Data Methodology

Product Market Concentration and Productivity in the UK

Oluwaseun Aguda, Ioannis Galanakis,
Kyung In Hwang, Anthony Savagar*

May 6, 2021

*a.savagar@kent.ac.uk.

Contents

1 Data Description

Our data source is the Business Structure Database (BSD). The BSD is a firm-level data-set provided by the UK Office of National Statistics (ONS) to accredited researchers. It includes basic information on the near population of UK firms and is annual. The record exists for taxation purposes.[1] The BSD is provided at enterprise level or local level. Enterprise units include a collection of local units.[2] We use enterprise units because it includes turnover data and is the definition of a firm that we would consider to compete with another firm. The advantage of the BSD is its near universal coverage of UK firms. The main variables of interest are employment, turnover, industry, birth and death year.

1.1 Coverage and Variables

The BSD is based on the Inter-Departmental Business Register (IDBR), which is a live register of data collected by HM Revenue and Customs (HMRC) via VAT and Pay As You Earn (PAYE) records. The IDBR is obtained from the HM Customs and Excise which registers traders for Value added Tax (VAT) purposes and the Inland Revenue which registers employers for Pay As You Earn (PAYE) purposes. The IDBR data are combined with data from ONS business surveys. If a business is liable for VAT (turnover exceeds the VAT threshold of £85,000 in 2018) and/or has at least one member of staff registered for the PAYE tax collection system, then the business will appear on the IDBR (and hence in the BSD).

The BSD has an extensive coverage of UK firm population. The ONS (2006) reports that in 2006 firms in the IDBR covered 99% of business activity in the UK.[3] Only very small businesses, such as the self-employed are not in the IDBR. The data is divided into 'enterprises' and 'local units'.[4] An enterprise is the overall business organisation which includes one or more local unit. A local unit is a 'plant', such as a factory, shop, branch or warehouse. ONS (2006) estimates that 97% of enterprises in the BSD consist of one local unit. For each enterprise, data are available on employment, turnover, foreign ownership, and industrial activity based on Standard Industrial Classification (SIC) 2003 or SIC 2007. Year of 'birth' (company start-up date) and 'death' (termination date) are also included, as well as postcodes for both enterprises and their local units.

[1] It captures any firm that qualifies for value-added tax (turnover exceeds £85,000 in 2018) or has at least one payroll employee.

[2] 97% of enterprise units have a single local unit, so for most firms the units are equivalent.

[3] Our calculations show that BSD employment accounts for 98% of aggregate UK employment data reported by the Office of National Statistics (ONS).

[4] The BSD includes data on group ownership. A group is the owner of several enterprises. Enterprises consist of one or more local units. The group information explains who own whom. However, data is collected at enterprise level.

1.2 Definition of a Firm

We define a firm as an enterprise unit (ENT). This is a single row in the dataset with a unique ID. ENT is the level of enterprise that is associated with VAT and PAYE accounts. This corresponds to most people's idea of a business. In total, the BSD comprises ENT_GROUP, ENT and LU.[5] The ENT_GROUP, also known as who owns whom reference, is mostly missing for early years and there is no data in recent years. LU is the local unit such as a shop, warehouse or plant.

Note on Classification of 'Firm' in ONS Data

There are four possible definitions of a firm across ONS datasets. Businesses decide themselves the most appropriate entity to act as a reporting unit. They know the best grouping of their business operations to respond to questionnaires. The following two examples show the possibilities. Firstly, an example of a firm organising RU by geography:

1. ENT_GROUP: Joe's Big Retail Consortium
2. ENT: Joe's Athletic Stores
3. RU: Joe's Athletic Stores, London
4. LU: Joe's Athletic Stores, Oxford Street

Secondly, an example of a firm organising RU by business function:

1. ENT_GROUP: Joe's Big Retail Consortium
2. ENT: Joe's Athletic Stores
3. RU: Joe's Discount Athletic Stores
4. LU: Joe's Discount Athletic Stores, Oxford Street.

