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

May 27, 2021

In Project TALENT and in the two NLSY cohorts, each survey respondent takes a battery of cognitive test and attitudinal scales that measure a respondent's non-cognitive abilities. Table 1 presents test components that we use to construct math, verbal, and social abilities that can be linked to skill counterparts. For each survey, we reduce the components into three composite dimensions using principal component analysis. Then, these composites are converted into percentile ranks among individuals.

| | Project TALENT | NLSY79 | NLSY97 |
|---|---|---|---|
| Math ability | Mathematics composite | Arithmetic reasoning<br>Mathematics knowledge | Arithmetic reasoning<br>Mathematics knowledge |
| Verbal ability | Verbal composite | Word knowledge<br>Paragraph comprehension | Word knowledge<br>Paragraph comprehension |
| Social ability | Impulsiveness, calmness, self-confidence, mature personality | Rotter locus of control scale<br>Rosenberg self-esteem scale | Organized, conscientious, dependable, thorough, trusting, disciplined, careful |

Table 1: Ability measures

O*NET contains information on ability, knowledge, and skills that characterize at total of 974 different occupations. Each of these occupations has scores for the importance of 277 descriptors. We are focusing on the occupational skill requirements corresponding to math, verbal, and social abilities.

Following Guvenen et al. (2020), we use 32 descriptors that are most closely related to our ability measures. First, we convert 26 O*NET skills into *Arithmetic Reasoning*, *Mathematics Knowledge*, *Word Knowledge*, and *Paragraph Comprehension* from ASVAB using the crosswalk created by the Defense Manpower Data Center. Then, we normalize the four components and reduce them into math and verbal skill requirements using principal component analysis. Next, we reduce six O*NET descriptors into a single social skill requirement using principal component analysis. Finally, these three composites are converted into percentile ranks among occupations.

We label the three dimensions of job requirements and worker skill with subscripts $m$ (math), $v$ (verbal), and $s$ (social), respectively.

Each individual is characterized by a triple $(a_m, a_v, a_s)$.

Similarly, each occupation $j$ is characterized by a vector of skill requirements $(b_m, b_v, b_s)$ based on the O*NET descriptors and the Guvenen et al. (2020) procedure above.

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.

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

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 transform this measure such that the support is $\mathbb{R}^+ \cup \{0\}$ by subtracting the minimum of all $\hat{a}_i^j$. In addition, we divide it by the standard deviation of $\hat{a}_i^j$ in order to normalize it.

$$\overline{a}_i^j = \frac{\hat{a}_i^j - \min\{\hat{a}_i^j\}_{i,j}}{\sigma(\hat{a}_i^j)} \tag{2}$$

Each individual $i$ in this economy now is endowed with a vector of occupation-specific abilities $\overline{a}_i = \{\overline{a}_i^j\}_{j \in J}$, where $J$ is the number of different occupations in the economy.

In the non-linear occupational choice environment of our model, both absolute and comparative advantage matter, which is why $\overline{a}_i$ takes into account both aspects.