Submitted: 2nd of October 2015
Published: 26th of January 2016
DOI: 10.26775/OQSPS.2016.01.26
ISSN: 2446-3868

Inequality among 32 London Boroughs: An S factor analysis

Emil O. W. Kirkegaard*

Open Quantitative Sociology & Political Science

Abstract

A dataset of 30 diverse socioeconomic variables was collected covering 32 London boroughs. Factor analysis of the data revealed a general socioeconomic factor. This factor was strongly related to GCSE (General Certificate of Secondary Education) scores (r's .683 to .786) and had weak to medium sized negative relationships to demographic variables related to immigrants (r's -.295 to -.558). Jensen's method indicated that these relationships were related to the underlying general factor, especially for GCSE (coefficients |.48| to |.69|). In multiple regression, about 60 % of the variance in S outcomes could be accounted for using GCSE and one variable related to immigrants.

Keywords: general socioeconomic factor, S factor, inequality, London, boroughs, United Kingdom, cognitive ability, IQ, intelligence, scholastic ability, GCSE, immigrants

1 Introduction

For an introduction to the study of the general socioeconomic factor (S), see earlier papers (e.g. Kirkegaard (2014b, 2015c,e); Sidhu & Pexman (2015); Kirkegaard (2014a)). This paper concerns a study of London's boroughs, similar to a previous study of census tracts of Boston (Kirkegaard, 2015d). There are 32 boroughs of London and 1 special zone (city of London).

2 Data source and selection

2.1 *London Borough Profiles*

The London DataStore (http://data.london.gov.uk/) publishes data concerning the greater London area. One dataset, *London Borough Profiles*, contains a wealth of data about the 33 administrative divisions of London. In total there are 71 fairly diverse socioeconomic measures. Listing them all would be too long, see the datafile for details (http://data.london.gov.uk/dataset/london-borough-profiles; or in the supplementary materials). Many of them however are not useful for a factor analysis, such as the name of the largest source country for immigrants or absolute number of cars.

29 variables were chosen by judgment call from the datafile for further analysis. The inclusion criteria were that the measure must measure a socioeconomically important matter and must not be heavily dependent on local natural environment. The following variables were chosen (abbreviated name used in the analysis first, full name from the datafile below):

1. Employment_rate_M
   Male employment rate (2014)

*University of Aarhus, Denmark, E-mail: emil@emilkirkegaard.dk

2. Employment_rate_F
   Female employment rate (2014)

3. Unemployment_rate Unemployment rate (2014)

4. Youth_unemployment_benefits
   Youth Unemployment (claimant) rate 18-24 (Dec-14)

5. Pct_youth_NEET
   Proportion of 16-18 year olds who are NEET (%) (2014) [NEET = "Not in Education, Employment, or Training"]

6. Pct_working_age_benefits
   Proportion of the working-age population who claim out-of-work benefits (%) (May-2014)

7. Disability
   % working-age with a disability (2014)

8. No_qualifications
   Proportion of working age people with no qualifications (%) 2014

9. Good_qualifications
   Proportion of working age with degree or equivalent and above (%) 2014

10. Annual_pay_M
    Gross Annual Pay - Male (2014)

11. Annual_pay_F
    Gross Annual Pay - Female (2014)

12. Pct_volunteering
    % adults that volunteered in past 12 months (2010/11 to 2012/13)

13. Pct_jobs_public_sector
    % of employment that is in public sector (2013)

14. Fire_rate
    Fires per thousand population (2014)

15. Ambulance_rate
    Ambulance incidents per hundred population (2014)

16. Median_house_price Median House Price, 2014

17. Home_ownership
    Homes Owned outright, (2014) %

18. Greenspace
    % of area that is Greenspace, 2005

19. Cars_rate
    Number of cars per household, (2011 Census)

20. Childcare_rate
    Rates of Children Looked After (2014)

21. Children_noWork_families
    % children living in out-of-work families (2013)

22. Life_expect_M
    Male life expectancy, (2011-13)

23. Life_expect_F
    Female life expectancy, (2011-13)

24. Teen_preg_rate
    Teenage conception rate (2013)
25. Life_satisfaction
    Life satisfaction score 2013-14 (out of 10)
26. Worthwhileness
    Worthwhileness score 2013-14 (out of 10)
27. Happiness
    Happiness score 2013-14 (out of 10)
28. Anxiety
    Anxiety score 2013-14 (out of 10)
29. Voter_turnout
    Turnout at 2014 local elections

Some of the variables are closely related, e.g. by gender. This was done on purpose to check whether they correlated strongly or not. The redundancy analysis will catch problematically closely correlated variables (Section 5).

Variables 25-28 were measured as part of the *Annual Population Survey* and based on self-ratings. The data is available in more detail at http://data.london.gov.uk/dataset/subjective-personal-well-being-borough.

More details on the variables can be found in the dataset in the supplementary materials (*london-borough-profiles.xls*).

2.2 Recorded Crime Summary Data

The London Datastore also publishes another dataset *Recorded Crime Summary Data for London: Borough Level*. http://data.london.gov.uk/dataset/recorded-crime-summary-data-london-borough-level. The dataset contains numbers of every recorded crime by month, the broad type of crime, and detailed type of crime and in which Borough it occurred. The types of crime are:

1. Burglary
   a) Burglary In A Dwelling
   b) Burglary In Other Buildings
2. Criminal Damage
   a) Criminal Damage To Dwelling
   b) Criminal Damage To Motor Vehicle
   c) Criminal Damage To Other Building
   d) Other Criminal Damage
3. Drugs
   a) Drug Trafficking
   b) Other Drugs
   c) Possession Of Drugs
4. Fraud & Forgery
   a) Counted Per Victim
   b) Other Fraud & Forgery
5. Other Notifiable Offences

   a) Going Equipped
   b) Other Notifiable
6. Robbery
   a) Business Property
   b) Personal Property
7. Sexual Offences
   a) Other Sexual
   b) Rape
8. Theft & Handling
   a) Handling Stolen Goods
   b) Motor Vehicle Interference & Tampering
   c) Other Theft
   d) Other Theft Person
   e) Theft From Motor Vehicle
   f) Theft From Shops
   g) Theft Person
   h) Theft/Taking Of Motor Vehicle
   i) Theft/Taking Of Pedal Cycle
9. Violence Against The Person
   a) Assault With Injury
   b) Common Assault
   c) Harassment
   d) Murder
   e) Offensive Weapon
   f) Other Violence
   g) Wounding/GBH (Grievous bodily harm)

The dataset listed on the website always covers the latest 2 years. At the time of writing, this means that it covers the period 2013-09 to 2015-08. However, I located an older version of the dataset covering the period 2011-04 to 2013-03 as well, which I combined with the first dataset. The types of crime overlapped almost perfectly.

