



Reconstructing G. Udny Yule's 1899 Pauperism Dataset

Ian Plewis and Ellie Schmidt

December 2015

The paper Udney Yule read to the Royal Statistical Society at the end of the nineteenth century (Yule, 1899) was a landmark in social statistics. He applied multiple regression analysis to a question of social policy, namely reforms to the 19th century system of poverty alleviation in England. To do this, Yule created a dataset from administrative and Census data. Yule's original dataset was not preserved, but because his data were drawn from public sources, it is possible to reconstruct it, albeit with some slight differences from the original. This report provides a description of how the dataset was reconstructed and how it varies from the one used in the 1899 paper. The reconstructed dataset, along with all the supporting material discussed in this report, are publicly available at the UK Data Archive:

<https://discover.ukdataservice.ac.uk/catalogue?sn=7822>

Enquiries about the data should be addressed to the Data Archive in the first instance. The data are used by Plewis (2015) who revisits Yule's original analysis and builds on it.

Original Data

The data Yule used are described on pp252-256 of his 1899 paper. Yule chose English poor law unions (PLU) as his unit of analysis. For each PLU, data were recorded for the Census of Population years 1871, 1881 and 1891. He focused on the following five variables:

1) Pauperism

Yule defined the pauperism rate in a PLU as the "percentage of the population in receipt of relief of any kind" (p. 252), excluding recipients who were classified as lunatic or vagrant.

2) "Outdoor" versus "Indoor" Relief

Central to the policy debate at the time was the question of limiting relief to those who resided 'indoors' in workhouses or other official facilities. Poverty relief delivered to individuals living in the community was termed 'outdoor' relief, while 'indoor' relief refers to benefits for those who were living in institutionalised settings.

3) Age distribution

Citizens aged 65 and older were used to calculate the 'Proportion of Old' (p. 253) within a PLU in order to control for variation in this variable across PLUs.

4) Population

Yule viewed population change over time as an indicator of the economic prosperity of a PLU and calculated these changes for (i) 1871 to 1881 and (ii) 1881 to 1891.

5) Population density

The population density per acre allowed Yule to construct a variable that characterised a PLU as rural, mixed, urban or metropolitan (London). He analysed each of these four groups of PLUs separately.

Yule assembled his data from three sources. Official reports to Parliament supplied counts of paupers for each PLU, broken down by the manner in which the relief was administered. The figures for age distribution and population were taken from the decennial Census of Population returns. The information about population density was taken from Booth (1894) who used 1891 Census data. Following requests by discussants of Yule's read paper, the Journal of the Royal Statistical Society published some of the data he had used: six tables of so-called frequency tables that provide the

cross-tabulations for all pairs of categorised 'percentage ratios' for variables z_k ; $k=1..4$ (i.e. $100(z_{kt}/z_{k,t-1})$, $t = 1,2,3$; Yule, 1899, p. 254) for the rural group for 1871-1881 (Tables I to VI), and raw data for the metropolitan group of PLUs ($n = 32$) for 1871 to 1881 (Table XIX). The published tables can still be found in the RSS archives but the others (Tables VII to XVIII, presumably for the mixed and urban groups of PLUs) appear to be lost as do all the data Yule used for 1881-1891. The bivariate frequency tables are not, of course, ideal for secondary analysis.

Data Reconstruction

Access to the historical documents used by Yule made it possible to reconstruct his dataset. A subscription database, House of Commons Parliamentary Papers, contains digitised copies of the pauperism reports submitted to Parliament. Nineteenth century Census data is archived at Histpop.org, a public website. The digitised versions of the original documents are specified here (note that hyperlinks for each title redirect to the online document for that title):

[1871 \(140B\) Poor rates and pauperism. Return \(B.\). Paupers relieved on 1st January 1871.](#) (login required)

[1881 \(60B\) Poor rates and pauperism. Return \(B.\). Paupers relieved on 1st January 1881.](#) (login required)

[1890-91 \(130B\) Pauperism \(England and Wales\). Return \(B.\). Paupers relieved on 1st January 1891.](#) (login required)

[Population abstracts, England and Wales, Vol. III, 1871.](#)

[Population, England and Wales. Vol. III. Age, marriage, occupation, birth-place, 1881.](#)

[Population. Age, marriage, occupation, birth-place, England and Wales, Vol. III, 1891.](#)

Booth's 1894 book can be purchased or found in some library collections. The tables on pp58-98 contain data on population density (persons per acre) for each PLU.

