



Munich Personal RePEc Archive

**Reconstructing entrepreneur and
business numbers for censuses 1851-81.
Working paper 9.**

Bennett, Robert and Montebruno, Piero and Smith, Harry
and van Lieshout, Carry

Department of Geography, University of Cambridge, Cambridge
Group for the History of Population and Social Structure (Campop)

February 2018

Online at <https://mpra.ub.uni-muenchen.de/103529/>
MPRA Paper No. 103529, posted 13 Oct 2020 21:00 UTC

Reconstructing entrepreneur and business numbers for censuses 1851-81

1. Introduction

This paper develops a reconstruction method to identify entrepreneurs 1851-81. Its aim is to reconstruct the numbers of employers and own account business people for 1851-1881, which are the years where employment status was not explicitly identified in the population censuses. The individuals are identified with varied levels of certainty. These will be invaluable resources for subsequent research and will be available in the entrepreneurs database deposited at UK Data Archive (UKDA), derived from the ESRC-supported project ES/M010953 *Drivers of Entrepreneurship and Small Businesses*. The final reconstruction is used for a wide range of aggregate analysis in the ESRC- *Drivers of Entrepreneurship and Small Businesses*. An overview of the project is provided in Working Paper 1, which shows how 1851-81 differ from the subsequent censuses. Working Paper 2 defines in detail the different censuses and the challenges they present for identifying entrepreneurs. Working Paper 3 describes how the data for 1851-81 were extracted from the censuses from different sources. A full list of Working Papers is given at the end of this paper.

The main source used in this paper is transcripts of the census, mainly as they are entered into the I-CeM electronic database for the censuses 1851-1911 produced by a team at the University of Essex, deposited at the UKDA: *The Integrated Census Microdata (I-CeM)*.¹

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>;

Also used are infills of gaps and truncations in I-CeM which have been added from a separate source (S&N: The Genealogist) for 1851 (about 14 % of all extracted records, and 26% of all employers) and for 1871 (100% of selected records). For 1871 the census records available are restricted in the project to those extracted from Groups 1-6 (as defined below); employers, masters and farmers as described in Working Paper 1. These sources provide transcriptions of the original Census Enumerators Books (CEBs)² as well as enhancing the data with various codes for household structure, relationships between people within households, and occupations. Within the data is information on employment status: whether individuals are *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.

A key aim of the ESRC project is to construct, as far as possible, an aligned, complete, quality-controlled, and consistent database of business proprietors/businesses for 1851-1911, for database deposit at UKDA to provide an open and generally accessible new resource for identifying the business population. It is sought to make this database as complete as possible in order to allow tracking sectors, geographical units, and different modes of business organisation over time on a consistent basis, and thus contribute a key resource for research on businesses, self-employment and economic development over time.

This paper seeks to fulfil this goal by reconstructing employers and own account for the early years where they are not explicitly recorded. The information on ‘employment status’ was collected by the censuses from 1891 onwards. In contrast, although the 1851-81 censuses did identify ‘employers’ and also collected information on their employees and the size of farm acreage, this is only partial for many occupations, and there was no explicit identification of ‘own account’, or explicit distinction from other statuses such as worker, unoccupied, etc.

The aim of this paper is to develop a methodology to ‘reconstruct’ the ‘own account’ and enhance the employers in order to provide a continuous series of data that can be used for analysis of evolution of employment status over the whole period 1851-1911. In practice, the

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).

² e.g. ‘General Instruction’, Census of England and Wales, Householder’s Schedule, 1851.

reconstruction of the own account cannot be undertaken independently of identifying (or confirming) the status of employers, workers, and others, so that a complete methodology for reconstructing or confirming ‘employment status’ as a whole is required. To make the reconstruction credible and reliable five objectives have to be satisfied:

1. The reconstruction should to be as accurate as possible, within specified probability bounds; with the target that overall accuracy (total numbers and assignments) will be at least as good as those where status is explicitly identified in the censuses over 1891-1911 and is encoded in I-CeM. I-CeM coding is generally accurate to within $\pm 5\%$, and this is used as a rough guide to some decisions.
2. The definitions should be aligned as far as possible between the reconstructed data for 1851-81 and the later series for 1891-1911 so that comparisons over time between the two are valid. This is a very important element in reconstruction decisions; however, the way in which some categories were originally defined or collected in the census, or the way they have been coded by I-CeM, result in some categories that have anomalies between census years. These are identified and recorded below.
3. There should be validation, as far as possible, that the reconstructions match with other sources of information distinguishing own account, employers and workers.

In addition, for further research and analysis, it is sought to satisfy two additional aims:

4. The data should be available at an acceptable confidence level for disaggregated spatial units (in practice this is sought at the level of Registration Sub-Districts: RSDs).
5. The data should be available at the level of individuals. This is the most ambitious objective, and it cannot be fully achieved; but it is sought to provide information at various levels of confidence: ranging from certain or ‘known’ identification of employment status, to those only identified with a given probability level.

Whilst every effort has been made to ensure accuracy, the aim of this paper is to provide reconstruction estimates for aggregates. Whilst the individual level is constructed for each economically active person in the censuses, many of these are identified only at a statistical level; some are identified by random selection. The individuals are identified in order to then provide a means to estimate accurate aggregates (for sectors, areas or other large groups). The

individual level should be used only with care, taking account of how the cases were estimated; generally fully accurate identification of an individual's employment status is only available for extraction Groups 1-6 described below).

The background to the cleaning and screening of the census data; the methods of extraction of those identified as 'employers' 1851-81; the identification, cleaning, screening and correction of all employment statuses 1891-1911; and the background to what the different censuses covered is provided in previous working papers (primarily WP1, WP2, WP3 and WP4). This paper focuses on the reconstruction method. It first overviews the methods for reconstruction (Section 2). Section 3 applies the methods to 1881; Section 4 then applies the same methods to 1851-71; Section 5 gives a brief overview of the main trends. In each case the results are presented at aggregate level, by occupational categories, and compared with the data extracted directly from the census CEBs. Wider validation against other sources is undertaken in other publications.

2. Reconstruction: overview of methodology

The reconstruction method follows a series of stages: (1) data preparation; (2) logit regression estimator for combined employers and own account; (3) validation against the 1891 census and inspection of residuals; (4) comparison with multinomial logit estimator of separated employers and own account; (5) estimates of reconstruction and comparisons with census extractions of employers and own account; (6) final choice of reconstruction method to give reconstructed aggregate numbers; (7) assignment of individuals to employment status as entrepreneurs or workers; and (8) assignment of individuals to employment status as employers or own account. Each is discussed in turn below, first for 1881, and then for the earlier years.

2.1 Data Preparation

2.1.1 Preparation of 1891-1911

First it is necessary to process and fully align the censuses 1891-1911, making adjustments where necessary to ensure that the earlier censuses can be made consistent. This process is

summarised in WP4, which details the pre-screening by age and occupation applied to the later census. Second, a method is developed to align and provide consistent data on employment status. This is specified in detail in Bennett (2018). Key features of this process which are critical to the reconstruction for earlier censuses are:

- Compensation through a method of occupational data cleaning and weighting for non-response bias 1891-1911 where a considerable number of people did not respond to the employment status question.
- Construction of a corrected version of 1891 employment status; its main features are to pre-screen those where their occupational descriptors were clearly workers; and to reassign those stating ‘employer’ who were more accurately own account.
- The full range of occupation categories (Occodes) defined in I-CeM have been checked and many individuals re-coded to their correct occupation categories.
- A set of 83 additional Sub-Occodes has been developed that help to better identify employment statuses; these are defined for those Sub-categories where the variance of individual status within an occupational category is large, and it is possible reliably to identify subgroups where the statuses of employer, own account or worker are more clearly defined. This is undertaken by using occupational descriptor strings as returned in the original CEBs to recode those in the Occode to: either to the new Sub-Occode, or to the residual of the original Occode. The new Sub-Occodes are listed in Appendix 1 with Occode numbering derived from the original I-CeM Occodes.
- Note that in all the subsequent discussion all those identified as ‘land owners’, ‘house owners’ and all those who were owners of non-business assets and who cannot be identified as definite business proprietors are kept as a separate category for separate analysis (see later below). These are distinct from owners of business assets such as boats or threshing machines, quarry owners, coal mine owners, boat owners, etc., who are included as definite business proprietors.

2.1.2 Preparation of 1881 census for reconstruction

The 1881 census is used as a pilot for the other earlier years. As the closest census to the 1891-1911 period, the patterns of entrepreneurship in that year should bear the closest resemblance to those observed in the later censuses and, consequently, the reconstruction

should be most accurate. For the 1881 census year an earlier e-version of the census transcribed by the Genealogical Society of Utah (GSU) was coded and deposited in the 1990s by Schürer and Woollard.³ Because extensive work had been done on these records to correct occupational coding errors, and further work had been undertaken specifically to identify and code employers and own account in a pilot for the current project,⁴ it was important to transfer this information across to the I-CeM codings. This was essential to maintain consistency of the 1881 research with that for the other census years, but at the same time it was necessary to recode the 1881 I-CeM version to include all corrections, developments and piloting work. The result is an updated I-CeM version of the 1881 census that includes any corrections to the GSU version made by the pilot project and not already corrected when I-CeM was created. Most of these changes involved occupation coding corrections. Additionally, all individuals were given a Sub-Occode as described above.

Other aspects of preparation of the 1881 data required involved:

- Pre-screening all entries by age, non-retired, unoccupied, non-scholar etc. using the same rules as for 1891-1911 so that the whole population is aligned with the same definitions as in 1891-1911 (except for the few special cases of retired giving employees). This ensures all economically active are included.
- As in 1891-1911, individuals in some Occodes were assumed to be workers, regardless of their response. These Occodes are detailed in Working Paper 4, but include those for domestic service, general labouring etc.
- Adding additional pre-screening to align Occodes and other coding between the earlier and later censuses. The main elements of this assignment are to identify employment status for a number of Occodes, or groups within an Occode, where they emerged as disruptive to the reconstruction. These are all included in the deposited database. The main aspects to note are:
 - Occode 145 Van guard, boy etc. were significantly mis-coded in I-CeM and after re-coding only a small residual (5 individuals) have non-worker status in

³ See Schürer, Kevin and Woollard, Matthew (University of Essex) (2000) 1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version) [computer file] UKDA, SN-4177, transcribed by Genealogical Society of Utah and Federation of Family History Societies.

⁴ See discussion of methods employed in the original pilot project: Bennett, Robert J. and Newton Gill (2015) Employers in the 1881 population census of England and Wales, *Local Population Studies*, 94, 29-49; also WP 3.

1891, which are assumed to be errors. The whole category is assumed to be workers.

- Occode 137 Motor garage proprietors was also significantly mis-coded in I-CeM; after re-coding there is a range of employment status across all categories.
- Occode 115 Accountants. All bookkeepers 1851-81 in this Occode were assigned to worker status. A very small proportion in 1891 were recorded as employer (0.3%) or as own account (0.8%). The methods described below consistently over-estimated the number of employers and own account individuals in this Occode, this recoding was done to help reduce the numbers so identified.
- Check coding of portfolios to ensure first coding is to main occupation (especially necessary for farmers). No attempt is made in the reconstruction to extrapolate portfolios; all are taken to the respondent's main Occode, but existence of a portfolio is used as part of the reconstruction in farming. Further research is developed with portfolios in subsequent papers.
- All economically active as defined by Occodes are included, no 'U' categories are removed where some data are blank in I-CeM (e.g. no gender, no age, etc.), except where there was a blank Occode for which economic status cannot be determined.

2.1.3 Preparation of 1851-71 censuses for reconstruction

All checks and data pre-screening used in 1881, as listed above, were also applied to the other early censuses (except the specifics deriving from the transformation between the GSU to I-CeM versions). This chiefly applies to 1851-61; the 1871 census records available are restricted in the project to those extracted from Groups 1-6 (as defined below).

2.1.4 Census extractions for 1851-81: Extracted Groups

The census instructions over 1851-81 allow many employers to be identified from the descriptions given of their occupations. These data were not coded in I-CeM but were mostly transcribed (and have been supplemented to infill where deficient transcriptions have been detected: see WP 1 and 3). The instructions to householders were to identify employers and

masters with their workforce, and additionally for farmers the acreage farmed. The 1881 instructions, which were almost identical to 1861 and 1871, were:⁵

‘In TRADES, MANUFACTURES or other Business, the Masters must, in all cases, be so designated. – *Example: ‘Carpenter – Master, employing 6 men and 2 boys;’* inserting always the number of persons in the trade in their employ at the time of the Census. In the case of Firms, the number of persons employed should be returned by *one partner only*’.

The farmer’s instructions were also similar over 1861-81:

‘FARMERS to state the number of acres occupied, and the number of men, women, and boys, employed on the farm at the time of the Census. – *Example: ‘Farmer of 317 Acres, employing 8 Labourers and 3 Boys.*’ Sons or daughters employed at home or on the farm should be returned – “*Farmer’s Son*”, “*Farmer’s Daughter*”. Men employed on the farm and sleeping in the Farmer’s house must be described in the schedule as *Farm Servants*.’

In 1851 the instruction was similar, but omitted the partnership question:⁶

‘In TRADES the Master is to be distinguished from the Journeyman and Apprentice, thus – “(*Carpenter – Master employing [6] men*);” inserting always the number of persons of the trade in his employ on March 31st.’

For farmers the 1851 instruction was almost identical to 1861-81 but was again slightly simpler:

‘The term FARMER to be applied only to the occupier of land, who is to be returned – “*Farmer of [317] Acres, employing [12] labourers;*” the number of acres, and of in or out-door labourers, on March 31st, being in all cases inserted. Sons or daughters employed at home or on the farm should be returned – “*Farmer’s Son*”, “*Farmer’s Daughter*”.’

Note that there is an important issue in the phrasing of these instructions of how far those not *explicitly* stated, as ‘trades’ or ‘trades and manufactures’ would reply; for example professionals, mining, commerce. This is particularly true where these other occupations had additional instructions. Although the general instructions regarding employers were supposed to have been read in conjunction with any specific instructions, it is evident that they were

⁵ ‘General Instruction’, Census of England and Wales, *Householder’s Schedule*, 1881; Parliamentary Papers 1883, No. 43, vol. LXXX, Appendix C.

⁶ ‘General Instruction’, Census of England and Wales, *Householder’s Schedule*, 1851.

sometimes ignored, by professionals in particular. The defective design of the question, from the point of view of covering all sectors of activity, is an issue that has to be addressed in the reconstruction, as discussed later.

The identification of people who responded to the instructions to give employees, acres, or report 'master' is complex. The information provided in the census is contained in the occupations descriptor. This is the string of alphanumeric characters the respondent gave at the time. It has not been coded in I-CeM except to extract the occupation; e.g. a respondent stating 'farmer with 10 men' would have been coded to the 'farmer' Occode in I-CeM. The rest of the information has to be recovered directly from the string. A method to identify and extract the employer information and their workforce numbers and acres was piloted for 1881 by Bennett and Newton (2015). In the original 1881 pilot there were three 'types' of entrepreneur identified: 1: employers, masters, and partners; 2: directors, landowners, mine/quarry owners, and shipowners; 3: general supplementation for own account, and farmers with acreage stated but with no employees reported. This has been adapted and given greater precision for the purposes here. Six Groups are defined from the occupational descriptor strings given by individual business proprietors: these were extracted as follows:

Group 1: *all employers and any others (such as masters, proprietors or owners) with stated employees*; farmers with stated employees; partners with stated employees.

Group 2: *all 'employers' with no employees*; 'masters' and anyone else who includes 'emp' in their occupation descriptor; and partners without stated employees.

Group 3: *master etc.* anyone including 'master' in their occupational descriptor but with no employees. Additionally, some 'masters' were not included, such as 'station masters', full details of these spurious masters are given in WP3.

Group 4: *'farmer'* not stating 'emp' or acres

Group 5: *farmer giving acres* but with no stated employees and had two or more acres of land (those with less than two acres with no employees were excluded; it was assumed that they work on other farms)

Group 6: *owners or proprietors of business assets*: mine/quarry owner, shipowner, barge owner and others with any business assets (other than land/housing)

Two further categories are extracted for further analysis at a later stage, but not included in reconstruction:

Group 7: ‘owners’ with no other information (not in 6), including landowners with no employees or only with acres, and house proprietors with no employee information. These are treated as a separate group and not included in reconstruction.

Group 8: directors. These are own account, but their companies are employers which are not otherwise identified in the census. These are supplemented in the database by data enrichment using external information which results in a separate database for companies and their directors with other information where found. Their business is not included in reconstruction, but they are included as part of those on own account in total counts.

No attempt is made to take account of partnership information in reconstruction, if it is given. Partners are coded for later analysis but otherwise are included in the different Groups 1 and 2, as above. For the purposes here, a partner with stated employees is an employer in Group 1; and with no stated employees is in Group 2. Partners thus count towards total numbers of business proprietors, but there will be minor double counting of number of firms (where two or more partners of the same firm are identifiable and were included in the census return). This applies the same assumptions for reconstruction as used in 1891-1911. Number of firms is not the purpose of reconstruction in this paper; the aim is to reconstruct the individual proprietors/entrepreneurs.

The six Groups listed above derive from the way the census data can be extracted. For the purposes of identifying employers and own account there are alternative assumptions that can be made of how these relate to the 6 Groups.

- Employers always include all Group 1. They do not include Group 2 which are ambiguous and include strings that mean ‘employed’. Transcription errors and illegible CEBs leave this Group as uncertain. However, Group 2 is usually clear for farmers so that they were accepted as own account; but no specific assumption was made for Group 2 non-farmers, who were allocated by the general reconstruction methods discussed later. Group 2 also includes partners who may or may not identify themselves as proprietors, with some senior partners perhaps viewing their junior partners as employees and are hence included in Group 1. However, the effect of this ambiguity is so small that it can be ignored in the reconstruction; but it is part of a more specific analysis of partnerships discussed in later publications.

- More ambiguous are farmers with no employees stated (Group 4 as well as Group 2); and farmers who stated no employees but gave acreages (Group 5). These are assessed below using their acreage to determine if it was large enough normally to require employees. It is assumed for reconstruction purposes that those in Groups 1 are employers in farming, Group 2 are own account, but Group 4 are farm workers (unless they have portfolios); those in Group 5 are assumed to be farm entrepreneurs, but have to be split between employers and own account. Hence for non-farmers all Groups 1, 3 and 5 are included as entrepreneurs, and for farmers all Groups 1, 2, and 5, some of Group 4, .
- Mine/quarry owners, shipowners, barge owners and others with any business assets may be regarded as proprietors even if they state no employees (especially again where family members may be employed); but their attribution as employers or own account needs separate analysis (see Section 3.6).
- Generally own account is more certain for those masters etc. who state no employees, and ‘farmers’ with no stated ‘emp’ but who had more than two acres.

From this it can be seen that there are several ways the extracted data may be aggregated, and these yield alternative estimates of the employer and own account categories. This varies between the censuses because the nature of the census instructions varied slightly, but also because of the ways the census data can be extracted. The choice of how the groups are treated is discussed below for each stage.

2.1.5 Aggregation of Occodes

The reconstruction is undertaken for 880 Sub-Occodes. Each of these is assessed in the reconstruction for each census year. However, in any one year some of these categories may be empty or contain very small numbers. To help in the assessment of the results of the reconstruction it is important to estimate not only the individual Sub-Occodes, but also aggregations. Aggregations allow groups of related activities to be assessed together to facilitate consistency of treatment if they are categories that are expected to have similar trends, or where the I-CeM coding may have been confused and individuals assigned to the wrong and/or related categories.

Two aggregations are used as the main guides to reconstruction. These focus on aggregating the Occodes into like groups of related activities, but also like groups of organisational structure (i.e. with similar proportions of employers, own account, and workers). These aggregations cover the entire economically active population and are defined at two levels: EA17 for 17 groups of Sub-Occodes; and EA51 for 51 groups of Sub-Occodes (Tables 1 and 2). These are based loosely on the approach to census coding developed by Booth-Armstrong, and are discussed and defined more fully in WP5.⁷

Aggregate EA17	Summary title
1	Farming, other agriculture and fishing [incl. farming labourers and family workers]
2	Mining & quarrying
3	Construction
4	Manufacturing
5	Makers and dealers
6	Retail & ironmongers
7	Transport
8	Professional & business services
9	Professional & personal services
10	Agricultural produce, drink & tobacco manuf.
11	Food retailing
12	Lodging & refreshment
13	Finance & commerce
14	Public admin, military, clergy
15	Domestic and service staff
16	Undefined general labourers
17	Persons of property with no stated occupation

Table 1. Aggregate industrial sector groups: entrepreneurship sector classification of economically active for EA 17 sector classification.

⁷ Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*.

Aggregate EA51	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 2. Detailed industrial sector groups: entrepreneurship sector classification of economically active for EA 51 sector classification.

It is recognised that all aggregations gain and lose information, and that for the period covered the census classifications make very imperfect distinctions between ‘makers’ and ‘dealers’. EA17 and EA51 keep makers and dealers together where they cannot be easily separated over the entire 1851-1911 period; this has significant benefits for the alignment of reconstructed numbers. But it is accepted that this is far from fully satisfactory; it is however, the best that can be done without significant assumptions about making or dealing that cannot always be validated at the individual level from the structure of the original historical records available in the census. However, once the reconstructions have been made, it is possible to aggregate the data to other classifications in any way subsequently desired: for example, using HISCO, PST, SIC or other occupational or industrial sector coding systems, as discussed in Working Paper 5.

2.1.6 Anomalies

One of the central aims of the reconstruction in this paper is to align definitions of Occodes and Sub-Occodes as far as possible over time, especially between the 1851-81 censuses and the later series for 1891-1911. This has generally been achieved by the methodology followed below. Hence in the last section of this paper it is possible to show how the reconstructions for 1851-1911 continue for the full period 1851-1911. The extension of the data in this way was also used as a guide to assess the reconstructions themselves: to check that any discontinuity or anomaly between the earlier and later periods was really supported by the census data itself, or was substantiated by any information on trends contained in the secondary literature or the Census report commentary.

