

Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881

Carry van Lieshout, Robert J. Bennett and Harry Smith

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

Working Paper 13:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Businesses

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

February 2019

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

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

Van Lieshout, Carry, Bennett, Robert J. and Smith, Harry (2019) *Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881*. Working Paper 13: ESRC project ES/M010953: 'Drivers of Entrepreneurship and Small Businesses', University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

Keywords: Entrepreneurship, Employers, Firm size, Farm employment, Census

JEL Codes: L26, L25, D13, D22

Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881

Carry van Lieshout, Robert J. Bennett and Harry Smith

Working Paper 13: Working paper series from ESRC project ES/M010953: Drivers of
Entrepreneurship and Small Businesses, University of Cambridge

1. Introduction

The household returns from the censuses between 1851 and 1881 collected information on the number of employees of each business proprietor and the acres of each farm. These censuses each followed a broadly similar structure and are the earliest large scale and most complete source of information on business size, farm employees and acreage. Unfortunately after 1881 this question was dropped, so that 1851-81 provides the only large-scale information on business proprietors by firm size. The 1851-81 information provided is contained in the original Census Enumerators Books (CEBs). These were part of the census process undertaken by the General Register Office (GRO) to count the population. The GRO itself undertook some analysis of the data on proprietors and farm acreage, but this was very limited. As a result most of the information has remained inaccessible for large-scale analysis until electronic versions of the census records have become available.

This Working Paper uses the data that can now be extracted from the electronic records and compares them with the tables published by the GRO. This serves two purposes: first, it provides a check on the quality of the data extracted, and second, it allows a check on the quality of GRO analysis at the time of the census. This is undertaken for three census years: 1851-71. Unfortunately, because the 1881 census reports by GRO gives no analysis of the equivalent census questions, there can be no equivalent assessment for that census year; although implications for 1881 are drawn in conclusion.

The process of identification and extraction of the data on proprietors and farm acreage from the original CEB household returns is described in Working Paper 3: *Identifying businesses and entrepreneurs in the Censuses 1851-1881*. Working Paper 2 outlines the different census questions and the challenges they present for identifying entrepreneurs. An overview of the research strategy and full data assembly process underpinning the database for entrepreneurs is given in Working Paper 1: *Drivers of Entrepreneurship and Small Businesses: Project overview and database documentation*. The final database is part of a UKDA data deposit: ‘British Business Census of Entrepreneurs, 1851-1911’. A full list of Working Papers is included at the end of this paper.

The database for entrepreneurs 1851-1911 referred to in this and other project Working Papers for ESRC project ES/M010953 *Drivers of Entrepreneurship and Small Businesses* is an amalgamation of several sources. The data referred to in this working paper for 1851-1881 is mainly derived from the Integrated Census Microdata (I-CeM) deposited at UK Data Archive (UKDA), which has been used in a revised and updated form.¹ The I-CeM records are derived from the transcriptions made by the commercial genealogy provider Find My Past (FMP) (part of BrightSolid) in conjunction with The National Archive (TNA).

However, because of major gaps in the required records for employers and other entrepreneurs in FMP and hence in I-CeM, additional data have been extracted to obtain the full records required to satisfy the target of a complete and consistent database. The additional material is derived from three sources: first, *for 1851 records* missing or truncated in FMP and I-CeM have been supplied by S&N Genealogy Supplies (TheGenealogist.co.uk) (hereafter: S&N); second, *for 1861 records* truncated in FMP and I-CeM have been obtained by direct inspection of the original census manuscript pages; third, *for 1871* there is no current version of I-CeM, and when it is available it will not contain any occupational information since this is not available from the FMP source of transcriptions – as a result an entirely separate source of 1871 transcriptions has been obtained entirely from S&N. The final data extractions are referred to as the I-CeM/S&N data.

¹ Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481, derived by FindMyPast using a variety of original FMP transcriptions. Version 2 of I-CeM includes a range of valuable additional inputs from colleagues at Campop; see Schürer, K., Higgs, E., Reid, A.M., Garrett, E.M., (2016) *Integrated Census Microdata V.2 (I-CeM.2)*.

The methods of data derivation from the original CEBs, and method of extraction both contain possible imperfections. Hence it is important to assess how far the extracted records match those analysed by GRO in their published reports. However, the GRO itself, which relied on a method of clerical extraction, was also subject to errors. Hence there are two issues which this working paper has to try and disentangle: the accuracy of data extractions now possible from the electronic records, and the accuracy of GRO analysis in its publications. However, the analysis that the GRO performed was unfortunately limited. The most extensive work was done on the results of the 1851 census, but this work was painstaking and expensive, and no similar analysis was performed on the other early censuses that contained the question on proprietors. Some limited analysis on farmers only was performed on the 1861 and 1871 censuses, but none at all for the 1881 census. While the published tables have been used to analyse or contextualise business size, particularly relating to farming, previous assessments of the published tables, and some case studies of limited areas, have raised questions about the reliability of the data.² This paper extends the previous assessments using the new data available from the full electronic records derived from the CEBs.

The Working Paper starts with comparisons with the 1851 census for farmers (Section 2). In Section 3 it extends the 1851 analysis to non-farmers. Sections 4 and 5 assess the GRO tabulations for 1861 and 1871, respectively, which are available for farmers only. Finally, Section 6 assesses the overall quality of the extracted data compared to the published tables and the analysis that can be based on them. The Working Paper demonstrates that although there are some deficiencies for the extractions possible from the I-CeM/S&N data, these can be mostly overcome for analysis purposes, and in many cases the new extractions are less ambiguous in definitions and more complete in some respects than achieved by GRO published tables, especially for larger firms and larger farms.

² E.g. Mills, D.R., 'Trouble with farms at the Census Office: an evaluation of farm statistics from the censuses of 1851-1881 in England and Wales', *The Agricultural History Review*, 47 (1999), pp. 58-77. Leigh Shaw-Taylor refutes some of his claims, and considers the published data as reliable but with certain caveats. Shaw-Taylor, L., 'Family Farms and Capitalist Farms in Mid Nineteenth-Century England', *The Agricultural History Review*, 53 (2005), pp. 158-191.

2. The 1851 extractions compared with the published figures: Farmers

As noted above, as a result of data problems, the 1851 extraction was based on several sources. The majority of extracted data has been derived from I-CeM, however, as there were gaps in the data, there has been substitution from S&N. The S&N supplementation was focused on ensuring there was a complete extraction of the employers with employees, and did not supplement entrepreneurs who returned no employees. This has implications for the completeness of the extracted data, which particularly affects farmers.

2.1 *The available data from the GRO*

In their 1851 report, the GRO compiled detailed tables of the number of farmers by acreage and employees as reported in the census, which can be used as a check on the extractions. These tables provide the number of labourers employed and the size of the farm holding by number of acres, available as a total for England and Wales as well as on division and county level. The tables included people who listed farming as a second or third occupation in addition to what they considered their main occupation, but excluded retired farmers. In contrast to the tables in later years, where the GRO stated the number of farmers who did not return employees or acreages for each county, in 1851 a footnote with the tables indicated that the difference between the acres/labourers tables and the general Occupation Tables was partly accounted for by farmers who returned no acres or labourers, and partly by the different inclusions of the retired and those with multiple occupations.³ The general Occupation Tables provided the total numbers of each occupation by county (as well as the England and Wales and Division totals), by gender, and by age. These tables do include retired farmers, but since they listed everyone under their main occupation, people who farmed as a secondary profession were not included in this total.

³ The footnote under the tables showing the farmers with their labourers and acres reads: ‘The number of Farmers appearing in this Table will not agree with the number as returned in the Occupation Tables, the difference being caused partly by the omission of, and partly by the fact that in the Occupation Tables *retired* Farmers are referred to their former occupation. On the other hand, a certain number of persons who, besides being engaged in farming, carried on some other business, are here included amongst “Farmers”, while in the Tables of Occupation they are referred to that other business, when it appeared to be their chief pursuit.’ 1851 Census of Great Britain, Population Tables II. Ages, Civil Condition, occupations and Birth-place of the People, Vol. I, *Parliamentary Papers*, LXXXVIII (1852-3), cclxxxii.

In order to be able to compare these tables, therefore, it is necessary to know how many farmers were retired or farmed as a secondary occupation. The extracted data used here excludes retired farmers (who were assumed inactive) unless they listed employees (where they were assumed to be de facto the lead operative even if stating retired), but the raw I-CeM data contained 10,835 farmers who were retired. Entrepreneurs with multiple occupations were extracted as part of the identification of portfolios, and the I-CeM/S&N database includes 12,190 people who were farming as their non-primary occupation in 1851. These numbers are roughly similar and make up 4-5% of the total, and as a result, the different inclusions in the GRO tables approximately cancel each other out, assuming that both retired and secondary farming is evenly distributed among the entrepreneurial types. This means that the remaining difference between the Occupation and Farmer tables is accounted for by the farmers who had neither labourers nor acres. Comparing both sets of tables for each county therefore allows a breakdown of farmers with employees, farmers with acres but no employees, and farmers with neither. For the purposes of the entrepreneurship project, these farmers have been classified as Employer (E), Own Account (OA), and Worker farmers (W), respectively. Table 1 shows the breakdown on county level, while Figure 1 maps the geographical spread of areas where employer or own account farming was more prevalent.

