ESCoE ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

escoe.ac.uk @ESCoEorg

# The New Wave: Technology Diffusion in the UK During the 2010s

Max Nathan, Centre for Advanced Spatial Analysis, University College London

Starting at 12.30 PM

**ESCoE ECONOMIC MEASUREMENT WEBINARS**

# The New Wave: Technology diffusion in the UK during the 2010s

Max Nathan (UCL and CEP)

(with Mirko Draca, Viet Nguyen-Tien, Juliana Oliveira-Cunha, Anna Rosso, Anna Valero and Shenxing Zhang)

max.nathan@ucl.ac.uk

@iammaxnathan

This version: November 2022

**Work in progress!**

Results are not final and may change. Please do not share or quote without permission.

# Background

- **Importance of tech diffusion for productivity ~> growth. But also ~> inequality** (Bryan and Williams 2021, Acemoglu 1998)
  - General Purpose Technologies (GPTs) (Bresnahan 2010)
  - Particular relevance *now* - productivity puzzle in the UK

- Large literature looking at ICT. But PCs and the WWW are now approaching middle age

- Like other revolutions before it (David 1990, Perez 2010) **the ICT revolution needs to be considered in overlapping stages, or 'waves'**

# Background

- **A 'new wave' of likely GPTs has emerged in the last decade – cloud and ML/AI in particular** (Goldfarb et al 2022, Brynjolfsson et al 2021)
  - Optimists: improved productivity growth across the economy once this wave diffuses (Brynjolfsson and McAfee 2014)
  - Pessimists are more sceptical (Gordon 2016)

- Importance of **understanding diffusion (and emerging economic impacts)** of this new wave of technologies <u>in the UK</u> – both for knowledge and for public policy

# Research questions

1) What are the distributions of key ‘new wave’ technologies – cloud & ML/AI – over firms and places?

2) What explains this?
   - Firm vs. sector differences
   - Localised learning and spillovers
   - Area complements (especially varieties of human capital)
   - Persistence / path-dependence (and overlapping tech waves)

3) *Later: what is the impact of the new wave on wage inequality across places?*

# Preview of findings

- Diffusion of ‘new wave’ tech in the 2010s is uneven: across occupations, firms, space

- **At area level, cloud and ML/AI adoption in 2010s is more STEM-biased than PC adoption in the 2000s**
- Cloud and ML/AI adoption is also much more London-biased; influenced by fast broadband provision; but not by PC rollout

- **Minority of most STEM-intensive firms adopt much faster than the rest, pull away from the pack**
- Evidence of persistence for areas, and for firms

# Framework

# Computers, skills and cities

- We start with the skill-biased technological change framework developed by Beaudry et al (2010)
  - **Beaudry and co-authors look at the spatial diffusion of PCs across US cities, 1980-2000** [and impacts on wage ratios]
  - **Key idea: complements drive diffusion**
    - Key complement in their framework: **area stocks of human capital**
    - Technology spread follows area-level comparative advantage – especially places where complements are abundant, cheap
  - Also: spillovers and localised clusters ('innovative city effects')
  - Also: path-dependence may shape an area's skill mix, other complements

# We build on this

- At this stage in the project, we focus on extending the framework empirically:
  - Wave 1: PC diffusion in the 2000s, directly measured
  - Wave 2: Cloud vs. ML/AI diffusion in the 2010s, adapting text-based methods from Bloom et al (2021)
  - Skills: distinguish general and technical human capital, proxied by the share of STEM workers (NESTA 2015)
  - Other correlates: look at city / location differences, wave 1 ~ wave 2 linkages, and role of broadband as an intermediate GPT
  - Scale: run analysis at both area and firm level

# Related literatures

- **Macro / spatial frameworks.** Technological paradigms, and contingent diffusion (David 1990, Perez 2010). Spatial patterns and dynamics of technology diffusion (Brezis and Krugman 1997, Duranton 2007, Berkes et al 2021)
- **Micro frameworks of technology diffusion**
  - Sector/firm variation in adoption cost/gains (Stoneman & Battisti 2010)
  - Localised learning, information asymmetries (Geroski 2000)
  - Complementarities at firm, area level (Bresnahan et al 2002, Beaudry et al 2010, Feng and Valero 2020)
  - Path-dependence (Nelson & Phelps 1966, Balland et al 2020)
- **Tracking diffusion using job ads** (Acemoglu et al 2022, Goldfarb et al 2022, Bloom et al 2021, Webb 2020 + others!)

# Data and build

# Data

- **Cloud + ML/AI adoption, 2010s –** firm-level online vacancy data from Burning Glass Technologies, 2012-2019 [more]
  - Use for adoption measures, also firm baseline characteristics

- **PC adoption, 2000s –** establishment-level survey data from Harte-Hanks, 2000-2002 [more]
- **Area human capital** – Output Area (OA) and local authority (LAD) data from the 1991/2011 Census, collapsed to Travel To Work Areas (TTWAs) [more]
- **Area broadband** – LAD-level speed data from Ofcom, 2011, crosswalked to TTWAs [more]
- **Area controls** – OA- and LAD-level data from 1991/2011 Census, crosswalked to TTWAs [more]

# Burning Glass data

- Burning Glass [BGT] scrape UK job ads from a range of online sources. Raw data ~60m vacancies, 2012-2019. Good coverage for our purposes.

- Use vacancies to proxy cloud / ML-AI adoption:

  – Vacancies ~ innovation: hiring around a given new tech **predicts future patenting patterns** in the same technology space (Goldfarb et al 2022)

  – Vacancies ~ adoption: hiring patterns reflect **tech-related shifts in labour demand** (Tambe and Hitt 2012)

  – Vacancies ~ adoption: new technologies **leave a vacancy ‘footprint’** (Acemoglu et al 2022)

# Ads ~> Adoption

- We adapt Bloom et al (2021), who identify 29 key 'disruptive technologies' as sets of text bigrams
  - Bloom and co use supervised learning on US patents and earning call text (1970-2020) [more]
  - We implicitly assume US~UK similarity (!)
  - For each tech, feed Bloom et al bigrams into cleaned job ad text
  - For each tech, flag a job ad as 'exposed' if ad text contains bigrams for that tech; otherwise 'not exposed'
  - For each tech, manually inspect posts with unusual SIC4 or SOC4 codes. Reassign to 'not exposed' in some cases
  - Use ONS E-Commerce Surveys to validate link between hiring and underlying tech adoption – at industry level [more]