The crime data was given in absolute numbers. I used the total population variable from the London Borough Profiles dataset to calculate a per capita version.

3 Preliminary analysis of crime data

Because the population of City of London was very small, this case was excluded due to excessive sampling error.

The London Borough Profiles also contains a variable covering crime, namely total crime rate per capita. As a preliminary analysis, I carried out factor analyses at both the broad and detailed type-levels using both standard form and rank form data and correlated the extracted factor scores with the crime rate given found in the London Borough Profiles dataset, also in standard form and in rank form. The use of case ranks is to reduce effects of outliers (Kirkegaard, 2016).

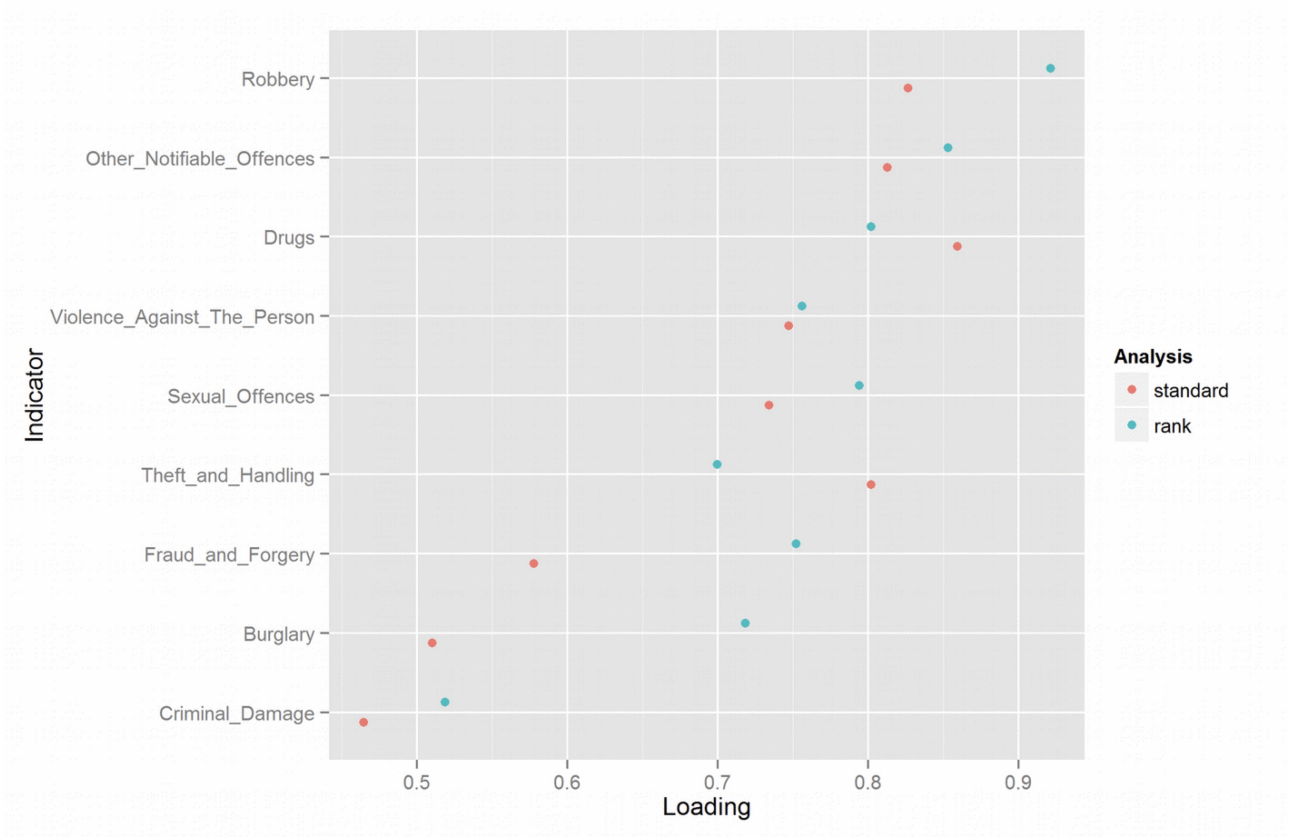

Figure 1: Factor loadings on the general crime factor using broad crime types.