Division	County	Has employees (E)	Has acres only (OA)	Neither employees nor acres (W)	Total Farmers	% E	% OA	% W
I	LONDON	219	53	495	767	29	7	65
II	SURREY	1522	366	53	1941	78	19	3
II	KENT	3742	948	547	5237	71	18	10
II	SUSSEX	3153	719	343	4215	75	17	8
II	HAMPSHIRE	2485	595	265	3345	74	18	8
II	BERKSHIRE	1655	184	327	2166	76	8	15
III	MIDDLESEX	683	131	28	842	81	16	3
III	HERTFORDSHIRE	1561	182	99	1842	85	10	5
III	BUCKINGHAMSHIRE	1598	212	223	2033	79	10	11
III	OXFORDSHIRE	1939	295	234	2468	79	12	9
III	NORTHAMPTONSHIRE	2438	382	209	3029	80	13	7
III	HUNTINGDONSHIRE	764	189	113	1066	72	18	11
III	BEDFORDSHIRE	1288	161	80	1529	84	11	5
III	CAMBRIDGESHIRE	2220	1071	330	3621	61	30	9
IV	ESSEX	3670	446	305	4421	83	10	7

IV	SUFFOLK	4343	749	549	5641	77	13	10
IV	NORFOLK	4868	1664	434	6966	70	24	6
V	WILTSHIRE	2563	517	402	3482	74	15	12
V	DORSET	1860	471	685	3016	62	16	23
V	DEVON	7414	3013	1537	11964	62	25	13
V	CORNWALL	4063	3154	848	8065	50	39	11
V	SOMERSET	5254	1936	1463	8653	61	22	17
VI	GLOUCESTERSHIRE	2640	946	653	4239	62	22	15
VI	HEREFORDSHIRE	1684	850	396	2930	57	29	14
VI	SHROPSHIRE	3357	1578	480	5415	62	29	9
VI	STAFFORDSHIRE	3128	2898	432	6458	48	45	7
VI	WORCESTERSHIRE	1985	859	349	3193	62	27	11
VI	WARWICKSHIRE	2467	949	612	4028	61	24	15
VII	LEICESTERSHIRE	2507	1237	93	3837	65	32	2
VII	RUTLAND	412	314	47	773	53	41	6
VII	LINCOLNSHIRE	5922	5095	31	11048	54	46	0
VII	NOTTINGHAMSHIRE	2526	1867	-67	4326	58	43	-2
VII	DERBYSHIRE	1834	2955	961	5750	32	51	17
VIII	CHESHIRE	3337	3326	581	7244	46	46	8
VIII	LANCASHIRE	5865	9585	2351	17801	33	54	13
IX	YORKSHIRE WEST RIDING	6602	10783	920	18305	36	59	5
IX	YORKSHIRE EAST RIDING	2683	1451	420	4554	59	32	9
IX	YORKSHIRE NORTH RIDING	3535	3272	647	7454	47	44	9
X	DURHAM	1839	1586	645	4070	45	39	16
X	NORTHUMBERLAND	1875	946	458	3279	57	29	14
X	CUMBERLAND	2517	2444	305	5266	48	46	6
X	WESTMORLAND	939	1425	185	2549	37	56	7
XI	MONMOUTHSHIRE	1276	1118	363	2757	46	41	13
XI	Total South Wales	7581	10656	2308	20545	37	52	11
XI	Total North Wales	7807	8120	1307	17234	45	47	8
	TOTAL E&W	133620	91698	24046	249364	54	37	10

Table 1. Published numbers of farmers by county, 1851

Note: Total Farmers is the total of Male Farmers, Female Farmers, and Male Graziers in the Occupations Table by county. (There were no Female Graziers). ‘Has acres only’ (OA) is derived from the Farmers employees/acres tables, and is the total of farmers with ‘no men employed or number not stated’. ‘Has employees’ (E) is calculated as the difference between the total in the Farmers employees/acres table, and the total of OA. ‘Neither employees nor acres’ (W) is the difference between total farmers from the Occupations Table, and the total from the employees/acres table. The only county where these numbers did not add up correctly was Nottinghamshire. In addition, the addition of totals for all counties is slightly below the total in the GRO England and Wales table.

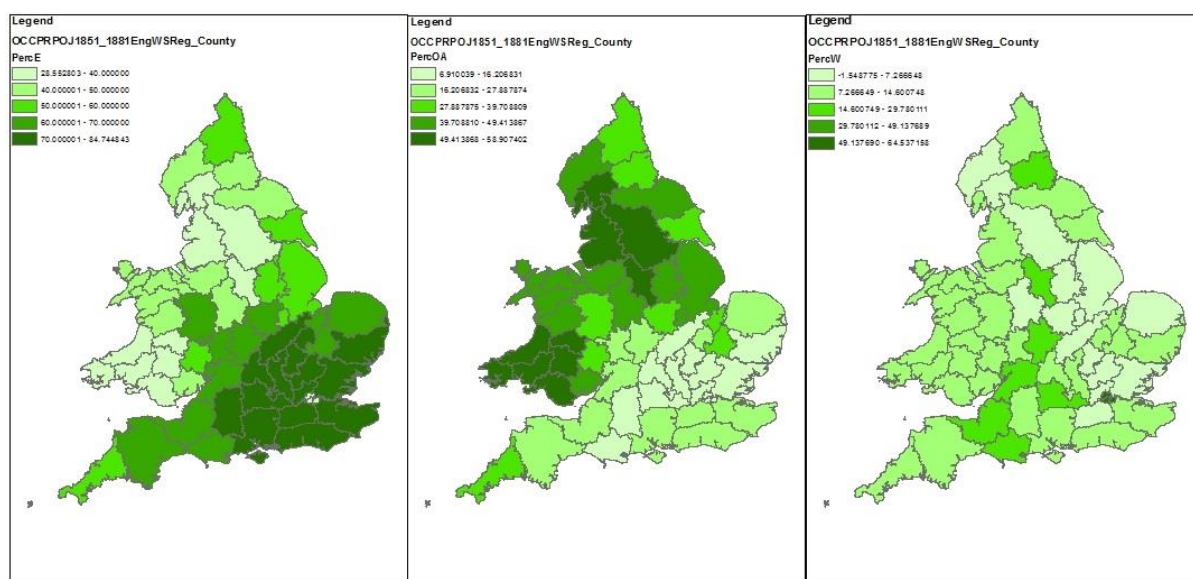


Figure 1. From left to right, percentage of Employer, Own account, and Worker farmers by county, based on published figures, 1851.

It appears that the GRO has some misgivings about what exactly they were measuring. The report stated that there was some uncertainty whether farmers returned all their in-door farm servants correctly, and it was recognised that in some cases household heads included women and boys in their counts but others did not, probably as a result of the gendered phrasing of the census question, which in 1851 asked for the return of ‘labourers’. Later years asked specifically about ‘men and boys’ (1861) or ‘men, women, and boys’ (1871 and 1881). In addition, the GRO believed that to ‘obtain the total number of persons who are employed on the farm, the farmer himself must be added, and frequently the farmer’s sons at home.’⁴ It notes that the 91,698 persons who called themselves farmers but who had apparently no labourers were probably either doing the manual labour themselves, did the labour with the assistance of their children, or only employed labourers for part of the year (and employed none at the census date). Finally, some of these would have had labourers but did not return them. In addition, there may have been a geographical component to the definition of farmer, as the report states that ‘in parts of the country, men who employ no workmen and have only a few acres of land, have always been called and returned at the Censuses as “Farmers”.’⁵

⁴ 1851 Census England & Wales, Population Tables II, Vol. I, p. lxxviii

⁵ 1851 Census England & Wales, Population Tables II, Vol. I, pp. lxxviii-lxxix

2.2 Comparing the I-CeM/S&N extractions with the published figures

Since the project extracted farmers as an occupational group in its entirety, the total numbers are comparable. Indeed, the I-CeM/S&N extractions hold up very well against the published Occupation Tables: the England and Wales published table lists 226,515 male farmers (including graziers), and 22,916 female farmers, totalling 249,431 people who had farming as the main occupation. The I-CeM/S&N extractions number 238,416 people who had farming as their main occupation, of which 24,494 were female and 212,995 male (the remainder are of unknown gender). The difference here is explained by the 10,835 retired or former farmers in the original I-CeM, who have been excluded from the extraction but were included in the published tables. This means the totals match almost exactly, although it shows that I-CeM/S&N probably includes some gender mis-codings.⁶

The published total of farmers returning acres and/or labourers for England and Wales is 225,318, of which 133,620 had employees, and 91,698 acres only. Using the same definition as the published tables, i.e. farming as either a main or secondary occupation, the I-CeM/S&N database only contains 183,048 farmers with acres and/or labourers, of which 118,322 were employers, and 64,726 returned acres only. This implies that the extractions missed about 15,000 employer farmers, and 26,000 farmers with acres only. The missing employers are likely to be a result of still incomplete transcriptions, even after using the S&N infill. The missing OA farmers, on the other hand, are partly due to incomplete transcriptions, and partly to the fact that the S&N infills were only concentrated on employers with stated employees; farmers giving only acreages (and no employees) were not extracted in the infills. Table 2 shows the county breakdown of the different types of farmers and comparisons with published. It shows that Sussex, Middlesex, Norfolk, Dorset, Gloucestershire, Warwickshire, Derbyshire, Cheshire, Lancashire, Northumberland, parts of Yorkshire and Wales had particularly poor extraction of OA farmers compared to published, even though all of these counties had significant S&N infills.

⁶ This was recognised at an early stage of the project, and resulted in manual CEB checks on female large employers, where the miscoding of a few women would distort analysis most. Many of those checked were indeed miscoded and have been corrected. However, to conduct CEB checks on over 20,000 female farmers would be out of the scope of this project.

		Published				Extracted I-CeM/S&N				S&N infill
Div	County	E	OA	W	Total	E	OA	W	Total	
I	LONDON	219	53	495	767	112	16	720	848	x
II	SURREY	1522	366	53	1941	1366	396	206	1968	
II	KENT	3742	948	547	5237	3488	836	1050	5374	x
II	SUSSEX	3153	719	343	4215	2696	48	1524	4268	x
II	HAMPSHIRE	2485	595	265	3345	2300	601	463	3364	x
II	BERKSHIRE	1655	184	327	2166	1665	206	199	2070	
III	MIDDLESEX	683	131	28	842	592	41	251	884	x
III	HERTFORDSHIRE	1561	182	99	1842	1509	217	136	1862	
III	BUCKINGHAMSHIRE	1598	212	223	2033	1451	239	188	1878	
III	OXFORDSHIRE	1939	295	234	2468	1912	231	321	2464	x
III	NORTHAMPTONSHIRE	2438	382	209	3029	2437	447	267	3151	
III	HUNTINGDONSHIRE	764	189	113	1066	685	165	256	1106	
III	BEDFORDSHIRE	1288	161	80	1529	1228	159	138	1525	
III	CAMBRIDGESHIRE	2220	1071	330	3621	2080	925	516	3521	x
IV	ESSEX	3670	446	305	4421	3179	422	350	3951	
IV	SUFFOLK	4343	749	549	5641	4279	874	397	5550	
IV	NORFOLK	4868	1664	434	6966	4217	120	3025	7362	x
V	WILTSHIRE	2563	517	402	3482	2439	486	406	3331	x
V	DORSET	1860	471	685	3016	1787	54	855	2696	x
V	DEVON	7414	3013	1537	11964	7255	3230	1042	11527	
V	CORNWALL	4063	3154	848	8065	3908	3207	829	7944	
V	SOMERSET	5254	1936	1463	8653	4256	1560	1955	7771	x
VI	GLOUCESTERSHIRE	2640	946	653	4239	2418	490	1248	4156	x
VI	HEREFORDSHIRE	1684	850	396	2930	1600	959	296	2855	

VI	SHROPSHIRE	3357	1578	480	5415	3074	1834	532	5440	
VI	STAFFORDSHIRE	3128	2898	432	6458	2951	3144	751	6846	
VI	WORCESTERSHIRE	1985	859	349	3193	1938	862	322	3122	
VI	WARWICKSHIRE	2467	949	612	4028	2217	254	1591	4062	x
VII	LEICESTERSHIRE	2507	1237	93	3837	2189	1024	845	4058	x
VII	RUTLAND	412	314	47	773	398	307	126	831	
VII	LINCOLNSHIRE	5922	5095	31	11048	5445	4134	1975	11554	x
VII	NOTTINGHAMSHIRE	2526	1867	-67	4326	2216	1691	358	4265	
VII	DERBYSHIRE	1834	2955	961	5750	1904	1678	2394	5976	x
VIII	CHESHIRE	3337	3326	581	7244	2605	126	4532	7263	x
VIII	LANCASHIRE	5865	9585	2351	17801	3164	523	15267	18954	x
IX	YORKSHIRE WEST RIDING	6602	10783	920	18305	5938	9581	3959	19478	x
IX	YORKSHIRE EAST RIDING	2683	1451	420	4554	2027	828	1438	4293	x
IX	YORKSHIRE NORTH RIDING	3535	3272	647	7454	3151	2701	1499	7351	x
X	DURHAM	1839	1586	645	4070	1603	1411	848	3862	x
X	NORTHUMBERLAND	1875	946	458	3279	1926	306	1080	3312	x
X	CUMBERLAND	2517	2444	305	5266	1917	2385	1178	5480	x
X	WESTMORLAND	939	1425	185	2549	762	1640	249	2651	
XI	MONMOUTHSHIRE	1276	1118	363	2757	1143	1224	438	2805	
XI	South Wales	7581	10656	2308	20545	6010	6827	7566	20403	x
XI	North Wales	7807	8120	1307	17234	6885	6317	3972	17174	x
	TOTAL E&W	133620	91698	24046	249364	118322	64726	67558	250606	

Table 2. Published and extracted farmers with employees and/or acres, 1851.