# Build

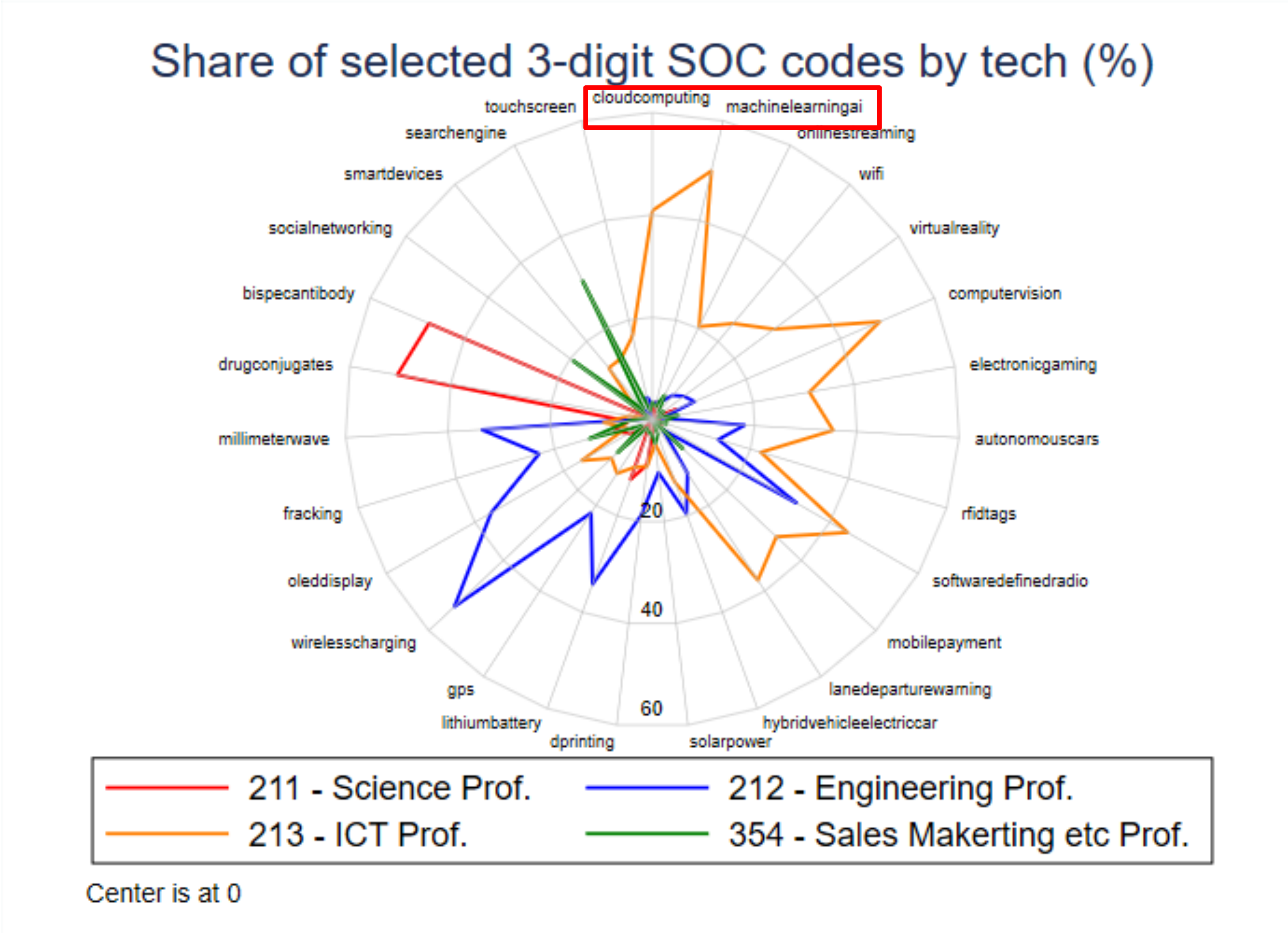
- **Firm-level dataset:**
  - Use employer name and SIC4 to identify firms, pooling establishments [HQ = plant with most vacancies]
  - Subset of larger firms with >100 posts per year
  - **Panel of 16,195 firm-year obs, for 1855 firms 2012-219**
  - Also atm running firm regressions with area controls
- **Area-level dataset:**
  - Aggregate firm-level data to Travel to Work Areas, add controls
  - BGT data kinks: merge London + Slough & Heathrow; Bournemouth + Poole
  - **Baseline: 206 TTWAs in Great Britain** [dstats]

# Challenges

- **Measurement error** – are we fully observing adoption? ‘New wave’ tech may be a) outsourced and/or b) embodied in software. Hiring data may better cover producers > users
  - We validate hiring ~ adoption links with ONS Survey data [more]
  - Drop producer (ICT), enabler (consultancy) SICs? Per Acemoglu et al 2022
- **Endogeneity** – our regressions are pretty basic atm
  - Yes … tho we use Control Function tests that are generally supportive
  - We would like exogenous measures of skills supply. What could these be?
- **Selection** – firm data is implicitly selected on / in BGT sample
  - We agree – it’s annoying
  - We could match to other firm-level data with known sample?

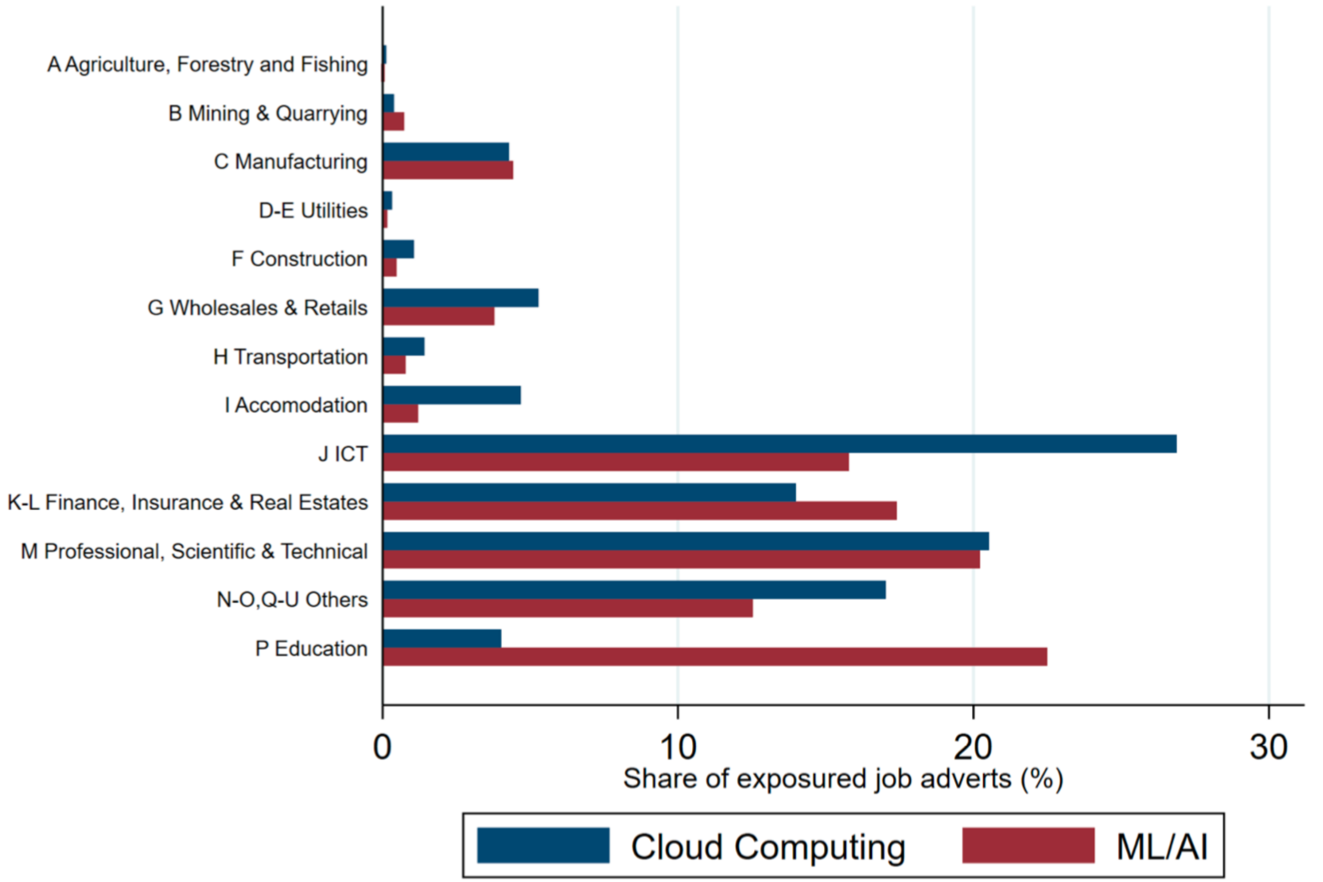
# Descriptives

# ‘Disruptive tech’ hiring mostly involves STEM occupations




BGT 2019 data, for vacancies in the 29 disruptive technologies from Bloom et al.
Vacancies tagged to SOC3 bins. Four biggest bins shown.

