The Career Atlas: Mathematical Notation

Cao Bittencourt

[a] *B. Sc. in Economics from EPGE (FGV), RJ, Brazil.*
[b] *Statistician at Atlas Career Guide Inc., FL, USA.*

Abstract

This is a brief document to define statistical methods for data-driven career choice and development. It deals with topics such as: career matching (i.e. vocational choice); estimation of competence, or overall skill level; estimation of skill set generality; versatility; skill set profitability; employability; labor market competitiveness; labor market taxonomy; optimal human resources acquistion and allocation; and so on and so forth. Each concept shall be explained at length in separate articles.

Keywords: Career choice; Career development; Matching algorithms; Competence; Similarity.

Contents

1. Basic Definitions

1.1. Skill Sets

The i-th professional attribute, or competency, of a person k is defined as:

$$a_i^k \in [0, 100], \tag{1}$$

where the interval $[0, 100]$ determines the bounds for every competency.[1]

The skill set, or career profile, of a person k is the vector of their m attributes:

$$\boldsymbol{a_k} = (a_1^k, \ldots, a_m^k). \tag{2}$$

The skill set matrix, or career profile matrix, is the collection of all n skill sets in the economy:

$$\mathbf{A} = \begin{bmatrix} a_1^1 & \ldots & a_m^1 \\ \vdots & \ddots & \vdots \\ a_1^n & \ldots & a_m^n \end{bmatrix}. \tag{3}$$

1.2. Skill Set Normalization

Normalization by the scale bounds is defined by the tilde operator:

$$\tilde{a}_i^k = \frac{a_i^k - 0}{100 - 0} = \frac{a_i^k}{100} \in [0, 1]; \tag{4}$$

$$\tilde{\boldsymbol{a}}_{\boldsymbol{k}} = (\tilde{a}_1^k, \ldots, \tilde{a}_m^k); \tag{5}$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{a}_1^1 & \ldots & \tilde{a}_m^1 \\ \vdots & \ddots & \vdots \\ \tilde{a}_1^n & \ldots & \tilde{a}_m^n \end{bmatrix}. \tag{6}$$

Normalization by a skill set's highest attribute is defined by the hat operator:

$$\hat{a}_i^k = \frac{a_i^k}{\max a_j^k} \in [0, 1]; \tag{7}$$

$$\hat{\boldsymbol{a}}_{\boldsymbol{k}} = (\hat{a}_1^k, \ldots, \hat{a}_m^k); \tag{8}$$

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{a}_1^1 & \ldots & \hat{a}_m^1 \\ \vdots & \ddots & \vdots \\ \hat{a}_1^n & \ldots & \hat{a}_m^n \end{bmatrix}. \tag{9}$$

[1] More generally, these could be defined as a_{lb} (the lower bound) and a_{ub} (the upper bound). Here, the interval $[0, 100]$ is used because of its ease of interpretation.

1.3. Preferences

1. individual attribute preference υ_i^k
2. attribute preference vector $\boldsymbol{\upsilon_k}$
3. attribute preference matrix $\boldsymbol{\Upsilon}$
4. most attribute operations apply to preferences
5. preference-adjusted skill sets

2. Basic Skill Set Models

2.1. Skill Set Generality

The generality of a skill set is the mean of its maxima-normalized attributes:

$$\gamma_k = \left(\frac{1}{m}\right) \sum_{i=1}^{m} \hat{a}_i^k \in [0,1]. \tag{10}$$

People with high γ_k scores are called *generalists*. Conversely, those with low γ_k scores are called *specialists*. Career profiles that are neither broad nor specialized are said to be *balanced*.

The generality vector of all n skill sets in the economy is:

$$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n). \tag{11}$$

2.2. Attribute Equivalence

The attribute equivalence of a particular attribute in a skill set measures the importance of that attribute relative to the skill set's highest attribute. It is calculated using skill set generality as both a midpoint and scaling parameter in the following linear-logistic classification function:

$$\mathrm{aeq}(\hat{a}_i^k, \gamma_k) = \hat{a}_i^k \left[1 + \gamma_k(1 - \hat{a}_i^k) \exp\left(\frac{\hat{a}_i^k - \gamma_k}{\gamma_k - 1}\right)\right]^{-\frac{\gamma_k}{\hat{a}_i^k}} \in [0,1]. \tag{12}$$

For a short-hand notation, the attribute equivalence can be denoted by the umlaut operator:

$$\ddot{a}_i^k = \mathrm{aeq}(\hat{a}_i^k, \gamma_k). \tag{13}$$

This is not to be confused with the Netwonian dot notation for partial derivatives, which we do not employ, instead preferring the more explicit $\frac{\partial x}{\partial y}$ derivative operator of Leibniz.

At any rate, attributes with high levels of $\ddot{a}_i^k$ are said to be equivalent to the skill set's most importante attribute. These are a career profile's *core* competencies. The remaining competencies are classified as either *important*, *auxiliary*, *minor*, or *unimportant*.