As a result of systematically missing some acres in the S&N data available, the extracted data cannot be compared against the published tables by acreage. A comparison with the employee numbers, although flawed, can be made, and sheds some light on the types of farmers that are missing from the extractions. Table 3 compares the number of farmers that returned a certain number of employees by division. Some divisions hold up better than others, with divisions VIII (Lancashire and Cheshire), IX (Yorkshire), and XI (Wales) being the worst. In all divisions however, the smallest farms are under-extracted by a much larger margin than larger ones, with the numbers of largest farms extracted usually being higher than the published numbers. Part of the over-count is due to the way the employees are attributed. This table included all employers who had farming as one of their businesses, but the total of employees has been taken. If, for instance, a colliery owner with 500 employees also described himself as a farmer, this table would include him with 500 employees as the database is rarely able to allocate the employees between the multiple occupations.⁷ The GRO clerks who compiled the tables by hand would have counted this person as a farmer with no employees (which was probably also inaccurate as a significant employer in another business would commonly have employees on their farm). If the same analysis is restricted to employers who returned employees as their main occupation only, the over-counts in the larger categories reduce, although they do not completely disappear. Some of the remaining over-counts are issues with the transcriptions. For example, the 'ing' in employing was often mis-transcribed as 7 or 9 and any '&' was often mis-transcribed as 3, 4 or 5. The data for the larger employer farmers were checked against original CEBs and cleaned down to 70 employees, but for employers below this size some errors of mis-transcription will remain. However, as further analysis shows, there were also large farmers who were missed by the GRO clerks.

⁷ This has been cleaned as much as possible in the final data deposit, which lists the occupation that employed most of the workers as main occupation as far as it is possible to determine. However, Table 3 includes farmers in first or additional occupation, and in the majority of cases it was not possible to separate the workforces.

	Pub	ICEM/SN	Pub	ICEM/SN	Pub	ICEM/SN	Pub	ICEM/SN	Pub	ICEM/SN	Pub	ICEM/SN
employees (not just labs)	I	I	II	II	III	III	IV	IV	V	V	VI	VI
1	25	7	1684	1545	1417	1397	1934	1758	5203	4830	3621	3334
2	26	17	1754	1613	1554	1503	1914	1700	4585	4176	3234	2868
3	25	10	1251	1100	1229	1143	1411	1219	2766	2541	2105	1884
4	23	14	1368	1219	1319	1240	1232	1163	2268	2103	1830	1682
5	17	10	828	734	864	787	899	782	1285	1191	962	862
6	9	1	742	687	906	863	752	692	1107	1007	809	810
7	6	5	505	445	632	588	520	492	636	586	524	513
8	9	5	597	554	633	614	548	504	583	549	485	486
9	8	2	413	364	464	442	443	384	358	331	254	232
10-	28	15	1622	1492	1783	1680	1673	1434	1121	1088	860	873
15-	7	6	707	681	764	748	549	613	478	457	267	284
20-	10	6	427	428	426	387	495	384	326	295	143	157
25-	3	4	245	229	185	185	135	193	137	143	58	64
30-	8	2	163	152	135	134	158	123	132	136	42	43
35-	3	3	68	72	53	42	43	64	42	59	22	31
40-	3	0	66	74	50	57	60	55	48	50	18	27
45-	2	1	28	31	18	26	27	30	14	21	3	5
50-	2	2	36	33	12	16	27	27	33	33	9	11
55-	0	0	8	12	19	10	17	13	10	13	4	5
60-	5	2	45	50	28	32	44	45	22	36	11	27

	Pub	ICEM/SN	pub	ICEM/SN	pub	ICEM/SN	Pub	ICEM/SN	Pub	ICEM/SN	Pub	ICEM/SN
employees (not just labs)	VII	VII	VIII	VIII	IX	IX	X	X	XI	XI	E&W	E&W
1	3944	3639	3377	2018	4511	3746	1988	1489	5860	4891	33564	28654
2	2896	2588	2497	1510	3155	2611	1793	1518	4541	3657	27949	23761
3	1826	1575	1300	820	1744	1502	1154	1022	2537	2051	17348	14867
4	1319	1231	862	538	1291	1130	853	759	1744	1493	14109	12572
5	675	617	399	268	667	577	343	314	683	637	7622	6779
6	622	570	278	165	452	420	283	262	489	455	6449	5932
7	300	301	146	93	239	207	126	104	215	206	3849	3540
8	322	301	93	73	196	201	134	134	206	197	3806	3618
9	161	165	53	40	107	113	66	64	96	108	2423	2245
10-	643	641	140	134	321	348	238	263	203	224	8632	8192
15-	226	229	29	28	67	85	86	96	41	39	3221	3266
20-	125	123	10	20	37	51	51	71	23	26	2073	1948
25-	40	49	5	13	14	19	17	19	11	13	850	931
30-	47	42	4	11	8	15	16	26	8	11	721	695
35-	16	19	3	6	0	6	6	11	0	7	256	320
40-	15	18	1	11	6	12	6	21	2	3	275	328
45-	9	9	1	2	2	6	1	9	1	2	106	142
50-	8	10	1	4	1	6	2	10	1	2	132	154
55-	4	2	1	1	1	4	1	2	0	1	65	63
60-	3	23	2	14	1	57	6	14	3	15	170	315

Table 3. Farmers with employees by division, published vs. I-CeM/S&N extractions, 1851.

A comparison with the non-farmers also raises questions about the quality of the data for smaller firms, as discussed in section 3. It remains to be seen, however, how much the missing 1851 data really matter in terms of analysis. First, as shown in Working Paper 9 on data reconstruction, good estimates for the actual numbers of E and OA farmers can be obtained irrespective of any small gaps in the data extraction. Second, comparisons with the GRO tables by farm size indicate that any biases are almost negligible. The GRO report stated that the 133,620 farmers reported a total of 665,651 labourer employees, or 5 labourers per farm on average. Two-thirds of these labourers were employed by 40,650 farmers, who employed 5 or more each, and 16,501 farmers had 10 or more labourers, employing a total of 311,707 labourers.⁸ The extractions resulted in only 118,322 farmer employers, employing a total of 639,484 labourers, which also averages out on 5 employees. There are 16,358 farmers with 10 or more employees, who together employed 325,003 labourers, and 38,468 farmers who had 5 or over employees, totalling 468,419 employees. These are all very similar to the GRO tables and result in similar averages to the GRO.

3. The 1851 extractions compared with the published figures: Non-Farmers

The extracted non-farmer employers can be compared to the employer tables published by GRO in the same way as for farmers. GRO published this table with the number of people employing a certain number of employees at division level. It was checked against this table of the original I-CeM extractions that led to the discovery of the missing employers' data, and hence use of S&N infill. As the infill was focused on fixing the employers with employees, it should have addressed the shortcomings of the data as identified in WP 3, and it does not suffer from a more restricted S&N extraction (which resulted in the missing acreage issue of the farmer analysis).

3.1. The available data from the GRO

Unfortunately comparisons with the GRO tables are made difficult because it is unclear how the GRO calculated their employer tables. The tables are titled 'Employers (with number of men)' and their header claims that the 'Table includes those persons *only*, who, in the

⁸ 1851 Census England & Wales, Population Tables II, Vol. I, p. lxxviii

Householders' Schedules, stated themselves to be *Masters*, in conformity with an instruction printed on every Schedule.⁹ The instruction was:

‘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.’¹⁰

The GRO recognised that many employers of workpeople had not adhered to the ‘master’ part of the instruction (‘and were in consequence, not distinguishable from journeymen’) even if they did return their workforce as per the second half of the instruction. The master-journeyman-apprentice system did not apply easily to employers of large factories or coal mines, and owners of these might not have considered themselves as being ‘masters’ and in a ‘trade’. Nevertheless, most still returned their workforce (for more discussion on this issue see WP 2). The GRO tabulations, however, seem to have aimed to stick to the full instruction only, as they claim the tables were ‘including all who described themselves as masters’, even though they recognised they were ‘far from complete’.¹¹

The GRO tabulated farmers and labourers separately, which implies they were not included in the employer tables. In addition, while the GRO text is not explicit about this, the title and instructions suggest that they only included the number of men returned, with separate tables for women and children returned either by masters or mistresses at the division level. However, the header also reads: ‘the facts are sufficiently numerous to enable deductions to be drawn as to the number of hands employed by masters’, and some of the division tables have footnotes for individual large employers, making clear they included people stating general descriptors such as hands (e.g. ‘One master employing 398 hands’ in division III).¹² It is unclear what happened to the other descriptors of employees (e.g. boys, males, journeymen, apprentice, or any other occupational title). Even the ‘masters only’ rule seems not to have been adhered to in a very strict manner: since, by comparing the listed large employers individually mentioned in the division tables’ footnotes to those extracted from I-

⁹ See e.g. Table XXX Employers (with number of men) in 1851 Census England & Wales, Population Tables I, Vol. II, pp. cclxxvi-cclxxix

¹⁰ ‘General Instruction’, Census of England and Wales, *Householder’s Schedule*, 1851.