# Cloud and ML/AI adoption is uneven across industries

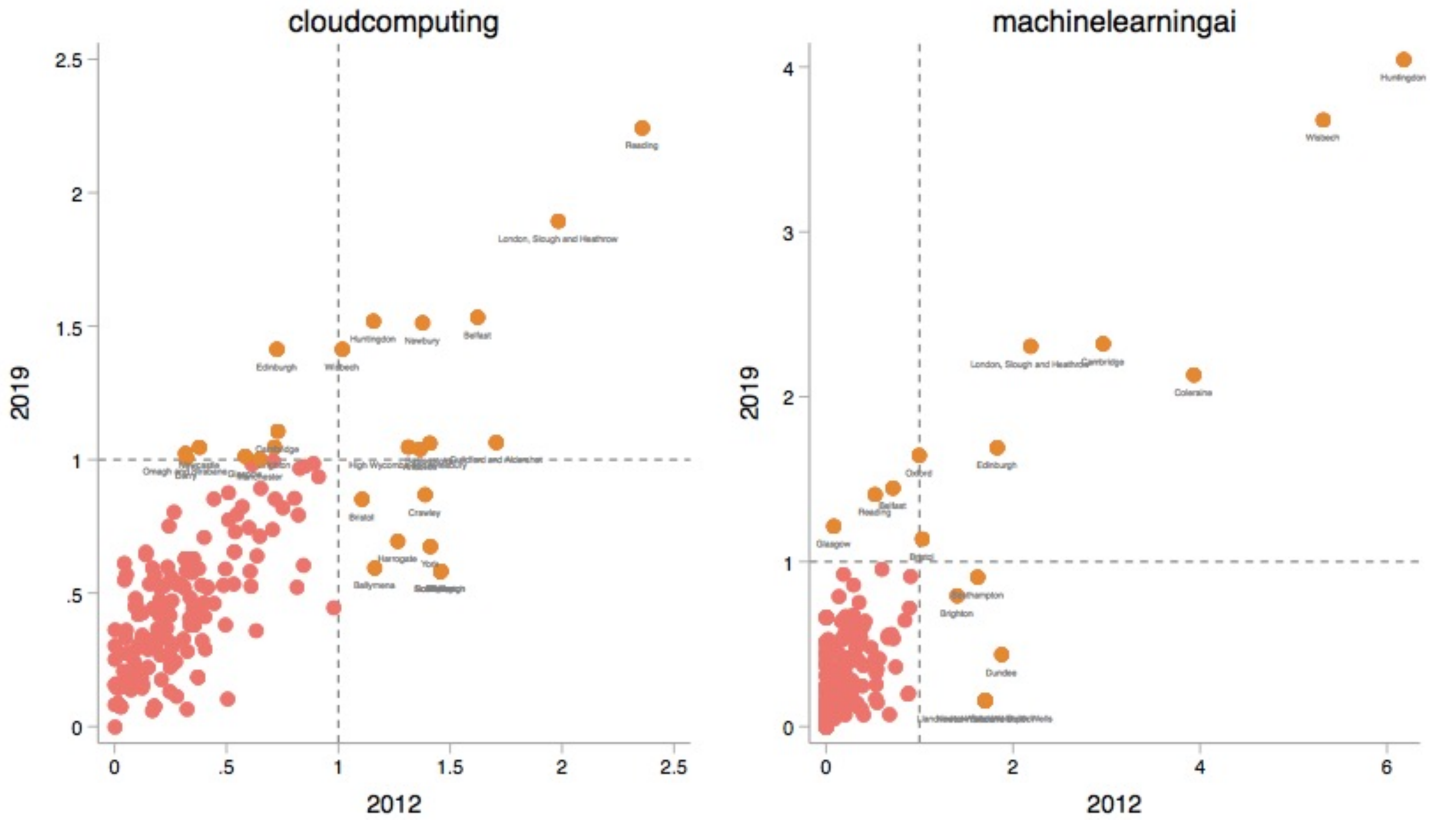



BGT 2019 data. Shares of all cloud and ML/AI-exposed jobs by SIC1 bin.

Sample = all ads where a SIC code is assigned.

# Cloud and ML/AI are highly clustered across physical space

Location quotients for 2012 vs 2019 [England and Wales only]




Source: Burning Glass. Graphs show the location qotients (LQs) of job ads in each technology.
LQs above 1 indicate that local ad shares are greater than national shares, implying clustering.
Each cell is a UK Travel to Work Area (TTWA). TTWAs in the bottom left quadrant have no clustering in that technology.
TTWAs in the top left have more clustering over time. TTWAs in the top right have clustering in both periods. TTWAs in the bottom right have de-clustering.

# Area results

# Area regressions

- For TTWA *j*, year *t*, we estimate:

$$Y_{jt} = a + b1\ln(S/U)_{j,t0} + b2\ln(STEM/nSTEM)_{j,t0} + \mathbf{X}c_{j,t0} + e_{jt}$$

- Where:
  - Y = adoption intensity in *j*, adjusted by firm size-SIC3-year cell [more]
    - Wave 1: PCs per employee
    - Wave 2: Cloud or ML/AI ads per 1,000 ads
  - S/U = ratio of graduates to non-graduates in *j* in base year *t0*
  - STEM/nSTEM = ratio of STEM / non-STEM workers in *j,t0*
  - **X** contains base-year TTWA controls: population density, London / Scotland / Wales dummies, % MF, % ILO unemployed. Plus in Wave 2 regressions, we can add share of super-fast broadband connections and PC adoption
- All regressions are weighted by TTWA working-age population

# Wave 2 is more skill-intensive than Wave 1, and more London-biased

| | (1) PC (2000s) | (2) Cloud (2019) | (3) ML/AI (2019) | (4) Cloud (2019) | (5) ML/AI (2019) |
|---|---|---|---|---|---|
| Baseline Gen. Skills (std.) | 0.30*** | 0.51*** | 0.31*** | 0.44*** | 0.24* |
| | (0.05) | (0.09) | (0.10) | (0.13) | (0.12) |
| Baseline Pop. Dens. (std.) | 0.03 | 0.09 | -0.04 | -0.09 | -0.10 |
| | (0.04) | (0.07) | (0.05) | (0.08) | (0.07) |
| London | 0.16 | 1.66*** | 1.95*** | 2.41*** | 2.34*** |
| | (0.23) | (0.47) | (0.28) | (0.43) | (0.35) |
| Scotland | -0.32* | -0.65** | 0.33* | -0.63** | 0.46** |
| | (0.18) | (0.27) | (0.20) | (0.29) | (0.22) |
| Wales | -0.43** | -0.47** | -0.18 | -0.25 | 0.01 |
| | (0.20) | (0.22) | (0.13) | (0.19) | (0.11) |
| Baseline Manuf. (std.) | -0.05 | 0.05 | 0.03 | 0.01 | 0.05 |
| | (0.05) | (0.07) | (0.04) | (0.07) | (0.06) |
| Baseline UR (std.) | -0.00 | 0.09 | 0.07 | 0.14 | 0.00 |
| | (0.07) | (0.09) | (0.06) | (0.13) | (0.09) |
| Broadband | | | | 1.07*** | 0.71*** |
| | | | | (0.35) | (0.23) |
| PC per Employees (adj win std) | | | | 0.13 | 0.05 |
| | | | | (0.09) | (0.06) |
| Observations | 206 | 206 | 206 | 206 | 206 |
| $R^2$ | 0.462 | 0.799 | 0.655 | 0.819 | 0.667 |