In general, most anomalies were overcome by checking and re-coding erroneous Occodes. But inevitably a few anomalies remained. As to be expected, the reconstructions appear to perform better, the higher the level of aggregations of the data since these amalgamate problematic groups. The ambition of aligning for 880 Sub-Occodes is clearly demanding. However, it is required because only at this level of disaggregation can the differences and anomalies be properly understood and a full grasp of the quality of reconstruction be achieved. But for analysis purposes fully accurate reconstruction of most Occodes or Sub-Occodes is not sought, and this needs to be born in mind by users. Instead, the disaggregate reconstructions provide the building blocks for more aggregate analysis. However, it is important to understand the potential issues that arise at the disaggregated level of individual Sub Occodes in order to understand the robustness of any aggregate analysis.

It is particularly important to note that convincing reconstruction at Occode or Sub-Occode level is not fully possible for some categories because of the way in which they were originally defined or collected in the census, or because of the coding in I-CeM, which lead to inevitable anomalies between census years. The main occupations affected are those where census respondents or enumerators gave insufficient information to assign to detailed Occodes (e.g. ‘manufacturer’, ‘labourer’, ‘weaver’, ‘cotton operative’). This is a greater problem in earlier than later census because GRO recognised the difficulties and later made greater efforts to get more precise occupational descriptors. However, problems recur even in the 1911 census where the addition of a question about industry led to some more generic responses to the occupation question. Textiles and female occupations in clothing and personal services proved particularly problematic. The difficulties emerged clearly in the reconstructions below, but also threw new light on the estimates developed for the 1891-1911 period when the data was viewed over the whole period. At the outset of the reconstruction it is valuable to bear in mind the main cases where anomalies occurred that make full alignment of the reconstructed data difficult or impossible. The following are the chief cases.

- EA51: 18 Woollen Cloth manufacturing. In this case two Occodes, and the related Sub-Occodes have larger numbers in 1861 than in adjacent years; these are 571 Woollen Cloth mf Undefined, 572 Worsted & Stuff mf Undefined, and related Sub-Occodes 842 and 843. This is caused by 1861, compared to 1851 or 1881, having many more generic occupation titles in wool and worsted production such as ‘worsted factory worker’ or ‘woollen cloth worker’ which are in Occodes 571 and 572. These cannot easily be re-coded.

- EA51: 19 Cotton & silk manufacturing (including ribbon, weaving, dyeing, bleaching etc.). It appears that in 1851 many of individuals in this group were classified to other codes but cannot be fully identified. Also, in 1861, similar to the woollen cloth manufacturing issue above, there were a great many more generic responses from those involved in cotton manufacturing and consequently Occode 555 Cotton & Cotton Goods Manufacture Undefined was much larger than in other years.
- EA51: 20 Other textile manufacturing. 1851 has a much larger number in this category as a result of heavier use of 606 Weavers (undefined) and to a lesser extent 607 Other work sundry fabrics undefined.
- EA51: 24 Drink & tobacco mf (maltsters, brewers, distillers, tobacco & pipes). Here there are large drops between 1851 and 1861 in two Occodes: 708 Maltsters, and 709 Brewers. There is some evidence from Inland Revenue statistics for a real decline of maltster & malt dealers over this period of about 20% because of changes in excise taxes on malt and sugar used in brewing. The reconstructions of employers and own account for both brewers and maltsters indicate a much sharper decline. This is assumed to be correct, but users should be aware that it may exaggerate the real picture.
- EA51: 26 General mf (manufacturers, mechanic, artisan, machinist). Two anomalies arise: one for 1851-61, where it appears that more people gave generic answers in 1861 than in 1851 leading to a large jump between the two years; second, a very large jump occurs for 1911. This is a generic classification which has a variety of different individual descriptors which have not been classified elsewhere. The 1861 problem is of a kind with those relating to textile production described above. The 1911 increase is caused by the large growth in the number of generic responses given such as ‘manufacturer’ or ‘manager’ in 1911. This growth was because individuals also gave the specific industry in which they were involved when responding to the newly introduced industry question rather than when answering the occupation question, as in previous years. In the current version of I-CeM the occupation coding takes no account of the industry question data and consequently there is an increase in the size of this Occode. Recoding individuals on the basis of their answers to the industry question is a non-trivial task. Hence the anomalies for EA51: 26 have not been overcome in this version of data reconstruction.
- EA51: 31 Clothing and dress dealing (drapers, hosiers, haberdashers). Here there is a very large drop in 1911 which is hard to explain. It may be that household heads were

less assiduous in returning the occupations of female household members than enumerators had been in previous years. The drop is mainly found in Occodes dominated by female employment, namely Occode 657 Dressmakers, 658 Stay and corset makers, and 659 Shirtmakers and Seamstresses.

- EA44 Personal services (washing & bathing, hairdressing, sweeps). This seems to be affected in the same way as EA51: 31, with an unusual drop in numbers 1911 in the female-dominated Sub-Occode105 Laundry work: washer, iron, etc. (not domestic)

2.3 Logit regression estimator for combined employers and own account

The preferred method for reconstruction is estimation of employment status using a robust logit regression model. This is based on the data contained in the 1891-1901 censuses where employment status was given explicitly. Estimates from 1891 and 1901 are both undertaken, and then compared. Both years are used because they offer different insights and overcome different problems. Whilst 1891 is generally to be preferred as it is closer to the 1881 and earlier censuses, it has problems of over-estimates of employer status that have to be corrected and may thus contain some inaccuracies. Also for 1891 there are some occupational categories that have very large proportions of non-respondents who do not give employment status at all, and for which it is difficult to develop a robust logit regression model. In these cases, 1901 may be a preferred data source. In any case 1901 may also be preferable for some occupational groups because 1891 was in some way aberrant. A cautious approach is therefore developed which assesses the value of both 1891 and 1901 as alternatives.

The logit regression method allows a wide range of the most significant explanatory variables for employer status in the later censuses to be taken into account. The variables used were the 880 Sub-Occodes for the occupational categories, gender interacted with marital status, population density of the Registration Sub-District, number of domestic servants, and household relationship codes (in I-CeM these are coded as a variable called RELA). These variables were determined after a range of experiments with alternatives. After estimation with the 1891 and 1901 data the coefficients are applied to the 1881 data and then the probability of being an employer for 1881 using 1891 and 1901 coefficients is calculated. Thus, employment status attribution is calculated using the 1881 independent variables values but with the estimated coefficients for the later censuses. This method has two outputs: (1A) the summation of the individual mass density to an aggregate-level (unrounded); and (1B) the

summation of the rounded numbers. (1A) is the preferred extrapolation of numbers because it has no bias. However, (1B) is the only method that gives individual-level employment status and thus can be compared with the other methods (see below).

Before applying this approach to 1881, the models were tested fully against their own years. That for 1891 is discussed here. It is necessary to be sure the regression model is producing a valid and unbiased estimation of the data for the year in which it is initially estimated. Thus, testing the predictions of the 1891 model on the 1891 data itself is a good test of whether any patterns exist in the residuals that are left unexplained. Initially a simplified model for 1891 was estimated which included only gender interacted with marital status, RSD population density, and number of servants. This had been found a satisfactory way of estimating corrections in 1891 data for upward bias. This produced a pseudo R^2 of 0.411 with $N=7,115,788$. Unfortunately, there were a number of difficulties with the estimate, mainly as a result of a high level of clustering of the residuals in either urban or remote rural areas; also, there were unsatisfactory estimations of many occupational categories where extensive in-house relationships existed for employment of family and others in a business. Particularly poor estimates arose for categories such as:

- 173. Farmer, Grazier
- 712. Lodging and Boarding-House Keepers
- 657. Dressmakers
- 713. Innkeepers, Hotel Keepers and Publicans
- 105. Laundry Work:

These all often have large numbers of in-household residents as employees (of family or assistants).

A number of experiments were undertaken to test alternatives estimators. One to replace number of servants as a covariate by household size produced little improvement in estimation quality and no change in residual patterns. Similarly, inclusion of an urban classification dummy variable did not fully remove the spatial patterning of residuals. Separate estimation for rural, urban and other areas in entirely different models also did not improve the estimators. Separating the model into different categories by Occode and using different suites of dummy variables for each also left the estimates virtually the same with similar spatial patterning. All these models had similar performance characteristics and were

not capable of capturing significant explanatory factors that resulted in clustering of residuals.

Finally, it was decided to adapt the methodology and replace modelling through a parsimonious estimator with a model that sought to include all the main spatial and demographic variables known to influence entrepreneurship in these historic data. This resulted in a much larger and more complex model, but one with stronger fit and, most importantly, eliminates almost all spatial patterning of residuals. The final model estimated included: the occupational categories as dummies for 880 Sub-Occodes; the density of the RSD and density squared to take account of known nonlinear density effects; density and density squared also interacted with a four-way urban classification (urban; rural; and two transition categories: see WP6); the age of the entrepreneur and age squared (to take account of known nonlinear effects over the working life); gender interacted with marital status; ten internal household codes to take account of each individual's position within the household (a simplification of the I-CeM RELA codes); and number of servants. This model is then used as the primary estimator for subsequent reconstructions. The estimates of the model are given in Appendix 3; similar estimates were produced from the same model using 1901 data, with some important differences for a few Sub-Occodes, but generally the models for both years were closely in line. The pseudo R^2 for the 1891 model improved to 0.456 compared to the previous estimator, with sample size slightly reduced as a result of a slightly greater extent of unknown data (due to gaps in the original census responses or their transcription) to $N=7,109,988$.

The spatial residuals from this estimation are shown in Figure 1. In general, there is a low level of clustering of outliers of high residuals, with the main areas of poor estimation in the upland hill farming areas of Wales and the Pennines. The best fit is achieved, with very low residuals, in London, the West Midlands, West Riding, and parts of Manchester. The spatial clustering of the residuals is tested in Figure 2. The primary index of the goodness of fit of the model used is the ratio of the sum of the rounded predicted individual probabilities over the sum of the individual that observed. A ratio of one, in a given area, means a perfect fit, or in other words, that the model explains the observed reality with high accuracy. The ratio can range from zero to infinity. An index greater than one indicates that the model predicts a higher density of entrepreneurs than the actual data show. An index under one means that the model under predicts what the data shows. To test the randomness in spatial terms the

clustering of similar residuals is tested as HH, LL, HL and LH clusters. This indicates areas where over predicting a variable occurs that decreases the probability of the fitted model; or where under predicting a variable increases the probability of the fitted model. Not only the geography but the sign of the new variable added must be understood. Using a dummy spatial variable alone is insufficient because it is an ad-hoc solution that is not dealing with the need for theoretical improvements of the model. The final model estimated is the outcome of this search of the residuals, with the only remaining strong clustering that is a concern being for hill farming areas. These relate almost exclusively to the structure farming entrepreneurship in these locations. The specific structure of these locations leads to use of a separate model for farming for reconstructed purposes, as discussed in Sections 3.5 and 4.1.

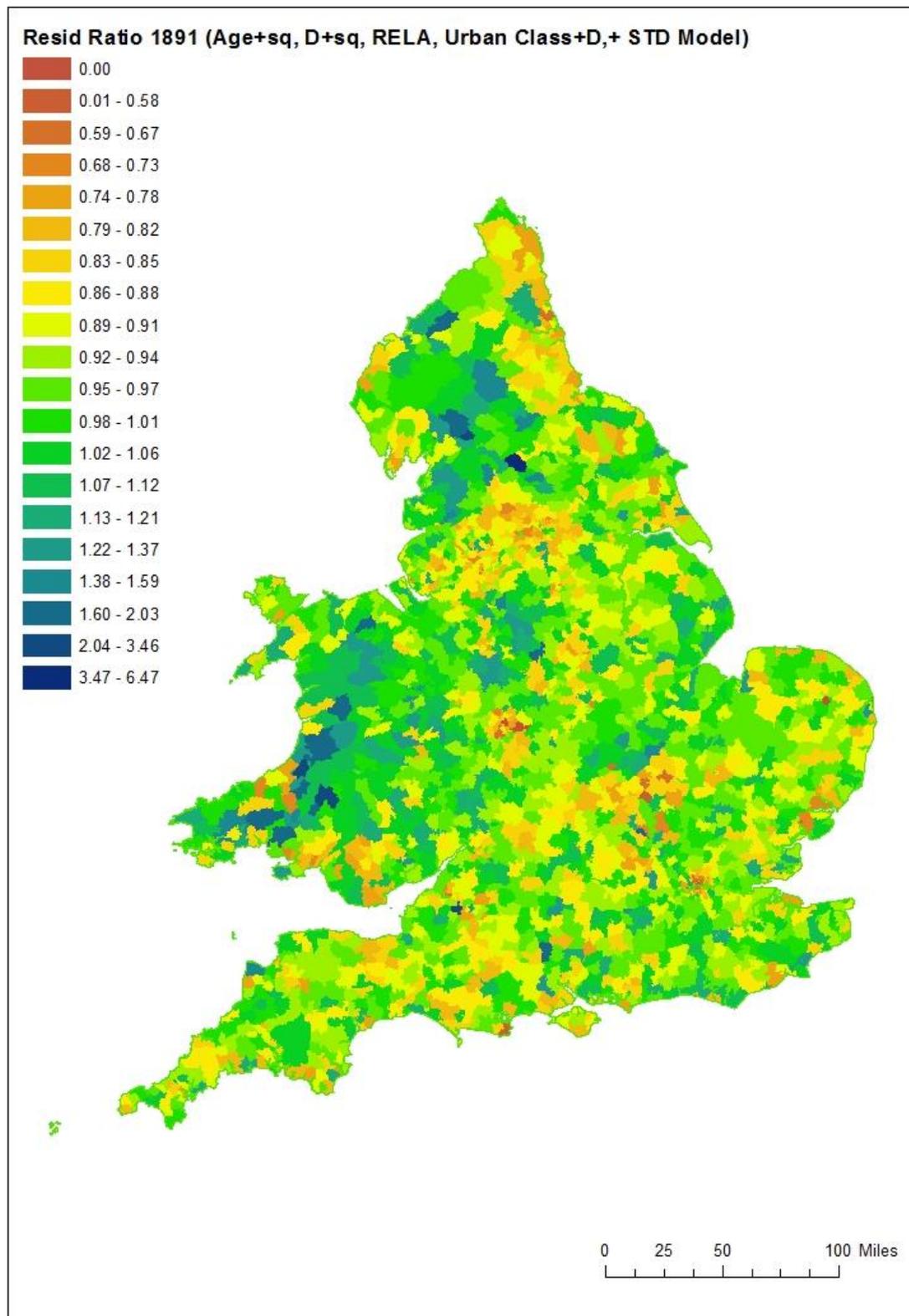


Figure 1. Mapping of residuals from the full model; the residuals are mapped for ratio of the observed/expected.

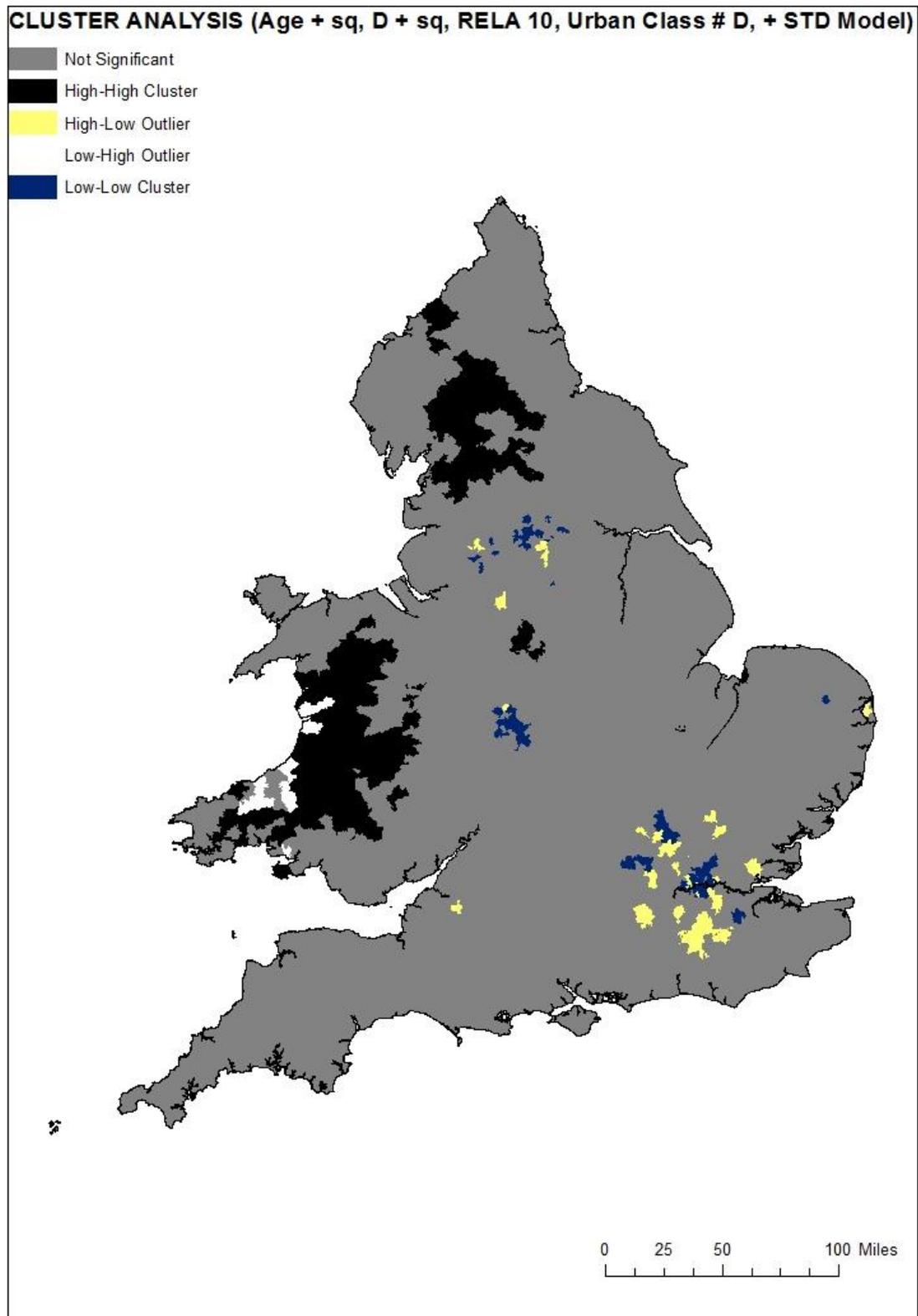


Figure 2. Mapping of spatial clustering of residuals from the full model.

2.4 Multinomial logit estimator of separated employers and own account

A check was made on the logit estimator by also estimating a multinomial logit for estimation of employment status as a single stage process for the three separate categories of: employers, own account, and workers. This was an alternative to the two-stage logic model of: first E+OA, then separating E from OA. The MNL model estimated was:

```
mlogit Type Age c.Age#c.Age Density_RSD i.Urban_Class i.Sex#i.Mar_short Servts
i.RELA_10 i.EA17, vce(robust).
```

This yielded very similar results to those for the logit regression:

```
Number of obs. = 10,806,930; Wald chi2(222) = 4257841.94; Prob > chi2 = 0.0000;
Log pseudolikelihood = -1379753; Pseudo R2 = 0.3585
```

Because they added little value to the simpler logit regression, had similar residuals, and the standard errors were uncertain, they were not used further in the analysis below. However, the estimation process confirmed that the logit regression was indeed robust compared to the alternative of the multinomial.

3. Estimates for reconstruction 1881

3.1 Alternative estimators for combined employers and own account

The method for reconstruction for the first stage for all entrepreneurs (E+OA) uses competing extrapolation methods which are assessed against each other. These are then checked against the information available from data extraction (Codes 1-6) (which are the partial ‘actual’ entrepreneurs), and then any other information that can be used for validation, such as commentary or data in census reports, other contemporary sources such as directories, and secondary sources. Five competing alternative methods are explored.

Method 1: 1891 logit regression. The logit regression estimator is the primary method by which it is sought to reconstruct the earlier census data for own account and other employment statuses. This model includes a wide range of the explanatory variables that fit

closely with the individual characteristics of entrepreneurship and hence should be strong differentiators of entrepreneurs from others. In Method 1 the 1881 prediction of entrepreneurs (E+OA) is based on the 1891 regression model of 1891 data corrected for upward bias and non-response bias (Bennett et al., 2018). As noted above, this model includes entrepreneurs age and age squared; RSD density and RSD density squared and interactions with urban classification; an interaction of gender and marital status; number of servants; ten household relationship codes (simplified RELA); and 880 Sub-Occode. The regression is weighted by the weights built from the regression for 1891 for corrected non-response bias on gender, ten household relationship codes (simplified RELA) and 17 economic activities (EA17) (Bennett et al., 2018). After saving the coefficients from this model, datasets are swapped with the 1891 coefficients applied to the 1881 data. The prediction can be the unrounded probability of being an entrepreneur or it can be the rounded 1s or 0s, with 1 indicating an individual was an entrepreneur and 0 that they were a worker. For individual predictions the rounded probabilities are normally the preferred choice. But for aggregate predictions the choice is not obvious and, for instance, in a category that has a high ratio of entrepreneurs to workers, the rounded sum of 1s will be higher than the unrounded sum of numbers between 0.5 and 1. In contrast, when the category has more workers, summing the 0s instead of numbers between 0 and 0.49 gives a lower aggregate summation. Detailed comparisons are needed to decide which is more accurate on an Occode-by-Occode basis depending on the ratio of E+OA to workers. Method 1 splits into two sub-methods: rounded and unrounded. Rounded is the base sub-method where each individual is or is not an entrepreneur according to the *rounded* probability. The unrounded sub-method simply adds the probability of being an entrepreneur without rounding. This sub-method adds up all the entrepreneurs predicted by the logit model. The drawback of the unrounded method is that it is not an individual level procedure, but just at aggregate level which needs a further step to allocate individuals to entrepreneur categories. The rounded logit method gives the floor of given individual numbers and above this we need to add or subtract the excess or deficit of chosen individuals by the given aggregate level procedure. This can only be done by random assignment. Say the logit rounded model gives 1,000 entrepreneurs, but the aggregate-level is chosen with unrounded logit model giving the number as 1,100. Then, we start from the 1,000 entrepreneurs from the rounded logit model and then from the pool of non-entrepreneurs we randomly select 100 individuals and allocate them to entrepreneurs. As a result, 1,000 entrepreneurs will come from the rounded model and will be predicted by this model and 100 will come from a random reallocation of the additional 100 workers recoded as entrepreneurs.

