Business sectors, occupations and aggregations of census data 1851-1911

Robert J. Bennett, Harry Smith, Carry van Lieshout and Gill Newton

rjb7@cam.ac.uk hjs57@cam.ac.uk cv313@cam.ac.uk ghn22@cam.ac.uk

Working Paper 5:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Businesses

University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

May 2017

Comments are welcomed on this paper: contact the authors as above.

© Robert J. Bennett, Harry Smith, Carry van Lieshout and Gill Newton, University of Cambridge, members of the Cambridge Group for the History of Population and Social Structure assert their legal and moral rights to be identified as the authors of this paper; it may be referenced provided full acknowledgement is made: ***Cite*** (Harvard format):

Bennett, Robert J., Smith, Harry, van Lieshout, Carry and Newton, Gill (2017) *Occupations and occupational aggregation of census data 1851-1911*. Working Paper 5: ESRC project ES/M010953: 'Drivers of Entrepreneurship and Small Businesses', University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

Keywords: Entrepreneurship, Employers, Self-employment, Small businesses, Census

JEL Codes: L26, L25, D13, D22

Business sectors, occupations and aggregations of census data 1851-1911

Robert J. Bennett, Harry Smith, Carry van Lieshout and Gill Newton

Working Paper 5: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Businesses, University of Cambridge.

1. Introduction

This paper discusses alternative sector aggregations that can be used for the ESRC-supported project ES/M010953 'Drivers of Entrepreneurship and Small Businesses'. Its aim is to produce simplified and aggregated classifications of the sectors of business activity that can be used for a wide range of analyses of entrepreneurs and business proprietors over 1851-1911, and which can be linked to other modern and historical classifications in a meaningful way. These aggregations are provided as codes in the entrepreneurs database deposited at UK Data Archive (UKDA).

The population census 1851-1911 undertaken by the General Register Office (GRO) allows individual employers and the self-employed to be identified. The main source used by the ESRC project is transcripts of the census, mainly derived for the I-CeM electronic database for 1851-1911 produced by a team at the University of Essex, deposited at the UKDA: *The Integrated Census Microdata (I-CeM)*.[1] This provides transcriptions of the original Census Enumerators Books (CEBs) as well as enhancing the data with various codes for household structure, relationships between people, and most important for the discussion here, coding of occupations. These codes are derived by Essex from classification of the descriptors given in the original responses to the census instruction to householders for 'filling up the column

[1] K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, (2016) [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1; enhanced; E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex, 2015).

headed "RANK, PROFESSION, or OCCUPATION"'. Household respondents also had to follow additional detailed instructions for various categories.[2]

Within the data is information on *employers* (those who employed others), *sole proprietor own account self-employed* (who employed no-one else), as well as employees, workers and the unoccupied. The information on employers and own account is the main subject of research in the 'Drivers of Entrepreneurship and Small Businesses' project, and hence an emphasis is given here to how codes for business sectors, and aggregations, relate to different employment status categories, especially to identify clearly those that are entrepreneurs and business proprietors.

The rest of the paper is divided into three sections. Section 2 discusses how sector aggregations can be constructed that classify business proprietors by the type of economic activity in which they were engaged, and also the sector of activity for those that were clearly not entrepreneurs. Section 3 compares this coding with that used for the PST occupational classification system devised by Tony Wrigley. Section 4 compares the codes with modern comparison and other mainstream classifications of sector and occupational groups.

2. Towards an a classification of business sectors

The central purpose here is to classify *business proprietors* by the sector of economic activity in which they are engaged. For the purposes of this working paper and the project 'Drivers of Entrepreneurship and Small Businesses', business proprietors and entrepreneurs are defined as widely as possible. Information and coding is added to the entrepreneurs database deposited at UKDA so that different groups and categories can be selected by users as required to answer different research questions using different definitions of business proprietor. The base definition is proprietors: those who were at the date of the census (or a nearby date where other data have to be used) *currently* the responsible individuals bearing the risks of running private business enterprises, however small, and even if they did this alone with no other individuals involved (as self-employed individuals with no employees, called 'own account' in the census). This definition is drawn very broadly to include all the

[2] e.g. 'General Instruction', Census of England and Wales, Householder's Schedule, 1851.

major types of proprietors that have been the focus of most previous research on business entrepreneurs.

The extraction of proprietors (and others) uses transcriptions of the original CEBs, which include the full descriptive string of information on occupations and business activity given by census respondents as recorded by the GRO enumerators. This information is valuable, but imperfect. As noted by Tony Wrigley, 'those making use of [published] nineteenth-century English censuses are constrained by the solution adopted by the census authorities'.[3] However, the use of the original CEBs now available in I-CeM significantly reduces this problem because it is possible to code to any format required, constrained only by what the enumerators recorded. This is usually a wider range of detail than after subsequent GRO processing and aggregation, as given in contemporary census publications.

Within the I-CeM database the 'rank, profession, or occupation' returned by all households has been coded to main occupation following a method developed by Woollard for 1881,[4] and then expanded to other census years by Schürer and Woollard.[5] For 1881 this provides 416 census occupational categories ('occode', which includes codes for illegible or uninterpretable responses). For the full censuses 1851-1911 in I-CeM there are 797 occodes. The occodes are an Essex interpretation of the CEBs, coded by algorithm with some hand checking. The Essex interpretation aims as closely as possible to mirror the occupational categories used by GRO in the census publications. The full 797 codes in I-CeM are composites running across all years 1851-1911; for any one census year they can be related to the specific census Registrar General's published categories through a dictionary;[6] for 1881 this the 399 categories used in that year's census report.[7] There are three ways to use the

[3] Wrigley, E. A. (2010) *The PST system of classifying occupations*, Working Paper 20, p. 5.
[4] Woollard, Matthew, 1999, *The Classification of Occupations in the 1881 Census of England and Wales*, Department of History, University of Essex; see also Wollard, 1997, Creating a Machine-Readable Version of the 1881 Census Enumerators' Books, pp. 98-101 in Charles Harvey and Jon Press, *Databases in Historical Research: Theory, Methods and Applications*, Macmillan, Basingstoke; Anderson, Michael (1988) Households, Families and Individuals: Some Preliminary Results from the National Sample from the 1851 Census of Great Britain, *Continuity and Change*, 3, 421-38; Morris, R. J. (1990) Occupational Coding: Principles and Examples, *Historical Social Research*, 15, 3-29..
[5] Schürer, Kevin and Woollard, Matthew (2002) *National Sample from the 1881 Census of Great Britain 5% Random Sample: Working Documentation Version 1.1*, Historical Censuses and Social Surveys Research Group, University of Essex.
[6] E. Higgs, C. Jones, K. Schürer and A. Wilkinson (2015) *Integrated Census Microdata (I-CeM) Guide.*
[7] The 416 occodes for the earlier version of 1881 was also based on the published census occupation categories; however, it included fifteen categories included in the census clerks' occupation dictionaries used to code the original data in 1881, but not included in the census report. It also includes two codes for illegible or blank occupations. This accounts for the difference between the classification created by Woollard for the earlier

occodes: (1) to use the 797 codes as a generic base for occupational categorisation, but this has the disadvantage for any year (especially the earlier ones) that many occodes are empty because that category was not used for that census year (with the individuals assigned elsewhere); (2) to use a reduced set for one census year and apply this to all years as a standard comparator, which has the advantage that there are fewer empty categories in any one year; or (3) use the occodes as a starting point but re-investigate the original transcribed strings to determine if there is further information available from the enumerator's records in the CEBs that helps to improve the occupational classification. This paper, and the ESRC project as a whole, follows a combination of methods (2) and (3): for some purposes it is valuable to use the I-CeM codes, but this is mainly done on the basis of a reduced set of census occodes, mainly using the 416 codes for 1881 as a base year; for other purposes, however, it is necessary to delve into the original CEB descriptor strings in order to better identify the business proprietors that are the focus of the research, and to ensure their separate identification from worker and employee categories.

