obtain the GIS data and construct the distance variables and the potential sources of measurement error and bias in these variables.

Table 2 shows some summary statistics on distances.[20] The first 8 rows show the mean distances from home and workplace to the nearest available point of embarkation for each of the four modes of public transport. As would be expected based on Figure 1, the average distance from both home and work is largest for the Underground and smallest for buses. The variance is also much higher for the Underground, due to its incomplete coverage. Table 2 also shows the universality of access to at least some form of public transport. A household two standard deviations above the mean distance from the nearest bus stop, would nevertheless still be within easy walking distance of a stop (520 meters). Only 0.02 percent of income earners in the sample lived more than one kilometre from the nearest available means of public transport.

The next 15 rows of Table 2 show the distribution of crow-flies distances between home and work. The mean and median distances were 3.21 and 1.94 kilometres, respectively; less than modern commutes, but considerably more than "working on the spot", which was typical in the 1890s (Ponsonby and Ruck 1930).[21] It is also evident that on average 1) men commuted greater distances than women, 2) commuting distance increased with skill, and 3) heads of households commuted further than others. The construction of the sample implies that it is likely that the average commute across the entire London population was almost certainly greater than shown in Table 2 because the NSLLL sample includes only the central part of the metropolis.

[20] Approximately eight percent of workers who report pay were itinerant, with no fixed place of work. We did not assign a commuting distance to these workers. See Appendix II for further details.
[21] According to figures from the *2011 UK Census*, the average commute for full-time workers in Greater London was 16.4 kilometres (Greater London Authority 2015).

Table 2 also outlines the direction of commuting, divided into four mutually exclusive and collectively exhaustive categories: inwards – workplace is at least one kilometre closer to the centre than home; outwards – home is at least one kilometre closer to the centre than workplace; local – distance travelled is less than one kilometre, and across – distance travelled is at least one kilometre but there is less than one kilometre difference in home and workplace centrality. The largest share of workers, 38 percent, commuted inward, followed by working locally, 29 percent. Around one third of workers in the sample commuted outward or across London. Employment was not confined to the central zone. Just 13 percent of all employment in the sample lay within 1 kilometre of Charing Cross/Bank of England and 31 percent within 2 kilometres. Work was, however, more centrally located than residence. Just 1 percent of workers lived within 1 kilometre of one of the two centres and 11 percent within 2 kilometres.

Figure 2 shows net flow rates by borough, defined as the difference between the number employed in and the number residing in the borough (from the sample data), divided by the number residing in the borough. The general pattern across all residential boroughs was that the largest share of workers either worked within their borough of residence or commuted inwards toward the centre. Consistent with evidence from earlier *Day Censuses* and the *1921 UK Census*, the City of London was the largest net recipient of commuters, with 3,412 employees (almost twice as many as the next largest borough) but no residents in the *New Survey* data. Wealthier boroughs north and west of the City, such as Westminster, St. Marylebone, and Holborn, were also net recipients. The exterior boroughs were typically “dormitory suburbs”, with far more residents than workers, although some had large employers that attracted many workers from other boroughs, such as the Arsenal at Woolwich or the docks at Bermondsey and Poplar.

4. Empirical Modelling – Commuting, Labour Force Participation, and Earnings

There exist several mechanisms by which lower commuting costs may increase the efficiency of labour markets. First, lower commuting costs enables workers to search across more potential employers. Manning (2003b) argues that the low arrival rate of new job opportunities in a given location is sufficient to initiate commuting across otherwise identical employers. Workers trade off any disutility of commuting for higher wages. Mortensen and Pissarides (1994) and Rogerson, et al. (2005) argue that if there is a match-specific component of productivity, increased search will lead to better matches between workers and firms and thus to higher productivity and earnings. In this framework, access to public transport lowers the cost of search and of daily commuting, giving workers the ability to accept further away jobs and choose between more potential employers. This, in turn, improves the average quality of employer/employee matches, thus leading to higher productivity and wages.

Second, lower commuting costs reduce employers' local monopsony power and thus improves workers' bargaining positions (Bhaskar and To 1999; Rotemberg and Saloner 2000; and Bhaskar, et al. 2002). If employers can perfectly price discriminate, they only need to pay employees at least their outside opportunity, defined by wages at an alternative employer minus the difference in commuting costs. Public transport lowers the cost of commuting to a more remote employer and thus increases the wage that the incumbent employer needs to pay to retain the worker. Relatedly, workers may distinguish between employers in terms of non-wage aspects of the job. Because workers differentiate between employers, the labour supply curve facing individual employers is upward sloping. To attract sufficient numbers of workers, employers may need to recruit outside their immediate area. The cost of commuting thus affects individual employers' labour supply (Bhaskar, et al. 2002).

Third, employers who are geographically isolated may need to pay higher wages as a compensating differential to attract workers. Improvements in public transport will reduce the cost of travelling to a previously isolated location, lowering the compensating differential necessary to attract workers (Gibbons and Machin 2006).

These frameworks offer predictions about the effects of centrality, commuting, and ability to commute on labour markets that we can examine using the *New Survey* data. Our objective is to examine London's labour market circa 1930 using these frameworks as guidance, not to test between models, which often offer similar predictions. The search and non-wage competition models imply a positive relationship between commuting and earnings, albeit by different mechanisms. The local monopsony model implies that a worker's access to public transport (measured by proximity to the nearest train, bus, tram, or Underground stop) has a positive effect on earnings, independent of whether they commute by public transport. Similarly, the monopsony model implies proximity to the centre would lead to a greater likelihood of labour force participation and higher earnings than residing in areas with less dense employment concentration. These models also imply that access to public transport will increase labour force participation. The compensating differentials model implies that more isolated employers will need to pay higher wages to attract workers, thus earnings will increase with employers' distance from public transport.

We examine the impact of centrality, access to public transport, and commuting distance on London labour markets using the *New Survey* data, augmented by the GIS data described in previous section. We run a series of regressions on labour force participation, extent of commuting, and earnings. The independent variables in these regressions include the distance variables and a vector of personal, locational, and employment characteristics. To examine labour force participation, we run probit regressions on employment – defined either as having an earner number in the NSLLL data or reporting non-zero working hours in the previous week. We control for the following crow-flies distances: 1) from the individual's residence to their workplace 2) from residence and workplace to the nearest of the two centres and 3) from residence and workplace to the nearest stop for each of the four public transport modes. We also control for the following personal and household characteristics: age, age squared, age not reported, sex, born in England, born in London, born in same borough as current residence, born in an adjacent borough to current borough of residence, wage income of other family members, non-wage income of the household, and borough of residence.

To examine the determinants of the extent of commuting, we run OLS regressions, defining the dependent variable in the following ways: crow-flies distance commuted, whether reporting positive commuting expenditures, whether distance commuted was less than one kilometre, and whether distance commuted was 3.2 kilometres or more.[22] The independent variables are the same as in the labour force participation regressions, except we add borough of workplace and do not include the two "labour supply" variables (wage income of other family members, non-wage income of the household). In one specification, we estimate extent of commuting conditional on participation in the labour force using the Heckman correction. To address the impact of commuting on earnings, we run modified Mincer-type wage regressions (Mincer, 1958 and 1974). The independent variables are the same as in the commuting distance regression, except we add either occupation or skill dummies.

4.1. Endogeneity of Location

An important econometric issue associated with these regressions is that the locations of both residence and workplace are potentially choice variables and thus there exists the possibility of reverse causality in a naïve OLS regression of wage on distance commuted. Income may determine commuting distance by affecting the set of available residential choices.

[22] The cutoff points of 1.0 and 3.2 kilometres are selected based on discussion about transport mode in Ponsonby and Ruck (1930). In the data, 92.1 percent of workers with non-missing expenditures who commuted less than 1.0 kilometres reported expenditures of exactly zero. By contrast, 89.8 percent of those commuting over 3.2 kilometres reported positive transport expenditures.

A simple monocentric city model implies that a process of self-selection will lead to wealthier households living further from the city centre than poorer households, although this pattern will reverse if the poor can afford transport to the centre or if the centre contains particularly valued amenities (Mills 1967; LeRoy and Sonstelie 1983; Brueckner, et al. 1999). Put simply, high earning individuals could choose to live either near their workplace or, alternatively, in distant residential suburbs and commute into work. Reverse causation would imply that the estimated coefficients on the distance variables in an OLS regression would be biased, and *a priori* the extent and even the direction of the bias is ambiguous. Llewellyn-Smith (1930b) argues that London housing rental markets were very tight circa 1930, and respondents in the NSLLL would not have had the extent of residential choice as do modern urban residents. Nevertheless, it is very likely that there was at least some degree of residential choice, and thus it is necessary to mitigate the associated potential biases.

We use three distinct approaches to addressing potential endogeneity of distance commuted: sample restrictions, household fixed effects, and instrumental variables estimation. First, we run OLS regressions restricting our sample to individuals who presumably had the least choice regarding residential location. There exists a literature in urban economics which assumes that households' residential choices revolve around the primary income earner (Kain 1962; O'Reagan and Quigley 1993; Rees and Shultz 1970). Non-relatives (such as lodgers) also likely had considerable choice over where to live.[23] Thus, our OLS regression specifications exclude heads of household and non-family members. Even among family members who were secondary earners within a household, there may have been differences in the extent of residential choice. Put simply, a son aged 28 could more easily leave home to be

[23] We use the relationship categories from the original record cards to decide whether individuals were related to the head of household. We have excluded individuals if there is ambiguity in the relationship ("single", "bachelor", "spinster"), as well as if it is clear that they were unrelated ("lodger").

nearer to work than could a daughter aged 16.[24] There is no formal econometric test for the appropriate sample and thus we run multiple specifications, progressively restricting the sample based on the plausible level of residential choice. As we show below, in practice the results are fairly robust to sample specification once heads of household and non-relatives are excluded.

In order for this strategy to identify a causal effect, it must be the case that there is no association between the workplace location of the head (who is excluded from the sample) and that of working relatives (who are in the sample). In such cases, the distance commuted variable would be proxying the real determinant of wages, household head influence. It is of course possible that some heads of households may have been able to help secure employment at the same workplace for other family members. The evidence, however, suggests that this was rare. Baines and Johnson (1999) show that it was fairly rare for fathers and sons to even work within the same trade. We expand on this by using the information in the data on workplace address to further split the sample, excluding households for which one or more relatives were employed at the same workplace as the head, as the estimates from these workers would be presumably more compromised by this type of endogeneity bias. In practice, only around 3 percent of individuals under the age of 25 worked with the same employer as the head and only 5 percent of individuals under the age of 25 lived in a household where someone had the same employer as the head. The removal of such a small group from the estimation sample is unlikely to make much difference to the OLS estimates and, thus, any estimation bias for the returns to commuting from this source is likely to be small.

[24] Arguably wives of household heads would have had among the least residential choice. However, we do not focus on them because working spouses are found in less than 8 percent of households with a male head in work. In contrast, 60 percent of women who were heads of household and under the age of 60 were working and 82 percent of women (and 83 percent of men) aged 14 and over but still living in the parental home were working. The median household in the sample contained three occupants and one earner. Among sample households, 42 percent of households had four or more occupants and 34 percent of sampled households had more than one earner.

Our second approach to controlling for endogeneity caused by unobserved heterogeneity in location choice in our OLS regressions is running a household fixed-effects specification. Including household fixed effects means that the distance variables are identified by within-household differences in commuting, effectively different household members travelling to different work locations. This will reveal whether individuals in the same household (and by extension the same location) received higher wages with increases in distance travelled. This approach will also mitigate against biases associated with any unobserved household level characteristics associated with location that also determine wages. For example if it were the case that more (less) "aspirational" households lived further from (nearer) the centre and "aspiration" is associated with higher earnings and the extent of "aspiration" is common within households, the household fixed effects should net this out.

Our third approach to identification is to use an instrumental variable (IV) strategy. The *New Survey* data contains information on the place of birth of most individuals. We can therefore calculate the distance between place of birth centroid and Charing Cross/Bank of England in the same way that we calculate the distance between home and work.[25] We use this birthplace distance to instrument the distance between home and work. Any IV must satisfy three exogeneity exclusion restrictions in order to produce consistent estimates: 1) relevance – the instrument must be "sufficiently" correlated with the endogenous regressors, 2) no direct influence – the instrument cannot plausibly influence the dependent variable directly through mechanisms other than its correlation with the endogenous variables, 3) monotonicity - the instrument should affect the endogenous variable of all individuals in the sample with the same sign (Angrist, et al.1996). We argue below and in Appendix III that this is indeed the case for our sample.

[25] Sometimes place of birth is recorded as a street. More often it is recorded as a local area. We use the centroid of the area as the basis for the distance calculation for the latter cases.

The logic behind the instrument's compliance with the "relevance" condition is that birthplace "fixes" residence and Charing Cross/Bank of England "fixes" workplace. Intuitively, this will hold if, first, birthplace and current location are correlated and, second, if workers tend to commute inwards toward the two centres. There is indeed a strong correlation between place of birth and place of residence. Among all income earners in the sample born in the *New Survey* area, 31.5 percent lived in the same borough that they were born in, and 11.9 percent resided in an adjacent borough. As we have shown in Section III, there is a strong tendency to commute inwards. The instrument may, however, be a "local" one, since current location may be more highly correlated with place of birth for those born in London than those born further away. As such any IV estimate can only be interpreted as causal for the group of compliers for whom place of birth affects current location, (the local average causal response in the case of a continuous endogenous variable). We address the "relevance" condition more formally in the first stage of our IV regressions.

There is no formal test for the "no direct influence" condition; however, it seems plausible that distance from birthplace to the nearest centre does not affect earnings, except through commuting distance. Neither birthplace nor the location of the centres are choice variables for individual workers, thus there can be no concerns about reverse causation.[26]

The third requirement of a valid instrumental variable approach, "monotonicity", implies in our case that those born further away from Charing Cross/Bank of England centroid commuted further. It is, however, possible that "local" London labour markets – concentrations of work outside the centre – may lead to a population of "defiers" in our sample. A large

[26] One possible threat to the IV strategy would be if distance from birthplace to the centre was an indicator of father's social status (e.g. the rich lived further from the centre, as per the monocentric city model) and there was low intergenerational mobility. In this case, the IV would be proxying head of household influence and thus be correlated with the error term. However, the evidence for both of these conditions is very weak. London has rich and poor areas, but they are more correlated with direction (north and west are wealthier than south and east) than with distance from the centre (Heblich, et al. 2021). Moreover, Baines and Johnson (1999) show that in the *New Survey* data there is a high degree of social mobility; sons rarely follow their fathers into a trade and there is a low correlation of Armstrong skill categories across generations.

population of defiers – e.g. individuals born some distance from the centre of London who nevertheless commute a short distance to work into one of these "local" labour markets – would undermine the IV strategy. In these circumstances, the IV estimates cannot be attributed as causal, local or otherwise. We examine the monotonicity assumption in detail in Appendix III. In brief, we find that for (banded) distances of up to about 10 kilometres from the centres, a stochastic version of monotonicity assumption holds, e.g. there are more "compliers" (individuals who commute inwards towards the centre) than defiers. Outside 10 kilometres, local labour markets become more important and the monotonicity assumption fails. In other words, local employers further from the centre, such as the Woolwich Arsenal, the East End docks, and the "new industries" concentrated just outside the *New Survey* area, become increasingly important for local residents and, thus, reduce average commuting distance. Thus, while we estimate the IV regressions for a variety of distances, we have far more confidence in the IV regressions using only workers born within 10 kilometres of one of the two centres. We use the sample of workers born within 10 kilometres of a centre as our baseline specification. This group comprises about 85 percent of workers in the sample with an identifiable birthplace.

5. Results

5.1 Effects of Public Transport on Labour Force Status and Commuting Distance

Table 3 shows results for the employment and distance commuted regressions for the sample who are related to the head of household and aged 14 (the school-leaving age in 1930) or over. The first two columns show results for the probability of being in work. We report the estimated marginal effects for the explanatory variables of interest.[27] The last five columns show results

[27] The estimated marginal effects for the other explanatory variables are given in Appendix IV, Table A.IV.1. results for the full sample, including heads of household and individuals not related to the heads, are shown in Appendix IV, Table A.IV.2.

for distance commuted. To ensure the robustness of our results, we change the dependent variable in the different specifications, and, in column 4, jointly estimate the effects of employment and commuting distance using the Heckman correction. In this regression, the first stage is identified by the standard "labour supply" variables, wage income of other family members and non-wage income of the household.

The estimated size of the distance effects are similar for the two participation specifications. The coefficient on distance from the nearer centre is negative and significantly different from zero in both the hours worked (column 1) and earner number specifications (column 2). Individuals residing more centrally were more likely to be employed, presumably because of the greater concentration of jobs in the central areas.[28]

The results of the commuting regressions are consistent across specification. Individuals residing near one of the centres commuted shorter distances (columns 3 & 4, row 1), were less likely to incur transport expenses (column 5 row 1), were more likely to work locally (column 6 row 1), and were less likely to commute medium to long distances (column 7 row 1). As with the employment regressions, the logical interpretation of these results is that labour markets were much thicker and thus there were more local employment opportunities near the two centres.