Optical Character Recognition

The online availability of the historical pauperism returns and Census abstracts facilitated the reconstruction of Yule's database. However, the digital documents are scans of the original sources and are stored in the form of Tagged Image File Format (TIF) images. The image files are not machine-readable, meaning that the letters and numbers on a page are not recognized as such, but rather as pixels in an image. For this reason, Optical Character Recognition (OCR) software was used in the effort to re-create Yule's data. Four OCR tools were tested for their accuracy in recognising pixel patterns as alphanumeric characters. Accuracy was assessed by comparing the results of the OCR reading with tables of data that were double entered by hand. Three OCR programs were too inaccurate to use in this project: Microsoft Document Imaging, FreeOCR and Microsoft Office OneNote. The OCR tool that had the most accurate results was ABBYY FineReader 11, and this was the piece of software used.

FineReader allows users to train the software to recognise patterns designated by the user. FineReader's built-in pattern recognition is based on modern fonts, so the recognition training feature was important for working with antiquated fonts. FineReader was trained to recognise numbers as they were printed in the 19th century sources. This was crucial to the software being able to distinguish reliably between numbers that appeared similar in the original images. (For example, in the old-fashioned font, the numbers 1 and 7 sometimes appear similar, as do other combinations of numbers and letters.) FineReader was trained to recognise the peculiarities of the fonts in each of the six documents. A pattern was trained separately for each source document since print blemish patterns differed from year to year.

FineReader also flags characters for which it is uncertain about the recognition. Users can move through an on-screen verification process to compare visually each character which FineReader flagged as being questionably recognised with the corresponding section of the original image. Printing irregularities from the antique documents, and shadows from scanning, meant that some sections of the documents were more difficult to recognise than others. FineReader's flagging of these sections facilitated detailed visual inspection of the harder-to-read pages. The verification process provided a fast and user-friendly way to improve the results of the OCR.

The accuracy of the character recognition was scrutinised for 169 PLUs. The first 68 of these PLUs (Kensington to Romney Marsh) were used in initial testing of the OCR software programs. The accuracy of FineReader's OCR for an additional 101 PLUs (Rye through Windsor and Bodmin through Stow on the Wold) was also closely assessed. The PLUs Rye through Windsor were chosen simply to complete the data entry on the PLUs in the South Eastern region. Then, as an additional test of the reliability of FineReader's OCR, a page (p.32) was selected at random from the 1871 pauperism return document containing PLUs from Bodmin through Stow on the Wold. Working with these PLUs, the accuracy of FineReader's OCR was assessed for the 1871, 1881 and 1891 pauperism returns, as well as the 1871, 1881 and 1891 Census abstracts. For the 169 PLUs which were used in testing accuracy, data from all six source documents were double hand-entered, and the data were also run through FineReader OCR. The files **hcpp_trained_and_verified.xlsx** and **histpop_trained_and_verified.xlsx** provide full details of accuracy testing.

The FineReader files were exported to Microsoft Excel to determine the accuracy rates of FineReader's OCR compared with hand-entered data. For pauperism returns from the House of Commons Parliamentary Papers collection, accuracy rates were based on 7348 cells of data. After training and verifying the data, 180 cells (2%) were recognised incorrectly. However, the majority of these errors could be easily identified and corrected. For instance, at times FineReader inserted the letter *i* into a cell, which was an incorrect reading of broken lines on the printed table, dust particles on the scan or other irregularities in the image. It was straightforward to assemble a list of non-numeric errors, then search for these in the Excel file and refer to the digital image for the correct value of the cell. After correcting these, 30 errors remained, an effective error rate of 0.41%.

