

Supplementary Appendix: Data Methodology

Market Concentration and Productivity: Evidence from the UK

Anthony Savagar*

Oluwaseun Aguda

Ioannis Galanakis

Jingwei Wu

May 2024

Contents

1	Data Description	2
1.1	Coverage and Variables	2
1.2	Definition of a Firm	2
1.3	UK SIC 2007 Industry Classifications	4
1.4	ONS ‘BSD Sector’ Classifications	9
1.5	Timing	10
2	Data Construction	11
2.1	Data Cleaning	11
2.2	Industry Reclassification: Linking SIC 2003 to SIC 2007	11
2.3	Deflating Variables	12
2.4	Variable Descriptions	12
3	Data Limitations	13
3.1	BSD Events	13
3.2	Double-counting	14
3.3	‘Accounting’ Firms	14

*a.savagar@kent.ac.uk.

1 Data Description

The Business Structure Database (BSD) is a firm-level dataset 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.

1.1 Coverage and Variables

The BSD is based on the Inter-Departmental Business Register (IDBR), which is a live register of firms collected for VAT and Pay As You Earn (PAYE) tax purposes by HMRC. 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. Our calculations show that BSD employment accounts for 98% of aggregate UK employment data reported by the Office of National Statistics (ONS). Only very small businesses, such as sole traders with low turnover are not in the IDBR. The data is divided into ‘enterprises’ and ‘local units’. There is also group ownership data, but with limited coverage. 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. The main variables of interest are employment, turnover, industry. We construct our own measures of birth and death based on economic activity.

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 com-

prises ENT_GROUP, ENT and LU.¹ 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. LUs do not include turnover information as financial accounts are not recorded at this level.

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 (RU). 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). 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. Local units (LUs) correspond to local footprints (e.g. factory, retail location, warehouses). An LU will not collect or report on specific financial data.

¹LUs 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 UK SIC 2007 Industry Classifications

Industrial classification follows UK SIC 2007. Full details on the [ONS website](#). 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. The single digits do not represent hierarchical codes. There are 88 divisions as some numbers between 01 and 99 are not used. 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
 - 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

SIC	SIC name	SIC	SIC name
1	Crop and animal production	47	Retail trade
2	Forestry and logging	49	Land transport and transport via pipelines
3	Fishing and aquaculture	50	Water transport
5	Mining of coal and lignite	51	Air transport
6	Extraction of crude petroleum and natural gas	52	Warehousing and transport support
7	Mining of metal ores	53	Postal and courier activities
8	Other mining and quarrying	55	Accommodation
9	Mining support service activities	56	Food services
10	Manufacture of food products	58	Publishing activities
11	Manufacture of beverages	59	Motion picture, video, sound and TV production
12	Manufacture of Tobacco	60	Programming and broadcasting activities
13	Manufacture of textiles	61	Telecommunications
14	Manufacture of wearing apparel	62	Computer programming, consultancy and related activities
15	Manufacture of leather and related products	63	Information service activities
16	Manufacture of wood and wood products	64	Financial service activities, except insurance and pension
17	Manufacture of paper and paper products	65	Insurance, reinsurance and pension funding
18	Printing and reproduction of recorded media	66	Activities auxiliary to financial and insurance
19	Manufacture of coke and refined petroleum products	68	Real estate activities
20	Manufacture of chemicals and chemical products	69	Legal and accounting activities
21	Manufacture of basic pharmaceutical products	70	Activities of head offices; management consultancy
22	Manufacture of rubber and plastic products	71	Architectural and engineering activities
23	Manufacture of other non-metallic mineral products	72	Scientific research and development
24	Manufacture of basic metals	73	Advertising and market research
25	Manufacture of fabricated metal products	74	Other professional, scientific, technical activities
26	Manufacture of computer, electronic and optical products	75	Veterinary activities
27	Manufacture of electrical equipment	77	Rental and leasing activities
28	Manufacture of machinery and equipment n.e.c.	78	Employment activities
29	Manufacture of motor vehicles, trailers	79	Travel agency; reservation services
30	Manufacture of other transport equipment	80	Security and investigation activities
31	Manufacture of furniture	81	Services to buildings and landscape activities
32	Other manufacturing	82	Office administrative and other business support activities
33	Repair and installation of machinery and equipment	84	Public administration and defence
35	Electricity, gas, steam and air conditioning supply	85	Education
36	Water collection, treatment and supply	86	Human health activities
37	Sewerage	87	Residential care activities
38	Waste collection, treatment and disposal	88	Social work activities
39	Remediation and other waste management services	90	Creative, arts and entertainment
41	Construction of buildings	91	Libraries, archives, museums and cultural activities
42	Civil engineering	92	Gambling and betting activities
43	Specialised construction activities	93	Sports, amusement and recreation activities
45	Motor trade and repair	94	Activities of membership organisations
46	Wholesale trade	95	Repair of computers and personal and household goods
		96	Other personal service activities

Table 1: 2-digit SIC07 names

1.4 ONS ‘BSD Sector’ Classifications

The ‘BSD sector’ definitions are specific to the ONS. They aggregate SIC 2-digit industries (“divisions”) into sectors that are of interest for the UK economy. Table 2 shows the relationship between ONS SIC 2007 section and BSD sector. There are 8 BSD Sectors, 19 1-digit sectors also known as Sections, and there are 85 2-digit sectors also known as divisions.