Top of the hierarchy is ENT_GROUP which are legal entities that control multiple distinct PAYE/VAT-paying enterprise units. Below them is an enterprise unit (ENT). This is what most people recognise as a firm. Each ENT can have one or multiple reporting units (RUs). Reporting units are relevant for ONS business surveys such as the Annual Business Survey (ABS) and Annual Respondents Database (ARD). They are the unit that returns a survey. The classification of RU is decided for administrative reasons by the ENT. Some ENTs organise RUs geographically, whereas others organise by business function. LUs correspond to local footprints (e.g. factory, retail location, warehouses). An LU will not collect or report on specific financial data.

[5] LUs are in a separate more granular dataset. BSD does not include reporting unit (RU). RU is used in the business survey datasets such as the ABS and ARDx databases.

1.3 Sector Classification: UK SIC 2007

Industrial classification follows UK SIC 2007.[6] The highest level of aggregation in the UK SIC 2007 system is a *'Section'* within sections are 2-digit *'Divisions'*, within divisions are 3-digit *'Groups'*, within groups are 4-digit *'Classes'* and (occasionally) within classes are 5-digit *'Sub-classes'*. For example:

- Section: P: EDUCATION
- Division 85: Education
- Group 85.4: Higher education
- Class 85.42: Tertiary education
- Sub-class 85.42/2: Post-graduate level higher education

There are 21 sections assigned letters A-U and 88 divisions assigned two-digit numbers between 01 and 99.[7] There are 831 5-digit industries. Our analysis excludes section T: Household production (divisions 97 and 98) and U: Extraterritorial activities (division 99). Therefore, there are 85 divisions and 19 sections in our analysis. The structure for sections and divisions is

- Section A: Agriculture, Forestry and Fishing
  - Division 01: Crop and Animal Production, Hunting and Related Service Activities
  - Division 02: Forestry and Logging
  - Division 03: Fishing and Aquaculture
- Section B: Mining and Quarrying
  - Division 05: Mining of Coal and Lignite
  - Division 06: Extraction of Crude Petroleum and Natural Gas
  - Division 07: Mining of Metal Ores
  - Division 08: Other Mining and Quarrying
  - Division 09: Mining Support Service Activities
- Section C: Manufacturing
  - Division 10: Manufacture of Food Products

[6] Full details on the ONS website.

[7] The single digits do not represent hierarchical codes. There are 88 divisions as some numbers between 01 and 99 are not used.

  - Division 11: Manufacture of Beverages
  - Division 12: Manufacture of Tobacco Products
  - Division 13: Manufacture of Textiles
  - Division 14: Manufacture of Wearing Apparel
  - Division 15: Manufacture of Leather and Related Products
  - Division 16: Manufacture of Wood and of Products of Wood and Cork, Except Furniture; Manufacture of Articles of Straw and Plaiting Materials
  - Division 17: Manufacture of Paper and Paper Products
  - Division 18: Printing and Reproduction of Recorded Media
  - Division 19: Manufacture of Coke and Refined Petroleum Products
  - Division 20: Manufacture of Chemicals and Chemical Products
  - Division 21: Manufacture of Basic Pharmaceutical Products and Pharmaceutical Preparations
  - Division 22: Manufacture of Rubber and Plastic Products
  - Division 23: Manufacture of Other Non-metallic Mineral Products
  - Division 24: Manufacture of Basic Metals
  - Division 25: Manufacture of Fabricated Metal Products, Except Machinery and Equipment
  - Division 26: Manufacture of Computer, Electronic and Optical Products
  - Division 27: Manufacture of Electrical Equipment
  - Division 28: Manufacture of Machinery and Equipment N.e.c.
  - Division 29: Manufacture of Motor Vehicles, Trailers and Semi-trailers
  - Division 30: Manufacture of Other Transport Equipment
  - Division 31: Manufacture of Furniture
  - Division 32: Other Manufacturing
  - Division 33: Repair and Installation of Machinery and Equipment
- Section D: Electricity, Gas, Steam and Air Conditioning Supply
  - Division 35: Electricity, Gas, Steam and Air Conditioning Supply
- Section E: Water Supply; Sewerage, Waste Management and Remediation Activities
  - Division 36: Water Collection, Treatment and Supply