It is expected that the Method 1 estimator will not provide a satisfactory approach for all Sub-Occodes because of unique developments in each sector. Hence, as a check on the logit regression, and as a test of the preferability of alternatives, several other methods of reconstruction were also calculated:

Method 2: 1901 logit regression. This is the same as Method 1, but instead of using the 1891 logit model an estimate based on the 1901 data is used, with the same list of covariates and weights belonging for the 1901 case. There are known uncertainties about the 1891 data. As shown in (Bennett et al., 2018) there was systematic upward bias in the numbers of employers identified in the 1891 census which has to be corrected by data cleaning and a regression estimator to reallocate over-estimated employers to own account. There were also categories with large non-responses where it is difficult to develop a robust logit model. The results of the 1891 estimation are believed to be good, but clearly they will suffer from some inaccuracies. Hence, it is valuable to test the 1891 logit regression against an alternative derived from 1901 to see if there are important differences. Because the census question was redesigned, the 1901 census data are known to be more accurate and contain little or no upward biases of the form found in 1891. Method 2 has also two sub-methods: rounded and unrounded. As above, the rounded sub-method gives estimates for individuals, whilst the unrounded sub-method requires a similar method of assignment of some individuals. . In general Method 1 is preferred over Method 2 because it uses 1891 data for the closest available year, but Method 2 may be preferred in some cases since 1901 offers the closest real, uncorrected (but screened and cleaned) available year.

Method 3: Entrepreneurship ratios for 1891. A valuable test of reconstruction is whether the same proportions of E and OA to workers existed in 1881 reconstructed data as in 1891 for the 880 Sub-Occodes. Other things being equal it can be expected that the entrepreneurship ratios will stay relatively constant in many occupational categories between censuses. This is likely because many of the businesses are actually the same businesses run by the same people in the same way; others are the same business run by new proprietors but in the same or similar situation (the business of a baker in one high street is sold on or occupied by a new baker, with no other change in local competitive conditions). Many other developments derive from ‘organic growth’ as business numbers and change with the growth of the population as a whole so that entrepreneurship ratios remain largely unaffected. The estimation of the numbers of employers or own account in a preceding year based on the

entrepreneurship ratios from following year is thus a valuable base-line for comparison; and where the regression proved unstable from methods 1 and 2, may be the best available reconstruction. Method 3 uses the 1891 ratios as an extrapolation.

Method 4: Entrepreneurship ratios for 1901. This is identical to Method 3, but uses the 1901 entrepreneurship ratios of employers to own account for the 880 Sub-Occodes. It will be preferable to Method 3 where the 1891 data give less accurate measures of entrepreneurs for the 1891 year because of upward and non-responses biases.

Method 5: Choices from extraction Groups 1-6. Accepting the actual Census responses in 1881; this is to be preferred where the 1881 extractions of employers and own account are thought to be complete for all individuals, where the occupational category is clear and unambiguous as to employer, own account or worker status, and respondents are believed to have been accurate. These are the real, unprocessed (but screened and cleaned) data recovered from the 1881 CEBs. They allow use of the Groups 1-6 derived from the extraction method, outlined above. Checks of this allocation were made against the same occupations in 1891-1911 to determine which occupational Sub-Occodes may be reliable. However, because of the different format of the census questions in 1881 very few of the occupational categories have extractions that are likely to be close to the actual numbers of employers and own account. But, fortunately, the largest of all employers and own account categories, farmers, is extracted with high accuracy for 1881. Groups 1-6 from the data extractions can also be used as a valuable test of the other estimators. A good reconstruction should be able to predict accurately the actual extracted employers and own account, even if these are themselves not complete in including all employers and own account.

Note that for Methods 3 and 4, and for any unrounded parts of Methods 1 and 2, the results are only aggregate-level reconstructions (not for individuals). To transform them into individual extrapolations for which separate E and OA can be generated, a random allocation method is used. This is undertaken by first subtracting the total of previously identified entrepreneurs for that sub Occode from rounded Methods 1 and 2 (i.e. the individual level method) from the selected Method unrounded 1 and 2, 3, and 4 (the aggregate level methods). The remainder is then randomly allocated using the closest individual-level extrapolation. To complete this stage, all individual extrapolated workers are first dropped before the final randomly allocation correction is performed.

As a further check, an additional method was trialled of a simple linear extrapolation of the *change in entrepreneurship ratio* of employers and own account between 1891 and 1901 to estimate 1881; this might be appropriate where a sector was experiencing continuous growth or decline at the same rate over the period. In the event, this method did not appear to offer a superior approach for any Occode and it is not reported here.

Each of the 880 Sub-Occodes has a different reconstruction estimate from each of the five methods. Methods 3 and 4 were used primarily as baseline tests: what would the reconstructed number of entrepreneurs be in a specific Sub-Occode in 1881 if there were constant ratios with nothing changed in sub-industry structure, business organisation, technologies of production, and overall production function. Methods 3 and 4 provide this no-change baseline for the alternative 1891 and 1901 ratios of E+OA to workers, or E to OA, where one may be preferred over the other because of the different issues arising in the way these two censuses were collected. The regression estimates for 1891 and 1901 provide the preferred actual estimates to be used for reconstruction. These identify the actual individuals that are most likely to have been entrepreneurs based on their personal and locational characteristics, within each of the 880 specific Sub-Occodes.

Decisions between the methods used for reconstruction were then made based on four rules as follows:

1. If *one* of the 1891 or 1901 regression estimates from Method 1 or 2 was within ± 10 per cent of the equivalent ratio extrapolations then that regression estimate was used. The ratio extrapolations were used only as the baseline to guide the choice of regression estimate. In effect this assumes that all methods converge on the same interpretation within ± 10 per cent. For example, Sub-Occode 645 straw plait manufacture in the 1891 regression estimate was 4.4 per cent greater than the 1891 ratio extrapolation, but the 1901 regression extrapolation was 62 per cent larger than the 1901 ratio extrapolation; the 1891 regression was used. This was a sector experiencing rapid change after 1891 so the outcome is to be expected.
2. If *both* 1891 and 1901 regression estimates were within ± 10 per cent of the equivalent ratio extrapolations, then the regression that was closer to the ratio extrapolation was used, e.g. for Sub-Occode 697 Grocers & Tea Dealers the 1891 regression was 8 per cent larger than the 1891 ratio extrapolation and the 1901 regression in 9.8 per cent larger than the 1901 ratio, so the 1891 regression estimate was used.

3. If none of the regression of estimates were within ± 10 per cent of the ratio extrapolations then the ratios were used, with a preference for 1891; for example Occode 183 Nurserymen, Seedsmen, Florists had no regression estimates close to the ratio extrapolations.
4. If the already extracted 1881 entrepreneurs were larger than any of the 1891 and 1901 estimates from rules 1-3 above, then the extracted entrepreneurs were used. Those accepted as true entrepreneurs were Groups 1, 3, 5, and 6. A special case of these is for farmers (which also assumes Group 2 are own account; see further below). The extractions were also accepted as correct for any cases where there was no alternative or the census data was otherwise so partial that no further progress can be made: a key example of this is Occode 777 Owners of Companies, where the census responses were clearly deficient but data enrichment is subsequently being used to allow more individuals to be identified.

In this process the comparisons over time between 1881 and 1911 threw up a number of problematic occupational codes in I-CeM that had not been previously detected. These were dealt with either by correcting and re-coding where extensive I-CeM occupational mis-codes occurred, or were noted as anomalies requiring separate treatment in subsequent analysis and interpretation, as outlined above in Section 2.1.6.

Once the choice of Method was made, the actual entrepreneurs were identified directly by using the regression estimates, but if there were actually identified individuals from the census extractions (Groups 1, 3, 5, 6; and Group2 for farmers) these were assigned first as ‘actual’ entrepreneurs, and then any deficit in numbers was added using the regression estimates as a top-up of ‘reconstructed’ entrepreneurs. The database deposit of reconstruction identifies the different sources for reconstruction used.

3.2 Rounded vs. unrounded

As noted above (Section 3.1) the regression estimates give either an unrounded probability in decimal points, or rounded to 1s or 0s. For individual-predictions the rounded probabilities are normally to be preferred. But for aggregate-predictions and some sectors the choice is not as clear. In the choice between methods the four rules outlined above take precedence. But where there is a choice or ambiguity, in most cases where workers were the majority in the

Sub-Occode the rounded result is chosen; this results in all the small decimals (between 0 and 0.5) being lost (they become zeros), and overall numbers of entrepreneurs estimated are lower. The choice of rounded then allocates to the majority class (workers) correctly. When employers plus own account were a majority, the reverse is true, and it would normally be preferable to preserve the smaller decimals parts, so that the unrounded results are to be preferred; resulting in the overall numbers of entrepreneurs estimated being higher. However, where the ratio of workers to entrepreneurs is around 50% the choice is made purely following the four rules above.

3.3 Farmers (all employers and own account)

For farmers an additional method is used. Whilst the five methods above all provide potential candidates for estimates of farmer entrepreneur numbers, they all lead to significant upward bias, over-estimating probable numbers of employers and own account compared to the later censuses where a significant proportion of those calling themselves ‘farmer’ identified themselves as workers. This was 18-21% of ‘farmers’ in 1891 depending on how data are estimated, 15% in 1901 and 11% in 1911. These provide ratio estimates that might be used for reconstruction for the 1881 and earlier censuses. But we also have other information collected by the census on farmers for 1881 and earlier censuses that give not only their workforce (and hence identify them as employers in Group 1 or not), but also their acres farmed.

The extractions for the earlier censuses, as discussed above Section 2.1.4, allow four specific Groups of farmers to be identified related to their employer and acre information:

Group 1: ‘farmer’ with stated employees.

Group 2: ‘farmers’ stating ‘emp’ but with no employees stated

Group 4: ‘farmer’ not stating ‘emp’ or acres

Group 5: ‘farmer’ giving acres but with no stated employees

Group 1 are accepted as definite actual employers. They should all be employers, but there will be a census non-response rate, or database misidentification from I-CeM and S&N because of transcription gaps or truncations, so that Group 1 will not include all farm employers; others have thus to be estimated by a reconstruction method. Group 2 is assumed

to be own account. Their number is very small in 1881, only 41 in total, but is larger and more ambiguous for earlier censuses.

Group 5 contains the valuable additional information that they farm a specific acreage. Those that had acreage of less than 2 acres and stated no employees are already preassigned as workers in the screening process. This assignment follows decisions made when agricultural statistics were collected, which assumed active farmers had to have 2 acres and over, with those below being 'smallholdings'. Smallholders were the subject of significant parliamentary attention, but were never assumed to be significant farmers. In the more modern period the Ministry of Agriculture attempted to define full-time holdings in 1968 with a cut-off of 4 hectares of agricultural land (which is almost 10 acres). There have been many technological changes in the meantime, but our assumption looks robust that farms below 2 acres were not viable businesses, but instead smallholders who were essentially subsistence farmers whilst also operating as farm labourers elsewhere as workers or other occupations (such as estate work). This removes those who could not normally have been own account or employer farmers. The rest of Group 5 are assumed to be entrepreneurs (either farm employers or own account), and the acreages can be used to separate employers from own account (see below, Section 3.5).

Group 4 state no other information than that they are 'farmers'. This is an occupational descriptor. Without the other information, that they employed workers or had an acreage, they have to be assumed to be farmers working on others' farms. Thus, all are assigned to worker status, but with one exception. The exception is those within Group 4 who have a portfolio of other activities as well as farming. Where an individual has a descriptor such as 'farmer and butcher', these are assigned to own account status. In such cases, it is assumed that their other occupation as butcher was as a small business trader, where the small acreage is either a subsistence element, or is adjoint to their other occupation. In this example, a butcher who had a field in which to keep animals prior to slaughter had a portfolio, but the 'farm' aspect was an adjoint to being a butcher. They could be re-coded as butchers, but to maintain a comparability of farm numbers with GRO definitions (which allocates them as farmers), the farm and butchery are assumed as adjoints: they are joint 'farm-butchers' and hence own account rather than just a butcher. Similarly, a grocer who had a small amount of land where they kept a cow to supply milk or hens for eggs to sell in their grocery shop is kept as a farmer, as a joint 'farm-grocer' on own account. This is an appropriate approach for

reconstruction. But in subsequent analysis of portfolio businesses this specification must be borne in mind, and an alternative assumption can be made if required: that in these cases the farm element is ignored and all are recoded to their other occupation.⁸ The proportions of portfolios within the different Groups in farming are shown in Table 3. This indicates clearly the significance of portfolios for Group 4, which is over twice as high as the average of all farmers, and nearly three times higher than in the main group of employers. Portfolios are thus an important element within Group 4, and they provide a good means to identify own account individuals within a group that otherwise appears to be only farm workers. For all other farm Groups, portfolios are not taken account of and they are assigned to own account or employer status on the basis of their Group or the extrapolation method outlined below.

Farmers	% with portfolios
Group 1	2.9
Group 2	2.4
Group 4	8.2
Group 5	3.8
Total	4.0

Table 3. Percentage of portfolios within each Group of farmers, 1881 census

For estimation of the total of farm entrepreneurs in 1881, therefore, we are able to use the sum of numbers of definite employers (Groups 1 and 2), Group 4 where they have portfolios, and Group 5 (where acres farmed are 2 or more); the rest of Group 4 are farm workers (that is ‘farmers’ working on other farms). Using these assignments, we thus have better information to estimate farm entrepreneurs, which is to be preferred to estimates based on the other extrapolation methods; though we can compare the Group allocation with the other estimates as part of the validation process. More detailed discussion of the specific case of farming entrepreneurs and how they interrelate with the census extraction Groups is given in other publications

⁸ Note that the extraction process for 1851-81 identifies portfolios only if individuals are already identified as master or employer; hence the equivalent of a farmer-worker with a trading portfolio, say a butcher or grocer, would not have been picked up as a portfolio if they did not mention being a master. For the later censuses 1891-1911 a wider definition of portfolios can be used. These differences are discussed in a subsequent Working Paper.

3.4 Estimation of separate statuses: Employers (E) and Own Account (OA)

Reconstruction follows a two-stage process: one for entrepreneurs as a whole (E+OA) as above, and then other differentiating employers and own account. To generate separated estimates for E or OA a similar method is followed to that above. But in this case the employers only are the subject of the estimation models. Then, once employers are estimated, the own account are the remainder from the previous estimates of all entrepreneurs (E+OA). The same five different extrapolation methods are used, plus an additional method for farmers (based on Groups 1, 2, 4, 5). These estimates are then tested against each other, the information available from data extraction (Groups 1-6), census reports and secondary sources.

Method 1: 1891 logit regression. The 1881 employers is constructed based on the 1891 regression of employers using the same full model as for all entrepreneurs including age, RSD density, urban class, gender, marital status, servants, relationships and 880 Sub-Occodes and squared terms, interactions and weights. As for the E+OA estimation, the datasets are then swapped with the 1891 coefficients applied to the 1881 data to give the employer estimates. The own account are then the residual from the total number of entrepreneurs. Again, unrounded and rounded versions of the predicted probability are generated as two sub-methods: rounded and unrounded. The former is then an individual level and the latter an aggregate level.

Method 2: 1901 logit regression. The same employer predictions are estimated as Method 1, but based on a 1901 data. The own account are then estimated by subtracting from the total of E+OA. Rounded and unrounded versions are generated.

Method 3: Employer ratios for 1891. Employer over own account self-employed ratio (E/OA) for 1891 is applied to the total number of entrepreneurs in 1881. Using this method means considering the closest available year.

Method 4: Employer ratios for 1901. The same as the 1891 ratio estimation, but using the 1901 data. Using this method means considering the closest real, uncorrected (but screened and cleaned) available year. In the event this method was not actually adopted, though it was used as a comparison with the alternative methods.

Method 5: Choices from extraction groups 1-6. Actual 1881 data. This is the real, unprocessed (but screened and cleaned) data recovered from the CEBs with groups 1-6. These are used as comparators in choice of methods. A special case of these is again for farmers, discussed below.

Again, only Methods 1 and 2 with the rounded estimates are individual-level methods. For the aggregate-level methods (Methods 1 and 2 unrounded, and Methods 3 and 4), after merging the chosen aggregate-level extrapolation, the excess and deficit must be randomly allocated by occupation Sub-Occode at the individual-level one.

As before, Methods 1 and 2 were the preferred for reconstruction (unless the actual is used under Method 5). Methods 3 and 4 were used primarily as baseline tests; though Method 4 (1901 ratios) was not actually adopted in any case. Methods 3 and 4 provide a no-change baseline for the alternative 1891 and 1901 ratios of E to OA. The rounded regression estimates (Methods 1 and 2) provide the preferred reconstruction. These identify the actual individuals that were, or were likely to have been entrepreneurs based on their personal and locational characteristics, within each of the 880 specific Sub-Occodes. As with the estimation of the total number of entrepreneurs a series of rules were used to determine which estimate to choose. However, in this case the margin used was +/-20% difference between the regression and the ratio estimates. The wider margin was used because of the greater degree of uncertainty over the accuracy of the employer/own account distinction in the later censuses.

3.5 Farm employers

Farm entrepreneurs were identified as those extracted from the CEBs coded as Groups 1, 2 and 5; Group 4 was assumed as workers who farm on the farms of others, unless they have a portfolio of other activities. Within Group 4, those identified as an entrepreneur through having a portfolio were assigned to own account status (none of them declared employees; had they done so they would be in Group 1). Within Groups 1 and 5 employers can be separately identified as the whole of Group 1, plus an element within Group 5 where their acreage is so large that they cannot have farmed this without employee or partner assistance.

The rest of Group 5 will then be own account. This requires a means to estimate the employer element within Group 5.

Group 5 is separated between E and OA by developing a model of the relationship between the probability of being an employer and the acres declared by Group 5 farmers. The model uses a logit to assign between employer and own account status, estimating the model for each of 630 Registration Districts (RDs) separately since there will be major differences between the employer/own account ratios for a given acreage within different parts of the country. RDs are chosen as the unit of analysis to create a large enough unit where an average ratio is appropriate and where there is sufficient data for the model to be estimated. The model allocates between individual farmers that reported employees and acres (Group 1), and farmers that reported only acres (Group 5) using the acreage which both groups reported. The latter who give no employees are assumed to contain a proportion of non-responses on employment that should have been reported. The aim is take account of non-response bias from some employers, who should give employees, by reassigning some Group 5 to the employer category because their acres are so large that they could not have been worked without employees. This model seeks to overcome downward bias in the number of employer farmers if all of Group 5 were otherwise assumed as own account. The model produces for each RD a sigmoid curve. The average estimate from the model for all RDs is 112 acres - the point at and above which a farmer can be expected to employ others than him or herself.

There are some non-convergences in the model caused by RDs with very small samples, or where there were major outliers where farms had very large acreages. Those exceeding 1,500 acres were excluded from the estimation; also excluded were locations where a large acreage estimate is clearly anomalous (as in several London-fringe areas). Checks on the CEBs for the cases of acreages over 1,500 in these areas suggest these are all absentee landowners who hold farms in other parts of the country; hence the model would give invalid results for the RDs with these people included, as their farm(s) really belong to a different location. The model is thus focused on estimating the average farm acreage character of an RD rather than all farmers reported in the census in that RD. For other RDs there were no relevant farmers (Group 5) or the acreages reported were so small or large that the allocation to employer or own account are completely determined to one or the other status for all farms of Group 5 in that RD; i.e. these RDs have only potential own account (where the sigmoid curve is formed

at positive infinity) or have only potential employers (where the sigmoid curve is formed at negative infinity).

This approach was validated in two ways. First, comparisons were made between the total of number of farm workers reported by farmers in the census and the total of all agricultural workers reported in the census. There were 16% fewer workers reported by employers than were returned in the census. This means that employers that reported agricultural labour as part of their own census declaration are downwardly biased by a proportion that is almost identical to the choice of cut-off made in logistic model. A second comparison was to consider the number of co-resident individuals involved in agriculture (whether they were farmers, agricultural labourers or fulfilling other functions) in a household where the farmer was a head, but who were not reported as employers. Co-resident farm workers indicate a strong probability of the head being their employer. However, it is also possible that all farmers in that household were workers on other farms. The proportion of co-resident workers with a farmer head of household in Group 5 is compared to those identified by the cut-offs in the logistic in Table 4. This confirms that farm heads with co-resident workers were more likely to be employers as indicated by the logistic (in 62% of cases); however, 45% of farmers with co-resident workers were more probably working on other farms as indicated by the logistic, so that the presence of co-resident workers is a likely but not definitive indicator compared to the logistic cut-off.

	Head without co-resident workers	Head with co-resident workers	Total
Under logistic threshold (OA)	26,728	21,959	48,687
%	54.9	45.1	100.0
Over logistic threshold (E)	18,765	30,750	49,515
%	37.9	62.1	100.0
Total N	45,493	52,709	98,202
%	46.3	53.7	100.0

Table 4. Estimates of the probable number of employers in Group 5 indicated by the cut-off compared with the number of farm heads with co-resident workers.