Access to the Underground – conditional on distance from the centre – was associated with longer commutes, a higher probability of incurring transport expenses, a lower probability of working locally, and a higher probability of a medium to longer commute (row 3, columns 3 to 7). There is a sharp contrast between access to the Underground and access to the train system, the two transport modes used to commute longer distances. The coefficients on access to the train are much smaller than on access to the Underground and are insignificant in all but

[28] As robustness tests we re-estimate the models using family fixed effects (Appendix IV, Table A.IV.2). We also replaced the distance from public transport variables with the number of train/Underground stations and bus/tram routes in the same 500 square meter grid (and one square kilometre grid) as the individual's residence. The results are qualitatively similar to those shown in Table 3.

one specification, suggesting either that commuting by train was fairly uncommon or that local employment near train stations more-or-less offset longer-distance travel by train. The coefficients on distance to bus and tram stops are generally insignificant and the estimated marginal effects small.

5.2. Effects of Access to Public Transport and Commuting on Earnings

Table 4 shows the OLS estimates of distance effects for the earnings regressions. In the main specification (column 1), we report estimates on the various distance measures, restricting the sample to relatives of the head of household, and thus exclude heads, lodgers, and other non-relatives. In column 2, we jointly estimate earnings and the probability of employment and report the Heckman selectivity corrected earnings estimates on distance. As robustness tests, we further restrict the sample to individuals under age 25 (column 3) and children of the head of household under age 25 (column 4).[29] We also run the regression using all individuals in the sample (column 5) and add household fixed effects (column 6). The control variables in the regressions are generally significant, have the expected sign, and are consistent with other studies estimating earnings (see Appendix IV, Table A.IV.3).

The strongest and most robust results in Table 4 pertain to the distance travelled from home to work. The estimated coefficients on the distance commuted and its square are large and strongly significant in every specification. The magnitude of the net effect is very similar across specifications and all household members. A one-kilometre increase in distance is associated with between a one to two percent increase in earnings, depending on distance travelled. The presence of the quadratic allows the distance effect to be non-linear. The coefficient on the quadratic term is negative suggesting that the marginal gains to commuting

[29] Children tended to live with their parents until their late-20s. Among individuals in the sample aged 25 or less, 78.6 percent lived in a household headed by one of their parents, 7.6 percent were the head of household, 10.1 percent were the spouse of the head of household, 2.9 percent lived in a household headed by another relative, and 0.1 lived in a household headed by a non-relative (usually as a lodger). At age 25, 45.4 percent of women and 54.5 percent of men in the sample were listed as the child of the head of household.

fall with distance.[30] The coefficients can be interpreted as semi-elasticities and imply that, when evaluated at the mean, earnings increased by about 1.5-2.0 percent for each kilometre commuted. [31]

Controlling for common unobserved attributes among working household members does not change the estimated coefficients appreciably. The inclusion of household fixed effects reduces the estimated returns to distance by around 0.5 percent a kilometre but they remain significantly different from zero. In Appendix II, we show that the measured crow-flies distance commuted is likely to be an over-estimate. This implies that there will be attenuation bias in the regression and thus the estimated returns to distance should be interpreted as a lower bound estimate of the true returns. The estimates in Table 4 use distance travelled as the primary independent variable, irrespective of the direction of commuting. White (1999) argues that the presence of employment centres outside the central zone can compromise estimated returns to distance. In Appendix Table A.IV.6 we test whether the estimates to distance commuted depend on direction of commute by interacting the distance travelled with dummy variables for commuting out and across London. There is no significant difference in the estimated returns to distance for the different commuting patterns.

While there is a strong effect for distance commuted on earnings, the effects of the other distance variables are far weaker. The coefficients on both home and workplace centrality are

[30] The implied distance before the marginal returns to commuting turn negative is around 20-25 kilometres depending on the sample used. Only about 0.4 percent of workers in the sample commute 20 kilometres or more, so the negative marginal returns is effectively an out-of-sample estimate.

[31] As further robustness tests, we have also run additional regressions which 1) replace the designated head with the highest income earner in each household, 2) replace the distance to public transport variables with the number of stops/stations within the same one kilometre squared grid as the workers' residence and workplace, 3) include only heads of household, 4) replace the continuous distance commuted variable with discrete categories, 5) include very long commutes (50+ kilometre), 6) exclude workers' occupation and workplace borough dummies, 7) replace occupation dummies with skill categories, 8) exclude observations collected by the most prolific enumerator, G.E. Bartlett, whose accuracy has been questioned in Abernathy (2017), 9) exclude individuals working for the same employer as their head of household, 10) exclude all members of households where at least one other member works for the same employer as the head. In all cases, the main results are qualitatively very similar to those presented in Table 4. These results are shown in Appendix IV, Table A.IV.4. Appendix IV, Table A.IV.5 replaces continuous distance with dummy variables for discrete distance intervals measured relative to a base category of a less than a 0.5-kilometre commute. The estimated effects rise montonically but nonlinearly with distance commuted.

insignificant in nearly every specification. The coefficients on the access to public transport variables are small, mostly insignificant, and not robust to specification. A likely explanation for this is that there was relatively little variation in access to public transport other than the Underground, with almost all individuals in the sample living within a 10-15 minute walk of some sort of transport.

5. 3 Instrumental Variable Estimates of Earnings

Since we only have one instrument and several distance variables, we estimate the wage returns model again solely as a function of the distance to work variable, along with age, gender, occupation and workplace borough dummies. The main results are given in Table 5, and full results are in Appendix IV, Table A.IV.6. We report OLS estimates for each distance group alongside their two stage least squares counterparts using the full sample of wage earners aged 14 and over. This enables us to determine the robustness of the IV estimates to different birth-distance thresholds in the results that follow. The OLS point estimates are quite stable, at around 0.014, no matter the distance cut-off sample, indicative of a 1.4 percent wage premium for each kilometre commuted.[32] In addition to being quite stable, the point estimate of the effect of distance commuted is very similar to that shown in Table 4.[33]

The IV estimates are also quite stable for the sample populations most likely to exhibit monotonicity behaviour. The point estimates for the first three IV sub-samples (born less than 6, 8, and 10 kilometres from the centre of London, respectively) are around 0.033 to 0.038, indicative that the average causal response wage effect may be somewhat higher than the OLS estimates suggest, at around 3.5 percent for each kilometre commuted. This is perhaps not

[32] Since we remove the quadratic in distance, the estimated coefficient on the distance variable is effectively the average marginal return over all distances and so is in line with the estimates in Table 4. The removal of the other distance variables does not affect the estimate of distance to work in Table 4.

[33] The magnitude of these estimates is somewhat larger than those found in the (rather rare) literature on contemporary commuting returns to distance. Mulallic et al (2013) report a wage-distance elasticity of around 0.015. If we estimate the wage equations using log of distance to work, the estimated wage-distance elasticities range from 0.04 (OLS) to 0.1 (IV).

surprising, given that we show in Appendix II that we are likely to overestimate distance travelled and thus underestimate the returns. The instrument is also highly relevant for these sub-samples. The first stage F statistic of the significance of the instrument is above 190 in column 9 (which implies a t statistic of about 13.8 on the instrument).[34] The estimated coefficient on the instrument in the first stage (not shown) suggests that distance commuted rises by around 0.2 kilometres for every kilometre further from the centre the individual was born. Extending the sample incrementally into sub-populations whose behavior is much less likely to exhibit monotonicity and/or compliance regarding the instrument, the IV point estimates start to rise, (Table 5, columns 10 to 12). This is consistent with the idea that the identification conditions required of an IV estimator may indeed be less likely to hold across the full sample. However, for the majority of the sample born within 10 kilometres of the centre of London, the IV estimates may be closer to describing the average causal response of distance on wages.

5.4 Commuting Costs

Commuting has monetary and non-monetary costs in addition to the benefits shown in Tables 4 and 5. Table 6 shows some simple back-of-the-envelope calculations of the returns and costs of 8 "standard" commutes, which are stylized versions of what we observe in the data. The table shows one-way commuting distances and the time, estimated monetary returns, and monetary and implied time costs for each of these commutes. We follow Ponsonby and Ruck (1930) in identifying the likely mode of travel based on distances. The monetary costs are typical costs over the distance taken from reported travel expenses in the *New Survey* data. We estimate the returns to commuting at the tenth, twenty fifth, and fiftieth percentiles of the weekly income distribution using the regression results from the first column of Table 4 and the ninth column of Table 5 (in parentheses). We estimate the implied time cost by calculating

[34] The standard errors in the first stage are clustered by "birthplace". There are 220 birthplaces in the sample.

the estimated time spent walking, waiting, and taking public transport for each commute and multiplying this by 50 or 100 percent of hourly earnings.[35] To do so, we assume a walking speed of four kilometres an hour, public transport speeds shown in Table 1, and waiting times of five minutes for bus and tram and eight minutes for train and Underground. Full details of these calculations are shown in Appendix V.

Table 6 shows that the monetary returns from commuting outweighed the monetary costs for all but the lowest earners (approximately the bottom 10-20 percent). Workers above this level of income would not have faced income constraints that prevented them from commuting, as their higher earnings due to commuting would have paid for the monetary costs of travel. Whether these workers would have chosen to commute would thus depend only on whether the total (monetary and non-monetary) returns outweighed the total costs. Table 6 also shows that time costs were substantial even for low earners, suggesting that most individuals would not have commuted unless there were non-monetary benefits, derived from greater choice over place of residence or non-pecuniary aspects of their job, as well as monetary returns. We do not observe non-monetary costs or benefits for individuals, and it is likely that there was substantial heterogeneity in both. This is consistent with the high standard deviation of commuting distances and sizable share of workers working locally despite the monetary returns to commuting outweighing the monetary costs for all but the lowest earners.

For at least some workers in the bottom quintile of the earnings distribution, the monetary costs of commuting outweighed the returns. If these workers were the primary source of income for their household, it is likely that they would have been income constrained, and unable to afford to commute even if they would have received substantial non-monetary returns from commuting. On the other hand, if these workers were secondary earners from wealthier

[35] This is a fairly typical range of estimated values of time travel savings in the urban economics and geography literatures (Wardman 1998; Zamparini and Reggiani 2007).

households, it is much less likely that they would have faced income constraints. These workers would have been able to commute, so long as there were sufficient non-monetary benefits. To determine whether low earners were typically from poor households, we construct household-specific poverty lines using the approach outlined in Hatton and Bailey (1998). The poverty lines are based on minimum required expenditure on food and clothing, rent, and fuel given the structure of each household and actual expenditure on National Insurance and transport. Further details of the construction of the poverty line are available in the appendix of Hatton and Bailey (1998) and Appendix V of this paper. We estimate the share of workers in the bottom 10 percent (25 percent) of the earnings distribution who resided in households under the poverty line to be 25.2 percent (17.3 percent). Although low earners were more likely than the sample as a whole to be below the poverty line, a majority of low earners were secondary earners in wealthier households. These figures imply that less than five percent of workers in the *New Survey* data faced income constraints that may have prevented commuting.

5.5. Comparison to the 1890s

Ponsonby and Ruck (1930, pp. 171, 191) argue that newly available modes of transport in the early twentieth century led to fundamental changes in the working-class labour market, stating, "No change in the last generation has had more far-reaching effects upon the life of the whole community in London than the improvement of transport facilities. … It must be remembered, above all, in this connection that by far the greatest proportion of the increase (in commuting) is due to working-class travel. In Charles Booth's time [the 1890s] workmen travelled but little, being generally employed on the spot." The evidence on commuting in companion volumes to the LLPL is largely indirect, but also suggests that most working-class employees in the 1890s did not commute. There are only a few direct references to working-class commuting in these volumes, generally pertaining to footloose occupations such as the building trades (Booth 1902, Vol. IX, p. 17, Vol. V, p. 125).[36] The LLPL volumes contain numerous mentions of outwork from home, the most extreme absence of commuting.[37] The LLPL volumes also make multiple mentions of workshops adjacent to or very near workers' residences, and to neighbourhoods of specific groups of workers, such as dock labourers, located near their workplace.

The primary reason for the lack of inner-city commuting in the late-nineteenth century was almost certainly under-developed infrastructure. As can be seen in Table 1, the bus, tram, and Underground networks had fewer route miles and vehicle miles and slower travel speeds in the early-twentieth century than at the time of the *New Survey*. The rail network was almost fully developed by the end of the nineteenth century, but, as shown in Section II, there were substantial barriers to rail commuting by the working-class. These supply-side issues were even greater in the 1890s, as most of the Underground lines had yet to be opened and buses and trams were still horse-drawn. Tables 1, 4, 5, and 6 highlight an additional demand-side explanation for absence of commuting in the earlier period, namely income constraints. The real monetary cost of travel was considerably higher in the first decade of the twentieth century than in 1930 (Table 1). It is likely that costs in the 1890s were higher still. Conversely, real weekly earnings were approximately 20 percent lower in the 1890s than circa 1930.[38] The regression results in Tables 4 and 5 suggest that the total return to commuting would have been substantially lower in 1890 than in 1903, as commuting distances were much smaller. The combination of lower income, lower returns to commuting, and higher cost of public transport suggests that a much higher proportion of workers in the 1890s would have faced income constraints than was the case in 1930.

[36] Booth (1902) refers to commuting by middle class workers, such as bankers and clerks (Booth 1902, vol. IX, p. 189). He also refers to working-class travel in the context of outworkers picking up raw materials and dropping off finished products. However, these trips were only made on a weekly or bi-weekly basis.
[37] Booth (1902), Vol. IX, p. 204-5, Vol. IV, pp. 19, 41-42, 60, 71, 73, 79, 117, 149, 160-1, 174, 204, 278, 295.
[38] Lewellyn-Smith (1930a, p. 19) states that the real weekly earnings of working-class Londoners increased by about 20 percent between 1890 and 1928. After adjusting for the decline in the workweek and making the comparison between like-for-like workers, he concludes that the real hourly increase was about a third. This is very similar to recent estimates of changes in real earnings over the period. Clark (2020) calculates real earnings increased 35.6 percent for the entire UK over the period 1895-1930.

Tables 1, 4, 5, and 6 and the discussion from Booth (1902) can be used to estimate additional earnings due to increased commuting distance between 1890 and 1930. Based on the discussion of co-location of residences and workplace and on the absence of discussion of longer commutes in Booth (1902), we believe that 200-500 meters would have covered the range of typical crow-flies distances from home to work in the 1890s. Using the 1930 distances shown in Table 2, this implies an increased commuting distance 2.5-3.0 kilometres each way or 25-36 kilometres per week. Taking the smaller figure of 2.5 kilometres and using the estimated returns per kilometre from Table 4, column 1 and Table 5, column 9, this increased distance would imply an increase in earnings (evaluated at the mean) of about 5.0 to 8.25 percent, using the OLS and IV estimates, respectively. This accounts for about 25-41 percent of the increase in real weekly earnings or 16.5-27.0 percent in real hourly earnings of the working-class over this period of time. The lower end of these figures is similar to Leunig's (2006) finding that social savings from railways accounts for about one sixth of economy-wide productivity growth during late-nineteenth and early-twentieth centuries.

6. Conclusions

In the 1890s the only available means of high-speed, low-cost transport was rail, which predominantly connected middle-class and professional workers from residential suburbs to the central areas of the metropolis. The working-class typically worked where they lived. By 1930, buses, trams, and the London underground had comprehensive networks running fast and relatively inexpensive routes throughout the inner-city area. Data from the *New Survey of London Life and Labour* show that these networks were widely used by the working-class. Over 70 percent of income earners in the data commuted at least one kilometre. Over twice as many workers reported positive transport expenditures as zero expenditures, although information about transport expenditures is frequently missing.

In this paper we have used data from the *New Survey* to examine the impact of access to public transport and commuting on working-class London labour markets, circa 1930. We GIS code home addresses and places of work to construct crow-flies distance commuted, distance from public transport, and distance from the metropolitan centres. We then use these distance variables in regressions on labour force participation and earnings. We control for the potential endogeneity of distance (caused by the link between income and residential choice) by 1) restricting the sample to individuals' whose residence was plausibly exogenous, 2) using household fixed-effects, and 3) instrumenting distance commuted with distance from birthplace to work. We find that residential centrality increased the probability of labour force participation. We also find that distance commuted had a strong impact on earnings, with a one-kilometre increase in commuting distance resulting in a plausibly causal estimate of about a 1.5-3.5 percent increase in earnings. However, we find little evidence to suggest that proximity of home or workplace to public transport had an independent effect on labour force participation or earnings, perhaps because virtually all areas covered in the *New Survey* were within easy walking distance of some form of public transport. All of these results are robust to a variety of specifications.

