The Business-to-Worker Register

Linking all businesses to all workers in the UK

Arnaud Dyèvre
HEC Paris & CenTax

joint with
Arun Advani
Sebastian Gazmuri-Barker
Helen Hughson
Sanaya Mahajan
Andy Summers

ESCoE Conference
May 2026

HMRC disclaimer

This work contains statistical data from HMRC which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information.

Building a universal Business-to-Worker Register

- **The need**: No **matched employer-employee data in the UK**
  - With data on firms *and* workers
  - UK clear outlier among rich nations
- **What we do**: Use administrative tax returns on firms and workers to create a universal register of UK businesses and employers matched to their workers, over two decades (2002–2022)
- The **Business-to-Worker Register** (BWR)
  1. Covering the entire UK working population, each year
  2. Not only corporations or employers (also partnerships, sole prop., NPOs, gov.)
  3. Not only employees (also partners, directors and sole prop.)
  4. Derive economically relevant variables from tax records: y_{it}, π_{ft}, I_{ft}, K_{ft} and δK_{ft}

→ We work with full populations, not samples

What makes the BWR possible

1. **Business IDs** — HMRC lookup tables link PAYE, VAT, CT600, and Companies House IDs over time
2. **Person IDs** — UTRs, NINos, and TRNs reconciled into one anonymised individual spine across datasets and years
3. **One secure environment** — the HMRC Datalab hosts the raw tax records used to link, harmonise, and enrich both sides

→ The BWR retains **all jobs within a year**, not just main job.

How the BWR sits among (selected) LEEDs

Country	Project	Years	Businesses	Workers	Distinctive strengths
UK	**BWR**	2002–2022	**7.0m** all types (corp., pship., sole prop., gov., NPOs)	**41.7m** emp., partners, directors, self-emp.	Universal coverage; non-residents; detailed worker variables; financial-accounting variables for non-incorp. firms
UK	**ESCoE LEED**	2014–2025	**IDBR** universe	**24m** employees	Near-real-time access; linkable to ONS products (BICS, LEO, LFS)
US	**LEHD**	1990s–	**6.2m** employers	**120m** UI-covered	Long history; full economy excl. self-employed and independent contractors
SE	**LISA / RAMS / FEK**	1985–	**1.8m**	**5.1m** one job per worker	Detailed demographics; firm balance-sheet data; researcher gold standard

Why non-employment work matters

15%

of UK workers are in non-employment relationships in 2021

66%

of private businesses have at least one non-employment relationship

- Standard employer–employee data miss **partners, sole proprietors, directors, and owner-managers**
- Important for top incomes and dynamics of inequality

  Advani & Summers (2024); Advani, Summers & Tarrant (2023); Miller, Pope & Smith (2024); Delestre et al. (2024)
- Concentrated where legal form, tax treatment, and labour supply interact

1. Building the BWR

Inputs

Individuals	NINo or UTR
Self Assessment	*97–22*
PAYE employees (P14)	*02–14*
PAYE employees (RTI)	*15–22*
Emp. 10% samples (COP, NPS)	*01–14*
Reported industries	*SIC codes*

Links	firm–worker and firm–firm
BLT CT600↔VAT↔PAYE↔CRN	*90–25*
IDBR VAT↔PAYE↔CT600↔FAME	*90–25*
Partners ↔ partnerships	*97–22*
Own link emp. ↔ partnerships	*97–22*
Person-ID crosswalk	UTR↔NINo↔TRN

Firms	CT600, SA800, UTR, CRN, PAYE
Corporations (CT600)	*01–23*
R&D tax credits	*01–23*
Partnerships (SA800)	*01–21*
Sole proprietors (SA103)	*97–22*
Partners (SA104)	*97–22*
VAT	*05–23*
FAME / ORBIS	*01–23*

Other	
Geography	*FAME*
Patents	*ORBIS IP*

→ All fiscal datasets are accessed through the HMRC Datalab;
any box on a tax return is reachable if the use case is justified.

From spines to jobs

2 spines

1. Individual spine

UTR, NINo and TRN records → one anonymised person ID over time.

2. Organisation spine

Tax identifiers and lookup tables → one organisation list.

5 job types

Job type	Main source	Link type	Notes
Employee	P14 / RTI / SA102	Administrative	Pension schemerefs filtered
Sole proprietor	Self Assessment	Person = business	Employee PAYE links incomplete
Partner	SA104 / SA800	Administrative	Sources overlap imperfectly
Director	SA102 + Directors Spine	Partial administrative	Some unremunerated directors omitted
Owner-manager	Directors + dividends	BWR classification	Depends on close-company status

Outputs

Business Register 2001–2022

Business outcomes each fiscal year

- Reconciliation of tax & financial accounts
- Turnover, profits (EBIT/EBITDA), investment, capital stock, depreciation, intermediate inputs
- R&D, patents, industry (SIC), geography (output area, TTWA), employees & wage bill, partner counts, trading and VAT status

Soon: ownership structure (Orbis).

Business-to-Worker Register 2002–2022

Who works for whom

- Multiple jobs per worker
- Total, earned & investment income; nine income types (employment, self-employment, partnership, pensions, dividends, property, savings, social security, other); realised capital gains; taxes & reliefs
- Age, sex, residence (output area), non-resident status, country of origin, migrant status

Soon: migrant visas.

2. Benchmarking against official statistics

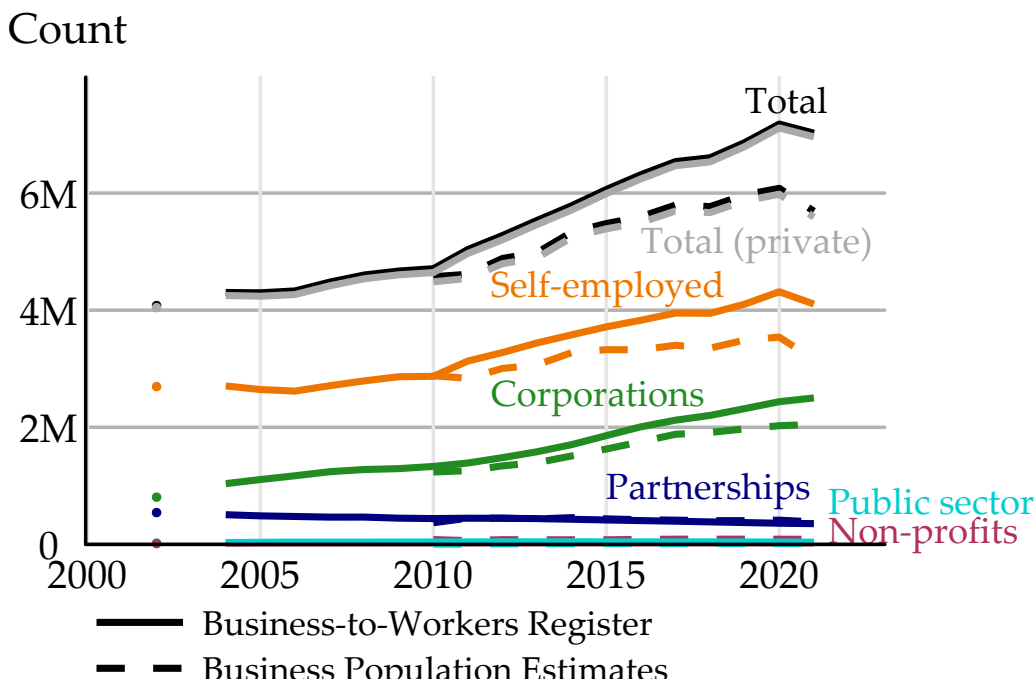
Business demography: BPE (DBT)

Approach	**Estimates** of active businesses, 2010–2024
Built from	VAT, PAYE & Corporation Tax **registrations** (IDBR)
Worker counts	**Extrapolated** from the Labour Force Survey
Outputs	Business & worker **estimates**; **jobs**, not workers
Activity window	Active **on 1 January** each year
"Active" means	A **registered** business