BSD Sector	Abbr.	SIC07 2-digit	SIC07 1-digit
Production	PD	1-39	Agriculture, Forestry and Fishing (A), Mining and Quarrying (B), Manufacturing (C), Electricity, Gas, Steam and A/C (D), Water Supply and Waste Management (E).
Construction	CN	41-43	Construction (F)
Catering	CA	55-56	Accommodation and Catering (I)
Motor trade	MT	45	Wholesale, Retail and Motor Trade (G)
Retail	RE	47	Wholesale, Retail and Motor Trade (G)
Wholesale	WH	46	Wholesale, Retail and Motor Trade (G)
Property	PR	68	Property (Real estate Activities (L))
Other services	ST	49-96*	Transport and Storage (H), Information and Communication (J), Financial and Insurance Services (K), Professional, Science and Tech (M), Administrative and Support Services (N), Public Administration (O) Education (P), Human Health and Social Work (Q), Arts, Entertainment and Recreation (R), Other Services (S).

Table 2: Two-digit SIC, Sector and Sub-sector classification

*Property (68) and Catering (55-56) are excluded from this range.

1.5 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.

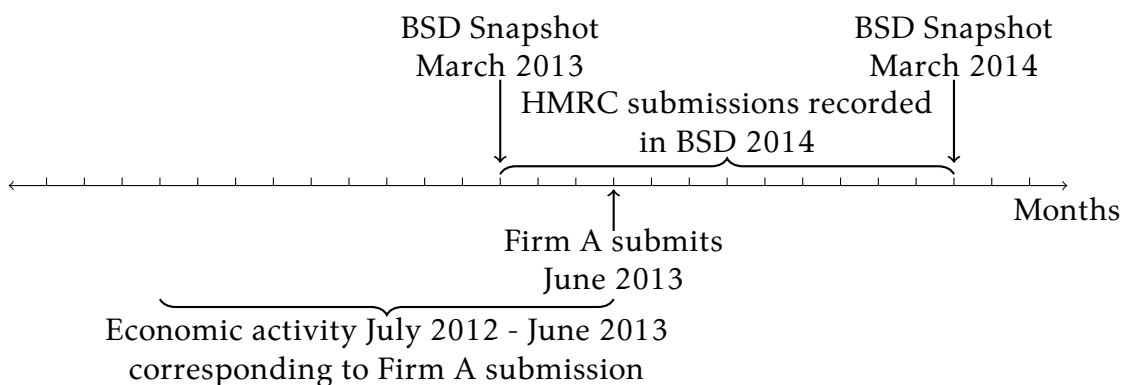


Figure 1: Time Period Covered by BSD Snapshots

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 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 would be 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. Du and Bonner (2016) also estimate the lag on the BSD turnover data may be approximately 2 years.

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

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. The 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 SIC 2003 code 85140 maps

to four different SIC 2007 codes at different percentages: 86900 (88%), 87100 (12%), 87200(0%), 87300 (0%), we convert all SIC 2003 firms with 85140 to SIC 2007 code 86900. For discontinued 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. 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. 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. The ONS deflator is available [here](#). We used the May 2019 version which is available at two-digit or three-digit SIC 2007 level. We 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.

2.4 Variable Descriptions

Number of firms

The number of firms is a count of establishments in the BSD. The BSD records all businesses which are registered for VAT or payroll tax (PAYE). This covers the vast majority of economic activity, but will exclude non-employers below the VAT threshold. Further details are available in '[UK business: activity, size and location Quality and Methodology Information](#)'. This documentation describes the Inter-departmental Business Register (IDBR) which is the main register which we have a snapshot of via the Business Structure Database (BSD).

Sales

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

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.² 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. Sales weighted and employment weighted labour productivity are calculated by taking labour productivity at the 2-digit SIC industry level and weighting it by the industry's share in total sales or employment.

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. HHI is the sum of squared market shares.

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.

3 Data Limitations

In this section we explain some limitations to the dataset.

3.1 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), however in the aggregate data we do not observe clear breaks following these events:

²Data available at [ONS website](#).

- 18% of all enterprises change SIC classification at least once between 1998 to 2006. This was due to ONS reclassifying firms' industrial sector due to change in the activity of the local unit of the firm.
- 2008 addition of PAYE-only enterprises. That is, firms with low turnover that have an employee.
- 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.

3.2 Double-counting

Sales data is subject to double-counting when there are long value chains. A value chain refers to a chain of production, 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. For example, firm 1 sells £100,000 widgets to firm 2, who are next in the chain. Firm 1 records £100,000 in revenue. 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. A solution to the double-counting problem for BSD data 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. However, in practice there are administrative difficulties to linking this data, which is not easily-accessible to researchers.

3.3 '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. However, in practice some firms may exist only for accounting purposes.

Special-purpose Entity / Vehicle

A special-purpose entity / vehicle is a legal entity (e.g. limited company or 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 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-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.

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 (May 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.