  - Division 37: Sewerage
  - Division 38: Waste Collection, Treatment and Disposal Activities; Materials Recovery
  - Division 39: Remediation Activities and Other Waste Management Services
- Section F: Construction
  - Division 41: Construction of Buildings
  - Division 42: Civil Engineering
  - Division 43: Specialised Construction Activities
- Section G: Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
  - Division 45: Wholesale and Retail Trade and Repair of Motor Vehicles and Motorcycles
  - Division 46: Wholesale Trade, Except of Motor Vehicles and Motorcycles
  - Division 47: Retail Trade, Except of Motor Vehicles and Motorcycles
- Section H: Transportation and Storage
  - Division 49: Land Transport and Transport via Pipelines
  - Division 50: Water Transport
  - Division 51: Air Transport
  - Division 52: Warehousing and Support Activities for Transportation
  - Division 53: Postal and Courier Activities
- Section I: Accommodation and Food Service Activities
  - Division 55: Accommodation
  - Division 56: Food and Beverage Service Activities
- Section J: Information and Communication
  - Division 58: Publishing Activities
  - Division 59: Motion Picture, Video and Television Programme Production, Sound Recording and Music Publishing Activities
  - Division 60: Programming and Broadcasting Activities
  - Division 61: Telecommunications
  - Division 62: Computer Programming, Consultancy and Related Activities

  - Division 63: Information Service Activities
- Section K: Financial and Insurance Activities
  - Division 64: Financial Service Activities, Except Insurance and Pension Funding
  - Division 65: Insurance, Reinsurance and Pension Funding, Except Compulsory Social Security
  - Division 66: Activities Auxiliary to Financial Services and Insurance Activities
- Section L: Real Estate Activities
  - Division 68: Real Estate Activities
- Section M: Professional, Scientific and Technical Activities
  - Division 69: Legal and Accounting Activities
  - Division 70: Activities of Head Offices; Management Consultancy Activities
  - Division 71: Architectural and Engineering Activities; Technical Testing and Analysis
  - Division 72: Scientific Research and Development
  - Division 73: Advertising and Market Research
  - Division 74: Other Professional, Scientific and Technical Activities
  - Division 75: Veterinary Activities
- Section N: Administrative and Support Service Activities
  - Division 77: Rental and Leasing Activities
  - Division 78: Employment Activities
  - Division 79: Travel Agency, Tour Operator and Other Reservation Service and Related Activities
  - Division 80: Security and Investigation Activities
  - Division 81: Services to Buildings and Landscape Activities
  - Division 82: Office Administrative, Office Support and Other Business Support Activities
- Section O: Public Administration and Defence; Compulsory Social Security
  - Division 84: Public Administration and Defence; Compulsory Social Security

- Section P: Education
  - Division 85: Education
- Section Q: Human Health and Social Work Activities
  - Division 86: Human Health Activities
  - Division 87: Residential Care Activities
  - Division 88: Social Work Activities Without Accommodation
- Section R: Arts, Entertainment and Recreation
  - Division 90: Creative, Arts and Entertainment Activities
  - Division 91: Libraries, Archives, Museums and Other Cultural Activities
  - Division 92: Gambling and Betting Activities
  - Division 93: Sports Activities and Amusement and Recreation Activities
- Section S: Other Service Activities
  - Division 94: Activities of Membership Organisations
  - Division 95: Repair of Computers and Personal and Household Goods
  - Division 96: Other Personal Service Activities
- Section T: Activities of Households as Employers; Undifferentiated Goods- and Services-producing Activities of Households for Own Use
  - Division 97: Activities of Households as Employers of Domestic Personnel
  - Division 98: Undifferentiated Goods- and Services-producing Activities of Private Households for Own Use
- Section U: Activities of Extraterritorial Organisations and Bodies
  - Division 99: Activities of Extraterritorial Organisations and Bodies

1.4 Timing

The BSD data for each year is captured from the IDBR in March and released in September. Figure 1 illustrates the period of economic activity recorded for a firm in a given BSD snapshot. A submission for a firm in a given BSD period will correspond to economic activity that pre-dates the BSD year. This is more problematic if the firm submits accounts to HMRC shortly after the BSD snapshot is taken.