¹¹ See e.g. Table XXX Employers (with number of men) in 1851 Census England & Wales, Population Tables I, Vol. II, pp. cclxxvi-cclxxix

¹² Employers Table, division III, in 1851 Census England & Wales, Population Tables I, Vol. II, p. 234

CeM/S&N, the only matches in a certain occupation, workforce size, and division revealed occupational strings without the word ‘master’.

3.2. Comparing the published data to the I-CeM/S&N extractions

The extracted employers with employees data were parsed into the different categories of employees; namely men, women, boys, girls, children, male (including son, or composite descriptor men & boys), female (including daughter, or composite descriptor women & girls), several occupational categories (labourers, journeymen, apprentices), and other (including other occupational descriptors, non-gendered descriptions such as ‘hands’, and composites such as ‘men, women, and children’). This allowed the data to be broken down in different ways. Table 4 shows the total non-farmer employers by division and number of employees, presented against the published figures, and compiled in four different ways. The first column includes non-farmer employers, whether they stated master or not, but only counting those who explicitly declared men. In the majority of the divisions, these numbers come pretty close to the published figures: with the exception of divisions I and VIII, most numbers are not more than 10% down, and in some cases over. The second column excludes the categories that were explicitly included in other sets of tables: labourers, women and children employees. This column only includes employers who declared their men, hands, and the ‘other’ category (which does include some women and children, but only in a phrasing where they could not be separated from the men). In most cases, this means the I-CeM/S&N extractions exceed the published totals, meaning that this either was not the way the GRO calculated their tables, or their calculations were not as good as our extractions. The third column includes all declared employees. The small differences between this and the previous column result from the number of women and children employees being relatively low, as are labourers in non-farmer occupations. These numbers show that the I-CeM/S&N extractions are higher everywhere than the GRO’s numbers, except in the London and the North Western Division (Lancashire and Cheshire), where large numbers of employers are missing. Finally, the table includes a calculation of what the GRO claimed to be doing in their table titles and headers: a computation of only masters who employed only men. The numbers here are dramatically low: the extractions only picked up between 15 to 50% of the published numbers, and almost none at all in the larger firm categories, which were often textile, steel or coal businesses that did not follow the master-apprentice system and whose owners tended

to return 'hands', 'workpeople' or a composite 'men, women and children'. However, as some of the examples of larger firms used by the GRO themselves in their footnotes clearly show, this was not how GRO tables were calculated.

After these comparisons it remains unclear what exactly the GRO was computing in the employer tables. It is certainly evident that the GRO did not fully follow the census instructions; nor did they fully follow the headings and footnotes of their published tables. As a result, it is very difficult to compare the I-CeM/S&N extractions to the published figures. It is clear, however, that there must be deficiencies in the I-CeM/S&N data for London and the North Western divisions, as these extractions are very low irrespective of how the published figures were calculated; nor do they match up well with the extracted patterns figures in other divisions. However, in other divisions it appears that the S&N/I-CeM extractions are often superior to the GRO clerical extractions. Indeed it is not surprising that GRO clerks missed some employer returns: they are often very small entries crammed into the space provided, overflowing into other parts of the CEB sheets (which results in split line issues which require additional treatment in our extractions: see WP 3); and it would be very easy for a GRO clerk who was turning many pages of the CEBs to miss the odd line where an employer appeared in a mass where no others occurred (all employers, particularly large employers, were rare).

	I. LONDON					II. SOUTH-EASTERN					III. SOUTH-MIDLAND				
NumberEmployees	Published	Men Only	NoLabWomenChildren	TotalEmployees	Masters Only	Published	Men Only	NoLabWomenChildren	TotalEmployees	Masters Only	Published	Men Only	NoLabWomenChildren	TotalEmployees	Masters Only
1	3182	1331	1660	1708	441	2664	2164	2620	2786	867	1746	1480	1807	1978	510
2	3092	1453	1644	1676	487	2190	1913	2178	2331	667	1256	1148	1300	1442	363
3	1922	926	1015	1036	291	1219	1084	1212	1324	354	698	655	716	795	205
4	1338	608	668	692	180	774	789	851	940	243	447	446	486	551	122
5	710	333	364	378	94	387	388	417	463	97	233	254	279	303	53
6	729	328	360	377	78	392	370	396	446	98	260	266	282	312	64
7	329	148	165	169	42	203	199	219	247	43	132	142	154	172	31
8	322	156	168	171	43	159	171	184	216	40	107	107	115	135	27
9	183	84	91	97	34	99	120	137	159	28	56	73	79	92	16
10-	985	502	577	595	122	484	540	619	710	113	299	341	391	473	79
20-	416	162	196	201	37	122	136	160	199	17	79	121	129	160	22
30-	183	76	95	98	15	58	54	68	79	13	48	50	59	68	7
40-	121	43	56	56	6	23	33	39	55	8	28	29	36	40	6
50-	100	48	62	63	7	35	41	54	65	8	30	27	37	44	4
75-	37	15	21	21	4	11	20	25	27	1	10	14	16	16	2
100-	39	20	27	27	3	12	14	19	21	5	9	8	12	14	0
150-	14	6	8	9	1	4	2	4	4	1	11	6	10	10	0
200-	10	5	7	7	1	3	1	1	1	0	2	2	2	2	0
250-	5	2	2	2	0	1	3	5	5	0	0	2	2	2	0
300-	5	3	5	5	1	2	3	4	4	0	3	0	1	1	0
350-	7	3	11	11	0	2	4	8	8	1	4	4	10	10	0

	IV. EASTERN					V. SOUTH-WESTERN					VI. SOUTHWEST-MIDLAND				
NumberE mployees	Publi shed	Men Only	NoLabWom enChildren	TotalE mplees	Mas ters Onl y	Publi shed	Men Only	NoLabWom enChildren	TotalE mplees	Mas ters Onl y	Publi shed	Men Only	NoLabWom enChildren	TotalE mplees	Mas ters Onl y
1	2331	1690	2102	2244	654	2115	1753	2407	2634	719	2536	2234	2813	3044	798
2	1638	1295	1482	1579	439	1837	1613	2071	2274	662	2026	1937	2302	2471	652
3	829	698	786	846	212	1124	1004	1171	1299	391	1235	1211	1370	1473	399
4	513	460	506	558	146	710	676	775	862	249	844	939	1016	1100	266
5	272	243	270	293	71	363	358	415	466	134	508	501	563	616	149
6	245	217	245	267	71	349	358	404	447	118	479	481	532	580	137
7	136	121	137	159	33	161	163	182	203	56	243	256	272	291	90
8	98	101	114	129	17	160	158	191	229	58	271	263	299	326	81
9	67	70	84	93	19	92	106	122	138	31	145	156	171	189	42
10-	288	305	354	420	68	462	461	569	645	156	935	846	991	1043	235
20-	63	91	103	124	22	117	100	127	160	24	283	242	301	322	66
30-	20	24	38	47	6	51	39	53	60	10	62	125	160	168	33
40-	10	9	14	19	0	23	26	36	41	2	42	72	101	103	20
50-	27	23	29	29	6	26	34	42	47	4	113	99	141	146	21
75-	7	9	15	15	3	15	10	15	21	2	53	47	75	77	10
100-	8	4	7	7	0	15	7	15	16	2	69	38	68	70	10
150-	3	1	3	3	0	13	4	10	10	0	26	30	44	44	5
200-	0	1	2	4	0	3	8	10	10	1	17	16	22	22	6
250-	1	1	1	1	0	2	1	2	3	0	13	7	12	12	2
300-	0	0	2	2	0	1	1	3	3	0	8	8	17	17	1
350-	5	5	8	8	1	9	10	20	21	0	28	18	37	37	6

	VII. NORTH-MIDLAND					VIII. NORTH-WESTERN					IX. YORKSHIRE				
NumberEmployees	Published	Men Only	NoLabWomenChildren	TotalEmployees	Masters Only	Published	Men Only	NoLabWomenChildren	TotalEmployees	Masters Only	Published	Men Only	NoLabWomenChildren	TotalEmployees	Masters Only
1	1899	1540	2022	2233	621	3216	1071	1498	1619	389	2079	1788	2590	2861	752
2	1246	1088	1343	1514	397	2682	1003	1228	1316	363	1249	1552	1983	2157	570
3	660	623	732	806	212	1626	660	755	800	230	725	1034	1217	1299	362
4	513	491	537	588	158	1264	495	557	597	167	513	760	865	954	248
5	299	264	294	318	89	676	278	304	320	92	265	417	468	512	124
6	260	271	301	337	77	769	307	339	352	97	325	486	547	588	149
7	151	142	163	175	50	425	152	175	189	52	132	260	284	298	69
8	140	105	124	139	28	455	185	204	216	69	135	229	259	270	68
9	81	73	81	84	28	222	88	110	117	24	86	155	170	178	37
10-	390	369	435	481	94	973	509	592	616	157	451	868	1018	1060	238
20-	122	109	141	151	29	388	209	243	245	60	183	303	368	374	85
30-	57	56	75	83	14	212	104	141	143	30	104	166	224	231	35
40-	35	34	42	45	10	145	66	87	90	19	54	86	128	133	19
50-	48	38	57	57	6	177	66	95	97	15	72	129	185	194	34
75-	28	30	40	40	5	143	45	66	67	10	26	79	106	109	14
100-	28	21	33	33	3	162	40	64	65	6	33	89	131	134	16
150-	14	10	20	20	2	106	26	52	52	3	24	45	54	54	8
200-	8	6	12	12	2	65	12	23	23	1	12	21	40	41	0
250-	3	4	6	6	1	50	12	16	16	2	10	19	26	27	3
300-	5	3	8	8	0	36	9	12	12	0	5	13	24	24	1
350-	11	8	19	19	2	141	27	60	60	1	18	63	98	98	3

	X. NORTHERN COUNTIES					XI. WALES				
NumberEmpl yees	Publish ed	MenOn ly	NoLabWomenChil dren	TotalEmpl ees	Maste rs Only	Publish ed	MenOn ly	NoLabWomenChil dren	TotalEmpl ees	Maste rs Only
1	1377	825	1332	1444	369	1200	986	1185	1342	1326
2	1125	824	1110	1178	325	902	706	803	896	884
3	693	527	669	728	189	446	406	444	487	483
4	409	348	401	439	140	299	263	278	314	310
5	233	201	222	243	69	162	139	152	168	167
6	213	200	227	241	79	157	144	154	166	162
7	123	93	104	115	30	58	55	60	69	67
8	118	86	99	108	20	79	65	71	78	73
9	71	59	65	72	14	38	50	54	60	60
10-	354	300	354	393	90	205	191	208	237	231
20-	101	86	106	112	25	55	47	53	60	59
30-	60	40	50	52	6	23	18	22	26	24
40-	22	29	34	35	10	11	15	16	18	17
50-	37	40	52	53	9	16	12	15	18	15
75-	13	15	20	20	3	5	7	9	9	9
100-	9	13	21	21	2	6	5	5	7	5
150-	20	9	15	15	1	1	4	4	6	4
200-	12	6	10	10	0	3	1	1	3	2
250-	3	2	3	3	0	0	1	1	1	1
300-	0	0	0	0	0	0	0	0	0	0
350-	3	5	7	7	1	0	0	0	0	0

Table 4. Non-farmers by business size (N of employees) by division, 1851. For each division shown are: the published total; employed men only; all employees except labourers, women, and children; total employees; and the employers who were masters only.

