**Data sources**

**RAIS**

To obtain information about the formal labor market in Brazil, we use the Brazilian matched employer-employee data set called RAIS (*Relação Anual de Informações Sociais*), which is an annual census of formal workers in the country. RAIS is an administrative data set administered by the Brazilian Ministry of Labor and that has detailed information about all formal workers in the country. Every year firms must submit information about their employees and are subject to fines if they do not comply with the submission deadlines. For the main analyses, we use the yearly RAIS data sets for the 1995-2021 period and the sectoral classification is based on CNAE 95 at the 5-digit level.[[1]](#footnote-1)[[2]](#footnote-2)

**Exports (MDIC)**

<Carlos to be describe that>

**Population census**

We use the Brazilian Population censuses data for 2000 and 2010 to analyze both the formal and informal labor markets at the microregion-sector level considering the 5-digit CNAE 95 categorization. The latest census data available refers to 2010 and the oldest one dates from 1960. Since the interest of our analysis is on the most recent period of exports expansion in Brazil, we restrict the analysis to the 2000 and 2010 census. The specific reason for not using the 1991 census is that the sectoral classification used in RAIS data is CNAE 95, which is was created after the 1991 census was fielded, and the Brazilian NSO does not publish a concordance from the 1991 census-specific categorization to CNAE 95. Therefore, one would need to make several assumptions in an ad-hoc concordance to produce such a mapping, and the quality of the resultant concordance is unknown, what could bring additional noise to the analysis. The definition of formal labor market is based on either having a formal job contract (i.e., “carteira assinada”) or whether the worker contributes to social security.[[3]](#footnote-3)

**Greenhouse gas emissions**

The information on GHG emissions by sector was extracted from Cirera and Martins-Neto (2021), which compiled this information at the 3-digit CNAE 95 level based on the Brazilian Initial National Communication to the United Nations Convention about Global Climate Change for the 1994 reference year. The unit of measure is tons of emissions and to use the same level sector classification from RAIS and census data, we converted the 3-digit CNAE 95 data into 5-digit CNAE 95 data.

**Appendix A**

**<To be revised>**

**Using a green/brown sector classification**

**1.1 RAIS data**

To use the green/brown sector classification proposed by Cirera and Martins-Neto (2021), we had to map the CNAE 2.0 classification at the 4-digit level to the 5-digit CNAE 95 level because the latter is the one used in RAIS data. To do so, we had to use two concordances published by IBGE, one from CNAE 2.0 to CNAE 1.0 and another from CNAE 1.0 to CNAE 95. Both concordances are at the 5 digit level. This implies that the first step of the analysis was to map the 4-digit CNAE 2.0 sectors and their green/brown classification – as defined by Cirera and Martins-Neto (2021) to the 5-digit CNAE 2.0 codes and then use the aforementioned two concordances.

One particular challenge in this process is that the concordances published by IBGE are not one-to-one mappings in that a given 5-digit CNAE 2.0 code, for example, is mapped to only one 5-digit CNAE 1.0 code. To ensure that we use all 4-digit CNAE 2.0 codes from Cirera and Martins-Neto (2021) for which we have the green/brown sector classification, we transformed the concordances into one-to-one mappings. In cases in which a given code in a given classification (e.g., CNAE 2.0) was mapped to more than one code in another classification (e.g., CNAE 1.0), we only keep the first entry as defined by IBGE.

**1.2 Census data**

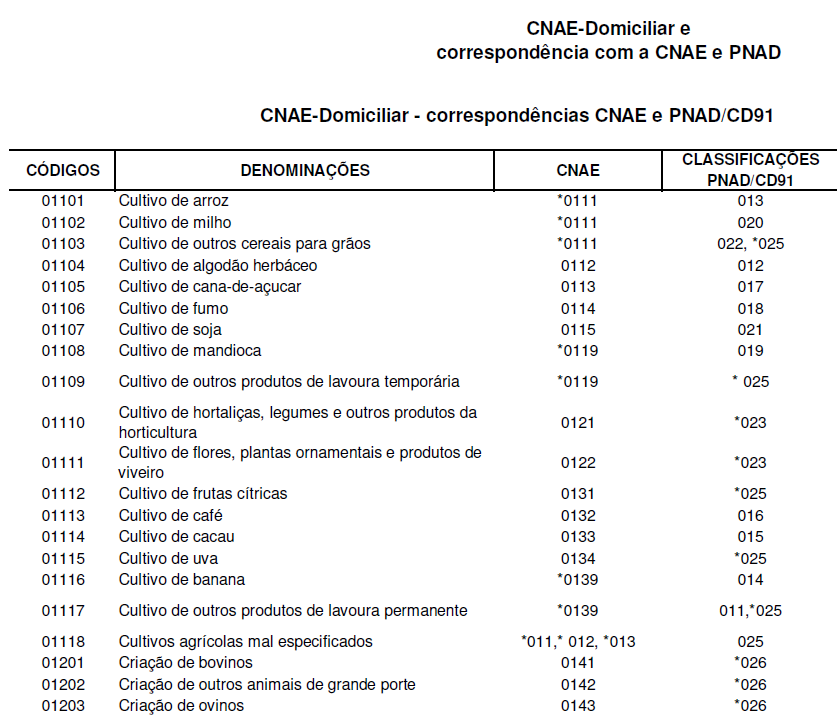
When using census data, we kept the same sectoral classification that was used in the RAIS data for the 1995-2021 period, i.e., CNAE 95 at the 5 digit level. The 2000 census uses CNAE Domiciliar 1.0, while the 2010 census uses CNAE Domiciliar 2.0 to classify the sectors at the XXX digit level. This entails that to use the same classification used in RAIS, we need to use two concordances published by IBGE: one from CNAE Domiciliar 2.0 to CNAE Domiciliar 1.0 and another one from CNAE Domiciliar 1.0 to CNAE 95. As in the case of RAIS data, we kept only the first entry as defined by IBGE when the mapping was not one-to-one.