The attribute equivalence vector of a skill set is given by the collection of their m umlauted attributes:

$$\ddot{\boldsymbol{a}}_{\boldsymbol{k}} = (\ddot{a}_1^k, ..., \ddot{a}_m^k). \tag{14}$$

Finally, the attribute equivalence matrix is the collection of all attribute equivalence vectors in the economy:

$$\ddot{\mathbf{A}} = \begin{bmatrix} \ddot{a}_1^1 & \dots & \ddot{a}_m^1 \\ \vdots & \ddots & \vdots \\ \ddot{a}_1^n & \dots & \ddot{a}_m^n \end{bmatrix}. \tag{15}$$

2.3. Skill Set Competence

The overall competence of a skill set is the mean of its scale-normalized attributes, weighted by each attribute's importance (i.e. its attribute equivalence):

$$c_k = \frac{\sum_{i=1}^m \ddot{a}_i^k \tilde{a}_i^k}{\sum_{i=1}^m \ddot{a}_i^k} \in [0, 1]. \tag{16}$$

Career profiles with high c_k are said to be competent. However, this adjective can be seen as offensive to some people; and, most importantly, it could also be misleading, because competence is often defined relative to the specific requirements of a particular job. Therefore, we opt for the more generic competence classification of *high level*, *mid level*, and *low level*, which is somewhat less ambiguous.

The competence vector of all n skill sets in the economy is:

$$\boldsymbol{c} = (c_1, \dots, c_n). \tag{17}$$

3. Comparative Models

3.1. Pairwise Comparative Models

3.1.1. Similarity

The most basic comparative model is that of Euclidean matching with linear weights:

$$s_{k,q} = s(\boldsymbol{a_k}, \boldsymbol{a_q}) = 1 - \tilde{d}(\boldsymbol{a_k}, \boldsymbol{a_q}) \in [0, 1], \tag{18}$$

where

$$\tilde{d}_{k,q} = \tilde{d}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \sqrt{\frac{\sum_{i=1}^m a_i^q (a_i^k - a_i^q)^2}{\sum_{i=1}^m a_i^q \max(100 - a_i^q, a_i^q)^2}} \in [0, 1]. \tag{19}$$

In this model, we compare a skill set $\boldsymbol{a_k}$ to a skill set $\boldsymbol{a_q}$ by calculating the weighted Euclidean distance from $\boldsymbol{a_k}$ to $\boldsymbol{a_q}$ normalized by the maximum theoretical distance to $\boldsymbol{a_q}$.

Other weighting systems can be employed in this type of matching model. We could, for instance, substitute the linear weights with either quadratic weights,

$$a_i^{q2} \in [0,1], \tag{20}$$

or speciality-root weights,

$$a_i^{q\,\frac{1}{1-\gamma_k}} \in [0,1]. \tag{21}$$

But the best and most interpretable results are obtained using attribute equivalence as the weighting function:

$$\tilde{d}_{k,q} = \tilde{d}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q (a_i^k - a_i^q)^2}{\sum_{i=1}^{m} \ddot{a}_i^q \max(100 - a_i^q, a_i^q)^2}} \in [0,1]. \tag{22}$$

We could also employ other matching methods instead of the “baseline” weighted Euclidean model. [detail each method later]:

1. logit regression matching
2. probit regression matching
3. bvls regression matching
4. tobit regression matching
5. pearson correlation matching
6. kendal nonparametric correlation matching
7. spearman nonparametric correlation matching

At last, similarity and normalized distance metrics determine the respective vectors and matrices, as follows:

$$\boldsymbol{s_k} = (s_{k,1}, \ldots, s_{k,n}); \tag{23}$$

$$\boldsymbol{\tilde{d}_k} = (\tilde{d}_{k,1}, \ldots, \tilde{d}_{k,n}); \tag{24}$$

$$\mathbf{S} = \begin{bmatrix} s_{1,1} & \ldots & s_{n,1} \\ \vdots & \ddots & \vdots \\ s_{1,n} & \ldots & s_{n,n} \end{bmatrix} = \begin{bmatrix} 1 & \ldots & s_{k,1} & \ldots & s_{n,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{1,k} & \ldots & 1 & \ldots & s_{n,k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{1,n} & \ldots & s_{k,n} & \ldots & 1 \end{bmatrix}; \tag{25}$$

$$\mathbf{D} = \begin{bmatrix} \tilde{d}_{1,1} & \ldots & \tilde{d}_{n,1} \\ \vdots & \ddots & \vdots \\ \tilde{d}_{1,n} & \ldots & \tilde{d}_{n,n} \end{bmatrix} = \begin{bmatrix} 0 & \ldots & \tilde{d}_{k,1} & \ldots & \tilde{d}_{n,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{d}_{1,k} & \ldots & 0 & \ldots & \tilde{d}_{n,k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{d}_{1,n} & \ldots & \tilde{d}_{k,n} & \ldots & 0 \end{bmatrix}. \tag{26}$$

3.1.2. Qualification

A closely related concept to matching is the qualification comparative model. In this family of functions, however, Euclidean matching is mandatory, as other matching methods do not make sense for this specific type of calculation. The reason for this is at that, here, we are not particularly interested in matching (i.e. a typical classification problem), but rather in the actual distances between comparison skill sets.

