Employability

Defining Employability Coefficients for Careers and Professional Attributes

Atlas Research Team

2023-06-08

Contents

1 Mathematical Definition of Employability

1.1 Initial Insights

1.1.1 Probability of Finding a Job

Employability, as the word itself indicates, is *the ability to find a job.* This is very obvious, indeed, but it bears mentioning in order to quantify the concept properly. From this tautological notion, then, we could define, in mathematical terms, that employability is not exactly the *ability* to find a job, but rather the *probability* of being employed, given a particular set of professional attributes. That is, if we measure professional attributes on a scale, say, of 0 to 100, calling each attribute a_i^u the i-th attribute of person u, such that $a_i^u \in [0, 100] \; \forall \; i, u$, then it follows that employability is:

$$\Lambda(\mathbf{a_u}) = \Pr(l = 1 \mid \mathbf{a_u}),$$

where $\mathbf{a_u} = (a_1^u, ..., a_m^u)$ is the vector of professional attributes a_i^u that characterizes person u, and $l \in \{0, 1\}$ is a dummy variable denoting whether the individual is employed or not.

Now, suppose every individual employed in an occupation q are characterized by the very same $\mathbf{a_q} = (a_1^q, ..., a_m^q)$ vector of professional attributes. Additionally, suppose we know the $n_q \geq 0$ employment level of this occupation, and the total labor force $\mathrm{N} = \sum_j n_j$. With this, it is then straightforward to estimate the probability $\Lambda(\mathbf{a_q})$ defined above as:

$$\begin{aligned} \Lambda(\mathbf{a_q}, n_q, \mathrm{N}) &= \Pr(l = 1 \mid \mathbf{a_q}) \\ &= \Pr(l = 1 \mid n_q, \mathrm{N}) \\ &= \frac{n_q}{\mathrm{N}} \end{aligned}$$

In other words, the employability of occupation q is the probability of any person in the total work force N being one of the n_q people employed in that occupation. Pretty simple.

However, this formulation fails to be an adequate employability metric for two main reasons. The first is that it yields values too small to be easily interpretable. Since any given occupation is but a really tiny fraction of the total labor market, Λ values calculated this way are proportionally small. The second problem is a more interesting one, and will be treated in the following section.

1.1.2 Adjusting for Professional Similarity

The second problem with this initial formulation of an employability coefficient is that people can find jobs in many different fields, provided they are sufficiently qualified. In other words, if an individual is employed in occupation q and occupation q is very similar to another occupation $\tilde{q}$, this individual could also find employment as a $\tilde{q}$. Thus, their employability coefficient would need to account for the probability of them being employed either as q or $\tilde{q}$ — again, supposing they are sufficiently qualified.

Mathematically, if occupations q and $\tilde{q}$ are said to be similar, then the probability of a person with a professional profile identical with q finding a job either as q or $\tilde{q}$ could be intuitively written as the probability of two mutually exclusive events, adjusted by the probability of two other independent events, namely: the probability of finding a job as a q or a $\tilde{q}$, adjusted by the probability of being *recognized* by the labor market as either one of these occupations to begin with.

The first part of this solution has to do with the amount of job posts available for each occupation, and it was already explained above. The second part has to do with professional similarity, as two occupations are recognized as identical by the labor market to the degree they are similar in the Euclidean sense. Hence, we may define a similarity metric like:

$$s(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) = 1 - \frac{d(\mathbf{a_q}, \mathbf{a_{\tilde{q}}})}{\max d(\mathbf{a_j}, \mathbf{a_k})} = 1 - \frac{\sum_{i=1}^m (a_i^q - a_i^{\tilde{q}})^2}{\max \sum_{i=1}^m (a_i^j - a_i^k)^2},$$

where $d(\mathbf{a_q}, \mathbf{a_{\tilde{q}}})$ is the Euclidean distance between the professional attribute vectors of occupations q and $\tilde{q}$, and $\max d(\mathbf{a_j}, \mathbf{a_k})$, the maximum Euclidean distance between any two occupations j and k. From this, we understand that the overlap between vectors $\mathbf{a_q}$ and $\mathbf{a_{\tilde{q}}}$ is proportional to the interchangeability of occupations q and $\tilde{q}$. And so, as $s(\mathbf{a_j}, \mathbf{a_k}) \in [0, 1] \ \forall \ j, k$, this metric could be understood as a probability of having sufficient professional overlap:

$$\Pr(j \approx k \mid \mathbf{a_j}, \mathbf{a_k}) = \mathrm{I}(\mathbf{a_j}, \mathbf{a_k}) = s(\mathbf{a_j}, \mathbf{a_k})$$

The coefficient above refers to the interchangeability of occupations, and can be defined as the similarity between them, or any increasing function of it. Its interpretation is as follows: if employers are interested in a particular subset of professional attributes, which can be attained by any person, regardless of their career path, then the probability of a person having the desired competencies is the overlap between their own competencies and those of a specific occupation, or professional profile, displaying these competencies to the desired degree; or, alternatively, the probability of two professional profiles being interchangeable is the percentage of overlap between them.