The GRO reported some further analysis on the non-farmers: the 87,270 masters employing one man or more had a total of 727,468 men in their employ, or an average of 8.33 men each.¹³ The I-CeM/S&N database counts 88,364 non-farmer employers (and only 26,937 ‘masters’ with employees), who employed a total of 870,370 people, giving an average of 9.85 employees, or 9.04 if calculated in the same manner as the GRO, which calculated the average within the larger employee-category ranges, not the actual employees numbers of each individual employer.

The overall comparisons indicate, therefore, that whilst the I-CeM/S&N database under-records the number of small employers, it gives higher and more accurate estimates of the larger employers than the GRO achieved. Finally, the GRO also computed the masters who did not return any employees, for which they counted 41,732. These can be extracted from I-CeM using a separate algorithm, and then cleaned for spurious masters such as station and postmasters (see WP 3). This means that the remaining masters in the I-CeM extractions should all be masters in a trade, just like the ones counted by the GRO (although no masters without employees were extracted from S&N). This number in I-CeM/S&N is 44,373 masters, implying that either the GRO definition of trade differs from the one used here (meaning they counted specific trades only, or considered more masters to be spurious) or they missed a number of masters. It is probably a mix of both explanations, although it is easier to miss occupational strings that include master but no employees than those which state employees, as these are usually substantially longer text strings. Nevertheless, overall it appears that the I-CeM/S&N database gives superior coverage to that achieved by GRO, except for those divisions where transcriptions remain deficient.

3.3. Assessing the 1851 extractions

The previous sections have pointed out the difficulties of comparing the extracted data to the published figures. In terms of assessing the accuracy of the extractions, they have proved to be less useful than initially hoped. However, they do shed some light on the quality and inherent biases of the extracted data.

¹³ 1851 Census England & Wales, Population Tables I, Vol. II, p. lxxviii.

Firstly, there are two divisions, London and the North West, that perform consistently worse on all firm sizes, and for farmers and non-farmers alike. This is likely related to the quality of transcriptions for these areas, as there are no significant gaps in the data preservation for 1851.

Second, the extractions underperform for estimates of smaller firm numbers. This is evident in both the farmers as well as the non-farmers. While this does not seem to affect farmer average firm size, it makes a difference in the non-farmer analysis. This is probably because the non-farm industries include more larger firms. However, data extraction issues are also mixed with respondent issues. All the CEB responses have ambiguity of how far the smallest employers fully responded to the master/employer instructions, and also how far enumerators fully recorded or administered this part of the census. There was census ambiguity about how family members (spouse, sons, daughters and others) should be included, and there was ambiguity about how respondents thought about their spouse and family, and indeed whether family were regarded as ‘occupied’ mainly in that employment or not. For this reason it is wise in subsequent use of the data to focus analysis on the categories that exclude employers with small firm sizes: either to exclude those with 1 employee, or those with 4 or 5 employees and under.

Third, for both farms and non-farm employers, the data extractions for large firms give a close fit with GRO tabulations in those areas which have complete transcriptions. Moreover, GRO appear to have missed a number of entries mainly for the larger firm and farm sizes. This means that for analysis of firm sizes, after excluding the smallest with only 1 employee, or those with 4 or 5 employees and under, the I-CeM/S&N data should give a very reliable estimate. However, for areas where transcriptions remain incomplete even after topping up from S&N (mainly London and the North West), subsequent analysis either needs to adopt a weighting process or focus on using proportions rather than absolute number counts.

4. The 1861 extractions

The 1861 data was derived from I-CeM, but with infill of transcription truncations by inspection of the original CEBs. Unfortunately, the GRO did not tabulate the employers with their employees at division level as they did in 1851, so full comparisons with published data

are impossible. However, they did tabulate some limited analysis for farmers. This was conducted on a group of 10 English counties: Buckinghamshire, Cambridgeshire, Cheshire, Cumberland, Lincolnshire, Norfolk, Shropshire, Sussex, Wiltshire, and the North Riding of Yorkshire.¹⁴ It is known that 3.7% of the 1861 records have not survived and have been lost since the GRO reports were published. As a result these are not included in I-CeM. However, comparisons of I-CeM with published data that are to be published in a subsequent working paper show that none of the 10 counties GRO tabulated was badly affected by data loss. In addition, for farmers the breakdown of entrepreneur types can be compared as well, since the GRO stated the number of ‘Worker Farmer’, i.e. those with no labourers or acreages, for every county. The comparisons with the extracted data are presented in Table 5. The main conclusion from this comparison is that, as for 1851, it is clear that the smaller farm sizes in particular have been under-recorded, while larger farms are more fully represented than GRO credited.

¹⁴ Census of England and Wales 1861: General Report, *Parliamentary Papers*, LIII (1863), 139-43.

	Buckinghamshire		Cambridgeshire		Cheshire		Cumberland		Lincolnshire	
labs	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM
1	123	52	583	226	1600	594	1351	390	2241	800
2	145	94	355	218	1224	629	983	344	1298	678
3	138	123	230	177	728	453	544	242	746	511
4	135	128	138	127	446	306	258	112	473	387
5	112	100	146	109	251	189	114	66	296	261
6	102	90	113	107	164	139	59	44	290	263
7	90	89	116	98	88	84	20	21	203	190
8	108	100	95	83	60	61	22	26	195	176
9	64	58	78	66	35	34	10	14	143	127
10-	277	248	270	238	75	94	22	24	468	406
15-	124	98	117	106	12	17	8	6	186	163
20-	43	37	79	89	7	13	1	6	118	106
25-	24	20	57	45	1	5	0	1	56	61
30-	12	10	45	40	1	5	0	0	33	27
35-	5	2	16	15	2	4	0	2	15	19
40-	2	3	12	17	0	1	0	0	12	13
45-	1	0	6	4	0	2	0	0	2	5
50-	0	0	4	2	0	0	0	0	3	4
55-	1	2	5	1	0	0	0	1	3	1
60+	2	4	10	13	0	2	0	2	10	12
Total E	1508	1258	2475	1781	4694	2632	3392	1301	6791	4210
OA	173	261	645	897	1488	2864	1091	2753	2061	3563
W	185	305	380	524	652	1239	470	763	1119	1717

Table 5

	Norfolk		Shropshire		Sussex		Wiltshire		Yorkshire E. Riding		Total	
labs	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM
1	970	469	996	452	482	240	356	204	1595	355	10297	3782
2	715	505	750	420	460	329	284	210	1176	307	7390	3734
3	480	413	561	368	341	249	236	193	709	259	4713	2988
4	383	337	408	287	276	255	183	156	341	244	3041	2339
5	306	273	321	241	222	204	155	140	191	167	2114	1750
6	233	224	207	182	175	143	151	143	148	157	1642	1492
7	198	156	147	111	128	111	102	92	71	84	1163	1036
8	181	158	123	118	129	119	113	104	64	81	1090	1026
9	143	141	66	62	100	89	76	74	31	63	746	728
10-	548	463	148	147	341	320	290	261	64	149	2503	2350
15-	251	231	41	44	145	145	172	155	13	48	1069	1013
20-	169	150	8	11	82	77	98	94	3	15	608	598
25-	102	78	1	8	55	45	61	63	1	13	358	339
30-	55	53	0	3	25	23	52	46	0	5	223	212
35-	39	34	0	0	17	16	27	29	1	1	122	122
40-	28	30	1	2	11	11	16	16	0	2	82	95
45-	20	22	0	1	6	4	10	14	0	5	45	57
50-	4	4	0	0	3	4	3	3	0	0	17	17
55-	2	2	0	0	2	1	2	5	0	2	15	15
60+	16	15	0	2	11	14	10	14	0	0	59	78
Total E	4843	3758	3778	2459	3011	2399	2397	2016	4408	1957	37297	23771
OA	1048	1427	850	1803	432	733	320	479	1786	1223	9894	16003
W	482	1015	796	1116	354	595	257	460	905	829	5600	8563

Table 5 (continued). Farmers with employees by county, published vs. I-CeM extractions, 1861.

	Buckinghamshire		Cambridgeshire		Cheshire		Cumberland		Lincolnshire	
Acres	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM
under5	24	20	89	68	246	202	63	41	200	140
5-	39	26	227	184	543	440	130	101	532	406
10-	77	81	387	353	936	821	271	252	1062	854
20-	65	66	246	241	644	605	236	237	757	709
30-	65	58	202	160	497	462	246	220	630	582
40-	78	77	169	154	412	377	305	283	502	460
50-	163	140	301	264	846	765	832	749	965	874
75-	101	88	184	137	529	436	496	451	565	490
100-	268	243	333	289	781	707	768	697	881	811
150-	183	160	194	165	368	285	417	367	568	489
200-	179	162	164	134	174	128	244	211	533	460
250-	134	114	121	97	86	73	116	98	375	298
300-	106	87	127	97	34	29	109	92	324	278
350-	50	40	78	61	7	9	33	39	163	147
400-	48	46	103	81	20	16	71	52	300	254
500-	26	23	62	48	9	9	29	34	162	139
600-	11	12	34	27	2	3	13	15	103	82
700-	7	6	24	19	0	0	16	14	51	43
800-	1	2	17	11	2	1	11	7	45	36
900-	4	3	8	8	0	4	6	5	19	14
1000-	1	1	10	5	0	1	7	2	30	24
1200-	1	1	8	5	0	0	3	4	16	11
1500-	1	1	1	1	0	0	4	3	8	6
2000 and up	0	0	0	1	0	0	3	2	5	6

Table 6.