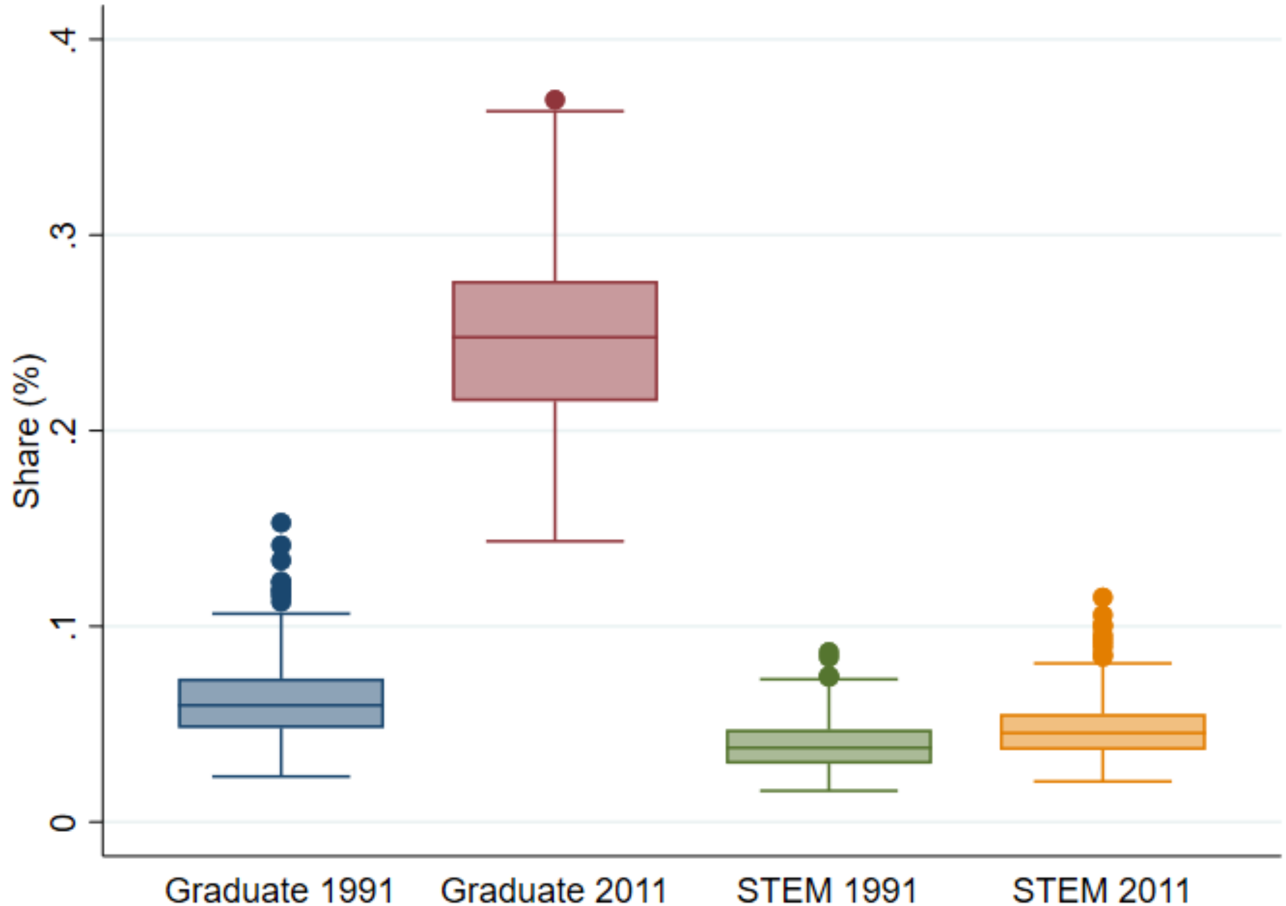
Robust standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01. Skills are general skills. Baseline = 1991 for PC and 2011 for Cloud and ML/AI. Manuf. is the share of employed working in manufacturing. UR is the unemployment rate. Broadband is the share of super-fast broadband coverage in 2011. PC per employees is winsorised at the top and bottom 1% and then adjusted by employee number (8 bins) interacted with industry (3-digit SIC codes) and year

# Wave 2 adoption is more STEM-intensive

| | (1) PC (2000s) | (2) Cloud (2019) | (3) ML/AI (2019) | (4) Cloud (2019) | (5) ML/AI (2019) |
|---|---|---|---|---|---|
| Baseline Gen. Skills (std.) | 0.15** | 0.23* | 0.04 | 0.28* | 0.03 |
| | (0.06) | (0.12) | (0.09) | (0.15) | (0.10) |
| Baseline STEM Skills (std.) | 0.27*** | 0.41*** | 0.40*** | 0.27* | 0.35** |
| | (0.08) | (0.12) | (0.12) | (0.15) | (0.14) |
| Baseline Pop. Dens. (std.) | -0.00 | 0.05 | -0.08 | -0.07 | -0.09 |
| | (0.04) | (0.07) | (0.06) | (0.09) | (0.07) |
| London | 0.50** | 2.17*** | 2.44*** | 2.57*** | 2.54*** |
| | (0.25) | (0.44) | (0.38) | (0.42) | (0.38) |
| Scotland | -0.34* | -0.70** | 0.29 | -0.70** | 0.36 |
| | (0.18) | (0.30) | (0.21) | (0.30) | (0.24) |
| Wales | -0.30 | -0.32 | -0.03 | -0.21 | 0.06 |
| | (0.20) | (0.21) | (0.11) | (0.19) | (0.11) |
| Baseline Manuf. (std.) | -0.09* | -0.03 | -0.05 | -0.06 | -0.04 |
| | (0.05) | (0.07) | (0.04) | (0.08) | (0.05) |
| Baseline UR (std.) | 0.01 | 0.13 | 0.10 | 0.16 | 0.02 |
| | (0.08) | (0.10) | (0.06) | (0.13) | (0.09) |
| Broadband | | | | 0.73* | 0.27 |
| | | | | (0.44) | (0.26) |
| PC per Employees (adj win std) | | | | 0.06 | -0.04 |
| | | | | (0.08) | (0.07) |
| Observations | 206 | 206 | 206 | 206 | 206 |
| $R^2$ | 0.502 | 0.818 | 0.687 | 0.825 | 0.684 |

Robust standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01. Skills are general skills. Baseline = 1991 for PC and 2011 for Cloud and ML/AI. Manuf. is the share of employed working in manufacturing. UR is the unemployment rate. Broadband is the share of super-fast broadband coverage in 2011. PC per employees is winsorised at the top and bottom 1% and then adjusted by employee number (8 bins) interacted with industry (3-digit SIC codes) and year