The accuracy assessment for the OCR of Histpop's population abstracts was based on 8313 cells of data. After training FineReader to recognise the characters and going through the verification process for the population abstract data, 1530 errors (18%) were found in FineReader's OCR of the tables. Such an error rate is clearly not acceptable. Some of these were conspicuous errors that could easily be corrected (such as recognising an *S* instead of an *8*). However, most of the errors occurred in the last two columns of data. FineReader did not cope well with reading the columns of data recording the population aged 95-99 and over 100.

-AGES of MALES and FEMALES—In REGISTRARS' SUB-DISTRICTS.†

and es.	Under 5 Years.	5-	10-	15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-	70-	75-	80-	85-	90-	95-	100 and upw*
731	3873	2664	2197	2217	2400	2402	2028	1657	1400	1112	961	665	534	338	244	110	55	19	.	1	.
907	3696	2835	2552	3328	4228	3700	2088	2407	1954	1615	1458	981	854	604	416	213	107	22	8	1	.
110	1590	1212	1050	1369	1567	1339	1148	951	800	771	700	519	428	276	190	88	40	15	2	.	.
1975	1507	1365	1358	2003	3395	3198	2378	1855	1641	1264	1047	755	654	403	285	168	69	23	4	.	.
740	5663	4491	3559	3414	3388	3334	3050	2265	2196	1778	1476	1021	848	571	388	198	90	17	7	2	.
896	5644	4803	4320	5587	6683	5063	4598	3646	3113	2489	2190	1463	1315	846	664	312	179	55	22	8	1
228	1402	1239	900	1037	1114	1055	878	790	714	566	519	350	245	186	125	62	28	6	3	.	.
426	1503	1232	1173	1850	2465	2096	1546	1280	1041	848	714	515	475	288	202	126	50	12	8	2	1
927	432	372	334	248	241	232	203	183	164	140	109	71	72	51	47	23	9	4	2	.	.
735	424	410	341	336	398	348	241	243	199	155	119	116	57	80	42	16	7	4	.	.	.
1377	2586	2123	1796	1469	1450	1421	1197	1069	998	747	589	379	306	174	146	81	30	12	1	1	.
1652	2619	2115	1896	1387	1383	1341	1566	1226	1147	895	772	546	467	317	280	154	101	27	6	1	.

Figure 1: Excerpt from a Population Abstract Table

As shown in Figure 1, the population abstracts used ‘.’ to indicate no data for a cell. When presented with columns of numbers mixed with full stops, FineReader had difficulty recognising the data in the table and often failed to recognise any data at all. As a result, a decision was made not to use OCR for the ‘95-’ and ‘100 and upw’ columns of data. Instead, these data were double entered by hand for all 599 PLUs in the reconstructed dataset. If conspicuous errors and errors in age 95+ columns of data were not considered, 244 mistakes were found, an effective error rate of 2.94%. This error rate was good but not sufficient to rely solely on OCR of population abstracts. As a result, the choice was made to use OCR to convert the tables of population aged 65-94, but each cell of data was also checked by hand.

Finally, the full documents were run through the FineReader software. Six separate FineReader files were built, one set of images for each of the original documents. Each file was read using the OCR pattern trained specifically for that document. All 202 pages of the original source material were systematically reviewed through the FineReader verification process. The results of FineReader’s OCR were exported into Microsoft Excel files. The data were then cleaned for the symbols and non-numeric characters shown in [Yule_cleaning.pdf](#) (available from the Data Archive on request).

Dataset Preparation

In setting up the data file, the master list of PLUs was taken from the pauperism returns archived in the House of Commons Parliamentary Papers. If a PLU did not appear on the list in all three waves of data collection, data were entered for the years available and left missing for the year(s) in which the data were not available. Additionally, some PLU boundaries changed during the period from 1871 to 1891. Thus, there are some PLUs with less than three years of data. Due to the variation across the decades in how PLUs were recorded on the pauperism returns, 599 PLUs appear in stage one of the reconstructed data set. The returns from 1891 required an additional step to correspond to the previous years. In 1891, the returns began recording additional detail for men and women receiving indoor relief and males receiving outdoor relief. The variables were combined in the following way for males (and in the same way for females):

Male able-bodied indoors: MABI91 = MABI91InHealth + MABI91TempDis

Male able-bodied outdoors: MABO91 = MABO91Sick + MABO91Other

Three separate Excel files containing pauperism data from 1871, 1881 and 1891 were combined into one file, [hcpp_complete.xlsx](#).

