TERRITORIAL DIAGNOSIS OF LABOR MARKET TENSIONS IN FRANCE

Unédic X Ecole Polytechnique, ENSAE, Télécom Paris

Alfonso Awadalla Carreño[1], Cynthia Francis[2], and Anahi Reyes Miguel[3]

March 27, 2025

Abstract

This paper presents the development of an interactive mapping tool designed to diagnose labor market tensions across French territories. Using both web-scraped and administrative data to construct the supply and demand sides of the labor market, the interactive tool highlights mismatches between job supply and demand at the employment zone level, disaggregated by sector and occupation. This tool aims to support more detailed territorial diagnoses for policymakers, researchers, and the general public.

Keywords: labor market, job vacancies, unemployment, France, spatial analysis

1. Introduction

The French labor market is characterized by significant tensions, with high recruitment difficulties and mismatches between labor supply and demand. Despite a high unemployment rate, many companies face challenges in finding suitable candidates, particularly in sectors such as construction, health, and social services (Bergeaud et al., 2022). These tensions vary significantly by region and profession across France, influenced by local economic dynamics and skill mismatches. For instance, during the 2020 COVID-19 pandemic, the sectors most affected were restauration, construction, and commerce (INSEE, 2021). Some regions experience severe shortages of skilled labor, while others struggle with high unemployment rates among low-skilled workers.

1. [1]alfonso.awadalla@polytechnique.edu, [2]cynthia.francis@polytechnique.edu, [3]anahi.reyes-miguel@polytechnique.edu

To help address this challenge, we developed an interactive and publicly accessible mapping tool in partnership with Unédic. The tool is designed to visualize labor market tensions across employment zones, disaggregated by sector and occupation. Unlike existing dashboards that focus on national or broad regional trends, our tool enables more granular territorial diagnoses by integrating multiple data sources and spatializing key indicators.

Specifically, we combine job vacancy data from the Job Offers Collection and Analysis System (JOCAS) by Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) with administrative unemployment data from the Statistiques Mensuelles du Marché du Travail (STMT) provided by France Travail. This integration allows us to approximate labor supply and demand and to construct a standardized Market Tightness Indicator (MTI), highlighting territorial mismatches between job offers and jobseekers at both the geographic and occupational levels. By visualizing these indicators spatially, stakeholders can identify areas with high recruitment tensions and gain detailed insights into the specific occupations affected.

This report documents the methodology behind the MTI, the technical implementation of the mapping tool, and its potential use cases.

2. Methodology

Labor Market Tightness (LMT) is a key concept in labor economics that captures the balance between labor demand and supply (Cahuc et al., 2016). It reflects how difficult it is for employers to find suitable candidates, or conversely, how hard it is for job seekers to find work. Our interactive tool measures LMT through a standardized MTI, enabling spatialized diagnoses of labor mismatches across occupations and territories.

2.1 Data

Job Offers Data

The Job Offers Collection and Analysis System (JOCAS) dataset compiles online job offers collected daily by DARES. It provides detailed information on job vacancies published online across France (excluding Mayotte) during 2020. The dataset aggregates postings from 18 different recruitment websites, covering a broad range of occupations, contract types, and qualification levels. The data is available at the commune level.

Job Seekers Data

The *Statistique Mensuelle du Marché du Travail* (STMT) dataset provides monthly statistics on jobseekers registered with France Travail. We use data corresponding to the year 2020. This dataset includes only jobseekers in Categories A, B, and C, as defined by INSEE (INSEE, 2023), which are defined as follows:

- **Category A:** Jobseekers who are unemployed and have no professional activity.
- **Category B:** Jobseekers who are unemployed but have carried out a reduced activity of up to 78 hours per month.
- **Category C:** Jobseekers who are unemployed but have carried out a reduced activity exceeding 78 hours per month.

The STMT data is available at the communal level but is limited to communes with over 5,000 inhabitants.

Geographic Reference Data

To perform the spatial analysis and link local data to territorial units, we rely on two geographic sources:

- **Communes:** We use the January 2022 shapefile of French communes from the administrative boundaries provided by OpenStreetMap, available via *data.gouv.fr* (OpenStreetMap contributors, 2022). This shapefile allows us to work with a consistent and up-to-date geographic reference for all municipalities in France, including overseas.
- **Employment Zones (Zones d'emploi):** We use the 2020 zoning of employment areas provided by INSEE, which reflects commuting patterns and local labor market structures. This territorial partition is tailored for local labor market analysis and serves as a reference for unemployment rates, job estimates, and spatial diagnostics (INSEE, 2022). Each commune is associated with its corresponding employment zone to enable aggregation and mapping.