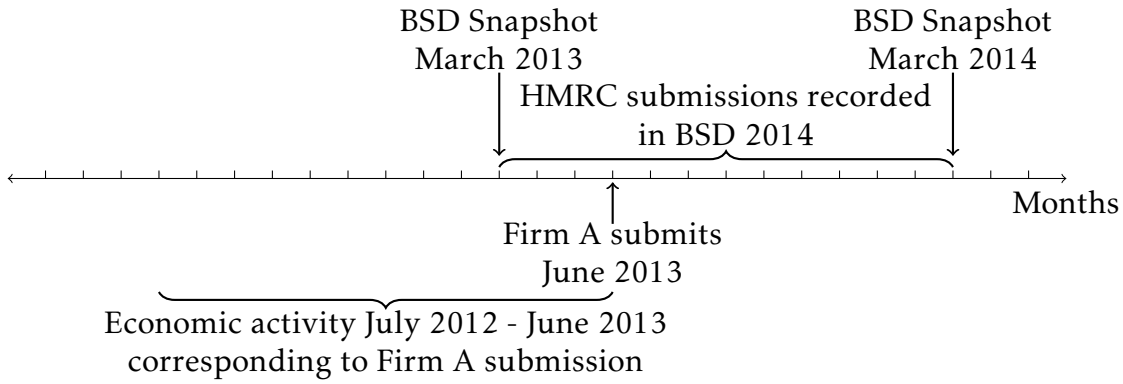

Figure 1: Time Period Covered by BSD Snapshots

In Figure 1 we assume that the accounts submitted by Firm A in June 2013 correspond to the immediately preceding 12 months. However, a firm may delay the submission of their accounts for up to 2 or more months after they have been completed. Therefore accounts submitted to HMRC will not necessarily represent economic activity up to the date of submission. For mismeasurement in our data, the worst case scenario is that a firm submits accounts in April 2013 – just after the BSD March 2013 snapshot – for accounts that were finalised in February or earlier. These accounts cover the 12 months of economic activity from March 2012 to February 2013 but would be recorded in BSD 2014. Therefore economic activity that took place in 2012 is recorded in BSD 2014.[8]

2 Data Construction

We follow Du and Bonner (2016) and the UKDS (2014) to generate a longitudinal BSD database for our analysis.

2.1 Data Cleaning

We exclude inactive firms, firms with no employees or turnover data and firms with no entry and exit year.[9]

2.2 Industry Reclassification: Linking SIC 2003 to SIC 2007

In 2007 UK SIC classification changed from SIC 2003 to SIC 2007. We match 5-digit SIC 2003 codes to 5-digit SIC 2007 codes using ONS look-up table [10]. For discontinued

[8]Du and Bonner (2016) also estimates the lag on the BSD turnover data may be approximately 2 years.

[9]We do not drop firms with no turnover or employees in the exit year of firms.

[10]ONS lookup table is available at here. We link 5-digit SIC 2003 codes to the 5-digit 2007 codes with the highest employment or sales weighting. e.g If 85140 sic 2003 code goes to 4 different 2007 sic

5-digit SIC 2003 codes with no corresponding 5-digit SIC 2007 codes, we allocate to the nearest 4-digit or 3-digit code. This leads to discontinuities at the 5-digit level. There are some discontinuities at the 2-digit level. These occur when activities are reclassified to another 2-digit sector. Also, the ONS re-classify enterprises from one 2-digit code to another due to change in the major activity of the firm.[11] In 2006, BSD notes that 18% of all enterprises have been re-classified at least once between 1998 and 2006. This is important for concentration studies if a large firm is reclassified.