A mixed strategy of using I-CeM occodes and investigating the original descriptors is also necessary because the I-CeM categories seek to follow as closely as possible the procedures of the census processing clerks at the GRO who were coding to achieve the published tables.[8] Thus I-CeM to some extent defeats the purpose of looking beyond the GRO to the original enumerator-mediated household responses which, as noted above, Wrigley identifies as potentially very fruitful. In addition, the census occupational instruction, as quoted in the Introduction above, included not only working activity but also social ranks, such as noble titles (Lord, Earl, etc.), Alderman, MP, captain of volunteers, etc.. This was valuable to many of the published census tabulations of categories that sought to count the numbers of different 'classes' of society, but it leads to a very confused field within CEBs, and hence within I-CeM, which muddles rank and actual occupation. For those of rank the occupation often omits their actual trade or professional activity (if any) in the original census. In addition, those of rank with complex strings often have transcription truncations in I-CeM which, as noted in earlier working papers, have to be infilled from the CEB originals.

version of the 1881 census, and the classification based on the 1881 report available in the I-CeM occupations dictionary; Woollard, *The Classification of Occupations*, p.42.

[8] Woollard, 1999, *Classification of Occupations,* pp. 6-8

This interrelates with a further drawback: that the resulting published tabulations by GRO categorised occupational 'classes' which often mixed rank, the industry in which employed, and occupational activity, rather than seeking to differentiate clearly industry and sector from skill or class.[9] This has caused frustration to researchers who are seeking insight into industries, industrial structure and business sectors. It is also an area which is believed to have more error in recording either by householders or enumerators.[10]

The occupational codes used by the GRO and reconstructed by I-CeM are a valuable contribution for analysis of occupations, but they are considerably less satisfactory for identification of business sectors. Sector classification requires the field of business activity or trade to be identified, whereas occupational classification focuses on the activities of the workforce. For example, the census identifies a shoemaker in a category 'Shoe, Boot - Maker, Dealer' (1881 definition), and this is reproduced in I-CeM. This case demonstrates the difficulty: the trade of shoe making is not separated into the different sectors of manufacture and wholesale or retail distribution; nor is it clear if individuals in this category are workers in someone else's enterprise or in their own. The combination of makers and dealers into single categories is one of the most pervasive difficulties of the historical census design, whilst differences in employment status (as workers, own account, or employers) is a key aspect of differentiation sought in the entrepreneurs data base. Similarly, the census category 'Nurseryman, Seedsman, Florist' covers three different sectors: production, processing and sales; whilst the census category 'Ironmonger, Hardware Dealer, Merchant' assumes both a wholesale and/or retail sector, and can also include major ironmongery manufacturers.

The difficulties of using GRO tabulations (and hence I-CeM codes) for industrial and sector analysis were recognised from an early date by Charles Booth who highlighted the inadequacies of the census for judging the relative size of different branches of industry, their change over time, and differences in geographical concentration.[11] His paper was taken up in

[9] See e.g. Woollard, 1999, *Classification of Occupations*, p. 2, 7; Higgs, 2004, *Life, Death and Statistics: Civil Registration, Censuses and the Work of the General Register Office, 1836-1952*,, pp. 56-64, 93-4; Higgs, 2005, *Making Sense of the Census Revisited: Census Records for England and Wales 1801-1901*.
[10] Woollard, 1997, Creating a Machine-Readable Version of the 1881 Census Enumerators' Books, p. 57; 1999, *Classification of Occupations* , p. 7-8.
[11] Booth, Charles, 1886, Occupations of the People of the United Kingdom 1801–1881, *Journal of the Royal Statistical Society*.

a wider memorial to the Committee on the Census which reported in 1890.[12] In this memorial by Booth and 45 other commentators, various suggestions were made for improvement. In the major review in 1890 by the Census Committee, evidence from several of the memorialists, including Alfred Marshall and Booth, argued that occupational classifications were unsatisfactory for the purposes of industrial and economic analysis, leading to a 'series of conundrums'.[13] However, the GRO did not change the fundamental approach throughout the period up to 1911, making only a few only minor modifications in 1891 and subsequently. The 1911 census was the first to attempt to develop some changes to meet the criticisms of Booth and others. Ironically, it is where the 1911 categories were expanded that it is now often impossible to go back and retrieve equivalent information from the CEBs: many of the I-CeM 797 occodes (which mainly reflect the expansions introduced in 1911) have no entries for the earlier censuses. This is why this project normally uses the 416 occodes for 1881 which can generally be aligned for each of the 1851-1911 years.

Because of the deficiencies in GRO classification and tabulations, Booth constructed a new classification which translated occupational codes into industrial sectors for all censuses 1851-81 which he then published. For 1881 this had about 350 categories in total, grouped into 51 subdivisions and 11 major categories. Many subsequent analyses have used this tabulation as a starting point, although others have developed an alternative sector format of PST, or have used the international comparative base of HISCO (discussed in sections 3 and 4 below, respectively). An important development was by Armstrong, who used Booth as a starting point, re-worked and simplified his classification, and extended it to include 1891 and tabulated summaries for 1851-91. He also re-examined Booth's original worksheets, checking them against the original census occupational returns and, using other comparisons, tested Booth's accuracy.[14] Armstrong found very few errors in Booth's classification, all of a minor nature. The Booth-Armstrong classification and tabulations provide the main starting point used in this paper for reclassifying and aggregating the census occupational descriptors to an industrial classification.[15]

[12] Acland, A. H. D and 45 others (1888) Memorandum on the Improvement of Census Returns, Especially as Regards Occupations and Industry, pp. 118-20 in *Report of the Committee Appointed by the Treasury to Inquire into Certain Questions Connected with the Taking of The Census*, C 6071 (1890).
[13] Booth, Questions 1354-1458, pp. 56-60; Marshall, Questions 1459-1566, pp. 60-8, quote in question 1462; *Report of the Committee on the Census*, 1890.
[14] Armstrong, W. A., 1972, The Use of Information about Occupations, pp. 191-210 in E. A. Wrigley (ed.) *Nineteenth-Century Society: Essays in the Use of Quantitative Methods for the Study of Social Data*.
[15] The details of the classification categories are given in full by Armstrong, 1972, Information about Occupations, pp. 255-81.

For the purposes here the Booth-Armstrong classification is used in a modified form. The modifications aim to maintain consistency with Booth-Armstrong, but also develop it, as far as possible, to be closer to modern classifications, and also to attempt to differentiate employer and own account individuals from workers. It should be noted that some of the Booth-Armstrong categories provide awkward grouping compared to modern classifications. For example, the category of 'industrial service' used by Booth (for banking, insurance and finance) overlaps with several modern business service categories; similarly Booth classified as 'public administration and professional services' diverse sectors that include a wide range of professions such as solicitors, barristers, architects, doctors, dentists etc.. The classification of professions in the census and in Booth-Armstrong is problematic since it did not differentiate between intermediate or final customer. Here it is expanded into two new categories. One identifies the growing range of 'personal service' professionals, such as doctors, dentists, artists, performers, etc., classified by Booth within very varied 'professions'. The second category identifies those serving business demands as 'professional and business services', which includes some in Booth's 'industrial service' as well as some 'professions' such as solicitors, barristers, and architects.