2.2 Measurement Strategy

In our tool, we measure LMT through a Market Tightness Indicator (MTI), defined as the exit rate from unemployment. We construct this indicator inspired by *Les tensions sur le marché du travail* (DARES, 2023), based on the standardized ratio between job vacancies and registered unemployed individuals.

This is proxied by the vacancy-to-unemployment ratio, denoted by θ (Cahuc et al., 2016). For each employment zone ze, occupation f, and month t, we define the indicator as:

$$\mathrm{MTI}^{(i)}_{ze,f,t} = \frac{\theta^{(i)}_{ze,f,t} - \mu^{(i)}_t}{\sigma^{(i)}_t} \tag{1}$$

where

$$\theta^{(i)}_{ze,f,t} = \frac{V_{ze,f,t}}{U^{(i)}_{ze,f,t}} \tag{2}$$

is the vacancy-to-unemployment ratio (*Job offers / Job seekers*) in category $i \in \{\mathrm{A}, \mathrm{ABC}\}$, where ABC is defined in section 3.3, and $\mu^{(i)}_t$, $\sigma^{(i)}_t$ are the monthly national mean and standard deviation of the ratio for the same category. This standardization removes seasonal effects and enables consistent comparisons across space and occupations.

The raw ratio $\theta^{(i)}_{ze,f,t}$ provides a straightforward interpretation:

- $\theta^{(i)}_{ze,f,t} = \frac{V_{ze,f,t}}{U^{(i)}_{ze,f,t}} > 1$: Tight market where there are more job offers than job seekers. This indicates that employers face difficulties in hiring, often leading to upward pressure on wages.
- $\theta^{(i)}_{ze,f,t} = \frac{V_{ze,f,t}}{U^{(i)}_{ze,f,t}} \leq 1$: Slack market where there are equal or more job seekers than job offers. This suggests greater competition among workers and reduced wage pressure.

Finally, to obtain a stable yearly indicator of tension per zone and occupation, we average these monthly scores over the full year:

$$\mathrm{MTI}^{(i)}_{ze,f} = \frac{1}{12}\sum_{t=1}^{12} \mathrm{MTI}^{(i)}_{ze,f,t} \tag{3}$$

This transformation captures the average monthly standardized tightness for each zone and occupation over the year. The resulting annual score is used for scoring and mapping, providing a summary measure of relative labor market tightness throughout the year. The score and categorization explained in 2.3 will allow for a more straightforward interpretation.

2.3 Scoring System

To make the standardized labor market tightness measures interpretable and comparable, we construct a scoring system that ranges from 1 (slackest) to 5 (tightest). The scores are derived from **quantile-based bins** applied to the annual average of the monthly standardized z-scores. Each score category corresponds to a specific range of raw supply/offer ratios, where a ratio greater than 1 indicates a tight market, and a ratio less than 1 indicates a slack market. Specifically, categories 1 to 3 correspond to markets with raw ratios consistently below 1, indicating slackness, while categories 4 and 5 correspond to raw ratios above 1, reflecting a relative labor shortage. This approach allows us to define five categories of labor market tightness, combining the standardized MTI and the raw ratio's clear interpretation. The methodology ensures a balanced distribution of scores and enables meaningful comparisons across occupations and geographic areas.

The use of five categories aligns with the methodology developed by DARES for their labor market tension indicators (DARES, 2022). The interpretation of each score is as follows:

Score	**Tightness Level**	**Category A Quantiles**	**Category ABC Quantiles**
1	Labor Surplus	0.0 to 0.12	0.0 to 0.09
2	Available Labor	0.12 to 0.36	0.09 to 0.28
3	Fragile Balance	0.36 to 1.06	0.28 to 1.0
4	Relative Labor Shortage	1.06 to 3.01	1.0 to 3.0
5	Severe Labor Shortage	3.01 to 122.86	3.0 to 122.86

Table 1: Labour Market Tightness Score Scale

3. Technical Implementation

This section outlines the technical workflow implemented to develop the MTIs and their integration into a user-friendly, interactive mapping tool. It covers data preprocessing, spatial matching of communes to employment zones, aggregation by occupation and month, the computation and standardization of tightness ratios, and the design of the final visualization interface.