# Graduates become abundant, but STEM workers remain rare




Box plots show the distributions of graduate and STEM shares in 1991 and 2011 across the TTWAs in our regressions

# Robustness

- Our framework suggests that recent skills supply predicts adoption of new wave technologies. Two challenges to that:
  - **Reverse causality:** tech adoption leads to skilled worker sorting
  - **Unobservables:** ‘innovative city’ effects raise demand for skills and for tech adoption, and these channels may be deeply path-dependent

- From Beaudry et al, a control-function style check:
  - In our TTWA regression, split skills variables into base level and change
  - Reverse causality => worker sorting => **change > base**
  - Unobservables => skills ‘path’ varies over time => **change =/= base**

# Control function test

| | Cloud Computing 19 | | ML/AI 19 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| General skills 2011 | 4.190* | | 0.249 | |
| | (2.258) | | (0.804) | |
| Skills 91 ($\alpha$) | | 3.815* | | 0.0265 |
| | | (2.253) | | (0.825) |
| Skills change 91-11 ($\alpha'$) | | 2.183 | | -0.889 |
| | | (2.959) | | (1.527) |
| STEM Skills 2011 | 3.372* | | 2.233** | |
| | (1.840) | | (0.933) | |
| STEM Skills 91 ($\beta$) | | 3.212* | | 2.148** |
| | | (1.842) | | (0.880) |
| STEM Skills change 91-11 ($\beta'$) | | 3.527 | | 2.454* |
| | | (2.560) | | (1.430) |
| Scotland | -2.842** | -2.861** | 0.761 | 0.751 |
| | (1.204) | (1.197) | (0.492) | (0.492) |
| Wales | -0.853 | -0.818 | 0.129 | 0.139 |
| | (0.751) | (0.741) | (0.221) | (0.246) |
| Manuf.11 | -7.315 | -6.675 | -2.464 | -2.160 |
| | (8.930) | (9.145) | (2.884) | (2.984) |
| UR11 | 38.96 | 39.35 | 3.108 | 3.410 |
| | (32.28) | (32.04) | (11.09) | (11.48) |
| Broadband | 2.947* | 2.903 | 0.561 | 0.535 |
| | (1.776) | (1.780) | (0.548) | (0.568) |
| PC per employees | 1.644 | 1.561 | -0.490 | -0.522 |
| (adj win std) | (2.133) | (2.148) | (0.938) | (0.934) |
| London & Density controls | Y | Y | Y | Y |
| Observations | 206 | 206 | 206 | 206 |
| p-value: $\alpha = \alpha'$ | | 0.367 | | 0.401 |
| p-value: $\beta = \beta'$ | | 0.834 | | 0.697 |

Change > base implies reverse causality

Significant *t*-test implies unobservables present

We run the control function test for general skills [α, α′] and STEM skills [β, β′]

In both cases, some evidence against reverse causality and unobservables

# Firm results

# Firm-level analysis

- **A quick flavour of what we're doing**
- **Firm growth literature emphasises**: huge variation in firm capacities, small subsets of firms drive aggregate patterns; firm capacities build gradually, show persistence ~> path-dependence
- We have **broadly analogous findings**:
  - On average, initial STEM intensity predicts Wave 2 adoption …
  - … but big differences in initial STEM-intensity across firms
  - Small minority of most STEM-intensive firms more likely to adopt Wave 2 tech in subsequent years, and to adopt more
  - Results also survive adding in area-level controls

# Firm regressions

- For firm *i,* area *j*, sector *k*, year *t*, we estimate:

$$\Delta Y_{ijkt} = a + b1 MID_{i,t0} + b2HIGH_{i,t0} + b3STEM_{i,t0} + \mathbf{Z}_{i,t0} + K_k + T_t + e_{ijkt}$$

- Where:
  - $\Delta$Y = change in firm adoption, adjusted by firm size-SIC3-year cell [more]
    - Extensive margin / intensive margin
    - Wave 2: Cloud or ML/AI ads per 1,000 ads
  - MID and HIGH = firm’s base-year shares of mid and high-skill vacancies as defined by ONS SOC1 typology
  - STEM = firm’s base year share of STEM-occ vacancies (NESTA 2015)
  - **Z** contains base-year firm controls – currently just firm size
  - K and T are SIC2 and year dummies, respectively

# Extensive margin

| | Any Cloud Computing Hire? | | | | Any ML/AI Hire? | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| % STEM vac. (2012) | 1.05*** | | 0.80*** | 0.71*** | 0.72*** | | 0.49*** | 0.56*** |
| | (0.076) | | (0.077) | (0.084) | (0.072) | | (0.081) | (0.097) |
| % High skill vac. (2012) | | 0.52*** | 0.33*** | 0.21*** | | 0.38*** | 0.26*** | 0.090** |
| | | (0.050) | (0.044) | (0.043) | | (0.071) | (0.090) | (0.036) |
| % Middle skill vac. (2012) | | 0.058 | 0.078** | 0.035 | | 0.0012 | 0.014 | -0.028 |
| | | (0.040) | (0.039) | (0.042) | | (0.029) | (0.029) | (0.041) |
| Year FE | x | x | x | x | x | x | x | x |
| Size (2012) control | | | | x | | | | x |
| SIC-2 control | | | | x | | | | x |
| Observations | 16695 | 16695 | 16695 | 16695 | 16695 | 16695 | 16695 | 16695 |
| ymean | | | | 0.34 | | | | 0.17 |

