Comparative and Absolute Advantage:
Ability in the Data and in the Model

May 25, 2021

In Project TALENT and in the two iterations of the NLSY, each survey respondent takes a battery of cognitive tests and we have test scores for three types of tests:

1. mathematics,

2. verbal (word knowledge and comprehension), and

3. social.

O*NET has skill requirements for each occupation and we are focusing on the skills corresponding most closely to the cognitive test scores, namely:

1. mathematics,

2. writing, and

3. social skills.

We'll label the three components m, w, and s, respectively.

Each individual is characterized by a triple (a_m, a_w, a_s) based on her test scores, which are normalized to the unit interval $[0, 1]$ to facilitate comparisons across cohorts in different surveys. Effectively, the triple keeps track of an individual's rank in each individual component.

Similarly, each occupation j is characterized by a vector of skill requirements (b_m, b_w, b_s) based on the O*NET descriptors.

Jobs are differentiated vertically as well as horizontally. For instance, some occupations have higher requirements in all three dimensions compared to other jobs. They require more skill

in an *absolute* sense. Other jobs have a different mix of requirements. For instance, one occupation may have low requirements for mathematics and high requirements for writing / communication while another job may have a "reverse" profile. This differentiation is mostly horizontal and hence the requirements vary in a *comparative* sense.

So how can we assign occupation-specific abilities to each individual with the test scores we have? In a nutshell, we are proposing a distance measure between an individual's test score profile and the skill requirement of each occupation that takes into account absolute as well as comparative differentiation.

Without loss of generality (at least I think so), the mathematical skill and requirement pair is the reference dimension in our measure of distance between individual i's skills and occupation j's requirements.

$$\overline{a}_i^j = \ln\left(\frac{a_{i,m}}{b_{j,m}}\right) - \left|\ln\left(\frac{a_{i,w}/b_{j,w}}{a_{i,m}/b_{j,m}}\right)\right| - \left|\ln\left(\frac{a_{i,s}/b_{j,s}}{a_{i,m}/b_{j,m}}\right)\right| \tag{1}$$

This measures the distance between an individual's skill set relative to the requirements of a particular occupation. The support of this measure is the real line.

For the purposes of our quantitative exercise, we need to normalize this measure such that the support is $\mathbb{R}^+$. I'm not sure what the best way to do this is.