Districts in the Census abstracts were matched to the master list of PLUs. Some names were not identical between the two sources but were assumed to refer to the same places, such as a place called *Wisbeach, Cambridgeshire, South Midlands* in the pauperism returns but *Wisbech, Cambridgeshire, South Midlands* in the Census abstracts. See **hcpp_and_histpop.xlsx** for a side-by-side comparison of the place names.

In 20 cases, a PLU was classified as a sub-district in the Census records. For example, Dorchester and Cerne are separate PLUs in the pauperism returns. In the Census abstracts, though, Cerne is listed as a sub-district under the Dorchester district. This was dealt with by referring to the sub-district pages in the Census abstracts to obtain the relevant data. To continue with the Dorchester/Cerne example, the population figures from Cerne were entered into a new line of data and then subtracted from the Dorchester population figures that appear in the district-level tables. All data entry and calculations for each of the 20 instances in which tables at the subdistrict level were consulted can be seen in the worksheet entitled *adding missing unions* in the file **histpop_complete.xlsx**. See also data note 1 at the end of this report.

Every data point in the aged population tables was checked. FineReader's output formatting was used as a starting point, then the data were entered by hand into an Excel spreadsheet. The cells of OCR-generated data were cross-checked with the cells of hand-entered data. Any discrepancies between the two were resolved by referring to the original image files.¹

Yule's dataset contained variables for the population aged 65 and older, but the source material provided a more granular level of detail, with tables detailing the counts of males and females in five year age brackets. The data tables retrieved from the Census abstracts needed to be converted into variables that matched with Yule's variables. Excel formulae were used to sum the population counts of the eight columns of age brackets for each gender within each union. Figure 2 illustrates how these formulae were set up. After calculating the required variables in Excel, the cells were pasted as values into the **histpop_complete.xlsx** file.

¹ See the *age* tabs in the following path locations. The data for unions that do not appear in these files had already been double-entered in the accuracy checking process. Note that hand-entered data is on the left of the sheet, and the FineReader output is on the right side of the sheet:

71_histpop_smid_to_northern.xlsx
81_histpop_smid_to_northern.xlsx
91_histpop_smid_to_northern.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Union	Region	County	M6571	F6571	65	70	75	80	85	90	95	100		check 95	check 100
2	Staines	South Mid	Middlesex	491	=SUM(F3:M3)	199	151	83	40	16	2					
3						203	165	90	50	15	3	2			2	
4	Uxbridge	South Mid	Middlesex	608		235	180	120	54	11	6	2			2	
5						347	237	118	71	24	13	1			1	
6	Brentford	South Mid	Middlesex	1367		561	429	223	120	26	8					
7						766	574	302	151	63	16	7			7	
8	Hendon	South Mid	Middlesex	556		237	174	77	48	16	4	2			2	
9						282	235	127	66	24	6	2			2	
10	Barnet	South Mid	Middlesex	587		245	184	91	48	15	4					
11						254	230	140	71	20	4	1			1	
12	Edmonton	South Mid	Middlesex	1712		709	547	288	123	34	11					
13						929	678	427	201	81	11	3			3	
14	Ware	South Mid	Hertfordshire	455		180	138	78	44	13	2					
15						201	184	90	55	22	5					
16	BishopStortford	South Mid	Hertfordshire	683		270	217	130	55	10	1					
17						266	202	117	58	23	6	1	1		1	1
18	Royston	South Mid	Hertfordshire	866		364	248	178	54	18	4					
19						334	241	142	99	24	5	3			3	
20	Hitchin	South Mid	Hertfordshire	703		263	257	118	52	12	1					
21						352	235	159	62	23	3	1			1	
22	Hertford	South Mid	Hertfordshire	468		198	152	71	33	11	2	1			1	
23						201	161	92	43	11	4	1			1	
24	Hatfield	South Mid	Hertfordshire	235		102	73	35	15	8	1	1			1	
25						95	76	48	26	12	1					
26	StAlbans	South Mid	Hertfordshire	489		213	147	83	36	9	1					
27						243	188	128	67	27	4	2			2	
28	Watford	South Mid	Hertfordshire	711		294	240	100	52	23	1	1			1	
29						371	260	151	64	25	4	2			2	
30	HemelHempstead	South Mid	Hertfordshire	352		157	105	62	21	5	1	1			1	
31						137	126	80	36	10	3	1			1	
32	Berkhamstead	South Mid	Hertfordshire	338		152	101	50	26	6	2	1			1	