The classification used also modifies Booth-Armstrong by not attempting to separate manufacturers and dealers, but instead introduces a new category of 'maker-dealer'. This joint grouping is not entirely satisfactory since it mixes manufacturers with retailers, but it accepts the actuality that the census instructions did not make this distinction at the time so that using the historical returns provides few clues to an accurate distinction. Even in 1911, when a distinction between makers and dealers began to be more fully introduced into the census instructions for some occupations, the results were far from accurate, and it was only in censuses from the 1950s that such a distinction is more reliably drawn. The difficulties were recognised at the time. For example, the GRO's statistician Ogle stated that 'though theoretically making and selling may be distinct occupations, in practice they are very generally combined. The man who makes shoes is also usually the man who retails shoes, and the man who bakes bread is also the man who sells loaves'.[16] Booth-Armstrong used various rules of thumb to ascribe proportions of several occupations to different industry sectors of manufactures or retailers. These rules have become fairly standard for some subsequent analyses. For example, goldsmiths, silversmiths and jewellers; tobacco

[16] Ogle, Question 142, *Report of the Committee on the Census*, 1890, p. 7

manufacturers and tobacconists; and general shopkeepers were each divided equally between manufactures and retailing.[17] Hat manufacturers were shared 75% to manufactures and 25% to retail.[18] On the other hand Booth-Armstrong made no attempt to reassign any ironmongers from dealing to manufacturing despite a significant proportion being manufacturers. Chemists and druggists are a particularly difficult case; they were assigned to professions by Booth-Armstrong but while having many of the characteristics of professions, were also retailers and manufacturers. Conversely, several agricultural production sectors classified by Booth-Armstrong as manufacturing contain significant numbers of retailers; e.g. milksellers, dairymen, cheesemongers, buttermen, provision curers. Similarly, Booth-Armstrong classified a wide range of occupations to manufacturing that were also retailers: e.g. bakers, confectioners, and pastry cooks; mustard, vinegar, spice, pickle makers; and ginger beer and mineral water manufacturers; apparel manufacturers although most were also dealers; e.g. tailors, milliners, dressmakers, stay makers, shawl manufacturers, shirt makers, seamstresses, hosiers, haberdashers, glovers, glove makers, button makers, shoe and boot makers, patten and clog makers, umbrella and parasol makers, and accoutrement makers. Also Booth-Armstrong took blacksmiths and whitesmiths as manufacturers even though most outside of large industrial complexes were also dealers selling directly to the consumer. These categories are all included in the joint maker-dealer sector below.

2.1 Aggregate codes for entrepreneurs: ID18 and ID50

One of the primary aims of the classification to be developed is to use the occodes as far as possible to separate sectors and occupations that were employers, own account or workers. This cannot be fully achieved, but a classification can be developed that separates out five large main groups that can be assumed to be entirely formed of workers. These are the occodes falling within these five groups: 'public administration, military, and the clergy' which are entirely in employment of state, church or related organisations,[19] 'clerks, students and apprentices, 'domestic and service staff', 'labourers and transport staff', workers on farms (including family Members), and 'persons of property with no stated occupation'. For most analysis of entrepreneurs they can be removed from the analysis entirely. This means

[17] Armstrong, 1972, Information about Occupations , p. 282, n. xiv, xxii.
[18] Armstrong, 1972, Information about Occupations , p. 282, n. xxi.
[19] For issues arising from clergy also receiving fees as a form of own account see working paper 1.

that the main part of the remaining occodes will contain a mix of employers, own account and workers that have to be managed in a different way.

Aggregate ID sector	**Detailed sector(s): see Table 2**	**Summary title**
1	1	Farming and fishing
2	2-3	Mining & quarrying
3	4	Construction
4	5-6, 9-20; parts of 7-8, 22-24, 26	Manufacturing
5	Parts of 7-8, 21-24, 26, 35-36	Makers and dealers
6	29-34, 37-38; parts of 35-36	Retail & ironmongers
7	27-28	Transport
8	39-42	Professional & business services
9	43-44	Professional & personal services
10	13-14, 23-24, 30	Agricultural produce, drink & tobacco manuf.
11	32-33	Food retailing
12	33	Lodging & refreshment
13	39	Finance & commerce
14	45-46	Public admin, military, clergy
15	47	Clerks, students and apprentices
16	48	Domestic and service staff
17	49	Labourers & transport staff, incl. family on farms
18	50	Persons of property with no stated occupation

Table 1. Aggregate industrial sector groups: ID18 entrepreneurship sector classification.

The resulting aggregate classification leads to a development from Booth-Armstrong of an entrepreneurship sector classification in eighteen groups, as shown in Table 1: ID18. In this classification it is the last five (ID18: 14-18) that contain no entrepreneurs as a main occupation on our definitions. The only major caveat to this is that category 18 ‘Persons of property with no stated occupation’ is extraordinarily difficult to interpret, and within it lay some prominent business proprietors where they owned large estates; but for this I-CeM

occode none have any stated employer or own account status as we would otherwise classify then elsewhere. Category 18 is discussed further in later papers where some in the category are brought into the entrepreneurs database though enrichment.

The more detailed content of the eighteen groups categories is addressed at a second level, as listed in Table 2 in fifty sub-categories: ID50. In this the last five categories (ID50: 45-50) again do not contain entrepreneurs as a main occupation, with ID50: 50 as the difficult category of 'persons of property'. Table 2 also gives the mapping between the categories used in the analysis here and, in column two, the original occupational classification of the census as re-classified by Booth-Armstrong. This shows how the various complexities of mapping the occupational categories into industrial categories have been managed. The most significant difficulty, as noted above, is the overlap between manufacturing and 'dealing' (selling directly to consumers as retailers and/or wholesaling).

Simplified groups ID used here	**Booth-Armstrong**	**Summary of census occupation descriptors (also used in I-CeM)**
1	A 1-4	Farming, fishing, market gardeners, horse breeding and keeping
2	M 1 part	Coal mining
3	M 1 part M 2, 3, 4,	Other mining & quarrying, brickmaking, gravel, salt works
4	B 2	Construction operatives (masons, bricklayers, thatcher, plumbers etc.)
5	MF 1	Machinery mf
6	MF 2	Tool & weapons mf
7	MF 4 part	Iron & steel mf, bolts and nails
8	MF 4 part	Blacksmiths
9	MF 5, 6	Other metal mf (copper, tin, brass, whitesmiths, etc.)
10	MF 3, 15	Ship, road & rail vehicle mf
11	MF 7	Earthenware & glass mf
12	MF 8, 9	Gas, coke and chemical mf
13	MF 10, 11, 12	Leather, fur, hair & bone mf
14	MF 13	Wood mf (sawyers, coopers, cane workers)
15	MF 14	Furnishing mf & dealers (cabinet makers, french polishers, undertakers)
16	MF 16, 30	Printing & paper mf (paper, cardboard, printers, bookbinders)
17	MF 17	Waterproof goods mf (floor & oil cloth, rubber etc.)
18	MF 18	Woollen mf (woollen goods, carpets, blanket, flannel)
19	MF 19, 22	Cotton & silk mf (incl. ribbon, weaving, dyeing, bleaching etc.)
20	MF 20, 21,	Other textile mf (flax, hemp, rope, jute, lace, tape, thread)
21	MF 23 part, 24 part	Clothing mf, (tailors, milliners, hosiery, hats, gloves, umbrellas, buttons, leather)

22	MF 23 part	Shoe, boot, clog mf
23	MF 25 part, 26 part	Agricultural produce mf (millers, refiners, bakers, confectioners)
24	MF 27 part, 28 part, D6 part	Drink & tobacco mf and dealing (maltsters, brewers, distillers, tobacco & pipes)
25	MF 29	Watch & instrument mf
26	MF 31 part	General mf (manufacturers, mechanic, artisan, machinist)
27	T 1, 2, 3	Ocean, inland and dock transport
28	T 4, 5	Road & rail transport
29	D 1	Coal dealing
30	D 2	Timber, hay, corn and agricultural produce dealing
31	D 3, 4	Clothing and dress dealing (drapers, hosiers, haberdashers)
32	D 5	Food sales (butchers, fishmongers, cheesemongers, milksellers, grocers)
33	D 7, 8	Lodging & drink sales (wine & spirits, hotels, inns, coffee ho)
34	D 10	Stationery dealing (stationers, publishers, newsagents)
35	D 11 part	Hhd ornaments dealer (earthenware, glass, jewellers)
36	D 11 part	Ironmongers
37	D 12	Other retail (general shopkeeper, huckster, hawker, pawnbroker)
38	P 8 part	Chemists, druggists,
39	IS 1 part	Merchants, banks, insurers and brokers
40	IS 1 part	Other commerce (accountants, salesmen, travellers, officers of Cos.)
41	B 1	Construction management (architects, builders and contractors)
42	PP 7, 8 part	Professions (barristers, solicitors, scientific pursuits)
43	PP 9, 10, 11	Professions (doctors, dentists, artists, performers, education)
44	DS 3 part	Personal services (washing & bathing, hairdressing, chimney sweeps)
45	PP 1-3, 6, 13-14	Public admin, military, clergy
46	PP 4-5	Military
47	IS 1 part	Clerks, weighers, telegraph, non-theological students, apprentices
48	DS 1, 2, 3 part	Domestic and service staff, cooks
49	B3, IS 2 and parts of others	Labourers & transport staff (including family on farms)
50	PO part	Persons of property with no stated occupation

Table 2. Detailed industrial sector groups, entrepreneurship sector classification ID50, and relation to Booth-Armstrong and census occupational classifications (Column 2: key in Armstrong, Information about Occupations, Appendix E, pp. 284-293).