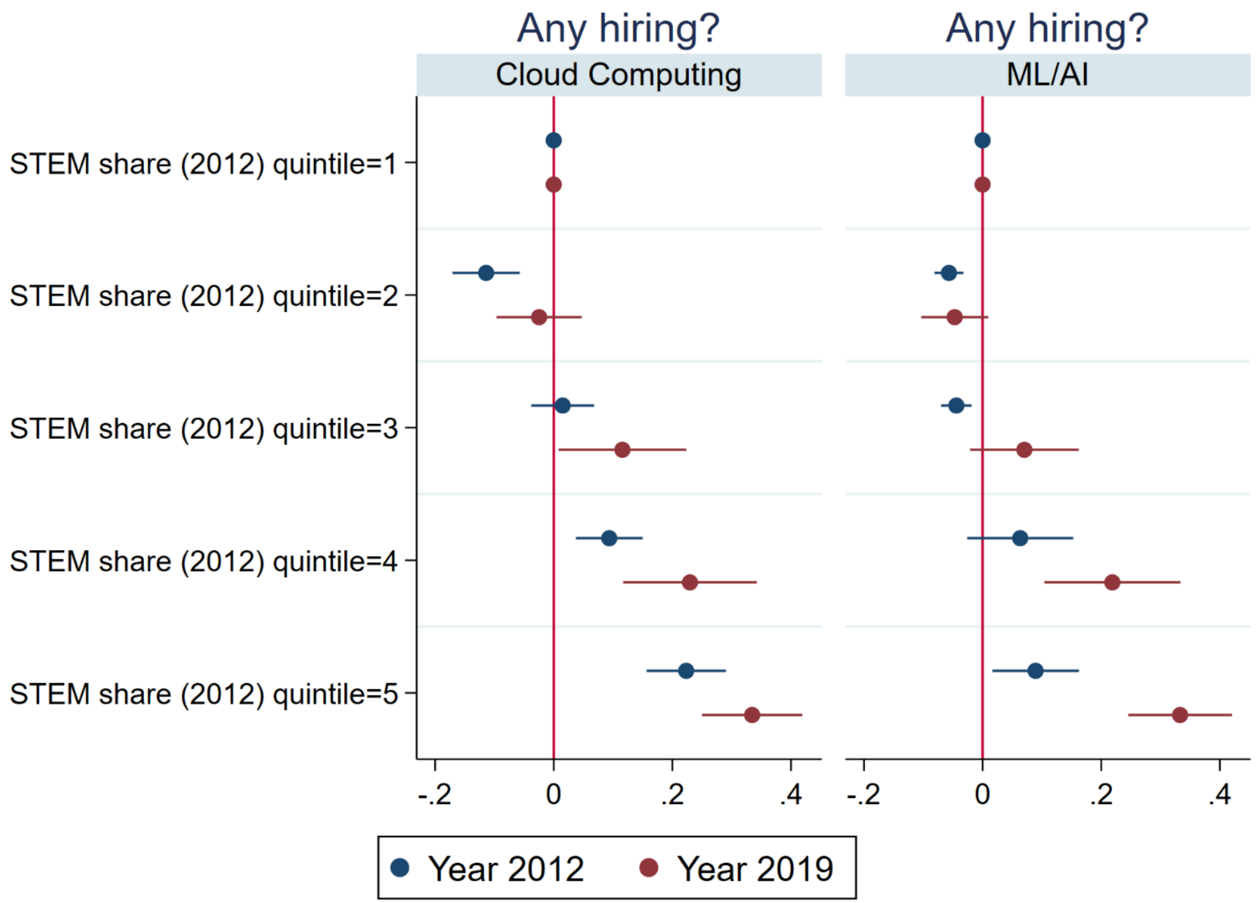
Robust standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01. Skill level is constructed by 1-digit SOC code with the reference group being labour skill level

# Intensive margin

| | Share of Cloud Computing Hires | | | | Share of ML/AI Hires | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| % STEM vac. (2012) | 0.09*** | | 0.09*** | 0.07*** | 0.02*** | | 0.02*** | 0.02*** |
| | (0.02) | | (0.02) | (0.01) | (0.003) | | (0.003) | (0.004) |
| % High skill vac. (2012) | | 0.03*** | 0.01*** | 0.010*** | | 0.01*** | 0.006** | 0.001 |
| | | (0.01) | (0.004) | (0.003) | | (0.002) | (0.002) | (0.001) |
| % Middle skill vac. (2012) | | 0.005 | 0.007** | 0.003 | | -0.0004 | 0.00010 | -0.001 |
| | | (0.003) | (0.004) | (0.004) | | (0.0009) | (0.0009) | (0.001) |
| Year FE | x | x | x | x | x | x | x | x |
| Size (2012) control | | | | x | | | | x |
| SIC-2 control | | | | x | | | | x |
| Observations | 16695 | 16695 | 16695 | 16695 | 16695 | 16695 | 16695 | 16695 |
| ymean | | | | 0.01 | | | | 0.004 |

Robust standard errors in parentheses. * p <0.1, ** p <0.05, *** p <0.01. Skill level is constructed by 1-digit SOC code with the reference group being labour skill level

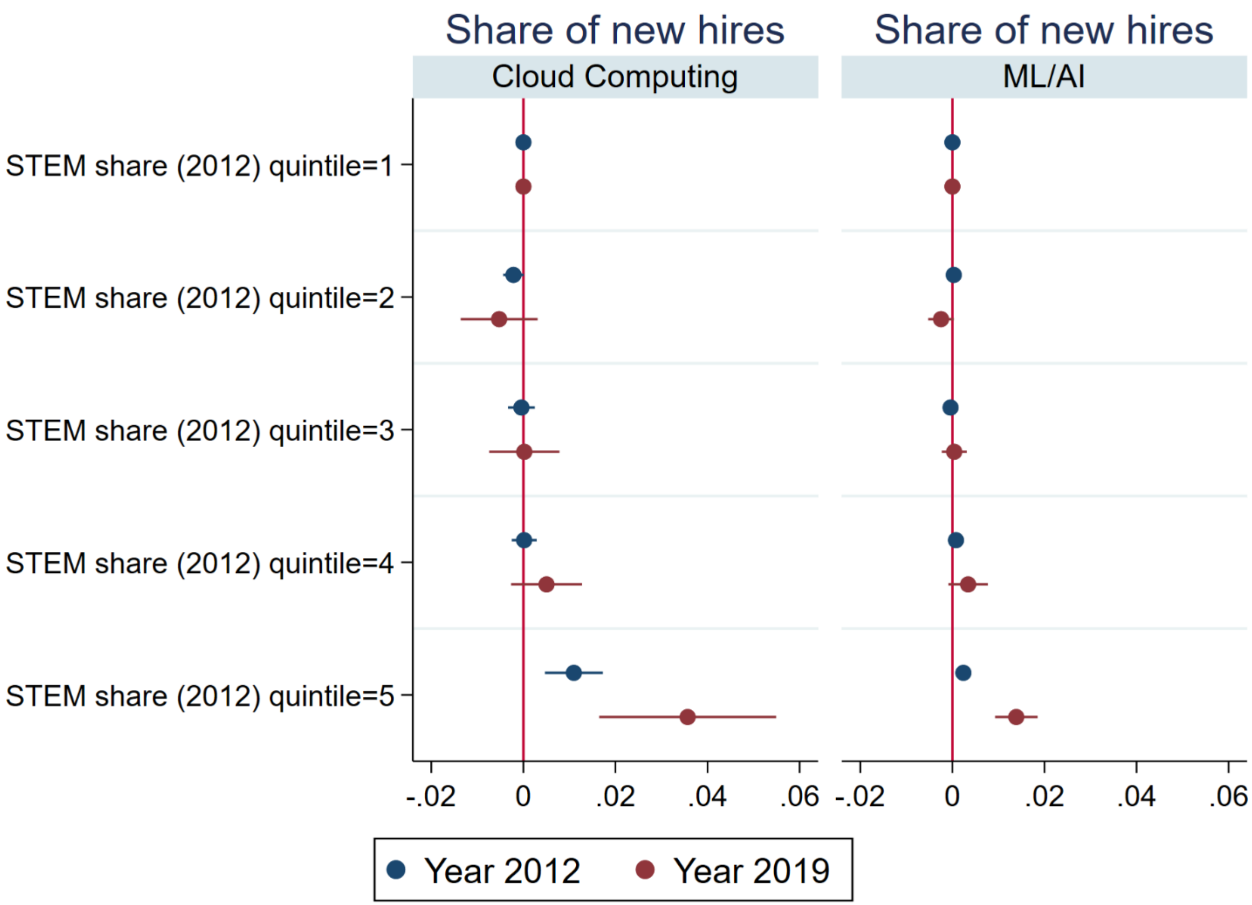
# Quantiles: extensive margin




The more STEM-intensive a firm in 2012, the more likely to make cloud, ML/AI hires in the following years

Cross-sectional regression of cloud / ML/AI hiring dummies on firms' STEM-intensity. STEM intensity is measured by share of STEM skills in job ads posted in 2012. Regressions control for initial firm size, SIC2 industry, skill contents

# Quantiles: intensive margin




Firms in the most STEM-intensive quintile in 2012, require more cloud and ML/AI jobs in the following years

Big difference in hiring between top quintile and the rest

Regression of shares of cloud / ML/AI hiring / all hires on firms' STEM-intensity. STEM intensity is measured by share of STEM skills in job ads posted in 2012. Regressions control for initial firm size, SIC2 industry, skill contents, year FE.