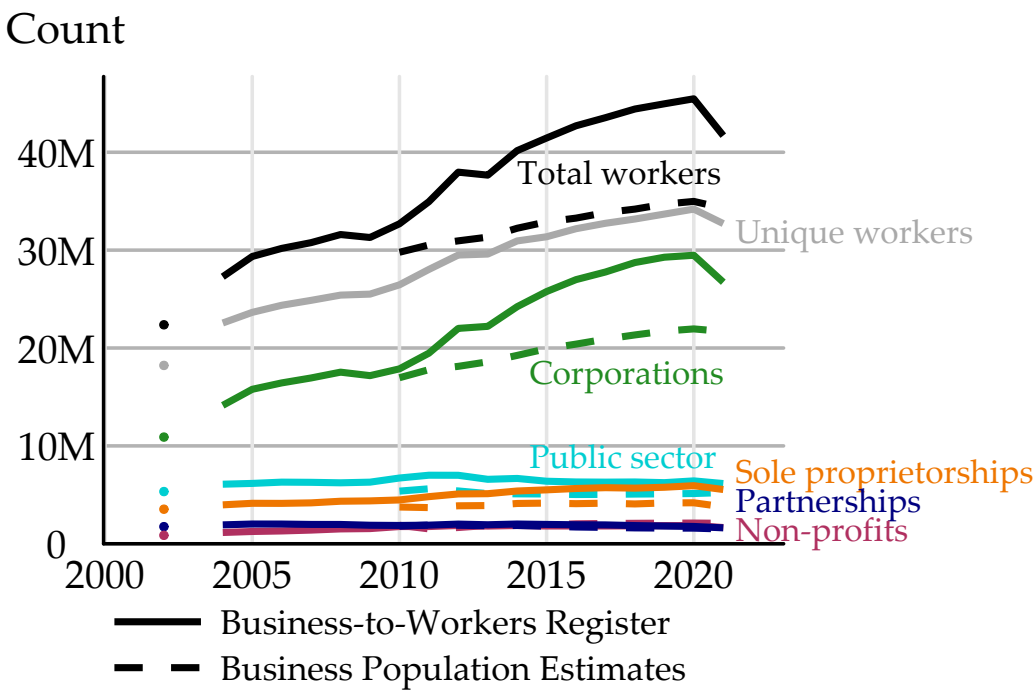
Business demography: **BWR**

Approach	**Counts** of active businesses, **2002–2022**
Built from	**All tax returns** filed by businesses + PAYE IDs of non-businesses
Worker counts	~~Extrapolated from the Labour Force Survey~~ **observed directly**
Outputs	Business & worker **counts**; jobs **and** workers
Activity window	Active at **any point** in the tax year
"Active" means	A business that **has revenue or workers**