3.1 Programming Environment

The analysis and dashboard development were conducted using the **Python** programming language. The following key libraries and tools were used:

- **pandas**: For loading, cleaning, transforming, and aggregating tabular data.
- **geopandas**: For handling spatial data and joining data frames with shapefiles.
- **plotly**: For building interactive graphs and maps.
- **dash**: To develop the interactive web application and connect filters, maps, and graphs.
- **matplotlib**: For static plotting and exporting figures.
- **unidecode** and **re**: For standardizing and cleaning textual data (e.g., commune names).

All scripts were developed in Python and integrated into an interactive dashboard using `Dash`. The resulting tool enables users to dynamically explore tightness indicators across space, occupations, and different jobseeker categories.

3.2 Data Preprocessing

3.2.1 Job Offers - JOCAS

The raw job offers data were organized by looping through source folders and their respective monthly subfolders. The cleaning process included standardizing commune names to ensure consistency, filtering for communes with populations exceeding 5,000 inhabitants, and removing entries with missing Répertoire Opérationnel des Métiers et des Emplois (ROME)[1] job codes.

The job codes from the ROME classification were aggregated to the broader FAP (Famille Professionnelle) system. FAP offers a higher level of aggregation, so grouping ROME codes into occupational families facilitates the analysis. It also aligns better with policy-relevant occupational categories, improving the clarity of the results and enabling easier interpretation for policymakers and nonspecialist users of the tool. This choice is

1. ROME is a classification system developed by Pôle emploi that categorizes jobs based on skill sets and professional domains

inspired by the methodology used by DARES in their labor Market Tension Indicators, which also rely on FAP as the main occupational classification (DARES, 2023).

To perform the aggregation from ROME codes to FAP families, we use the official correspondence table published by DARES (2024). Each ROME code is mapped to its corresponding FAP87 and FAP22 category using this table. The FAP classification system groups occupations into professional families based on shared characteristics such as required skills or work environments. We use both FAP22 (a more aggregated classification into 22 broad families of occupations) and FAP87 (a finer classification into 87 detailed families of occupations) to enable flexibility in the analysis. FAP22 allows for higher-level overviews, while FAP87 offers more granular insights into specific occupational tensions. The resulting dataset was then aggregated by commune, FAP category, and month, with job offers summed for each grouping.

To ensure the robustness of the processed dataset we conducted the following frequency checks. The frequency of job offers totaled 963,800 across all records. The dataset achieved full coverage of all 2,230 communes meeting the population threshold and included all 86 FAP87 job categories. Furthermore, temporal consistency was verified across all 12 months of data in 2020. These preprocessing steps ensured a clean and reliable dataset for subsequent analysis.

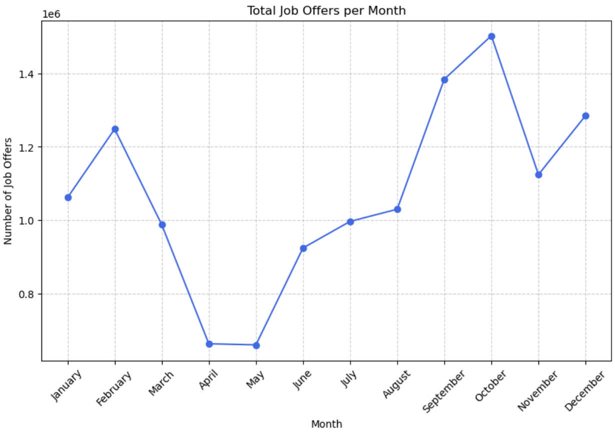
To ensure our data is robust, we compared the JOCAS dataset to a Pôle emploi report on total job vacancies in France for the same year (Figure 1). The graph from the JOCAS dataset (Figure 1a) illustrates the monthly progression of job offers, showing fluctuations throughout the year. This pattern reflects seasonal trends in job postings, likely influenced by hiring cycles and economic factors. The Pôle Emploi graph (Figure 1b) shows the number of job offers disseminated across different time periods from 2015 to early 2021. The bars are divided into quarters (T1, T2, T3, T4) for each year.