Figure 2: Summing Aged Population Data for Histpop Documents

Data from the **hcpp_complete.xlsx** and **histpop_complete.xlsx** files were pasted into the file called **yule_with_formulas_for_ratios.xlsx**. At this point, following Yule, derived variables such as ratios and proportions, were computed from the raw data. The formulae used in constructing the derived variables are visible by clicking on cells in the **yule_with_formulas_for_ratios.xlsx** file. At this point, the dataset was nearly completely reconstructed, and the data were pasted as values into the **yule_final.xlsx** file. Cleaning was carried out for cells with Excel warnings about derived variables (such as ratios which contained a zero in the denominator), and missing data were replaced with ‘.’.

Next, Booth’s population density data were double entered by hand from a copy of Booth (1894). The spreadsheet containing these data can be found at **booth_data.xlsx**. Booth’s listing of PLUs’ population densities was reorganised to match the master list of PLUs, and the data were pasted into the **yule_final.xlsx** file. Yule used these data to group the PLUs into four types: rural, mixed, urban and metropolitan. As shown on p.255 of his paper, Yule’s used population density thresholds to designate the type of PLU, and this was replicated in the reconstructed file. PLUs with a population density between 0.02 and 0.30 persons per acre were classified as rural. PLUs with a population density of 0.31 – 1.00 were classified as mixed. PLUs with a population density of 1.01 to 88.00 were classified as urban. Yule set 32 PLUs in Greater London apart by grouping them into their own type, metropolitan.

Generating the Reconstructed Dataset: Stage One

Once the full dataset was assembled, it held three years of data: for 1871, 1881 and 1891. The reconstructed dataset contained data on 103 variables for 599 PLUs in England during that time frame. In order to prepare the dataset for statistical analysis, a Stata .dta file was created. The file **importing_yule_excel_file.do** (available from the Data Archive on request) gives a record of the commands used to import the data from Excel into Stata. The same file also shows how the variables were set up, organised and labelled. The stage one data file can be found at **yule.dta**.

Although there are 599 PLUs on the file, three (Witham, Shrewsbury, Radford) did not exist in 1891 and so could not be categorised in terms of population density, and one (Seisdon) did not appear in Booth's list. The distribution of the remaining 595 PLUs by type of PLU can be compared with Yule's data (his Table A, p.255) as follows (Table 1):

Table 1: Number of PLUs by Type

PLU Type	Yule: 1871-1881	Yule: 1881-1891	Reconstructed data (1)	Reconstructed data (2)
Rural	236	236	240	235
Mixed	206	207	220	209
Urban	103	105	103	103
Metropolitan	32	32	32	32
Total	577	580	595	579

Table 1 shows some differences between this stage of the reconstructed data (1) and Yule's original tabulations, notably in the number of mixed type PLUs.

Generating the Reconstructed Dataset: Stage Two

Further exploration of the histories of PLUs led to the omission of another 16 PLUs at stage two. Three (Dore, Kington and Great Boughton, later Tarvin) appeared, in some years, to be at least partially in Wales so were omitted. In addition, Booth (and presumably Yule too) combined some smaller PLUs into single entities (see data note 1) leading to a further 13 PLUs being omitted. The final column of Table 1 gives the distribution of PLUs in the file used for analysis (**yule_ip1.dta**) and these show only small differences from Yule's figures.

It was also apparent that some of the population data in the Census tables and in the PLU returns were incorrect so these were rectified as set out in data note 2. Also, two PLUs (Todmorden and Woolwich) did not have workhouses in 1871, necessitating some further adjustments as set out in data note 3.

