Labor Problem Set 5

Colin Aitken, Fern Ramoutar, Jordan Rosenthal-Kay, and Yixin Sun

February 17, 2021

Question 1. Define "complexity" as it relates to jobs in the modern labor market (see Caines et al. , 2017). Discuss the evidence on the rise in the payment ("return") to complexity and compare with Deming's (2017) evidence on social skills and to Kosse et al (2020) on the payment to prosociality. Specifically, discuss:

(a) How these skills are measured.

The Caines et al. (2017) paper measures an occupation's complexity in the following way take survey measured descriptions of task components from the O*NET skills and abilities data, calling it X_o, perform one-dimensional principal component analysis on the task content, which means that they seek a vector γ such that it minimizes the residual variance $\|X - \gamma' X \gamma\|$, then define "complexity" C_o as the projected value, $C_o = \gamma' X_o$ for occupation o.

The Deming (2017) paper measures social skills using the "social module" of the O*NET survey, taking the average skill/task "intensity" over four questions.

The Kosse et al. (2020) paper measures prosociality in children with a three-pronged questionaire (with modules on trust, altruism, and "other regarding behavior") as well as "lab" experiments (e.g., incentivized dictator game) to validate these measures.

(b) The sensitivity of the evidence to alternative definitions of these skills.

O*NET-based measures like those in the Caines et al. and Deming studies will be highly sensitive to alternative definitions. There are several nodes in the data production process where the choices the researcher makes can significantly alter that data:

1. In data collection, sensitivities arise in the **choice of survey measures and modules** from the O*NET data base are used in construction. For example, the Caines et al. paper selected specific questions from the skills and abilities modules in O*NET but, for example, omitted responses from the "Interests," "Knowledge," or "Work Context" questionnaires. Because they use PCA to collapse their data, the inclusion of additional questions would fundamentally reshape the variance/covariance structure of their data and the loadings of each component onto the principal vector in the decomposition. Deming's paper uses the social "skills" data but omits content from the "work activities/interacting with others" module.

2. Results may be sensitive to the **scaling** of the intensity measure. E.g., Deming states that he "convert[s] average scores by occupation on O*NET questions to a 0–10 scale that reflects their weighted percentile rank in the 1980 distribution of task inputs," while Caines et al. use PCA's linear projection to covert the likert-scaled data to a numerical value. Because the underlying data does not correspond to any natural measure, any monotone transformation of the data is effectively as good as another, e.g., turning data into z-scores, quantiles, weighted averages, etc, and these transformations will necessarily affect the results.

Kosse et al. (2020) use a tailored questionnaire that is validated by "experiments," which by consequence is less likely to be sensitive to alternative definitions of the data, since the researchers know precisely what it measures and what those measures correspond to. Nonetheless they still use aggregation across questions to produce a single score, for some regressions, which gives room for result sensitivity to aggregation method. This concern is mitigated as they report results for each of their underlying measures which show qualitiatively the same thing.

> (c) How does any of this relate to the discussion of the "hollowing out" of wages?

First, we need to define this term. 'Hollowing out' is when the labor market experiences a reduction in the share of jobs in the middle of an initial wage distribution, and an increase in the share of jobs at the bottom and top of this distribution. Typically, this is attributed to the fact that technology and globalization are making it easier for employers to replace workers who perform routine tasks, and jobs like these are concentrated in the middle of the wage distribution. However, our ability to accurately document such a phenomenon hinges on how we measure skills, since this this will in turn determine what constitutes a routine task, and which tasks are replaceable by technology. As we mentioned above, idiosyncratic choices of surveys and scaling of the intensity measures will lead to potentially inconsistent definitions of how routine or complex certain tasks are. Moreover, middle wage occupations may involve routine or non-routine tasks, and as Autor et al. (2003) argue, technology can replace human labor in routine tasks much better than in non-routine tasks. If the complexity of skills is mis-measured, then we might incorrectly associate observed hollowing out with task or skill-biased technical as opposed to a different demand or supply-side factor.

> (d) How does "complexity" fit in with Mandelbrot's (1962) definition of an occupation?