As is clear from Table 2, the matching to Booth-Armstrong is often complex because they did not differentiate types of skill level (e.g. banker and bank clerk; railway driver and railway

labourer; farmer and farm labourer; builder, carpenter and building labourer). Nor did they attempt to differentiate categories that might be business proprietors and entrepreneurs.

It is accepted that all classifications have difficulties, particularly separating 'business services' from 'personal services' and in handling 'dealing'. Hence all groupings must be treated as indicative rather than definitive, and the best that can be done given the structure of the original historical record. The main approach followed here has been to make decisions on the most appropriate allocation to an aggregated sector based on the sector field of activity. In manufactures this mainly follows the material being worked on, which is largely that used by the census; and in services it is the main category of service provided. This approach is similar to modern definitions (SIC, see below). Within I-CeM occodes checks were made against the full CEB descriptor strings to determine the most frequent uses of the terminology, and how this could be classified at a more aggregate level.

For the purpose of sensitivity to status as business proprietor, the main guidance on the expected scope of employers and own account as distinct from those of employee or worker status was derived from the statements of employer status given in the census over 1891-1911. This allows categories to be distinguished which have clear business status, from those that were workers and no employer or own account status. The non-business categories were grouped together in categories 45-49, with the problematic category 50 coded separately to allow more complex analysis if required. For categories 45-49 a detailed cleaning of all descriptor strings was undertaken which achieved two objectives: first, to remove a large number of people who were spuriously coded as employers or own account that had entered either because of census errors or I-CeM miscodes; the I-CeM miscodes were then corrected and the corrections applied to the whole database; second, the categories judged to be incompatible with employer or own account status which were re-assigned to worker status (e.g. domestic servants). The full range of decisions taken on re-coding of occupations and status is described in Working Paper 3 and 4. It must be recalled that the CEBs do not have the benefit of the checks that were used by the GRO to correct household or enumerator errors, and so some of this has to be re-imposed in the processing. In identifying the occupations allocated to categories 45-49, in addition to the information provided by the Booth-Armstrong classification and the instructions to census clerks about classifying occupations, the proportions in each occupation returned as employers, own account and

workers in 1891-1911 were used to identify occupations that were predominately workers or, indeed, economically inactive.

In categories ID18: 1-13 and ID50: 1-44, in contrast, there is expected to be a mix of employers, own account and workers which have to be distinguished by more complex methods *within* occodes. This is discussed at greater length in later publications. The main limitations to be noted are those arising from very generalised descriptions for many occupations that give insufficient detail to be certain of industry sector, most commonly arising from categories such as 'General labourer', 'Office keeper (not Government)', 'Commercial clerk', and problems that result from separating 'business services' from 'personal services' for individuals such as lawyers or auctioneers that often serve both, or cleaners that operate in business premises as well as homes.

Tables 1 and 2 summarise the categorisation developed for 18 and 50 groups and the relation to Booth-Armstrong. In the entrepreneurs database, each individual has a number of different occupation codes: the original I-CeM occode, an 18ID, 50ID, a PST code and a modern SIC code. Therefore, the entrepreneurs can be organised and analysed according to the particular classification any user prefers. A look-up matrix between I-CeM's occodes, 18ID, 50ID, PST and the modern SIC is also available with the database. For analysis purely of entrepreneurs the last 5 categories include no entrepreneurs as a main occupation, so that for some analysis a reduced list of categories for ID13 and ID45 is often more appropriate.

2.2 Aggregate codes for Economically Active: EA17 and EA51

The advantage of the ID18 and ID50 classifications is that a very large number of workers are automatically and precisely classified as workers leaving the difficult question of separating employers own account and workers for the analysis of the remaining sectors. For analysis of entrepreneurs this has many advantages. But for analysis of the entire economically active it has the consequence of separating labourers and others who are definitionally workers from those in the rest of the sector. As a result entrepreneurship ratios and total employment numbers in each sector are adjusted and become misleading. For analysis of the whole economically active, therefore, another classification is developed: for the economically active: EA. This is also based on Booth-Armstrong and follows the main categories of the ID classifications used above, but with the important difference that as far as possible all

workers are placed into the same categories as the rest of that business sector; e.g. all farm labourers and family working on farms are in farming; all transport labourers are in transport, all factory hands and textile labourers to manufacturing. At the most aggregate level, EA17, a classification is developed as shown in Table 3.

Aggregate EA sector	**Detailed sector(s) as in ID50**	**Summary title**
1	1, 49	Farming and fishing [with farming labourers and family workers]
2	2-3	Mining & quarrying
3	4	Construction
4	5-6, 9-20; parts of 7-8, 22-24, 26	Manufacturing
5	Parts of 7-8, 21-24, 26, 35-36	Makers and dealers
6	29-34, 37-38; parts of 35-36	Retail & ironmongers
7	27-28	Transport
8	39-42	Professional & business services
9	43-44	Professional & personal services
10	13-14, 23-24, 30	Agricultural produce, drink & tobacco manuf.
11	32-33	Food retailing
12	33	Lodging & refreshment
13	39, 47	Finance & commerce
14	45-46	Public admin, military, clergy
15	48	Domestic and service staff
16	49	Undefined general labourers
17	50	Persons of property with no stated occupation

Table 3. Aggregate industrial sector groups: EA 17 sector classification and relation to ID50.

It can be seen that EA17 is very similar ID18 but reallocates all or part of the ID18 categories 15 and 17, for ‘clerks, weighers, telegraph, non-theological students and apprentices’, and ‘Labourers & transport staff (including family on farms and factory hands and textile labourers)’. This unites all recognisable workers within the sectors for farming and

commerce, but unfortunately still leaves a residual of general labours that cannot be attributed reliably to any one aggregate sector. They are probably mostly construction labourers, but there will be some in other parts of manufactures, transport, and also in making and dealing. This means that analysis of these sectors has the constraint that the general labourers have to be brought back into consideration at the appropriate point. It is possible in some cases to develop plausible estimates of distinction of general labourers between sectors, but otherwise the constraints must be dealt with along with other specifics in each different type of analysis: as discussed in subsequent publications for each situation.

The disaggregated version of the classification of economically active, EA51 is shown in Table 4. The main distinctions between EA51 and ID50 are: the attribution of workers as far as possible to the main sectors, reducing the numbers assigned into categories EA 48-50; the separation of textile and other factory hands/labourers; and the reduction of the unassigned commercial solely to 'commercial clerks'. General labourers remain unassigned. It is noteworthy that because of its greater aggregation, EA17 allows the unassigned commercial clerks to be allocated to a single commercial sector, but this is not possible in EA 51; similarly general and textile factory hands cannot be allocated between groups in EA51, but are all in manufacturing in EA17.

Simplified EA groups used here	**Summary of census occupation descriptors**
1	Farming, fishing, market gardeners, horse breeding and keeping
2	Coal mining
3	Other mining & quarrying, brickmaking, gravel, salt works
4	Construction operatives (masons, bricklayers, thatcher, plumbers etc.)
5	Machinery mf
6	Tool & weapons mf
7	Iron & steel mf, bolts and nails
8	Blacksmiths
9	Other metal mf (copper, tin, brass, whitesmiths, etc.)
10	Ship, road & rail vehicle mf
11	Earthenware & glass mf
12	Gas, coke, water and chemical mf
13	Leather, fur, hair & bone mf
14	Wood mf (sawyers, coopers, cane workers)
15	Furnishing mf (cabinet makers, french polishers, undertakers)
16	Printing & paper mf (paper, cardboard, printers, bookbinders)
17	Waterproof goods mf (floor & oil cloth, rubber etc.)
18	Woollen mf (woollen goods, carpets, blanket, flannel)
19	Cotton & silk mf (incl. ribbon, weaving, dyeing, bleaching etc.)