*Mapping the CNAE 95 classification to ISCI 3.0 at the X-digit level*

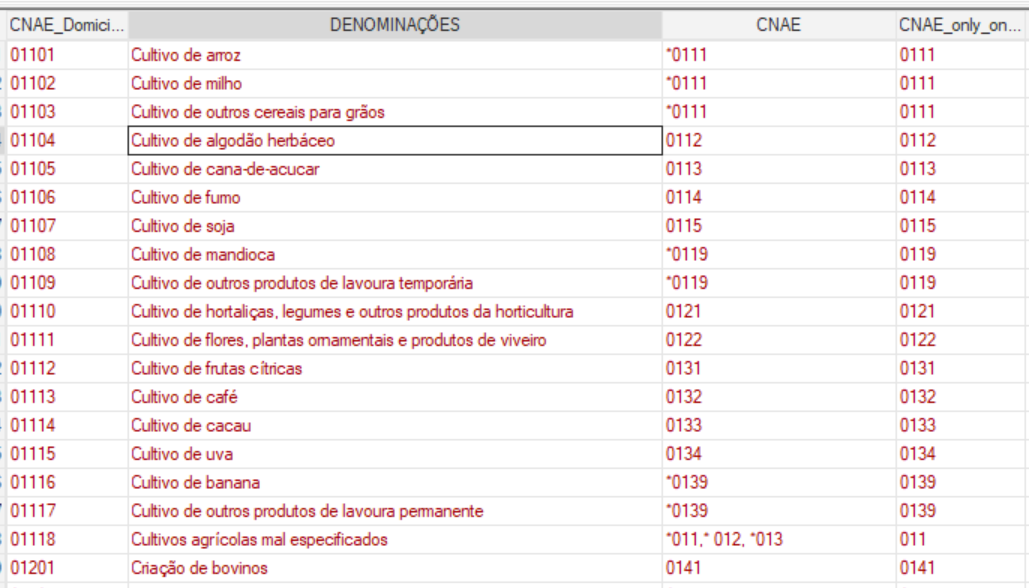
*<Carlos to complete>*

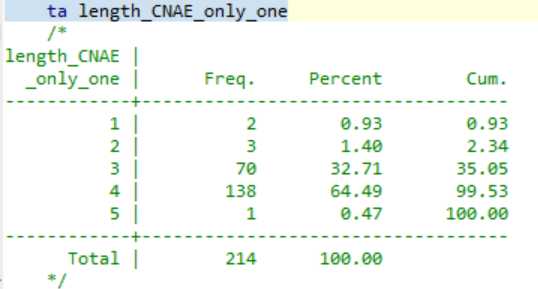
*Transforming 2000 census classification into the one used in RAIS*