To define these models, we first have to define the gap function, which measures only positive competency gaps:

$$\delta_{k,q}^{i} = \delta(a_i^k, a_i^q) = \max(a_i^k - a_i^q, 0) \in [0, 100]. \tag{27}$$

Having defined the gap function, we can write the underqualification model:

$$\tilde{\delta}_{k,q}^{<} = \mathrm{uqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^q, a_i^k)^2}{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^q, 0)^2}} = \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^q, a_i^k)^2}{\sum_{i=1}^{m} \ddot{a}_i^q {a_i^q}^2}}. \tag{28}$$

And, analogously, the overqualification model is given by:

$$\tilde{\delta}_{k,q}^{\geq} = \mathrm{oqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^k, a_i^q)^2}{\sum_{i=1}^{m} \ddot{a}_i^q \delta(100, a_i^q)^2}} = \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^k, a_i^q)^2}{\sum_{i=1}^{m} \ddot{a}_i^q (100 - a_i^q)^2}}. \tag{29}$$

The final set of "sufficient qualification" is, evidently, the complement of the underqualification model:

$$s_{k,q}^{\geq} = \mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = 1 - \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^q, a_i^k)^2}{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^q, 0)^2}} = 1 - \mathrm{uqa}(\boldsymbol{a_k}, \boldsymbol{a_q}). \tag{30}$$

As with the similarity and normalized distance statistics described above, these three qualification models are bounded to the $[0, 1]$ interval. Likewise, they also determine qualification vectors:

$$\boldsymbol{\tilde{\delta}_k^{<}} = (\tilde{\delta}_{k,1}^{<}, \ldots, \tilde{\delta}_{k,n}^{<}); \tag{31}$$

$$\boldsymbol{\tilde{\delta}_k^{\geq}} = (\tilde{\delta}_{k,1}^{\geq}, \ldots, \tilde{\delta}_{k,n}^{\geq}); \tag{32}$$

$$\boldsymbol{s_k^{\geq}} = (s_{k,1}^{\geq}, \ldots, s_{k,n}^{\geq}); \tag{33}$$

and matrices

$$\tilde{\Delta}_{<} = \begin{bmatrix} \tilde{\delta}_{1,1}^{<} & \ldots & \tilde{\delta}_{n,1}^{<} \\ \vdots & \ddots & \vdots \\ \tilde{\delta}_{1,n}^{<} & \ldots & \tilde{\delta}_{n,n}^{<} \end{bmatrix} = \begin{bmatrix} 0 & \ldots & \tilde{\delta}_{k,1}^{<} & \ldots & \tilde{\delta}_{n,1}^{<} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\delta}_{1,k}^{<} & \ldots & 0 & \ldots & \tilde{\delta}_{n,k}^{<} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\delta}_{1,n}^{<} & \ldots & \tilde{\delta}_{k,n}^{<} & \ldots & 0 \end{bmatrix}; \tag{34}$$

$$\tilde{\Delta}_{\geq} = \begin{bmatrix} \tilde{\delta}^{\geq}_{1,1} & \dots & \tilde{\delta}^{\geq}_{n,1} \\ \vdots & \ddots & \vdots \\ \tilde{\delta}^{\geq}_{1,n} & \dots & \tilde{\delta}^{\geq}_{n,n} \end{bmatrix} = \begin{bmatrix} 0 & \dots & \tilde{\delta}^{\geq}_{k,1} & \dots & \tilde{\delta}^{\geq}_{n,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\delta}^{\geq}_{1,k} & \dots & 0 & \dots & \tilde{\delta}^{\geq}_{n,k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\delta}^{\geq}_{1,n} & \dots & \tilde{\delta}^{\geq}_{k,n} & \dots & 0 \end{bmatrix}; \tag{35}$$

$$\mathbf{S}_{\geq} = \begin{bmatrix} s^{\geq}_{1,1} & \dots & s^{\geq}_{n,1} \\ \vdots & \ddots & \vdots \\ s^{\geq}_{1,n} & \dots & s^{\geq}_{n,n} \end{bmatrix} = \begin{bmatrix} 1 & \dots & s^{\geq}_{k,1} & \dots & s^{\geq}_{n,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s^{\geq}_{1,k} & \dots & 1 & \dots & s^{\geq}_{n,k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s^{\geq}_{1,n} & \dots & s^{\geq}_{k,n} & \dots & 1 \end{bmatrix}. \tag{36}$$

P.S.: should qualification be unweighted?

$$\tilde{\delta}^{<}_{k,q} = \mathrm{uqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \sqrt{\frac{\sum_{i=1}^{m} \delta(a^q_i, a^k_i)^2}{\sum_{i=1}^{m} \delta(a^q_i, 0)^2}} = \sqrt{\frac{\sum_{i=1}^{m} \delta(a^q_i, a^k_i)^2}{\sum_{i=1}^{m} {a^q_i}^2}}; \tag{37}$$

$$\tilde{\delta}^{\geq}_{k,q} = \mathrm{oqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \sqrt{\frac{\sum_{i=1}^{m} \delta(a^k_i, a^q_i)^2}{\sum_{i=1}^{m} \delta(100, a^q_i)^2}} = \sqrt{\frac{\sum_{i=1}^{m} \delta(a^k_i, a^q_i)^2}{\sum_{i=1}^{m} (100 - a^q_i)^2}}; \tag{38}$$

$$s^{\geq}_{k,q} = \mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = 1 - \sqrt{\frac{\sum_{i=1}^{m} \delta(a^q_i, a^k_i)^2}{\sum_{i=1}^{m} \delta(a^q_i, 0)^2}} = 1 - \mathrm{uqa}(\boldsymbol{a_k}, \boldsymbol{a_q}). \tag{39}$$