Finally, although the focus of this paper is on labour markets, we end by noting that the commuting revolution likely had additional consequences for the standard of living of working-class Londoners. Economic historians have generally accepted that living standards rapidly increased from the late-nineteenth century, across a range of metrics including income, health, leisure time, education, etc. (Maddison 1964; Horrell 2000; Clark 2005; Chapman 2019). Our focus on labour markets has shown that commuting was an important contributor to increased income. It is likely that, independent of income effects, it also contributed to improvements in health by reducing crowding. Commuting likely facilitated outwards residential movement, reducing crowding in inner London. Most of the working-class boroughs in the central and middle rings of the metropolis reached their maximum *UK Census* population in 1901, just before the major improvements to public transport infrastructure and technology (London County Council, 1928). Conversely, population in all but two boroughs in the outer ring peaked in 1921 or later. The *New Survey* included nine municipal boroughs and urban districts which were absent from the *Life and Labour of the People of London* precisely because there had been substantial working-class movement into these areas between 1890 and 1930 (Llewellyn-Smith 1930). Ponsonby and Ruck (1930) argue that commuting was the direct cause of this outwards movement, although it is beyond the scope of this paper to formally test this. In addition to reducing crowding within areas, it is likely that commuting also reduced crowding within dwellings by at least two mechanisms. First, the income increases associated with commuting would have led to families consume more housing, leading to larger dwellings. Second, the outwards movement of the working-class population described above, meant more were living in the outer ring of the metropolis where dwellings were typically larger than in the central areas (Llewellyn-Smith 1930).

In addition to these benefits, increased commuting had one additional non-monetary cost through its affect on time use. The 200-500 meter walk typical of the 1890s would have taken perhaps 3-7.5 minutes, whereas a 3-kilometre bus trip typical of 1930 would likely have taken about 9 minutes plus another 10-15 minutes walking at either end and waiting for the bus. Assuming a five and a half or six-day work week, there would have been about four hours weekly difference in the typical commuting times between the 1890s and 1930s. Huberman and Minns (2007) estimate that work time in the UK declined from 56.3 hours per week in 1890 to 47.0 hours in 1929. However, this estimate is “clock in to clock out”, not “front door to front door”. For the worker, commuting time may resemble work time much more closely than leisure time and, thus, our estimates of increased commuting times suggest a much smaller decrease in effective working time than has previously been believed.

References

Abernethy, S. (2017). “Deceptive data? The role of the investigators in the New Survey of London Life and Labour 1928–32.” Historical Methods: a Journal of Quantitative and Interdisciplinary History, 50(1):1-15.

Aldred, R. (2014). “The commute.” In: Adey, P.; Bissell, D.; Hannam; K., Merriman; P. Sheller; M., eds. *Routledge Handbook of Mobilities*. (Abington: Routledge), pp. 450-9.

Allen, R. (2009). “Engels’ pause: technical change, capital accumulation, and inequality in the British industrial revolution.” Explorations in Economic History, 46(4): 418-35.
References

Angrist, J.D., Imbens, G.W. & Donald B. Rubin, D. B. (1996). “Identification of causal effects using instrumental variables.” Journal of the American Statistical Association, 91(434): 444-55.

Baines, Dudley and Paul Johnson, Paul (1999). “In search of the 'traditional' working class: social mobility and occupational continuity in interwar London.” Economic History Review, 52(4): 692-713.

Bhaskar, V. and To, Ted (1999). “Minimum wages for Ronald McDonald monopsonies: a theory of monopsonistic competition.” Economic Journal, 109(455): 190-203.

Bhaskar, V.; Manning, A, and To, T., (2002). “Oligopsony and monopsonistic competition in labour markets.” Journal of Economic Perspectives, 16(2): 155-74.

Booth, C. (1902). *Life and Labour of the People of London*. (London: Macmillan).

Brueckner, Jan K. Thisse, Jacques-François, and Zenoubd, Yves (1999). “Why is central Paris rich and downtown Detroit poor?: An amenity-based theory.” European Economic Review, 43(1): 91-107.

Chapman, J., (2019). “The contribution of infrastructure investment to Britain’s urban mortality decline, 1861–1900.” Economic History Review, 72(1): 233-59.

City of London (1866, 1881, 1891, 1911). *Report of the Day Census*. (London: Simpkin, Marshall, Hamilton, Kent & Co.).

Clark, G., (2005). “The condition of the working class in England, 1209–2004.” Journal of Political Economy, 113(6): 1307-40.

Clark, G., (2020). “What were the British earnings and prices then? (new series).” Measuring Worth, http://www.measuringworth.com/ukearncpi/ [Accessed 25/9/2020].

Crafts, N., (1997). “The human development index and changes in standards of living: some historical comparisons.” European Review of Economic History, 1(3): 299-323.

Crafts, N. and Leunig, T., (2005). “The historical significance of transport for economic growth and productivity.” Eddington Transport Study: Research Annexes Vol. 1.

Dyos, H. J. (1953). “Workmen’s fares in south London, 1860 – 1914.” Journal of Transport History, fs-1(1): 3-19.

Feinstein, C., (1998). “Pessimism perpetuated: real wages and the standard of living in Britain during and after the Industrial Revolution.” Journal of Economic History, 58(3): 625-58.

Gibbons, S. and Machin, S., (2006). “Transport and labour market linkages: empirical evidence, implications for policy and scope for further UK research.” Department of Transport.

Greater London Authority (2015). *Characteristics of Commuters*. (London: Greater London Authority).

Hamilton, Bruce W. (1982). “Wasteful commuting.” Journal of Political Economy, 90(5): 1035-105.

Hatton, Timothy and Bailey, Roy (1998). “Poverty and the welfare state in interwar London.” Oxford Economic Papers, 50(4): 574-606.

Heblich, Stephan, Redding, Stephen, and Sturm, Daniel (2020). “The making of the modern metropolis: evidence from London.” The Quarterly Journal of Economics, 135(4): 2059–133.

Heblich, Stephan, Trew, Alex and Zylberberg, Yanos (2021) “East-side story: historical pollution and persistent neighborhood sorting.” Journal of Political Economy, 129(5): 1508-52.

Horrell, Sara (2000). “Living standards in Britain 1900-2000: women's century?” National Institute Economic Review, 172: 62-77

Horrell, Sara and Humphries, Jane (1992). “Old questions, new data, and alternative perspectives: families' living standards in the Industrial Revolution.” Journal of Economic History, 52(4): 849-80.

Huberman, Michael and Minns, Christopher (2007). “The times they are not changin’: days and hours of work in Old and New Worlds, 1870–2000.” Explorations in Economic History, 44(4): 538-67.

Johnson, Paul, Bailey, Roy, Baines, Dudley, Raspin, Anna, Hatton, Timothy (1999). *New Survey of London Life and Labour, 1929-1931*. [data collection]. 2nd Edition. UK Data Service. SN: 3758, http://doi.org/10.5255/UKDA-SN-3758-1. https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=3758 [Accessed 25/9/2020].

Kain, John (1962). “The journey-to-work as a determinant of residential location.” Papers in Regional Science, 9(1): 137-60.

Komlos, John (1998). “Shrinking in a growing economy? The mystery of physical stature during the Industrial Revolution.” Journal of Economic History, 58(3): 779-802.

LeRoy, Stephen F. and Sonstelie, Jon (1983). “Paradise lost and regained: transportation innovation, income, and residential location.” Journal of Urban Economics, 13(1): 67-89.

Leunig, Tim (2006). “Time is money: a re-assessment of the passenger social savings from Victorian British railways.” Journal of Economic History, 66(3): 635-73.

Lindert, P. H. and Williamson, J. G. (1983). “English workers' living standards during the Industrial Revolution: a new look.” Economic History Review, 36(1): 1-25.

Llewellyn-Smith, Hubert (1930a). “Introduction.” In: Llewellyn-Smith, Hubert (ed.) *The New Survey of London Life and Labour* (London: P. S. King), pp. 1-57.

Llewellyn-Smith, Hubert (1930b). “House rents and overcrowding.” In: Llewellyn-Smith, Hubert (ed.) *The New Survey of London Life and Labour* (London: P. S. King), pp. 143-70.

London County Council (various years). *London Statistics*. (London: P. S. King and Son).

London School of Economics (2020). *Charles Booth's London: Poverty Maps and Police Notebooks*. https://booth.lse.ac.uk/map/14/-0.1174/51.5064/100/0 [accessed 13/9/2020].

London Transport Museum (2020). “London's transport history: a social history of London's public transport, 1829-2000.” https://web.archive.org/web/20070609023730/http://www.ltmuseum.co.uk/learning/online_resources/ecobus_omnibus/pg/intro.htm [Accessed 25/9/2020].

Maddison, A., (1964). *Economic Growth in the West*. (New York: Twentieth Century Fund).

Manning, A., (2003a). *Monopsony in Motion*. (Princeton: Princeton University Press).

Manning, A., (2003b). “The real thin theory: monopsony in modern labour markets.” Labour Economics, 10(2): 105-31.

Measuring Worth (2020). “Five ways to compute the relative value of a U.K. dollar amount, 1270 to present.” https://www.measuringworth.com/calculators/ukcompare/

Mills, Edwin S. (1967). “An aggregative model of resource allocation in a metropolitan area.” American Economic Review, 57(2): 197-210.

Mincer, J., (1958). “Investment in human capital and personal income distribution.” Journal of Political Economy, 66(4): 281–302.

Mincer, J., (1974). *Schooling, Experience and Earnings*. (New York: National Bureau of Economic Research).

Morgan, Mary, Bhullar, Inderbir, Wise, Sarah, Power, Anne, Garner, Katie, Reid, Aileen and Field, Jacob (2019). *Charles Booth's London Poverty Maps*. (London: Thames & Hudson).

Mortensen, Dale and Pissarides, Christopher (1994). “Job creation and job destruction in the theory of unemployment.” The Review of Economic Studies, 61(3): 397-415.

Mulalic, I., Van Ommeren, J. and Pilegaard, N. (2014), “Wages and commuting: quasi-natural experiments' evidence from firms that relocate.” Economic Journal, 124: 1086–105.

Munby, D. L. (1978). *Inland Transport Statistics: Railways; Public Road Passenger Transport: London's Transport v. 1: Great Britain, 1900-70*. (Oxford: Clarendon Press).

National Library of Scotland (2020). *Side-by-Side Map Images*. https://maps.nls.uk/ [Accessed 25/9/2020].

Orford, Scott, Dorling, Danny, Mitchell, Richard, Shaw, Mary and Davey Smith, George 2002. "Life and death of the people of London: a historical GIS of Charles Booth's inquiry." *Health & Place*, 8(1): 25-35.

O'Regan, Katherine and Quigley, John (1993). "Family networks and youth access to jobs." Journal of Urban Economics, 34(2): 230-48.

Paul, Jonathan (2016). "High-resolution geological maps of central London, UK: comparisons with the London Underground." Geoscience Frontiers, 7(2): 273-86.

Polasky, Janet (2010). *Reforming Urban Labour: Routes to the City, Roots in the Country* (Ithaca: Cornell University Press).

Ponsonby, G. and Ruck, S. K. (1930). "Travel and mobility". In: Llewellyn-Smith, Hubert, eds. *The New Survey of London Life and Labour*. (London: P. S. King), pp. 171-99.

Prados de la Escosura, Leandro (2015). "World human development: 1870–2007." Review of Income and Wealth, 61(2): 220-47.

Redding, Stephen and Sturm, Daniel (2016). "Estimating neighborhood effects: evidence from war-time destruction in London." Mimeo, London School of Economics.

Rees, Albert and Schultz, George (1970). *Workers in an Urban Labour Market*. (Chicago: University of Chicago Press).

Rogerson, Richard; Shimer, Robert; and Wright, Randall (2005). "Search-theoretic models of the labour market: a survey." Journal of Economic Literature, 43(4): 959-88.

Rotemberg, Julio and Saloner, Garth (2000). "Competition and human capital accumulation: a theory of interregional specialization and trade." Regional Science and Urban Economics, 30(4): 373-404.

Seltzer, Andrew (2020a). *London Bus Map, 1931-32*. http://sharemap.org/ajseltzer1/london_bus_1931#!webgl [accessed 25/9/2020].

Seltzer, Andrew (2020b). *London Underground Map, 1931*. http://sharemap.org/ajseltzer1/underground_1#!webgl [accessed 25/9/2020].

Seltzer, Andrew (2020c). *London Train Map, 1931*. http://sharemap.org/ajseltzer1/London_Railways#!webgl [accessed 25/9/2020].

Seltzer, Andrew (2020d). *London Tram Map, 1931*. http://sharemap.org/ajseltzer1/trams_1#!webgl [accessed 25/9/2020].

Streetmap (2020). https://www.streetmap.co.uk/ [accessed 25/9/2020].

Szreter, Simon and Hardy, Anne (2001). "Urban fertility and mortality patterns." In: Martin Daunton, ed. *The Cambridge Urban History of Britain*. (Cambridge: Cambridge University Press), pp. 629-72.

van Ommeren, Jos, Rietveld, Piet, and Nijkamp, Peter (1997). "Commuting: in search of jobs and residences." Journal of Urban Economics, 42(3): 402-21.

Voth, Hans-Joachim (1998). "Time and work in eighteenth-century London." *Journal of Economic History*, 58(1): 29-58.

Wardman, Mark (1998). "The value of travel time: a review of British evidence." Journal of Transport Economics and Policy, 32(3): 285-316.

White, M.J. (1988b). "Urban commuting is not 'wasteful'." **Journal of Political Economy**, 96(5):1097-110.

Zamparini, Luca and Reggiani, Aura (2007). "Meta-analysis and the value of travel time savings: a transatlantic perspective in passenger transport." Networks and Spatial Economics, 7(4): 377-96.

Table 1. Public Transportation Statistics for 1900-07, 1913, and 1929

		1900-07	1913	1929
Local and Underground railways	Route miles	95.2	110.6	120.6
	Train miles (1,000,000s)	5.01	18.45	24.13
	Passengers (1,000,000s)	214.5	474.7	648.8
	Average fare per mile (1930 pence)	1.143	0.989	0.790
	Average scheduled speed (in kilometres per hour)	24.1	28.8	31.1
Mainline railways	Route miles	561.2	558.2	534.1
	Daily trains	2,097	NA	2,799
	Average scheduled speed (kilometres per hour)	32.35	NA	NA
	Passengers (1,000,000s)	233	250	415
Buses	Route miles	300	467	1,170
	Car miles (1,000,000s)	42.0	110.0	215.7
	Average scheduled speed (kilometres per hour)	8.0	13.7	15.3
	Passengers (1,000,000s)	264.5	735.7	1,912.1
	Average fare per mile (1930 pence)	1.879	1.061	0.960
	Average seats per vehicle	23	34	50
Tramways	Miles of roadway	221.7	350.3	345.5
	Car miles (1,000,000s)	47.9	95.9	104.3
	Average scheduled speed (kilometres per hour)	NA	14.2	16.1
	Passengers (1,000,000s)	340.2	812.1	1,076.3
	Average fare per mile (1930 pence)	1.062	0.899	0.680
	Average seats per vehicle	38	67	67

Notes: The figures for the first column are from 1900, 1905, 1906, and 1907. See Ponsonby and Ruck (1930), p. 194 for details. Figures for fares are converted into 1930 prices using O'Donoghue, et al. (2004). The reported scheduled speed for the London Underground is for the Metropolitan and District Line.

Sources: Ponsonby and Ruck (1930), p. 194 and Munby (1978), p. 537.

Table 2. Summary Statistics on Commuting

	Mean (standard deviation)
Distance from home to nearest train station	0.64 (0.35)
Distance from home to nearest Underground station	1.17 (1.21)
Distance from home to nearest bus stop	0.20 (0.16)
Distance from home to nearest tram stop	0.35 (0.31)
Distance from workplace to nearest train station	0.60 (0.40)
Distance from workplace to nearest Underground station	0.95 (1.19)
Distance from workplace to nearest bus stop	0.16 (0.19)
Distance from workplace to nearest tram stop	0.39 (0.39)
Distance from home to work	3.21 (6.11)
10th percentile	0.40
Median	1.94
75th percentile	4.17
90th percentile	7.29
95th percentile	9.44
99th percentile	15.14
Heads of household	3.34 (7.44)
Others	3.08 (4.35)
Male	3.42 (6.91)
Female	2.74 (3.68)
Skill Category = professional	7.65 (10.25)
Skill Category = middling	4.43 (6.75)
Skill Category = skilled	3.61 (6.76)
Skill Category = semi-skilled	2.75 (5.97)
Skill Category = unskilled	2.84 (4.66)
Direction of Commute	
Commutes inwards	37.9%
Commutes outwards	16.0%
Works locally, does not commute	29.4%
Commutes across	16.7%
Distance from Nearest Centre	
Living < 1km	0.9%
Living < 2km	11.2%
Living < 5km	50.0%
Working < 1 km	12.7%
Working < 2km	30.7%
Working < 5km	63.3%

Notes: Standard deviation are shown in parentheses for continuous variables. Sample sizes: 49,361 with an earner number (used to estimate distances from home to public transportation), 34,972 with an identifiable workplace (used for all other figures). Distances from workplace to public transport is only reported for workers employed in the *New Survey* area.