Comparing the Reconstructed Dataset with Yule

Yule's Table XIX gives the raw data for his percentage changes from 1871 to 1881 for the metropolitan group. Table 2 compares his data with the final reconstructed data. Most of the disagreements can be attributed to differences in rounding accuracy except for the data for Greenwich, Lewisham and Woolwich (see data note 3) and what appears to be a transcription error by Yule for pauperism for Croydon.

Table 3 compares the means and standard deviations and Table 4 the correlations for Yule's percentages ratios (his Tables 1 and 2) with those from the reconstructed dataset. Despite Yule basing his computations on grouped values of his percentage ratios, the differences between the two datasets for the means and standard deviations (Table 3) are generally very small. The differences between the correlations (Table 4), although not substantial, are larger notably for the rural group for 1881-1891, where it is possible that Yule omitted one observation as an outlier.

Table 2: Raw data, Metropolitan group, 1871-1881.

PLU	Pauperism	Out-relief Ratio	Proportion of Old	Population
Kensington	27	5	104	136
Paddington	47	12/13	115	111
Fulham	31	21	85	174
Chelsea	64	21	81	124
St. George's	46	18/19	113	96
Westminster	52	27/26	105	91
Marylebone	81	36	100	97
St. John, Hampstead	61	39	103	141
St. Pancras	61	35	101	107
Islington	59	35	101	132
Hackney	33	22	91	150/149
St. Giles'	76	30/29	103	85
Strand	64	27	97	81
Holborn	79	33/34	95	93
City	79	64	113	68
Shoreditch	52	21	108	100
Bethnal Green	46	19	102	106
Whitechapel	35	6	93	93/94
St. George's East	37	6	98/99	98
Stepney	34	10	87	101
Mile End	43	15/13	102	113
Poplar	37/39	20	102/103	135/134
St. Saviour's	52	22	100	111
St. Olave's	57	32	102	110
Lambeth	57/56	38	99/98	122
Wandsworth	23	18	91	168
Camberwell	30	14	83	168
Greenwich	55	37/38	94	131/130
Lewisham	41/40	24/27	100	142/141
Woolwich	76/77	20/15	119	110
Croydon	38/48	29	101	142
West Ham	38	49	86	203

Notes

1. Unshaded cells – agreement between Yule and reconstruction.
2. Shaded cells - disagreements with Yule followed by reconstruction (a/b).

Table 3: Means (Standard Deviations) of Percentage Ratios

Period	Group	Source	Pauperism	Out-relief Ratio	Proportion of Old	Population	
1871-1881	Rural	Yule	65.4 (16.2)	72.1 (25.2)	103.6 (8.9)	97.5 (7.3)	
		Recon.	66.0 (15.5)	71.7 (25.4)	103.6 (8.4)	97.8 (7.6)	
	Mixed	Yule	66.1 (17.2)	70.0 (24.6)	99.0 (9.8)	111.0 (14.5)	
		Recon.	66.0 (16.5)	69.5 (24.3)	99.2 (9.9)	111.2 (14.0)	
	Urban	Yule	74.0 (25.3)	64.4 (25.4)	98.5 (7.9)	121.5 (16.2)	
		Recon.	74.5 (25.8)	64.6 (25.5)	98.1 (8.1)	121.5 (16.0)	
	Metropolitan	Yule	50.3 (16.2)	25.2 (12.6)	99.2 (9.0)	120.0 (29.8)	
		Recon.	50.6 (16.3)	25.1 (13.0)	99.2 (9.1)	119.9 (30.1)	
	1881-1891	Rural	Yule	91.7 (19.5)	115.5 (43.2)	107.5 (7.5)	97.7 (6.6)
			Recon.	91.5 (20.5)	116.5 (48.4)	107.6 (8.1)	97.5 (6.8)
Mixed		Yule	91.1 (18.0)	107.1 (34.4)	105.6 (8.2)	108.9 (11.8)	
		Recon.	90.7 (17.3)	107.3 (33.5)	105.8 (8.3)	108.7 (12.1)	
Urban		Yule	85.8 (20.9)	98.0 (36.3)	105.9 (7.5)	117.5 (15.9)	
		Recon.	85.3 (20.6)	97.2 (33.8)	106.5 (7.7)	117.8 (16.0)	
Metropolitan		Yule	104.7 (29.2)	90.6 (41.7)	107.7 (5.5)	111.3 (23.8)	
		Recon.	104.7 (29.8)	90.8 (42.9)	107.6 (5.6)	111.3 (24.3)	