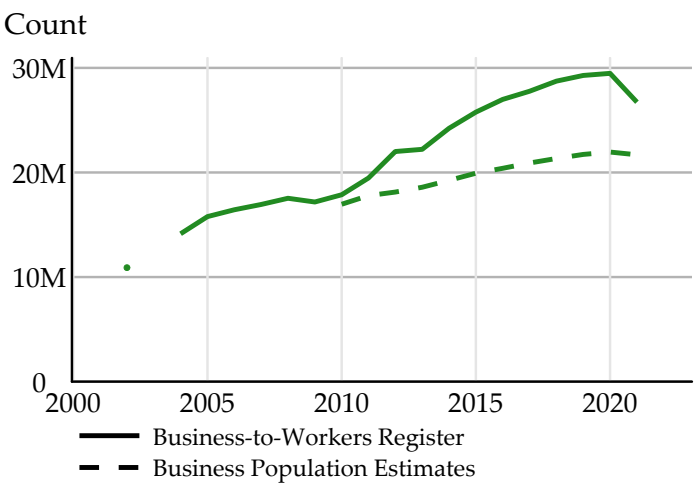
Organisation counts: BWR vs. BPE



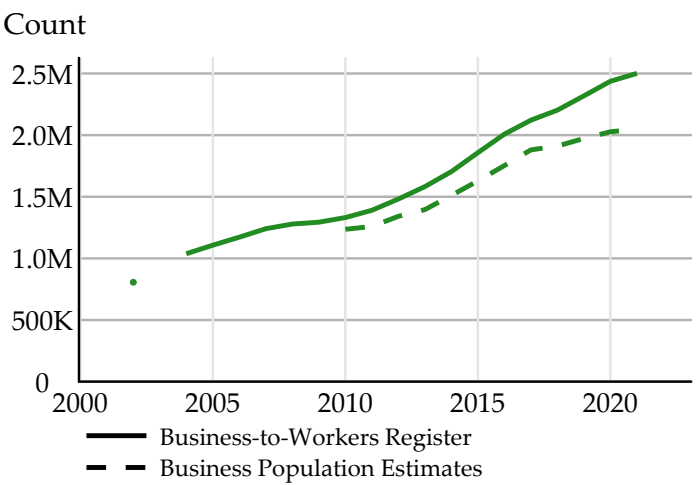
Worker counts: BWR vs. BPE



Corporations

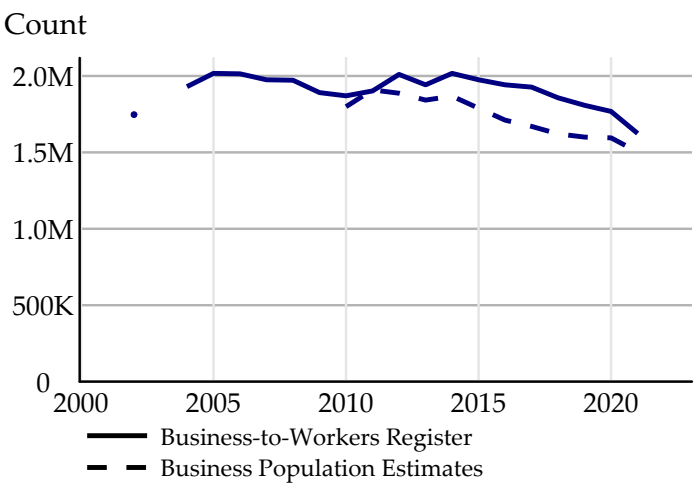


Workers

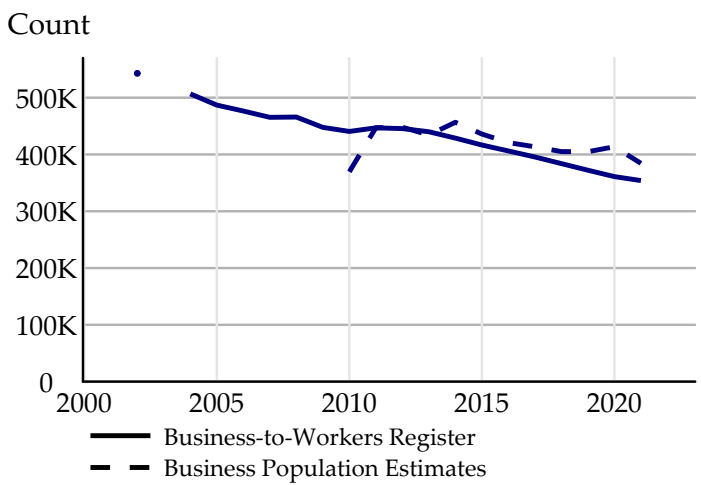


Organisations

Partnerships



Workers



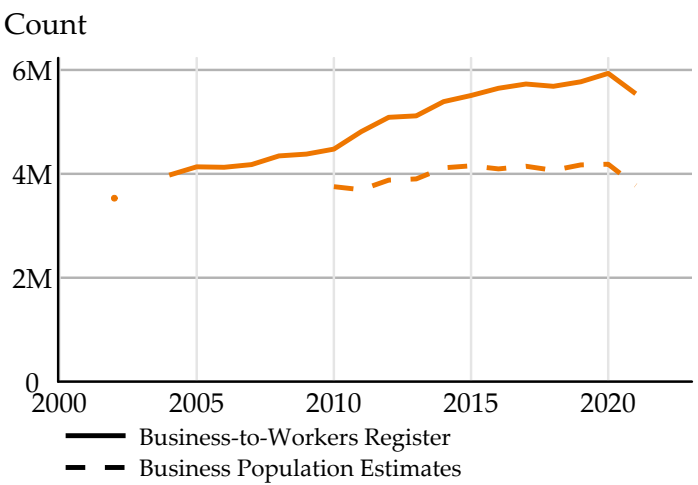
Organisations

Non-profits

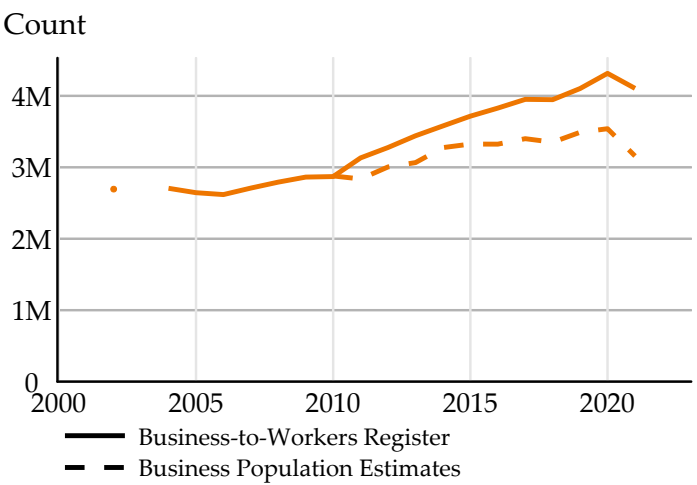
Workers

Organisations

Sole proprietors



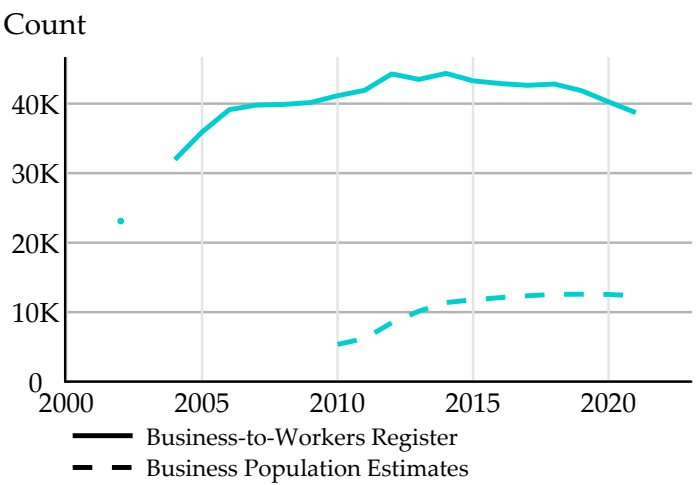
Workers



Organisations

Central & local government

Workers



Organisations

Sources of discrepancies

1. **Labour Force Survey**: declining response rates (noisier BPE)
2. **Extrapolation** vs. **direct observation** (noisier BPE)
3. **Jobs** vs. **individuals** (more in the BWR)
4. **Registration**: all CT600 filers, not only PAYE/VAT (more in the BWR)
5. Activity at **any point**, not the 1 Jan snapshot (more in the BWR)
6. Stricter **"active"**: non-zero turnover or wage bill (less in the BWR)
7. **Disaggregated IDs**: public bodies split by payroll (more in the BWR)

3. Constructing and validating firm-level variables

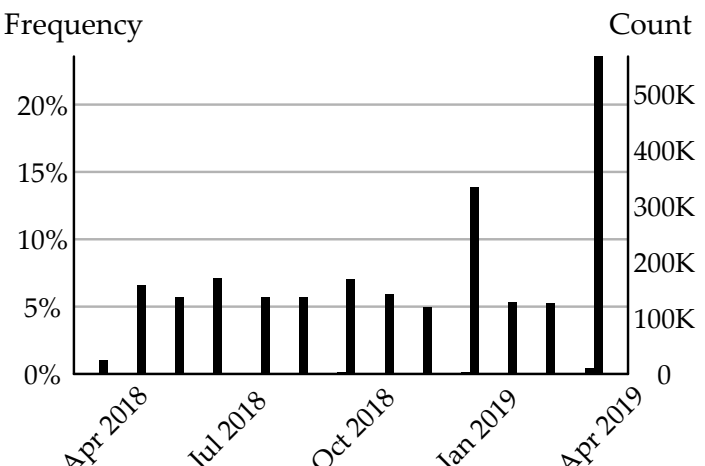
From tax returns to π_{ft}, I_{ft}, K_{ft} and δK_{ft}

Challenges:

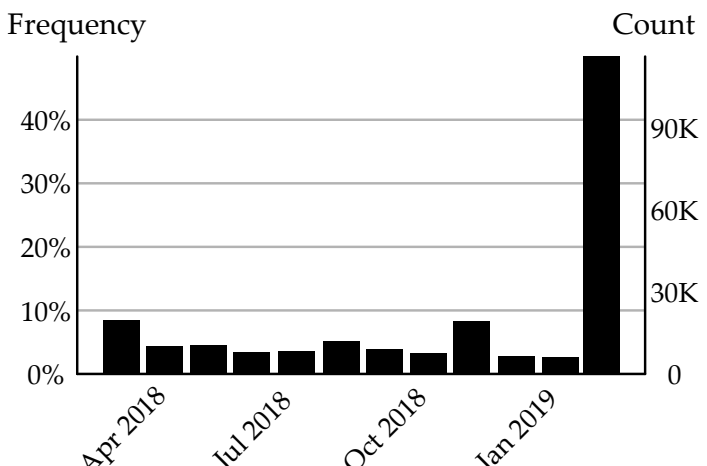
1. π_{ft}: taxable profits on tax returns $\neq$ financial-accounting profits
2. δK_{ft}: tax depreciation rate $\neq$ true capital depreciation
3. I_{ft}: no investment reported (only capital allowances & balancing charges)
4. K_{ft}: no capital stock reported $\rightarrow$ estimated by perpetual inventory
5. Different reporting periods across firms (need not fall neatly in the April–March tax year)

Reporting periods are not aligned

End-of-period dates in tax returns (tax year 2020): about one quarter of corporations file on the tax-year cycle; most do not. We **apportion** each variable to the April–March tax year.



Corporations (daily)



Partnerships (monthly)

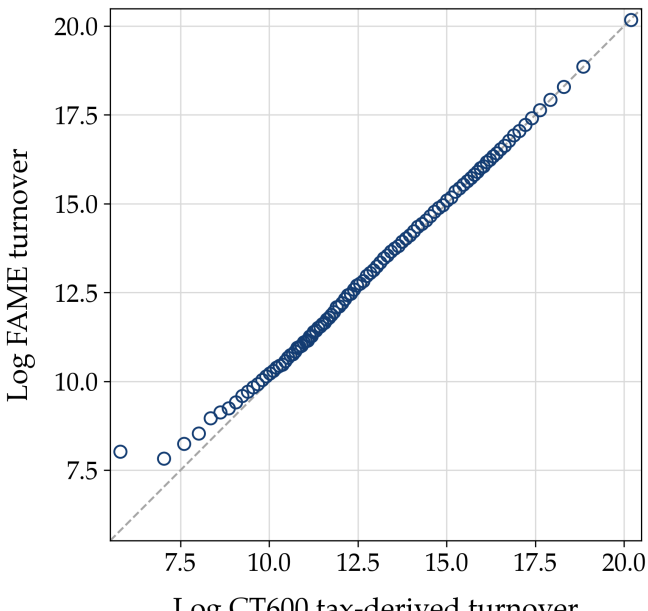
From tax returns to π_{ft}, I_{ft}, K_{ft} and δK_{ft}

Solution: reverse-engineer π_{ft}, I_{ft}, K_{ft} and δK_{ft} by:

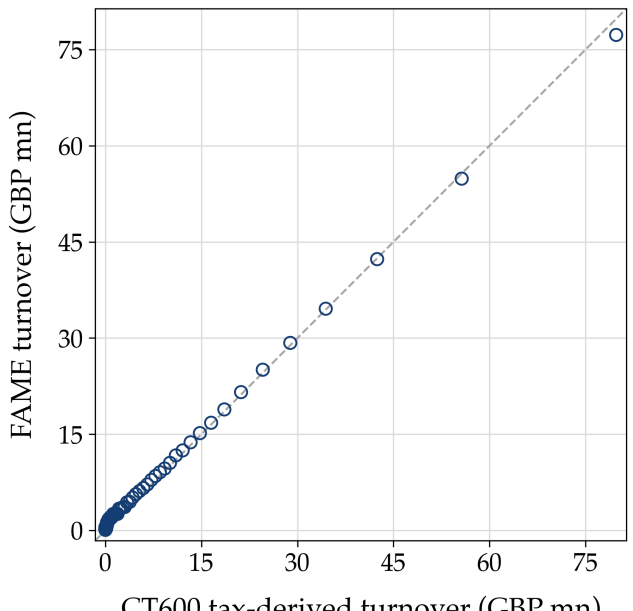
- Homogenising reporting periods, so t = tax year
- Extracting I_{ft} from capital allowances and balancing charges
- Estimating K_{ft} by initialising and capitalising it:
  - Initial value from capital allowances and balancing charges
  - Specific δ's for different types of K_{ft}
  - I_{ft} from the step above
- Estimating π_{ft} by purging capital allowances & balancing charges and applying a proper δK_{ft} (EBIT & EBITDA-style)

$\rightarrow$ Quality assessment: compare our variables to FAME ("ground truth").