Sources: Johnson, et al. (1999) and authors' calculations.

Table 3: Effects of Distance on Labour Force Participation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Hours>0	Has earner num.	Dist. to Work	Dist. to Work	Trans. Costs=0	Dist. work<1 Km	Dist. work> 3.2Km
Estimation technique	Probit	Probit	OLS	Heckman	Probit	Probit	Probit
Distance-Centre	-0.007* (0.002)	-0.009* (0.002)	0.459* (0.031)	0.470* (0.038)	-0.060* (0.005)	-0.062* (0.004)	0.066* (0.004)
Distance-Train	0.010 (0.006)	0.001 (0.006)	-0.165 (0.108)	-0.167 (0.109)	0.032* (0.013)	0.013 (0.012)	-0.002 (0.011)
Distance-Underground	0.005 (0.004)	-0.003 (0.004)	-0.201* (0.070)	-0.192* (0.068)	0.052* (0.009)	0.057* (0.007)	-0.038* (0.007)
Distance-Tram	-0.007 (0.008)	-0.003 (0.007)	0.123 (0.145)	0.128 (0.136)	-0.016 (0.017)	-0.062* (0.015)	0.001 (0.014)
Distance-Bus	-0.007 (0.013)	-0.016 (0.012)	-0.028 (0.235)	-0.007 (0.224)	0.037 (0.028)	-0.027 (0.024)	-0.030 (0.023)
Observations	41,319	41,319	16,910	41,319	14,212	16,901	16,901
R-squared	0.320	0.352	0.069		0.057	0.048	0.082
Sample mean	0.399	0.449	3.067	3.067	0.395	0.290	0.329
χ^2	11,568.6*	12,086.2*		1,250.3*	1,206.3*	864.5*	1,613.4*
F			36.1*				

Notes: The sample consists of individuals aged 14 or more who were related to the head of household. Robust standard errors in brackets. * = significant at a 5% level. Marginal effects evaluated at sample means reported for probit estimates. Pseudo-R^2 reported for all probits. The Heckman regression in column 4 uses the probit on reporting positive hours in the previous week (column 1) as the first stage. See Appendix IV, Tables A.IV.1 and A.IV.2 for additional results.

Sources: Johnson, et al. (1999) and authors' calculations.

Table 4. Effects of Distance on Log Weekly Earnings

	(1) Related to Head: OLS	(2) Related to Head: Heckman	(3) Related to Head & Age<25	(4) Children of Head & Age<25	(5) All	(6) All: HH Fixed Effects
Distance to work	2.577* (0.233)	2.574* (0.231)	2.144* (0.244)	2.017* (0.247)	2.185* (0.167)	2.134* (0.229)
Distance squared	-0.0489* (0.0138)	-0.0490* (0.0137)	-0.0486* (0.0148)	-0.0448* (0.015)	-0.0605* (0.0113)	-0.0551* (0.0125)
Home Centrality	-0.062 (0.355)	-0.066 (0.354)	-0.335 (0.351)	-0.156 (0.357)	1.105* (0.236)	
Home-Train	1.659 (0.978)	1.510 (0.977)	2.133* (0.957)	2.601* (0.976)	0.736 (0.644)	
Home-Underground	0.768 (0.614)	0.763 (0.612)	0.701 (0.615)	0.792 (0.620)	-0.221 (0.384)	
Home-Tram	2.405 (1.255)	2.380 (1.250)	1.098 (1.222)	1.505 (1.236)	1.380 (0.773)	
Home-Bus	-2.882 (2.082)	-2.826 (2.074)	-1.687 (2.050)	-1.995 (2.074)	-1.620 (1.261)	
Work Centrality	-0.139 (0.348)	-0.136 (0.347)	-0.213 (0.341)	-0.040 (0.345)	0.194 (0.230)	0.242 (0.369)
Work-Train	-0.533 (0.943)	-0.529 (0.938)	-1.459 (0.908)	-1.171 (0.914)	0.513 (0.566)	-0.111 (0.911)
Work-Underground	0.929 (0.483)	0.930 (0.481)	0.955 (0.539)	0.961 (0.543)	0.363 (0.298)	-0.025 (0.558)
Work-Tram	0.773 (0.804)	0.774 (0.800)	0.444 (0.832)	0.213 (0.861)	0.552 (0.493)	-0.962 (0.816)
Work-Bus	0.229 (1.120)	0.227 (1.112)	0.980 (1.302)	0.716 (1.296)	0.241 (0.707)	3.887* (1.195)
Observations	15,436	15,436	11,997	11,354	31,668	31,668
R-squared	0.572	.	0.632	0.645	0.687	0.891
Sample Mean	10.359	10.359	10.273	10.271	10.756	10.756
F statistic	173.123*	.	.	.	601.849*	370.387*

Notes: The dependent variable is the natural log of earnings (in hundredths of pence) in the previous week. All regression coefficients and standard errors multiplied by 100. The first stage controls in the Heckman regressions are as in Table 3, column 1. Household characteristics are omitted in the household fixed effects specification. Robust standard errors in brackets. * indicates significance at a 5% level. See Appendix IV, Tables A.IV.3, A.IV.4, A.IV.5 for additional results.

Sources: Johnson, et al. (1999) and authors' calculations.

Table 5. Ordinary Least Squares and Instrumental Variable Estimates of the Wage Returns to Distance

	OLS							IV					
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)	(10)	(11)	(12)
Sample dist. from centre	< 6k	< 8k	< 10k	< 15k	< 30k	All		< 6k	< 8k	< 10k	< 15k	< 30k	All
Distance to work	1.54* (0.14)	1.52* (0.13)	1.39* (0.12)	1.34* (0.10)	1.33* (0.10)	1.34* (0.10)		3.87* (1.69)	3.61* (1.10)	3.33* (0.80)	3.84* (0.69)	4.33* (0.80)	3.55 (2.38)
Observations	12,430	15,131	17,035	18,233	18,581	19,680		12,430	15,131	17,035	18,233	18,581	19,680
R-squared	0.62	0.62	0.62	0.62	0.62	0.62		0.61	0.61	0.61	0.61	0.61	0.61
1st Stage F Statistic								4.75*	12.64*	14.80*	16.68*	17.05*	9.24*
F statistic	303.4*	367.2*	420.1*	452.2*	458.1*	477.1*							
χ^2								21,796.0*	26,605.0*	30,411.1*	32,201.9*	32,073.9*	34,202.5*

Notes: The sample includes everyone in paid work aged 14 and over who 1) reported earnings in the previous week and 2) had an identifiable birthplace. The dependent variable is the natural log of earnings (in hundredths of pence) in the previous week. All regression coefficients and standard errors multiplied by 100. Robust standard errors in brackets in OLS and IV regressions. * indicates significance at 5% level. The first stage F-statistic is for the relevance of the instrument net of controls. Standard errors o the first stage clustered by birthplace. Full regression results in Table A.IV.6.

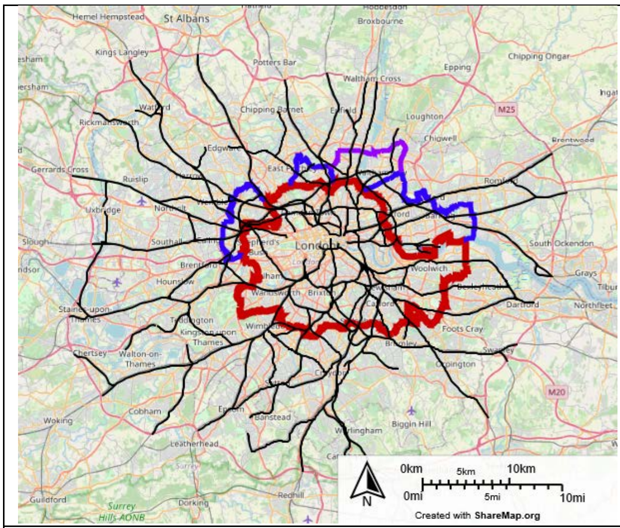
Table 6. Costs and Returns of Commuting

Crow-flies dist. (mtrs)	Trans. mode	Dist. walked	Dist. public trans.	Walk time (mins)	Public trans. time (mins)	Implied return 10th percentile	Implied return 25th percentile	implied return (median)	Monetary cost (d)	implied time cost, 10th percentile (d)	implied time cost, median (d)
500	Walk	625	0	9.38	0	2.3 (2.9)	3.9 (4.8)	7.5 (9.8)	0	3.5-7.0	11.4-22.8
1,000	Walk	1,250	0	18.75	0	4.6 (5.8)	7.7 (9.6)	14.9 (19.8)	0	7.0-14.1	22.8-45.6
2,000	Bus	300	2,200	4.5	8.62	9.2 (11.7)	15.3 (19.5)	29.7 (40.2)	6-12	6.5-13.0	21.1-42.2
2,000	Tram	300	2,200	4.5	8.20	9.2 (11.7)	15.3 (19.5)	29.7 (40.2)	6-12	6.4-12.7	20.6-41.2
4,000	UG	600	4,400	9	7.83	18.0 (24.2)	30.0 (40.3)	58.3 (83.1)	12-24	9.1-18.1	29.3-58.6
8,000	UG	600	8,800	9	15.66	34.4 (51.6)	57.3 (86.0)	111.5 (178.1)	18-30	11.7-23.4	38.0-76.0
8,000	Train	600	8,800	9	15.08	34.4 (51.6)	57.3 (86.0)	111.5 (178.1)	18-30	11.5-23.0	37.3-74.6
16,000	Train	600	17,600	9	30.17	59.9 (118.0)	99.8 (196.6)	194.0 (410.7)	30-36	16.7-33.3	54.0-108.0

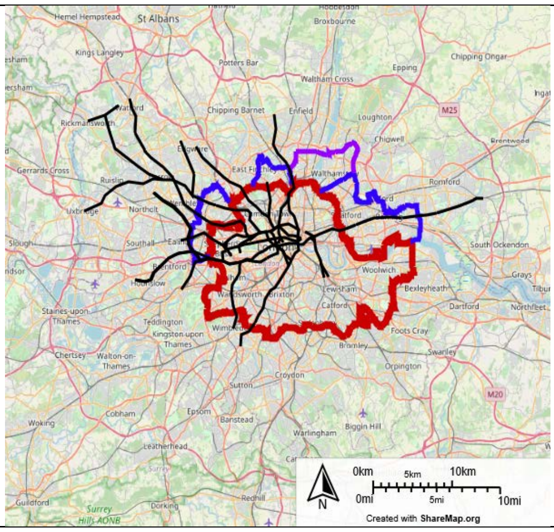
Notes: The tenth, twenty fifth, and fiftieth percentile of weekly earnings were 180d, 300d, and 583.5d, respectively. Implied returns are estimated using Table 4, column 1 and (Table 5, column 9).

Sources: Ponsonby and Ruck (1930); Johnson, et al. (1999); and authors' calculations.

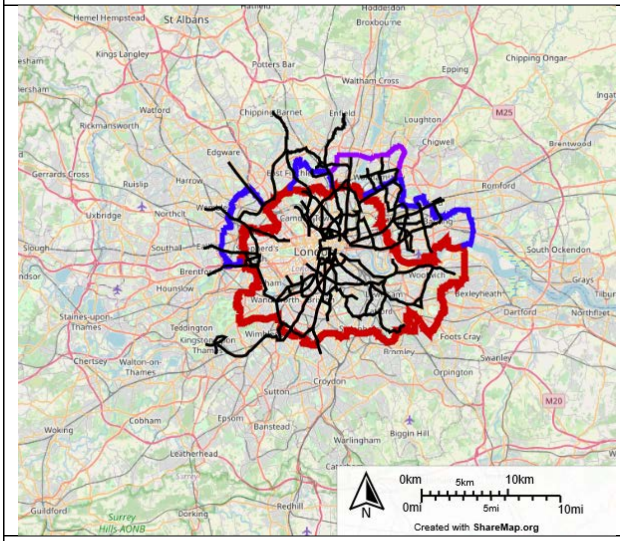
Figure 1. Public Transport Networks, Circa 1931



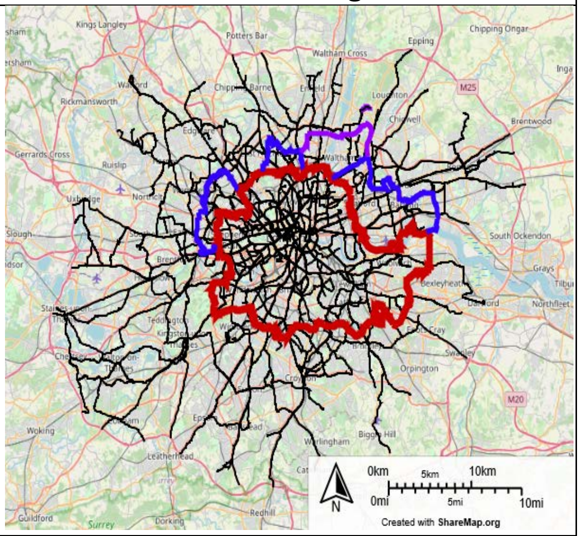
A. Rail



B. London Underground



C. Trams

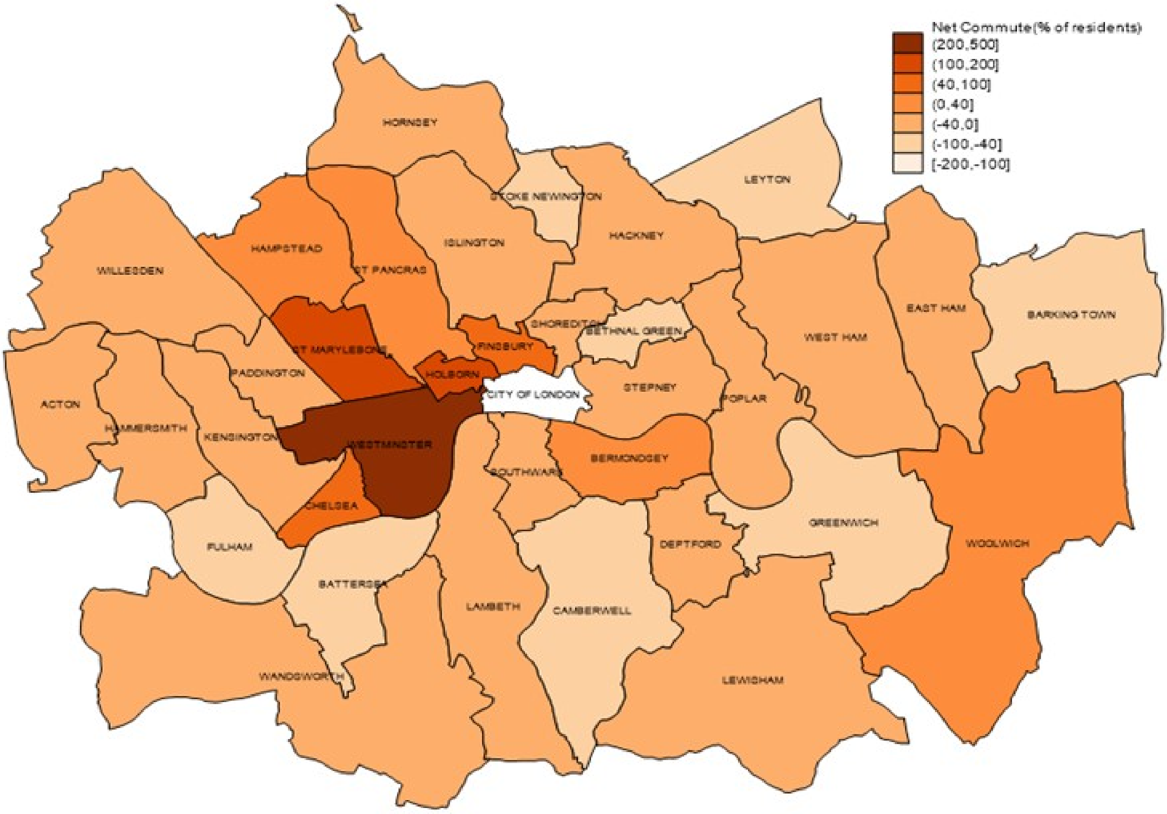


D. Buses

Notes: The maps show the transport networks (circa 1931) within the Greater London area (inside the modern ring road, the M25). The borders of the County of London and the *New Survey* catchment area are shown to provide scale. Scalable versions of these maps can be found at Seltzer (2020a, 2020b, 2020c, 2020d).