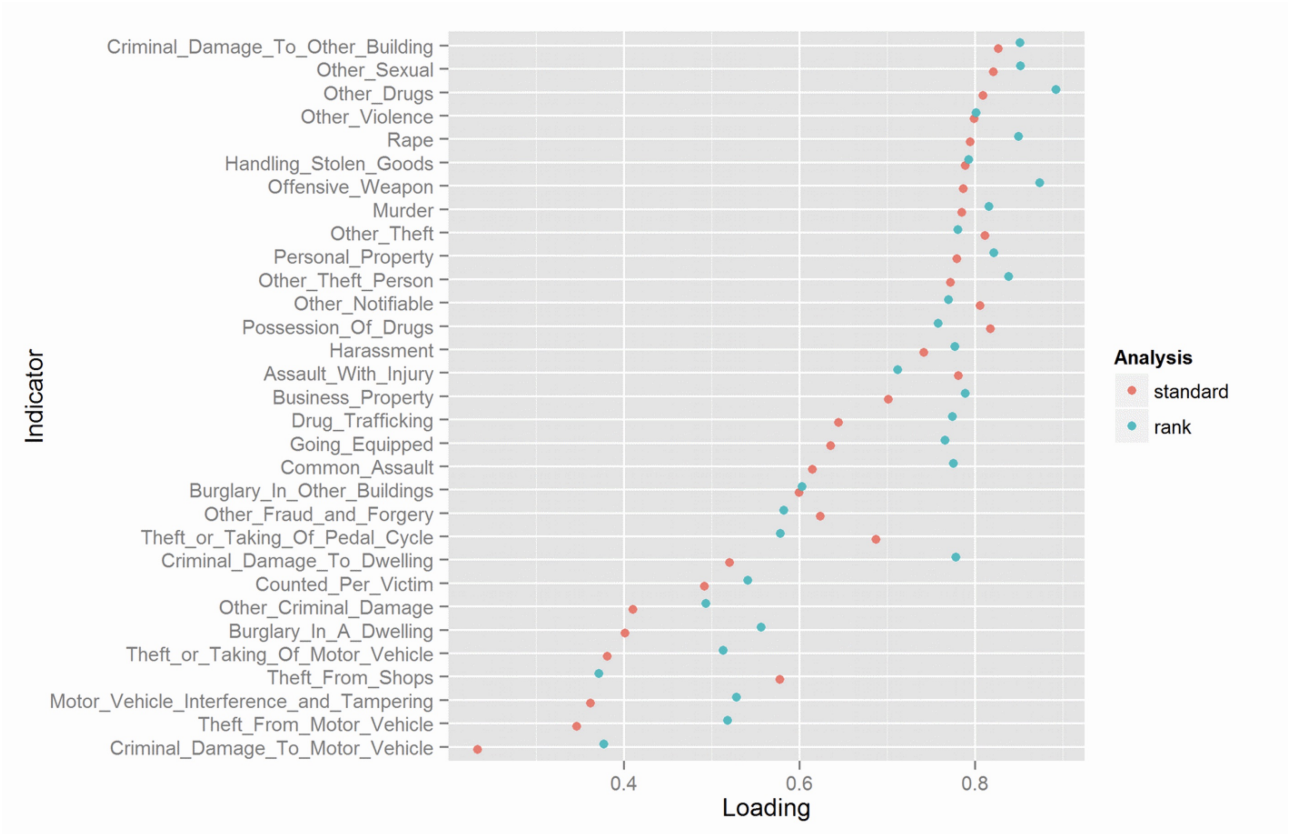

Figure 2: Factor loadings on the general crime factor using specific crime types.

Before the factor analysis was run, I excluded variables that correlated at |r| > .9 with another variable to reduce the 'coloring' of the general factor by group factor variance (Kirkegaard, 2016). This resulted in the removal of two variables (Theft_Person, Wounding/GBH) in the case of the analysis using detailed types.

Figures 1 and 2 show the loadings in the two factor analyses.

All loadings were positive at both levels. Results using standard form and rank order form were similar. Congruence coefficients were .99 in both cases.

Table 1 shows the intercorrelations.

Table 1: Intercorrelations between general crime factors and total crime rate using standard form and rank form.

	broad_std	broad_rank	fine_std	fine_rank	crime_rate_std	crime_rate_rank
broad_std	1	0.944	0.969	0.907	0.744	0.691
broad_rank	0.944	1	0.947	0.96	0.567	0.664
fine_std	0.969	0.947	1	0.956	0.718	0.732
fine_rank	0.907	0.96	0.956	1	0.561	0.711
crime_rate_std	0.744	0.567	0.718	0.561	1	0.792
crime_rate_rank	0.691	0.664	0.732	0.711	0.792	1

As expected, all correlations were strong. However, the correlations between the extracted scores and the total crime rate were surprisingly weak, only around .66 to .79.

4 Missing data

For the main analysis, the detailed crime type dataset was combined with the selected socioeconomic variables to form a dataset with 32 cases and 62 variables. The City of London was excluded as before.

A few cases had 1 or 2 missing datapoints which were imputed. This was done using deterministic imputation with the *irmi()* function from the **VIM** package (Templ et al., 2015). Deterministic imputation was used to make the results reproducible and because there was not a strong interest in calculating precise standard errors (Donders et al., 2006).

5 Redundant variables

To avoid 'coloring' the extracted general factor, a redundancy analysis was run (Kirkegaard, 2016). This resulted in the same two variables being excluded as in the preliminary analysis. Thus, the final dataset contained 60 variables.

6 Stability across extraction and scoring methods

S factors can be unstable across methods of extraction (Kirkegaard, 2015b). For this reason an S factor was extracted using every combination of extraction and scoring method available in the *fa()* function in the **psych** package (Revelle, 2015). All methods gave nearly identical results, all r > .99. In line with previous studies, Bartlett's scoring method (and least squares extraction/MinRes) was used for all the following analyses.

7 Indicator sampling reliability

Because we don't have access to all possible indicators of S for London boroughs, there is necessarily some sampling error (indicator sampling error). A general factor can be considerably 'colored' if an unrepresentative sampling of indicators is chosen. To examine this, the dataset was repeatedly split up randomly into two sets of indicators and factor analysis was run independently within the two subsamples (Kirkegaard, 2016). Then the

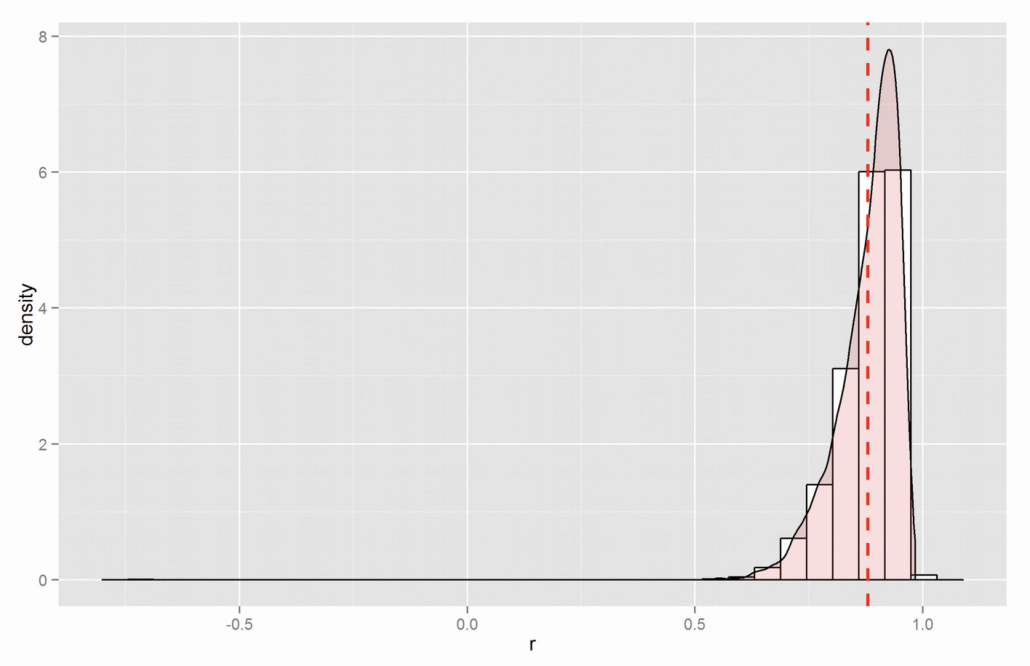

Figure 3: Distribution of indicator sampling reliability correlations. N=5000.

factor scores from the one subsample was correlated with those from the other and the obtained correlation saved. Figure 3 shows the distribution of the obtained correlations.

In about 1 out of 5000 of the runs, the extracted factors were reversed. If this is accounted for (using absolute values), the mean split-half reliability was .88 indicating that indicator sampling error was a small problem for this dataset.

8 Mixedness

Mixedness/structural outlierness was examined using five measures:

1) mean absolute residual,

2) change in factor size,

3) absolute change in factor size,

4) mean absolute change in factor loadings,

5) max absolute change in factor loadings (Kirkegaard, 2016).

This analysis revealed that Westminster was a moderate outlier. Using the indicator residuals, one can see which variables the case is an outlier in. The variables for which Westminster has a |residual| > 2.5 are: Theft_From_Shops (3.85), Other_Theft (3.12), Ambulance_rate (2.84), Other_Notifiable (2.76) and Annual_pay_M (2.64). Three of these are crime variables and another is plausible a consequence of violent crime (Ambulance_rate). Thus it seems that the case is highly mixed because crime rates are generally higher than one would expect based on the other indicators. Westminster is a tourist area with many shops so this may explain the oddity.

A parallel dataset without the outlier was created.

9 Main factor analyses

Three main factor analyses were run:

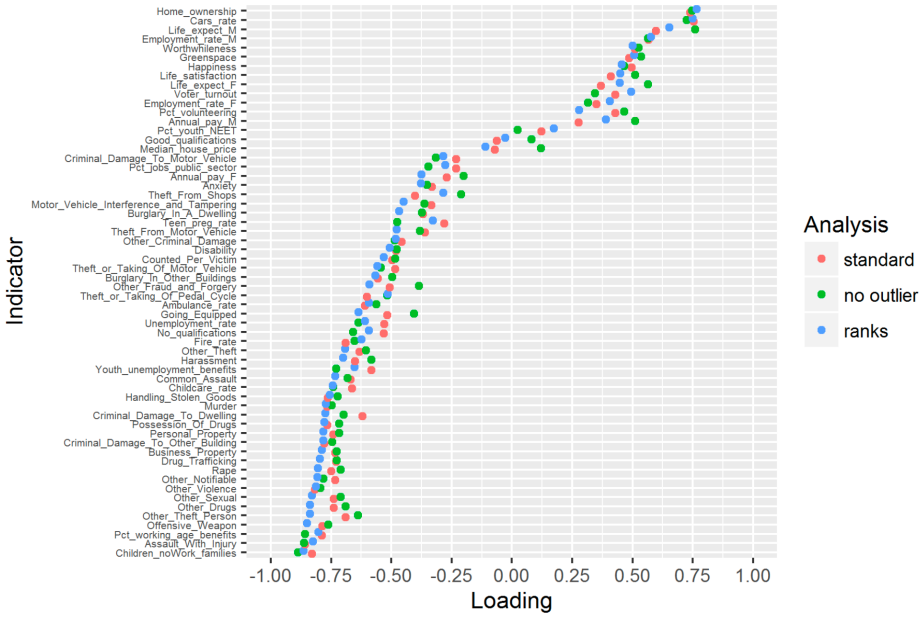

Figure 4: Factor loadings of the S factor.

1) classical/standard factor analysis with Westminster,

2) the same but without Westminster, and

3) factor analysis based on the rank-order transformed dataset.

The purpose of the last method is to not rely on the assumption that the indicators are normally distributed, which is often not the case (Kirkegaard, 2016).

Figure 4 shows the loadings from the three analyses.

In general, loadings were very similar across analyses. Factor congruence coefficients were all .99 (Lorenzo-Seva & Ten Berge, 2006).

A few variables had loadings in the direction opposite to that expected. Female annual pay had negative loadings between -20 and -.37. This is curious because male annual pay had positive loadings of .28 to . 51. The correlation of the two annual pay variables was, however, only .51. It is uncommon to see variables diverge by gender this much, so further research should look into the matter.

Pct_youth_NEET had small positive loadings between .02 and .17. This is strange because Youth_unemployment_benefits had strong negative loadings -.58 and -.73 and they seem to measure something similar. Their correlation is only .22.

Good qualification had loadings between -.06 and .08 whereas a stronger positive loading was expected. Strangely, the opposite variable No_qualifications did show strong negative loadings between -.59 and -.66. The correlation between the two is -.58.

Median_house_price had a strong positive loading around .80 in a prior S analysis of Boston census districts (Kirkegaard, 2015d), but in the present analysis the loading was near zero: -.11, to .12.

10 Criteria analysis

The dataset contains a few variables one might expect to have causal relationships to the S factor. The variables are:

- % of resident population born abroad (2014)

- % of population from BAME groups (2013)
  "BME/BAME – Black and Minority Ethnic or Black, Asian and Minority Ethnic is the terminology normally used in the UK to describe people of non-white descent." (Institute of Race Relations, 2015).
- % people aged 3+ whose main language is not English (2011 census)
- % of pupils whose first language is not English (2014)
- All Pupils at the End of KS4 Achieving 5+ A* - C Including English and Mathematics

All of which are demographic variables except the last which is scholastic. A reviewer (Noah Carl) pointed out some problems with using this variable. Instead, a related variable (Average GCSE and Equivalent Point Score Per Pupil at the End of KS4) from another dataset, GCSE Results by Location of Pupil Residence, Borough (2013/2014) http://data.london.gov.uk/dataset/gcse-results-location-pupil- residence-borough (also available in supplementary materials), was used. See Appendix A for more details.

In line with much other research (Herrnstein & Murray, 1994; Kirkegaard & Tranberg, 2015), one would expect higher cognitive ability to lead to higher S. The GCSE scores are not exactly an IQ test (16,17), but it has been found that at the national-level, scholastic ability and cognitive ability as measured by traditional IQ tests are nearly perfectly correlated (18,19). This suggests that the GCSE scores may serve as a good proxy for cognitive ability at the borough-level, but it need not be the case. Prior research using similar data has found strong relationships between scholastic/cognitive ability and S, so a correlation in the vicinity of .40 to .90 would be expected here.

The expected outcome of the immigrant variables depends on which kind of immigrants it is: where they are from and how they were selected (Kirkegaard, 2014a; Kirkegaard & Fuerst, 2014). Unfortunately, little information is given regarding the countries of origin. The datafile does contain information regarding the three largest countries of origin and their percentages for each borough, but because the countries vary from case to case, one cannot construct a variable with complete data from this. For instance, for Camden, the three largest source countries are United States (2.8 %), Bangladesh (2.7 %) and Ireland (2.4 %), while for Lambeth the they are Jamaica (3.2 %), Portugal (2.3 %) and Poland (2.3 %). Because the composition of the immigrants vary from case to case and because immigrants vary so much in their socioeconomic performance, no strong predictions regarding the correlations can be made.

One would however expect the non-English speaking immigrants fare worse in society than the English speaking immigrants because immigrants from English-speaking countries have been found to perform well in Denmark, Norway, Finland and the Netherlands (Kirkegaard, 2014a; Kirkegaard & Fuerst, 2014; Kirkegaard, 2015a). Since the born-abroad variable includes all immigrants and the other three variables includes mainly or exclusively those who don't speak English, the correlation are expected to be more negative for the latter. This does not imply that the correlations are expected to be negative. That fact depends on the average performance of the immigrant population, which can be either better or worse than the natives (Rindermann & Thompson, 2016).

10.1 Correlations

The first step is to examine the correlations. Table 2 shows the correlations between the criteria variables and the S scores.

Table 2: Correlations between S scores and criteria variables. Weighted correlations below the diagonal. The square root of population was used as weight.

	S	S_no	S_rank	GCSE	born_abroad	Pct_BAME	NonEnglish	Pupils_nonEnglish
S		0.996	0.983	0.693	−0.372	−0.511	−0.48	−0.539
S_no	0.996		0.99	0.786	−0.295	−0.558	−0.442	−0.49
S_rank	0.983	0.99		0.724	−0.354	−0.545	−0.492	−0.539
GCSE	0.683	0.774	0.714		−0.177	−0.404	−0.279	−0.313
born_abroad	−0.374	−0.302	−0.358	−0.179		0.706	0.9	0.883
Pct_BAME	−0.504	−0.551	−0.539	−0.393	0.726		0.808	0.799
NonEnglish	−0.471	−0.437	−0.485	−0.277	0.903	0.813		0.952
Pupils_nonEnglish	−0.524	−0.479	−0.526	−0.302	0.886	0.8	0.951	

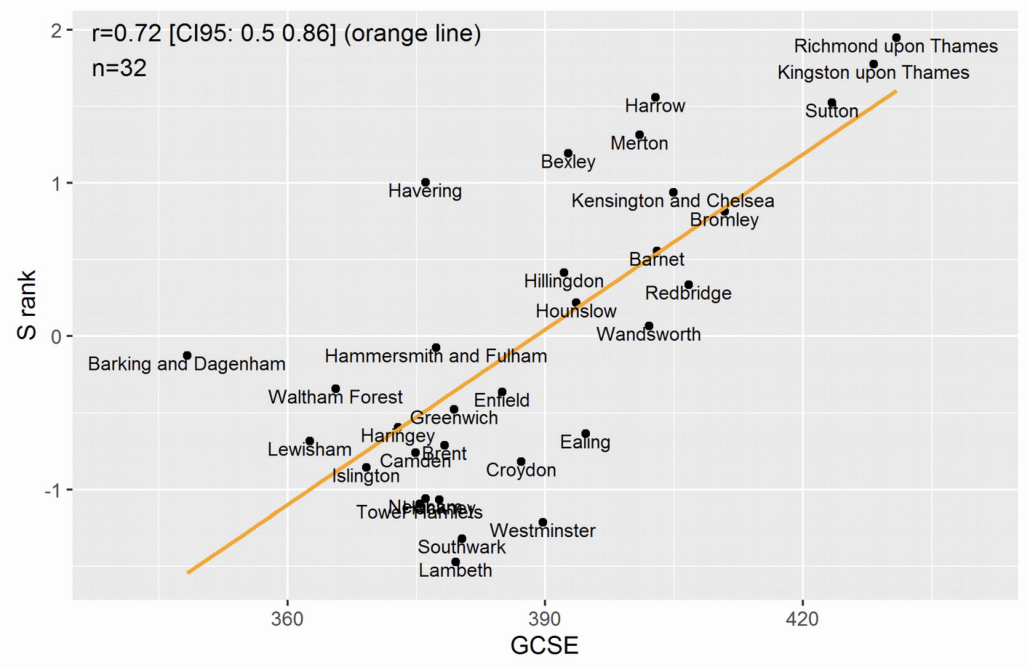

Figure 5: Scatter plot of S_rank and GCSE. Unweighted.

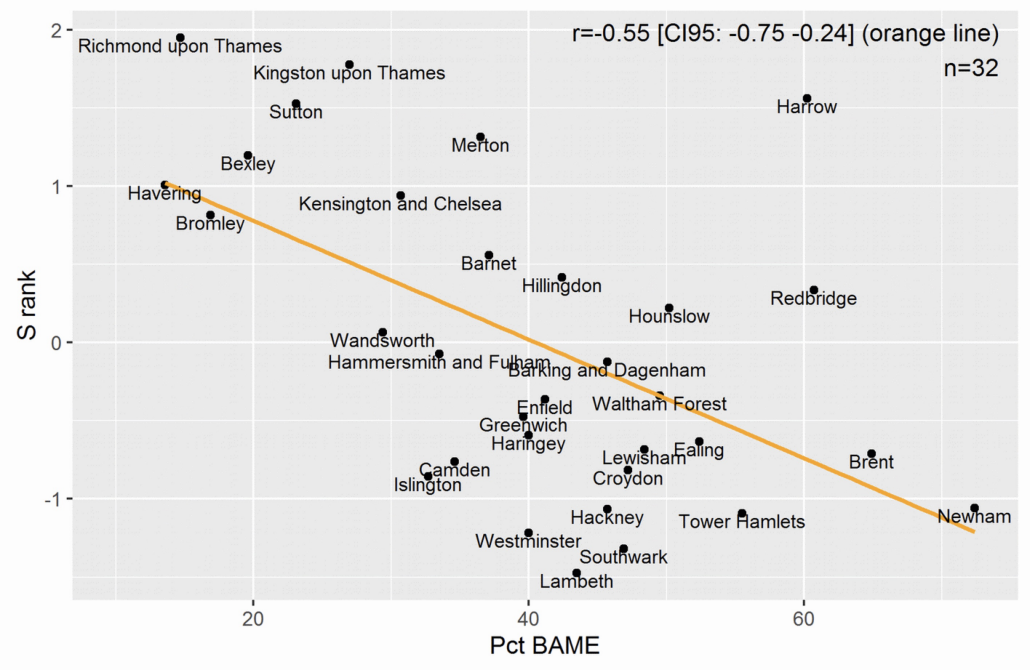

Figure 6: Scatter plot of S_rank and Pct_BAME. Unweighted.

As expected, the S scores correlate very strongly across methods. They are also very strongly related to GCSE scores. The correlations are almost identical when using weights. The immigrant variables have weak to medium sized negative correlations to S scores and the expected relative values of born abroad versus the others was found. Figure 5 shows a scatter plot of the relationship between S and GCSE, and Figure 6 between S and Pct_BAME.

The immigrant variables show somewhat weaker relationships to the GCSE. One model of the relationship posits a negative causal effect from immigrants to S which is mostly mediated by cognitive performance. Such a model can be tested with semi-partial correlations. First S is regressed on GCSE and the residuals extracted. These are then correlated with the immigrant variables. A complete mediation model predicts that the semi-partial correlations should be near 0. Table 3 shows the results.

Table 3: Semi-partial correlations for S, GCSE and immigrant variables. Weighted by square root of population.

Variable	Orig. cor	Semi-partial cor
born_abroad	−0.358	−0.329
Pct_BAME	−0.539	−0.369
NonEnglish	−0.485	−0.41
Pupils_nonEnglish	−0.526	−0.444

The semi-partial correlations are only a little smaller, suggesting only weak mediation. This could indicate that immigrants have a negative effect that is not related to their lower cognitive ability. Alternatively, it may mean that the GCSE variable has measurement bias in favor of immigrants. Further research is needed on this question.

10.2 Multiple regression

Given the fairly low intercorrelations between the immigrant variables and the cognitive variable, one would expect that they could be fruitfully combined in multiple regression. To see if this was the case, I fitted all the possible models using the GCSE and the 4 immigrant variables (best subsets method; 31 models). The best models judging from AIC, BIC and R^2 adj. included the GCSE and one or two immigrant variables. The top 5 models according to R^2 adj. are shown in Table 4.

Table 4: Top 5 model fits for S_rank.

Model#	GCSE	born_abroad	Pct_BAME	NonEnglish	Pupils_nonEnglish	AIC	BIC	r2.adj.
18	0.595	0.257			−0.566	66.705	74.033	0.59
9	0.621				−0.333	65.896	71.759	0.589
27	0.577	0.272	−0.1		−0.505	68.405	77.2	0.579
20	0.608		−0.074		−0.278	67.736	75.064	0.577
21	0.62			0.093	−0.423	67.822	75.151	0.576

The best models explained about 60 % of the variance. Interestingly, model #18 posits a positive beta for being born abroad. The correlation for this variable itself with the S_rank is negative (-.35). The best subsets method, however, tends to overfit the data (James et al., 2013). For this reason, lasso regression was also used. The tuning (shrinkage) parameter was chosen by cross-validation as recommended by James et al. (2013). The fitting procedure was run 500 times. Table 5 shows the results:

Table 5: Lasso regression results for S_rank. N_{runs}=500.

	GCSE	born_abroad	Pct_BAME	NonEnglish	Pupils_nonEnglish
mean	0.405	0	−0.008	0	−0.121
median	0.4	0	−0.005	0	−0.118
sd	0.024	0	0.006	0	0.02
fraction zero	0	1	0.198	1	0.002

We see that GCSE was always a useful predictor, while born_abroad and NonEnglish were never. Pct_BAME and Pupils_nonEnglish were useful predictors most of the time.

10.3 Jensen's method

For a discussion of Jensen's method applied to S factor data, see Kirkegaard (2016). Figure 7 shows Jensen's method applied to the S x GCSE relationship. The S derived from rank-ordered data was used because this was considered the best estimate that included all cases.