Corporation turnover



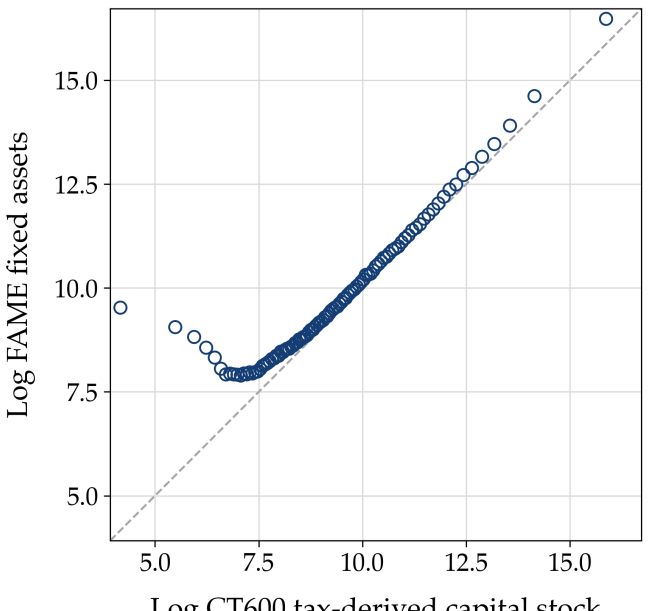
Log scale



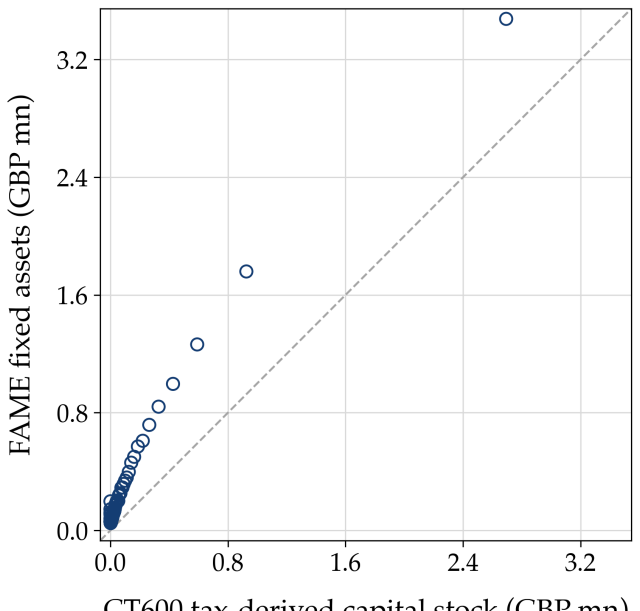
Levels

CT600 (horizontal) vs. FAME (vertical); 100 equally-sized bins; 45-degree line.

Corporation capital stock



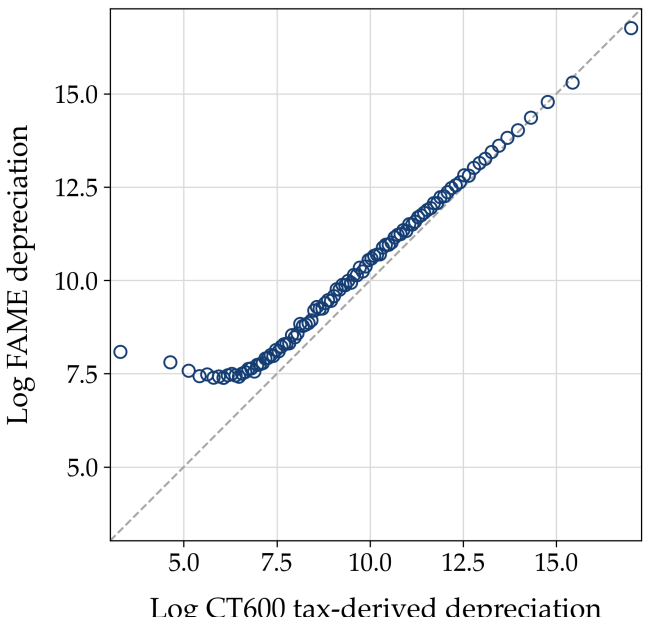
Log scale



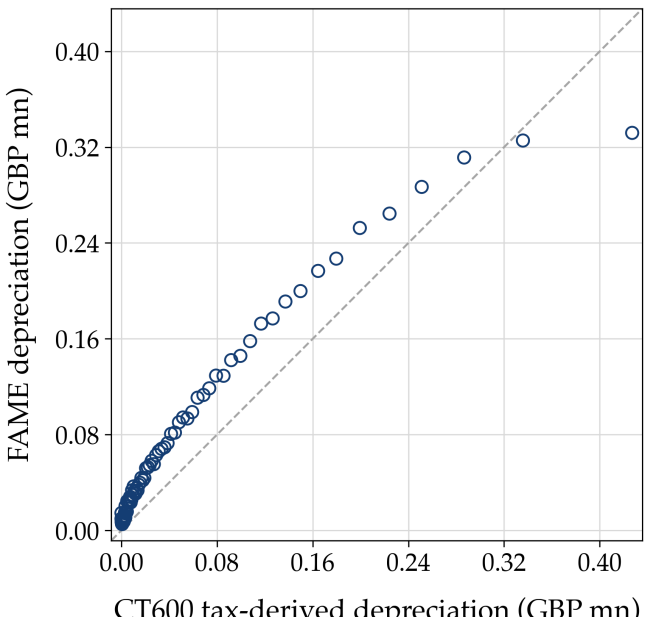
Levels

Tax-derived capital stock (capitalised from CT600 investment flows) vs. FAME fixed assets.

Corporation depreciation



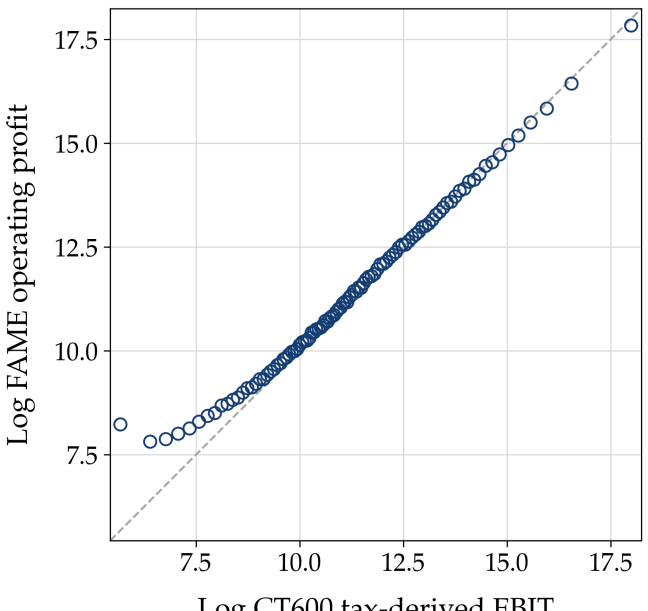
Log scale



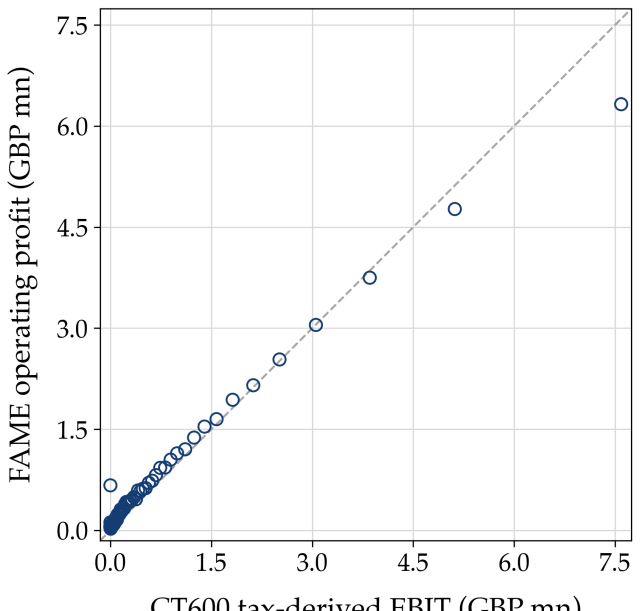
Levels

Tax-derived depreciation vs. FAME depreciation benchmark.