Sources: National Library of Scotland, Map Images, https://maps.nls.uk/geo/explore/side-by-side; Omnibus Society, *Motor Omnibus Routes in London*, vol 10a, 1930-31; http://sharemap.org/public/Trams_in_London#!webgl; https://londonist.com/2016/05/the-history-of-the-tube-map.

Figure 2. Net Commuting Flows by Borough of Residence



Notes: The map shows the net commuter flow by borough:

$$\left(\frac{100 * (individuals\ employed\ in\ borough\ \ i - employees\ residing\ in\ borough\ i)}{employees\ residing\ in\ borough\ i}\right)$$

The City of London is blank on the map as its residents were not included in the *New Survey*.

Sources: Johnson, et al. (1999) and authors' calculations.

Appendix I: Information in the *New Survey of London Life and Labour*

Front of Card

Name (Not recorded in Johnson, et al. (1999))
Address (street address)
Borough
For each wage earner
- Relationship to head of household
- Age
- Occupation – see Johnson, et al. (2002) for a description
- Employer
- Place of Work
- Cost of transport weekly
- Hours last week
- Hours full time
- Earnings last week (in s. and d.)
- Earnings full time (in s. and d.)
- State Insurance deductions (in s. and d.)

For each non-wage earner
- Sex
- Age
- Relationship to head of household

Income from other sources
- Source
- Amount (in s. and d.)

Date of Interview
Interviewer

Back of Card

Birthplace of Adults
Rent, weekly (includes rates)

Persons and accommodation
- No. of persons
- No. of bedrooms

Parlour (yes or no)
Kitchen (number)
Scullery (yes or no)
Pantry or Larder (yes or no)
Bath (yes or no)
Yard (yes or no)
Garden (yes or no)
Allotment (yes or no)
Remarks on accommodation

References

Johnson, Paul, Bailey, Roy, Baines, Dudley, Raspin, Anna, Hatton, Timothy (1999). New Survey of London Life and Labour, 1929-1931. [data collection]. 2nd Edition. UK Data Service. SN: 3758, http://doi.org/10.5255/UKDA-SN-3758-1. https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=3758 [Accessed 25/9/2020].

Bailey, Roy and Leith, Anna (2002). “New Survey of London Life and Labour: codebook for the data files.” (Release 2.0), unpublished manuscript.

Table A.I.1
Summary Statistics

A. Individual Characteristics

	All Individuals	Employed in the Previous Week
Male (%)	48.7	70.5
Age	28.4 (19.8)	32.6 (14.4)
Born outside England (%)	2.7	2.5
Born in London (%)	84.2	86.2
Living in borough of birth (%)	33.3	36.6
Living in borough adjacent to borough of birth (%)	11.8	11.4
N	93,891	35,282

B. Household/Dwelling Characteristics

Persons in the Household	3.5 (1.9)
One-person household (%)	11.0
Two-person household (%)	24.6
Three-person household (%)	22.5
Four-person household (%)	17.2
Five + person household (%)	24.7
Non-wage income (d per week)	0.38 (0.65)
Minimum distance to a central point (km)	5.4 (2.8)
Nearest central point = Charing Cross (%)	42.9
Inner ring (%)	30.1
Middle ring (%)	20.6
Exterior ring (%)	49.3
County of London (%)	81.8
N	26,915

Table A.I.1
Summary Statistics, Continued

C. Employment Characteristics

Hours working in previous week	46.1 (8.3)
% age 16-65 reporting hours in previous week >0	54.9
Worked at least 40 hours (%)	90.8
Worked at least 48 hours (%)	57.1
Hours worked in a full-time week	46.9 (7.2)
% age 16-65 reporting hours in a full time week >0	56.1
Earnings in previous week (£)	2.34 (1.17)
Earnings in a full-time week (£)	2.38 (1.17)
Occupation = metal worker (%)	8.0
Occupation = electrical (%)	2.4
Occupation = makers of textile goods (%)	8.5
Occupation = food, drinks & tobacco (%)	3.1
Occupation = wood & furniture (%)	4.9
Occupation = printers & photographers (%)	2.5
Occupation = building trades (%)	4.4
Occupation = painters & decorators (%)	2.8
Occupation = transport and communications (%)	17.7
Occupation = Commerce, finance, insurance (%)	8.6
Occupation = personal service (%)	11.4
Occupation = clerk (%)	5.5
Occupation = warehousemen, storekeepers (%)	4.7
Occupation = other or unknown (%)	15.5
Armstrong skill category = professional (%)	0.1
Armstrong skill category = middling (%)	0.7
Armstrong skill category = skilled (%)	49.0
Armstrong skill category = semi-skilled (%)	24.0
Armstrong skill category = unskilled (%)	26.1
N	35,353

Notes: Standard errors are reported in parentheses for continuous variables. Age is reported as zero in 9.2 percent of observations. We have excluded these observations from our calculations in Table A.I.1 and included a dummy variable for the regressions reported in Tables 3 and 4 and Appendix IV.

Sources: Johnson, et al. (1999) and authors' calculations.

Appendix II: Coding GIS Coordinates and Distances

This appendix outlines our approach to GIS coding of home street addresses, workplace locations, and the public transport network. We also outline how we use these GIS coordinates to construct our measures of distance commuted, access to public transport, and centrality. Finally, we examine measurement error and biases that are likely to result from our approach and the likely implications for our results.

I. Residential Addresses

Our approach to GIS coding home addresses is as follows. First, we have entered the street name into streetmap.co.uk. If there exists exactly one modern street with the same name that is located within the historic Metropolitan Borough listed on the original record we assumed this to be the residential address. We then took the GIS coordinates from streetmap.co.uk. Occasionally there are multiple streets of the same name within the same historic borough, e.g. two towns have a "High St.". Normally in these cases, the records themselves indicate the correct street. For example, the streets may be listed as "High St., Woolwich" or "High St., Plumstead". In cases such as this, the street is clearly identified even though the official name of both streets is just "High St." and both Woolwich and Plumstead were in the Metropolitan Borough of Woolwich. In the small handful of cases where there remained ambiguity, we have looked at additional information from the record cards to determine the most likely correct address (e.g. whether household members were working in Plumstead or Woolwich).

For about a quarter of the observations, we were unable to find the address listed on the record card in streetmap.co.uk due to changes of street names. London was extensively bombed during the Second World War. Many homes and even entire neighbourhoods were damaged beyond repair (Ward 2015). After the War, London's urban planners "cleared" many War-damaged areas and other urban slums. The clearances disproportionately affected working-class areas, as wealthier areas which suffered minor bomb damage were quickly rebuilt. The clearances often changed the physical layout of the area, for example replacing low-rise dwellings with high-rise council housing (Sturm and Redding 2016).

To locate no-longer-extent streets, we began by searching the online indexed maps from the LLPL (London School of Economics 2020). Because of the similarity of coverage between the two surveys, most residential streets within the County of London appearing in the NSLLL previously appeared in the LLPL. We were thus normally able to find residential streets on the LLPL map and obtain GIS coordinates from Ordinance Survey maps (National Library of Scotland 2020). In cases where a street was not included in the LLPL index (particularly in the outer boroughs not surveyed in the LLPL), we have searched other on-line resources such as *Medical Officer of Health Reports*, *Census Street Index*, and various genealogical web sites (Welcome Library 2021; Family Search 2021). Often, we were able to find an exact or at least an approximate location for a residential street, usually based on known locations for nearby streets that were listed in the same source. Once we found the location, we then found the GIS coordinates using Ordinance Survey maps (National Library of Scotland 2020).[39]

When entering GIS coordinates, we have used a single centroid for each home address in the data.[40] The centroid will either be the location used by Streetmap.co.uk (for still-extent streets) or at approximately the middle of the street (for no-longer-extent streets). The approach of using a single centroid facilitates checking for inconsistencies in the data. It also makes it possible to replicate our GIS coding procedure, as it avoids non-replicable *ad hoc* assumptions about individual locations.

We believe that home addresses were very accurately recorded by the NSLLL enumerators. The enumerators were instructed to visit each individual household; hence they actually set foot on the residential street. We have been able to locate well over 99 percent of home addresses in the data. Although our use of a single centroid for each address will inevitably create some measurement error, we believe that this measurement error is likely to be small because working-class residential streets tended to be fairly short.

II. Workplace addresses

Our approach to obtaining GIS coordinates for workplaces closely follows that of home addresses. We searched streetmap.co.uk, followed by the indexed LLPL maps, followed by other on-line sources to identify workplace streets or place names. If this failed to produce a likely match, we searched old Ordinance Survey maps (National Library of Scotland 2020) for similarly named streets nearby the place of residence.[41]

There are several sources of measurement error for workplaces that are not present for residences. First, the survey question for workplace is less precise than residence, asking for “place of work” rather than “address”. Accordingly, the responses were more varied than for home addresses, ranging from an exact address, to just a street name, to a broader place name. About half the responses are place names. Even place names can be imprecise. For example, “Greenwich” is the name of both a Metropolitan Borough and the main town within the borough. Where a place name is reported, we enter a common GIS coordinate using the centroid of the smallest plausible geographic unit (e.g. town, rather than borough). If the record card provides a general, but very broad area (e.g. “London” or “East End”), we treated place of work as missing unless we could identify a more specific location based on the employer. The lack of precision in workplace locations implies that our GIS coordinates are inherently subject to more measurement error than for home addresses. In addition, commercial streets

[39] National Library of Scotland (2020) contains a variety of scalable modern and historic street maps which can be uploaded side-by-side. Our GIS coding has primarily relied on OpenStreetMap; Ordinance Survey (OS), 25 inches, 1892-1914; and OS 1/2500, 1944-1967.

[40] We have used the street/borough pair when assigning centroids. Most streets were entirely contained within a single borough and we assigned a single set of GIS coordinates to these streets. In cases were a street passed through multiple boroughs, we assigned a single centroid for each borough covered by the street. Street names tended to change at borough boundaries, so there are very few residential streets with multiple coordinates in the data.

[41] Since enumerators relied on residents to provide workplace addresses, there were fairly frequent transcription errors or spelling mistakes on the original records. We were often able to find very similar (or identically pronounced) workplace street names by searching the map around the home address.

tended to be longer than residential streets and thus there is likely to be more measurement error for workplaces than residences, even in cases where a street is listed for both. A final issue is that in about eight percent of observations with pay reported, the place of work is listed as "various" or "casual". We assume these workers to either be footloose (such as in the building trades) or itinerant. We do not assign workplace locations to these workers.

A second difficulty identifying workplaces is that, unlike the home addresses, the original record cards do not contain boroughs for workplace.[42] This makes it more difficult to identify the workplace location for common London street names, unless the location is given on the original record card, e.g. "High St., Plumstead". In cases where a workplace address was ambiguous, we used other data from the record such as home address, name of employer, and travel costs to identify the most plausible location.[43]

A final issue results from the fact that enumerators never visited places of work, instead they relied on information supplied by interviewees. At best, this meant that the address was from second-hand information from the worker, rather than directly from the enumerator. However, it is likely that the information was often supplied by another member of the household. Although the *New Survey* enumerators were explicitly instructed to make repeated visits to households in order to get employment information from the income earner themselves, it is known that Arthur Bowley, the overseer of the NSLLL, was willing to "sacrifice accuracy to speed and simplicity" (Abernathy 2017; Hennock 1991).[44] Responses from someone other than individual workers themselves were probably widely tolerated.[45]

Although the issues raised above imply that workplace addresses are more prone to measurement error than home addresses, we believe that they are nevertheless fairly accurate. We were able to obtain GIS coordinates for about 98 percent of observations where a street or place name is given for place of work. When we were unable to obtain workplace GIS coordinates, it was typically because either the worker was itinerant or the original respondent did not supply the necessary information. In about 12 percent of observations reporting

[42] The Johnson, et al. (1999) data contains a variable for workplace borough. However, this has been constructed by the researchers, not transcribed from the original record cards. We believe that there are numerous coding errors in this variable for observations in which only a street address is given for place of work. We have created a new variable for workplace borough by mapping the NSLLL area into approximately 500 square meter grids. We map each observation into a cell using the GIS coordinates. We then map the grid cells into boroughs. In cases where a grid cell is divided between more than one borough, we mapped workplace addresses within the cell into boroughs by hand using Ordinance Survey maps (OS 25 inches, 1892-1914).

[43] Locations of larger employers often turned up in on-line searches, and we were often able to identify precise GIS coordinates using this information.

[44] It was noted in the original instructions to enumerators that "Vague estimates of husband's earnings by wife, of child's by parent, or of lodger's by landlady, should not be entered until an effort has been made to see the wage earner concerned" (*New Survey*, instructions issued to investigators, quoted in Abernathy (2017).

[45] Missing or imprecise workplace information is much more common in the NSLLL data for lodgers than family members. It is difficult to reconcile this with earners supplying their own information, but consistent with a single (non-working) resident supplying information for all household members. In addition, the original record cards often provide relationships within households vis-à-vis someone other than the likely head of household. Johnson, et al. (1999) have reclassified household relationships using the age, gender, and earnings data to identify the head. The fact that the head of household was frequently not correctly identified by the enumerators is strongly suggestive that they made a single visit to the household and collected all information from the person who answered the door.

earnings in the previous week the information is either missing or unusable ("X", "refused", "London", etc.).

III. Public transport

We have compiled a list of railway and London Underground stations in 1929 using historic Underground maps (Graham-Smith 2018), historic Ordinance Survey maps (National Library of Scotland 2020), and Wikipedia lists of current and historic stations (Wikipedia 2020a, 2020b, 2020c, 2020d). Wikipedia usually provides GIS coordinates for rail and Underground stations and we cross-checked these using historical Ordinance Survey maps (National Library of Scotland 2020). We thus believe that these GIS coordinates are very accurate.

We obtained detailed information on bus and tram routes from London Historical Research Group (2014) and Public (2020), respectively. However, neither source indicates where vehicles stopped along the route. We have used OpenStreetMap to find the location of modern bus stops and assumed that these correspond to stops on the historic routes. Where the historic routes do not coincide with modern routes, we have assumed that stops were 300-500 meters apart in central areas and slightly further apart in outer areas, as with modern routes. We start with a known stop on route, such as the route terminus or a railway station, and assign stops approximately equidistant from this point. In addition to official stops, it was generally possible for able-bodied passengers to board or leave a bus or tram at any point where the vehicle was stopped and thus we also classify major intersections as stops.[46] There will be some measurement error in this approach that is absent in our calculations for rail and Underground (for which we know the exact location for each station), but this is likely to be fairly small, as tram and bus stops were generally fairly close together.

IV. Calculating Distances

We have used the GIS coordinates to calculate crow-flies distances between home, work, the city centres, and public transport for each employed individual in the sample. Conceptually, these distances are 1) the distance commuted (home to work), 2) the centrality of their home or workplace (minimum distance to Charing Cross or the Bank of England), and 3) access to public transport (distance to bus, tram, Underground, or train). Figure A.II.1 shows these distances for one individual. On the map, the residence is denoted H, the workplace is denoted W, the nearest Underground stop to home is denoted U, and Charing Cross is denoted CX. The black line shows the crow-flies distance between home and work (commuting distance). The green line shows the distance between home and Charing Cross (centrality). The solid red line between H and U shows the distance from home to the nearest underground station. The dotted and solid red lines connecting H to W show the most plausible transport route to work (by underground for two stops and a short walk to the workplace at the end). For ease of exposition, we have not shown the other distances on this map.

To calculate distances, we use the great circle distance formula:

[46] The buses and trams of the 1930s were "routemaster" design with an open entrance at the back. Passengers could embark or disembark at any point on route when the bus was stopped.

$$d = R*acos(sin(lat\_a)*sin(lat\_b) + cos(lat\_a)*cos(lat\_b)*cos(d\_lon));$$

where: R = radius of the earth (6365 kilometres); lat_a, lat_b = latitudes of points a and b; d_lon = difference in longitude between the two points.

There were a few exceptions to these principles in our calculations of distance. As mentioned above, if a worker was deemed to be itinerant, we did not fix workplace coordinates, and thus could not calculate distances. In addition, if the original record card listed workplace as some variant of "local" or "nearby", we assume a commuting distance of 0.5 kilometres and that workplace centrality is the same as home centrality. These cases account for approximately 2.4 percent of observations.

There are also observations for which the distances between home and workplace were very large. It is likely that these workers were stationed remotely and did not commute on a daily basis. Thus, in our main analysis, we exclude the 76 observations where the distance between home and work was over 50 kilometres. We have also set the cut-off at 20 kilometres (which excludes an additional 117 observations) and included long commutes in the analysis as robustness checks. Our results are not particularly sensitive to the rule used for exclusion of outliers.