Table 4: Correlations of Percentage Ratios

Period	Group	Source	Pauper. & Out-relief	Pauper. & Prop. Old	Pauper. & Pop.	Out-relief & Prop. Old	Out-relief & Pop.	Pop. & Prop. Old
1871-1881	Rural	Yule	0.484	0.194	-0.094	0.131	-0.079	-0.575
		Recon.	0.417	0.226	-0.060	0.089	-0.055	-0.578
	Mixed	Yule	0.412	0.090	0.038	0.018	0.097	-0.593
		Recon.	0.388	0.139	0.046	0.017	0.082	-0.544
	Urban	Yule	0.575	0.053	-0.104	0.111	-0.129	-0.431
		Recon.	0.650	0.093	-0.049	0.239	-0.075	-0.472
	Metropolitan	Yule	0.594	0.395	-0.593	0.167	-0.012	-0.528
		Recon.	0.573	0.389	-0.586	0.066	-0.009	-0.531
1881-1891	Rural	Yule	0.528	0.053	-0.001	-0.087	0.001	-0.479
		Recon.	0.504	0.228	-0.107	-0.024	0.006	-0.609
	Mixed	Yule	0.307	0.265	-0.218	-0.054	0.073	-0.555
		Recon.	0.292	0.282	-0.226	-0.075	0.093	-0.548
	Urban	Yule	0.558	0.207	-0.158	-0.148	-0.007	-0.352
		Recon.	0.521	0.218	-0.194	-0.122	-0.025	-0.320
	Metropolitan	Yule	0.516	0.412	-0.135	0.486	0.230	0.254
		Recon.	0.515	0.388	-0.134	0.471	0.227	0.252

References

Booth, C. (1894) *The Aged Poor in England and Wales*. Macmillan: London.

Plewis, I. (2015) Multiple regression, longitudinal data and welfare in nineteenth century England: reflections on Yule (1899).

Yule, G. U. (1899) An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades (Part I) (with discussion). *Journal of the Royal Statistical Society*, 62, 249 – 295.

Data Notes

1. Booth's amalgamation of small PLUs:

- (i) Kirkby Moorside included with Helmsley
- (ii) Welwyn included with Hatfield
- (iii) Buntingford included with Royston
- (iv) Penistone included with Wortley
- (v) Cerne included with Dorchester
- (vi) Hursley included with New Winchester
- (vii) Chailey, West Firle, Newhaven all included with Lewes
- (viii) Hawarden included with Chester
- (ix) Sedgefield included with Stockton
- (x) Glossop included with Hayfield
- (xi) Grimsby included with Caistor for 1891.

2. The final reconstructed dataset contains the following five population variables:

Popn1871: data from the Census of that year.

Popn1871_2: 1871 Census data from 1881, allowing for boundary changes between 1871 and 1881.

Popn1881: data from the Census of that year.

Popn1881_2: 1881 Census data from 1891, allowing for boundary changes between 1881 and 1891.

Popn1891: data from the Census of that year.

Yule's percentage ratios were based on (i) Popn1871_2 and Popn1881_2, (ii) Popn1881_2 and Popn1891.

The following adjustments were made:

(a) Popn1881_2 for Toxteth Park was clearly wrong; the data for West Derby and Toxteth Park had been swapped round.

(b) North Bierley and Bradford: North Bierley was not in the Census tables for 1871 and 1881 so Popn1871 and Popn1881 were incorrect. Popn1871_2 and Popn1881_2 were substituted and the figures for Bradford adjusted accordingly.

(c) Prestwich and Manchester: Prestwich was not in the Census tables for 1871 so Popn1871_2 was substituted and Manchester adjusted accordingly.

3. All 'indoor' variables were zero for Todmorden in 1871. The Woolwich PLU workhouse was not built in 1871 so Greenwich and Lewisham pauper data include residents of Woolwich and the ratio of 'indoor' to 'outdoor' for Woolwich for 1871 is an outlier.