20	Other textile mf (flax, hemp, rope, jute, lace, tape, thread)
21	Clothing mf (tailors, milliners, hosiery, hats, gloves, umbrellas, buttons, leather)
22	Shoe, boot, clog mf
23	Agric. produce mf (millers, refiners, bakers, confectioners)
24	Drink & tobacco mf (maltsters, brewers, distillers, tobacco & pipes)
25	Watch & instrument mf
26	General mf (manufacturers, mechanic, artisan, machinist)
27	Ocean, inland and dock transport
28	Road & rail transport
29	Coal dealing
30	Timber, hay, corn and agric. produce dealing
31	Clothing and dress dealing (drapers, hosiers, haberdashers)
32	Food sales (butchers, fishmongers, cheesemongers, milksellers, grocers)
33	Lodging & drink sales (wine & spirits, hotels, inns, coffee ho)
34	Communications (publishing, newsagents, stationers and telecoms)
35	H/h & personal goods dealer (earthenware, glass, jewellers)
36	Ironmongers
37	Other retail (general shopkeeper, huckster, hawker)
38	Chemists, druggists
39	Merchants, banks, insurers and brokers
40	Other commerce (accountants, salesmen, travellers, officers of cos.)
41	Construction management (builders and contractors)
42	Professions (barristers, solicitors, scientific pursuits)
43	Professions (doctors, dentists, artists, performers, education)
44	Personal services (washing & bathing, hairdressing, chimney sweeps)
45	Public admin, clergy
46	Military
47	Domestic service
48	Undefined general &factory labourers
49	Factory hand (textile, undefined)
50	Commercial clerks
51	Persons of property

Table 4. Detailed industrial sector groups, entrepreneurship sector classification EA51.

It should be noted that the classifications developed here need further detailed attention when applied to a specific year, since each census year has slightly adjusted instructions that lead to some categories being recorded in different ways, some people being more fully included or excluded, etc. The changes in instructions affecting business proprietors are summarised in Working Paper 2. The impact of these changes is that the classifications have to draw on further detailed re-codes using either the occodes themselves, and/or the descriptor strings. The adjustments required are summarised in later working papers.

In addition, as noted at the outset, all classifications have difficulties and are at best approximations. For the census classifications that can be developed for 1851-1911 the three greatest difficulties are, first, being unable easily to allocate general labourers, factory hands and commercial clerks; second, separating 'business services' from 'personal services'; and third, managing any distinction between 'makers' and 'dealers'. The difficulties of handling these groups are discussed in later working papers. Our ID and EA groupings are not held out as more than indicative, and the best that can be done given the structure of the original historical record.

3. Comparisons with PST

One of the other main occupational classifications, which also seeks to identify sectors and is widely used to analyse the Victorian census, is the PST (primary, secondary, tertiary) system devised by Tony Wrigley.[20] Wrigley's PST system provides a range of occupational codes for British historical occupational descriptors encountered in the census, parish registers and other sources. The classification and coding tables are available from the Cambridge Group for the History of Population and Social Structure (Campop).[21] PST categorises occupations according to a four point system, which in the current 'Feb2010' version contains 1,635 distinct occupational categories. There is a dictionary of 2,614 distinct occupations and their PST code and a look-up table containing 31,991 variant spellings which allows it to be compared with other groupings. Each four-part code indicates, hierarchically, (1) the PST main **sector**, (2) general industry **group**, (3) specific industry or **section**, and (4) the specific **occupation**. For example, 2,52, 1, 2 is the code for a watch maker; where the code 2 indicates the secondary sector, 52 is the instrument-making industry group, 1 is the clock and watch making section of that industry, and 2 the occupation of making watches.[22]

PST was designed by Wrigley to capture sectoral transformation through occupational change during the industrial revolution, which he interpret as the transition from an 'organic economy' where much of the population must work in farming to sustain themselves to an industrial society in which constraints on growth are overcome, and rising wages increase

[20] Wrigley, E. A. (2010), *The PST system* .
[21] See: http://www.campop.geog.cam.ac.uk/research/projects/occupations/britain19c/pst.html
[22] Wrigley, *The PST system*, p. 13-17.

demand for manufactured goods and, especially, services, creating positive feedback that incrementally alters the occupational structure. Wrigley's work on occupational classification takes as its starting point the three sector concept identified by Colin Clark, who had used the writings of the early English political economist William Petty to understand the transition process between the primary (raw materials) and secondary (manufacturing) sectors.[23] PST was designed to identify shifts in the labour force (rather than industries) from primary production towards secondary sector work converting agricultural output and raw materials into finished products, and to tertiary sector service roles 'downstream' of manufacture such as in transport, wholesale and retail, as well as employment less associated with primary or secondary sector activity, such as the professions.[24] The PST system allows measurement of the changing relative contribution of primary, secondary, and tertiary occupations over the period of transition: in general, the primary sector declined, whilst the secondary and especially the tertiary sectors grew.[25]

PST takes as its evidence base both occupational descriptors assigned to individuals (as in parish baptism registers), and the most disaggregated categories of occupations used in original census reports, aligning them to a common format. Although it is strongly orientated toward the three sectors of primary, secondary, and tertiary, after this initial sectoral split the main focus is on bottom-up coding of *occupational* titles, reflecting the origins of the classification system as a tool for measuring occupational change over time. As a result the categorisation does not always align easily to different economic sectors below the 3-sector starting point; nor does it align to categories of entrepreneurial status (except in farming). Moreover, most of the use of PST for the 1851-1911 censuses is based on encoding the published census occupational categories rather than individual occupational descriptors.[26] Thus, as with I-CeM, there are difficulties of identifying correct assignment independently of GRO definitions. This interrelates with the greatest difficulty of using PST for identifying sectors: that the attempt to allocate between P, S and T which is essential to the PST methodology, requires arbitrary decisions to allocate maker-dealers into makers and dealers, if the census has not collected the information on these different categories separately in the first place. In many analyses at aggregate level it has been assumed that allocating one group of maker-dealers to one sector is compensated by allocating another group of maker-dealers

[23] Clark, *Conditions of economic progress*, p. 395
[24] Wrigley, *The PST system*, p. 9.
[25] Shaw-Taylor and Wrigley (2014) *Occupational structure and population change*.
[26] See e.g. http://www.campop.geog.cam.ac.uk/research/projects/occupations/britain19c/pst.html

to the other sectors. Thus Shaw-Taylor and Wrigley assume that maker-dealer occupations such as bakers assigned to secondary are compensated by assigning occupations like beer selling and brewing to tertiary.[27]

Hence, while PST has the advantage of identifying sectors, it was designed for a different purpose using different sources alongside the census reports over a very long timespan. Here, since we are targeting a specific group, entrepreneurs, over a shorter period of more uniform source material, a tailored refocussing is more tractable. The preferred method developed in section 2 above is to live with the constraints of the census collection process and identify an occupational category of 'maker-dealers' where the census does not allow differentiation, and not attempt to separate them. PST codes are nonetheless provided in the database deposited from the 'Drivers of Entrepreneurship and Small Businesses' project so that it is possible to translate between the occupational codes in the database and PST codes as may be desired for comparative purposes.

Certain aspects of PST also offer useful guidance for sector aggregations. Only the highest level of aggregation, PST sector, is specifically a 'sector' code. For enhanced flexibility there are in fact seven categories rather than three for PST sector (see Table 5), although 3- 6 may be taken collectively as constituting the tertiary sector. The second level of PST is the group, which has 139 categories which are partly sectoral but mainly occupational; the next two lower levels of PST for section and occupation are occupational.

PST sector	**Sector category name**
1	primary
2	secondary
3	tertiary dealers
4	tertiary sellers
5	tertiary services & professions
6	transport & communications
99	unstated or unoccupied

Table 5. PST classification at top level, PST sector.

[27] Shaw-Taylor and Wrigley (2014) *Occupational structure and population change*, p. 60.