2.3 Deflating Variables

We deflate turnover using the ONS deflator [12] which is available at two-digit or three-digit SIC 2007 level [13].

2.4 Full Sample and Sub-sample

We consider a full-sample and a sub-sample of the dataset.

1. *Full-sample*: Includes all 18 subsectors.
2. *Sub-sample*: Excludes 9 subsectors. The following are excluded: Financial sector; Agriculture; Mining; Electricity; Water; Real Estate; Public Administration and Defense; Education; Human Health and Social Work Activities.

The subsample excludes subsectors that are known to be poorly measured or in which using turnover to represent output is misleading.[14]

3 Variable Construction and Robustness

In this section we describe our variables, explain how they are constructed, and comment on their consistency with other aggregate data.

Number of firms

The number of firms is a count of establishments in the BSD.

codes at different percentages: 86900 (88%), 87100 (12%), 87200(0%), 87300 (0%),we convert all firms with 85140sic 2003 to 86900 sic 2007 code.

[11] For example, an enterprise producing food and beverages can be classified into 2-digit SIC code for food *or* beverages. Based on its sales in each category the ONS may re-classify year-on-year.

[12] ONS deflator is available at here. We used the May 2019 version.

[13] We mostly used two-digit deflators except in sector divisions 10, 11, 20, 23, 24, 25, 30, 33, 35, 49, 68 and 69 where three-digit deflators are available.

[14] The subsample contributes half of aggregate sales and a third of employment, see figures. The financial sector accounts for the largest turnover in aggregate UK BSD turnover followed by the Wholesale sector. In the case of employees, Education, Public Administration, Human Health represent a significant portion of employment and accounts for the difference between the full sample and subsample aggregate employees.

Sales

Sales measure the nominal value of sales at each enterprise unit in the BSD. We deflate nominal values into 2016 prices.[15]

Employment

Employment measures the number of employees at each enterprise unit in the BSD. Comparing BSD employment data to aggregate UK employment data shows that firms in the BSD cover 98% of total UK aggregate employment. In the first quarter of 2018, UK employment was 32.34 million.[16] BSD employment data captured in March 2018 was 31.64 million.

Labour productivity

We measure labour productivity as turnover per employee. Simple average labour productivity is calculated as aggregate real turnover divided by aggregate employees in a given year. We also calculate sales weighted and employment weighted labour productivity.[17]

Concentration

We measure concentration by the Hirfindahl-Hirschman Index (HHI) or concentration ratio of the top N firms (CRN). Concentration ratios (CRN) represent the market share of the biggest N firms. For example, CR10 is the market share of the ten largest firms.

Entry and Exit

We generate an entry and exit variable for firms. Entry is the first year that a firm is recorded as being active and records employees and turnover as non-zero or not missing. Exit is the first year the firm is recorded as being inactive having been active the previous year or the first year a firm records turnover and employees as zero.

[15] Price adjustment for the real series are applied at the 2-digit level before aggregation. The price deflators for each 2-digit industry are given by the ONS.

[16] Data available at ONS website.

[17] Labour productivity in each 2-digit SIC industry is weighted by the percentage of its sales/employment in aggregate BSD sales/employment.

4 Limitations

4.1 BSD Events

BSD coverage has changed over time. The BSD is affected when a large number of firms is added in a given year. The following BSD measurement events are noteworthy (ONS 2017):

- 18% of all enterprises change SIC classification at least once between 1998 to 2006.[18]
- 2008 addition of PAYE based only enterprises.
- 2012 addition of enterprises due to improvements in HMRC computer systems.
- 2013 & 2014 increase in the number of enterprises registered for PAYE.
- 2015 addition of previously excluded PAYE based enterprises.