However, our JOCAS dataset reports overall lower values than Pôle Emploi due to several methodological differences:

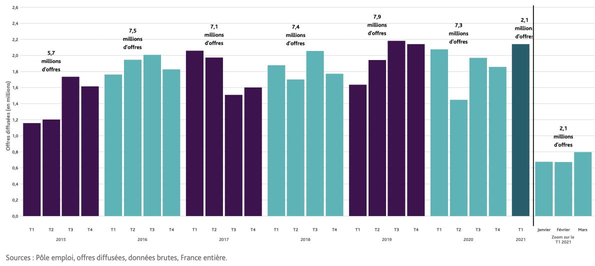
- **Sample Size**: The JOCAS data focuses exclusively on communes with more than 5,000 inhabitants, which inherently limits geographic scope but intensifies data concentration. The Pôle Emploi dataset covers the entire French terrirory.
- **Mapping Methodology**: Transitioning from ROME codes to FAP codes introduces some duplication of job offers due to overlaps in classification systems.
- **Online Posting Bias**: Online job ads tend to disproportionately represent higher-skilled positions, as employers are more likely to post vacancies for managerial or

professional roles online than for low-skilled or elementary positions. This may lead to an underrepresentation of low-skilled job offers in the dataset.

These methodological differences explain discrepancies in absolute values between the datasets. While their trends are slightly similar, both datasets show seasonal fluctuations in job offers. Notably, there is a visible drop in job offers during T2, coinciding with the onset of COVID-19 restrictions, which impacted hiring cycles across France.



(a) Source: JOCAS Dataset

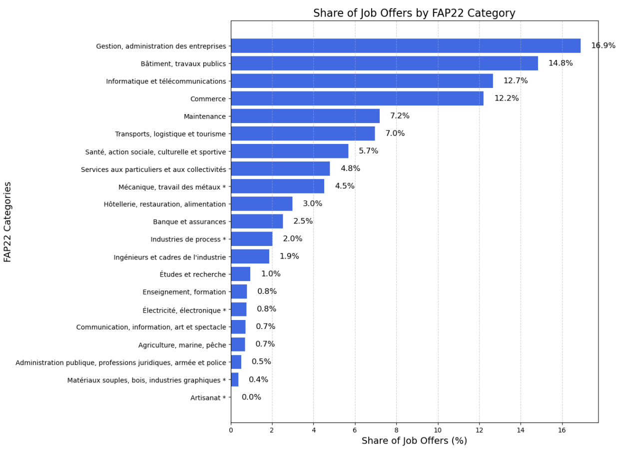


(b) Source: Pôle Emploi (France Travail)

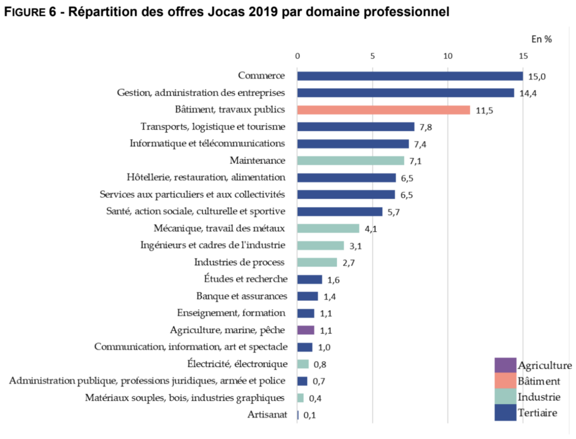
Figure 1: Comparison of job offers from JOCAS and Pôle Emploi datasets.

Another robustness check we conducted was by comparing job categories between our JOCAS dataset and the DARES dataset (Figure 2). The JOCAS dataset (Figure 2a) illustrates the distribution of job offers by FAP22 categories within French communes with more than 5,000 inhabitants in 2020. The top categories include "Gestion, administration des entreprises," "Bâtiment, travaux publics," and "Informatiques et télécommunications,' which collectively account for a significant share of postings. This distribution reflects urban hiring trends in sectors that rely heavily on online recruitment platforms. In contrast, the DARES dataset (Figure 2b) represents job offers across all French territories in 2019. It covers similar professional domains, including “Gestion, administration des entreprises,” and “Bâtiment, travaux publics.”

Despite differences in scope (communes with more than 5000 inhabitants vs. national) and year (2020 vs. 2019), both datasets exhibit similar patterns in the relative proportions of job categories.



(a) Source: JOCAS Dataset



(b) Source: DARES

Figure 2: Comparison of job offers from JOCAS and DARES datasets.

3.2.2 Job Seekers - STMT

Similarly, to prepare the STMT dataset for analysis, files were processed to standardize commune names, restructure jobseeker data, and merge it with job classification codes (FAP87) to enable comparisons with job supply data. The cleaned dataset was aggregated by commune, FAP87 categories, and month, providing a structured overview of jobseeker trends. Robustness checks confirmed the reliability of the dataset: it included 808,627 jobseeker records, achieved full coverage of all 2,230 communes meeting the population threshold, and represented all 86 FAP87 job categories. Furthermore, temporal consistency was verified across all 12 months of data.