PST sector is a valuable guide to the aggregation developed for the ID18 and ID17 groups. However, there are difficulties in using it for our purposes without customisation:

(1) In PST sector category 5 (tertiary services & professions) there is no separation at this top level of public administration, military, students and others from professionals such as lawyers, engineers, architects, musicians or photographers.

(2) At the top PST sector level there is no separation between workers or operatives, and more skilled staff or entrepreneurs. Thus, for example, PST sector 6 for transport and communications, represents 'Omnibus, Coach, Cab, Owner - Livery Stable Keeper' but also 'Carman, Carrier, Carter, Haulier' , and 'Ship Steward, Cook' or 'Messenger, Porter, Watchman'. Similarly PST sector 1 includes farmers, 'Agricultural Student' and 'Farmer's, Grazier's - Son, Grandson, Brother, Nephew'; The differences are coded at lower levels in PST, but the top level aggregations combine very unlike statuses and skills.

(3) The main difficulty for aggregation is that PST jumps from 3-7 categories at the top level PST sector to 139 subcategories at the second point in the hierarchy, PST group. These 139 groups are quite disaggregated and again do not align well with employment statuses. There are additional problems of under-aggregation as a consequence of the varying detail and lack of direct cross-comparability of occupational categories from one census to the next, which are not fully resolved at any level of the PST classification hierarchy.

(4) Conversely, other categories in PST are not disaggregated at any level, even though they cover a wide range of statuses and skills in practice across primary and secondary; notably the very large categories of 210801 'Blacksmith' and 210802 'Whitesmith' which both sit within PST sector 'secondary'. While it is common for many persons to be enumerated under these traditional occupational titles, this precludes more nuanced classification where the source data do permit. For those entrepreneurs whose employment of others is declared and whose economic interests are described at greater length in the CEBs, a more detailed occupational title more often occurs than for other, non-entrepreneur persons.

Many of these difficulties are inherent in the way occupations are described or categorised in the census data, but the PST approach is at times both too aggregated and too disaggregated. Unfortunately these differences are often critical for the entrepreneurial classification.

Links between PST sector and ID18 and EA17.

PST sector categories can be used as a guide the ID18 and EA17 classifications. This is particularly useful are for the makers and dealers whose contemporary census occupational category is considered partially or wholly equivalent across most censuses until 1911 and later to a census occupational category(ies) that would be classified to a separate PST sector. As far as possible all makers and dealers in PST sectors 2 and 3 so affected have been classified to 5 'maker-dealer' in ID18 and EA17. The exceptions to this are where the ID18 uses a separate code that is more specific: e.g. 'Wine, Spirit - Merchant, Agent', equivalent titles of which in other censuses are attributed to both PST sector codes 2 and 3, but which is coded to 12 refreshment in ID18 and EA17; also 'Cheesemonger, Butterman', and 'Fishmonger' which have PST sector codes of 2 and 4, and 'Confectioner, Pastrycook' and 'Provision Curer, Dealer' which have PST sector codes of 2 and 3, all of which are coded to 11 'food sales (butchers, fishmongers, cheesemongers, milksellers, grocers)' in ID18 and EA17.

Conversely some PST sector categories that are coded as purely 2 manufactures, are given a code of 5 'maker-dealer' in ID18 and EA17; e.g. 'Hatter, Hat Manufacture', 'Shawl Manufacture', 'Shirt Maker, Seamstress', and 'Hosier, Haberdasher' as many were dealers as well as makers, as confirmed in the subdivision of these classifications in later censuses. Blacksmith and whitesmith, both coded to PST sector 2, are coded to ID48 5 'maker-dealer'. The difficult category of 210806 'Ironmonger, Hardware Dealer, Merchant' which is assigned to PST sector 4 'tertiary sellers' is kept in the retail category in ID18 and EA17, although it is recognised that some ironmongers were manufacturers.

PST also makes an important distinction between sectors 3 and 4, tertiary dealers and tertiary sellers; which are wholesalers and retailers. The censuses over 1851-1911 make this distinction very difficult to determine. In the ID18 and EA17 coding, dealers that are believed to have been essentially wholesalers of their own manufactures are coded to manufactures, since every producer has to sell their products, and no distinction is made between wholesalers and retailers. There is an important group in textiles which PST sector codes as manufacturers and tertiary dealers, but are coded to manufactures in ID18 and EA17: e.g. Flax, Linen - Manufacturer, Dealer; Lace Manufacturer, Dealer; Fustian Manufacturer, Dealer; Thread Manufacturer, Dealer; Rope, Twine, Cord - Maker, Dealer;

Mat Maker, Seller; and Sacking, Sack, Bag - Maker, Dealer. The ID18 and EA17 coding is chosen because it is believed in this category none (or very few) were selling direct to the public, and very few were solely wholesale dealers. This decision also aligns with the modern SIC, which does not distinguish between wholesale and retail at the aggregate section level (section G: see below).

Overall therefore ID 18 and ID50, or EA17 and EA51 are mainly consistent with the distinctions made in PST sector, but not in all details, and both schemes are more fine-grained than PST sector but less disaggregated than PST group. The rule employed is to assess the original census questions and classification to determine what it mainly is identifying for each census category; these are then checked against the full CEB descriptor strings to determine the most frequent occupational descriptors encompassed by the classification terminology. It is believed in this way that the best fit to an aggregate classification is achieved, consistent to what respondents actually recorded, and within the constraints of the census instructions and coding that often did not differentiate some important categories (of which makers and dealers are the main cases).

The database deposited from the 'Drivers of Entrepreneurship and Small Businesses' project gives PST codes so that it is possible to for users to translate between the occupational codes in the database if that is desired.

4. Links to modern classifications

It is important for the database deposit of the ESRC 'Drivers of Entrepreneurship and Small Businesses' project to provide the capacity to link the historical data for 1851-1911 to modern definitions. This allows tracking over time and offers a means to align definitions either to modern or earlier periods to facilitate comparisons. There are various classifications that can be used. PST, as discussed above, is one such classification that has been used for long term comparisons for both modern and earlier periods than 1851-1911. There are three others that are discussed here. These cover the main alternatives, which together with PST, offer means to track over time. The one that is the mainly used in the ESRC project is the SIC. In addition other census classifications and the HISCO system are also discussed below.

4.1 The Standard Industrial Classification (SIC)

The Standard Industrial Classification (SIC) is the classification almost universally used for modern analysis of industry and sectors of the UK economy since 1948, and with adaptation is the main source for international sector comparisons of the UK with other countries. A comparable SIC was established in the USA in 1937, and the same approach is followed in most countries. The SIC established in the UK in 1948 followed the UN structure for SIC also launched in 1948.[28] This has been incorporated into the Nomenclature des activites économiques dans la Communauté européenne (NACE) for the EU since 1990.[29] The most recent version of these codes (SIC 2007) has been used since 1 January 2008.[30] The SIC is also used by Companies House in a condensed version for company registration purposes, thus extending its value to company as well as non-incorporated business comparisons.[31] Even a company that is dormant or non-trading still has a SIC code. Alternative occupational classifications are only adopted when the object of study is skills, labour relations, wage levels, seniority and managerial hierarchies, etc..

The SIC focuses on classifying business establishments and units by the type of economic activity in which they are engaged. These, in general, align with the definition used here of business proprietorship (but see discussions of some important differences in later publications). There are 21 main section codes in the SIC, lettered A-U, 88 divisions, 272 groups, 615 classes and 191 subclasses. It is a 4-digit NACE code but with a few sections taken to five digits where this has been thought useful for UK purposes. To illustrate the SIC, ONS quotes the example of carpet and rug manufacture:[32]

Section C	Manufacturing (comprising Divisions 10 to 33)
Division 13	Manufacture of textiles
Group 13.9	Manufacture of other textiles
Class 13.93	Manufacture of carpets and rugs
Subclass 13.93/1	Manufacture of woven or tufted carpets and rugs

[28] ONS (2009) *UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007)* https://www.ons.gov.uk/methodology/classificationsandstandards/ukstandardindustrialclassificationofeconomicactivities/uksicarchive

[29] *NACE Rev. 1 Regulation No 3037/1990* , Official Journal of the European Communities L 293 Volume 33, 24 October 1990 (ISSN 0378-6978); minor revision, NACE Rev. 1.1, Official Journal of the European Communities L6/3 10 January 2002.