# Discussion

- Diffusion of ‘new wave’ technologies like cloud and ML/AI is very uneven: across occupations, space, and firms

- Skill-biased adoption, with Wave 2 more skill-biased than Wave 1
- Also: London location; intermediate technologies like superfast broadband; but not previous diffusion of PCs

- Firm heterogeneity: the most STEM-intensive firms more likely to adopt, adopt more, pull away from the rest

- Theory: firm/sector differences, localised learning, area complements, path-dependence/persistence. Our story finds roles for *all* of these!

# Planned next steps

- **Short term**
  - Run richer firm-area panel regressions [doing this now]
  - Improve STEM occupation measures; check STEM vs. managerial occs
  - Wave 1 firm-level analysis with KIBS dummies / industry STEM shares, which we can then compare to Wave 2
- **Longer term**
  - Identify exogenous skills supply [lags, IV or policy shifter?]
  - Test links from Wave 1, Wave 2 adoption to area wage inequality (graduate / non-graduate wage ratio)
  - Develop a framework for theorising STEM bias in adoption
- **We welcome your suggestions!**

# Thanks!

max.nathan@ucl.ac.uk
@iammaxnathan

# Appendix

[back]

# A1: Burning Glass data

- Burning Glass [BGT] scrape the UK's universe of job ads from a range of online sources
- **Raw data:** 59.9m UK-based, 2012-2019 with county/UA identifier
- **Cleaning:** standardise job text; remove generic phrases referencing cloud, ML/AI
- **Missing data:** employer name (66.8% coverage), SIC (67.1%)

[back]

# A2: Harte-Hanks data

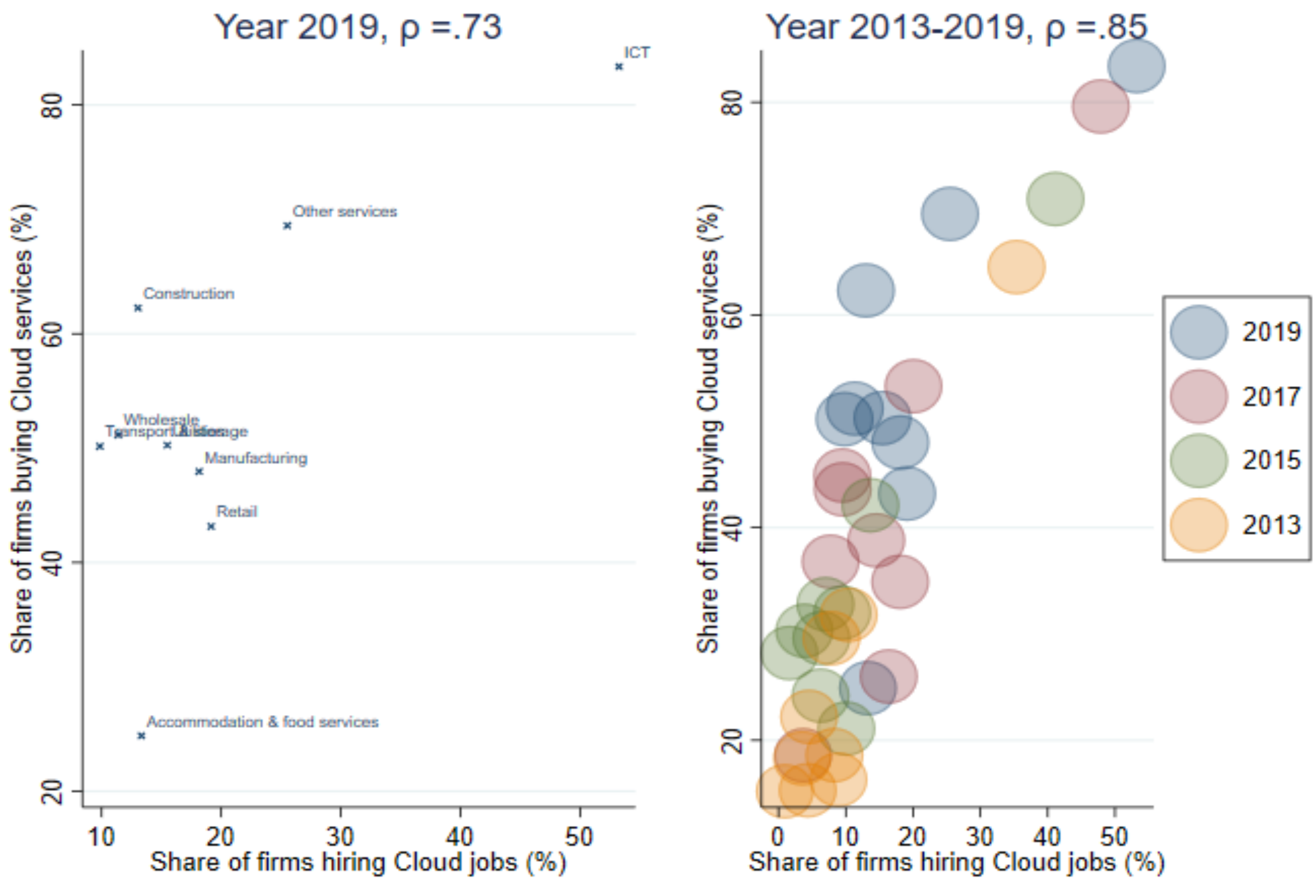
- We measure PC adoption using data provided by Harte-Hanks (HH), a multinational company
- The data is designed for the commercial use of large IT firms (e.g., IBM, Cisco, and Dell) (Bloom et al., 2015)
- It surveys establishments of large firms (with at least 100 employees across the country) on an annual basis
- We look at data for the UK in 3 years 2000-02 and map postcodes of establishments to TTWAs using the National Statistics Postcode Lookup (NSPL) crosswalk
- The variable of interest is PC per employee adjusted by size, industry and year fixed effects at the TTWA.

[back]