Corporation EBIT



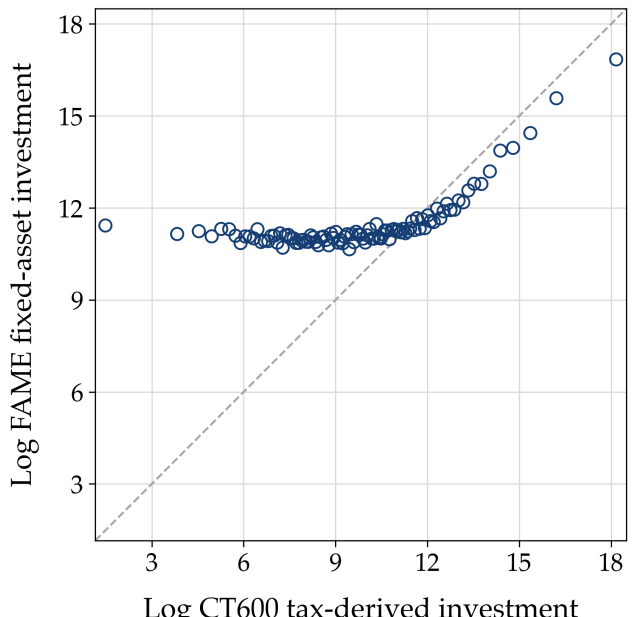
Log scale



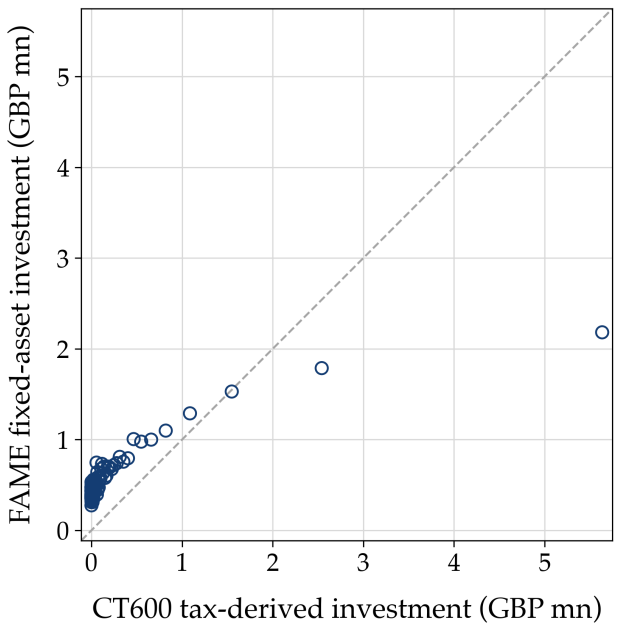
Levels

Tax-derived EBIT vs. FAME operating-profit benchmark; log scale excludes loss-makers.

Corporation investment



Log scale



Levels

Tax-derived investment vs. Orbis fixed-asset additions.

Partnership benchmarks →

4. Going forward and Conclusion

From one link to an economic network

Layer 1: work

BWR today
Worker ↔ organisation links, with worker and business outcomes.

Layer 2: ownership

Next
Shareholder and corporate-group links from Orbis ownership and HMRC group data.

Layer 3: households

Next
Links between people sharing resources and making joint economic decisions.

Additional enrichments: migrant visa data; better unincorporated-business payroll links; international trade; consumer prices; matched NHS–PAYE hospital records.

Conclusion

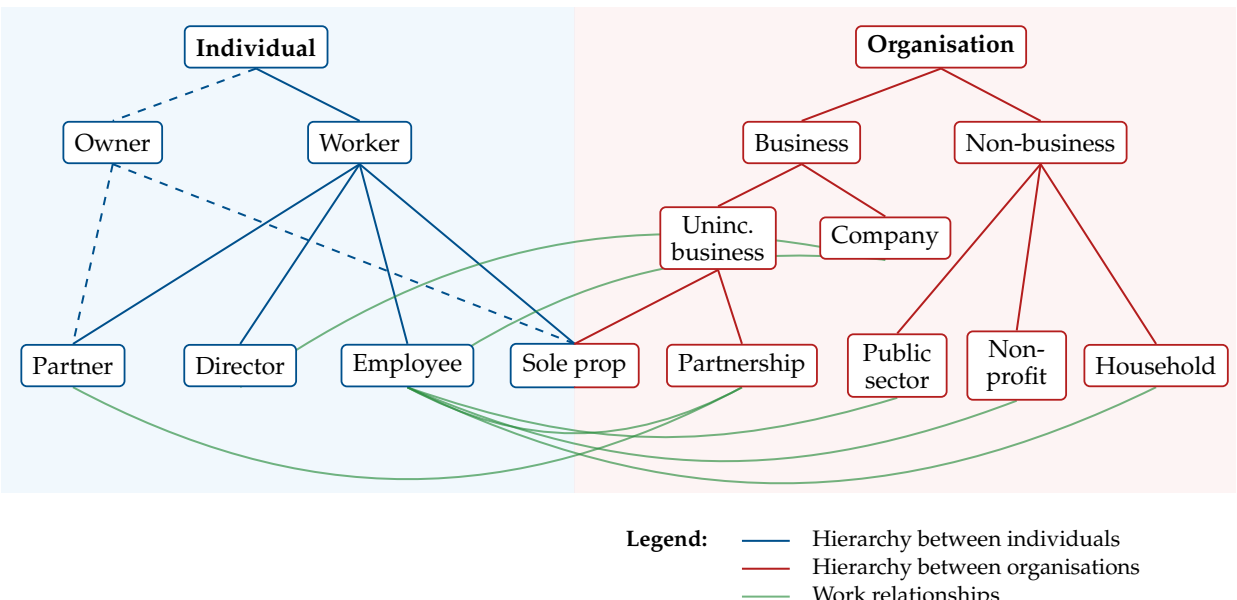
1. A **universal** matched workers-to-businesses register for the UK
   - **All** workers, **all** businesses — not only employers and employees
2. **Vast array** of business- and worker-level variables, derived from tax forms
3. Designed to **complement and enhance official statistics**
4. Many applications — firm size distribution, dynamics, disaggregated growth, migrant impacts, spillovers

Thank you.

dyevre@hec.fr a.advani.1@warwick.ac.uk a.d.summers@lse.ac.uk

Appendix

Individuals and organisations in the BWR



Detailed source data

7 business tax filings

CT600; R&D tax-credit claims; VAT100; SA800 main pages; SA800PS; SA801; SA802.

3 linkage tables

Business Lookup Table; legacy IDBR; person-identifier crosswalk across UTR, NINo, and TRN.

7 individual tax filings

RTI; COP; NPS; P14; SA102; SA104; Self Assessment Valid View extracts.

4 external inputs

FAME; Orbis IP; ONS industrial-sector crosswalks; ONS geography crosswalks.

21 raw data types, disaggregated into annual extracts over the BWR coverage window.

Individual spine vs. SPI

SPI comparator

HMRC's Survey of Personal Incomes is a stratified sample drawn from SA and PAYE records for income-statistics production.

BWR individual spine

The BWR builds a full person-year spine across UTR, NINo, and TRN identifiers, then links all available tax records.

- **Late SA filers**: BWR uses custom extracts beyond the Valid View cut-off.
- **SA/PAYE de-duplication**: BWR reconciles different identifier systems directly.
- **Additional records**: people seen in other HMRC sources can remain in the spine even before they link to a job.