The reasoning behind this apparently hasty conclusion is simple: all the job activities of an occupation require of individuals a particular set of skills, which is quantified by their professional profile. If a person has all these competencies, then the probability of being unable to perform a given task on the job is, *ceteris paribus*, precisely zero. But, conversely, if they lack any of the traits required to perform these tasks, then the probability they are unable to perform them properly increases by the exact measure of their under-qualification. Thus, the interchangeability of professional profiles is proportional to their similarity.

Having understood this, at last we can estimate the probability of finding a job as either q or $\tilde{q}$, adjusted by the similarity between these occupations:

$$\begin{gathered}
\Lambda(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}, n_q, n_{\tilde{q}}, \mathrm{N}) = \Pr(l = 1 \mid \mathbf{a_q}) \\
= \Pr(q \approx q \mid \mathbf{a_q}, \mathbf{a_q}) \Pr(l = 1 \mid n_q, \mathrm{N}) + \Pr(q \approx \tilde{q} \mid \mathbf{a_q}, \mathbf{a_{\tilde{q}}}) \Pr(l = 1 \mid n_{\tilde{q}}, \mathrm{N}) \\
= \mathrm{I}(\mathbf{a_q}, \mathbf{a_q}) \Pr(l = 1 \mid n_q, \mathrm{N}) + \mathrm{I}(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) \Pr(l = 1 \mid n_{\tilde{q}}, \mathrm{N}) \\
= s(\mathbf{a_q}, \mathbf{a_q}) \Pr(l = 1 \mid n_q, \mathrm{N}) + s(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) \Pr(l = 1 \mid n_{\tilde{q}}, \mathrm{N}) \\
= 1 \times \left(\frac{n_q}{\mathrm{N}}\right) + s(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) \times \left(\frac{n_{\tilde{q}}}{\mathrm{N}}\right) \\
= \frac{n_q + s(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) n_{\tilde{q}}}{\mathrm{N}}
\end{gathered}$$

Put simply, to the exact measure that occupations q and $\tilde{q}$ are interchangeable, employment opportunities for either occupation should include the other: as their professional attributes overlap, so too their employment opportunities also overlap. In fact, we could say people with professional profiles $\mathbf{a_q}$ and $\mathbf{a_{\tilde{q}}}$ compete for the same $n_q + n_{\tilde{q}}$ job posts.

In this sense, we notice, for example, that occupations which are completely dissimilar have no employment overlap, as

$$s(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) = 0 \implies \Lambda(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}, n_q, n_{\tilde{q}}, \mathrm{N}) = \frac{n_q + 0 \times n_{\tilde{q}}}{\mathrm{N}} = \frac{n_q}{\mathrm{N}},$$

and, on the other hand, that completely similar occupations have a complete overlap of employment opportunities as well:

$$s(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}) = 1 \implies \Lambda(\mathbf{a_q}, \mathbf{a_{\tilde{q}}}, n_q, n_{\tilde{q}}, \mathrm{N}) = \frac{n_q + 1 \times n_{\tilde{q}}}{\mathrm{N}} = \frac{n_q + n_{\tilde{q}}}{\mathrm{N}}$$

This said, it is important to note that this so-called "overlap" of employment opportunities in no way implies the total work force actually increases beyond N, since N is calculated by summing up the n_j *workers* employed in each occupation, not the total amount of *people* in the work force. That is, if a person chooses to work another job, they are counted as a separate worker in that particular occupation. Thus, N already accounts for people with multiple careers. This is important because, otherwise, we would need to adjust the denominator in this mathematical formulation, and in doing so we could fall into the first problem mentioned above (viz. coefficients of employability being too small).

The employability coefficient, therefore, only indicates which percentage of the N available jobs would be *suitable* to a given person, adjusting for their similarity with each occupation. It does not mean that said person in fact ends up working at the associated jobs, nor that the total amount of available job posts increases simply because workers could be employed elsewhere. Put another way, while employment opportunities do overlap, actual employment does not.