[30] ONS (2009) *UK Standard Industrial Classification*, p. 1.

[31] Companies House (2007) *Standard Industrial classification of economic activities (SIC).*

[32] ONS (2009) *UK Standard Industrial Classification*, pp. 1-2.

In this case a carpet and rug manufacturer would have the four digit code C.13.93, with the addition of a fifth digit in this case (C.13.93/1) for woven or tufted carpets and rugs. The full list of SIC at section level is shown in Table 6

Number code	Section SIC Code	Section
1	A	Agriculture, Forestry and Fishing
2	B	Mining and Quarrying
3	C	Manufacturing
4	D	Electricity, gas, steam and air conditioning supply
5	E	Water supply, sewerage, waste management and remediation activities
6	F	Construction
7	G	Wholesale and retail trade; repair of motor vehicles and motorcycles
8	H	Transportation and storage
9	I	Accommodation and food service activities
10	J	Information and communication
11	K	Financial and insurance activities
12	L	Real estate activities
13	M	Professional, scientific and technical activities
14	N	Administrative and support service activities
15	O	Public administration and defence; compulsory social security
16	P	Education
17	Q	Human health and social work activities
18	R	Arts, entertainment and recreation
19	S	Other service activities
20	T	Activities of households as employers; undifferentiated goods and service-producing activities of households for own use
21	U	Activities of extraterritorial organisations and bodies

Table 6. SIC 2007, main section codes.

The advantage of the SIC is its widespread international use, availability in the modern censuses, coverage of all sectors, clearly differentiation of different stages of activity, and application to all individuals (employers, own account and workers). It is the coding manly used in the ESRC project to relate the historic data to modern classifications, though modifications in detail are necessary because of incompatibilities with historical coding, most notably the aggregation of makers and dealers in the early censuses. In practice this means that when tracking historical changes over time after 1911 a few very aggregate sectors have to be used.

4.2 Other census and ONS census classifications

The main focus of the ESRC project and coding is for 1851-1911. For each of the censuses after 1911 changes occurred in occupational classification and the format of the questions used. The changes make long-term comparisons with the modern period difficult.

However, the definition of own account and employers in government surveys (including the census) stabilised by the mid-20th century and remained almost constant so that, with some caveats over detail, the self-employment counts 1961-2011 are relatively consistent. Employers in government activities, nationalised industries and of domestic-staff employers were excluded throughout. The main potential differences are (i) Some employers may be included in own account; (ii) Those working for an employer in their own houses as 'out-workers' were sometimes classified as own account; (iii) Part-timers were included in the occupation questions from 1931, and the treatment and definition of part-time varied; (iv) Managers and foremen were included with employers in some tabulations.

The question of which 'industry' individuals were occupied in was introduced into the general census instructions in 1891, phrased more fully as 'industry or service with which connected' in 1901. Employer's names were requested partially in 1911 and fully from 1921, when the first full industry tables were published. Industry was not extensively classified other than through occupations until 1921 (with a preliminary version used in 1911). From 1921 a varied range of industry groupings termed 'orders' were used in the census that related to national economic statistics. These changed in detail significantly over time and make long term comparisons difficult. After the Standard Industrial Classification (SIC) was introduced in 1948, and used in census tabulation for industry tables from 1951 (see above),

comparisons become easier. However, the SIC itself also changed over time so that long term comparisons must be treated with care.

The census also began to use the Standard Occupational Classification (SOC), which is a classification through the jobs performed. The workers classified to a particular industry fall into a number of different categories of SOC and similarly the workers in some SOCs may be in many different industries. The SOC was introduced in its modern form in 1990, and used in the census form 1991. It now has nine 'major groups' (see Table 7), 25 sub-major groups, 90 minor groups and 369 unit groups. The core to the SOC is the concept of a 'job', defined as a set of tasks or duties to be carried out and representing a basic element in the employment relationship.[33] It assumes that jobs are usually structured by employers, professional bodies, employer and/or worker organisations and government, and by workers themselves for self-employment. Jobs are recognised primarily by the associated job title, which are classified into groups according to 'skill level' and 'skill specialisation', defined by the duration of training/qualification and/or work experience, and in some areas by the type of work performed or materials or tools used.

	Major Group
1	Managers, directors and senior officials
2	Professional occupations
3	Associate professional and technical occupations
4	Administrative and secretarial occupations
5	Skilled trades occupations
6	Caring, leisure and other service occupations
7	Sales and customer service occupations
8	Process, plant and machine operatives
9	Elementary occupations

Table 7. Major groups in the SOC.

[33] http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-1-structure-and-descriptions-of-unit-groups/index.html

The SOC bears some resemblance to the nineteenth-century occupation classification, with its similar focus on the materials used; however, its focus on skill level and the qualifications required is significantly different. Consequently, conversion of the historical census data into SOC classifications would be difficult. While the presence of the occupation titles in I-CeM means that they could be recoded to SOC, albeit through a time-consuming process, the changes in the structure of the labour force, technology and in the education system between the nineteenth and the late twentieth century suggest that such a comparison would be of suspect validity.

4.3 Other national and international sector classifications

Three other sector and industry classifications are relevant to long term comparisons and can be aligned through guides from the Office for National Statistics (ONS), usually based on adaptations of the SIC.

First, is the Sector Classification for the National Accounts.[34] This allocates each business and economic unit to five institutional sectors: General government, Financial corporations, Non-financial corporations (private and public), Non-profit institutions serving households, and the 'Rest of the world'. In principle, the classification embraces all economic units engaging in transactions in goods and services and financial assets, including people and households and overseas concerns, corporations and public bodies. This industrial classification brings together units engaged in similar activities, irrespective of ownership. The SIC code in conjunction with the legal status and UK/non-UK ownership status of a unit provides an approximation to the sector code.

Second, is the individual product classification, where each product (good or service) is, in general, classified to the industry in which it is mainly produced. The UN Central Product Classification (CPC) is an international versions of this.[35] The UN documentation of the CPC provides direct links to the Harmonized Commodity Description and Coding System (HS) and the Standard International Trade Classification (SITC) described below. The EU

[34] The UK *Sector Classification for the National Accounts*; see ONS *Sector classifications Guide MA 23*.
[35] CPC (*Central Product Classification*) (2006) *CPC ver.2*, adopted by UN Statistical Commission March.

Classification of Products by Activity (CPA) is similar, but closer to the NACE and SIC.[36] The CPA 2008 relates directly to the classification structure in NACE Rev. 2, with the first four digits almost always the same, and is linked to the CPC for the fifth and sixth digit. The EU's CPA is also linked to PRODCOM, the EU system of production statistics for mining and manufacturing (excluding services, other than 'industrial services').[37]

Third are classifications of import and export records. The UN Standard International Trade Classification (SITC) and the World Customs Organisation Harmonized Commodity Description and Coding System (HS) give the worldwide basis for national classifications of tariff and trade statistics used by WTO and most other analyses. The EU has a more detailed Combined Nomenclature (CN) based on HS that is used for trade data and is collected customs procedures.

4.4 The HISCO system

The 'historical international standard classification of occupations' (HISCO) system was adapted from ISCO (version ISCO68), originally developed by the International Labour Organisation (ILO). They sought to 'provide a systematic basis for presentation of occupational data relating to different countries in order to facilitate international comparisons …[and] provide an international standard classification system which countries might use in developing their national occupational classifications', allowing for occupations to be described in more one language. This was not designed from scratch, but sought 'to historicise a system with proven comparative credentials'. [38]

HISCO uses a three-digit system (major groups, minor groups, and unit groups). The seven major groups are: (1) professional, technical and related workers; (2) administrative and managerial workers; (3) clerical and related workers; (4) sales workers; (5) service workers; (6) agricultural, animal husbandry and forestry workers, fisherman and hunters; (7) production and related workers, transport equipment operators and labourers. These are based

[36] *Classification of Products by Activity* (CPA) Regulation (EEC) No 3696/93, Official Journal of the European Communities L342 Volume 36, 31 December 1993; various subsequent revisions.
[37] *PRODCOM Regulation (EEC) No 3924/91*, Official Journal of the European Communities L 374/1, 31 December 1991
[38] Van Leeuwen, M. H .D., Maas, I., and Miles, A. (2000) *HISCO: Historical international standard classification of occupation*, Leuven., p. 10; also see summary in Wrigley, *The PST system*, p. 20-22.

as much on seniority and supervisory status as on industry sector. It is like the original census, more of an occupational and social status classification than an industry or sector one.