Additional robustness checks were conducted on the demand datasets. Figure 3a illustrating jobseeker trends in major French communes for 2020 exhibits similar patterns to the France Travail report (Figure 3b) on total job vacancies for the same year. Despite these similarities, discrepancies in jobseeker figures arise from several key factors:

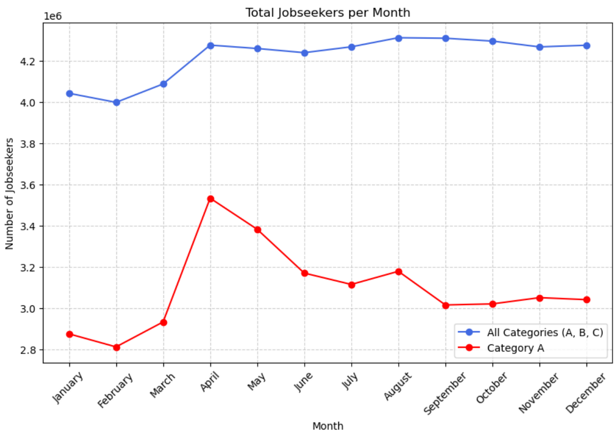
- **Exclusion of Small Communes**: Our analysis focuses exclusively on communes with more than 5,000 inhabitants. This exclusion leads to a downward bias in our numbers, as smaller communes contribute significantly to the overall jobseeker count. For instance, in January, these smaller communes accounted for approximately 1 million jobseekers out of a total of 3.7 million. This omission results in underestimation compared to reports that include all communes.

- **Raw vs. Adjusted Data**: The figures from France Travail are adjusted for seasonal variations and working-day effects, which can significantly impact the reported numbers. In contrast, our dataset consists of raw values due to limitations in data availability. This difference in data processing introduces uncertainty in comparing the two datasets directly.

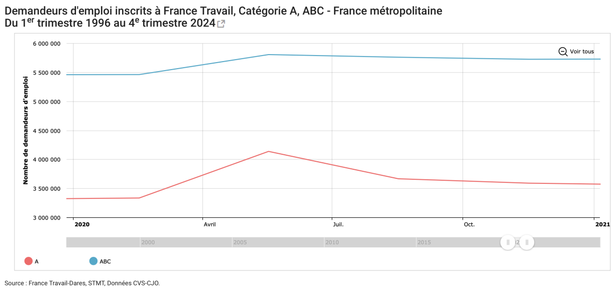
- **Mapping ROME to FAP Codes**: The transition from ROME codes to FAP codes introduces potential duplication of jobseekers. Since a single ROME code can correspond to multiple FAP codes, our assumption that individuals with the same ROME code can be linked to any matching FAP may inflate the jobseeker count.

- **Data Availability and Limitations**: The publicly available data from France Travail provides the number of jobseekers per commune and ROME code, but it may not fully align with our dataset due to differences in data processing and availability.

These discrepancies highlight methodological variations between the datasets but do not detract from their overall alignment in trends.



(a) Source: STMT Dataset



(b) Source: France Travail

Figure 3: Comparison of job seekers from STMT and France Travail datasets.

3.3 Ratio and Score Computation

In line with the methodological choices made by DARES, we restrict our analysis to unemployed workers (U) falling under Categories A, B, and C of registered jobseekers with France Travail, all of whom are required to actively seek employment. These individuals may or may not be receiving unemployment benefits and may still be engaged in some form of employment (INSEE, 2023). The categories used for the calculation are defined as follows:

- **Category A:** Jobseekers who are unemployed, actively seeking, and immediately available for employment.
- **Category ABC:** The combined group of Categories A, B, and C, representing all individuals actively seeking employment, regardless of reduced activity.

We compute two versions of the MTI: one using only Category A, and another using the broader ABC category to capture a wider segment of the jobseeking population. This is done through the following steps:

- **Handling zeros and missing values:** To prevent division errors, zero jobseeker counts are replaced with one. Missing values in job offers or jobseeker data are filled with zero.
- **Standardization:** To enable comparison across time, space, and occupations, monthly z-scores of the tightness ratio are computed for each employment zone and occupational category. This adjustment corrects for structural and seasonal variations.
- **Annual aggregation:** For each zone–occupation pair, we compute the average of monthly z-scores over the year. This provides a synthetic, normalized indicator of annual labor market tightness.
- **Scoring:** These yearly averages are transformed into a five-point ordinal scale using quantile-based bins. A score of 1 indicates the slackest markets, and 5 the tightest. This approach, inspired by the DARES methodology, ensures interpretability across occupations and regions.