Mandelbrot (1962) stated that an occupation was a bundle of multiple tasks (???), and that earnings were rewards to multiple tasks which varied across markets. Complexity is relevant to this definition in three ways. First, complexity helps us understand which tasks will be grouped together to form an occupation, and what the nature of the occupation will be. A series of complex tasks form are more likely to form a high-skill occupation, whereas a series of simple tasks are more likely to form a low-skill occupation. However, an occupation may also feature tasks of varying complexity levels, and in that case, the overall mix of complexities will determine the overall skill-level of the job. Second, we can also use complexity in this context to think about the complementarity of cognitive and non-cognitive skills for performing the tasks that make up an occupation. Cognitive skills may be more important in jobs with many complex tasks (e.g. scientists) relative to jobs with fewer complex tasks (e.g. clowns), whereas diligence may be equally important in both occupations. Third, given Madelbrot's definition, the complexity of tasks in an occupation will be a determinant of the overall earnings associated with the occupation in a given market. As a result, Heckman (2019) suggests that people will sort into different occupations given their endowment of skills "in pursuit of their comparative advantage" in performing different tasks. Deming (2017) argues that the returns to social skills has increased while the returns to cognitive skills has decreased over time. However, his estimates may be sensitive to data choices. More consistent with Mandelbrot's definition of occupation as a bundle of tasks is Caines et al. (2017), which argues that there have been high returns to specific bundles of cognitive and non-cognitive skills.

> **Question 2.** Define statistical discrimination. Does the Bayesian theory of optimal decision-making vindicate this practice? Why or why not? (Discuss not only the application of priors in Bayesian decision-making, but also their formation.) Specifically, how could taste-based discrimination affect statistical discrimination?

As discussed in the presentation, statistical discrimination can be framed as a signal extraction problem, where labor market discrimination is the outcome of an employers expectation or variance of productivity, given a person's group-membership. Let us discuss how this plays out in the setup from Coate and Loury (1993).

There are two tasks the firm needs: task 0 which can be performed by any worker, and task 1 which needs a qualified worker. There are two groups, A and B, that workers belong to. The sequence of decisions between workers and the firm is as follows:

1. Nature chooses workers group membership
2. Workers make investment decisions, which qualify them to perform task 1. This decision depends on the cost of the investment and the extent to which this investment increases their chance of being assigned to task 1 (task 1 pays a wage premium).
3. Workers are matched with employers randomly.
4. Workers take a test or interview, which translates to a noisy signal of their qualification that the employer observes
5. The employer has a prior belief, based exclusively on group membership, of the probability the worker is qualified. They update this belief after observing the noisy signal, and assigns the worker to task 1 if this posterior belief is above a certain threshold.

The equilibrium outcome is a pair of beliefs about the qualifications of the two groups. In this setup, this belief is formed both on the basis of how much workers choose to invest, and on the thresholds the employer picks for each group. It is then possible that because employers believe there are less qualified people in group A, then workers in group A will choose to invest less in attaining higher qualifications. In this way, the employees beliefs causes a self fulfilling prophecy, and there is discrimination in the labor market even though the employer is updating their belief rationally. The key here is that is the reason group-membership conveys any information is because the employer expects them to, and thus attributes the resulting disparity in outcomes to disparity in worker ability within a group.

Thinking about this sort of hiring process in a dynamic sense, it is straightforward to think of taste-based discrimination leading to statistical discrimination. In the first period, firms hire not only based on expected productivity of workers, but based on prejudicial beliefs. This causes distortions in workers investment decisions. In the second second period, even if the firms now no longer hold prejudicial beliefs, statistical discrimination can arise out of the firms' prior beliefs about the number of qualified workers in each group, and workers understand that their investment in attaining qualifications will yield a lower return on their wages.

Question 3. Many economists estimate "value-added regressions" based on test scores or discuss achievement gaps in terms of convergence of or divergence of test scores (see, e.g., Derek Neal's handout ("Black-White Inequality") on the class website).

(a) What is the "natural scale" of test scores?

In most empirical work, researchers treat test scores as cardinal scales, when in reality, they should be on an ordinal scale. There is no "natural scale" of test scores, but rather these test scores are subject to arbitrary coding and scale transformations; we cannot say that the distance gap a 90 and a 91 on a test is equal to the gap between 70 and 71. Whether we think an achieve gap has narrowed depends on the scale that we use. Additionally, what we care about is not actually test scores, but rather test scores as a summary of an array of traits that lead to later life outcomes. Thus we must anchor test scores to outcomes such as educational attainment or future earnings in order to truly evaluate the return on investment in education.

> (b) Why is a monotone transformation of a test score not a test score?