# A3: Other data

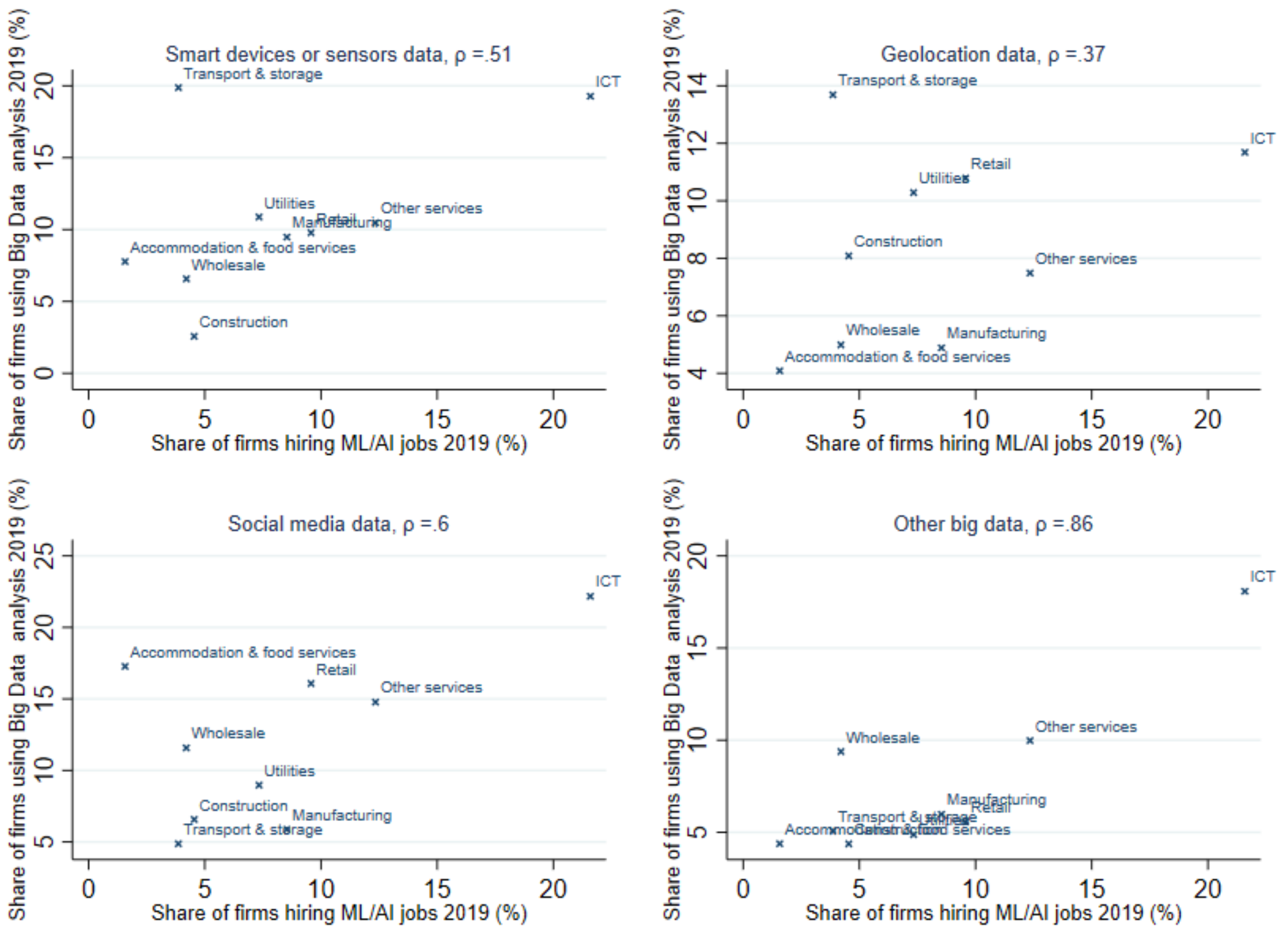
- **Human capital measures** - from 1991/2001/2011 Census
  - % graduates
  - % STEM workers (science and engineering professionals and associate professionals, following NESTA (2015))
- **Broadband** - 2011 LAD-level speed data from Ofcom
  - % superfast broadband connections / all connections
- **Current, very basic control set** from 1991/2001/2011 Censuses
  - Population density, 1991 and 2011
  - London dummy
  - Oxbridge dummy [Oxford + Cambridge TTWAs]
  - Share of manufacturing, 1991 and 2011
  - ILO unemployment rate, 1991 and 2011
- **Note:** microdata crosswalked to 2011 TTWAs using NSPD postcode-weighting crosswalk
- **Note: regression-adjusted dependent variables** also control for firm size, 3-digit industry and year

[BGT] [tracking] [ID]




**This figure compares the industry shares of firms buying Cloud Services reported in ONS E-Commerce and ICT activity surveys (vertical axes) and firms hiring Cloud Computing jobs computed from BGT (horizontal axes).** Correlation coefficients (ρ) are reported for 2019 (Left panel) and 2013/15/17/19 (Right Panel). We include firms with at least 10 employees / posting at least 10 online vacancies in each year. We assume firms in BGT can be identified by employer name and SIC code.

[BGT] [tracking] [ID]




**This figure compares the industry shares of firms using Big Data Analytics reported in ONS E-Commerce and ICT activity surveys (vertical axes) and the share of firms hiring ML/AI jobs computed from BGT (horizontal axes).** Correlation coefficients (ρ) are reported for 2019 for each category of Big Data Analytics. We include firms with at least 10 employees / posting at least 10 online vacancies in each year. We assume firms in BGT can be identified by employer name and SIC code.

[back]

| | All TTWA | | Estimation sample | |
|---|---|---|---|---|
| | mean | sd | mean | sd |
| PC per employee | 0.765 | 0.689 | 0.721 | 0.377 |
| Cloud Adoption 2019 | 0.010 | 0.007 | 0.010 | 0.007 |
| ML/AI Adoption 2019 | 0.002 | 0.003 | 0.002 | 0.003 |
| Share graduates 1991 | 0.062 | 0.021 | 0.062 | 0.021 |
| Share graduates 2011 | 0.249 | 0.049 | 0.250 | 0.049 |
| Share STEM 1991 | 0.038 | 0.015 | 0.039 | 0.014 |
| Share STEM 2011 | 0.049 | 0.016 | 0.050 | 0.016 |
| Pop. density 91 | 2.367 | 2.961 | 2.476 | 2.989 |
| Pop. density 11 | 2.687 | 3.386 | 2.895 | 3.471 |
| London | 0.004 | 0.067 | 0.005 | 0.070 |
| Oxbridge | 0.009 | 0.094 | 0.010 | 0.098 |
| Share Manuf.1991 | 0.167 | 0.060 | 0.172 | 0.057 |
| UR 1991 | 0.084 | 0.024 | 0.084 | 0.024 |
| Share Manuf.2011 | 0.099 | 0.036 | 0.100 | 0.035 |
| UR 2011 | 0.058 | 0.017 | 0.057 | 0.016 |
| Broadband | 0.393 | 0.312 | 0.366 | 0.287 |
| Observations | 226 | | 206 | |

Cloud and ML/AI Adoption rates per 1,000 vacancies. Share of Manuf.91/11 is the share of employed working in manufacturing in 1991/2011. UR91/11 is the unemployment rate in 1991/2011. Broadband is the share of super-fast broadband coverage in 2011.

[back]

# A6: Bloom et al method

- Bloom et al. (2021) propose a method to track the diffusion of 'Disruptive Technologies':
  - Intersect USPTO highly-cited patents (1976-2016) with company earnings call text (2002-2020)
  - Identify the ~300 patent bigrams most common in earnings calls, i.e. both scientifically and economically important
  - Use supervised machine learning to group these bigrams into 29 technologies
  - Validation using word2vec and human audit.

- Claim: these technologies reflect recent advances in innovation that largely impact businesses and employment within the last two decades.

[back]

# A7: Regression adjustment

- Following Beaudry et al (2010), we calculate TTWA-level PC and ML/AI intensity by regression to remove heterogeneity along establishment size, industry and year(after winsorizing the top and bottom 1% outliers).
- E.g. for PCs, we estimate for TTWA *j*, year *t*:

$$pcpe_{jt} = \Phi Ind_j * \o\ Size_j + \Psi Year_t + \Omega TTWA_j + \varepsilon_{jt}$$

- Where:
  - $pcpe_{jt}$ is the the count of PCs/establishment in TTWA *j* and year *t*
  - Ind, Size, Year and TTWA are vectors of dummies for SIC3 industry, establishment size (8 bins) year and establishment TTWA .