3.4 Visualization Tools

We visualise the computed indicators through both static and interactive components, with a focus on spatial and temporal variations in labor market tightness.

Dash (Plotly): An interactive dashboard is built using Dash, allowing users to explore labor market tightness based on multiple filters:

- *Job Category:* Choose between FAP 22 (broad professional sectors) or FAP 87 (specific professions).
- *Jobseeker Type:* Select between Category A (fully unemployed) or Categories A, B, and C (including part-time unemployment or training).
- *Profession or Sector:* Narrow down to a specific job family within the selected classification.
- *Geographic Focus:* Zoom into a department to view localized tightness indicators.

Interactive Outputs: The application displays a dynamic map of tightness scores and an accompanying time series chart showing monthly evolution within the selected zone d'emploi. All outputs update in real-time based on user interaction.

Export and Access: Users can download the underlying data used for the selected area. The dashboard is fully deployable on web platforms for broader dissemination.

4. Limitations and Considerations

While the labor market tightness tool provides valuable insights, several limitations should be acknowledged:

- **Geographic and data coverage gaps:** Some communes or employment zones are excluded due to missing data or unmatched geographic references. In particular, our analysis is restricted to communes with more than 5,000 inhabitants, as job vacancy data below this threshold is often sparse or unavailable. This introduces a coverage bias, especially in rural or less populated areas. Additionally, four employment zones are missing entirely from our dataset: Château-Gontier, Ghisonaccia, Mayotte, and Propriano.
- **Aggregation uncertainty:** Assigning communes employment zones introduces some imprecision, particularly when a commune spans multiple zones.
- **Interpretation over time:** Since tightness scores are standardized within each year, they are not directly comparable across different years. If future updates of the tool include multi-year data, we recommend applying standardization at the yearly level to enable consistent temporal comparisons.

- **Descriptive tool only:** The tool is exploratory and descriptive in nature. It does not support causal inference and should be used accordingly.

5. Conclusion

This tool provides a systematic, spatially-explicit approach for measuring and visualizing labor market tightness across France. By combining granular job offer and jobseeker data with a standardized scoring methodology, it facilitates improved monitoring of regional and sectoral labor market imbalances. It offers valuable support for employment policy planning and vocational training, and lays a strong foundation for further research and public data dissemination.

Looking ahead, several improvements could enhance the analytical depth and usability of the tool. A key extension would be to incorporate complementary indicators—such as recruitment delays, contract quality, or skill mismatches—to move beyond descriptive scores and enable a multidimensional diagnostic of labor market tensions. Additionally, integrating demographic and contract-related filters, including qualification level, age, gender, and contract type, could allow for more targeted and nuanced interpretations of labor dynamics across territories.

Ongoing development efforts could focus on refining geographic matching procedures, expanding dashboard interactivity, and integrating new data sources to broaden the tool's coverage and policy relevance.

References

Bergeaud, A., Cette, G., and Stary, J. (2022). Recruitment difficulties and firms' characteristics: An analysis of french company data. *Economie et Statistique / Economics and Statistics*, 534-35:43–59.

Cahuc, P., Charlot, O., Malherbet, F., Benghalem, H., and Limon, E. (2016). Taxation of temporary jobs: Good intentions with bad outcomes? IZA Discussion Paper 10352, Institute for the Study of Labor (IZA).

DARES (2022). Les tensions sur le marché du travail en 2022. Technical report, Ministère du Travail, du Plein Emploi et de l'Insertion. Dernière consultation: mars 2025.

DARES (2023). Note méthodologique sur l'analyse des tensions sur le marché du travail. *Dares*. Accessed March 2025.

DARES (2024). La nomenclature des familles professionnelles 2021. *Dares*. Accessed March 2025.

INSEE (2021). Quels métiers sont les plus touchés par la crise sanitaire ? Accessed: March 26, 2025.

INSEE (2022). Zones d'emploi 2020 – contours géographiques. Accessed March 24, 2025.

INSEE (2023). Jobseeker - definition. `https://www.insee.fr/en/metadonnees/definition/c2010`. Accessed: 2024-03-25.

OpenStreetMap contributors (2022). Découpage administratif communal français issu d'openstreetmap. Accessed March 24, 2025.