HISCO is an important international comparative classification that is widely used in historical and other research. However, it has a number of drawbacks for application in the ESRC project. First, as noted by Wrigley, the focus on seniority results in splitting many occupations in the same sector between the different major groups. Thus those in senior positions in one industry or profession are classified with those in comparable positions in other industries, and are separated from those in more junior clerical or manual occupations. As a result Wrigley concludes that HISCO and PST 'are so different that it is fruitless to attempt' to compare between the major categories, although 'difficulties become progressively less serious when starting from more detailed categories,' with census-derived categories often difficult to align.[39] Second, for the purpose of focusing on sectors, makers and dealers remain confused. Third, entrepreneur and business proprietor categories are not separated.

For these reasons HISCO is not used in the ESRC project. However, HISCO codes are given in I-CeM. As the database deposited from the 'Drivers of Entrepreneurship and Small Businesses' project retains all the crucial attributes of I-CeM, especially the ID number of each individual in each household, it is fully I-CeM compatible and it is possible to translate between the occupational codes in the database and HISCO codes if that is desired.

5. Conclusion

This paper has outlined the method by which two levels of aggregation of I-CeM occupations have been undertaken to achieve classifications that approximate industry sectors for ID18 and ID50, or EA17 and EA51 categories for the censuses 1851-1911. The relation of these classifications to PST and the modern SIC are given in the entrepreneurs database deposited at UK Data Archive (UKDA).

[39] Wrigley, *The PST system*, p. 21-22.

The starting point of the 1851-1911 censuses is constrained by the information provided by householders and what was recorded by the census enumerators at the time. The main limitations are those arising from very generalised descriptions that give insufficient detail to be certain of industry, most commonly arising from categories such as ‘general labourer’, problems that result from separating ‘business services’ from ‘personal services’, and coding ‘makers’ and ‘dealing’ when the descriptor given is very brief.

The classifications developed also need detailed adjustment when applied to a specific year to take account of any adjusted census instructions that led to some categories being recorded in slightly different ways; the adjustments necessary are summarised in later working papers. It is also accepted that all classifications have difficulties and this is certainly the case with the early censuses. Hence the grouping presented here must be treated as indicative rather than definitive, and with the aim of being best that can be done for the purpose given the structure of the original historical record.

Acknowledgments:

This research has been supported by the ESRC under project grant ES/M010953: **Drivers of Entrepreneurship and Small Businesses.** Piloting of the research for 1881 draws from Leverhulme Trust grant RG66385: **The long-term evolution of Small and Medium-Sized Enterprises (SMEs).**

The database used for 1891 and 1901-11 derives from K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2),* (2016) [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1; enhanced; E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex, 2015).

The GIS boundary files for RSDs were constructed by Joe Day for the ESRC fertility project directed by Alice Reid:
http://www.geog.cam.ac.uk/research/projects/victorianfertilitydecline/publications.html

These used as a starting point the GIS parish files of Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., Wrigley E.A. (2006) *1851 England and Wales census parishes, townships and places*, 2006, ESRC RES-000-23-1579, supported by Leverhulme Trust and the British Academy; Satchell, A.E.M. (2015) *England and Wales census parishes,*

townships and places; which is an enhanced and corrected version of Burton, N, Westwood J., and Carter P. (2014) *GIS of the ancient parishes of England and Wales, 1500-1850*, UKDA, SN 4828; which is a GIS version of Kain, R.J.P., and Oliver, R.R. (2001) *Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata*, UKDA, SN 4348.

A special acknowledgement of thanks is made to Kevin Schürer for advice and all his help in developing improved versions of I-CeM, and to Alice Reid, Eilidh Garrett, Joe Day, Hanna Jaadla, Xuesheng You, Leigh Shaw-Taylor and other members of the Campop I-CeM group who, with the current authors, have collectively worked on the new versions of I-CeM.

References.

Acland, A. H. D and 45 others (1888) Memorandum on the Improvement of Census Returns, Especially as Regards Occupations and Industry, pp. 118-20 in *Report of the Committee Appointed by the Treasury to Inquire into Certain Questions Connected with the Taking of The Census*, C 6071 (1890), HMSO, London.

Anderson, Michael, (1988) Households, Families and Individuals: Some Preliminary Results from the National Sample from the 1851 Census of Great Britain, *Continuity and Change*, 3, 421-38.

Armstrong, W. A. (1972) The Use of Information about Occupations, pp. 191-210 in E. A. Wrigley (ed.) *Nineteenth-Century Society: Essays in the Use of Quantitative Methods for the Study of Social Data*, Cambridge University Press.

Booth, Charles (1886) Occupations of the People of the United Kingdom 1801–1881, *Journal of the Royal Statistical Society*, 49, 2, 314-435.

Clark, C., (1951) *The Conditions of Economic Progress*, 2nd ed., London.

Companies House (2007) *Standard industrial classification of economic activities (SIC)*, Companies House, Cardiff.

Higgs, Edward (1995) Occupational Censuses and the Agricultural Workforce in Victorian England and Wales, *Economic History Review*, 49, 700-16.

Higgs, Edward (2004) *Life, Death and Statistics: Civil Registration, Censuses and the Work of the General Register Office, 1836-1952*, Local Population Studies, Supplement, Hatfield.

Higgs, Edward (2005) *Making Sense of the Census Revisited: Census Records for England and Wales 1801-1901*, Institute of Historical Research, National Archives, London.

Morris, R. J. (1990) Occupational Coding: Principles and Examples, *Historical Social Research*, 15, 3-29.

ONS (2009) *UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007): Structure and explanatory notes*, Office for National Statistics, Palgrave Macmillan.

PP (1890) *Report of the Committee Appointed by the Treasury to Inquire into Certain Questions Connected with the Taking of The Census*, C 6071, Parliamentary papers, London.

Schürer, Kevin and Woollard, Matthew (2002) *National Sample from the 1881 Census of Great Britain 5% Random Sample: Working Documentation Version 1.1*, Historical Censuses and Social Surveys Research Group, University of Essex.

Schürer, Kevin, Higgs, Eddy, Reid, Alice M., Garrett, Eilidh M. (2016) *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, [data collection].

Shaw-Taylor, Leigh and Wrigley, E. A. (2014) Occupational structure and population change, pp. 53-88 in Floud R, Humphries J, Johnson P. (eds.) *The Cambridge Economic History of Modern Britain: Volume I, 1700-1870*, 4th ed. Cambridge: Cambridge University Press.

Van Leeuwen, M. H. D., Maas, I., and Miles, A. (2000) *HISCO: Historical International Standard Classification of Occupations*, Leuven.

Wollard, Matthew (1997) Creating a Machine-Readable Version of the 1881 Census Enumerators' Books, pp. 98-101 in Charles Harvey and Jon Press, *Databases in Historical Research: Theory, Methods and Applications*, Macmillan, Basingstoke.

Wollard, Matthew (1997) 'Shooting the Nets': A Note on the Reliability of the 1881 Census Enumerators Books, *Local Population Studies*, 59, 54-7.

Woollard, Matthew (1999) *The Classification of Occupations in the 1881 Census of England and Wales*, Department of History, University of Essex.

Wrigley, E. Anthony (2010) *The PST system of classifying occupations*, Working Paper 20, Cambridge Group for the History of Population and Social Structure, http://www.campop.geog.cam.ac.uk/research/projects/occupations/britain19c/papers/paper1.pdf

Other Working Papers:

Working paper series: ESRC project ES/M010953: *'Drivers of Entrepreneurship and Small Business',* University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design.*

WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution.*

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*.

WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911*.

WP 5: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*.

Full list of all current Working Papers available at:
http://www.geog.cam.ac.uk/research/projects/historyofentrepreneurship/