Acres	Norfolk		Shropshire		Sussex		Wiltshire		Yorkshire E. Riding		Total	
	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM	Pub	ICEM
under5	178	158	60	54	63	55	59	49	104	46	1086	833
5-	413	360	212	168	144	120	88	74	328	127	2656	2006
10-	573	540	422	389	248	237	180	171	520	271	4676	3969
20-	441	442	272	279	204	196	151	138	418	234	3434	3147
30-	379	362	237	231	172	162	132	130	395	168	2955	2535
40-	322	274	192	180	177	165	121	112	343	148	2621	2230
50-	628	551	439	425	378	331	248	231	809	316	5609	4646
75-	382	322	341	299	207	179	172	158	553	206	3530	2766
100-	590	535	660	613	516	473	356	323	1052	433	6205	5124
150-	356	302	517	449	324	283	237	209	653	309	3817	3018
200-	322	288	446	406	242	218	187	181	424	277	2915	2465
250-	195	157	276	236	142	116	113	99	193	149	1751	1437
300-	221	183	227	203	157	136	110	103	137	118	1552	1326
350-	102	81	93	84	75	62	61	52	71	67	733	642
400-	164	143	102	84	102	86	104	90	82	93	1096	945
500-	132	110	40	30	66	63	86	75	28	64	640	595
600-	102	79	11	11	37	34	74	60	15	31	402	354
700-	44	37	6	4	33	29	50	45	12	15	243	212
800-	46	35	4	4	24	17	28	30	6	11	184	154
900-	24	16	0	0	23	18	24	19	5	4	113	91
1000-	30	25	1	1	24	16	38	29	6	6	147	110
1200-	20	8	0	0	15	14	25	20	2	4	90	67
1500-	17	14	0	0	1	0	11	8	2	0	45	33
2000 and up	7	4	0	0	2	2	4	3	5	3	26	21

Table 6 (continued): Farmers by acreage by county, published vs. I-CeM extractions, 1861.

The 1861 extractions also allow us to compare the acres; see Table 6. While many acres records are missing in I-CeM compared to GRO, the distribution here is more evenly across categories, with between 10-25% missing farmers in each acreage category. There is some variations between the counties, with Yorkshire East Riding in particular standing out; however, this is a county that performed worse on reported workforces as well. In general, the more even distribution across size and county suggests that the missing data is a transcription deficiency that was systematic in affecting a fairly standard proportion, but random between areas. It appears that a proportion of transcribers, or a proportion of each transcriber's effort, resulted in leaving out some farmers' attributes regardless of their acreage. However, the acreage comparisons are generally far better than the workforce comparisons. If it is indeed the case that the transcribers left out farmer attributes equally across the acreage size range then it is likely that the more systematic gap for workforces for smaller farms is less a result of transcriber deficiencies than other deficiencies: small farms in terms of workforce would be likely to be small in acreage as well. The main possible alternative explanation is that the GRO had a different way of calculating farm workforce size, potentially including family members who were counted in the smaller farms. It is unclear whether this is indeed what happened. While the 1861 report states that 'To obtain the total number of persons who are employed on the farm, the farmer himself must be added, and frequently the farmer's sons at home', but there is no definite indication for the tables that this is how they were computed.¹⁵

5. The 1871 extractions

The 1871 data was derived and extracted solely from S&N; it is not available in I-CeM. The extraction and data processing are set out in WP 12, which also contains some evaluation of the quality of extractions. The GRO did not tabulate the 1871 employers with their employees at division level as they did for 1851, so no detailed comparisons with published data are possible. However, comparisons of coverage between censuses show that the extracted number of employers who reported employees was lower than expected if 1871 had the same proportions as in 1851-61 and 1881. The national employer entrepreneurship rate (% of employers in the whole population) for the other census years was 0.8-0.9%, but only

¹⁵ 1861 Census England & Wales, General Report, p. 29.

0.61% in the 1871 data. Hence, based on expectations in line with 1861 and 1881, the 1871 extracted employers are about 20% under-estimates. However, the discrepancy is not evenly distributed. In 112 Registration Sub-Districts (RSDs) there were no employers at all, and another 116 RSDs which had an employer entrepreneurship rate of less than 0.05%, as shown in Figure 2. Analysis of other early census years has shown that while there are a few areas that genuinely have 0.1 to 0.5% employers, such as south Wales, parts of east London, and parts of the north east, overall a coverage of at least 0.5% employers would be expected. Hence in Figure 2, all red and black areas, and most of the green and yellow ones, are likely to be deficient in coverage from the data extractions. This is due to transcription gaps in the S&N data that cannot be remedied. However, in at least 11 counties there is complete or nearly complete coverage judged by equalling or exceeding 0.5%: the light blue RSDs. A check on these counties shows they are mainly in line with the expected numbers of employers based on 1861 and 1881. A further 9-10 counties have near complete coverage of rural areas but a few deficiencies in some of the urban centres; these give reliable estimates for farmers, but are less reliable for non-farmers. The least reliable are Caernarvon, Cardigan, Cornwall, Cumberland, Durham, Glamorgan, Hampshire, Leicester, Monmouth, Norfolk, Northamptonshire, Sussex, and parts of Lancashire and the West Riding of Yorkshire.

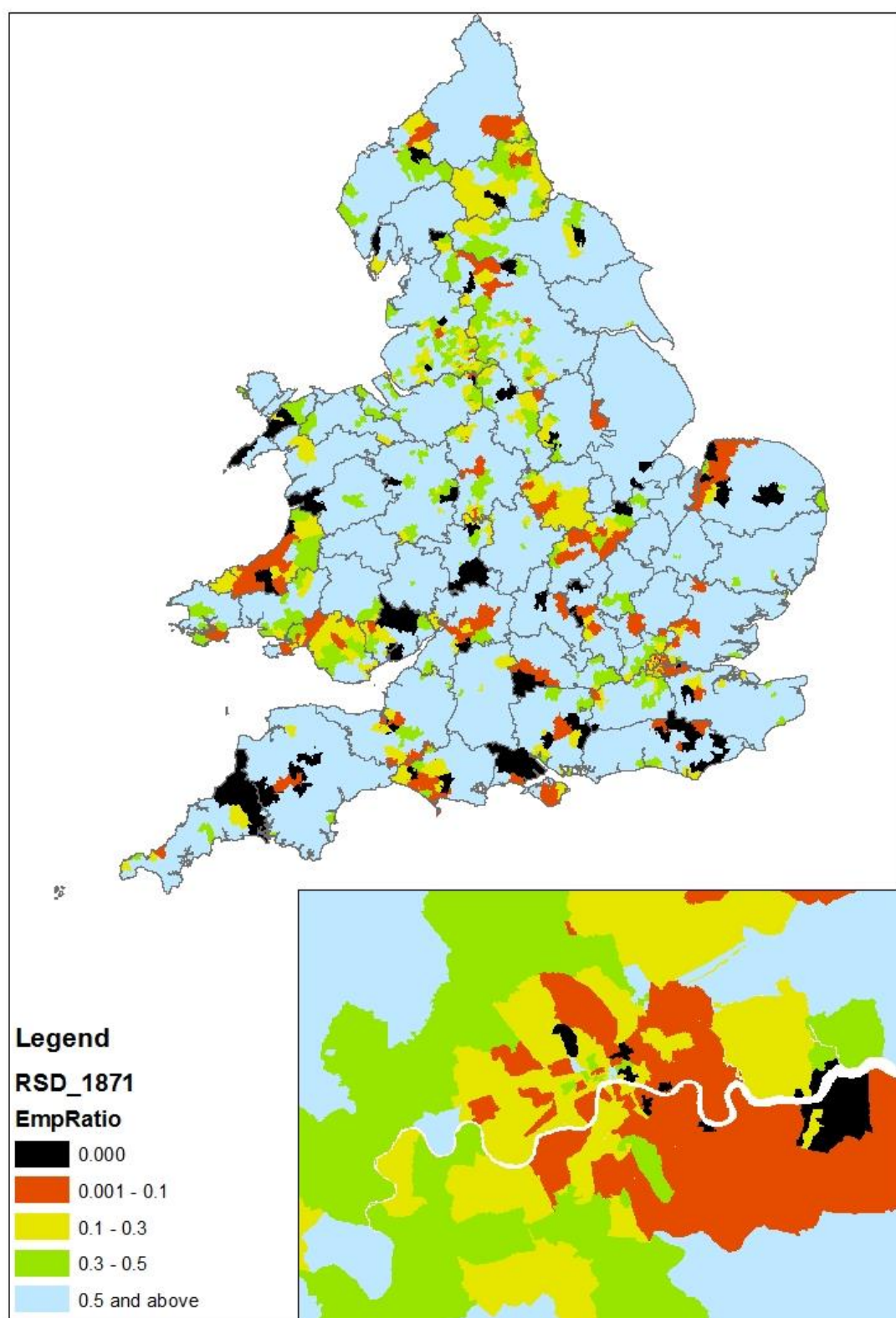


Figure 2. Employer entrepreneurship rate by RSD in 1871; used to indicate possible gaps in transcription coverage; RSDs below 0.5 indicate possible omissions.

5.1. The available data from the GRO report

Although the GRO did not tabulate the employers with their employees, they did tabulate a limited analysis for farmers, and the report contained some further analysis on farm size and workforce numbers. This was conducted on a group of 17 ‘representative’ counties in England, only 4 of which overlapped with the counties GRO chose in 1861. The counties for 1871 were: Surrey (Extra-Metropolitan), Kent (Extra-Metropolitan), Sussex, Hampshire, Berkshire, Essex, Suffolk, Norfolk, Leicestershire, Rutland, Lincolnshire, Nottinghamshire, Derbyshire, Durham, Northumberland, Cumberland, and Westmorland.¹⁶ It should be noted that while some of these are amongst the better-transcribed counties when it comes to employers (e.g. Lincolnshire, Suffolk), there are also some of the poorer ones, such as Durham, Hampshire and Leicestershire.

The 1871 published report calculated that of the farmers who employed one or more labourers, the average number of employees was fewer than 6.¹⁷ The average number of employees in the extracted data is 7.8, an indication that the extraction picked up more large rather than small farms, and this seems to be more skewed than the 1851 farmer extractions. This is also evident in the breakdown by size between the published and the extracted farms, as shown in Tables 7 and 8. Since the extraction method picked up the farmers without employees through a different algorithm, the results for farmers with zero employees have not been included in the total in Table 8.