Once the cut-off thresholds were defined for each RD, they were applied to all the Group 5 farmers in that RD: any individuals over the specified threshold in that location were then assigned as employers. This compensates for non-response bias by estimating the additional group 5 farmers that are estimated to be employers, which are added to the extracted 'employers' for Group 1 to give the final number of all farm employers. The model is good at identifying average employer proportions, provided that there is a sufficient number of farmers in each RD to allow the estimates to converge.

It should be noted that the farm model using this methodology is a superior way of controlling for the specifics of farm employer status compared to the model based on general logistic regressions. Referring back to Figures 1 and 2 it is clear that the general reconstruction model performs well in almost all cases, except in hill farming areas where farms have large acreages and have a different pattern of entrepreneurship from almost all other locations. This arises because of the specifics of the farm community in these locations, where a very large area is required for farming but few or no labourers or other farm workers are needed; hence most are own account. These locations have very high entrepreneurship ratios because almost the only adults resident in the area are entrepreneurs, but employee levels are very low. The farm model using the logit for acreage cut-off captures this effect and allows it to be used in reconstruction to correctly identify the levels of employers and own account farmers.

3.6 Mining

The other sector which must be dealt with in a different manner is mining. Mining presents several difficulties. First, mines were often owned by individuals, particularly landowners, who did not return themselves as mine owners in the census. Secondly, when individuals did return themselves as entrepreneurs involved in mining, they did not always report what type of mining they were involved in, whether coal, tin, copper etc. Thirdly, while the published *Mineral Statistics* are available to check the numbers produced by the census, they are organised by mine rather than by mining company, which adds further complication.

These potential issues were confirmed when the reconstruction process was carried out. The raw trend in the number of employers was suspect, with numbers level between 1861 and 1881, dropping in 1891 and then rising substantially from 1891 to 1901. Whilst some of these

trends are real, some are spurious. To address this problem, four Sub-Occodes were identified in which any entrepreneurs identified by the reconstruction methods were assumed to be all employers; the Sub Occodes in question were 202 Mine – Owners, Agent, Manager minus Sub Occode 809 those extracted individuals from 202; Sub-Occode 809 itself; 220 Stone Quarry Owners; and 221 Slate Quarry Owners. This method is similar to the assumptions made for farming: that a priori indicators can be used to separate employers from own account. The coding is supported by the vast majority of the occupational descriptor strings in these categories. The re-coding reduces the apparent error and brings the estimates in line with the numbers reported in *Mineral Statistics*; however, there remain concerns that a drop in the number of employers between 1881 and 1891, which is mainly caused by a fall in the number of employers in 220 Stone Quarry Owners, may not be accurate. Unfortunately, *Mineral Statistics* is less accurate as a source for quarries than mines, mainly because of the dominance by small proprietors and quarrying being an adjunct to other aspects of land-holding (such as farming). However, a large scale directory of the building industry covering the period 1870-1911 includes ‘quarry owners’. This confirms that a slow-down in the growth of numbers of mining and quarrying businesses occurred in the later 1880s and early 1890s.⁹ Hence it is believed that the trends reconstructed in these Sub-Occodes and the sectors EA17 2, and EA51 3, as a whole are accurate estimates.

3.7 Aggregate estimates by extracted Group

3.7.1 All entrepreneurs: employers plus own account

An important test of the reconstruction methods was made by comparing the logit estimator for 1891 and 1901 (Methods 1 and 2) with the extracted data from the census Groups 1-6, although Group 6 is not used as a definitive ‘actual’ but rather an indicator of likely entrepreneur. If the logit regression provides an effective method of reconstruction it should be able to correctly predict those identified from the census by direct extractions. Table 5 shows this comparison. In the first section of the table (General), for all sectors of entrepreneur, Group 1 is correctly predicted for 74% of cases of all entrepreneurs; for Group 5 they are predicted correctly for 96% of cases, and for Group 3 (masters) 54% are correctly

⁹ *The Post Office* (afterwards *Kelly's*) *Directory of the Building Trades ...*, Kelly & Co., London.

predicted. These are extremely good results. However, there are poorer predictions for Groups 4 and 2, which is to be expected: Group 4 because it is believed to be mainly a worker code, and Group 2 where some are believed to be mis-transcriptions, where ‘employed by’ or a similar term has been transcribed as ‘employs’. Hence the regression estimates are very good predictors of the cases we expect. In general, similar, though not quite so accurate results were given by the regression model using 1901 data.

When these predictions are separated between the two categories of Non-farm and Farm we see that the accuracy of the predictions varies considerably between the two. For farmers the accuracy rises to 100% because we are using the known and extracted farmers, only reassigning those that are believed to be workers. But for non-farmers the predictions are much lower, with Groups 3 and 6 best predicted, followed by Groups 1 and 5. It is clear that for non-farmers the extraction Groups are a very partial guide. This is to be expected given the defective question design of the census Instructions, where only ‘trades’ or ‘trades and manufactures’ were explicitly stated to reply, and those in professions, mining or commerce may have ignored the question. This reiterates that reconstruction estimates, especially for non-farmers, must be treated as aggregate level guides and should not be treated as valid at the individual level. They can be disaggregated to large groups (such as sectors like the EA17 and EA51 groups) or large locations, but cannot be used as accurate individual level assignments or even for single Occodes unless they are very large and well coded.

1. General

Group	0	1	Total
1	47,131	134,406	181,537
	25.96%	74.04%	100.00%
2	813	432	1,245
	65.30%	34.70%	100.00%
3	21,679	25,753	47,432
	45.71%	54.29%	100.00%
4	36,877	489	37,366
	98.69%	1.31%	100.00%
5	3,998	102,189	106,187
	3.77%	96.23%	100.00%
6	2,040	6,846	8,886
	22.96%	77.04%	100.00%
Total	112,538	270,115	382,653
	29.41%	70.59%	100.00%

2. Non-farmers

Group	0	1	Total
1	47,083	45,142	92,225
	51.05%	48.95%	100.00%
2	813	391	1,204
	67.52%	32.48%	100.00%
3	21,679	25,753	47,432
	45.71%	54.29%	100.00%
4	1,018	489	1,507
	67.55%	32.45%	100.00%
5	3,988	3,986	7,984
	50.08%	49.92%	100.00%
6	2,040	6,846	8,886
	22.96%	77.04%	100.00%
Total	76,631	82,607	159,238
	48.12%	51.88%	100.00%

3. Farmers

Group	0	1	Total
1	48	89,264	89,312
	0.05%	99.95%	100.00%
2	0	41	41
	0.00%	100.00%	100.00%
4	35,859	0	35,859
	100.00%	0.00%	100.00%
5	0	98,203	98,203
	0.00%	100.00%	100.00%
Total	35,907	187,508	223,415

Table 5. Regression estimates using the 1891 and 1901 model for the 1881 groups 1-6 extracted directly from the census.

In the actual reconstruction, the extraction Groups are imposed on categories where we are certain of their accuracy: Groups 1, 2, 3, and 5 that are assigned as definite entrepreneurs. We impose these Groups because they reflect the descriptions people gave of themselves in the census. After imposing these conditions, the reconstruction gives the results which are used in the rest of the analysis below.

3.7.2 Employers only

The same comparisons against the extracted Groups are shown in Appendix 2 employers only. In general the methods give better predictions for non-farmers: for employers than for all entrepreneurs, and for own account.

3.8 Final choice of reconstruction method

3.8.1 All entrepreneurs

The validation of the regression models for 1891 and 1901 as predictors of entrepreneur status in 1881 (Table 5 and Appendix 2) suggests that in aggregate the model-based approach offers good approximation for reconstruction purposes, though much better for farmers than non-farmers. However, at the level of each of the 880 Sub-Occodes there is great variety, and sensitivity to small sample sizes in many categories, which makes estimation more difficult. Hence, in the final reconstruction full reliance cannot be placed on the regression estimates so that other Methods 3-5 are also used a for the final reconstructions for 1881.

It is important to note that the use of Sub-Occodes for the 83 split categories where there is a residual Sub-Occode generally results in very close estimates against extraction Groups 1-6, against the extrapolation ratios used as a basic validation test, and against external comparator data. Almost all the main part of the split Sub-Occodes had close fits to these comparison sources (most within ± 5 per cent, almost all within ± 10 per cent), whereas the residuals of these Sub-Occodes were often the poorest and most difficult to make decisions on for reconstruction purposes. This is because individuals with occupation descriptors which were common among employers and own account individuals in the later censuses have already been separated between the Sub-Occodes, the residual having a much larger heterogeneous mix of unknown employers, unknown own account and a large number of workers. The residual Sub-Occodes were treated as any other sub Occode, and may contain significant numbers of employer and own account individuals. However, by separating out those who had occupational descriptors which were likely to have been used by employers or own account individuals we aimed to improve extraction in a number of sectors which were often characterised by low entrepreneurship ratios, notably in manufacturing, consequently

improving the accuracy of the regression estimations for at least a portion of these difficult categories.

In the final choice of reconstruction, as well as validation against Groups 1-6, comparison is made against other primary and secondary sources and census commentaries. A full discussion of comparisons with other sources is contained in a subsequent Working Paper. The comparisons result in a choice of methods that is heavily focused towards Method 1 (the preferred method using the 1891 regression as the closest possible date), but also with use of 1901 regression, and a small use of ratios from 1891 and 1901. Table 6 summarises the number and proportion of the different methods used, and the number of entrepreneurs reconstructed by that method. Generally use of a large number of unrounded methods was required.

Method	N. of Sub-Occodes using this method	% of Sub-Occodes	N. of entrepreneurs assigned	N. of randomly allocated
1 1891 rounded regression	23	3	189,102	0
2 1891 unrounded regression	340	48	601,798	193,318
3 1901 rounded regression	27	4	103,727	0
4 1901 unrounded regression	184	26	256,362	110,120
5 1891 ratio	91	13	122,336	58,179
6 1901 ratio	13	2	4,723	4,545
7 Actual 1891	35	5	5,951	0

Table 6. Methods of reconstruction used for identifying all entrepreneurs (employers plus own account) in the reconstruction for 1881.

3.8.2. Employers only

The equivalent choice of methods for employers separated from own account is shown in Table 7. As for all entrepreneurs, the choice is heavily focused towards Method 1 but also uses the 1901 regression, but with less use of ratios. Again a large number of unrounded

methods were required. Comparison of the reconstruction and the extraction Groups 1-6 are given in Appendix 2.

Method	N. of Sub-Occodes using this method	% of Sub-Occodes	N. of entrepreneurs assigned	N. of randomly allocated
1 1891 rounded regression	59	8	18,465	0
2 1891 unrounded regression	301	42	331,572	132,433
3 1901 rounded regression	75	11	18,942	0
4 1901 unrounded regression	169	24	87,214	36,586
5 1891 ratio	108	15	43,380	21,861

Table 7. Methods of reconstruction used for identifying employers in the reconstruction for 1881.

3.8.3 Own account only

Once the employers are estimated by the reconstruction, the own account are derived by subtracting the employers from the total of all entrepreneurs. This can be done at the aggregate level, and later at the individual level.

4. Reconstruction for 1851-71

The same methods were applied to the full data available for 1851 and 1861. The only exception to the same methods (as applied to 1881) is for farmers in 1851 where a small adjustment has to be made. The methods again use a combination of the logit regression and extrapolation ratios for all occupations, except farming where a separate farm model is used based on the extraction Groups. The reconstructions result in similar fits compared with the extraction Groups 1-6, again better for farmers in all cases, given the quality of the extraction process, and better for employers than for non-farmers and all entrepreneurs. The comparisons are given in Appendix 2.

4.1 Farm model 1851

In 1851 the data available from I-CeM is seriously deficient. As noted in the Introduction, I-CeM has serious truncation of transcriptions of occupational strings in 1851, which resulted in truncation of employee descriptors for about 26% of employer strings. These were subsequently infilled from a separate source (S&N: The Genealogist). The affected records of 55,000 employers are highly spatially concentrated: affecting the whole of 14 counties, and one or more Registration Districts within 18 other counties (see WP 3). The infills from S&N were generally very successful in identifying the missing employer transcriptions, but for farmers it was found that transcription strings stating the number of acres were often truncated for this element. While this was not an issue for those farmers mentioning employees as well as acres, for some farmers who did not mention employees their acreage might be lost. This resulted in the loss of acreage data for what should be Group 5 farmers in the counties where there had been substantial S&N infill: these individuals had been transcribed as just ‘farmer’, meaning that they became coded as Group 4 farmers rather than their correct Group 5. The extent to which counties have missing information on farm acres can be estimated by making comparing with the GRO published records, since for the 1851 census only the GRO published for all counties the totals for farms with employees and/or acres. All 32 counties with S&N infill were somewhat affected by over-counts in Group 4 farmers and undercounts in Group 5 farmers, although 7 counties stood out as being severely affected: Lancashire, Cheshire, London, Middlesex, Norfolk, Sussex, and Warwickshire, all of which had had S&N substantial infills.

Reconstruction of farm entrepreneurship for 1851 thus has to follow a modified route. For all counties with complete data derived entirely from I-CeM, or I-CeM supplemented by S&N, the same farm model as used in 1881 with logit estimator and cut-off for each RSD can be estimated in the same way, as it is based on the farmers who provided both acres and employees. However, for the 7 counties with severe Group 5 under-counts there were insufficient Group 5 farmers available to feed into the model, and as a result a hybrid model must be used. First, the logit model was used to re-assign the Group 5 farmers who were available. Next, the remaining deficiency in Group 1 farmers for these counties was calculated from the GRO tables, and a corresponding number of Group 4 farmers was re-assigned to Group 1. These Group 4 farmers were allocated randomly by RD, in the same ratio that a given RD contributed farmers to the county’s total. In addition, around 25,000

farmers were moved from Group 4 to Group 5 in 157 RDs in the 32 counties affected by S&N infill, according to the ratio of E/OA/W farmers in the published tables for that county. Actual reallocations by individual could only be assigned by random selection.

The details of the method of calculation and reassignment are summarised in Table 8 for the counties affected.

Division	County	S&N infill	Missing employers	Move 4 to 5		Move 4 to 1	Final employers
					Of which 5 to 1		
I	LONDON	x	-124			124	230
II	SURREY		161				1619
II	KENT	x	314	233			3949
II	SUSSEX	x	-489	645	489		3020
II	HAMPSHIRE	x	226	82			2580
II	BERKSHIRE		240				1742
III	MIDDLESEX	x	-78	97	78		664
III	HERTFORDSHIRE		96				1560
III	BUCKINGHAMSHIRE		184				1589
III	OXFORDSHIRE	x	182	62			2017
III	NORTHAMPTONSHIRE		256				2573
III	HUNTINGDONSHIRE	x	31	49			773
III	BEDFORDSHIRE		81				1286
III	CAMBRIDGESHIRE	x	162	205			2205
IV	ESSEX		283				3394
IV	SUFFOLK		579				4648
IV	NORFOLK	x	-856	1582	856		4889
V	WILTSHIRE	x	365	21			2672
V	DORSET	x	138	539			1717
V	DEVON		3022				9903
V	CORNWALL		1836				5638
V	SOMERSET	x	775	398			5347
VI	GLOUCESTERSHIRE	x	226	352			2717
VI	HEREFORDSHIRE		759				2353
VI	SHROPSHIRE		977				4193
VI	STAFFORDSHIRE		1146				4213
VI	WORCESTERSHIRE		717				2561
VI	WARWICKSHIRE	x	-138	583	138		2333
VII	LEICESTERSHIRE	x	99	120			2608
VII	RUTLAND		73				480
VII	LINCOLNSHIRE	x	624	809			6530
VII	NOTTINGHAMSHIRE		601				2919
VII	DERBYSHIRE	x	395	1153			2154

VIII	CHESHIRE	x	-714	3071	714		3203
VIII	LANCASHIRE	x	-2778	6478	1811	966	5986
IX	YORKSHIRE WEST RIDING	x	1503	492			7810
IX	YORKSHIRE EAST RIDING	x	-12	431			2366
IX	YORKSHIRE NORTH RIDING	x	1112	727			4372
X	DURHAM	x	712	52			2297
X	NORTHUMBERLAND	x	244	454			2017
X	CUMBERLAND	x	1340	462			3823
X	WESTMORLAND		511				1431
XI	MONMOUTHSHIRE		662				1911
XI	South Wales	x	833	4132			8029
XI	North Wales	x	4159	1749			11662
	Total E&W		20435	24978	5053	124	151983

Table 8. Reallocation of farmers by county, 1851.

4.2 Reconstructions 1851-61

The decisions on each method for 1851 are shown in Table 9. Comparison of the reconstruction and the extraction Groups 1-6 are given in Appendix 2.

Method	N. of Sub-Occodes using this method	% of Sub-Occodes	N. of entrepreneurs assigned	N. of randomly allocated
1 1891 rounded regression	40	6	36,747	0
2 1891 unrounded regression	315	44	796,723	181,650
3 1901 rounded regression	46	6	62,718	0
4 1901 unrounded regression	168	23	210,925	68,383
5 1891 ratio	149	21	44,817	23,036

Table 9. Methods of reconstruction used for identifying all entrepreneurs (employers and own account) in the reconstruction for 1851.

For employers only the choice of methods is summarised in Table 10, with comparison of the reconstruction and the extraction Groups 1-6 are given in Appendix 2.

Method	N. of Sub-Occodes using this method	% of Sub-Occodes	N. of entrepreneurs assigned	N. of randomly allocated
1 1891 rounded regression	55	8	25,114	0
2 1891 unrounded regression	320	45	339,366	97,597
3 1901 rounded regression	72	10	37,494	0
4 1901 unrounded regression	164	23	31,078	13,749
5 1891 ratio	93	13	40,343	18,255

Table 10. Methods of reconstruction used for identifying employers in the reconstruction for 1851.

The decisions on each method for 1861 are shown in Tables 11 and 12 for E+OA and E only. Comparison of the reconstruction and the extraction Groups 1-6 are given in Appendix 2.

Method	N. of Sub-Occodes using this method	% of Sub-Occodes	N. of entrepreneurs assigned	N. of randomly allocated
1 1891 rounded regression	50	7	148,054	0
2 1891 unrounded regression	243	34	343,872	106,981
3 1901 rounded regression	69	10	102,699	0
4 1901 unrounded regression	131	18	479,606	55,279
5 1891 ratio	218	31	119,705	70,756

Table 11. Methods of reconstruction used for identifying all entrepreneurs (employers and own account) in the reconstruction for 1861.

Method	N. of Sub-Occodes using this method	% of Sub-Occodes	N. of entrepreneurs assigned	N. of randomly allocated
1 1891 rounded regression	69	10	18,189	0
2 1891 unrounded regression	294	42	346,087	106,764
3 1901 rounded regression	75	11	37,327	0
4 1901 unrounded regression	152	22	40,558	15,243
5 1891 ratio	117	17	48,160	19,641

Table 12. Methods of reconstruction used for identifying employers in the reconstruction for 1861.

5. Database for reconstructed estimates of employment status and overview of trends

The reconstructions developed in this paper are developed at an aggregate level for everyone in an Occode and Sub-Occode. But they can also be used to assign employment status to individuals within Occodes and Sub-Occodes. Whilst it is not intended that these individuals be analysed as individuals (but only in aggregate categories), the estimated employment statuses of all individuals are provided in the database deposit. It must be borne in mind that the ambition of aligning at Sub-Occode level is clearly demanding, results in revealing anomalies in census and I-CeM coding, and analysis using the reconstructed data is only possible at aggregated levels.

The aggregate reconstruction choices made using the Methods 1-5 outlined in Sections 3 and 4 are the starting point to identify each individual that can be assigned to their most likely status. For those reconstructions that use Method 5, for individuals extracted on the basis of their strings (i.e. Groups 1-6), these employment statuses are certain. This also applies to the categories of farmer and mine proprietors, as outlined in Sections 3.3, 3.5 and 3.6. For those identified by the regression Methods 1 and 2, each individual identified as an entrepreneur is estimated by either the rounded or unrounded probabilities derived from the regression. For those identified by ratios using Methods 3-4, and for Methods 1 and 2 unrounded, only

aggregates are estimated by the method chosen and the individuals have to be assigned by random selection.

The data base deposit identifies each economically active person for each census year and ascribed them an estimated employment status code (worker, employer, or own account). This gives the census 1851-81 and equivalent coding to those for 1891-1911. In addition the database deposit shows how each individual was estimated using the above methods. From this it is possible for the researcher to select only those with known status from extractions, or those where the logit individual level has been used. This would avoid using those assigned only through random selection.

From the reconstructions it is possible to give a brief overview of the main trends that are evident. These trends are the subject of more detailed assessment in subsequent publications. The main trends over 1851-1911 are shown in Figure 5, for the total of all entrepreneurs, and own account and employers separately. Data for 1871 are not available, except for the case of farming. The patterns of entrepreneurs as a whole and employers increased in a broadly similar manner throughout this period. However, the number of entrepreneurs increased at a faster rate than employers between 1861 and 1881, and between 1891 and 1901, suggesting that these were periods in which there was increasing self-employment and slower development of concentration, two processes which may well have been connected in some sectors, notably farming. After 1901 the trends show a major reversal: own account fall absolutely and employer numbers begin to increase rapidly. Overall numbers of all entrepreneurs show a marked slowdown in rate of growth. It shows a profound re-balancing between the smallest sole proprietors with no employees in favour of those employing others.