Since test scores are arbitrarily scaled and normalized, we can think of them as being on an ordinal scale rather than a cardinal scale. Then, similar to utils, any monotonic transformation of a test score preserves the uniqueness of the scale, but in practice this is a transformation of a variable with an unknown scale. Thus, the danger in transforming scores to mean zero and a variance of one means that intervals between points on this scale have no interpretation. If test-scores are normalized each year, then the value added, or the change in test scores over time, is not invariant to monotonic transformations.

> (c) Discuss evaluation of teachers through use of value-added regressions.
> Test score at grade $g : T(g)$: test score at grade $g-1 : T(g-1)$
> Value added regression for student j with teacher $i : T_j(g) - T_j(g-1) = \alpha + \beta I(i,j) + \varepsilon_j$ where ε_j is an error term and $I(i,j)$ indicates the teacher that student j had in grade g and test scores are taken at the end of the school year with teacher i.

Economic research has shown that teachers matter, but there is also strong evidence that teacher effects fade out quickly [1]. It is easy to see how both these effects could be true when test scores are rescaled. Suppose student learning was the sum of teachers' effects through time. Assume that the variance of teacher quality is constant. Then the variance of student learning will increase over time. However, if test scores are rescaled, then the variance in learning will appear constant over time.

So what happens when we run the regression from this question? The value add of increased teacher quality will appear smaller in later grades, even if its impact on actual learning is quite large. Similarly, the effect of a high quality teacher in first grade will appear to decline quickly over time, even if the teacher had large and persistent effects on their students.

[1] Lang, Kevin. "Measurement Matters: Perspectives on Education Policy from an Economist and School Board Member." Journal of Economic Perspectives, vol. 24, no. 3, 2010, pp. 167–182., doi:10.1257/jep.24.3.167.

As stated above, the score should instead be anchored to something like educational attainment or future earnings. If log wages are not linear functions of test scores, then increases in test scores do not provide full information on gains in socioeconomic outcomes.

Question 4. When are prices (wages) linear in attributes? Identical across firms? Across sectors? Is nonlinearity in prices (wages) evidence of monopoly or monopsony power?

A typical model in which characteristics influence wage setting is one in which jobs j have characteristics z_j and individuals have characteristics x_i. Labor supply L_{ij} comes from indirect utility maximization, taken wages as given,

$$L_{i^*(j),j} = \text{argmax}_{L_{ij}}\, v(w_{ij}L_{ij}, L_{ij}, z_j, x_i, \xi_i) \quad (1)$$

where consumption C has been substituted for the budget constraint $w_{ij}L_{ij}$, where there is a disutility or time cost to the provision of labor and utility in consumption, $v_c > 0$, $v_L < 0$. The term ξ_i is idiosyncratic and potentially unobservable to the firm. Labor demand comes from profit maximization where firms account for characteristics, i.e., firms solve,

$$\max_i \pi_j(w_{ij}L_{ij}, z_j, x_i) \quad (2)$$

First order conditions on these problems define labor supply and demand curves, and equilibrium is defined as a price vector w_{ij} and an assignment of workers to firms i^*, that is consistent with maximization and labor market clearing.

While e.g., the FOC on (2) shows that $d\pi/dz_j$ will be proportional to the shadow price of that characteristic in equilibrium, there is no reason for the equilibrium wage equation to be linear in attributes. Linearity arises from functional form assumptions on utility and profits.

By consequence, nonlinearity of attributes in the wage equation, e.g., $w_{ij} = \alpha x_i + \beta z_j + \gamma x_i z_j$ is not necessarily evidence of market power and can be completely consistent with equilibrium in a competitive environment.

Even if firms do not have preferences over worker type, i.e.,

$$\pi_j(w_{ij}L_{ij}, z_j, x_i) = \pi_j(w_{ij}, z_j)$$

the resulting wage equation may still depend on individual characteristics because equilibrium wages give rise to indifference for the marginal worker across occupations. If workers of type x' have strong distaste for jobs of type z', the wage must offset utility differences, i.e., it must compensate.

Finally, if the profit equation does not depend on workers characteristics, and the utility equation does not depend on job characteristics, even linearity in attributes in the wage equation may then be evidence of market imperfections. In particular, the pricing of worker attributes then reflects monopsonistic price discrimination, where otherwise identical workers who vary only in x_i receive different wages not because of the logic of compensating differentials, but because the firm is able to lower their wages to the point of indifference with their outside option. If women employed in a given occupation, for example, have worse outside options than men – because of differential preferences in commuting or the task content of occupations – but are otherwise identical, then a gender wage gap within that occupation may be evidence of firms' market power.