4.2 Double-counting

Sales data is subject to double-counting when there are long value chains. A value chain refers to a chain of procedures, performed by independent producers, modifying a product until it transforms from 'raw materials' to final output. Each company in a value chain counts all sales as revenue. The addition of that product to the economy (defined as total revenue) explodes as the value chain gets longer.[19] For example, firm 1 makes £100,000 widgets recording £100,000 in revenue. It then sells these to the next firm in the chain (firm 2). Firm 2 markets the widgets adding £50,000 of marketing value. Firm 2 sells the marketed widgets on for £150,000. Therefore the original £100,000 is recorded twice if we total revenues. With value-added data we would record £100,000 for firm 1 and £50,000 for firm 2, so output is recorded as £150,000. But with sales data we record £100,000 for firm 1 and £150,000 for firm 2, so output is recorded as £250,000. In addition to double-counting created by long value chains, double-counting can also be caused by off-payroll working. See the discussion below.

4.3 'Accounting' Firms

Economists typically think of a firm as a production function. That is, an entity which combines inputs to produce output – labour to produce turnover in our data. How-

[18] This was due to ONS reclassifying firms' industrial sector due to change in the activity of the local unit of the firm.

[19] A solution to the double-counting problem is to use HMRC datalab VAT data. This alternative dataset includes VAT inputs. Therefore, revenues from the BSD less inputs gets closer to value-added.

ever, in practice some firms may exist only for accounting purposes.

Special-purpose Entity / Vehicle

A special-purpose entity (US) or vehicle (UK) is a legal entity (usually a limited company of some type or, sometimes, a limited partnership) created to fulfill narrow, specific or temporary objectives. SPEs are typically used by companies to isolate the firm from financial risk. Since these entities appear in the data as new firms they can lead to misleading results – they are not firms in the way economists think of a firm (as a production function), which can lead to misleading results. They typically appear as single employee firms with a huge amount of turnover in one year.

Disguised Employees, Contracting, Off-Payroll Working and IR35

A common practice in the UK is off-payroll working. The rules relating to this are known as IR35 to the tax department (HMRC). This occurs when an employee works at a single firm but sets themselves up as a company that charges the employing firm for their labour. In the data on revenues this output will be recorded twice. Once as the individual employee's firm since they record their pay as revenues for their business. However, their work also adds to the revenues of the company they are working for, whereas with value-added data this contribution of the worker would be accounted for as it would be deducted as a cost by the employing firm. If the employee worked for the firm as a standard employee, their production would only be added to the employing business's revenue. There would be no separate firm contributing to aggregate revenue measures.

5 Outliers

5.1 Small Firms

We investigate the effect of small firms in our data as they are more likely to be used for taxation purposes. We conduct two sensitivity checks.

1. First, we exclude single employee firms. Single employee firms account for 40% of the total number of firms in the BSD, but they represent 3% of total BSD sales and employment. Dropping them has no significant effect on aggregate trends.

2. Second, we exclude firms with less than 5 employees. These firms represent 16% of total BSD sales and employment and excluding them affects aggregate trends in the earlier and later periods: 1997-1999 and 2017-2018. We also exclude firms with less than 10 employees which represent 21% of BSD sales and employment.

After excluding these firms, we find similar trends with aggregate BSD data in most periods except in 2017-2018.

5.2 Labour Productivity

We drop outliers in labour productivity by removing the top and bottom tails of the distribution. We remove firm year observations in which labour productivity exceeds the outlier threshold. Plots of labour productivity suggest it is problematic. This occurs because of single employee firms with exceptionally high revenue.

References

Du, Jun and Karen Bonner (2016). "Decomposing UK aggregate labour productivity and growth: 1998-2013 using the ONS Business Structure Database data". In: *ERC Research Paper* 48.

ONS (2006). "Business Structure Database User Guide". In:

— (2017). "UK business activity, size and location: methodology". In:

UKDS (2014). "Code for compilation of longitudinal BSD 1997-2013". In: