



# The S factor in China

EMIL O. W. KIRKEGAARD

READ REVIEWS

WRITE A REVIEW

**CORRESPONDENCE:**

[emil@emilkirkegaard.dk](mailto:emil@emilkirkegaard.dk)

**DATE RECEIVED:**

June 10, 2015

© Kirkegaard This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



## Abstract

I analyze the S factor in Chinese states using data obtained from Lynn and Cheng as well as new data obtained from the Chinese statistical agency. I find that S correlates .42 with IQ and .48 with ethnic Han%.

## Introduction

Richard Lynn has been publishing a number of papers on IQ in regions/areas/states within countries along with various socioeconomic correlates. However, usually his and co-authors analysis is limited to reporting the correlation matrix. This is a pity, because the data allow for a more interesting analysis with the S factor (see Kirkegaard, 2014). I have previously re-analyzed Lynn and Yadav (2015) [in a blogpost](#) to be published in a journal sometime 'soon'. In this post I re-analyze the data reported in Lynn and Cheng (2013) as well as more data I obtained from the [official Chinese statistical database](#).

## Data sources

In their paper, they report 6 variables: 1) IQ, 2) sample size for IQ measurement, 3) % of population Ethnic Han, 4) years of education, 5) percent of higher education (percent with higher education?), and 6) GDP per capita. This only includes 3 socioeconomic variables — the bare minimum for S factor analyses — so I decided to see if I could find some more.

I spent some time on the database and found various useful variables:

- Higher education per capita for 6 years
- Medical technical personnel for 5 years
- Life expectancy for 1 year
- Proportion of population illiterate for 9 years
- Internet users per capita for 10 years
- Invention patents per capita for 10 years
- Proportion of population urban for 9 years
- Scientific personnel for 8 years

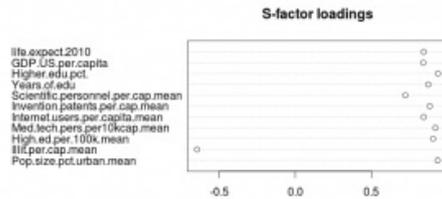
I used all available data for the last 10 years in all cases. This was done to increase reliability of the measurement, just in case there was some and reduce transient effects. In general tho regional differences were very consistent thruout the years, so this had little effect. One could do factor analysis and get the factor scores, but this would make the score hard to understand for the reader.

For the variable with data for multiple years, I calculated the average yearly intercorrelation to see how reliable the measure were. In all but one case, the average intercorrelation was  $\geq .94$  and the last case

it was .86. There would be little to gain from factor analyzing these data and using the scores instead of just averaging the years preserves interpretable data. Thus, I averaged the variables for each year to produce one variable. This left me with 11 socioeconomic variables.

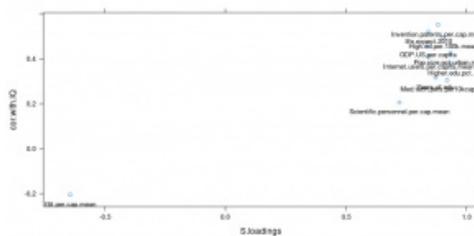
**Examining the S factor and MCV**

Next step was to factor analyze the 11 variables and see if one general factor emerged with the right direction of loadings. It did in fact, the loadings are as follows:



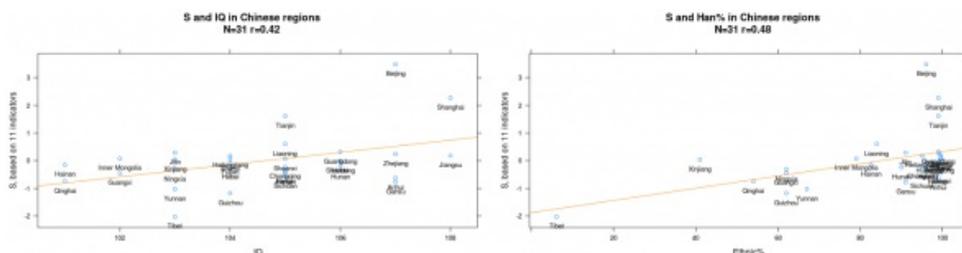
All the loadings are in the expected direction. Aside from the one negative loading (illiteracy), they are all fairly strong. This means that MCV (method of correlated vectors) analysis is rather useless, since there is little inter-loading variation. One could probably fix this by going back to the databank and fetching some variables that are worse measures of S and that varies more.

Doing the MCV anyway results in  $r=.89$  (inflated by the one negative loading). Excluding the negative loading gives  $r=.38$ , which is however solely due to the scientific personnel datapoint. To properly test it, one needs to fetch more data that varies more in its S loading.



**S and IQ and Han%**

We are now ready for the two main results, i.e. correlation of S with IQs and % ethnic Han.



Correlations are of moderate strength,  $r=.42$  and  $r=.48$ . This is somewhat lower than found in analyses of Danish and Norwegian immigrant groups (Kirkegaard 2014,  $r$ 's about .55) and much lower than that found between countries ( $r=.86$ ) and lower than that found in India ( $r=.61$ ). The IQ result is mostly due to the two large cities areas of Beijing and Shanghai, so the results are not that convincing. But they are tentative and consistent with previous results.

Han ethnicity seems to be a somewhat more reasonable predictor in this dataset. It may not be due to higher general intelligence, they may have some other qualities that cause them to do well. Perhaps more conscientious, or more rule-conforming which is arguably rather important in authoritarian

societies like China.

### **Supplementary material**

The R code and datasets are available at the *Open Science Foundation repository* for this study.

### **References**

Kirkegaard, E. O. W., & Fuerst, J. (2014). Educational attainment, income, use of social benefits, crime rate and the general socioeconomic factor among 71 immigrant groups in Denmark. *Open Differential Psychology*.

Kirkegaard, E. O. W. (2014). The international general socioeconomic factor: Factor analyzing international rankings. *Open Differential Psychology*.

Lynn, R., & Cheng, H. (2013). Differences in intelligence across thirty-one regions of China and their economic and demographic correlates. *Intelligence*, 41(5), 553-559.