IBGE published the mapping from CNAE Domiciliar 1.0 (2000 census) to CNAE (RAIS classification) in a pdf. So the 1st step was to use the ExtractTable tool, do some manual corrections to the Excel file and import it into Stata. To my surprise, the 1st challenge is that the mapping is not 1 to 1, but 1 to many in several cases. The 1st challenge is that we have 1 to m in the mapping from the 5-digit CNAE Domiciliar 1.0 (2000 census) to the 5-digit RAIS classification, which means that we should decide how to transform the 1:m mapping into a 1:1 concordance. In the screenshot below, for example, you can see that the category 01118 maps to 3 different CNAE codes in RAIS. What is even more challenging is that the 5-digit CNAE Domiciliar 1.0 classification is mapped to 3 different 3-digit CNAE codes in RAIS. This means that the mapping is not from a 5-digit classification to another 5-digit classification. In particular, the 5-digit classification in 2000 census maps to a 5-digit classification in RAIS for 65% of the observations, but in 32% of the cases it is mapped to a 4-digit classification in RAIS. So there are many decisions to be made and that could affect the results of the trade paper because the shock takes place at the microregion-sector-year level and depending on the sector to which we assign a given 5-digit 2000 census classification it will be artificially changing the results (compared to a situation in which we have a 1:1 perfect 5-digit to 5-digit mapping). The 1st solution that I considered was to assign the category 01118, for example, to the 1st 3-digit RAIS code to which it is mapped but this is clearly imperfect because by doing this we will be artificially dismissing the other 2 3-digit RAIS codes to which 01118 is mapped and at the same time giving a “higher” weight for the 1st RAIS code. Also, since we the 1st destination RAIS code is at the 3 digit level we would need to define to which 5-digit level RAIS code we would map the 01118 code from the 2000 census classification (we can also decide for simplicity to assign it to the 1st 5-digit valid RAIS code). So please tell me if you can think of a better solution.



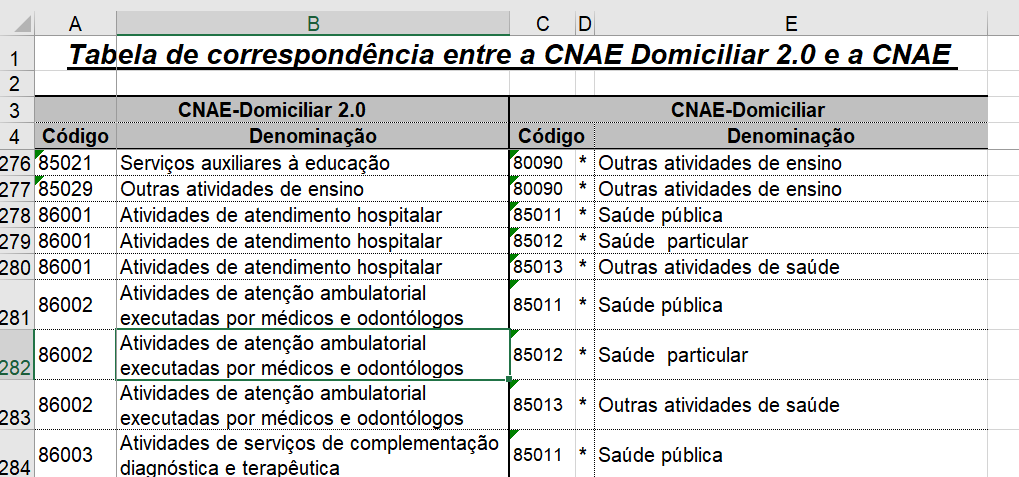
See how the data looks like in Stata:

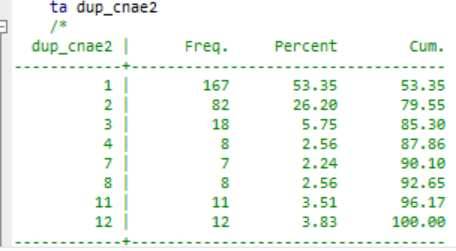




*Transforming 2010 census classification into the one used in RAIS*

To transform the 2010 census classification into the one used in RAIS we need to 2 concordances, one from CNAE Domiciliar 2.0 to CNAE Domiciliar 1.0 and then use the concordance done above (i.e., from 2000 census to RAIS) to map the CNAE Domiciliar 1.0 to CNAE. For the 1st concordance, IBGE fortunately publishes a spreadsheet (screenshot below). The only challenge related to this concordance is that it is m:1 for some categories and not fully 1:1, but the good news is that the concordance is from a 5-digit classification to another 5-digit classification for all categories. In particular, there is a category in the RHS with up to 12 different RAIS codes in the LHS (2nd screenshot). So we need to decide how to transform the 1:m mapping into a 1:1 for such categories.





1. RAIS 2022 was available at the time this paper was written but due to methodological changes in the way RAIS was produced in 2022, the Brazilian Ministry of Labor does not ensure comparability with previous years (for details, see the following link: https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/estatisticas-trabalho/rais/rais-2022/nota-tecnica-rais-2022.pdf). For this reason, we restricted the analysis to the most recent year immediately before 2022. [↑](#footnote-ref-1)
2. Several steps were followed to transform the 2000 census- (i.e., CNAE Domiciliar 1.0) and the 2010 census-specific (i.e., CNAE Domiciliar 2.0) sectoral classifications into CNAE 95. These are described in the Appendix. [↑](#footnote-ref-2)
3. More specifically, the formal labor market is composed of both private- and public-sector waged workers, including domestic ones, with a formal job contract and employers that contribute to social security. The informal labor market is made up of the remaining categories of workers in censuses data. [↑](#footnote-ref-3)