For example, Economists, Accountants, Executives, and Actuaries have a reasonably high professional similarity. And in practice, we do find these occupations competing for the same job posts. Engineers, and other STEM-related careers, likewise also have significant overlap in their employment opportunities. And the pattern repeats itself for every group of similar occupations. However, in any of these cases, no matter how many people are suitable for a given job post, only one of them is actually hired. So, despite the sum of overlapping employment opportunities being indeed much higher than available jobs, when each individual is finally employed, the total work force equals N.

Thus, at least so far, this metric seems intuitive and consistent. Therefore, we can write a provisional generalization for any number of occupations:

$$\begin{aligned}\Lambda(\mathbf{a_u}, \mathbf{A}, \mathbf{n}, \mathrm{N}) &= \Pr(l = 1 \mid \mathbf{a_u}) \\ &= \sum_j \Pr(u \approx j) \Pr(l = 1 \mid n_j, \mathrm{N}) \\ &= \frac{\sum_j \mathrm{I}(\mathbf{a_u}, \mathbf{a_j}) n_j}{\mathrm{N}} \\ &= \frac{\sum_j s(\mathbf{a_u}, \mathbf{a_j}) n_j}{\mathrm{N}},\end{aligned}$$

where $\mathbf{A}$ is the matrix of all $\mathbf{a_j}$ professional profiles, and $\mathbf{n} = (n_1, ..., n_j)$, the vector of job posts for the corresponding occupations, such that $\mathrm{N} = \sum_j n_j$.

1.2 Additional Metrics

Now, even though this adjusted metric is much more adequate than the previous one, it still has some minor flaws. For instance, it isn't reasonable to assume linear behavior in the professional interchangeability coefficient. For assuming such a relationship would imply that occupations with small similarity to other occupations, say in the 0 to 30% range, could be seen by the labor market as "still somewhat similar", and thus compete for a fraction of available job posts equal to their similarity coefficients; while, in practice, job applicants with such small similarities would be immediately dismissed by HR recruiters.

Thus, a scaling factor should be applied to the interchangeability metric, in order to reduce the probability of being recognized as "similar" to a given occupation, specially in the smaller ranges of professional similarity. Again, as $s(\mathbf{a_j}, \mathbf{a_k}) \in [0, 1]\ \forall\ j, k$, a simple way of implementing this adjustment would be:

$$\Pr(j \approx k \mid \mathbf{a_j}, \mathbf{a_k}) = \mathrm{I}(\mathbf{a_j}, \mathbf{a_k}) = s(\mathbf{a_j}, \mathbf{a_k})^{\alpha},\ \ \alpha > 1$$

as $x^{\alpha} \leq x\ \forall\ \alpha > 1,\ x \in [0, 1]$. However, this metric is also inadequate, as it can be "overly severe" in the higher ranges of similarity. A linear interpolation then is applied so as to smooth the transition between the adjusted equation above and the linear one already described:

$$\Pr(j \approx k \mid \mathbf{a_j}, \mathbf{a_k}) = \mathrm{I}(\mathbf{a_j}, \mathbf{a_k}) = s(\mathbf{a_j}, \mathbf{a_k}) s(\mathbf{a_j}, \mathbf{a_k}) + [1 - s(\mathbf{a_j}, \mathbf{a_k})] s(\mathbf{a_j}, \mathbf{a_k})^{\alpha},\ \ \alpha > 1$$