V. Measurement error and bias

For the most part, errors in our distance variables will simply be classical measurement error, resulting from our inability to identify precise locations for home, workplace, bus stops, and tram stops. Our assumption that residences and workplaces are at the centre of the street (or broader location) implies that, on average, our GIS coordinates will be very close to correct, but there will be a variance around the point estimate. As noted above, the extent of this measurement error is likely to be larger for workplaces than for residences or public transport.

In addition, there exist two likely sources of systematic bias in our variable for distance commuted. First, we are less likely to find workplace locations for individuals who had longer commutes. When we could not find a workplace through other means, our final approach was to search the map in proximity to the worker's residence. This approach helped locate numerous workplaces, but it also implies that we are more likely to be missing data for workplace location if a worker's residence was further from home, and thus had a longer commute. This will bias the estimated travel distances in Section VI downwards. However, we do not feel that this bias is likely to be large because we have been able to locate all but about two percent of workplaces named on the original record cards.

A more serious potential concern results from our use of a single centroid for each residential street and workplace location. As noted above, the use of a single centroid is likely to result only in measurement error for residential addresses or workplace locations, *taken* *individually*. However, in this context the location of an individual's workplace is not independent of their residence. For example, a worker who resides on a short north-south street that intersects a lengthy east-west workplace street near its eastern end is more likely to work around the corner at the eastern end of the workplace street than a couple kilometres away at the centroid of the street. Similarly, a worker whose workplace is reported as a borough adjacent to their borough of residence is more likely to be employed near the border of the two boroughs than at the centroid of the workplace borough. Consequently, it is possible that overestimate the distance commuted. In this case the measurement is not classical but instead the covariance between the measurement error and (true) distance commuted is positive. A positive covariance generally reinforces the attenuation bias stemming from classical measurement error (Bound, Brown and Mathiowetz 2001).

Taken together, the likely consequence of these types of measurement error will be attenuation bias in our estimates of the effects of distances on earnings. In other words, our estimated returns to commuting in Section VII are likely to be biased downwards. The extent of this bias is likely to be lower in the IV estimates of the returns to commuting (Table 5) than the OLS estimates (Table 4) because the birthplace location and nearest city centre are much more likely to be independent of each other than residence and workplace. On the other hand, there is likely to be little bias in our estimated returns to access to public transport, as these GIS coordinates are measured with little error. As a result, we think our results in the OLS regressions in Tables 4 and Appendix IV represent a lower bound on the returns to distance commuted but are fairly accurate on the returns to access to public transport.

References

Abernethy, Simon T. (2017). "Deceptive data? The role of the investigators in the New Survey of London Life and Labour 1928–32." Historical Methods A Journal of Quantitative and Interdisciplinary History, 50(1):1-15.

Bound, John, Brown, Charles, Mathiowetz (2001). "Measurement error in survey data." In: James Heckman and Edward Leamer, Eds. *Handbook of Econometrics*, Volume 5, Chapter 59, pp. 3705-843.

Family Search (2021). *Census Street Index*. https://www.familysearch.org/wiki/en/Census_Street_Indexes [Accessed 25/9/2020].

Graham-Smith, Darien (2018). "The history of the tube map". *Londonist*. https://londonist.com/2016/05/the-history-of-the-tube-map [Accessed 25/9/2020].

Hennock, E. P. (1991). "Concepts of poverty in the British social surveys from Charles Booth to Arthur Bowley." In: M. Bulmer, K. Bales, and K. K. Sklar, eds. *The Social Survey in Historical Perspective*. (Cambridge: Cambridge University Press), pp. 189-217.

Johnson, Paul A., Bailey, Roy E., Baines, Dudley E., Raspin, Anna, Hatton, Timothy J. (1999). *New Survey of London Life and Labour, 1929-1931*. [data collection]. 2nd Edition. UK Data Service. SN: 3758, http://doi.org/10.5255/UKDA-SN-3758-1. https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=3758 [Accessed 25/9/2020].

London Historical Research Group, The Omnibus Society (2014): *Motor Omnibus Routes in London, vol. 10A (1st January 1931 to 31st December 1932)* (Sevenoaks: London Historical Research Group).

London School of Economics (2020). *Charles Booth's London: Poverty Maps and Police Notebooks*. https://booth.lse.ac.uk/map/14/-0.1174/51.5064/100/0 [accessed 13/9/2020].

National Library of Scotland (2020). *Side-by-Side Map Images*. https://maps.nls.uk/ [Accessed 25/9/2020].

Public (2020). "Trams in London." http://sharemap.org/public/Trams_in_London#!webgl [Accessed 25/9/2020].

Redding, Stephen and Sturm, Daniel (2016). "Estimating neighbourhood effects: evidence from War-time destruction in London." Mimeo, London School of Economics.

Streetmap (2020). https://www.streetmap.co.uk/ [accessed 25/9/2020].

Ward, Laurence (2018). *The London County Council Bomb Damage Maps, 1939-1945*. (London: Thames & Hudson).

Welcome Library (2021). *The Medical Officer of Health Reports*. https://wellcomelibrary.org/moh/about-the-reports/about-the-medical-officer-of-health-reports/ [accessed 25/9/2020].

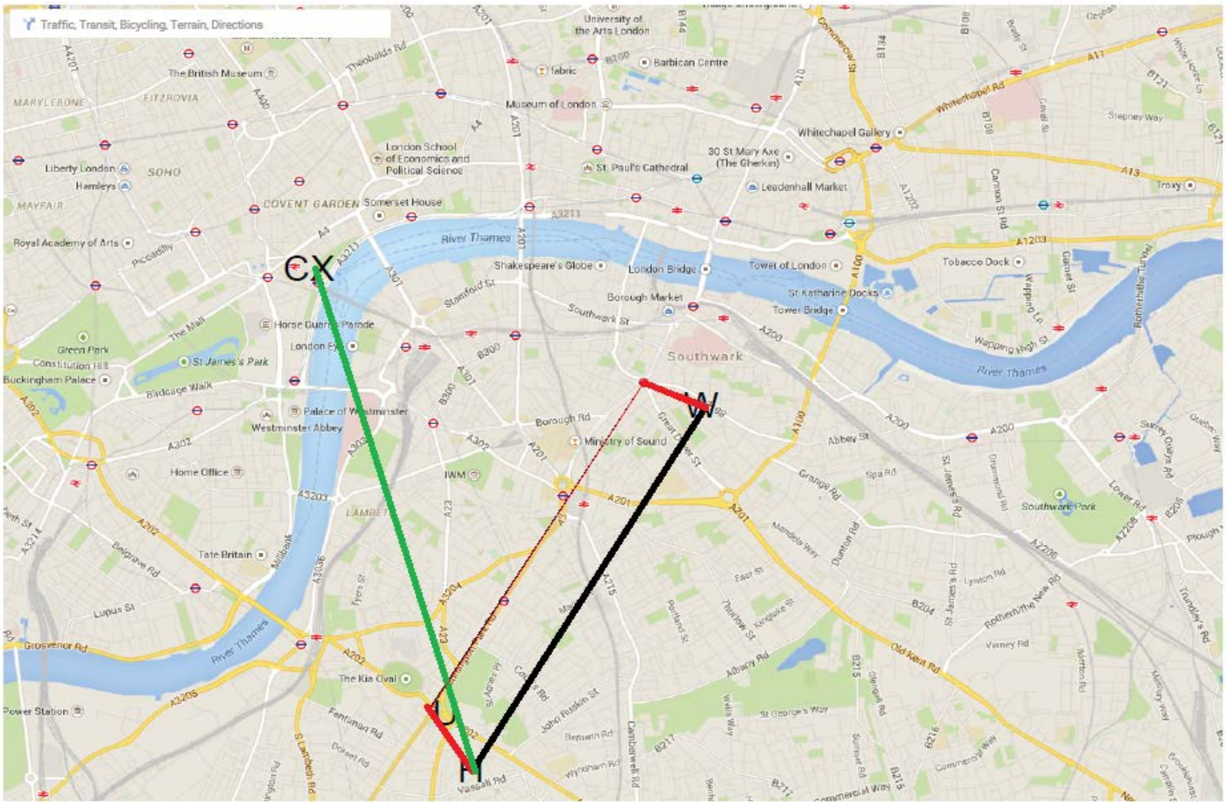
Wikipedia (2020a). "List of London railway stations." https://en.wikipedia.org/wiki/List_of_London_railway_stations [accessed 25/9/2020].

Wikipedia (2020b). "List of closed railway stations in London." https://en.wikipedia.org/wiki/List_of_closed_railway_stations_in_London [accessed 25/9/2020].

Wikipedia (2020c). "List of London Underground stations." https://en.wikipedia.org/wiki/List_of_London_Underground_stations [accessed 25/9/2020].

Wikipedia (2020d). "List of former and unopened London Underground stations." https://en.wikipedia.org/wiki/List_of_former_and_unopened_London_Underground_stations [accessed 25/9/2020].

Figure A.II.1
Distances for One Individual



Appendix III: Monotonicity in the Instrumental Variable Regressions

The monotonicity assumption implies that commuting distance should be increasing with distance from birthplace to the nearest centre. Strict monotonicity implies this should hold for all employees in the sample. More realistically, a stochastic version of monotonicty may hold, if there are more "compliers" than "defiers" or if the expected value of distance commuted is increasing with the distance of birthplace to the nearest centre. We summarise the theory behind the monotonicity assumption and the relevant evidence in the *New Survey* data below.

There have been several recent attempts to address what instrumental variable regressions can estimate if the assumptions of strict monotonicity are violated. Small, et al. (2017), re-purposed the idea of "stochastic monotonicity" to show that if in each sub-group of the population having the same value of the outcome variable, there are more compliers than defiers, then 2SLS estimates a weighted difference of the average treatment effect, where the weights reflect the size of compliers and defiers in the population. Chaisemartin (2017) argues that something close to a treatment effect can be identified if there are at least as many compliers as defiers and that the "treatment" variable has the same effect on both groups. In our case, this would require that a larger distance commuted has the same effect on wages for both compliers and defiers. If so, then the 2SLS estimates will give the local average treatment effect of the remaining (excess) compliers in the sample. Dahl, et al. (2017) show that a local average treatment effect (LATE) for compliers (and defiers) can be estimated when there are only one or the other population in some range of the outcome variable for a given level of the instrument. Under this "local monotonicity" each LATE of the compliers is then locally identified in the range of the outcome variable where they are the only group.

Many of these results are, however, developed on the assumption of a binary endogenous variable and/or instrument. In our case, both endogenous variable (commute distance) and the instrument (distance from birthplace to the nearest centre) are continuous variables. As such any appeal to the weaker versions of montonicity above can only be suggestive. Counting the numbers of compliers and defiers is harder when the endogenous variable is continuous rather than discrete. In a review of the recent literature, Fiorini and Stevens (2021) recommend looking at "stochastic dominance" of the instrument first proposed by Angrist and Imbens (1995) as a necessary but not sufficient condition for monotonicity. The idea here is to look at the cumulative distribution function of the endogenous variable conditional on the level of the instrument. If monotonicity holds, the *expected value* of the distance commuted for someone with a birthplace distance of k+1 should be higher than the expected value of distance commuted for someone with a birthplace distance of k. Equivalently the (cumulative) probability of a given distance commuted should be lower among individuals with a lower value of the instrument (birthplace distance) than the (cumulative) probability of the distances commuted among those with a higher value of the instrument. With continuous variables this is a necessary but not sufficient test of deterministic monotonicity – since the continuous variables have to be drawn into discrete groups. A graph of distance commuted for each of the (banded) levels of the instrument is one way of assessing whether monotoncity is likely to hold. This can bring both compliers and defiers into the same discrete group. Nevertheless, a graph that is suggestive of stochastic dominance is at least suggestive of (local) monotonicity of the instrument. If the cumulative distances commuted for the sample with birthplace to the centre distance k is always above the cumulative distances travelled for the sample born at distance K+1 from the centre, then this suggests monotonicity may hold.

The visual tests of stochastic dominance are given in Figure A.III.1. When the instrument is banded into discrete distance intervals, the figure suggests that stochastic dominance may hold for (banded) values of the instrument below 10 kilometres from the centre of London. At values of the instrument greater than 10 kilometres, stochastic dominance no longer holds. Some individuals born outside the centre of London may find work in more local labour markets, so that distance from birthplace to the centre of London is negatively correlated with distance to work. This suggests that we may be able to identify something close to a local causal response for a restricted range of the instrument. In practice, some 85 percent of the sample of wage earners was born within 10 kilometres of the London centroids.

If we use other distance metrics, for example distance from birthplace to the 8 major railway stations that lie on the outer fringes of central London, or distance from birthplace to 10 town halls in inner London, the monotonicity condition is much more obviously violated in stochastic dominance graphs. This indicates that many more people living in the outer suburbs commute than work locally.

References

Angrist, Joshua D. and Imbens, Guido W. (1995). "Two-stage least squares estimation of average causal effects in models with variable treatment intensity." Journal of the American Statistical Association, 90(430): 431-42.

Chaisemartin, Clément de (2017). "Tolerating defiance? Local average treatment effects without monotonicity." Quantitative Economics, 8: 367-96.

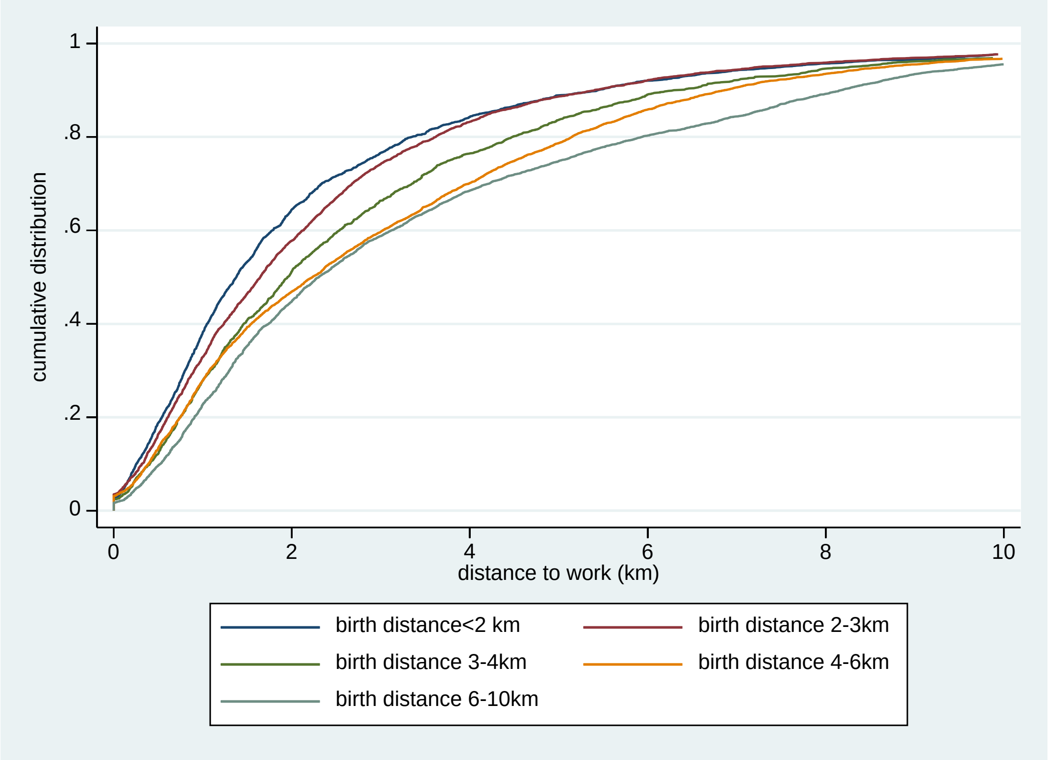
Dahl, Christian M., Huber, Martin & Mellace, Giovanni (2017). "It's never too LATE: a new look at local average treatment effects with or without defiers." *Discussion Papers on Economics* 2/2017, University of Southern Denmark, Department of Economics.

Fiorini, Mario and Stevens, Katrien (2021). "Scrutinizing the monotonicity assumption in IV and fuzzy RD designs." Oxford Bulletin of Economics and Statistics, 83(6): 1475-526.