¹⁶ 1871 Census of England and Wales, *General Report, Parliamentary Papers, LXXI (1873), 124-9.*

¹⁷ 1871 Census England & Wales, General Report, , p. xlviii

labourers	Surrey		Kent		Sussex		Hampshire		Berkshire		Essex		Suffolk		Norfolk		Leicestershire	
	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N
0	299	247	934	871	962	987	721	687	254	228	400	310	753	985	1692	2197	1220	698
1	130	86	428	278	423	150	295	76	131	62	281	141	579	318	879	422	502	127
2	142	73	454	312	422	196	285	111	164	58	389	191	548	333	625	391	386	119
3	126	81	347	234	300	189	230	107	128	61	307	185	468	313	462	326	272	94
4	88	80	294	211	235	153	170	83	103	57	279	151	434	290	333	240	216	80
5	73	75	219	198	206	107	153	87	98	49	264	168	300	245	236	190	118	76
6	64	46	234	191	193	107	134	60	103	44	214	153	281	226	226	160	102	40
7	60	50	129	133	130	98	104	61	71	43	162	150	219	179	171	137	59	29
8	55	42	157	139	125	98	137	61	112	53	199	118	220	167	183	134	41	21
9	35	32	114	102	84	72	75	51	52	35	145	114	132	140	124	108	25	23
10-	105	129	329	379	259	256	263	188	200	146	469	500	417	475	415	360	74	45
15-	37	51	132	180	87	128	84	101	89	98	209	255	163	220	193	175	12	16
20-	16	30	89	96	52	76	52	71	57	69	131	153	69	114	85	122	3	7
25-	6	12	44	76	21	30	30	24	20	33	42	87	37	59	53	54	2	4
30-	3	9	26	46	17	30	10	20	10	17	31	50	17	46	33	36	0	1
35-	1	6	10	23	4	12	2	11	2	26	13	34	11	23	15	20	0	3
40-	2	3	12	16	4	4	3	4	6	11	16	18	9	12	6	11	1	1
45-	0	1	5	9	0	7	0	3	3	0	4	14	2	11	7	7	0	0
50-	1	1	6	5	3	0	0	1	0	2	5	8	3	5	1	6	0	0
55-	0	0	3	3	2	3	0	0	1	3	4	8	2	4	4	5	0	0
60-	0	2	11	20	3	4	1	0	0	3	10	20	6	12	5	12	0	0
total	1243	1056	3977	3522	3532	2708	2749	1807	1604	1098	3574	2828	4670	4177	5748	5113	3033	1384

Labourers	Rutland		Lincolnshire		Nottinghamshire		Derbyshire		Durham		Northumberland		Cumberland		Westmorland	
	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N	Pub	S&N
0	196	157	5144	4940	1923	1411	4068	2661	2412	1679	1268	1017	1610	2644	1762	1677
1	91	36	1251	747	503	213	650	383	303	133	206	118	385	242	161	107
2	65	33	969	681	400	222	354	241	302	167	273	175	322	219	107	97
3	52	45	614	500	220	163	204	190	178	107	231	120	166	123	45	49
4	39	35	444	363	184	120	84	109	104	53	182	121	95	81	33	30
5	24	32	326	299	86	87	31	67	47	44	92	86	54	45	11	12
6	13	17	255	263	77	79	11	30	35	24	102	64	22	25	11	14
7	9	21	181	224	54	66	7	23	22	15	52	45	14	15	6	2
8	11	19	189	164	40	51	7	21	28	16	65	54	17	11	3	6
9	7	7	115	138	23	32	2	5	7	16	32	34	4	4	1	4
10-	22	21	370	491	74	116	11	18	15	23	89	109	15	25	2	6
15-	7	16	112	222	14	31	2	8	3	8	34	40	1	5	2	3
20-	2	2	67	123	2	16	1	3	0	4	13	26	2	3	0	0
25-	0	2	27	50	0	4	1	1	1	4	3	16	1	0	0	0
30-	0	0	10	24	0	2	0	3	0	3	4	15	0	0	0	0
35-	0	0	4	17	0	0	0	1	0	0	4	7	0	0	0	1
40-	0	1	9	19	1	0	1	0	0	0	1	4	0	1	0	0
45-	0	1	3	5	1	1	1	0	0	0	0	1	0	1	0	0
50-	0	0	0	5	1	1	0	0	0	0	1	1	0	0	0	0
55-	0	0	1	2	0	0	0	0	0	0	0	1	0	0	0	0
60-	0	0	2	6	0	1	0	0	0	1	2	4	0	0	0	0
total	538	445	10093	9283	3603	2616	5435	3764	3457	2297	2654	2058	2708	3444	2144	2008

Table 7. Published vs extracted farm data in 1871 by number of employees for the 17 ‘representative’ counties.

	Total 17 counties		Comparison	
labourers	Published	S&N	Difference	%Difference
0	25618	23396	-2222	-8.7
1	7198	3639	-3559	-49.4
2	6207	3619	-2588	-41.7
3	4350	2887	-1463	-33.6
4	3317	2257	-1060	-32.0
5	2338	1867	-471	-20.1
6	2077	1543	-534	-25.7
7	1450	1291	-159	-11.0
8	1589	1176	-413	-26.0
9	977	917	-60	-6.1
10-	3129	3287	158	5.0
15-	1181	1557	376	31.8
20-	641	915	274	42.7
25-	288	456	168	58.3
30-	161	302	141	87.6
35-	66	184	118	178.8
40-	71	105	34	47.9
45-	26	61	35	134.6
50-	21	35	14	66.7
55-	17	29	12	70.6
60-	40	85	45	112.5
Total (excl 0)	35144	26212	-8932	-25.4

Table 8. Published vs extracted farm data in 1871 by number of employees for the 17 ‘representative’ counties.

Table 8 shows a total underestimate of 25% of employees for farms with employees. However, these are not evenly distributed among the firm sizes. Small farms, with fewer than 10 employees, are up to 49% lower, while the larger farms have larger numbers than the GRO reported, as also occurred in 1851 and 1861. On top of the usual cleaning process involving CEB checks on farms with over 70 employees, for the 1871 data some further CEB checks have been performed on the largest farms in order to test whether these were transcription errors or reflect undercounts by the GRO’s clerks. This process is described in WP 12. It used careful checking of all farms of over 50 employees, with 149 found in the extraction, where the GRO only listed 78 – roughly half of the farmer employers present in the data. Hence, GRO again missed an high proportion of the larger farms. Taking into account that the extracted sample includes counties from which parts are known to be

missing, the ‘real’ number of larger farms is likely to be even higher, implying that the GRO missed over half the large farms in the census returns. While these tables were not produced by the same set of clerks as those of 1851 and 1861 (although a proportion of staff would have been the same) it is likely that similar mistakes were made in 1851 and 1861. Manual CEB checks further down the farm size range are infeasible due to numbers, but the implication of the larger farm checks is either that GRO misallocated the missing large farms as small farms, and/or the extraction possibly misses out many small farms. The total of GRO gaps for larger farms in Table 8 is 1,375 farms (those with over 10 labourers). This is 19.6 percent of the 7,016 included in transcriptions from the S&N data. This is a very high clerical error rate; and if S&N had any transcripts gaps for these size groups the error rate would be even higher. However, the 1,375 that GRO missed for the larger farms is far below the under-estimate for the farms below 10 employees which is 10,307. So it is likely that the GRO clerical errors were probably mostly included in the smaller categories as clerical mistakes. The under-estimates suggest that the S&N data may have many transcription gaps for the smaller farms; although these will be some extent lesser than indicated because GRO will have included some non-main farms (for portfolios) in the farm tables. A rough bound might be guessed at by assuming the GRO included the 1,375 large farms they missed in the small categories ($10,307 - 1,375 = 8,932$), and that 10% of these were portfolios that should have been attributed to non-farm businesses ($8,932 - 893 = 8,039$). The portfolio percentage is an average for the number of portfolios over the period. This suggests that S&N may have missed about 27 percent of the 29,503 farms below 10 employees in the 17 counties that GRO tabulated. However, S&N believe their transcriptions were of far higher quality than this, and they made quality control checks which did not pick up significant errors. This leaves the interpretation of the GRO data difficult and suggests it had various other clerical or compilation discrepancies compared with what can now be found in the CEBs.

5.2. Acreage checks 1871 farmers

Unlike the 1851 S&N infill extractions, for 1871 the S&N extractions did include the farmers who stated their acreage only, which allows a comparison with the breakdown in acreage in the published tables. The GRO calculated that the average farm size in the ‘representative’ counties was 152 acres, with more than a fifth of farms occupying less than 20 acres.¹⁸ This

¹⁸ 1871 Census England & Wales, General Report., p. xlvii

corresponds exactly with the extracted data, which has an average of 152.4 acres. Table 9 shows the comparison between the published and extracted acreages broken down by farm size, for the 17 counties grouped together. The 19% missing acreage overall corresponds with the estimated missing people due to transcription error, and is higher than the overall missing acreage in 1861 at 16%. As in 1861, the missing acreages are distributed more evenly across the farm sizes than the missing employers, although the percentage of missing small farms is much higher, proportionally twice as much as the overall missing acreage.

Acres	Published	Extracted	Difference	%Difference
Under 5	1984	1198	786	39.6
5-	4017	2845	1172	29.2
10-	6074	4731	1343	22.1
20-	4193	3416	777	18.5
30-	3363	2795	568	16.9
40-	3048	2597	451	14.8
50-	6370	5401	969	15.2
75-	4113	3486	627	15.2
100-	7341	6351	990	13.5
150-	4706	3834	872	18.5
200-	3927	3205	722	18.4
250-	2324	1902	422	18.2
300-	2226	1819	407	18.3
350-	1166	947	219	18.8
400-	1824	1493	331	18.1
500-	1098	875	223	20.3
600-	666	565	101	15.2
700-	390	304	86	22.1
800-	270	218	52	19.3
900-	188	159	29	15.4
1000-	249	186	63	25.3
1200-	159	118	41	25.8
1500-	84	61	23	27.4
2000 upwards	90	85	5	5.6
Total	59870	48591	11279	18.8

Table 9. Published vs extracted farm data by acreage.

6. Assessment and conclusion

The purpose of this paper has been to use the published tables from GRO to assess both how far the extractions from the electronic 1851-71 versions of the census give reliable estimates of the population of employers, and what the extractions also tell us about the reliability of the GRO process that was used for producing published tables of these data.

There are four main conclusions on the GRO published tables for the years 1851-71. First, the GRO tabulations of the CEB data contain errors and omissions which have not been previously identified. There was a tendency for clerks to miss a substantial proportion of the (smaller number of) larger farms and non-farm firms that exist in the CEBs. Second, GRO usually had a higher count of the smaller farms and firms; some of these larger firms will have been misallocated, but some probably indicate deficiencies in transcriptions and ability to identify employers through the extraction method used. Third, however, it is difficult to be sure exactly what the GRO tables actually tabulated as the definitions used in the instructions and the way the tables claim to cover the data are slightly different. As shown in section 3, comparisons using the various interpretations of the same definitions stated by GRO for 1851 fail to produce results that confirm exactly what GRO did. It is likely that the clerks, when confronted with the complexity and scale of the tabulation challenge with these data, were not fully consistent nor able to fully distinguish all the different elements of the question. This results in discrepancies in the published tables between what they claim to cover, and the data themselves.

Fourth, although we have not commented above, the analysis also indicates that the GRO process for collecting and tabulating in 1851 included more employees in the workforce headcount for the smallest firms than the later censuses. This is likely to derive from a more thorough effort to include spouse and other family in 1851 that was not fully repeated in other years. In any case, all the censuses have some ambiguity about how family members were included, as well as ambiguity about how respondents thought about their spouse and family, and hence whether their occupations were fully included or not. For these reasons there is inconsistency between the censuses in the coverage of the very smallest firms, with 1851 including more than later years. To obtain consistent comparisons over time subsequent

analysis should therefore exclude the smallest firms with only 1 employee, or those with 4 or 5 employees and under.