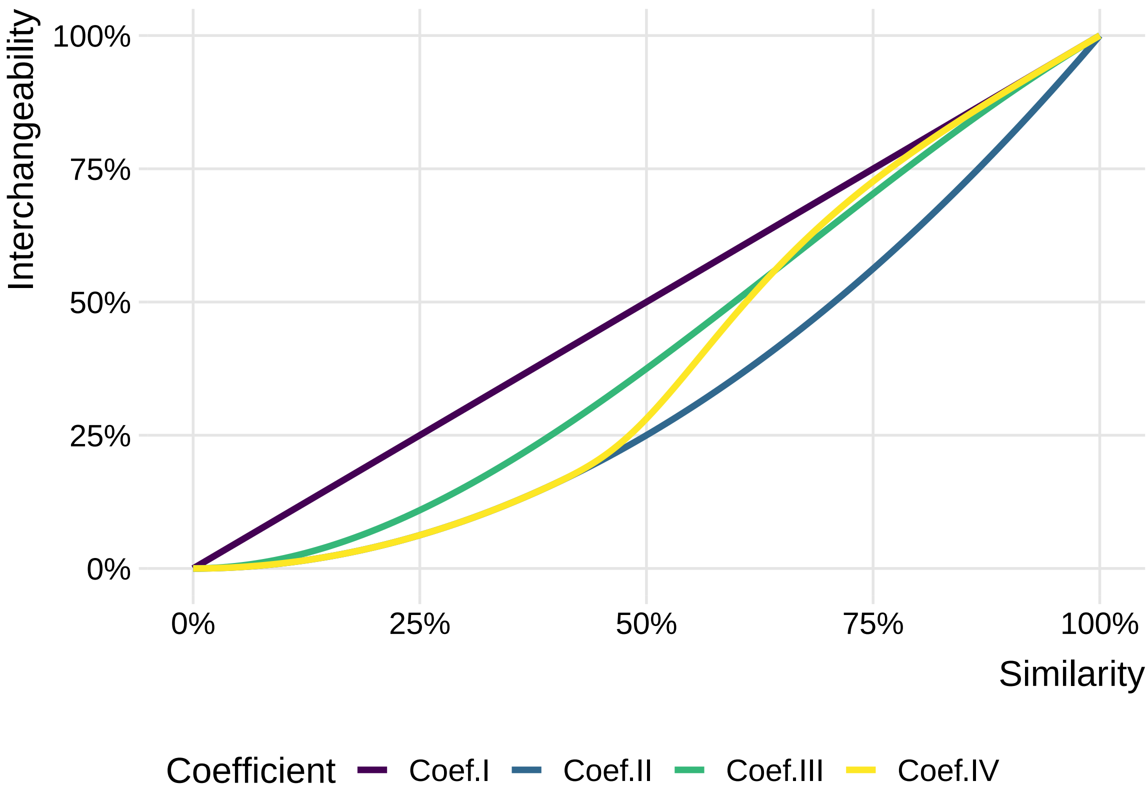
And an additional adjustment could provide an even more realistic behavior to the interchangeability function:

$$\Pr(j \approx k \mid \mathbf{a_j}, \mathbf{a_k}) = \mathrm{I}(\mathbf{a_j}, \mathbf{a_k}) = s(\mathbf{a_j}, \mathbf{a_k}) s(\mathbf{a_j}, \mathbf{a_k}) + [1 - s(\mathbf{a_j}, \mathbf{a_k})] s(\mathbf{a_j}, \mathbf{a_k})^{\left[\frac{1}{s(\mathbf{a_j}, \mathbf{a_k})}\right]^{\left[\frac{1}{s(\mathbf{a_j}, \mathbf{a_k})}\right]}}$$

Here, the “severity” of the adjustment is entirely proportional to the similarity between occupations, and does not depend on arbitrarily set parameters. In either case, the more similarity approaches 1, the more interchangeability becomes linear.

Finally, these interchangeability metrics are illustrated in the diagram below:

Figure 1: Interchangeability Coefficients I–IV



From this figure, we immediately conclude that coefficients of interchangeability III and IV exhibit a very natural behavior, allowing for some overlap between occupations, without being “too severe” or “too lenient” in their compensation. Therefore, these mathematical formulations appear to adequately express what we refereed to in the beginning of section 1.2. as “sufficient qualification”. However, it is evident that coefficient IV is indeed the most realistic of them all.

2 Estimating Employability Coefficients

We now estimate the employability metrics developed above in three instances: for occupations; for professional attributes; and for any professional profile. In order to estimate the employability of occupations we need only apply the $\Lambda(\mathbf{a_j}, \mathbf{A}, \mathbf{n}, \mathrm{N})$ function for every $\mathbf{a_j}$ vector of already registered career competencies, utilizing one of the four $\mathrm{I}(\mathbf{a_j}, \mathbf{a_k})$ interchangeability coefficients illustraded in Figure 1. The estimation of employability coefficients for individual professional attributes involves carrying out a linear regression of some sort (either OLS or NNLS). And, provided this model is reliable, we can proceed to calculate employability coefficients for any arbitrary $\mathbf{a_u}$ vector of professional attributes (i.e. a professional profile) by multiplying the model’s parameters with individuals’s attribute scores.

2.1 For Occupations

dsds

2.2 For Professional Attributes

dsds

2.3 For Any Professional Profile

dsds