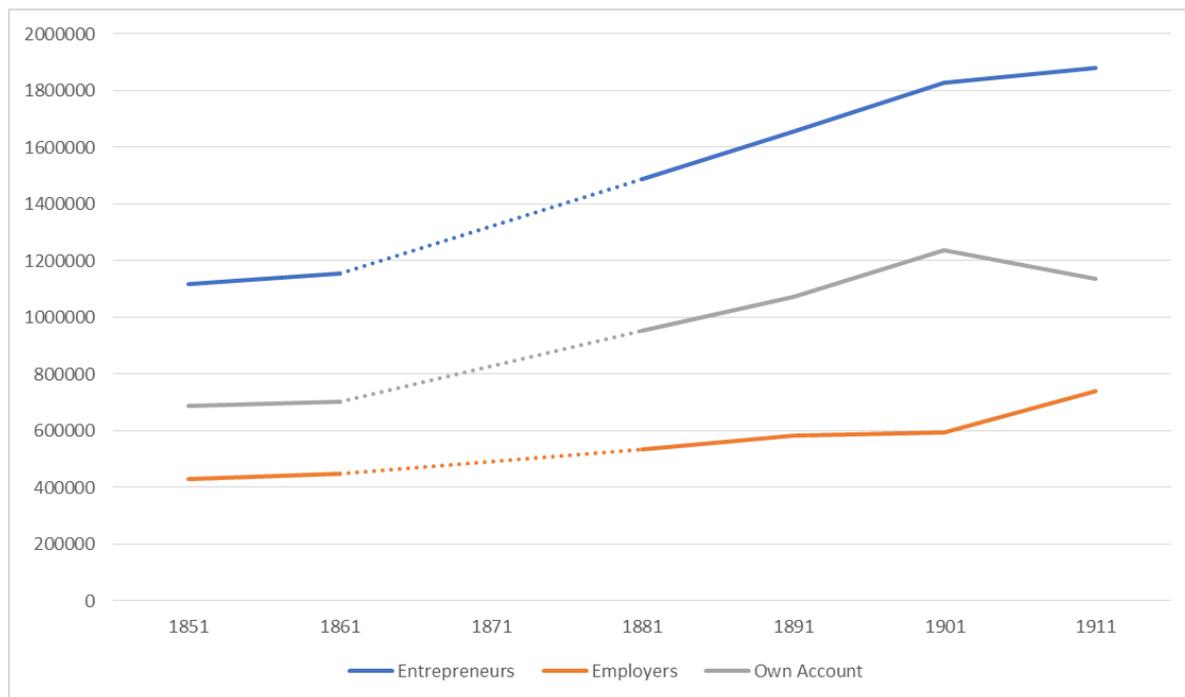


Figure 5. Reconstructed total numbers of own account, employers and all entrepreneurs, 1851-1911.

In different types of industry the trends differ. For the total number of entrepreneurs for the thirteen occupational categories within EA17 that contain entrepreneurs the trends are shown in Figure 6. This shows contrasts within the general pattern of steady increase in numbers across the period. Maker-dealers were the largest group throughout this period, and by the later censuses the gap in size between those businesses and others was substantial up to 1901. But after 1901 they are the group that experiences the most marked downturn at the end of the growth period. They were experiencing competition from two directions, reflecting their hybrid status across areas of manufacturing and retail, from metal working to jewellery to very substantial numbers involved in the production of clothing. Manufacturing was now experiencing considerable consolidation into larger firms as a result of increasing mechanisation of some processes requiring greater capital investment. Retail was also beginning to see expansion of the multiple stores and large operations with branches across the country. Although the sectors were not declining in total employment, the number of independent businesses was under challenge, especially the smallest. Similar trends also occurred in agricultural food processing and personal services, though they were less marked than for maker-dealers. A very small contraction in business numbers also affected transport. These sectors were all subject to increasing consolidation into larger businesses, with the

personal services sector also showing impacts from the increasing activity in charitable, municipal and state sectors. These were expanding into large scale provision in hospitals, state schools and other sectors that absorbed or displaced a previously buoyant area for entrepreneurship as private nursing, medical practitioners, and private schools proprietorships, which had been particularly important fields for female businesses.

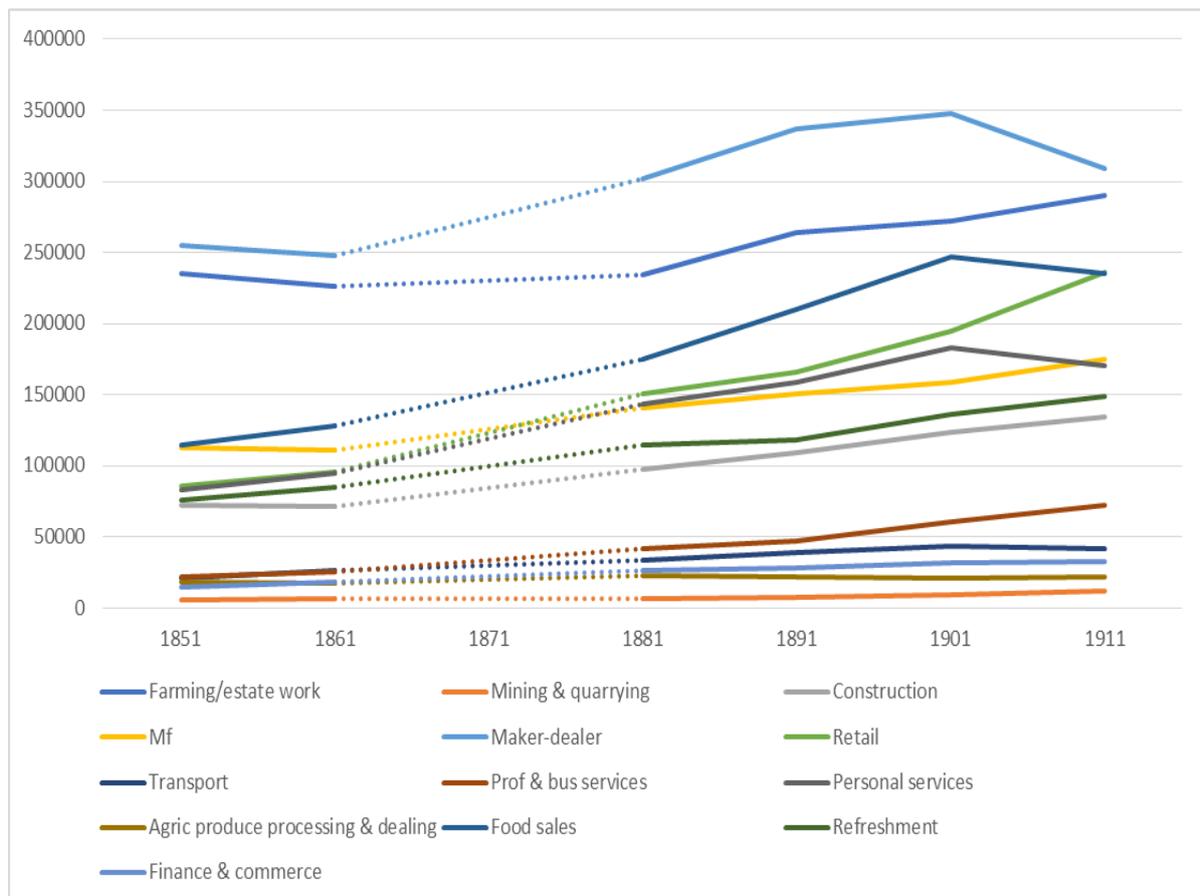


Figure 6. Reconstructed total entrepreneur numbers, 1851-1911, for 13 aggregated business sectors.

In farming and the rest of agriculture the trends were more mixed. Although farm labour had been steadily declining over the century, especially after the agricultural depression in 1873, the number of independent proprietors was tending to grow. New machinery made it possible to manage the land with a smaller number of labourers, or as a sole proprietor with no employees. This is reflected in the slow growth in the number of farming businesses up to 1911, and which had been relatively rapid 1881-91 as a result of consolidation after the depression.

Retail was the most rapidly growing sector in terms of business numbers across the whole period. Construction business numbers also grew rapidly and consistently across the period. Professional and business services, refreshments, and finance and commerce also saw continuous growth. Other sectors had more uneven profiles. For example, manufacturing had declined 1851-61, but then grew rapidly before slowing down over 1891-1901, after which the sector again expanded. Mining had very little change in business numbers until after 1891.

In all these cases the reconstruction process allows us to say a great deal more about the evolution of business proprietors in England and Wales than was previously possible. The reconstructed census information for 1851-81 can be joined up with corrected census data for 1891-1911 which also allow us for the first time to join up entrepreneurship information with modern census series, and to track changes from 1851 to the current period on a consistent and uniform basis. The link with modern information is developed in subsequent publications.

6. Conclusion

This paper has sought to develop reconstructed estimates of the employment status of individuals (employer, own account or worker) aligned between the 1851-81 censuses definitions of Occodes and Sub-Occodes, and the explicitly identified employment status available in the later censuses for 1891-1911. This has generally been achieved by the methodology followed but, as to be expected, the reconstructions perform better the higher the level of aggregations of the data since these amalgamate problematic groups. The ambition of aligning at Sub-Occode level is demanding. For the purposes of analysis, it is not recommended that most individual Occodes or Sub-Occodes are used. Instead the disaggregate reconstructions should be used as building blocks for more aggregate sector, geographical or occupational analysis.

Researchers are cautioned that whilst every effort has been made to ensure accuracy, the aim of this paper and the related database deposit is to provide reconstruction estimates for aggregates. Whilst the individual level is constructed for each economically active person in

the censuses, many of these are identified only at a statistical level; some are identified by random selection. The individuals are identified in order to provide a means to estimate aggregates. The individual level should be used only with care, taking account of how the cases were estimated; generally fully accurate identification of an individual's employment status is only available for extraction Groups 1-6; and with lesser confidence those identified by the logit regression with rounded estimates. Those individuals identified by the regression with unrounded estimates, and by the extrapolation Methods 3 and 4 derive purely from random selection.

It is also important to note that reconstructions at Occode or Sub-Occode level are not reliable for very small groups Occode categories or where changes occurred in the way information was originally defined or collected in the census, or is coded in I-CeM. This particularly affects census respondents or enumerators that gave insufficient information (e.g. general categories of 'manufacturer', 'labourer', or 'cotton operative') and many large-scale female occupations, especially for textiles in early censuses, and female occupations more generally for clothing and personal service. Generally, these constraints are greater in earlier than later census years because GRO improved the precision of occupational descriptors.

Despite these caveats, the reconstructions provide total numbers of employers and own account business proprietors for the 1851, 1861 and 1881 censuses which can be compared to those available for the 1891, 1901 and 1911 censuses. As figures 4 and 5 show, the estimates fit with general patterns of sectoral development and population growth. Further publications are developing the analysis of these data. The data will be available in the final database deposit and will allow other scholars to examine the entirety of England and Wales' business population in a manner never before possible.

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 data used for the 1881 pilot and in this paper derives from Schürer, Kevin and Woollard, Matthew (University of Essex) (2000) *1881 Census for England and Wales, the*

Channel Islands and the Isle of Man (Enhanced Version) [computer file] UKDA, SN-4177, supported by ESRC, derived from the manuscript census enumerators' books, transcribed by Genealogical Society of Utah and Federation of Family History Societies.

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 authors, have collectively worked on the new versions of I-CeM.

References

- Higgs, E., C. Jones, K. Schürer and A. Wilkinson (2015) *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex).
- Bennett, Robert J. and Newton Gill (2015) Employers in the 1881 population census of England and Wales, *Local Population Studies*, 94, 29-49.
- Bennett, Robert J., Smith, Harry and Montebruno, Piero (2018) *The Population of Non-Corporate Business Proprietors in England and Wales 1891-1911*.
- 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.

Schürer, Kevin and Woollard, Matthew (University of Essex) (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)* [computer file] UKDA, SN-4177, transcribed by Genealogical Society of Utah and Federation of Family History Societies.

Schürer, K., E. Higgs, A.M. Reid, E.M Garrett (2016) *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, [data collection]. UK Data Service, SN: 7481, <http://dx.doi.org/10.5255/UKDA-SN-7481-1>; enhanced.

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.*

WP 6: Smith, Harry J. and Bennett, Robert J. (2017) *Urban-Rural Classification using Census data, 1851-1911.*

WP 7: Smith, Harry, Bennett, Robert J., and Radicic, Dragana (2017) *Classification of towns in 1891 using factor analysis.*

WP 8: Bennett, Robert J., Smith, Harry, and Radicic, Dragana (2017) *Classification of occupations for economically active: Factor analysis of Registration Sub-Districts (RSDs) in 1891.*

WP 9: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2018) *Reconstructing entrepreneurship and business numbers for censuses 1851-81/*

Appendix 1: Additional Sub-Occodes to better identify employment statuses

Each of 83 new Sub-Occodes 800-882 is defined from the original I-CeM Occodes, which number 797 in total. This gives 880 Sub-Occodes in total. -Occodes were created by examining the proportions of employers, own account and workers in the most common occupation descriptors in 1891. Strings which were used by more than 100 people and which were over 80 per cent entrepreneurs were identified. The effect of removing these strings from the existing Occodes was then calculated and if they substantially reduced the number of employers and own account individuals left in the parent Occode they were used to create a new sub Occode. For example, individuals in Occode 473, Earthenware, China, Glass – dealers, with either ‘dealer’ or ‘merchant’ in their occupations were found to be 80 per cent employers and own account in 1891. Removing individuals with those strings from the original Occode removed 76 per cent of the entrepreneurs in Occode 473 and 64 per cent of the total Occode. Consequently, Occode 473 was split into two sub Occodes, one (sub Occode 828) which included individuals that were highly likely, according to their strings, to be entrepreneurs and another (sub Occode 473) which was made up of the remainder. The remainder sub Occode (473) did include some employers and some own account, and may have included a greater proportion in earlier censuses if the occupational descriptors used by entrepreneurs in earthenware, china and glass distribution were different in earlier censuses.

New Sub-Occode	Occupational category from which separated, with original I-CeM Occode
800	Dental E/OA Occode 43
801	Subordinate Medical Service E/OA Occode 49
802	Schoolmasters And Teachers E/OA Occode 52
803	Authors Ends Journalists & Creative Advert Worker E/OA Occ 56
804	Musician, Music Master, Singer (Not Teachers) E/OA Occode 76
805	Laundry Work: Washer, Iron Mangle (Not Dom) E/OA Occ 105
806	Broker Factor Comm Agent (Not - Mine Quarry Ins) E/OA Occ 112
807	Surveyors (House Agents) E/OA Occode 117
808	Ag Machine--Proprietor, Attendant E/OA Occode 186
809	Mine - Owner, Agent, Manager E/OA Occode 202
810	Pig Iron Manufacture (Blast Furnace) E/OA Occode 241
811	Steel--Manufacture, Smelting, Founding E/OA Occode 244
812	Blacksmiths E/OA Occode 262
813	File Makers E/OA Occode 292
814	Cutlery And Scissor Makers E/OA Occode 294
815	Lock, Key Makers E/OA Occode 315
816	Bicycle Makers, Bicycle Repairers E/OA Occode 362
817	Ironmonger; Hardware-Dlr, Merchant E/OA Occode 369
818	Metal Dealers E/OA Occode 380
819	Watch And Clock Makers E/OA Occode 386

820	Carpenter, Joiner E/OA Occode 409
821	Cabinet Makers E/OA Occode 437
822	Dealer In Works Of Art E/OA Occode 450
823	Furniture Brokers Dealers E/OA Occode 451
824	House & Shop Fittings Dealers E/OA Occode 452
825	Brick, Plain Tile, Terra-Cotta--Maker E/OA Occode 464
826	Earthenware, China, Porcelain--Mf E/OA Occode 467
827	Other Workers In Glass Manufacture E/OA Occode 470
828	Earthenware, China, Glass--Dealer E/OA Occode 473
829	Manufacturing Chemist E/OA Occode 480
830	Chemists Druggists E/OA Occode 482
831	Soap Boilers And Makers E/OA Occode 487
832	Oil And Colourmen E/OA Occode 498
833	Brush And Broom Makers E/OA Occode 512
834	Dealers In Leather E/OA Occode 516
835	Stationers (All Assumed Retail/Law Strs) E/OA Occ 529
836	Paper Dealers E/OA Occode 530
837	Printers So Described E/OA Occode 536
838	Publish, Bookseller & Library (Exc Public Libr) E/OA Occode 543
839	Newspaper Publishers E/OA Occode 546
840	Cotton & Cotton Goods Mf Spinning E/OA Occode 549
841	Flannel Manufacture (Various) E/OA Occode 563
842	Woollen Cloth Mf Undefined E/OA Occode 571
843	Worsted And Stuff Mf Undefined E/OA Occode 572
844	Silk Workers - Undefined E/OA Occode 579
845	Rope, Twine, Cord – Makers E/OA Occode 584
846	Hosiery Manufacture E/OA Occode 592
847	Lace Manufacture E/OA Occode 593
848	Fancy Goods Textile Mf (Various) E/OA Occode 599
849	Other Work Sundry Fabrics Undefined E/OA Occode 607
850	Drapers Linen Drapers Mercers E/OA Occode 628
851	Woolstapler Wool Dealer Broker E/OA Occode 630
852	Cloth Worsted & Stuff Merchants Dlrs E/OA Occode 631
853	Yarn Merchant E/OA Occode 632
854	Waste Merchant E/OA Occode 633
855	Silk Merchants Dealers E/OA Occode 634
856	Cotton & Calico Dealers E/OA Occode 638
857	Hatters E/OA Occode 651
858	Tailors (Not Merchants) - Default E/OA Occode 653
859	Clothier & Outfitter (Dlrs & Merchants) E/OA Occode 655
860	Glove Manufacture E/OA Occode 661
861	Hosiers Haberdashers E/OA Occode 662
862	Shoe And Boot Makers (And Repairers) E/OA Occode 663
863	Footwear Dealers E/OA Occode 666
864	Cowkeepers, Milksellers E/OA Occode 677
865	Cheesemongers Buttermen E/OA Occode 679
866	Butchers And Meat Salesmen E/OA Occode 682
867	Poulterers Game Dealers E/OA Occode 684

868	Fishmongers E/OA Occode 685
869	Corn Flour And Seed Merchants Dealers E/OA Occode 687
870	Bakers (Dealers) E/OA Occode 691
871	Biscuit Cake Dealers (Confectionists) E/OA Occode 692
872	Grocers Tea Dealers E/OA Occode 697
873	Hop Merchants Dealers E/OA Occode 704
874	Tobacconists E/OA Occode 707
875	Coffee And Eating-House Keepers E/OA Occode 711
876	Wine And Spirit Merchants E/OA Occode 722
877	Advertising Agents E/OA Occode 730
878	Others About Animals (Not Agric) E/OA Occode 735
879	Rag Dealers E/OA Occode 744
880	General Shopkeepers Dealers E/OA Occode 758
881	Pawnbrokers E/OA Occode 759
882	Mfrs, Mngers, Superintdnts, Foremen (Unspec) E/OA Occode 761

Table A.1 Definition of Sub-Occodes use for reconstruction and their relation to I-CeM Occodes.

Appendix 2: Summary of reconstructions compared to extraction Groups.

An important test of the reconstruction methods is comparison of the logit estimator using 1891 and 1901 models, with the extracted data from the census Groups 1-6 for each type of potential entrepreneur. The extraction Groups derive from the original census respondent's comments and hence are a direct validation of the model. If the logit regression provides an effective method of reconstruction it should be able to correctly predict those identified from the census by direct extractions. The comparisons are summarised here for all the reconstruction models used for 1881, and then 1851 and 1861; in each case first for E+OA, and then for E only.