Job-link source matrix

Relationship	Main source	Observed or built?	Main limitation
Employee	P14 / RTI / SA102	Administrative	Pension schemerefs must be filtered; P14 missing in 2003
Partner	SA104 / SA800	Administrative	SA104 and SA800 overlap imperfectly
Sole proprietor	SA Valid View	Identity link	Owner observed; employee PAYE link still incomplete
Director	SA102 + Directors Spine	Partial administrative	Unremunerated directors observed but omitted from worker totals
Owner-manager	Directors + close-company and dividend information	BWR classification	Depends on close-company status and income composition
Off-payroll / inside IR35	PSCs, umbrellas, client links	Not yet included	Client-side relationship unavailable in current Datalab sources

Partnership-to-PAYE schemeref links for employees are inferred from partner payroll records and career histories.

Why FAME is not full ground truth

Public filing gaps

- Micro-entity category starts in 2013
- Small companies generally do not file public P&L accounts
- Turnover and operating profit are therefore missing for much of the lower tail

What remains useful

- Medium and large companies provide stronger P&L benchmarks
- Fixed-asset notes support capital-expenditure and depreciation checks
- Filing thresholds help interpret lower-tail deviations

Benchmark gaps can reflect **Companies House reporting rules**, not only errors in tax-derived BWR measures.

Known limitations

- No education or occupation variables
- No reliable hours worked series
- No untaxed benefits, except where taxable payments enter HMRC data
- No informal, grey-market, or black-market work
- No household links yet
- No systematic corporate-group aggregation yet
- Some balance-sheet variables are inferred
- Sole-proprietor employee links and off-payroll client links remain incomplete

Business Register: beyond accounting variables

Industry and geography

SIC2007 codes, worker-based output areas, travel-to-work areas, and postcode-derived locations.

Innovation

R&D tax-credit claims, qualifying expenditure, patents, claims, and forward citations.

Business dynamics

Activity flags for wage bills, VAT turnover, CT600 outcomes, FAME status, and legal-form-specific signals.

Intermediate inputs

VAT Box 7 purchases and other inputs, with VAT-registration links apportioned to tax entities where needed.

Four key challenges

1. Disambiguating individual IDs

NINOs, UTRs, TRNs → **one consistent ID** over time. (previously solved by CenTax)

2. Extracting employer–employee links

From **P14 forms** (2002–2014), where the firm–worker linkage was not previously exploited.

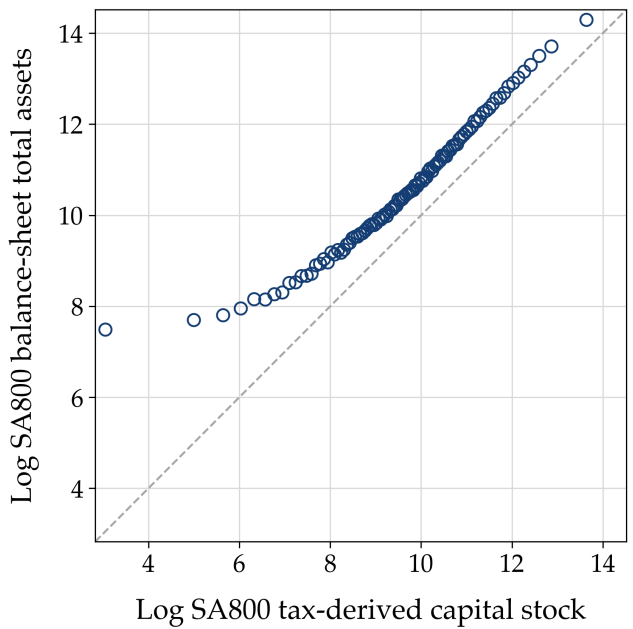
3. Linking partnerships to employees

No native link between `schemeref` and partnership ID.
→ Infer from **partner career trajectories** (earlier as employee, later as partner).

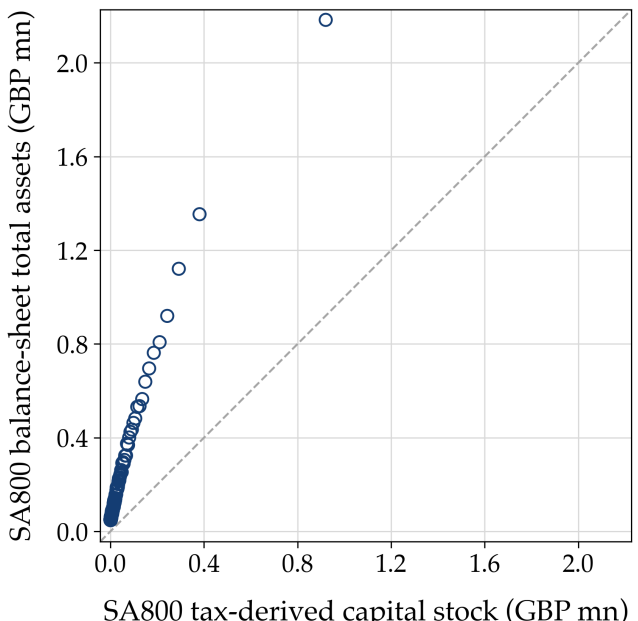
4. Tax data → economic variables

Taxable profits → π_{ft};
capital allowances → $I_{ft}, K_{ft}, \delta K_{ft}$.

Partnership capital



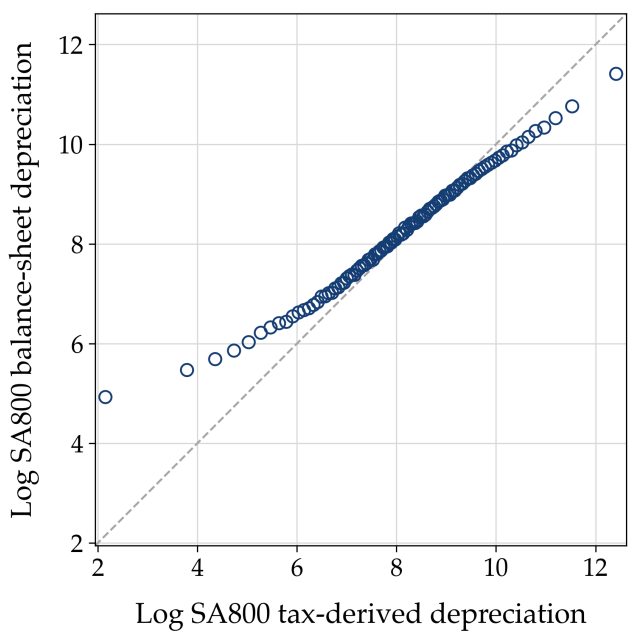
Log scale



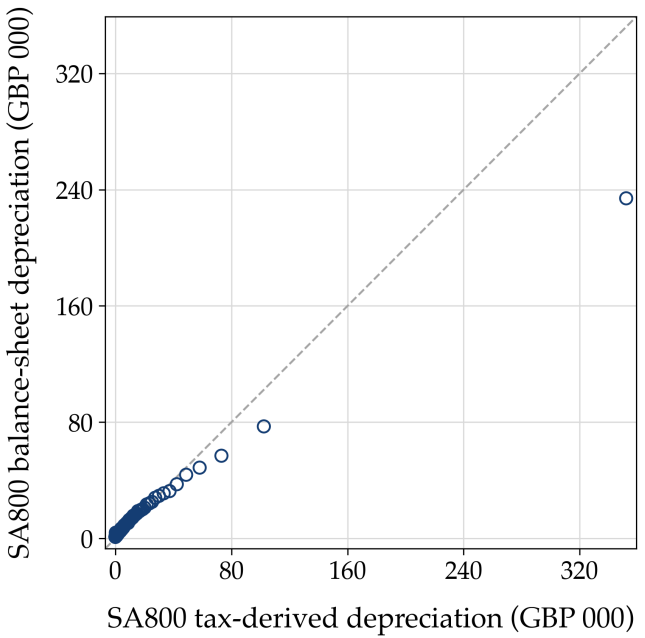
Levels

SA800-derived partnership capital vs. balance sheet total assets.