Or even

$$\tilde{\delta}^{<}_{k,q} = \mathrm{uqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \frac{\sum_{i=1}^{m} \delta(a^q_i, a^k_i)}{\sum_{i=1}^{m} \delta(a^q_i, 0)} = \frac{\sum_{i=1}^{m} \delta(a^q_i, a^k_i)}{\sum_{i=1}^{m} a^q_i}; \tag{40}$$

$$\tilde{\delta}^{\geq}_{k,q} = \mathrm{oqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = \frac{\sum_{i=1}^{m} \delta(a^k_i, a^q_i)}{\sum_{i=1}^{m} \delta(100, a^q_i)} = \frac{\sum_{i=1}^{m} \delta(a^k_i, a^q_i)}{\sum_{i=1}^{m} (100 - a^q_i)}; \tag{41}$$

$$s^{\geq}_{k,q} = \mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = 1 - \frac{\sum_{i=1}^{m} \delta(a^q_i, a^k_i)}{\sum_{i=1}^{m} \delta(a^q_i, 0)} = 1 - \mathrm{uqa}(\boldsymbol{a_k}, \boldsymbol{a_q})? \tag{42}$$

P.S.: unweighted qualification metrics can be overly strict. probably use umlauted weights when calculating qualification for the hireability coefficient. however, absolute qualification metrics (i.e. unweighted) can be useful, specially in career development. the strictness of unweighted qualification only becomes a probably in assessing hireability. therefore, it seems appropriate to make a weighted / unweighted qualification metric distinction. for example:

$$\ddot{s}^{\geq}_{k,q} = 1 - \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}^q_i \delta(a^q_i, a^k_i)^2}{\sum_{i=1}^{m} \ddot{a}^q_i {a^q_i}^2}} = 1 - \ddot{\delta}^{<}_{k,q} \tag{43}$$

$$s_{k,q}^{\geq} = \mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = 1 - \frac{\sum_{i=1}^{m} \delta(a_i^q, a_i^k)}{\sum_{i=1}^{m} a_i^q} = 1 - \tilde{\delta}_{k,q}^{<} \tag{44}$$

or, instead, define the weighted model as the default, and the unweighted as the "scale-normalized" model (as it is normalized by the actual theoretical limit, not the weighted theoretical limit):

$$s_{k,q}^{\geq} = \mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}, \boldsymbol{\ddot{a}_k}, \boldsymbol{\ddot{a}_q}) = 1 - \sqrt{\frac{\sum_{i=1}^{m} \ddot{a}_i^q \delta(a_i^q, a_i^k)^2}{\sum_{i=1}^{m} \ddot{a}_i^q {a_i^q}^2}} = 1 - \delta_{k,q}^{<} \tag{45}$$

$$\tilde{s}_{k,q}^{\geq} = \mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) = 1 - \frac{\sum_{i=1}^{m} \delta(a_i^q, a_i^k)}{\sum_{i=1}^{m} a_i^q} = 1 - \tilde{\delta}_{k,q}^{<} \tag{46}$$

3.1.3. Interchangeability

1. sufficient similarity
2. scaled similarity
3. similarity in the strong sense
4. s.s. $\rightarrow ß_{k,q}$
5. alternatively, $ß_{k,q} = \mathrm{seq}(...) = \ddot{s}_{k,q}$
6. interchangeability vector $\boldsymbol{ß_k}$
7. interchangeability matrix $\mathbf{ß}$

alternatively,

$$\begin{aligned}
ß_{k,q} = \mathrm{seq}(\boldsymbol{a_k}, \boldsymbol{a_q}) &= s_{k,q} \left\{ 1 + c_q \tilde{d}_{k,q} \exp\left[-\frac{d_{\max}^q (\tilde{d}_{k,q} - c_q)}{1 - c_q} \right] \right\}^{-\frac{c_q}{\tilde{d}_{k,q}}} \\
&= s_{k,q} \left\{ 1 + c_q (1 - s_{k,q}) \exp\left[\frac{c_q d_{\max}^q - d_{k,q}}{1 - c_q} \right] \right\}^{-c_q/\tilde{d}_{k,q}} \\
&= (1 - \tilde{d}_{k,q}) \times \left[1 + c_q \tilde{d}_{k,q} \exp\left(\frac{c_q \bar{d}_q - d_{k,q}}{1 - c_q} \right) \right]^{-c_q/\tilde{d}_{k,q}} \\
&= (1 - \tilde{d}_{k,q}) \left[1 + c_q \tilde{d}_{k,q} \exp\left(\frac{c_q \bar{d}_q - d_{k,q}}{1 - c_q} \right) \right]^{-c_q/\tilde{d}_{k,q}} \\
&= s_{k,q} \left[1 + c_q \tilde{d}_{k,q} \exp\left(\frac{c_q \bar{d}_q - d_{k,q}}{1 - c_q} \right) \right]^{-c_q/\tilde{d}_{k,q}} \\
&= s_{k,q} \left[1 + c_q \tilde{d}_{k,q} \exp\left(\frac{c_q d_{\mathrm{ub}}^q - d_{k,q}}{1 - c_q} \right) \right]^{-c_q/\tilde{d}_{k,q}} \\
&= s_{k,q} \left\{ 1 + c_q \tilde{d}_{k,q} \exp\left[-\frac{\bar{d}_q (\tilde{d}_{k,q} - c_q)}{1 - c_q} \right] \right\}^{-\frac{c_q}{\tilde{d}_{k,q}}}
\end{aligned}$$