Reconstruction for 1881 E+OA

1.General

Group	0	1	Total
1	47,131	134,406	181,537
	25.96%	74.04%	100.00%
2	813	432	1,245
	65.30%	34.70%	100.00%
3	21,679	25,753	47,432
	45.71%	54.29%	100.00%
4	36,877	489	37,366
	98.69%	1.31%	100.00%
5	3,998	102,189	106,187
	3.77%	96.23%	100.00%
6	2,040	6,846	8,886
	22.96%	77.04%	100.00%
Total	112,538	270,115	382,653
	29.41%	70.59%	100.00%

2. Non-farmers

Group	0	1	Total
1	47,083	45,142	92,225
	51.05%	48.95%	100.00%
2	813	391	1,204
	67.52%	32.48%	100.00%
3	21,679	25,753	47,432
	45.71%	54.29%	100.00%
4	1,018	489	1,507
	67.55%	32.45%	100.00%
5	3,988	3,986	7,984
	50.08%	49.92%	100.00%

6	2,040	6,846	8,886
	22.96%	77.04%	100.00%
Total	76,631	82,607	159,238
	48.12%	51.88%	100.00%

3. Farmers

Group	0	1	Total
1	48	89,294	89,312
	0.05%	99.95%	100.00%
2	0	41	41
	0.00%	100.00%	100.00%
4	35,859	0	35,859
	100.00%	0.00%	100.00%
5	0	98,203	98,203
	0.00%	100.00%	100.00%
Total	35,907	187,508	223,445
	16.07%	83.93%	100.00%

Reconstruction for 1881 E Only

1. General

Group	0	1	Total
1	64,182	117,249	181,431
	35.38%	64.62%	100.00%
2	172	209	381
	45.14%	54.86%	100.00%
3	33,874	13,558	47,432
	71.42%	28.58%	100.00%
4	261	197	458
	56.99%	43.01%	100.00%
5	39,148	66,930	106,078
	36.90%	63.10%	100.00%
6	3,079	3,663	6,742
	45.67%	54.33%	100.00%
Total	140,716	201,806	342,522
	41.08%	58.92%	100.00%

2. Non-farmers

Group	0	1	Total
1	40,020	52,205	92,225
	43.39%	56.61%	100.00%
2	148	193	341
	43.40%	56.60%	100.00%
3	33,874	13,558	47,432

	71.42%	28.58%	100.00%
4	261	197	458
	56.99%	43.01%	100.00%
5	4,978	3,006	7,984
	62.35%	37.65%	100.00%
6	3,079	3,663	6,742
	45.67%	54.33%	100.00%
Total	82,360	72,822	155,182
	53.07%	46.93%	100.00%

3. Farmers

Group	0	1	Total
1	24,162	65,044	89,206
	27.09%	72.91%	100.00%
2	24	16	40
	60.00%	40.00%	100.00%
5	34,170	63,924	98,094
	34.83%	65.17%	100.00%
Total	58,356	128,984	187,340
	31.15%	68.85%	100.00%

Reconstruction for 1851 E+OA

1. General

Group	0	1	Total
1	36,537	128,714	165,251
	22.11%	77.89%	100.00%
2	745	3,779	4,524
	16.47%	83.53%	100.00%
3	16,344	18,364	34,708
	47.09%	52.91%	100.00%
4	5,368	59,030	64,398
	8.34%	91.66%	100.00%
5	3,824	51,539	55,363
	6.91%	93.09%	100.00%
6	695	2,659	3,354
	20.72%	79.28%	100.00%
Total	63,513	264,085	327,598
	19.39%	80.61%	100.00%

2. Non-farmers

Group	0	1	Total
1	30,081	27,193	57,274
	52.52%	47.48%	100.00%

2	572	481	1,053
	54.32%	45.68%	100.00%
3	16,344	18,359	34,703
	47.10%	52.90%	100.00%
4	7	4	11
	63.64%	36.36%	100.00%
5	309	205	514
	60.12%	39.88%	100.00%
6	695	2,659	3,354
	20.72%	79.28%	100.00%
Total	48,008	48,901	96,909
	49.54%	50.46%	100.00%

3. Farmers

Group	0	1	Total
1	6,456	101,521	107,977
	5.98%	94.02%	100.00%
2	173	3,298	3,471
	4.98%	95.02%	100.00%
3	0	5	5
	0.00%	100.00%	100.00%
4	5,361	59,026	64,387
	8.33%	91.67%	100.00%
5	3,515	51,334	54,849
	6.41%	93.59%	100.00%
Total	15,505	215,184	230,689
	6.72%	93.28%	100.00%

Reconstruction for 1851 E only

1. General

Group	0	1	Total
1	53,769	111,482	165,251
	32.54%	67.46%	100.00%
2	1,454	2,493	3,947
	36.84%	63.16%	100.00%
3	25,836	8,872	34,708
	74.44%	25.56%	100.00%
4	19,418	39,612	59,030
	32.90%	67.10%	100.00%
5	17,049	38,314	55,363
	30.79%	69.21%	100.00%
6	977	1,682	2,659
	36.74%	63.26%	100.00%

Total	118,503	202,455	320,958
	36.92%	63.08%	100.00%

2. Non-farmers

Group	0	1	Total
1	29,001	28,273	57,274
	50.64%	49.36%	100.00%
2	354	122	476
	74.37%	25.63%	100.00%
3	25,835	8,868	34,703
	74.45%	25.55%	100.00%
4	1	3	4
	25.00%	75.00%	100.00%
5	289	225	514
	56.23%	43.77%	100.00%
6	977	1,682	2,659
	36.74%	63.26%	100.00%
Total	56,457	39,173	95,630
	59.04%	40.96%	100.00%

3. Farmers

Group	0	1	Total
1	24,768	83,209	107,977
	22.94%	77.06%	100.00%
2	1,100	2,371	3,471
	31.69%	68.31%	100.00%
3	1	4	5
	20.00%	80.00%	100.00%
4	19,417	39,609	59,026
	32.90%	67.10%	100.00%
5	16,760	38,089	54,849
	30.56%	69.44%	100.00%
Total	62,046	163,282	225,328
	27.54%	72.46%	100.00%
	ReadyE_51		
Type	0	1	Total

Reconstruction for 1861 E+OA

1. General

Group	0	1	Total
-------	---	---	-------

1	48,040	127,282	175,322
	27.40%	72.60%	100.00%
2	1,250	1,936	3,186
	39.23%	60.77%	100.00%
3	21,193	18,123	39,316
	53.90%	46.10%	100.00%
4	17,911	39,575	57,486
	31.16%	68.84%	100.00%
5	6,184	81,797	87,981
	7.03%	92.97%	100.00%
6	2,342	5,405	7,747
	30.23%	69.77%	100.00%
Total	96,920	274,118	371,038
	26.12%	73.88%	100.00%

2. Non-farmers

Group	0	1	Total
1	44,048	34,335	78,383
	56.20%	43.80%	100.00%
2	1,188	536	1,724
	68.91%	31.09%	100.00%
3	21,193	18,114	39,307
	53.92%	46.08%	100.00%
4	14,651	1,468	16,119
	90.89%	9.11%	100.00%
5	2,844	2,499	5,343
	53.23%	46.77%	100.00%
6	2,342	5,383	7,725
	30.32%	69.68%	100.00%
Total	86,266	62,335	148,601
	58.05%	41.95%	100.00%

3. Farmers

Group	0	1	Total
1	3,992	92,947	96,939
	4.12%	95.88%	100.00%
2	62	1,400	1,462
	4.24%	95.76%	100.00%

3	0	9	9
	0.00%	100.00%	100.00%
4	3,260	38,107	41,367
	7.88%	92.12%	100.00%
5	3,340	79,298	82,638
	4.04%	95.96%	100.00%
6	0	22	22
	0.00%	100.00%	100.00%
Total	10,654	211,783	222,437
	4.79%	95.21%	100.00%

Reconstruction for 1861 E only

1. General

Group	0	1	Total
1	58,548	116,787	175,335
	33.39%	66.61%	100.00%
2	612	1,386	1,998
	30.63%	69.37%	100.00%
3	25,161	14,158	39,319
	63.99%	36.01%	100.00%
4	13,826	25,749	39,575
	34.94%	65.06%	100.00%
5	29,150	58,831	87,981
	33.13%	66.87%	100.00%
6	2,337	3,068	5,405
	43.24%	56.76%	100.00%
Total	129,634	219,979	349,613
	37.08%	62.92%	100.00%

2. Non-farmers

Group	0	1	Total
1	35,329	43,067	78,396
	45.06%	54.94%	100.00%
2	218	318	536
	40.67%	59.33%	100.00%
3	25,159	14,151	39,310
	64.00%	36.00%	100.00%
4	754	714	1,468
	51.36%	48.64%	100.00%
5	3,204	2,139	5,343

	59.97%	40.03%	100.00%
6	2,324	3,059	5,383
	43.17%	56.83%	100.00%
Total	66,988	63,448	130,436
	51.36%	48.64%	100.00%

3. Farmers

Group	0	1	Total
1	23,219	73,720	96,939
	23.95%	76.05%	100.00%
2	394	1,068	1,462
	26.95%	73.05%	100.00%
3	2	7	9
	22.22%	77.78%	100.00%
4	13,072	25,035	38,107
	34.30%	65.70%	100.00%
5	25,946	56,692	82,638
	31.40%	68.60%	100.00%
6	13	9	22
	59.09%	40.91%	100.00%
Total	62,646	156,531	219,177
	28.58%	71.42%	100.00%

Appendix 3. Estimates of the full model for 1891 to estimate 1891.

INDIVIDUAL-LEVEL 1891: Weighted logit model for the probability of being an entrepreneur (E+OA) compared to a worker with age, age squared, population density, population density squared, urban classification and population density interaction, urban classification and population density squared interaction, marital status and sex interaction, number of servants, relationship to head and 882 occupation subcategories. Density at RSD level. Only simple interactions and no levels for all interactions; Base categories urban (Urban areas), Male and Single (Sex and Marital status) and Librarians (Sub-Occode 16). Data from 1891 census corrected for upward bias and non-response bias.

Estimation of Entrepreneur = 1		Household RELA codes:	
Age	0.135*** (245.76)	1. Head	0 (.)
Age # Age	-0.00102*** (-170.21)	2. CFU member	-0.825*** (-139.54)
Density and urban dummies:		3. Older generation	-0.903*** (-53.14)
Density RSD	-0.00821*** (-139.63)	4. Siblings	-0.721*** (-71.67)
Density RSD # Density RSD	0.0000224* ** (90.38)	5. Other family	-1.065*** (-77.37)
Urban # Density RSD	0 (.)	6. Servants	-3.300*** (-76.88)
Hinterland 1 # Density RSD	0.00648*** (3.56)	7. Working title	-2.829*** (-80.60)
Hinterland 2 # Density RSD	0.00678*** (4.34)	8. Lodgers/boarders	-1.187*** (-162.21)
Rural # Density RSD	0.179*** (36.09)	9. Non-household	-1.460*** (-55.05)
Urban # Density RSD # Density RSD	0 (.)	10. Unknown RELA	-0.601*** (-48.34)
Hinterland 1 # Density RSD # Density RSD	-0.000163 (-1.88)	Occupational Sub-Occodes:	
Hinterland 2 # Density RSD # Density RSD	0.0000109 (0.66)	16. Librarians	0 (.)
Rural # Density RSD # Density RSD	-0.00323*** (-27.54)	38. Barrister, Advocate	3.857*** (24.85)
Gender and marital status:		39. Solicitor	3.748*** (27.70)
1. Male # 1. Single	0 (.)	41. Others Connected With Law	-0.699*** (-4.21)
1. Male # 2. Married	-0.100*** (-16.64)	42. Physician, Surgeon, Gen Practitioner, Incl Poor Law Hosp Doctors Etc	3.921*** (28.93)
1. Male # 4. Widowed	-0.00860 (-1.01)	43. Dentists And Dentists Assistants Minus Sub-Occode 800	2.725*** (19.82)
2. Female # 1. Single	-0.0576*** (-8.96)	44. Veterinary Surgeons	3.598*** (25.36)
2. Female # 2. Married	0.144*** (18.82)	45. Midwives	-5.928*** (-5.88)
2. Female # 4. Widowed	-0.0430*** (-5.07)	49. Subord Medical Ser & Med Services - Minus Sub-Occode 801	2.387*** (16.44)
No. of Servants	0.532*** (156.02)	52. Schoolmasters And Teachers (Default) Minus Sub-Occode 802	1.039*** (7.82)

53. University Teachers	3.283 ^{***} (22.61)	80. Performers Showmen Exhibition Service	2.836 ^{***} (19.76)
56. Authors Eds Journalists & Creative Advert Workers Minus Sub-Occode 803	1.657 ^{***} (11.95)	81. Billiards Cricket And Other Games Service	1.914 ^{***} (13.68)
57. Reporters And Shorthand Writers	1.006 ^{***} (6.60)	96. Hospital Servants' Registry Office Keepers	1.826 ^{***} (8.19)
58. Others In Writing	1.780 ^{***} (11.57)	98. Cooks (Not Domestic)	0.511 ^{***} (3.65)
59. Engaged In Scientific Pursuits (Inc Inventors & Scientific Qualifiers)	1.973 ^{***} (12.70)	104. Charwomen	0.257 (1.93)
60. Others Scientific	2.019 ^{***} (12.26)	105. Laundry Wrk: Washer, Iron, Etc. (Not Dom) Minus Sub-Occode 805	2.128 ^{***} (16.06)
61. Others Connected With Literature, Etc - Lit & Scientific Instit Service	0.182 (1.07)	106. Bath And Wash-House Service	0.854 ^{***} (5.65)
62. Others Connected With Literature, Etc - Political & Election Agents	1.311 ^{***} (8.81)	110. Window Cleaners	1.428 ^{***} (9.05)
63. Librarians (Literary, Scientific Society)	0.147 (0.93)	111. Merchant -- Commodity Undefined	3.553 ^{***} (26.25)
64. Mining Engineers	1.509 ^{***} (10.71)	112. Broker Factor Comm Agent (Not Mine Quarry Ins) Minus Sub-Occy 806	2.631 ^{***} (19.62)
65. Civil Engineers	1.566 ^{***} (11.37)	113. Salesmen And Buyers (Not Otherwise Described)	0.773 ^{***} (5.48)
66. Land, House, Ship--Surveyor	1.888 ^{***} (13.69)	114. Commercial Travelers And Manufacturers' Agents	0.398 ^{**} (2.98)
67. Mining Engineers Assistant	-1.416 ^{**} (-2.96)	115. Accountants	1.114 ^{***} (8.34)
69. Surveyors Assistant	-1.822 ^{***} (-4.75)	116. Auctioneers Appraisers Valuers House Agents	2.839 ^{***} (21.04)
70. Painters (Artists)	4.023 ^{***} (29.64)	117. Surveyors (House Agents) Minus Sub-Occode 807	2.289 ^{***} (16.04)
71. Sculptors (Artists)	3.444 ^{***} (23.30)	118. Officers Of Commercial And Others Companies	-0.220 (-1.51)
72. Architects	3.000 ^{***} (22.13)	120. Bankers	-0.362 [*] (-2.57)
73. Engravers (Artists) And Lithographic Artists	2.113 ^{***} (15.35)	121. Bank Service, Bank Officials And Clerks	-2.968 ^{***} (-16.83)
74. Map Chart Geographical Engravers	0.722 (1.27)	122. Bill Brokers, Agents And Discounters	3.276 ^{***} (20.82)
75. Photographers	2.945 ^{***} (21.91)	123. Insurance Officials Clerks	0.272 [*] (1.96)
76. Musician, Music Mstr, Singer (Not Teachers) Minus Sub-Occode 804	3.943 ^{***} (29.56)	124. Insurance Agents	0.385 ^{**} (2.87)
77. Actors	1.649 ^{***} (11.72)	135. Livery Stable Keepers, Cab And Bus Owners	4.333 ^{***} (31.80)
79. Art, Music, Theatre, Cinema & Service Etc	1.250 ^{***} (8.48)		

136. Horse Proprietor, Breeder, Dealer	3.661 ^{***}	Watermen (& Canal Workers Generally)	
	(26.44)		(3.52)
137. Motor Garage ---Proprietor, Worker	0.670 ^{***}	164. Ship And Boat Owners	3.196 ^{***}
	(4.40)		(22.46)
138. Coachman (Not Domestic): Cabman	-0.158	165. Steam Navigation Service	0.0132
	(-1.18)		(0.08)
139. Horsekeeper, Groom, Stableman (Not Domestic)	-0.598 ^{***}	167. Harbour, Dock, Wharf, Lighthouse--Officials And Servants	-0.695 ^{***}
	(-4.43)		(-4.41)
140. Motor Car Driver (Not Domestic); Motor Cab Driver	1.350 ^{***}	168. Warehousemen (Not Manchester Goods)	-1.142 ^{***}
	(4.87)		(-8.27)
141. Carmen Carriers Carters And Draymen	0.699 ^{***}	169. Meters, Weighers Engaged In Storage	-1.566 ^{***}
	(5.27)		(-7.35)
142. Hauliers (Mine)	0.534 ^{***}	171. Messenger, Porter, Watchman (Not Railway Or Government)	-0.465 ^{***}
	(3.73)		(-3.46)
143. Water Carriers	1.008 ^{***}	173. Farmer, Grazier	4.400 ^{***}
	(7.27)		(33.19)
144. Motor Van Etc. Drivers	-1.243 ^{***}	174. Crofters	2.926 ^{***}
	(-5.30)		(20.61)
145. Van, Etc.---Guard, Boy	-0.0849	177. Farm - Bailiffs Stewards Foremen	-1.549 ^{***}
	(-0.41)		(-11.16)
146. Others Connected With Carriage/Cartage (Not Mine)	0.659 ^{***}	182. Woodman	0.147
	(3.40)		(1.07)
147. Others Connected With Carriage/Cartage (Mine)	0.703 ^{***}	183. Nurseryman, Seedsman, Florist	2.533 ^{***}
	(4.41)		(18.90)
148. Omnibus Horse Drivers (Default)	-0.718 ^{***}	184. Market Gardener (Including Labourer)	3.282 ^{***}
	(-4.43)		(24.52)
149. Omnibus Motor Drivers	1.174 ^{***}	185. Other Gardener (Not Domestic)	0.458 ^{***}
	(3.79)		(3.45)
151. Omnibus Service --Others	-1.005 ^{***}	186. Agric Machine--Proprietor, Attentd Minus Sub-Occode 808	1.068 ^{***}
	(-3.39)		(7.59)
152. Tramway Service Drivers	-2.130 ^{***}	187. Others In Agriculture	2.348 ^{***}
	(-7.12)		(14.17)
153. Tramway Service Conductors	-1.859 ^{***}	188. Land Drainage Service	-0.163
	(-5.82)		(-0.88)
154. Tramway Service ---Others	-1.231 ^{***}	189. Rabbit Catchers Trappers Destroyers (On Farm)	0.893 [*]
	(-4.31)		(2.30)
156. Others On Roads - Wheelchair Proprietors, Attendants	3.055 ^{***}	190. Others Engaged About Animals (Non-Agriculture)	0.850 ^{***}
	(20.41)		(5.39)
157. Merchant Service; Seaman-Navigating Department	0.0431	191. Vermin Destroyers (Agriculture)	1.666 ^{***}
	(0.32)		(11.26)
158. Merchant Service; Seaman--Engineering Department	-0.792 ^{***}	192. Thatchers (Agriculture)	1.592 ^{***}
	(-5.49)		(9.35)
160. Boatmen On Seas (Seamen General)	0.694 ^{***}	193. Others Coal Coke Peat&Charcoal(Inc Agricl Peat/Turf Work)	1.707 ^{***}
	(5.12)		(10.64)
161. Pilots Of Ships	1.281 ^{***}	194. Fishermen	2.445 ^{***}
	(9.10)		(18.32)
162. Bargemen Lightermen	0.474 ^{***}	195. Fishermen Crofter	2.434 ^{***}
			(11.23)

196. Coal Miners - Hewers, Workers At The Coal Face	-2.946*** (-21.75)	227. Slate Quarriers	-2.208*** (-12.22)
198. Coal Miners - Others Below Ground & Undefined	-2.567*** (-17.87)	228. Slate Workers Cutters Dressers	-0.00990 (-0.05)
200. Coal Miners Above Ground	-2.405*** (-12.42)	229. Lime Burners	-0.0448 (-0.30)
201. Mine Service Engineers	-1.015*** (-5.26)	230. Sand Flint Clay Gravel Chalk Workers	-1.620*** (-9.22)
202. Mine - Owner, Agent, Manager Minus Sub-Occode 809	-0.408** (-2.78)	231. Fossil Coprolite Diggers	-2.717** (-2.70)
203. Mine Service Others (Coal)	-0.305 (-1.95)	232. Other Workers In Quarries	-1.486*** (-4.90)
204. Coke Burners	-1.887*** (-10.29)	233. Emery, Sand Paper Maker	0.972*** (4.86)
206. Patent Fuel Manufacture	-0.162 (-0.55)	234. Asbestos Maker	1.162 (1.13)
207. Iron--Miner, Quarrier	-2.616*** (-15.52)	235. Coal Merchants And Dealers	3.228*** (24.22)
208. Copper Miner	-1.118*** (-4.91)	236. Coke Dealers	4.663*** (7.13)
209. Tin Miner	-1.132*** (-7.42)	237. Stone Dealers	3.794*** (24.45)
210. Lead Miner	-1.354*** (-8.52)	238. Slate Dealers	2.763*** (14.82)
211. Miners In Other Or Undefined Minerals	-1.457*** (-8.98)	239. Sand Flint Clay Gravel Dealers	3.211*** (17.35)
212. Gold Miner	0.576 (1.89)	241. Pig Iron Mf (Blast Furnace) Minus Sub-Occode 810	-1.350*** (-8.63)
213. Silver Miner	1.792* (2.46)	242. Puddling Furnaces; Iron And Steel Rolling Mills	-1.656*** (-11.16)
214. Metalliferous Mine--Owner, General Manager, Captain	1.621*** (11.33)	243. Tube Manufacture	0.425* (2.36)
216. Stone Quarry Managers Agents	1.142*** (6.33)	244. Steel--Mf, Smelting Founding Minus Sub-Occode 811	-2.130*** (-12.84)
217. Slate Quarry Managers Agents	0.219 (0.71)	245. Galvanized Sheet Manufacture	-0.866* (-2.24)
218. Sand Flint Clay Gravel Chalk Managers Agents	-0.502 (-0.89)	246. Tinplate Manufacture	-0.0242 (-0.18)
219. Other Quarries Managers Agents	1.654*** (9.82)	247. Copper Manufacture	-1.880*** (-9.42)
220. Stone Quarry Mine Owners	2.919*** (15.62)	248. Lead Manufacture	-1.149*** (-5.71)
221. Slate Quarry Mine Owners	2.682*** (13.47)	249. Zinc Manufacture	1.007*** (6.28)
222. Sand Flint Clay Gravel Quarry Mine Owners	6.641*** (9.20)	250. Brass, Bronze-Manufacture	0.830*** (4.36)
224. Stone Quarriers	-1.333*** (-9.65)	251. Gold Silver Smelters Etc (Extraction)	0.418* (2.10)
225. Limestone Quarrier	-1.633*** (-8.08)	252. Manufacture Of Other Or Unspecified Metals	-0.856*** (-4.28)
226. Stone Cutters Dressers	-0.250 (-1.72)	253. Patternmaker (Engine & Machine Making (Undefined - Not Textile)	-0.505*** (-3.47)