Partnership depreciation



Log scale



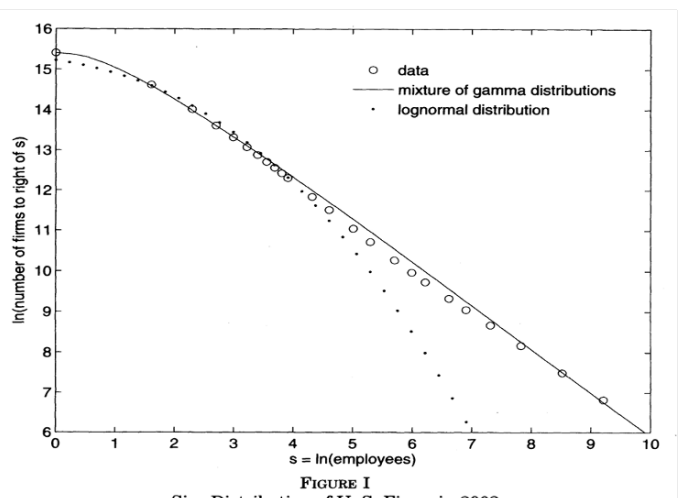
Levels

Tax-derived partnership depreciation vs balance-sheet depreciation.

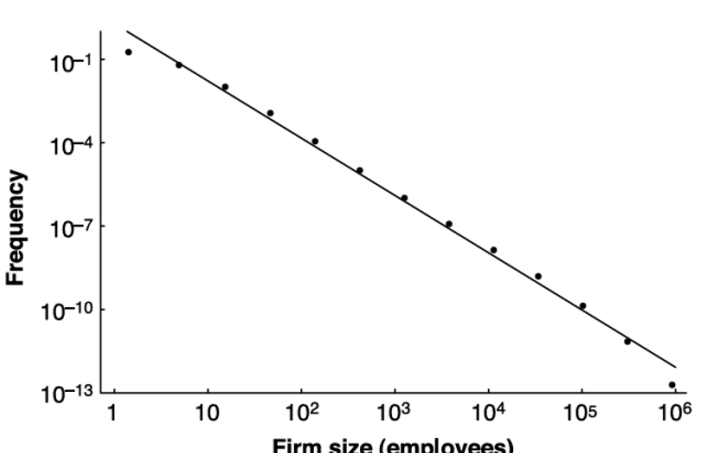
← Back to corporation benchmarks

Firm size distribution — what the literature finds

- Key empirical and theoretical object (macro, public economics, IO, trade)
- Estimation complicated by **partial coverage** of existing datasets
- Typical result: mixture of **log-normal** and **Pareto** distributions



Luttmer (2007)



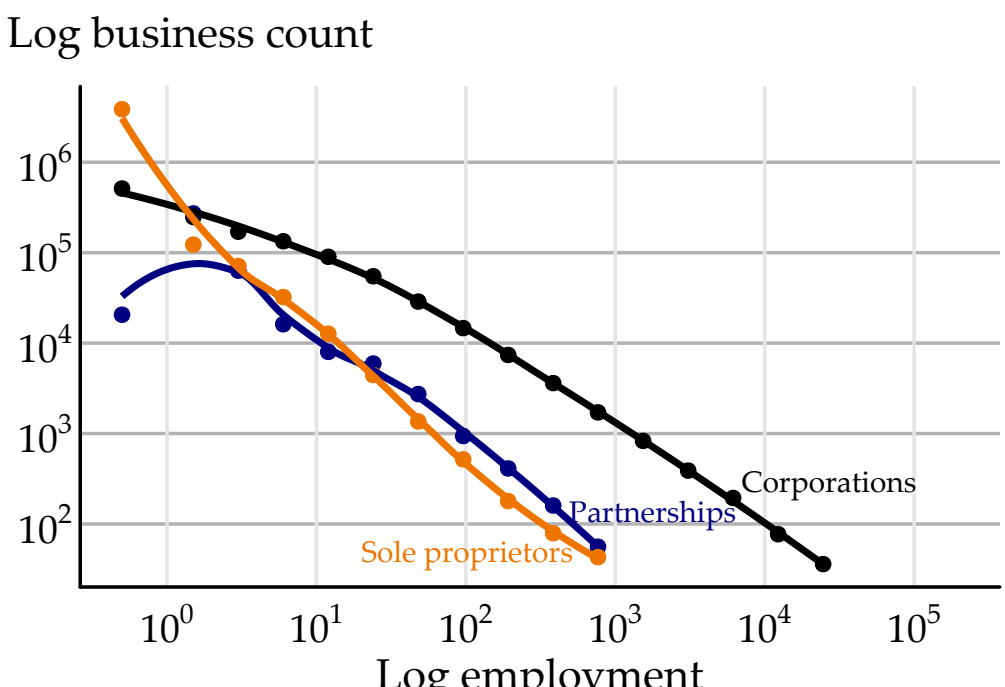
Axtell (2001)

Firm size distribution — Employment

With the BWR:

- Full coverage of small businesses
- Direct measure (no zombie firms)
- Breakdown by business type

→ **Pareto throughout.**

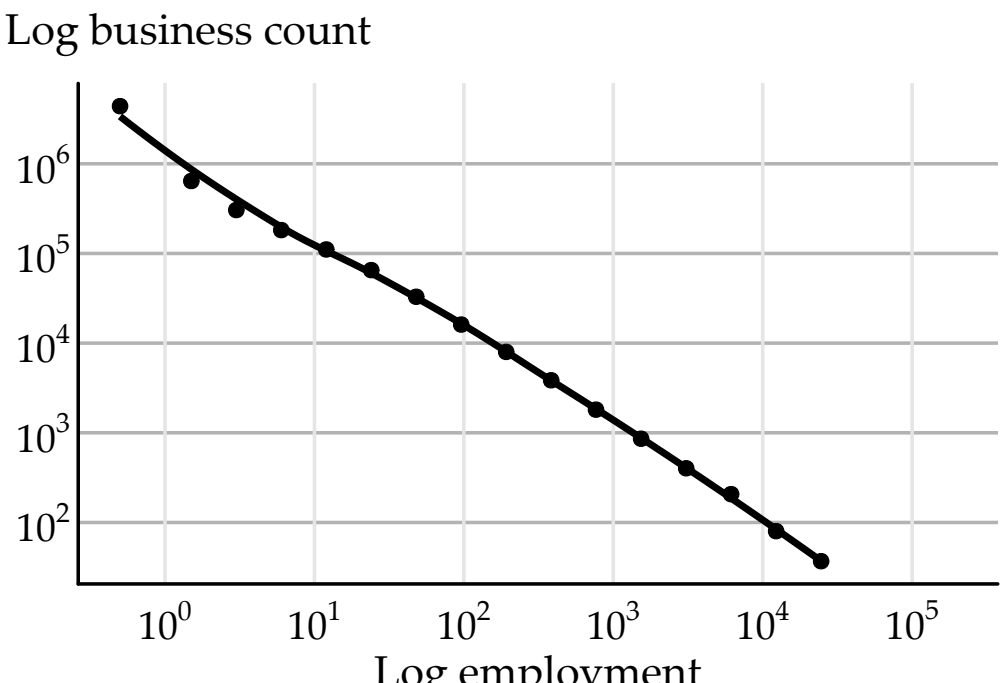


Firm size distribution — all businesses

Pooling all business types:

- Corporations, partnerships, and sole proprietors combined
- The "missing mass" of small businesses is restored

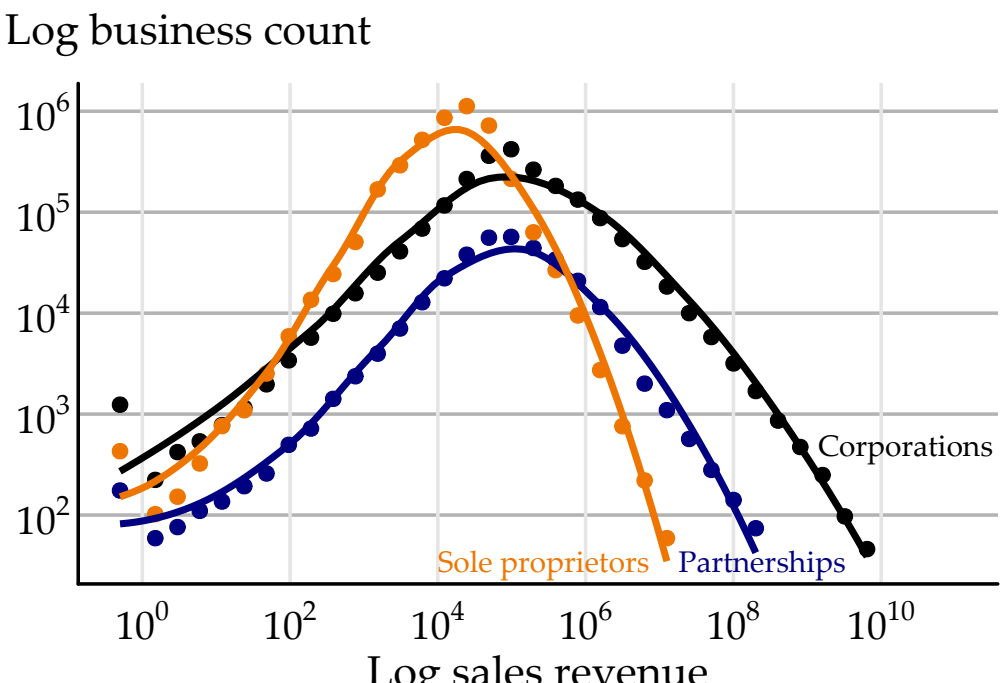
→ A single Pareto law across the whole size range.