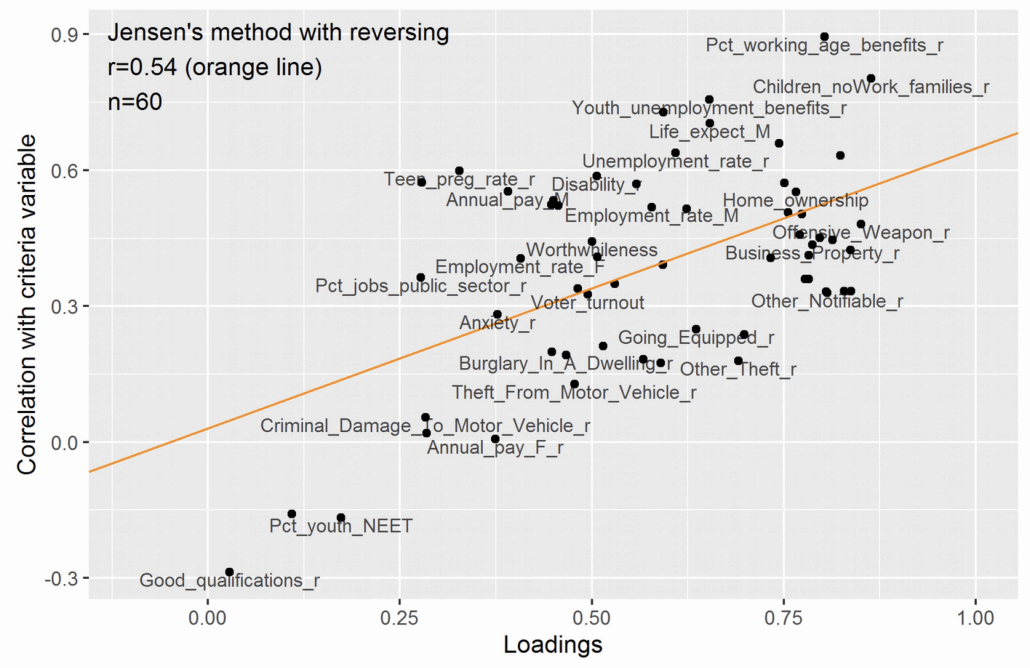

Figure 7: Jensen's method for S x GCSE.

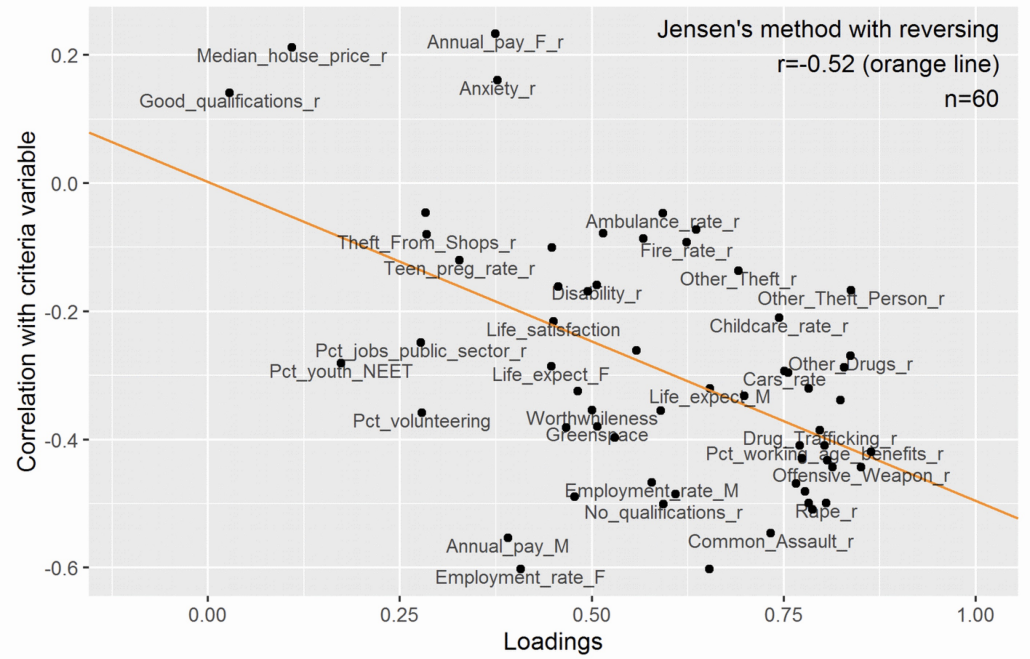

Figure 8: Jensen's method for S x BAME %.

Figure 8 shows that for S x BAME %.

The results using the other S estimates were similar, a bit stronger for S_no. In general, results indicate that the criteria variables are probably related to the underlying S factor to a substantial degree but that there are strong group factors in the dataset.

11 Discussion and conclusion

Overall results were similar to previous S factor studies. Only one prior study has examined administrative divisions of a major city before and it too found similar results (Kirkegaard, 2015d). The odd finding of a negative loading for female annual pay in the rank-order analysis deserves more attention in the future, as does the general case where gendered versions of a variable give markedly different results.

11.1 Limitations

The cognitive variable was a scholastic variable, which can bias the results in an unknown direction.

The demographic variables were not optimal. It would be better if the size of each country of origin could be found for all boroughs because this would allow one to make quantitative predictions for how immigrant % would be related to S and GCSE/cognitive scores (Kirkegaard & Tranberg, 2015; Kirkegaard, 2013).

Supplementary material and acknowledgments

All data, figures and R source code available at the OSF repository: https://osf.io/f4uc2/

Thanks to Noah Carl, Kenya Kura and L.J. Zigerell for reviewing this paper. Peer review can be found at: http://openpsych.net/forum/showthread.php?tid=249

See also discussion in the peer review thread: http://openpsych.net/forum/showthread.php?tid=249

References

Donders, A., van der Heijden, G., Stijnen, T., & KGM, M. (2006, October 10). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 1087—1091. doi: 10.1016/j.jclinepi.2006.01.014

Herrnstein, R., & Murray, C. (1994). *The bell curve: Intelligence and class structure in american life*. New York: Free Press.