$$= s_{k,q}\left\{1 + c_q\tilde{d}_{k,q}\exp\left[\frac{\bar{d}_q(c_q - \tilde{d}_{k,q})}{1 - c_q}\right]\right\}^{-\frac{c_q}{\tilde{d}_{k,q}}}$$
$$= s_{k,q}\left\{1 + c_q\tilde{d}_{k,q}\exp\left[\frac{\bar{d}_q(c_q - \tilde{d}_{k,q})}{1 - c_q}\right]\right\}^{-c_q/\tilde{d}_{k,q}}$$
$$= s_{k,q}\left\{1 + c_q(1 - s_{k,q})\exp\left[(c_q\bar{d}_q - d_{k,q})/(1 - c_q)\right]\right\}^{-c_q/\tilde{d}_{k,q}}$$

$$ß_{k,q} = \mathrm{seq}(s_{k,q}, c_q) = s_{k,q}^{\frac{1}{1-c_q}} \tag{47}$$

3.1.4. Education and Experience Equivalence

1. years of education τ_k, τ_q
2. years of education vector $\boldsymbol{\tau}$
3. $\ddot{\tau}_{k,q} = \mathrm{eeq}(...) = \mathrm{teq}(\tau_k, \tau_q, ...)$
4. $\ddot{\tau}^{\theta}_{k,q} = \mathrm{eeq}(...) = \mathrm{teq}(\tau_k, \tau_q, ...) \times s^{\theta}_{k,q}$
5. field similarity (direction): $s^{\theta}_{k,q} = \cos\theta_{k,q} = \frac{\boldsymbol{\ddot{a}_k} \cdot \boldsymbol{\ddot{a}_q}}{\|\boldsymbol{\ddot{a}_k}\|\|\boldsymbol{\ddot{a}_q}\|}$
6. $s^{\theta}_{k,q} \in [0, 1]$, because cosine cannot be negative, as attributes are truncated
7. interpretation: equivalent education and experience in the same field
8. vector: $\boldsymbol{s^{\theta}_k}$
9. matrix: $\mathbf{S_{\Theta}}$
10. eeq vector: $\boldsymbol{\ddot{\tau}^{\theta}_k}$
11. eeq matrix: $\mathbf{\ddot{T}_{\Theta}}$

$$\ddot{\tau}_{k,q} = \mathrm{teq}(\tau_k, \tau_q, c_q) = \left\{1 + c_q\tau_q\exp\left[-\tau_q(\tau_k - \tau_q + 1 - c_q)\right]\right\}^{-c_q} \tag{48}$$

$$s^{\theta}_{k,q} = \cos\theta_{k,q} = \frac{\boldsymbol{\ddot{a}_k} \cdot \boldsymbol{\ddot{a}_q}}{\|\boldsymbol{\ddot{a}_k}\|\|\boldsymbol{\ddot{a}_k}\|} = \frac{\sum_{i=1}^{m}\ddot{a}_i^k\ddot{a}_i^q}{\sqrt{\sum_{i=1}^{m}\ddot{a}_i^{k^2}}\sqrt{\sum_{i=1}^{m}\ddot{a}_i^{q^2}}} \tag{49}$$

$$\ddot{\tau}^{\theta}_{k,q} = \ddot{\tau}_{k,q} \times s^{\theta}_{k,q} \tag{50}$$

3.1.5. Hireability

1. to be hireable is to be:
2. sq: sufficiently qualified (basic skill level)
3. se: sufficiently educated / experienced (in the field)
4. ss: sufficiently similar (good fit, same shape)
5. i.e. hireability pressuposes all the previous metrics
6. if *ß* is similarity in the strong sense, then hireability is similarity in the "strongest sense", as it even pressuposes interchangeability (*ß*)
7. $s^{h}_{k,q} = \mathrm{sq}(\boldsymbol{a_k}, \boldsymbol{a_q}) \times \mathrm{se}(\boldsymbol{a_k}, \boldsymbol{a_q}) \times \mathrm{ss}(\boldsymbol{a_k}, \boldsymbol{a_q}) =$
   $\mathrm{sqa}(\boldsymbol{a_k}, \boldsymbol{a_q}) \times \mathrm{eeq}(\boldsymbol{a_k}, \boldsymbol{a_q}) \times \mathrm{seq}(\boldsymbol{a_k}, \boldsymbol{a_q}) =$
   $s^{\geq}_{k,q} \times \ddot{\tau}^{\theta}_{k,q} \times \text{\textit{ß}}_{k,q} =$
   $s^{\geq}_{k,q} \times \ddot{\tau}_{k,q} \times s^{\theta}_{k,q} \times \ddot{s}_{k,q}$
8. p.s.: this metric is too strict. suppose a candidate scored a 0.9 on every metric: $0.9^4 = 0.6561$
9. candidate is hireable if ${}^{1}_{0}[s^{h}_{k,q}]$
10. candidate is hireable for ${}^{1}_{0}[s^{h}_{k,q}] \times s^{h}_{k,q}$ percent of jobs (requires an axiom)
11. hireability vector: $\boldsymbol{s_k^h}$
12. hireability matrix: $\mathbf{S}_{\boldsymbol{h}}$
13. evaluated hireability vector: ${}^{1}_{0}[\boldsymbol{s_k^h}]$
14. evaluated hireability matrix: ${}^{1}_{0}[\mathbf{S}_{\boldsymbol{h}}]$
15. ?percent of hireable jobs vector: ${}^{1}_{0}[\boldsymbol{s_k^h}]\boldsymbol{s_k^h}$
16. ?percent of hireable jobs matrix: ${}^{1}_{0}[\mathbf{S}_{\boldsymbol{h}}]^{\top}\mathbf{S}_{\boldsymbol{h}}$

3.1.6. Utility Equivalence

1. normalized wages $\hat{y}_k = y_k / \max y_q$
2. normalized wages vector $\hat{\boldsymbol{y}}$
3. wage utility function $u(\hat{y}_k)$
4. utility equivalence function $\ddot{u}_{k,q} = \mathrm{ueq}(u(\hat{y}_k), u(\hat{y}_q), ...)$
5. static utility u_k vs comparative utility $u_{k,q}$
6. incorporate preferences in the utility function like in the eeq function? (def = 1 for O*NET career profiles)
7. e.g. $\ddot{u}_{k,q} = \mathrm{ueq}(u(\hat{y}_k), u(\hat{y}_q), ...) \times s^{\upsilon}_{k,q}$

3.2. Axioms

3.3. Fundamental Axioms

In order for derive quantitative career choice and career development methods, we need some basic assumptions.

3.3.1. Skill Set Sufficiency Axiom

1. SSSA (weak): skill $\implies$ activities

2. SSSA (strong): skill $\iff$ activities

The Skill Set Sufficiency Axiom (SSSA) defines that the activities a person is capable of performing follow directly from their skill set. That is to say, activities are essentially the application of an underlying capacity, so that having a certain skill level in a competency necessarily translates into being able to perform every task associated with that competency at that level (and, of course, at the previous levels).

The most intuitive way to understand and mathematically guarantee the Skill Set Sufficiency Axiom is by conceptualizing human capital as a sum of successes on binary outcome variables representing tasks of progressive difficulty:

$$a_i^k = \sum_{\ell=0}^{\ell_i} T_\ell^k, \tag{51}$$

$$T_\ell^k = \begin{cases} 1, \text{if } k \text{ succedes at difficulty level } \ell; \\ 0, \text{otherwise.} \end{cases} \tag{52}$$

This, in turn, pressuposes the following:

1. task difficulty is objectively measureable (see Fundamental Axioms)

2. the skill scale is uniform (see Fundamental Axioms)

   2.1. each competency has both easy and difficult levels

   2.2. the difficulty of each scale is the same as the difficulty of the other scales

   2.3. i.e. performing the tasks associated with attribute a_i at the level $a_i = \bar{a}_\ell$ is, objectively, just as difficult as performing the tasks of attribute a_j at the same $a_i = a_j = \bar{a}_\ell$ level

   2.4. this has nothing to do with the distribution of task difficulty for each attribute: it only means that tasks at the same difficulty level are just as hard; it does not mean that some attributes are not more *frequently* harder than other attributes

3. the skill scale is truncated (see Fundamental Axioms)

4. anyone who can perform a task of difficulty $\bar{a}_{\ell_1}$ can also perform a task of difficulty $\bar{a}_{\ell_2} \leq \bar{a}_{\ell_1}$ (see Fundamental Axioms)

5. a person k can only perform a task of difficulty $\bar{a}_\ell$ if they have a skill level of $a_i^k \geq \bar{a}_\ell$

Thus, we can define a person's skill level as the sum of their successful trials on tasks of increasing difficulty. And assuming the scales are truncated, we can also interpret these metrics as the portion of tasks one is able to perform out of all difficulty levels for that skill.

1. e.g. skill level = 0: cannot perform even the most basic of tasks associated with that skill

2. e.g. skill level = 1: can perform only the bottom 1% of tasks associated with that skill (in order of difficulty)

3. e.g. skill level = 10: can perform the bottom 10% of tasks associated with that skill (in order of difficulty)

4. e.g. skill level = 50: can perform the (easiest) half of the tasks associated with that skill

5. e.g. skill level = 100: can perform all the tasks associated with that skill

Additionally, we can interpret skill levels as the normalization of task difficulty by the most objectively difficult task for each skill.

1. e.g. skill level = 0: cannot perform even the most basic of tasks associated with that skill

2. e.g. skill level = 1: can only perform tasks of up to 1% the difficulty of that skill's most difficult task

3. e.g. skill level = 10: can only perform tasks of up to 10% the difficulty of that skill's most difficult task

4. e.g. skill level = 50: can perform tasks of up to half of the difficulty of that skill's most difficult task

5. e.g. skill level = 100: can perform all the tasks associated with that skill

Now, because we assume the scales are truncated, this latter interpretation implies, and is implied, by the former. For if a task is of the same difficulty as another, then they are proportionately just as difficult in relation to that skill's most difficult task (i.e. they require the same percentage of the scale's upper limit), and, likewise, are also included in the same difficulty "bracket" (i.e. they are equivalent to the same skill test in the previous aggregate binary outcome interpretation).

So, however one decides to interpret skill levels, the conclusion remains the same: to be skilled in an attribute is to be able to perform the activities associated with that attribute. Put simply, the capacity to act follows virtue, for virtue is, itself, the capacity to act.

3.3.2. Skill Set Composition Axiom

A corollary of both these axioms is that having the sufficient mastery of an occupation' skill set automatically means one is able to perform the job activities of that occupation. As the first axiom's name implies, skills are sufficient.

3.3.3.

3.4. Aggregate Comparative Models

3.4.1. Employability

1. employment levels w_k, w_q
2. total workforce $W = \sum_{q=1}^{n} w_q$
3. $\Lambda_k = \left(\frac{1}{W}\right) \sum_{q=1}^{n} {}_{0}^{1}[s^h_{k,q}] s^h_{k,q} w_q$
4. alternatively, $\Lambda_k = \left(\frac{1}{W}\right) \sum_{q=1}^{n} {}_{0}^{1}[s^h_{k,q}] s^{\geq}_{k,q} w_q$
5. employability = percentage of jobs in the economy on which a candidate could be hired
6. employability can also be defined at specific aggregation levels of the economy (see Labor Economic Models)
7. $\Lambda^?_k = \left(\frac{1}{W}\right) \sum_{?} {}_{0}^{1}[s^h_{k,q}] s^h_{k,q} w_q$
8. ? is the economic aggregation level (e.g. sector, industry, market, role)
9. employability vector $\mathbf{\Lambda}$
10. need to define an axiom to go from x percent of hireability in one job of a class to hireability in x percent of jobs of that class

3.4.2. Competitiveness

1. employment levels w_k, w_q
2. total workforce $W = \sum_{q=1}^{n} w_q$
3. $\mathrm{vs}_k = \left(\frac{1}{W}\right) \sum_{q=1}^{n} {}_{0}^{1}[\ddot{u}_{q,k}]{}_{0}^{1}[s^h_{k,q}] s^h_{k,q} w_q$
4. alternatively, $\mathrm{vs}_k = \left(\frac{1}{W}\right) \sum_{q=1}^{n} {}_{0}^{1}[\ddot{u}_{q,k}]{}_{0}^{1}[s^h_{k,q}] s^{\geq}_{q,k} w_q$
5. competitiveness (versus) = percentage of workers in the economy that could (and would like to) be hired for a particular job
6. competitiveness can also be defined at specific aggregation levels of the economy (see Labor Economic Models)
7. $\mathrm{vs}^?_k = \left(\frac{1}{W}\right) \sum_{?} {}_{0}^{1}[\ddot{u}_{q,k}]{}_{0}^{1}[s^h_{k,q}] s^h_{k,q} w_q$
8. ? is the economic aggregation level (e.g. sector, industry, market, role)
9. competitiveness vector $\mathbf{vs}$

3.5. Labor Economic Models

3.5.1. Economic Taxonomy

1. hierarchical clustering of hireability coefficients
2. number of levels = 1 (i.e. trivial, economy) + L levels, where L is the number for which the levels L and L - 1 (i.e. the previous aggregation) are identical
3. alternatively, use optimization methods to choose number of aggregation levels
4. alternatively, use the same number of aggregation levels as popular economic taxonomies (e.g. ISIC, NAICS, SIC, NACE, OKVED)

3.5.2. Competition in Labor Markets

1. every hireable skill set competes at every level / market on which it is hireable
2. market value, industry value, etc = percent of money in an aggregation level
3. market value, industry employability, etc = weighted employability of workers at that level
4. market value, industry competitiveness, etc = weighted competitiveness of workers at that level
5. etc etc (define every variable for labor market aggregations)

4. Microeconomic Models

4.1. Marginal Human Capital Models

4.1.1. Marginal Compensation or Market Prices

1. wages y_k
2. NNLS attributes vs wages $\rightarrow p_i$
3. p_i is the marginal compensation (i.e. market price) of a point in an attribute
4. market prices vector $\mathbf{p}$

4.1.2. Marginal Time Investment

1. years of education and experience τ_k
2. NNLS attributes vs years of education and experience $\rightarrow \eta_i$
3. η_i is the marginal time investment (expected time of arrival) to gain a point in an attribute
4. marginal time investment (ETA) vector $\boldsymbol{\eta}$

4.1.3. Micro-Flexibility

1. NNLS attributes vs attributes $\rightarrow \phi_{i,j}$
2. micro-flexibility = expected gain on other attributes for an additional point in an attribute
3. !requires additional statistical pressupositions for regression coefficients to be interpretable as $\phi_{i,j}$
4. !potential for recursion
5. !affects ETA
6. define a geometric progression learning model?
7. attribute micro-flexibility vector $\boldsymbol{\phi_i}$
8. attribute micro-flexibility matrix $\boldsymbol{\phi}$

4.2. Aggregate Human Capital Models

4.2.1. Skill Set Value

1. expected compensation based on market prices $\mathbf{p}$ and skill set $\boldsymbol{a_k}$
2. $y_k^{\mathrm{E}} = \mathrm{E}[y_k|\boldsymbol{a_k}, \mathbf{p}] = \sum_{i=1}^{m} a_i^k p_i$
3. use another notation for this?
4. probably define a value variable

4.2.2. Macro-Flexibility

1. attribute macro-flexibility $\Phi_i = \left(\frac{1}{W}\right) \sum_{k=1}^{n} \tilde{a}_i^k w_k$
2. attribute macro-flexibility vector $\mathbf{\Phi}$

5. Additional Comparative Models

5.1. Total Time Investment (ETA)

1. aggregate expected time investment to fill competency gaps
2. uppercase eta η
3. $\mathrm{H}_{k,q} = \sum_{i=1}^{m} \delta(a_i^k, a_i^q)\eta_i$
4. vector $\mathbf{H}_{\boldsymbol{k}}$
5. matrix $\mathbf{H}$
6. p.s.: notation conflicts with hireability matrix (probably change hireability notation)

5.2. Letter-Shaped Skill Sets

1. a generalization of "T-Shaped Skills"

2. refer to the atlas.letters R package

6. Additional Skill Set Models

6.1. Versatility

1. the versatility of a skill set is the weighted sum of its macro-flexibility

2. $\Phi_k = \frac{\sum_{i=1}^{m} a_i^k \Phi_i}{\sum_{i=1}^{m} a_i^k}$

6.2. Leverage? [aggregate micro-flexibility]

1. define a name for aggregate micro-flexibility (leverage?)

2. same procedure as above

7. Roadmap Models

7.1. Career Recommendation

1. this theory needs to be further developed, as I never got the chance to write it down completely

2. see quantitative roadmap sketch for an idea of what this should be

7.1.1. Macro-Strategies

7.1.2. Preference-Adjusted Strategic Matching

7.2. Competency Training Recommendation

7.2.1. Micro-Strategies

7.2.2. Preference-Adjusted Strategic Training

8. Factor-Analytic Models

*8.1. Exploratory Factor Analysis of the O*NET Database*

1. use standard factor-analytic notation (I already reserved many latin and greek characters for this reason)

8.2. Psychometric Questionnaire Optimization

8.2.1. Minimal Factor Representation

8.2.2. Retained Variance

8.2.3. Retained Matching Accuracy and Precision

8.3. Factor-Analytic Intelligence Quotient (IQ) Approximation

1. in psychometrics, general intelligence (IQ) is denoted as g
2. define a expected value notation (also solves the expected compensation, or skill set value, problem above)
3. $g_k^{\mathrm{E}} = \mathrm{E}[g_k | \boldsymbol{a_k}, \boldsymbol{\Psi_?}] = \left(\frac{1}{m_{\psi}?}\right) \sum_{\psi?} \tilde{a}_i^k$
4. $\psi?$ are the subset of items which are proxies for intelligence
5. alternatively, write this coefficient as factor scores (if there is a notation for factor scores)

8.4. Factor-Analytic Comparative Statics

1. there isn't a mathematical notation for this method yet
2. but most or all of its variables and functions have already been defined above
3. refer to the atlas.fstatics R package to understand how this model works

8.4.1. Unbounded Exogenous Impacts

8.4.2. Truncated Exogenous Impacts

8.5. Career Type

1. refer to the atlas.acti R package

9. Writing Plan

9.1. Introduction to Quantitative Career Matching: How to Statistically Pick the Right Occupation

1. introduce the much neglected topic of data-driven career choice and development
2. estimate the baseline Euclidean matching model
3. for simplicity's sake
4. to benchmark matching methods

9.2. Equivalence Estimators for Quantitative Career Matching

1. aeq, seq functions
2. eeq function?
3. take γ_k and c_k parameters as given
4. explain how to calculate generality and competence on the next paper
5. recalculate the Euclidean matching model from paper 1 with these new methods

9.3. Generalists vs Specialists: Who Are the Most Competent Workers?

1. explain how to calculate generality and competence
2. weighted correlation between generality and competence

9.4. Alternative/Advanced Career Matching Methods

1. explain all career matching methods besides the baseline Euclidean model
2. compare all models against one another
3. benchmark models against the baseline Euclidean model
4. select one matching method

9.5. Employability and Competitiveness

1. having defined all the statistical pressupositions for these models in the previous papers, calculate employability and competitiveness coefficients
2. choose another name for this paper?

9.6. A Hierarchical Clustering Labor Market Taxonomy

1. define the hireability taxonomic model
2. choose another name for this paper?

9.7. Competition in Labor Markets

1. continue from where the taxonomy paper left off
2. choose another name for this paper

9.8. Human Capital Flexibility

1. calculate macro-flexibility
2. calculate micro-flexibility
3. aggregate macro-flexibility for each skill set (versatility)
4. aggregate micro-flexibility for each skill set (?)

9.9. Time Investment and Expected Returns of Human Capital

1. choose another name for this paper?
2. calculate marginal compensation of human capital
3. calculate marginal time investment of human capital
4. aggregate marginal compensation of human capital (skill set value)
5. aggregate marginal time investment of human capital (ETA)
6. calculate which careers are most efficient to aim for in the short run (starting from the average skill set)
7. calculate which careers are most effective to aim for in the long run (starting from the average skill set)

9.10. The Career Roadmap: A Mathematical Map to Career Choice and Career Development

1. choose another name for this paper?
2. The Career Roadmap: Statistical Methods for Optimal Career Choice and Career Development?
3. define the concept of macro-strategies
4. combine all statistical metrics so far as the career recommendation coefficient
5. define the concept of micro-strategies
6. combine all statistical metrics so far as the training recommendation coefficient
7. calculate which careers are most efficient to aim for in the short run (starting from the average skill set)
8. calculate which careers are most effective to aim for in the long run (starting from the average skill set)

9.11. Letter-Shaped Career Profiles: A Generalization of "T-Shaped Skills"

1. define the methodology for converting hersheys fonts to career profiles
2. match career profiles against the latin, greek, and cyrillic alphabets
3. find a way to include "from A-to-Z" in this paper's name
4. or from "A-to-Ω"

9.12. Factor-Analyzing the Occupational Information Network Database

1. EFA on O*NET db

9.13. Career Types for the Occupational Information Network Database

1. calculate career types based on the factor model from the EFA paper and the statistics already defined in the previous papers

2. career type molecules

9.14. Factor-Analytic Comparative Statics: Estimating Exogenous Impacts on the Labor Market

1. ai impact analysis

2. some other impact analysis

3. aggregate on the attribute level

4. aggregate on the occupation level

5. aggregate on each taxonomic level defined in the economic taxonomy paper (including the entire economy)

9.15. Factor-Analytic Intelligence Quotient Approximation

1. based on the factor model from the EFA paper, choose a proxy

2. calculate factor-analytic approximation of IQ

3. compare against data