254. Patternmaker (Spinning Weaving Machinery Making)	0.846 ^{***} (5.16)	278. Engine And Machine Makers (Undefined - Not Textile)	-0.0978 (-0.70)
255. Patternmaker (Agricultural Machine & Implement Making)	-0.205 (-1.05)	279. Domestic Machinery Makers	-0.000223 (-0.00)
256. Patternmaker (Domestic Machinery Making)	0.672 (0.65)	280. Agricultural Machine And Implement Makers	3.137 ^{***} (22.51)
257. Patternmaker (Undefined)	-0.262 (-1.51)	281. Railway - Signal Switch Turntable Makers	0.200 (1.03)
258. Millwright	0.382 ^{**} (2.75)	282. Riveter Rigger Metal Plate Worker (Not Ship, Textile, Boiler)	-1.492 ^{***} (-6.90)
259. Ironfounder -- Moulders, Core Makers, Fetter, Cupola, Over-Man, Foundry Labourer, Other & Undefined	-0.992 ^{***} (-7.33)	284. Electrical Cable Makers	0.583 ^{**} (3.23)
260. Brass Founders	0.883 ^{***} (6.42)	285. Electrical Lamp Makers	0.276 (0.84)
261. Farriers	1.664 ^{***} (12.09)	286. Other Electrical Apparatus Makers	-0.431 ^{**} (-2.89)
262. Blacksmiths Minus Sub-Occode 812	0.938 ^{***} (7.08)	287. Electricians (Undefined)	1.586 ^{***} (11.34)
263. Fitters, Turners (Engine And Machine)	-1.259 ^{***} (-9.25)	289. Toolmakers (Agricultural Machine & Implement Making)	1.040 ^{***} (7.10)
264. Colliery Fitters	-2.556 ^{***} (-4.99)	290. Toolmakers	0.498 ^{***} (3.62)
265. Railway - Signal Switch Turntable Fitters	-1.611 ^{***} (-10.53)	291. Domestic Implement Makers - Toolmakers	1.111 ^{***} (6.52)
267. Brass Finishers	-0.160 (-1.12)	292. File Makers Minus Sub-Occode 813	-0.520 ^{***} (-3.51)
268. Coppersmiths (Various)	-0.00299 (-0.02)	293. Saw Makers	0.981 ^{***} (6.63)
269. Metal Machinist (Engine And Machine Making)	0.329 [*] (2.33)	294. Cutlery & Scissor Makers Minus Sub-Occode 814	0.541 ^{***} (3.98)
270. Metal Machinist (Spinning Weaving Machinery Making)	0.417 [*] (2.51)	295. Needle Makers	0.549 ^{***} (3.75)
271. Metal Machinist (Agricultural Machine & Implement Making)	0.0422 (0.24)	296. Pin Makers	-0.138 (-0.73)
272. Metal Machinist (Domestic Machinery Making)	-1.126 ^{***} (-7.42)	297. Steel Pen Makers	-0.00345 (-0.02)
273. Undefined Engine And Machine Making (Not Textile)	-0.0398 (-0.25)	298. Roller Engravers Block Cutters (For Textile Printing)	0.100 (0.55)
275. Boiler Makers	-1.118 ^{***} (-8.05)	299. Type Founders	0.505 ^{**} (2.86)
276. Other Workers (Spinning Weaving Machinery Making)	-0.338 (-1.91)	300. Die, Seal, Coin, Medal & Maker	1.696 ^{***} (10.94)
277. Other Workers In Textile Machinery Fittings	0.513 ^{***} (3.65)	301. Gunsmiths	0.648 ^{***} (4.67)
		302. Ordnance Manufacture	-2.876 ^{***} (-5.05)
		303. Arms And Ordnance - Others	-0.578 [*] (-2.13)
		304. Sword And Bayonet Makers	0.454

	(1.46)	330. Iron Fence And Gate Maker	0.796 ^{***}
305. Nail Manufacture	0.120		(4.43)
	(0.86)	331. Other Iron Goods Makers	-0.672 ^{***}
306. Bolt Nut Rivet And Screw Manufacture	-0.133		(-4.82)
	(-0.91)	333. Spring Maker	1.081 ^{***}
307. Anchor & Chain Manufacture	0.132		(6.81)
	(0.92)	334. Iron Workers (Undefined)	-0.901 ^{***}
308. Gas Stove Makers	0.942 ^{***}		(-6.49)
	(5.79)	335. Other Implement Makers	1.345 [*]
309. Stove Grate Range Fire-Iron Makers	-0.201		(2.48)
	(-1.24)	336. Tin Workers	1.228 ^{***}
310. Brass Bedstead Makers	-0.236		(8.24)
	(-0.67)	337. Metal Burnishers (Undefined)	0.0998
311. Iron Bedstead Makers	-0.264		(0.52)
	(-1.55)	338. Fancy Chain Ring Gilt Toy Makers	1.974 ^{***}
312. Wire/Spring Mattress Makers	1.135 ^{***}		(8.02)
	(7.25)	339. Other Metal Workers	0.230
313. Wire Makers Workers Weavers	0.532 ^{***}		(1.57)
	(3.83)	340. Mixed Or Unspecified Metals - Tube Manufacture	-0.785 ^{***}
314. Wire Fencer	1.194 ^{***}		(-4.09)
	(7.68)	342. Clasp Buckle Hinge Makers - Not Brass Or Iron	0.308
315. Lock, Key Makers Minus Sub-Occode 815	0.726 ^{***}		(1.46)
	(5.20)	343. Whitesmiths	0.895 ^{***}
316. Gas Fittings Makers	0.747 ^{***}		(6.57)
	(4.41)	344. Shackle Makers	-0.157
317. Lamp, Lantern, Candlestick--Maker	0.955 ^{***}		(-0.59)
	(6.49)	345. Ship Boat Platers Rivetters	-0.566 ^{***}
318. Silver-Smiths, Buffers, Burnishers, Finishers	0.908 ^{***}		(-3.75)
	(6.64)	347. Fitters (Ships)	-0.293
319. Pewter White Metal Plated Ware Manufacture	0.876 ^{***}		(-1.62)
	(5.29)	348. Shipwrights - Wood Ships	-0.590 ^{***}
320. Metal Refiners Workers	-0.163		(-4.32)
	(-0.59)	349. Shipwrights - Metal Ships	0.963 ^{***}
321. Tinsplate Goods Manufacture	1.082 ^{***}		(6.90)
	(8.02)	350. Wood Ships - Worker In Wood	-0.637
322. Copper Workers	0.966 [*]		(-1.86)
	(2.53)	351. Mast, Yard, Oar, Block Maker	-0.287
323. Leaden Goods Manufacture (Various)	0.833 ^{***}		(-0.94)
	(4.35)	352. Ship Boat Painters (Wood)	-0.0156
324. Zinc Goods Workers	0.508		(-0.09)
	(1.93)	353. Ship Boat Painters (Iron)	2.861 ^{**}
325. Brass Bronze Implement Makers	0.894 ^{***}		(2.92)
	(3.70)	355. Others In Ship/Boat Building - Wood (Default)	0.670 ^{***}
326. Brass, Bronze Goods Workers	-0.265		(4.29)
	(-1.81)	356. Others In Ship/Boat Building - Metal	-3.141 ^{***}
327. Brass Clasp Buckle Hinge Makers	-0.00450		(-4.35)
	(-0.03)	357. Riggers (Ships)	0.0565
328. Iron Domestic Implement Makers	0.875 [*]		(0.37)
	(1.97)	358. Ship Chandler	2.534 ^{***}
329. Iron Clasp Buckle Hinge Makers	-1.494 ^{***}		(15.34)
	(-7.96)	359. Sailmakers	0.633 ^{***}
			(4.51)
		360. Railway--Coach, Wagon Maker	-1.655 ^{***}
			(-9.50)
		361. Tram Car Maker	-0.654

362. Bicycle Makers & Repairers Minus Sub-Occode 816	(-0.63) 0.878 ^{***}	389. Weighing And Measuring Machine Makers	0.904 ^{***} (6.13)
363. Motor Car Chassis Maker: Motor Car Mechanic	(6.20) 0.0623	390. Tinmen (Meter Making)	2.285 ^{***} (13.65)
364. Motor Car Body Maker	(0.26) 1.046 ^{***}	391. Dental Instrument And Apparatus Makers	2.359 ^{***} (3.83)
365. Coach, Carriage--Maker	(3.29) 1.054 ^{***}	392. Surgical Instrument And Apparatus Makers	1.529 ^{***} (10.32)
366. Wheelwright	(7.89) 1.655 ^{***}	393. Piano Organ Makers	1.537 ^{***} (11.31)
367. Cartwrights	(12.43) 0.376 [*]	394. Other Musical Instrument Makers	1.991 ^{***} (12.35)
368. Perambulator Maker	(2.44) 1.256 ^{***}	395. Toy Makers	2.047 ^{***} (13.54)
369. Ironmonger; Hardware-Dlr, Mercht Minus Sub-Occode 817	(7.71) 2.169 ^{***}	396. Fishing Tackle Makers	1.538 ^{***} (9.79)
370. General Household Appliance Dealers	(16.24) 2.464 ^{***}	397. Apparatus For Other Games Makers	1.221 ^{***} (8.19)
372. Dealers In Electrical Apparatus	(17.69) 0.492 ^{**}	398. Gold And Silversmiths Jewellers (Dealers)	2.543 ^{***} (17.02)
374. Bicycle Dealers	(3.25) 0.0795	399. Musical Instrument Dealers	3.439 ^{***} (22.67)
376. Dealers In Copper	(0.52) 2.935 ^{***}	400. Toy Dealers	3.568 ^{***} (25.42)
377. Dealers In Tin, Tinplate	(17.43) 1.491 ^{***}	401. Fishing Tackle Dealers	2.032 ^{***} (10.10)
378. Zinc Dealers	(9.26) 1.672 ^{***}	402. Apparatus For Other Games Dealers	4.383 ^{***} (3.55)
379. Lead Dealers	(10.34) 3.344 ^{***}	403. Photographic Apparatus, Phil Instrument (Scient & Optical)- Dlrs	2.493 ^{***} (8.70)
380. Metal Dealers Minus Sub- Occode 818	(20.99) 1.367 ^{***}	404. Weighing And Measuring Machine Dealers	3.640 ^{***} (16.77)
381. Other Dealers In Machines, Implements Etc.	(8.55) 0.775 ^{***}	405. Builders	3.355 ^{***} (25.20)
382. Gold And Silversmiths Jewellers (Not Dealers)	(5.15) 2.139 ^{***}	406. Thatchers (Not Agriculture)	1.867 ^{***} (13.31)
383. Lapidaries And Other Precious Material Workers	(15.93) 0.918 ^{***}	408. Builders' Excavators	0.632 [*] (2.25)
384. Gold And Silver Beaters	(5.10) 0.877 ^{***}	409. Carpenter, Joiner Minus Sub- Occode 820	0.674 ^{***} (5.09)
385. Other Workers Gold Silver Jewellery	(3.97) 1.499 ^{***}	411. Stove Setters Furnace Oven Liners	0.371 (1.64)
386. Watch And Clock Makers Minus Sub-Occode 819	(10.60) 2.503 ^{***}	412. Bricklayer	0.189 (1.42)
387. Philosophical Instrument Maker (Scientific & Optical)	(18.75) 2.060 ^{***}	414. Mason	0.520 ^{***} (3.91)
388. Photographic Apparatus Makers	(14.50) 1.926 ^{***}	416. Slaters & Tilers	1.254 ^{***} (9.16)
	(10.69)	417. Plasterer	0.886 ^{***}

	(6.61)		(16.02)
419. Whitewashers	1.223 ^{***}	448. Cane, Rush, Straw - Other Workers	2.015 ^{***}
	(6.50)		(13.44)
420. Paperhangers	1.969 ^{***}	449. Willow Cane And Rush Dealers	2.455 ^{***}
	(14.34)		(16.57)
421. Painters, Distemperers & Decorators	1.074 ^{***}	450. Dealer In Works Of Art Minus Sub-Occode 822	3.040 ^{***}
	(8.10)		(20.32)
422. Glaziers Lead Window Makers	2.179 ^{***}	451. Furniture Brokers Dealers Minus Sub-Occode 823	1.741 ^{***}
	(15.22)		(12.54)
423. Architectural, Monumental--Carver, Sculptor	2.552 ^{***}	452. House & Shop Fittings Dealers Minus Sub-Occode 824	1.072 ^{***}
	(16.05)		(7.33)
424. Monumental Mason	2.713 ^{***}	453. Sawyer; Wood Cutting Machinist	-0.689 ^{***}
	(18.50)		(-5.02)
425. Plumber, Pipe Fitter	1.778 ^{***}	454. Lath, Wooden Fence, Hurdle-Maker	0.869 ^{***}
	(13.37)		(5.94)
426. Gasfitter	0.972 ^{***}	455. Wood Turner	0.970 ^{***}
	(7.14)		(7.12)
427. Locksmith, Bellhanger	1.586 ^{***}	456. Wooden Box, Packing Case--Maker	0.915 ^{***}
	(8.99)		(6.72)
428. Contractor: Public Works: Sewer Drainage Dock Harbour Canal	1.533 ^{***}	457. Cooper; Hoop--Maker, Bender	0.589 ^{***}
	(10.13)		(4.37)
429. Railway Contractors	0.0940	458. Cork, Bark-- Cutter, Worker	1.299 ^{***}
	(0.61)		(9.05)
430. Pond Reservoir Makers	2.269 ^{***}	459. Timber Wood Workers	-0.659 ^{***}
	(12.91)		(-4.12)
433. Well, Mine--Sinkers, Borer	0.223	460. Besom Maker	1.261
	(1.44)		(1.86)
434. Road--Contractor, Surveyor, Inspector	1.876 ^{***}	461. Other Workers In Wood	1.487 ^{***}
	(13.00)		(10.94)
435. Paviers	0.280	462. Timber Merchants Wood Dealers	3.283 ^{***}
	(1.92)		(24.12)
437. Cabinet Makers Minus Sub-Occode 821	1.652 ^{***}	463. Cork And Bark Dealers	3.308 ^{***}
	(12.40)		(20.18)
438. Furniture And Fittings - Others	1.571 ^{***}	464. Brick, Plain Tile, Terra-Cotta--Maker Minus Sub-Occode 825	-0.638 ^{***}
	(11.68)		(-4.71)
439. French Polisher	1.186 ^{***}	465. Plaster, Cement--Manufacture	-0.572 ^{***}
	(8.75)		(-3.80)
440. Upholsterer	1.737 ^{***}	466. Earthenware Figure And Image Makers	-0.497 ^{**}
	(12.96)		(-2.67)
441. Hot Water Apparatus Pipe Jointers	0.979 ^{***}	467. Earthenware, China, Porcelain--Mf Minus Sub-Occode 826	-0.988 ^{***}
	(6.36)		(-7.17)
442. House & Shop Fittings Makers	1.909 ^{***}	468. Sheet, Plate--Glass Manufacture	0.221
	(12.52)		(1.47)
443. Refrigerator Maker	0.945 ^{***}	469. Glass Bottle Manufacture	-0.160
	(6.34)		(-1.07)
444. Undertaker; Funeral Furniture Maker	2.817 ^{***}	470. Other Workers In Glass Mf Minus Sub-Occode 827	0.0431
	(20.13)		(0.31)
445. Wood Carver	2.487 ^{***}	471. Brick Dealers	2.274 ^{***}
	(17.95)		
446. Wood Gilder	1.996 ^{***}		
	(14.47)		
447. Willow Cane And Rush Workers	2.162 ^{***}		

	(13.20)		(14.28)
472. Plaster And Cement Dealers	3.512***	498. Oil And Colourmen Minus Sub-Occode 832	2.781***
	(14.58)		(20.34)
473. Earthenware, China, Glass-Dealer Minus Sub-Occode 828	1.662***	499. Soap Dealers	1.067***
	(11.45)		(4.53)
474. Dye And Paint Manufacture	0.279	500. Other Dealers In Chemicals Etc	1.884***
	(1.80)		(11.80)
475. Ink And Blacking Manufacture	1.252***	501. Floorcloth And Oilcloth Dealers	0.193
	(8.09)		(1.19)
476. Gunpowder, Guncotton, Explosive Substance--Mf	-1.133***	502. India Rubber And Gutta Percha Dealers	1.975***
	(-5.16)		(13.04)
477. Cartridge, Fireworks, Explosive Article--Manufacture	-0.889***	503. Waterproof Goods Dealers	1.570***
	(-4.29)		(5.47)
478. Lucifer Match Manufacture	-0.181	505. Skinners Furriers	0.944***
	(-0.99)		(6.79)
479. Salt Makers And Dealers	0.211	506. Tanners Fellmongers	0.132
	(1.42)		(0.95)
480. Manufacturing Chemist Minus Sub-Occode 829	-0.437**	507. Curriers	0.576***
	(-3.14)		(4.20)
481. Alkali Manufacture	-1.979***	508. Leather Goods Manufacture (Belts Etc)	0.705***
	(-8.93)		(5.06)
482. Chemists Druggists Minus Sub-Occode 830	2.271***	509. Parchment, Vellum - Maker	1.503***
	(17.01)		(11.05)
483. Drysalters	2.412***	510. Saddle Harness And Whip Makers	1.924***
	(17.08)		(14.43)
484. Oil Millers Oil Cake Makers	0.533***	511. Hair And Bristle Workers	1.089***
	(3.72)		(7.31)
485. Mineral Oil Worker	-1.356***	512. Brush And Broom Makers Minus Sub-Occode 833	0.951***
	(-5.70)		(6.96)
486. Tallow Chandlers Candle And Grease Manufacture	1.766***	513. Hair Plume Makers	1.364***
	(12.49)		(6.31)
487. Soap Boilers And Makers Minus Sub-Occode 831	-0.227	514. Quill And Feathers Dressers	2.136***
	(-1.36)		(14.12)
488. Manure Manufacture Bone Boilers	0.214	515. Dealers In Skins	2.124***
	(1.23)		(13.04)
489. India Rubber And Gutta Percha Makers	0.118	516. Dealers In Leather Minus Sub-Occode 834	1.723***
	(0.75)		(12.19)
490. Waterproof Goods Makers	0.623***	517. Quill And Feathers Dealers	2.372***
	(3.90)		(11.49)
491. Glue Size And Gelatine Manufacture	0.814***	518. Hair And Bristle Dealers	2.688***
	(3.74)		(13.10)
492. Other Workers In Chemicals Etc	1.517***	519. Paper Manufacture	-0.957***
	(8.63)		(-6.72)
493. Gum Manufacture	1.112***	520. Paper Stainer	0.640***
	(4.04)		(4.11)
494. Resin Manufacture	0.400	521. Pencil Makers	0.793***
	(0.97)		(3.45)
496. Varnish Maker	1.355***	522. Card, Stationery Makers	1.010***
	(8.14)		(6.70)
497. Oil Cake Dealers	2.176***	524. Envelope Manufacture	-0.267
			(-1.18)
		525. Paper Bag Makers	1.280***
			(8.80)
		526. Cardboard Box Maker	0.997***
			(8.80)

	(6.81)	551. Cotton & Cotton Good Mf Weaving Processes	-2.772***
527. Ticket Writers	3.226***		
	(22.95)		(-19.49)
528. Other Paper Workers	0.777***	552. Cotton & Cotton Good Mf Other Processes	-1.190***
	(4.09)		
529. Stationer (All Assumed Retail/Law Strs) Minus Sub-Occode 835	2.033***		(-6.39)
	(15.18)	553. Sheeting Manufacture	0.487*
530. Paper Dealers Minus Sub-Occode 836	2.705***		(1.99)
	(16.81)	554. Muslin Embroiderer	0.522*
531. Hand Compositors (Default)	-1.015***		(2.19)
	(-7.18)	555. Cotton & Cotton Goods Manufacture Undefined	-0.552***
532. Machine Compositors	-0.137		(-4.06)
	(-0.29)	556. Fustian Manufacture	-0.0694
533. Printing Machine Minders	-0.511**		(-0.46)
	(-2.72)	557. Woollen Cloth Manufacture Sorting	-1.253***
534. Stereotypers, Electrotypers	0.303		(-7.72)
	(1.48)	558. Woollen Cloth Manufacture Carding Combing	-1.247***
535. Music Printers	2.057***		(-7.73)
	(10.40)	559. Wool Spinners Wool Piecers	-0.893***
536. Printers So Described Minus Sub-Occode 837	1.325***		(-5.72)
	(9.88)	560. Worsted And Stuff Manufacture Spinners Piecers	-0.0374
537. Others In Printing	-1.398***		(-0.27)
	(-9.18)	561. Wool Winders Wool Warpens Wool Weavers	-1.549***
538. Paper Ruler	0.203		(-10.90)
	(1.05)	562. Worsted And Stuff Manufacture Winders Warpens Weavers	-2.819***
539. Lithographers	1.000***		(-15.56)
	(7.23)	563. Flannel Manufacture (Various) Minus Sub-Occode 841	0.625***
540. Copper Plate Printers	0.932***		(3.81)
	(5.54)	564. Blanket Manufacture (Various)	-0.749***
541. Map Makers	1.192***		(-4.35)
	(5.86)	565. Other Weaving Processes (Wool)	-0.231
542. Bookbinders	0.667***		(-1.22)
	(4.90)	566. Woollen Cloth Manufacture Other Processes	-0.687***
543. Publish, Bookseller & Library(Exc Public Libr) Minus Sub-Occy 838	1.905***		(-4.34)
	(14.03)	567. Tartan And Wincey Manufacture	-1.571
544. Music Publishers, Music Sellers	2.556***		(-1.57)
	(17.80)	568. Worsted And Stuff Manufacture Other Processes	-1.260***
545. Map Publishers, Map & Print Sellers	2.590***		(-4.75)
	(15.64)	569. Knitter (Woolen Articles)	0.820***
546. Newspaper Publishers Minus Sub-Occode 839	1.278***		(4.06)
	(7.54)	570. Crimpers	-0.986*
547. Newspaper Agents	3.752***		(-1.98)
	(27.58)	571. Woollen Cloth Mf Undefined Minus Sub-Occy 842	-1.088***
548. Cotton & Cotton Good Mf Card Blowing Room Procs	-2.607***		(-6.77)
	(-15.64)	572. Worsted & Stuff Mf Undefined Minus Sub-Occ 843	-2.210***
549. Cotton & Cotton Good Mf Spin Procs Minus Sub-Occy 840	-1.018***		(-10.18)
	(-7.47)	573. Fuller	-1.197***
550. Cotton & Cotton Good Mf Winding Warming Procs	-2.537***		
	(-16.70)		