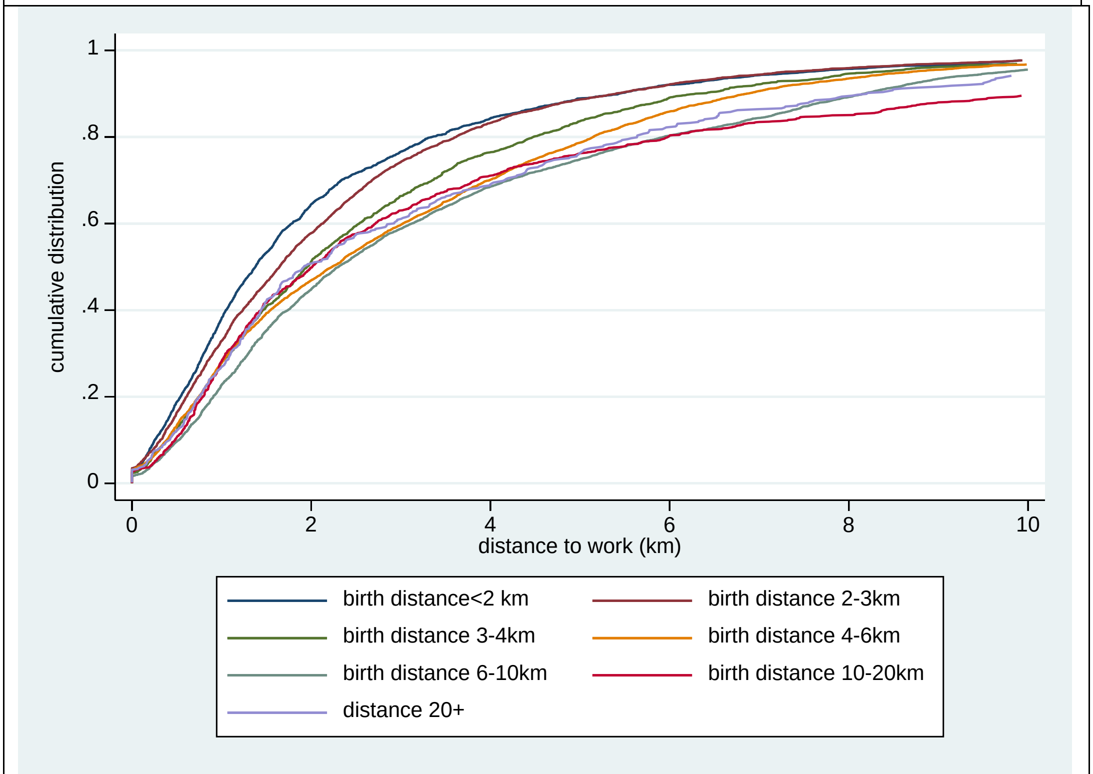
Small, Dylan S., Tan, Zhiqiang, Ramsahai, Roland R., Lorch, Scott A. and Brookhart, M. Alan (2017). "Instrumental variable estimation with a stochastic monotonicity assumption." Statistical Science, 32(4): 561–79.

Figure A.III.1. Stochastic Dominance Inspection Test: Distance to Work by Birthplace Distance Groups

Panel A



Panel B



Appendix IV: Additional Tables and Figures

Table A.IV.1: Estimated Effects of Control Variables on Labour Force Participation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hours>0	In Work	Distance	Distance (Heckman)	Transport Costs=0	Distance <1 Km	Distance >3.2 Km
Age	-1.558* (0.056)	-1.630* (0.053)	10.195* (1.478)	6.873* (1.771)	-1.130* (0.001)	-0.633* (0.122)	1.104* (0.127)
Age squared	0.009* (0.001)	0.009* (0.001)	-0.144* (0.020)	-0.129* (0.020)	0.170* (0.020)	0.011* (0.002)	-0.016* (0.002)
Age reported as zero	-0.599* (0.012)	-0.570* (0.011)	1.825* (0.276)	0.518 (0.484)	-0.131* (0.028)	-0.064* (0.024)	0.186* (0.024)
Female	-0.341* (0.003)	-0.380* (0.003)	-0.418* (0.062)	-1.041* (0.200)	-0.015* (0.007)	0.027* (0.006)	-0.046* (0.006)
Born in England	0.077* (0.007)	0.070* (0.007)	-0.176 (0.185)	-0.004 (0.152)	0.019 (0.014)	-0.001 (0.012)	-0.022 (0.012)
Born in London	0.188* (0.006)	0.192* (0.006)	-0.214 (0.139)	0.163 (0.159)	-0.009 (0.011)	0.001 (0.009)	-0.014 (0.009)
Born and lives in the same borough	-0.029* (0.005)	-0.037* (0.004)	-0.165* (0.081)	-0.187* (0.090)	0.031* (0.009)	0.010 (0.007)	-0.023* (0.008)
Borough is adjacent to birth borough	-0.055* (0.006)	-0.071* (0.006)	-0.209* (0.102)	-0.280* (0.138)	-0.004 (0.013)	-0.019 (0.011)	0.009 (0.012)
Other family pay	-0.000* (0.000)	-0.001* (0.000)					
Non-labour income household	-0.002* (0.000)	-0.002* (0.000)					
Observations	51,970	51,970	23,254	51,970	19,147	23,254	23,254

Notes: Coefficients on Age and Age squared are multiplied by 100. See Table 3 for additional results and notes.

Sources: Johnson, et al. (1999) and authors' calculations.

Table A.IV.2: Distance Effects on Labour Force Participation: All Family Members in a Household

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hours>0: Probit	In Work: Probit	Dist. to Work: OLS	Dist. to Work: Heckman	Trans. Costs=0: Probit	Dist. to Work<1Km: Probit	Dist. to Work>3.2Km: Probit
Distance-Centre	-0.001	-0.005*	0.441*	0.410*	-0.052*	-0.057*	0.060*
	(0.002)	(0.002)	(0.035)	(0.038)	(0.003)	(0.003)	(0.003)
Distance Home-Train	0.024*	0.007	-0.100	-0.072	0.011	-0.010	-0.008
	(0.005)	(0.005)	(0.105)	(0.109)	(0.009)	(0.008)	(0.008)
Distance Home-Underground	0.005	-0.003	-0.182*	-0.179*	0.039*	0.041*	-0.026*
	(0.003)	(0.003)	(0.080)	(0.067)	(0.006)	(0.005)	(0.005)
Distance Home-Tram	-0.007	-0.003	-0.063	-0.102	-0.007	-0.040*	0.001
	(0.006)	(0.006)	(0.129)	(0.133)	(0.012)	(0.010)	(0.010)
Distance Home-Bus	-0.004	-0.026*	-0.064	-0.060	-0.009	-0.043*	0.006
	(0.010)	(0.009)	(0.189)	(0.220)	(0.019)	(0.016)	(0.016)
Observations	67,584	67,584	34,342	67,584	28,880	34,294	34,294
R-squared	0.278	0.351	0.029		0.047	0.042	0.063
Sample mean	0.512	0.582	3.22	3.22	0.418	0.294	0.332

Notes: The sample in these regressions is the same as in Table 3, except heads of households and non-relatives are included. See other notes in Table 3.

Sources: Johnson, et al. (1999) and authors’ calculations.

Table A.IV.3 Estimated Effects of Control Variables on Earnings

	(1)	(2)	(3)	(4)	(5)	(6)
	Related to Head	Related to Head: Heckman	Related to Head: Age<25	Children of Head: Age<25	All	All: HH Fixed Effects
Age	0.146* (0.003)	0.146* (0.003)	0.491* (0.015)	0.495* (0.015)	0.098* (0.001)	0.110* (0.001)
Age^2	-0.002* (0.001)	-0.002* (0.001)	-0.010* (0.001)	-0.010* (0.001)	-0.001* (0.001)	-0.001* (0.001)
Missing Age	2.586* (0.039)	2.595* (0.038)	6.131* (0.140)	6.246* (0.143)	2.054* (0.022)	2.187* (0.030)
Male	0.266* (0.007)	0.266* (0.007)	0.211* (0.007)	0.202* (0.007)	0.321* (0.006)	0.272* (0.008)
Born in England	0.042* (0.016)	0.043* (0.016)	0.006 (0.018)	0.007 (0.019)	0.039* (0.008)	0.058* (0.014)
Born in London	-0.001 (0.010)	-0.001 (0.010)	0.016 (0.009)	0.018* (0.009)	0.003 (0.007)	-0.024 (0.014)
Born and lives in same borough	0.102* (0.008)	0.102* (0.008)	0.015 (0.008)	0.014 (0.008)	0.074* (0.005)	0.161* (0.010)
Lives in an adjacent borough	0.019 (0.013)	0.019 (0.013)	0.004 (0.013)	0.013 (0.013)	0.024* (0.007)	0.082* (0.014)
Hours last week	0.017* (0.001)	0.016* (0.001)	0.010* (0.001)	0.009* (0.001)	0.016* (0.000)	0.018* (0.000)
Hours missing	0.615* (0.036)	0.615* (0.036)	0.351* (0.041)	0.305* (0.041)	0.635* (0.024)	0.672* (0.028)
Quarrying	0.056 (0.132)	0.056 (0.132)	0.136 (0.127)	0.129 (0.134)	0.167* (0.054)	0.132 (0.120)
Brick/Glass worker	0.106 (0.099)	0.105 (0.099)	0.139 (0.087)	0.117 (0.089)	0.165* (0.048)	0.198 (0.084)
Chemicals	0.178 (0.095)	0.179 (0.094)	0.184 (0.080)	0.178 (0.082)	0.170* (0.042)	0.196 (0.078)
Metals	0.144 (0.093)	0.144 (0.093)	0.134 (0.078)	0.132 (0.080)	0.174* (0.040)	0.165* (0.073)
Electro-plate	0.090 (0.121)	0.090 (0.120)	0.107 (0.111)	0.079 (0.116)	0.240* (0.062)	0.144 (0.099)
Electricians	0.115 (0.093)	0.115 (0.093)	0.103 (0.079)	0.095 (0.081)	0.185* (0.041)	0.124 (0.075)
Watchmakers	0.151 (0.115)	0.152 (0.114)	0.113 (0.090)	0.107 (0.092)	0.226* (0.055)	0.096 (0.098)
Leather	0.104 (0.095)	0.105 (0.094)	0.131 (0.079)	0.111 (0.081)	0.175* (0.043)	0.130 (0.077)
Textiles	0.036 (0.096)	0.035 (0.096)	0.047 (0.083)	0.038 (0.085)	0.092 (0.048)	0.102 (0.085)
Dressmakers	0.108 (0.093)	0.108 (0.092)	0.112 (0.078)	0.101 (0.080)	0.166* (0.040)	0.103 (0.073)
Food/Drink	0.131 (0.093)	0.132 (0.093)	0.148 (0.079)	0.142 (0.081)	0.169* (0.040)	0.170 (0.074)
Wood	0.157 (0.094)	0.158 (0.093)	0.168* (0.079)	0.165* (0.081)	0.240* (0.040)	0.175* (0.073)
Paper	0.068 (0.094)	0.068 (0.093)	0.109 (0.079)	0.096 (0.081)	0.112* (0.042)	0.106 (0.075)

Printer	0.167 (0.094)	0.168 (0.094)	0.165* (0.079)	0.156 (0.081)	0.300* (0.041)	0.185* (0.075)
Builders	0.319* (0.095)	0.320* (0.094)	0.334* (0.081)	0.327* (0.083)	0.210* (0.040)	0.247* (0.074)
Painters	0.245* (0.095)	0.245* (0.095)	0.236* (0.080)	0.227* (0.082)	0.253* (0.041)	0.212* (0.075)
Other Materials	0.088 (0.095)	0.089 (0.094)	0.082 (0.080)	0.059 (0.082)	0.131* (0.043)	0.125 (0.079)
Other	0.049 (0.095)	0.049 (0.095)	0.072 (0.081)	0.065 (0.083)	0.121* (0.044	0.072 (0.077)
Transport	0.084 (0.093)	0.085 (0.092)	0.136 (0.077)	0.131 (0.079)	0.127* (0.039)	0.112 (0.072)
Finance	0.074 (0.093)	0.075 (0.092)	0.098 (0.078)	0.099 (0.080)	0.109* (0.040)	0.064 (0.073)
Public Admin.	0.556* (0.136)	0.560* (0.136)	0.728* (0.207)	0.647* (0.208)	0.453* (0.042)	0.381* (0.095)
Professional	0.402* (0.108)	0.402* (0.108)	0.285* (0.099)	0.259* (0.101)	0.391* (0.055)	0.343* (0.082)
Entertainment	0.178 (0.136)	0.178 (0.135)	0.214 (0.136)	0.211 (0.137)	0.180* (0.063)	0.049 (0.095)
Personal Services	-0.148 (0.093)	-0.148 (0.093)	-0.032 (0.078)	-0.002 (0.080)	-0.124* (0.040)	-0.125 (0.073)
Clerks	0.279* (0.093)	0.279* (0.092)	0.249* (0.078)	0.238* (0.080)	0.333* (0.040)	0.224* (0.073)
Warehouse	0.120 (0.093)	0.120 (0.093)	0.134 (0.078)	0.128 (0.080)	0.159* (0.040)	0.136 (0.073)
Drivers	0.297* (0.101)	0.299* (0.101)	0.269* (0.094)	0.265* (0.098)	0.167* (0.041)	0.200* (0.077)
Other	0.126 (0.094)	0.127 (0.094)	0.141 (0.079)	0.137 (0.081)	0.108* (0.040)	0.108 (0.074)
Missing	-0.090 (0.308)	-0.086 (0.307)	0.153 (0.175)	0.138 (0.182)	0.040 (0.121)	-0.034 (0.142)
Observations	15,436	16,566	11,997	11,354	31,668	31,668

Notes: The omitted occupation is agriculture. See Table 4 for additional results and notes.

Sources: Johnson, et al. (1999) and authors' calculations.

Table A.IV.4: Robustness Checks on Distance Estimates in Pay Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	4.738* (0.273)	3.359* (0.165)	2.447* (0.188)	2.451* (0.203)	2.577* (0.233)	1.858* (0.314)	2.992* (0.319)	2.540* (0.236)
$Distance^2$	-0.075* (0.018)	-0.057* (0.009)	-0.039* (0.011)	-0.035* (0.009)	-0.049* (0.014)	-0.026 (0.017)	-0.083* (0.023)	-0.046* (0.014)
Demographic	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Work Location	No	No	No	Yes	Yes	Yes	Yes	Yes
Residence	No	No	No	Yes	Yes	Yes	Yes	Yes
Other Distance	No	No	No	No	Yes	Yes	Yes	Yes
Household Fixed Effects	No	No	No	No	No	Yes	No	No
One Km^2 Grids	No	No	No	No	No	No	Yes	No
Excludes influential heads	No	No	No	No	No	No	No	Yes
Observations	15,436	15,436	15,436	15,436	15,436	15,436	15,185	14,679
R-squared	0.046	0.521	0.559	0.571	0.572	0.869	0.571	0.572
Mean Dep. Var.	10.359	10.359	10.359	10.359	10.359	10.359	10.355	10.355
F statistic	381.7*	1,396.4*	482.6*	186.8*	173.1*	107.8*	167.7*	164.4*

Notes: Estimates based on sample of family members, excluding head of household (column 1, Table 4). See Table 4 for additional notes. Column 8 excludes workers in households with influential heads – e.g. households where at least one additional member works for the same employer as the head.

Sources: Johnson, et al. (1999) and authors' calculations.

Table A.IV.5: Further Robustness Checks on Distance Estimates in Pay Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Unlimited Distance	Armstrong Skill	Exclude highest earner	No Bartlett sample	Discrete Distance: Non-Head Family	Discrete Distance: All Family	Only Heads	Discrete Distance: Only Heads
Distance	2.337* (0.177)	2.999* (0.239)	3.040* (0.267)	2.595* (0.258)			1.272* (0.167)	
$Distance^2$	-0.0304* (0.0093)	-0.0600* (0.0143)	-0.0734* (0.0154)	-0.0506* (0.0153)			-0.0418* (0.0118)	
Ref: Distance<0.5 Km								
Distance: 0.5-1					0.042* (0.014)	0.029* (0.009)		0.025* (0.010)
Distance: 1-2					0.058* (0.013)	0.042* (0.008)		0.041* (0.009)
Distance: 2-3					0.078* (0.014)	0.061* (0.009)		0.047* (0.010)
Distance: 3-4					0.083* (0.015)	0.074* (0.009)		0.062* (0.011)
Distance: 4-5					0.143* (0.017)	0.102* (0.010)		0.057* (0.012)
Distance: 5-10					0.176* (0.016)	0.126* (0.009)		0.071* (0.010)
Distance: 10+					0.237* (0.021)	0.183* (0.013)		0.117* (0.013)
Observations	15,441	15,436	12,480	13,298	13,298	31,195	15,758	15,758
R-squared	0.572	0.553	0.511	0.567	0.567	0.686	0.512	0.512
Mean Dep. Var.	10.359	10.359	10.259	10.349	10.349	10.754	11.141	11.141
F statistic	172.9	199.8	99.2	147.6	141.8	574.5	68.9	67.4

Notes: See notes in Table 4.

Sources: Johnson, et al. (1999) and authors' calculations.