There are also three main conclusions about the data that is now available in the I-CeM/S&N version assessed here. First, the overall level of coverage compared to published tables is generally good, especially for the larger firms and farms. The numbers identified in the electronic records and the published tables generally match well (once lost data and transcription omissions are taken into account). This should give confidence that the electronic records capture what the census actually recorded. Second, it is clear that for all years 1851-71 there are gaps in transcriptions. These occur despite the substantial efforts made to infill identified gaps. The gaps mainly arise from truncation of lines by transcribers so that the crucial information on employee numbers and/or acres was omitted, and also from total omission of some individuals' occupation strings. Some of the transcription omission is a random loss where individual transcribers, who were otherwise thorough, missed odd entries. This probably accounts for less than one percent of missing entries (based on previous assessments of transcription errors in I-CeM for the extracted employer data these were under 0.5 percent for age and gender data).¹⁹ However, unfortunately, the biggest gaps occur in large geographical blocks. Because transcription was usually undertaken in batches for geographical blocks of CEBs, a batch could end up with a very poor transcriber who omitted all the required data on employees and acres, resulting in non-random geographical concentrations in certain counties or smaller units. This affects data derived from both FMP and S&N. These gaps are far worse for 1871 than the other years; relying on S&N as the only source gives poorer coverage than the combination of FMP and S&N used in 1851, or FMP alone in 1861.

Third, because the electronic versions of the census now available have been assembled from different sources by different processes the quality of coverage varies by census year. This needs to be taken into account in subsequent analysis. The main differences can be summarised as:

1851 census: The FMP transcriptions for this census had significant amounts of material lost through truncation. Much of this has been infilled from S&N data, which has overcome most

¹⁹ See Bennett, R.J. and Newton, G. 'Employers and the 1881 population census of England and Wales', *Local Population Studies*, 94 (2015), 29-49.

gaps; but total coverage is still incomplete. FMP transcriptions are most incomplete for non-farm employers, even after topping up from S&N, in London and the North West. Other areas can probably be treated as have random omissions, except among the smallest firm size categories as noted above. Farm acreage transcriptions are most incomplete in Sussex, Middlesex, Norfolk, Dorset, Gloucestershire, Warwickshire, Derbyshire, Cheshire, Lancashire, Northumberland, parts of Yorkshire and Wales.

1861 census: There is a general problem from all sources that can be used for the electronic census that 3.7% of the 1861 records have not survived and hence cannot be included in I-CeM or S&N. These have a concentrations in Wales, small parts of London, and a scatter of isolated locations across the country. In addition the FMP transcriptions in I-CeM analysed here show a small number of further gaps in a few isolated locations mainly in London and surrounding districts, parts of Monmouth and Glamorgan, and a few districts elsewhere. However, for 1861, FMP transcriptions are usually complete and the main missing coverage comes from lost records. Indeed 1861 records can generally be taken as the most complete for the three years 1851-71.

1871 census: This has the poorest transcriptions. However, at least 11 counties have complete or nearly complete coverage, and a further 9-10 counties have good coverage of rural areas with deficiencies mainly in a few urban centres. Hence for much of the country there is adequate coverage, and best for farmers. The least reliable counties are Caernarvon, Cardigan, Cornwall, Cumberland, Durham, Glamorgan, Hampshire, Leicester, Monmouth, Norfolk, Northamptonshire, Sussex, and parts of Lancashire and the West Riding of Yorkshire.

For the 1881 census, although no comparisons can be made against GRO tables since none were published, we can draw conclusions based on the three earlier censuses since the 1881 questions were administered in exactly the same format. This census should give the fullest and most accurate records of all the censuses 1851-81. Hence, it should provide the best benchmark for what constitutes full coverage. This is because the transcription was undertaken in an entirely different way to the other censuses. It was transcribed by the Genealogy Society of Utah (GSU) by individuals with a commitment and interest in the accuracy of records for genealogy that is not reflected in the commercial and other motives

applied to the transcription process used by FMP or S&N. It is also believed that much was double keyed and hence had an inbuilt accuracy check.²⁰ The GSU transcripts were deposited at UKDA, and were incorporated into I-CeM. Given its origin it is believed that this is by far the most accurate set of electronic records of the census years in I-CeM. Checks made during the ESRC *Drivers of Entrepreneurship and Small Businesses* project confirm this conclusion: there were virtually no truncations (and those found have all been infilled), very few keying errors detected against CEBs where they have been fully checked, and no systematic omissions by RSD were detected.²¹ There will of course remain enumerator and response error, but 1881 should be almost totally free of transcriber error.

Comparisons of the 1881 data with that for 1851-71 show very comparable coverage by firm size, with the same relatively low counts for small firms and strong coverage of larger firms.²² A detailed comparison of the frequency distribution by firm size also shows that the same mathematical curves fit closely to all the years (but with 1871 as expected far weaker).²³ Their characteristic is a ‘heavy tail’ for the larger firm sizes, and this specific feature is reproduced by the same mathematical functions for each year. This gives strong confirmation that the main part of the firm-size distribution is captured by the electronic records and extraction methods used.

The central conclusion of this working paper is, therefore, that although there are deficiencies of the transcriptions and extractions possible from the I-CeM/S&N data, these can be mostly overcome for analysis purposes, and in many cases the new extractions are less ambiguous in definitions and more complete in some respects than achieved by GRO published tables, especially for larger firms and larger farms. However, subsequent analysis needs to take account of the deficiencies as follows:

- Either adopt a weighting process to allow for data omissions,
- or use proportions rather than absolute number counts.

²⁰ 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. Although the two versions for 1881 are nominally the same, the version used includes many corrections and updates to occupational and other codes made at Campop.

²¹ See Bennett, R.J. and Newton, G. ‘Employers and the 1881 population census of England and Wales’, *Local Population Studies*, 94 (2015), 29-49.

²² Bennett et al., *The Age of Entrepreneurship*, (Routledge: 2019), chapter 5.

²³ Montebruno et al., ‘A tale of two tails’, (2019).

- To obtain consistent comparisons over time, exclude the smallest firms with only 1 employee, or those with 4 or 5 employees and under, depending on the purposes required.

Acknowledgements.

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

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

The GIS boundary files for RSDs were constructed by Joe Day for the ESRC fertility project directed by Alice Reid:

<http://www.geog.cam.ac.uk/research/projects/victorianfertilitydecline/publications.html>

These, and GIS parish files, develop from 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.

References.

- Bennett, Robert J. and Newton Gill (2015) Employers and the 1881 population census of England and Wales, *Local Population Studies*, 94, 29-49.
- Bennett, Robert J., Smith, Harry, van Lieshout, Carry, Montebruno, Piero, and Newton, Gill (2019) *The Age of Entrepreneurship: Business proprietors, self-employment and corporations since 1851*, Routledge, London.
- Higgs, Edward and Schürer, Kevin (2014) *The Integrated Census Microdata (I-CeM)* UKDA, University of Essex, SN-7481.
- Mills, D. R., (1999) Trouble with farms at the Census Office: an evaluation of farm statistics from the censuses of 1851-1881 in England and Wales, *The Agricultural History Review*, 4), 58-77. Montebruno, P., Bennett, R.J., van Lieshout, C. and Smith, H

- (2019a) Shifts in agrarian entrepreneurship in mid-Victorian England and Wales, *The Agricultural History Review*, 67(1).
- Monteburno, P., Bennett, R. J., van Lieshout, C. and Smith, H (2019b) A tale of two tails: Do Power Law and Lognormal models fit firm-size distributions in the Mid-Victorian era?
- Schürer, Kevin and Woollard, Matthew (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)*, UKDA, University of Essex, SN-4177.
- Schürer, Kevin, Higgs, Edward, Reid, Alice M., Garrett, Eilidh M. (2016) *Integrated Census Microdata V.2 (I-CeM.2)* [data collection].
- Schürer, Kevin and Woollard, Matthew (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)* [computer file] UKDA, SN-4177.
- Shaw-Taylor, Leigh, (2005) Family Farms and Capitalist Farms in Mid Nineteenth-Century England, *The Agricultural History Review*, 53, 158-191.

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, van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design*.
<https://doi.org/10.17863/CAM.9508>

WP 2: Bennett, Robert J., Smith Harry 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*.
<https://doi.org/10.17863/CAM.9640>

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*.
<https://doi.org/10.17863/CAM.9639>

WP 4: Smith, Harry, Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911*. <https://doi.org/10.17863/CAM.9638>

WP 5: Bennett, Robert J., Smith Harry, van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*.
<https://doi.org/10.17863/CAM.9874>

Data download of classification file: <https://doi.org/10.17863/CAM.9874>

- WP 6: Smith, Harry and Bennett, Robert J. (2017) *Urban-Rural Classification using Census data, 1851-1911*. <https://doi.org/10.17863/CAM.15763>
- WP 7: Smith, Harry, Bennett, Robert J., and Radicic, Dragana (2017) *Classification of towns in 1891 using factor analysis*. <https://doi.org/10.17863/CAM.15767>
- 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*. <https://doi.org/10.17863/CAM.15764>
- WP 9: Bennett, Robert, J., Montebruno, Piero, Smith, Harry, and van Lieshout, Carry (2018) *Reconstructing entrepreneurship and business numbers for censuses 1851-81*.
- WP 10: Bennett, Robert, J., Smith, Harry and Radicic, Dragana (2018) *Classification of environments of entrepreneurship: Factor analysis of Registration Sub-Districts (RSDs) in 1891*. <https://doi.org/10.17863/CAM.26386>
- WP 11: Montebruno, Piero (2018) *Adjustment Weights 1891-1911: Weights to adjust entrepreneur numbers for non-response and misallocation bias in Censuses 1891-1911*. <https://doi.org/10.17863/CAM.26378>
Adjustment weights: <https://doi.org/10.17863/CAM.26376>
- WP 12: van Lieshout, Carry, Day, Joseph, Montebruno, Piero and Bennett Robert J. (2018) *Extraction of data on Entrepreneurs from the 1871 Census to supplement I-CeM*. <https://doi.org/10.17863/CAM.27488>
- WP 13: van Lieshout, Carry, Bennett, Robert J. and Smith, Harry (2019) *Extracted data on employers and farmers compared with published tables in the Census General Reports, 1851-1881*.
- WP 14: van Lieshout, Carry, Bennett Robert J. and Montebruno, Piero (2019) *Company Directors: Directory and Census record linkage*.
- WP 15: Bennett, Robert, J., Montebruno, Piero, Smith, Harry and van Lieshout, Carry (2019) *Entrepreneurial discrete choice: Modelling decisions between self-employment, employer and worker status*.

Full list of all current Working Papers available at:

<https://www.campop.geog.cam.ac.uk/research/projects/driversofentrepreneurship/>