	(-6.02)	(Various) Minus Sub-Occ 848	
574. Silk Workers - Spinners	-0.689 ^{***}		(15.51)
	(-4.47)	600. Shawl Makers	0.763 ^{***}
575. Crepe Gauze Manufacture (Various)	-2.129 ^{***}		(3.77)
	(-4.03)	601. Knitter (Undefined)	1.612 ^{***}
576. Silk Workers - Weaving Processes	-1.380 ^{***}		(10.94)
	(-9.54)	602. Embroiderers	2.191 ^{***}
577. Ribbon Manufacture	-0.0879		(15.19)
	(-0.50)	603. Curtain Makers	1.504 ^{***}
578. Silk Workers - Other Processes	-0.628 ^{**}		(7.62)
	(-3.22)	604. Tapestry Manufacture	-2.552 ^{***}
579. Silk Workers - Undefined Minus Sub-Occode 844	-1.806 ^{***}		(-3.54)
	(-8.91)	605. Weavers Sundry Fabrics	-1.575 ^{***}
580. Flax Linen And Damask Manufacture (Various)	-0.648 ^{***}		(-6.34)
	(-4.07)	606. Weavers (Undefined)	-0.931 ^{***}
581. Hemp Manufacture (Various)	0.145		(-6.54)
	(0.68)	607. Other Worker Sundry Fabric Undefined Minus Sub-Occ 849	-1.009 ^{***}
582. Jute Manufacture (Various)	-0.288		(-6.75)
	(-1.27)	609. Wool Woollen Goods Bleachers	-1.130
583. Cocoa Fibre Manufacture (Various)	-4.446 ^{***}		(-1.52)
	(-6.17)	610. Silk Bleachers	-2.505 ^{***}
584. Rope, Twine, Cord " Makers Minus Sub-Occode 845	0.791 ^{***}		(-5.37)
	(5.74)	611. Cotton & Calico Bleachers	-0.434 [*]
585. Mat Manufacture (Various)	0.825 ^{***}		(-2.32)
	(5.47)	612. Flax Linen Bleachers	0.574 ^{**}
586. Net Manufacture (Various)	0.750 ^{***}		(3.02)
	(4.48)	613. Undefined Textile Bleachers	-0.526 ^{**}
587. Canvas And Sailcloth Manufacture (Various)	0.481 [*]		(-3.39)
	(2.53)	614. Wool Woollen Goods Printers	-0.0949
588. Sacking Manufacture (Various)	0.624 ^{***}		(-0.49)
	(3.76)	615. Silk Printers	-0.867 [*]
589. Tent Maker	1.167 ^{***}		(-2.08)
	(6.25)	616. Cotton & Calico Printers	-0.259
591. Thread Manufacture (Various)	-1.457 ^{***}		(-1.70)
	(-5.34)	617. Undefined Textile Printers	-1.356 ^{***}
592. Hosiery Manufacture Minus Sub-Occode 846	0.0977		(-4.50)
	(0.72)	618. Wool Woollen Goods Dyers	-0.434 [*]
593. Lace Manufacture Minus Sub- Occode 847	0.807 ^{***}		(-2.41)
	(6.01)	619. Silk Dyers	-0.323
594. Elastic Web Makers	-0.362 [*]		(-1.77)
	(-2.02)	620. Cotton & Calico Dyers	-0.855 ^{***}
595. Felt Manufacture (Various)	-0.746 [*]		(-4.80)
	(-2.30)	622. Undefined Textile Dyers	0.424 ^{**}
596. Carpet Rug Manufacture (Various)	-0.376 ^{**}		(3.13)
	(-2.63)	623. Wool Woollen Goods Finishers	-0.751 ^{***}
597. Trimming Makers	0.946 ^{***}		(-4.00)
	(6.65)	624. Silk Finishers	-0.155
598. Tape Manufacture (Various)	-1.107 ^{***}		(-0.50)
	(-5.79)	625. Cotton & Calico Finishers	-0.970 ^{***}
599. Fancy Goods Textile Mf	2.183 ^{***}		(-4.69)
	(-2.63)	626. Flax Linen Finishers	0.816 ^{***}
	(-2.63)	627. Undefined Textile Finishers	-0.822 ^{***}
	(-2.63)		(-5.28)
	(-2.63)	628. Drapers Linen Drapers Mercers Minus Sub-Occode 850	2.041 ^{***}
	(-2.63)		(15.39)
	(-2.63)	629. Cotton Planter Grower	1.833 ^{**}
	(-2.63)		(2.67)

630. Woolstapler Wool Dealer Broker Minus Sub-Occ 851	1.131 ^{***} (8.04)	Merchants) Minus Sub-Occ 859	(16.99) 3.940 ^{***}
631. Cloth Worsted & Stuff Merchants Dlr's Minus Sub-Occ 852	0.154 (0.91)	656. Clothes Dealers	(26.62) 3.875 ^{***}
632. Yarn Merchant Minus Sub- Occode 853	1.246 ^{***} (8.28)	657. Dressmakers	(29.25) 1.457 ^{***}
633. Waste Merchant Minus Sub- Occode 854	2.091 ^{***} (9.00)	658. Stay And Corset Makers	(10.69) 2.056 ^{***}
634. Silk Merchants Dealers Minus Sub-Occode 855	1.458 ^{***} (9.64)	659. Shirtmakers And Seamstresses	(15.46) 0.598 ^{***}
635. Rope Twine Cord - Dealers	2.552 ^{***} (14.34)	660. Button Makers	(3.97) -0.638 ^{***}
636. Mat Sellers	1.749 ^{***} (5.54)	661. Glove Manufacture Minus Sub- Occode 860	(-4.35) 2.204 ^{***}
637. Other Fibrous Materials Dealers	4.357 ^{***} (15.93)	662. Hosiers Haberdashers Minus Sub-Occode 861	(16.35) 1.840 ^{***}
638. Cotton & Calico Dealers Minus Sub-Occode 856	0.882 ^{***} (6.01)	663. Shoe & Boot Maker (& Repairer) Minus Sub-Occode 862	(13.89) 1.930 ^{***}
639. Flax Linen Dealers	2.327 ^{***} (11.91)	664. Slipper Makers	(13.27) 2.392 ^{***}
640. Lace Dealers	1.934 ^{***} (12.57)	665. Patten, Clog Makers	(17.64) 1.141 ^{***}
641. Fustian Dealers	0.773 ^{**} (3.26)	666. Footwear Dealers Minus Sub- Occode 863	(8.22) 1.725 ^{***}
643. Manchester Goods Warehousemen	0.00701 (0.04)	667. Artificial Flower Makers	(12.23) 3.646 ^{***}
644. Dealers Sundry Fabrics	3.001 ^{***} (18.74)	668. Hairdressers	(27.37) 2.042 ^{***}
645. Straw Plait Manufacture	3.770 ^{***} (26.94)	669. Umbrella Parasol And Stick Makers	(15.06) 1.317 ^{***}
646. Straw Hat Manufacture	1.140 ^{***} (8.37)	670. Accoutrement Makers	(8.91) 0.806 ^{***}
647. Felt Hat Manufacture Wollen Bonnet Manufacture	-0.773 ^{***} (-4.70)	671. Glove Dealers	(3.94) 1.445 ^{***}
648. Cloth Hat Cap Manufacture	1.334 ^{***} (9.18)	673. Umbrella Parasol And Stick Dealers	(6.72) 2.572 ^{***}
649. Hat Cap (Not Cloth Felt Straw) Manufacture	-0.0782 (-0.53)	674. Other Dealers In Dress	(15.79) 3.662 ^{***}
650. Milliners (Not Retail - Workshop/Factory)	3.024 ^{***} (22.74)	675. Artificial Flower Dealers	(20.25) 1.704 ^{***}
651. Hatters Minus Sub-Occode 857	1.070 ^{***} (7.82)	676. Creamery Workers	(8.95) 2.272 ^{***}
652. Milliners (Retail)	1.478 ^{***} (9.80)	677. Cowkeepers, Milksellers Minus Sub-Occode 864	(17.07) 2.485 ^{***}
653. Tailors Not Merchants- Default Minus Sub-Occ 858	1.579 ^{***} (11.91)	678. Provision Curers	(17.69) 2.016 ^{***}
654. Others Clothing Manufacturers	-0.378 ^{**} (-2.68)	679. Cheesemongers Buttermen Minus Sub-Occode 865	(14.58) 3.358 ^{***}
655. Clothier & Outfitter (Dlr And	2.280 ^{***}	680. Provision Dealers	(25.04) 0.968 ^{***}
		681. Slaughterer	(6.48)

682. Butchers & Meat Salesmen Minus Sub-Occode 866	2.656 ^{***} (20.03)	707. Tobacconists Minus Sub- Occode 874	(3.90) 3.452 ^{***}
683. Fish Curers	2.082 ^{***} (14.51)	708. Maltsters	(25.70) 0.124
684. Poulterers Game Dealers Minus Sub-Occode 867	2.166 ^{***} (15.47)	709. Brewers	(0.90) 0.197
685. Fishmongers Minus Sub-Occode 868	3.194 ^{***} (23.93)	710. Distillers And Rectifiers	0.778 ^{***} (4.17)
686. Corn Millers	1.282 ^{***} (9.61)	711. Coffee & Eating-House Keepers Minus Sub-Occ 875	1.653 ^{***} (11.97)
687. Corn Flour & Seed Merchants Dlrs Minus Sub-Occ 869	1.054 ^{***} (7.50)	712. Lodging And Boarding-House Keepers	3.855 ^{***} (28.80)
688. Hay Straw And Chaff Cutters	0.948 ^{***} (5.25)	713. Innkeepers, Hotel Keepers And Publicans	3.564 ^{***} (26.84)
689. Hay Straw And Chaff Dealers	3.714 ^{***} (26.31)	714. Beersellers	3.755 ^{***} (27.99)
690. Biscuit Cake Bread Makers	1.421 ^{***} (10.10)	715. Beer Bottlers	-0.131 (-0.73)
691. Bakers (Dealers) Minus Sub- Occode 870	1.989 ^{***} (14.99)	716. Cellarmen	-1.891 ^{***} (-9.07)
692. Biscuit Cake Dealer (Confectionist) Minus Sub-Occ 871	2.995 ^{***} (22.53)	722. Wine And Spirit Merchants Minus Sub-Occode 876	1.267 ^{***} (9.08)
693. Sugar Refiners	0.927 ^{***} (6.42)	723. Gas Works Service	-1.757 ^{***} (-12.28)
694. Jam Preserve Sweet Makers	1.438 ^{***} (9.65)	724. Waterworks Service	-0.719 ^{***} (-4.86)
695. Chocolate Cocoa Makers	0.468 ^{**} (3.09)	725. Other Electricity Supply	-0.617 ^{**} (-2.95)
696. Coffee Tea Planter Grower	2.015 ^{***} (11.22)	726. Electricity Generation Distribution	-2.332 ^{***} (-8.16)
697. Grocers Tea Dealers Minus Sub- Occode 872	0.813 ^{***} (6.11)	727. Town Drainage	-1.216 ^{***} (-5.75)
698. Greengrocers Fruiterers Potato Dealers	3.763 ^{***} (28.31)	728. Scavenging, Street Cleaners, Crossing Sweepers, Dustmen	1.015 ^{***} (7.25)
699. Ginger Beer Mineral Water Manufacture Others	2.366 ^{***} (17.41)	729. Circular Envelope Addressers	1.526 [*] (2.41)
700. Mustard Vinegar Spice Pickle Makers	1.285 ^{***} (8.03)	730. Advertising Agents Minus Sub- Occode 877	2.268 ^{***} (14.92)
701. Ice Makers	-1.129 ^{**} (-2.89)	731. Billstickers	1.765 ^{***} (12.41)
702. Sugar Planter Grower	0.871 (1.88)	732. Sandwichmen, Bill Distributors	0.795 ^{***} (3.75)
703. Others Dealing In Food	3.348 ^{***} (23.08)	733. Cattle Sheep Etc Salesmen	4.034 ^{***} (29.07)
704. Hop Merchants Dealers Minus Sub-Occode 873	1.531 ^{***} (8.19)	734. Drovers	1.012 ^{***} (6.84)
705. Ice Dealers Importers Vendors	4.846 ^{***} (9.71)	735. Others Engaged Animals (Not Agric) Minus Sub-Occ 878	1.631 ^{***} (8.59)
706. Tobacco Manufacture	0.544 ^{***}	736. Knacker Catsmeat Dealers	3.934 ^{***} (25.15)

737. Celluloid Makers, Workers	0.145 (0.61)	762. Contractors - Public Works Etc	3.283 ^{***} (23.71)
738. Tobacco Pipe And Snuff Box Manufacture	1.388 ^{***} (8.79)	763. News Boys Vendors	3.304 ^{***} (18.88)
739. Bone Horn Ivory Tortoise-Shell Workers	1.382 ^{***} (9.26)	767. Draughtsmen (Undefined)	0.682 ^{***} (4.39)
740. Comb Makers	0.854 ^{***} (4.71)	768. Artisans Mechanics	-0.792 ^{***} (-5.74)
741. Floorcloth And Oilcloth Manufacture	-0.472 [*] (-2.04)	771. Machinists Machine Workers Undefined	-1.455 ^{***} (-9.83)
742. Japanners	0.397 [*] (2.46)	777. Owners Of Companies	1.495 ^{***} (6.85)
743. Chimney Sweepers	3.825 ^{***} (28.14)	800. Dental E/OA Occode 43	4.164 ^{***} (23.90)
744. Rag Dealers Minus Sub-Occode 879	0.192 (1.28)	801. Subordinate Medical Service E/OA Occode49	4.903 ^{***} (26.48)
745. Organ Grinders	2.424 ^{***} (15.22)	802. Schoolmasters And Teachers E/OA Occode52	4.086 ^{***} (29.19)
746. Animal, Bird Preservers Taxidermists	3.470 ^{***} (22.34)	803. Authors Eds Journalists & Creative Advert Workr E/OA Occ 56	5.214 ^{***} (20.45)
747. Figure And Image Makers Not Earthenware	1.961 ^{***} (11.78)	804. Musician, Music Master, Singer (Not Teachers) E/OA Occode 76	5.481 ^{***} (35.18)
748. Gut Manufacture	2.002 ^{***} (7.64)	805. Laundry Work: Washer, Iron Mangle (Not Dom) E/OA Occ 105	4.256 ^{***} (26.99)
749. Charcoal Burners	0.413 (1.94)	806. Broker Factor Comm Agent (Not - Mine Quarry Ins) E/OA Occ 112	3.520 ^{***} (25.82)
750. Other Workers In Sundry Industries	0.512 ^{***} (3.55)	807. Surveyors (House Agents) E/OA Occode 117	3.337 ^{***} (11.73)
751. Dealers In Sundry Materials	3.462 ^{***} (25.08)	808. Ag Machine--Proprietor, Attendant E/OA Occode 186	4.846 ^{***} (11.89)
753. Sponge Dealer	3.277 ^{***} (11.61)	809. Mine - Owner, Agent, Manager E/OA Occode 202	3.814 ^{***} (21.13)
754. Charcoal Dealers	2.149 ^{***} (5.29)	810. Pig Iron Manufacture (Blast Furnace) E/OA Occode 241	4.303 ^{***} (12.91)
755. Water Dealers	1.433 ^{***} (6.92)	811. Steel--Manufacture, Smelting, Founding E/OA Occode 244	3.342 ^{***} (15.35)
756. Receiving Shop, Receiving Office, Keepers, Assistants	0.476 ^{**} (2.61)	812. Blacksmiths E/OA Occode 262	4.599 ^{***} (23.68)
757. Multiple Shop Keepers	0.283 (0.65)	813. File Makers E/OA Occode 292	4.037 ^{***} (15.94)
758. General Shopkeeper Dealers Minus Sub-Occode 880	0.962 ^{***} (7.11)	814. Cutlery And Scissor Makers E/OA Occode 294	4.993 ^{***} (19.32)
759. Pawnbrokers Minus Sub-Occode 881	0.428 ^{**} (2.97)	815. Lock, Key Makers E/OA Occode 315	4.736 ^{***} (17.99)
760. Hawkers Hucksters Costers	4.926 ^{***} (36.94)	816. Bicycle Makers, Bicycle	4.983 ^{***}
761. Mfrs, Mngers, Superintdnts, Foremen (Unspec) Minus Sub-Occ 882	0.196 (1.43)		

Repairers E/OA Occode 362	(22.86)	837. Printers So Described E/OA Occode 536	(21.37) 2.625***
817. Ironmonger; Hardware-Dlr, Merchant E/OA Occode 369	4.501***		(19.23) 3.640***
818. Metal Dealers E/OA Occode 380	(29.00) 4.882***	838. Publish, Bookseller & Library(Exc Public Libr) E/OA Occode 543	(24.21) 4.365***
819. Watch And Clock Makers E/OA Occode 386	(17.63) 3.882***	839. Newspaper Publishers E/OA Occode 546	(20.92) 3.801***
820. Carpenter, Joiner E/OA Occode 409	(27.60) 3.653***	840. Cotton & Cotton Goods Mf Spinning E/OA Occode 549	(14.05) 4.366***
821. Cabinet Makers E/OA Occode 437	(26.61) 4.633***	841. Flannel Manufacture (Various) E/OA Occode 563	(13.28) 3.757***
822. Dealer In Works Of Art E/OA Occode 450	(31.42) 4.602***	842. Woollen Cloth Mf Undefined E/OA Occode 571	(24.84) 3.814***
823. Furniture Brokers Dealers E/OA Occode 451	(21.54) 4.808***	843. Worsted And Stuff Mf Undefined E/OA Occode 572	(19.42) 3.694***
824. House & Shop Fittings Dealers E/OA Occode 452	(33.56) 3.350***	844. Silk Workers - Undefined E/OA Occode 579	(16.94) 4.132***
825. Brick, Plain Tile, Terra-Cotta-- Maker E/OA Occode 464	(21.99) 3.865***	845. Rope, Twine, Cord " Makers E/OA Occode 584	(20.15) 4.467***
826. Earthenware, China, Porcelain-- Mf E/OA Occode 467	(22.70) 4.416***	846. Hosiery Manufacture E/OA Occode 592	(23.98) 4.019***
827. Other Workers In Glass Manufacture E/OA Occode 470	(23.52) 4.250***	847. Lace Manufacture E/OA Occode 593	(24.86) 4.629***
828. Earthenware, China, Glass-- Dealer E/OA Occode 473	(15.85) 4.224***	848. Fancy Goods Textile Mf (Various) E/OA Occode 599	(20.46) 4.134***
829. Manufacturing Chemist E/OA Occode 480	(30.31) 3.581***	849. Other Work Sundry Fabrics Undefined E/OA Occode 607	(18.00) 4.598***
830. Chemists Druggists E/OA Occode 482	(18.14) 3.994***	850. Drapers Linen Drapers Mercers E/OA Occode 628	(30.54) 4.175***
831. Soap Boilers And Makers E/OA Occode 487	(26.96) 3.497***	851. Woolstapler Wool Dealer Broker E/OA Occode 630	(26.09) 3.670***
832. Oil And Colourmen E/OA Occode 498	(15.30) 5.119***	852. Cloth Worsted & Stuff Merchants Dlrs E/OA Occode 631	(19.84) 4.138***
833. Brush And Broom Makers E/OA Occode 512	(27.20) 4.150***	853. Yarn Merchant E/OA Occode 632	(16.45) 3.638***
834. Dealers In Leather E/OA Occode 516	(22.76) 4.176***	854. Waste Merchant E/OA Occode 633	(19.14) 3.750***
835. Stationers (All Assumed Retail/Law Strs) E/OA Occ 529	(26.42) 3.896***	855. Silk Merchants Dealers E/OA Occode 634	(15.35) 3.791***
836. Paper Dealers E/OA Occode 530	(26.31) 4.450***	856. Cotton & Calico Dealers E/OA	

Occode 638		691	
	(21.61)		(27.79)
857. Hatters E/OA Occode 651	4.332***	871. Biscuit Cake Dealers	4.534***
	(23.36)	(Confectionists) E/OA Occode 692	
858. Tailors (Not Merchants) -	3.876***		(25.96)
Default E/OA Occode 653		872. Grocers Tea Dealers E/OA	4.070***
	(28.33)	Occode 697	
859. Clothier & Outfitter (Dlrs	4.183***		(30.66)
& Merchants) E/OA Occode 655		873. Hop Merchants Dealers E/OA	3.995***
	(30.35)	Occode 704	
860. Glove Manufacture E/OA	4.464***	874. Tobacconists E/OA Occode 707	(14.31)
Occode 661			4.466***
	(14.31)	875. Coffee And Eating-House	(29.51)
861. Hosiers Haberdashers E/OA	4.714***	Keepers E/OA Occode 711	4.501***
Occode 662			(32.38)
	(20.48)	876. Wine And Spirit Merchants	3.613***
862. Shoe And Boot Makers (And	4.132***	E/OA Occode 722	
Repairers) E/OA Occode 663			(25.76)
	(29.86)	877. Advertising Agents E/OA	4.107***
863. Footwear Dealers E/OA Occode	3.968***	Occode 730	
666			(16.44)
	(26.70)	878. Others About Animals (Not	4.827***
864. Cowkeepers, Milksellers E/OA	4.266***	Agric) E/OA Occode 735	
Occode 677			(18.86)
	(31.02)	879. Rag Dealers E/OA Occode 744	4.663***
865. Cheesemongers Buttermen	4.727***		(28.91)
E/OA Occode 679		880. General Shopkeepers Dealers	4.218***
	(15.03)	E/OA Occode 758	
866. Butchers And Meat Salesmen	4.279***		(31.63)
E/OA Occode 682		881. Pawnbrokers E/OA Occode 759	4.204***
	(29.92)		(29.36)
867. Poulterers Game Dealers E/OA	4.315***	882. Mfrs, Mngers, Superintdnts,	2.832***
Occode 684		Foremen (Unspec) E/OA Occode 761	
	(28.56)		(19.16)
868. Fishmongers E/OA Occode 685	4.568***	Constant	-5.932***
	(32.77)		(-44.62)
869. Corn Flour And Seed Merchants	3.588***	Observations	7,109,988
Dealers E/OA Occode 687		Pseudo R^2	0.456
	(26.35)		
870. Bakers (Dealers) E/OA Occode	3.748***		