Table A.IV.6: Effects of Cross and Out-Commuting on Distance Effects on Pay

	Related to Head	Related to Head & Age <25	Age <25 & Child of Head	Head only	All
	(1)	(2)	(3)	(4)	(5)
Distance	2.811* (0.425)	2.468* (0.428)	2.318* (0.431)	1.335* (0.307)	2.284* (0.275)
Distance squared	-7.185* (2.730)	-7.445* (2.665)	-6.633* (2.698)	-4.107* (1.764)	-6.754* (1.706)
Commute Out * Distance	-0.323 (0.564)	-0.512 (0.573)	-0.501 (0.577)	-0.102 (0.391)	-0.143 (0.360)
Commute Out * $Distance^2$	2.959 (2.942)	3.688 (2.918)	3.110 (2.961)	-0.109 (2.010)	0.858 (1.951)
Commute Across * Distance	0.594 (0.467)	0.628 (0.469)	0.563 (0.479)	-0.306 (0.358)	-0.008 (0.320)
Commute Across * $Distance^2$	-0.717 (3.901)	-2.102 (4.029)	-1.321 (4.114)	0.369 (3.077)	0.402 (3.191)
Observations	15,436	11,997	11,354	15,758	31,668
R-squared	0.572	0.632	0.645	0.512	0.687
Mean Dep. Var.	10.359	10.273	10.271	11.141	10.756
F statistic	168.0*	157.9*	158.1*	127.1*	533.4*

Notes: See notes in Table 4.

Sources: Johnson, et al. (1999) and authors’ calculations.

Table A.IV.7 Full Ordinary Least Squares and Instrumental Variable Estimates of the Wage Returns to Distance (Table 5)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS <6k	2nd Stage IV <6k	OLS <8k	2nd Stage IV <8k	OLS <10k	2nd Stage IV <10k	OLS <15k	2nd Stage IV <15k	OLS <30k	2nd Stage IV <30k	OLS: All	2nd Stage: IV All
Distance to work	1.540* (0.137)		1.520* (0.130)		1.393* (0.115)		1.342* (0.104)		1.331* (0.101)		1.337* (0.097)	
Distance to work IV (Birthplace centrality)		3.87* (1.69)		3.61* (1.10)		3.33* (0.80)		3.84* (0.68)		4.33* (0.80)		3.55 (2.38)
Age	0.0929* (0.0019)	0.0917* (0.0021)	0.0955* (0.0018)	0.0946* (0.0019)	0.0971* (0.0017)	0.0962* (0.0017)	0.0967* (0.0016)	0.0954* (0.0017)	0.0963* (0.0016)	0.0948* (0.0017)	0.0939* (0.0016)	0.0927* (0.0020)
Age^2	-0.00107* (0.00002)	-0.00106* (0.00003)	-0.00111* (0.00002)	-0.00109* (0.00002)	-0.00113* (0.00002)	-0.00112* (0.00002)	-0.00112* (0.00002)	-0.00111* (0.00002)	-0.00112* (0.00002)	-0.00110* (0.00002)	-0.00109* (0.00002)	-0.00107* (0.00003)
Missing Age	1.9121* (0.036)	1.8820* (0.0428)	1.9540* (0.0336)	1.9309* (0.0360)	1.9830* (0.0317)	1.9629* (0.0330)	1.9760* (0.0308)	1.9483* (0.0321)	1.9703* (0.0305)	1.9370* (0.0321)	1.9305* (0.0299)	1.9061* (0.0391)
Male	0.4069* (0.0110)	0.3987* (0.0122)	0.4006* (0.0100)	0.3929* (0.0107)	0.3992* (0.0092)	0.3925* (0.0096)	0.4037* (0.0090)	0.3958* (0.0092)	0.4042* (0.0089)	0.3946* (0.0093)	0.4128* (0.0088)	0.4049* (0.0123)
Head of household	0.1631* (0.0120)	0.1621* (0.0122)	0.1679* (0.0110)	0.1662* (0.0112)	0.1690* (0.0103)	0.1676* (0.0104)	0.1633* (0.0099)	0.1620* (0.0100)	0.1632* (0.0098)	0.1614* (0.0010)	0.1660* (0.0096)	0.1651* (0.0098)
Observations	12,430	12,430	15,131	15,131	17,035	17,035	18,233	18,233	18,581	18,581	19,680	19,680
R-squared	0.618	0.610	0.617	0.610	0.620	0.614	0.622	0.610	0.622	0.605	0.622	0.613
F	303.37*		367.15*		420.14*		452.19*		458.07*		477.11*	
Wald χ^2		21,796.10*		26,605.01*		30,411.04*		32,201.95*		32,073.98*		34,202.49*
1st Stage F statistic		4.75*		12.64*		14.80*		16.68*		17.05*		9.24*

Notes: See notes in Table 5.

Sources: Johnson, et al. (1999) and authors' calculations.

Appendix V: Costs and Benefits of Commuting and Calculating Poverty Lines

I. Costs and benefits of Commuting

Table 6 shows the costs and benefits of commuting. The calculations in this table are very much back-of-the-envelope, but are informed by stylized facts from Ponsonby and Ruck (1930) and from the *New Survey* data. In this appendix we describe the assumptions behind these calculations and examine their historical basis.

Crow-flies distances:

The distances in the first column are arbitrary round numbers, which cover the range of distances commuted. The range of distances covers virtually all workers who were not working “on the spot” (the term used by Ponsonby and Ruck (1930) to describe the proximity of employment to residence in the 1890s). Approximately 86 percent of income earners in the sample travelled at least 500 meters; approximately 99 percent travelled at most 16 kilometres.

Transport mode:

Ponsonby and Ruck (1930) state that workers typically walked up to 1.6 kilometres, buses and trams were used interchangeably for distances of 1.6 to 3.2 kilometres, the Underground was typically used for distances of 3.2 to 19.3 kilometres, and trains were used for longer distances or as a replacement for the Underground in places where it was not available. Approximately 82.9 percent of workers with non-missing transport costs who travelled distances of up to 1.6 kilometres, reported costs of exactly zero, and thus must have walked or cycled. Approximately 83.7 percent of individuals with non-missing transport costs who commuted at least 1.61 kilometres, reported positive costs, and thus must have used public transport. We do not observe the mode of transport, but the relationship between proximity to stops/stations and commuting distances shown in the regressions in Table 3 is consistent with the pattern reported by Ponsonby and Ruck (1930).

Distance walked:

We construct *as the crow-flies*, distances i.e. in a straight line. The urban layout rarely allows this to be the actual route, thus actual distance travelled must be greater than crow-flies distance. There exists a substantial literature in geography on the difference between crow-flies and actual distances (Rietveld, et al. 1999 and Underhill 2020). We assume, somewhat arbitrarily, that walking-only journeys were 25 percent longer than the crow-flies distance. Any journey involving public transport would have also involved walking from home to transport and from transport to work. Table 2 shows the average distance from home and work to each public transport mode. However, workers would have chosen their mode of transport at least partly based on proximity, so the walking distance to the chosen mode of transport will have on average been less than for modes not used. We assume 300 meters of total walking at both ends for a one-way bus journey, about 75 percent of the sum of average distances from the nearest bus stop to home and workplace. We also assume 300 meters of walking for a one-way tram journey. This is considerably less than the average distance shown in Table 2; however, one would expect that the distance walked by tram users would have been approximately the same as bus users, as the two modes of transport were used interchangeably. We assume 600 meters of walking at both ends for a one-way Underground or train journey, about 75 percent of the maximum distance that geographers have argued that commuters are willing to walk to these modes of transport (Daniels and Mulley 2013).

Distance on public transport:

As with walking, public transport generally does not travel in a straight line and this adds to the total distance. On the other hand, in many cases workers can choose between stops which are approximately equidistant to home (work). If there are two approximately equidistant stops from home, a worker would have been more likely to use the one closer to their work in order to minimize total travel time. This will at least partly offset the effect of added distance due to non-linear transport routes, and thus we assume that the travel distance for public transport is only 10 percent more than the crow-flies distance of total travel.

With each of our assumptions about distance, there is likely to be considerable heterogeneity across individuals and locations. The assumptions are not verifiable in the data, so it is also possible that there is some error on average. However, it is unlikely that modest errors in either direction will have a substantial impact on our conclusions in Section VI.

Walking and public transport time:

Time is calculated as distance (from columns 3 and 4 in Table 6) divided by speed. Following Leunig (2006), we use 4 kilometres per hour as a typical urban walking speed. We take public transport speeds from Table 1. Neither Ponsonby and Ruck (1930) nor *London Statistics* provide train speeds after the first World War and we assume that average speeds in 1930 were 10 percent faster than those reported in *London Statistics* for 1907-08, similar to the increase in rail speed calculated by Leunig (2006) for "minor routes" – intercity rail routes which stopped at most or all on-route stations.

Implied return:

We use the regressions in Table 4, column 1 and Table 5, column 9 evaluated at the tenth, twenty fifth, and fiftieth percentiles of the weekly earnings distribution (180d, 300d, and 583.5d) to calculate the returns to commuting the distances shown in the first column. Specifically, we estimate the returns as:

$$\text{Exp}[\ln(WE) + 0.02577d - 0.000489d^2] - \text{Exp}[\ln(WE)] \quad \text{OLS estimates}$$
$$\text{Exp}[\ln(WE) + 0.0333d] - \text{Exp}[\ln(WE)] \quad \text{IV estimates}$$

where:

WE = weekly earnings (in pence), evaluated at the 10^{th}, 25^{th}, and 50^{th} percentile
d = distance commuted, in kilometres

Monetary costs:

We use reported transport expenses from the *New Survey* data and the description of travel costs from Ponsonby and Ruck (1930) to determine a "typical" cost for the journey in each row. Public transport fares were set according to travel zones, which were imperfectly correlated to distance. For any given distance and mode of transport there may have been multiple fares, depending on the embarkation and disembarkation stations. The monetary costs reported in Table 6 are the range of typical fares reported in the *New Survey* data for workers commuting distances within 500 meters of the crow-flies distances in column 1.[47]

Implied time costs:

[47] The *New Survey* data typically reports weekly expenditures on transport. We divide this by 12 to find one-way fares. In each case, the range reported in Table 6 covers both the median and modal fare.

The total time spent commuting is the sum of walking time, transport time, and waiting time. We assume five minutes waiting time for the bus and tram and eight minutes for the train and Underground. We assume a longer time for train and Underground because the platform was physically removed from the entrance to the station. Following an extensive literature on the value of travel time saved, we use values of 50 and 100 percent of salary as the lower and upper bounds of the implied cost of commuting time (Wardman 1998; Zamparini and Reggiani 2007). In the second column of Table 6, we assume a single mode of transport. In practice, some commuters, particularly those with longer commutes, may have needed to transfer between modes and thus incurred additional waiting time.

II. The Hatton-Bailey Poverty Line

The appendix in Hatton and Bailey (1998) outlines their approach to constructing household poverty lines. To briefly summarize, they allocate minimum required expenditures on food and clothing, rent, and fuel. The minimum required expenditure on food and clothing is based on age and sex of the individual and ranges from 36d per week for a child under age 1 to 102d per week for males aged 18 and over. The minimum required rental expenditure is based on a standard of no more than two individuals to a room and a cost of 60d per week for one room, 102d per week for two rooms, 126d per week for three rooms, and 30d per week for each additional room. The minimum required expenditure on fuel is 36d per week, plus an additional 2d in South London. A household is classified as poor if the sum of these minimum expenditures and actual household expenditures on transport and National Insurance is greater than their total income from all sources. The *New Survey* reports income in previous week and in a full-time week and Hatton and Bailey (1998) use income from the previous week for the poverty calculations. The estimated poverty lines using this approach are 198, 292, 392, 588 pence per week for a household with only a single adult male, a married couple, a couple with one child, and a couple with three children, respectively. Hatton and Bailey (1998) estimate that 12.11 percent of households and 12.0 percent of individuals fell below the poverty line.

We have made two adjustments to their calculations. First, approximately 9.2 percent of individuals in the *New Survey* have a reported age of exactly zero. This is an implausibly large number and almost certainly indicates that age was not reported in most of these observations. We have reclassified these individuals as adults and adjusted required expenditure on food and clothing if 1) they had an occupation or reported earnings or hours worked or 2) their relationship to the head of household indicates they must have been an adult (e.g. wife or grandfather). If the individual was not an income earner and was plausibly a child (e.g. son or nephew of the head), we use a value of zero for age and assign minimum expenditure accordingly. The reclassification of children age zero to adults (age 18+) increases the minimum expenditure on food and clothing from 38d per week to 102d per week for 3742 men and to 94d per week for 4792 women. It is likely that many of those we still classify as age zero were actually older and thus would require greater expenditures than for an infant, so our upwards adjustment to the poverty line is a lower-bound. Secondly, both National Insurance contributions and transport expenses are frequently missing in the *New Survey* data. Among individuals reporting earnings or hours in the previous week, approximately 30 percent are missing data for transport costs and approximately 20 percent are missing data for National Insurance contributions. We have handled missing data in two ways: constructing a lower-bound poverty line where missing observations are replaced by a value of zero and an expected value poverty line where missing observations are replaced by the overall sample mean. These reclassifications increase the poverty line for 4730 households and increase the number of individuals classified as poor from 12.0 percent of the sample (Hatton and Bailey 1998, p. 584) to 16.7 percent of the sample (lower-bound poverty line) or 20.6 percent of the sample (expected value poverty line).

References

Daniels, Rhonda and Mulley, Corinne (2013). “Explaining walking distance to public transport: the dominance of public transport supply.” Journal of Transport and Land Use, 6(2): 5-20.

Leunig, Tim (2006). “Time is money: a re-assessment of the passenger social savings from Victorian British railways.” Journal of Economic History, 66(3): 635-73.

London County Council (various years). *London Statistics*. (London: P. S. King and Son).

Ponsonby, G. and Ruck, S. K. (1930). “Travel and mobility.” In: Llewellyn-Smith, Hubert, eds. *The New Survey of London Life and Labour*. (London: P. S. King), pp. 171-99.

Rietveld, Piet, Zwart, Bert, van Wee, Bert, van den Hoorn, Toon (1999). “On the relationship between travel time and travel distance of commuters.” Annals of Regional Science, 33(3): 269-87.

Underhill, Keziah (2020). “December visualisation blog: crow flies vs. OD paths.” https://www.basemap.co.uk/december-visualisation-blog-crow-flies-vs-od-paths/ [accessed 23/02/2021].

Wardman, Mark (1998). “The value of travel time: a review of British evidence.” Journal of Transport Economics and Policy, 32(3): 285-316.

Zamparini, Luca and Reggiani, Aura (2007). “Meta-analysis and the value of travel time savings: a transatlantic perspective in passenger transport.” Networks and Spatial Economics, 7(4): 377-96.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1867	Emanuel Ornelas John L. Turner	The costs and benefits of rules of origin in modern free trade agreements
1866	Giammario Impullitti Syed Kazmi	Globalization and market power
1865	Nicolás González-Pampillón Gonzalo Nunez-Chaim Henry G. Overman	The economic impacts of the UK's eat out to help out scheme
1864	Rocco Macchiavello Ameet Morjaria	Acquisitions, management and efficiency in Rwanda's coffee industry
1863	Andrew E. Clark Conchita D'Ambrosio Jan-Emmanuel De Neve Niccolò Gentile Caspar Kaiser Ekaterina Oparina Alexandre Tkatchenko	Human wellbeing and machine learning
1862	Fabrizio Leone Rocco Macchiavello Tristan Reed	Market size, markups and international price dispersion in the cement industry
1861	Elias Einio Josh Feng Xavier Jaravel	Social push and the direction of innovation
1860	Xiao Chen Hanwei Huang Jiandong Ju Ruoyan Sun Jialiang Zhang	Endogenous cross-region human mobility and pandemics
1859	Xavier Jaravel Danial Lashkari	Nonparametric measurement of long-run growth in consumer welfare

1858	Leonardo Bursztyn Jonathan Kolstad Aakaash Rao Pietro Tebaldi Noam Yuchtman	Political adverse selection
1857	Attila Lindner Balázs Muraközy Balázs Reizer Ragnhild Schreiner	Firm-level technological change and skill demand
1856	Jeremiah Dittmar Ralf R. Meisenzahl	The university, invention and industry: evidence from German history
1855	Donna Brown Jonathan Wadsworth	Accidents will happen: (de)regulation of health and safety legislation, workplace accidents and self employment
1854	Fabrizio Leone	Foreign ownership and robot adoption
1853	Andrew E. Clark Conchita D'Ambrosio Niccoló Gentile Alexandre Tkatchenko	What makes a satisfying life? Prediction and interpretation with machine-learning algorithms
1852	Xiang Ding Teresa C. Fort Stephen J. Redding Peter K. Schott	Structural change within versus across firms: evidence from the United States
1851	Christos Genakos Eleni Kyrkopoulou	Social policy gone bad educationally: unintended peer effects from transferred students
1850	Antonin Bergeaud Cyril Verluise	A new dataset to study a century of innovation in Europe and in the US
1849	Jo Blanden Mattias Doepke Jan Stuhler	Education inequality

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE