Share of selected ISIC (or ISCO) categories at the global/regional level

Estimation procedure for countries without ISIC (or ISCO) information at the 4-digit level

Authors Bescond David, Kapsos Steve, Limani Donika

Summary: This procedure estimates employment in selected economic sector or occupation categories of interest (ISIC/ISCO) for countries with missing information by leveraging data in countries with detailed information. It calculates the shares of employment in selected categories for countries where data is available and applies these shares to the rest of the countries. The method iteratively estimates weighted shares of selected sectors/occupations for each broad category at all ISIC/ISCO levels. These average shares are weighted based on a country's total employment relative to the total employment in its country income group. The approach assumes that the employment distribution of countries with detailed ISIC/ISCO information is relatively similar to that of countries belonging to the same income group, but that lack this detailed information.

The step-by-step procedure is explained below using a mock-example.

1. **Group of Countries and ISCO Information:**

Let's consider Income Group 1 countries, denoted by $n=\{A,B,C,D,E,F\}$. Among these countries, only two (A and B) have detailed ISCO 4-digit information. Country C has 3-digit data, D has 2-digit data, E has 1-digit data, and F lacks ISCO information (only the total number of employed people is known).

Let us name the digits with d and the subscript l represent the digit levels (d_l). For simplicity, let's assume that the list of 4 digits contains only the following categories:

$$d_4=\{\{1121,1122\},1131,1411,\{2111,2112\}\}$$

The ISCO structure allows us to deduce the rest: $d_3=\{112,113,141,211\}$; $d_2=\{11,14,21\}$; $d_1=\{1,2\}$.

2. **ISCO Categories of Interest**

We have a list of ISCO categories of interest (in this case, automated jobs). This list is available only at 4 digits and defined as: $a=\{1121,2111\}\in d_4$.

At this point we cannot know the share of automated jobs in income group 1, because we only have numbers of employed people at this detailed level for country A and B. The missing data for the rest of the countries will be estimated using all the available information (3 digits structure in C, 2-digit structure in D, and so on.)

3. **Estimation procedure**

Let us note the number of people employed in country n and category d_4 as e_{nd_4}. Step one is to calculate the share of automated jobs per 3-digit category for country A and B.

$$s_{nd_3} = \frac{A_{nd_4}}{E_{nd_3}} = \frac{\sum_{i \in a} e_{¿}}{\sum_{j \in d_4 e_{nj}} ¿¿}$$

Where A_{nd_4} is just the sum of people employed in the selected ISCO 4-digit categories of interest that belong to the same 3-digit category in country n and $E_{nd_3}$is the total employment in that respective ISCO 3-digit category in the same country. For instance, the share of augmented jobs in ISCO 3-digit category 112 for country A would be take the following form:

$$s_{A,112} = \frac{e_{A,1121}}{e_{A,1121} + e_{A,1122}}$$

We would do this for each category and each country, ensuring that 3-digit categories include only relevant 3-digit categories from the ISCO structure (i.e., first 3 digits match). We can then get a weighted average of the 2 countries share for each 3-digit category as follows, where the weight applied is their share of employment in the total employment of countries with ISCO information at 4-digits. The weight is calculated based on the countries with available information at 4 digits as follows: $w_{n4} = \frac{\sum_{i \in d_4} e_i}{\sum_{\substack{j \in \{A,B\} \\ f \in d_4}} e_{jf}}$. For country A, it would be:

$$w_{A,4} = \frac{e_{A,1121} + \ldots + e_{A,2112}}{(e_{A,1121} + \ldots + e_{A,2112}) + (e_{B,1121} + \ldots + e_{B,2112})}$$

The imputed weighted share of automated jobs in each 3-digit ISCO categories is then just the sum of the product of each country share and their respective weights as follows:

$$S_{d_3} = \sum_{i \in \{A,B\}} s_{id3} * w_{i4}$$

We can now multiply this share to the number of people employed in that d_3 category in country C to get the number of people with automated jobs, which we can call A_{C,d_3}, as follows: $A_{Cd_3} = S_{d_3} * e_{C,d_3}$

We do this for all 3-digit categories, sum the numbers up and we have the total number of people employed in the selected ISCO categories (in this case, automated jobs) in country C.

$$A_C = \sum_{i \in d_3} S_i * e_{Ci}$$

We repeat the steps above to include next country with 2-digit occupation information (country D). Now instead of calculating the share of automated (4-digit) jobs in the 3-digit category, we check the share in the 2-digit category. The share equation above becomes:

$$s_{nd_2} = \frac{A_{nd_3}}{E_{nd_2}} = \frac{\sum_{i \in a} e_{¿}}{\sum_{j \in d_2 e_{nj}} ¿¿}$$

Since there is no 4-digit information for country C (i.e., missing $e_{C,1121}$), we can instead use the estimated number of augmented jobs A_{C,d_3} (defined above) when computing the weighted average share of automated jobs in 2-digit categories. For instance, for country C: $s_{C,11}=\frac{A_{C,112}+A_{C,113}}{E_{C,11}}$. The weights also change to reflect the addition of a new country estimate in our calculations. $w_{n3}=\frac{\sum_{i \in d_3} e_i}{\sum_{\substack{j \in \{A,B,C\} \\ f \in d_3}} e_{jf}}$. The estimated share of automated jobs in 2-digit categories then becomes:

$$S_{d_2}=\sum_{i \in \{A,B,C\}} s_{id2} * w_{i3}$$

We can now multiply this share to the number of people employed in that $d_2$category in country D to get the number of people with automated jobs in that category as follows: $A_{Dd_2}=S_{d_2} * e_{Dd_2}$. If we do this for all 2-digit categories, sum the numbers up and we have the number of all automated jobs in country D.

$$A_D=\sum_{i \in d_2} S_i * e_{Di}$$

The same logic applies to countries with 1-digit. Again, we repeat the steps above to include next country with 1-digit ISCO information. So now instead of calculating the share of automated (4-digit) jobs in the 3- or 2-digit category, we check the share in the 1-digit category. The first equation above becomes:

$$s_{nd_1}=\frac{A_{nd_2}}{E_{nd_1}}=\frac{\sum_{i \in a} e_{¿}}{\sum_{j \in d_1 ¿} e_{nj} ¿¿}$$

For the country with only 2-digit information, we would use the estimated number to calculate its share in the 1 digit like we did for country C above, for instance $s_{D,1}=\frac{A_{D,11}}{E_{D,1}}$. Again, the weights change to reflect the addition of a new country estimate (Country D) in our calculations: $w_{n2}=\frac{\sum_{i \in d_2} e_i}{\sum_{\substack{j \in \{A,B,C,D\} \\ f \in d_2}} e_{if}}$. The estimated share of automated jobs in 1-digit categories then becomes:

$$S_{d_1}=\sum_{i \in \{A,B,C,D\}} s_{id1} * w_{i2}$$

We can now multiply this share to the number of people employed in that d_1 category in country E to get the number of people with automated jobs as follows: $A_{Ed_1}=S_{d_1} * e_{Ed_1}$. We do this for all 1-digit categories, sum the numbers up and we have the number of all automated jobs in country E.

$$A_E=\sum_{i \in d_1} S_i * e_{Ei}$$

Finally, we have to estimate employment in specific categories for the country with no ISCO information. In this case, we can impute the figure by using simply the average of the share of automated jobs in overall employment. This weighted share is based on the data from countries with ISCO 4 and the imputed figures for countries which had ISCO at 3,2, or 1 digit.

$$s_n = \frac{A_n}{E_n}\ ;\ w_{n1} = \frac{E_n}{\sum_{i \in \{A,B,C,D,E\}} E_i}$$

$$S = \sum_{i \in \{A,B,C,D,E\}} s_i * w_{i1}$$

We finally apply this weighted mean to the overall employed population of country F in order to get the overall augmented employed people in country F.

$$A_F = S * E_F$$

Total number of automated jobs in income group 1 is then just the sum of all the jobs in each of the countries (actual and imputed).

$$A = \sum_{i \in \{A,B,C,D,E,F\}} A_i$$

An excel sheet with formulas corresponding to this mock example is attached for demonstration purposes.