Institute of Race Relations. (2015, September 9). *Definitions*. Retrieved from http://www.irr.org.uk/research/statistics/definitions/

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. New York: Springer.

Jensen, A. (1998). *The g factor: the science of mental ability*. Westport, Conn.: Praeger.

Kirkegaard, E. O. W. (2013). Predicting immigrant iq from their countries of origin, and lynn's national iqs: A case study from denmark. *Mankind Quarterly*, *54*(2), 151–167. doi: 10.46469/mq.2013.54.2.2

Kirkegaard, E. O. W. (2014a, oct). Crime, income, educational attainment and employment among immigrant groups in norway and finland. *Open Differential Psychology*. Retrieved from https://doi.org/10.26775%2Fodp.2014.10.09 doi: 10.26775/odp.2014.10.09

Kirkegaard, E. O. W. (2014b, sep). The international general socioeconomic factor: Factor analyzing international rankings. *Open Differential Psychology*. Retrieved from https://doi.org/10.26775%2Fodp.2014.09.08 doi: 10.26775/odp.2014.09.08

Kirkegaard, E. O. W. (2015a, oct). Crime among dutch immigrant groups is predictable from country-level variables. *Open Differential Psychology*. Retrieved from https://doi.org/10.26775%2Fodp.2015.10.04 doi: 10.26775/odp.2015.10.04

Kirkegaard, E. O. W. (2015b, April 19). Examining the S factor in Mexican states. *The Winnower [Internet]*. Retrieved from https://thewinnower.com/papers/examining-the-s-factor-in-mexican-states

Kirkegaard, E. O. W. (2015c, June 26). IQ and socioeconomic development across Regions of the UK: a reanalysis. *The Winnower [Internet]*. Retrieved from https://thewinnower.com/papers/1419-iq-and-socioeconomic-development-across-regions-of-the-uk-a-reanalysis

Kirkegaard, E. O. W. (2015d). An S factor among census tracts of Boston. *The Winnower*. Retrieved from https://thewinnower.com/papers/an-s-factor-among-census-tracts-of-boston

Kirkegaard, E. O. W. (2015e, March 28). The S factor in the British Isles: A reanalysis of Lynn (1979). *The Winnower [Internet]*. Retrieved from https://thewinnower.com/papers/the-s-factor-in-the-british-isles-a-reanalysis-of-lynn-1979

Kirkegaard, E. O. W. (2016, nov). Some new methods for exploratory factor analysis of socioeconomic data. *Open Quantitative Sociology & Political Science*. Retrieved from https://doi.org/10.26775%2Foqsps.2016.11.07 doi: 10.26775/oqsps.2016.11.07

Kirkegaard, E. O. W., & Fuerst, J. (2014, may). Educational attainment, income, use of social benefits, crime rate and the general socioeconomic factor among 70 immigrant groups in denmark. *Open Differential Psychology*. Retrieved from https://doi.org/10.26775%2Fodp.2014.05.12a doi: 10.26775/odp.2014.05.12a

Kirkegaard, E. O. W., & Tranberg, B. (2015, mar). Increasing inequality in general intelligence and socioeconomic status as a result of immigration in denmark 1980-2014. *Open Differential Psychology*. Retrieved from https://doi.org/10.26775%2Fodp.2015.03.03 doi: 10.26775/odp.2015.03.03

Lorenzo-Seva, U., & Ten Berge, J. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64.

Revelle, W. (2015). Procedures for Psychological, Psychometric, and Personality Research. *psych*. Retrieved from http://cran.r-project.org/web/packages/psych/index.html

Rindermann, H., & Thompson, J. (2016). The cognitive competences of immigrant and native students across the world: An analysis of gaps, possible causes and impact. *Journal of Biosocial Science*, *48*(1), 66–93. doi: 10.1017/S0021932014000480

Sidhu, D., & Pexman, P. (2015, May 27). What's in a Name? Sound Symbolism and Gender in First Names. *PLoS ONE*, *10*(5), e0126809.

Templ, M., Alfons, A., Kowarik, A., & Prantner, B. (2015). Vim: Visualization and imputation of missing values. *CRAN*. Retrieved from http://cran.r-project.org/web/packages/VIM/index.html

Appendix A Scholastic variables

The dataset, GCSE Results by Location of Pupil Residence, Borough http://data.london.gov.uk/dataset/gcse-results-location-pupil-residence-borough, contains a datafile with a number of scholastic variables:

1. All Pupils at the End of KS4 Achieving 5+ A* - C
2. All Pupils at the End of KS4 Achieving 5+ A* - G
3. All Pupils at the End of KS4 Achieving 5+ A* - C Including English and Mathematics
4. All Pupils at the End of KS4 Achieving 5+ A* - G Including English and Mathematics
5. All Pupils at the End of KS4 Achieving the Basics
6. All Pupils at the End of KS4 Entering the English Baccalaureate
7. All pupils at the End of KS4 Achieving the English Baccalaureate
8. Average GCSE and Equivalent Point Score Per Pupil at the End of KS4
9. Average Capped GCSE and Equivalent Point Score Per Pupil at the End of KS4

The datafile also contains the gendered versions of the above as well as total counts of students by borough.

All the variables concern grades but some are non-continuous transformations of other variables. Variables (1-7) all appear to be variables concerning students meeting certain minimum thresholds defined in various ways. Donders et al. (2006) seems to be a capped version of Templ et al. (2015). Templ et al. (2015) seems to be the closest to a mean total score similar to other scholastic measures such as PISA or TIMMS.

In general, all variables can be expected to measure the same construct of scholastic ability, so one could factor analyze them, which I did. All variables load on a general factor with loadings from .69 to .98.

Since we have 9 variables measuring the same construct, one could use Jensen's method (method of correlated vectors; (Jensen, 1998)), to see if there is a relationship between being a better measure of the underlying construct and being related to the main criteria variable, S. This is in fact the case, Jensen coefficients are .94 and .95 depending on which vector of S scores is used as the criterion. Furthermore, the GCSE variable with the strongest correlation to the criterion variable is Templ et al. (2015). This variable was chosen for use in the main analysis.