Firm size distribution — Sales

With the BWR:

- Full coverage of small businesses
- Direct measure (no zombie firms)
- Breakdown by business type

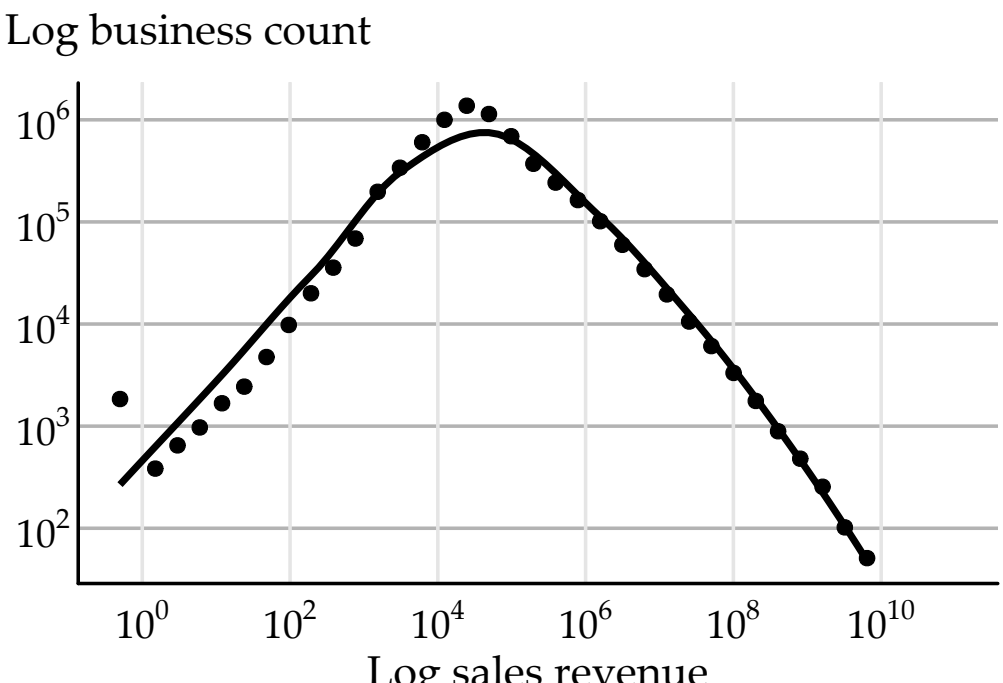


Firm size distribution — Sales, all businesses

Pooling all business types:

- Corporations, partnerships, and sole proprietors combined
- The "missing mass" of small businesses is restored

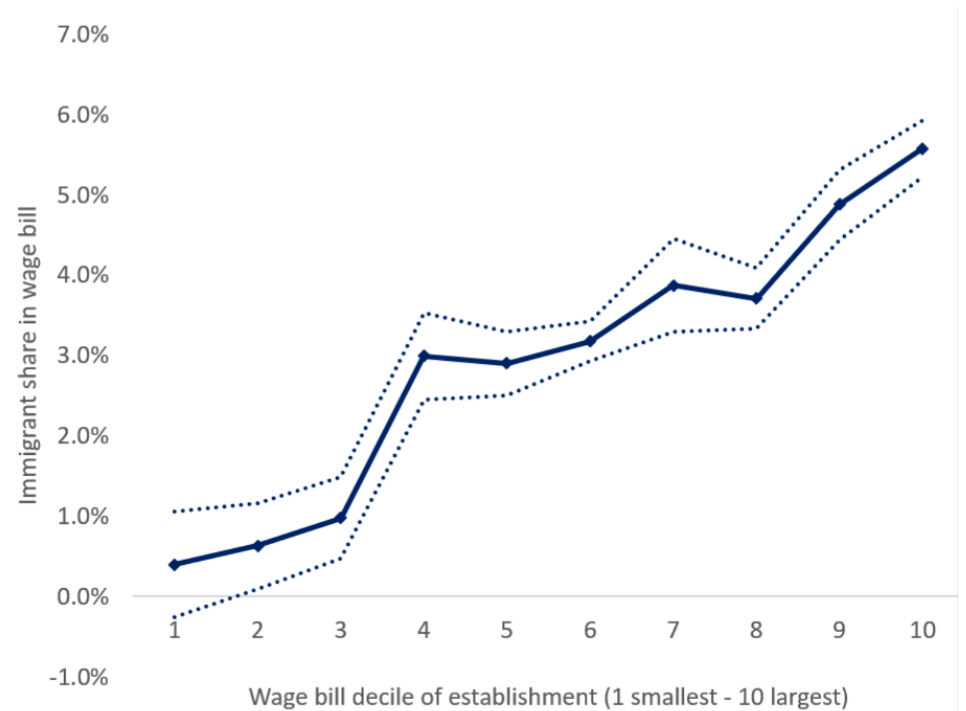
→ Pareto upper tail, with a log-normal-shaped body.



Migrants and firm size — precedent

Immigrant share in the wage bill of German firms, by wage-bill decile (firms with 10+ employees).

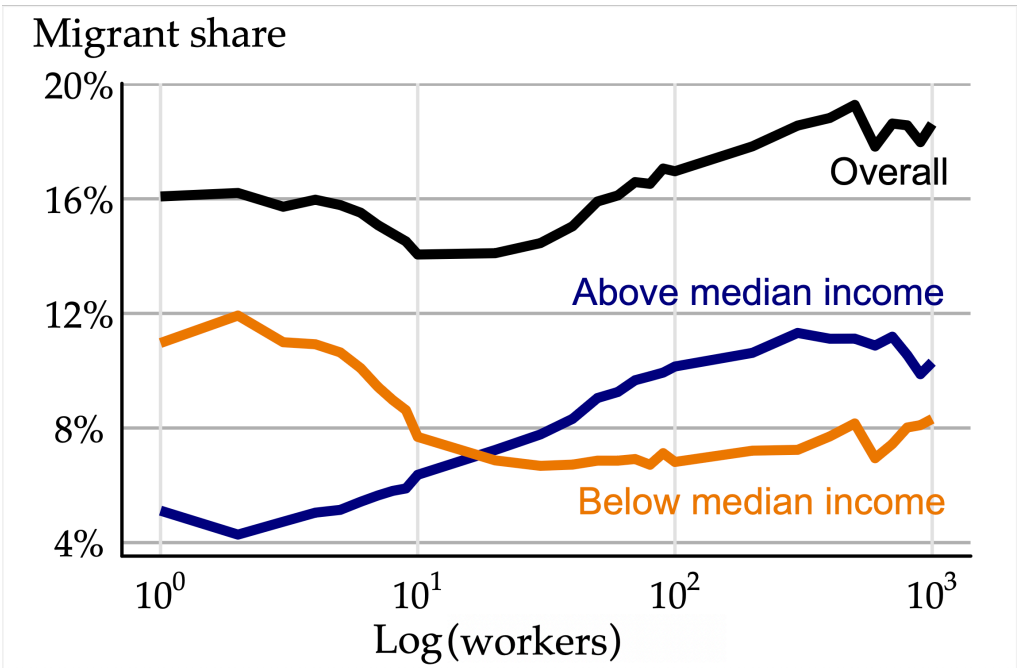
Brinatti & Morales (2023), Figure 1. Robust to industry and geography controls.



Migrants and firm size — BWR

Our result

Migrant share of workers by firm size (employment), all businesses. Splits above-/below-median income.



Joint distribution: worker income × firm revenue