# Methods: Iron and steel labor modeling

The following describes the parameterization and formulation of the labor module.

## Employment and activity data

The following describes historical data on employment and activity (e.g., production, capacity) for the iron and steel industry used to derive employment factors.

The U.S. Bureau of Labor Statistics (BLS) reports annual, county- and state-level employment1. Specifically, we collect data for the year 2022 for iron and steel mills and ferroalloy manufacturing, which is categorized under the North American Industry Classification System (NAICS) code 3311. While the data allow for regional analysis of employment trends, some county-level data are not reported to preserve confidentiality, which has implications for deriving employment factors.

The Mineral Commodity Summaries of the United States Geological Survey (USGS) report annual, national-level employment in the iron and steel industry2. Particularly, we collect data from 1998 to 2022, including the value of produced raw steel [in real reporting year United States dollars (USD)] the number of companies and mills involved in production, production capacity [million metric tons (MMT)], production volumes [million metric tons (MMT)] by technologies including Basic Oxygen Furnace (BOF) and Electric Arc Furnace (EAF), and annual average of monthly employment for the iron and steel industry. The USGS collects data from various sources, such as production and shipment data from the American Iron and Steel Institute and employment data from the BLS.

The Global Energy Monitor (GEM) reports annual, facility-level employment3. We collect data for 2022, including plant location, operating status, workforce size, ownership, process technology, and steel production capacity [thousand tonnes per annum (TTPA)]. The GEM dataset provides a more granular view of the industry, and includes data for plants with capacities exceeding 0.5 MMT. GEM data are based on publicly available sources like government reports, company filings, news articles, and regulatory filings. However, these data required substantial cleaning, cross-referencing, and overall quality control, given concerns regarding data sources, collection methods, reporting accuracy, timeliness, and regional differences (compared to other data sources).

The Facility Level Information on Greenhouse Gases Tool (FLIGHT) from the U.S. Environmental Protection Agency (EPA) supplements plant identification by providing additional information on emissions, fuel types, and locations of iron and steel mill plants.

Each dataset offers unique insights, with the USGS, BLS, and GEM datasets reported at the national-, regional-, and plant-levels, respectively. By combining these datasets, we better triangulate employment in the U.S. iron and steel industry, and in part address some of the limitations of each dataset (e.g., missing and reported data). To ensure the accuracy and consistency of our data, we compare the employment and activity data across datasets. We find that national steel capacity based on the GEM dataset is approximately 7% higher than those from USGS, despite GEM excluding smaller steel mills.

Figure 1 provides a comparison of national-level employment datasets for the iron and steel industry over time. Employment estimates from USGS are generally higher than those from BLS, with more pronounced differences between datasets prior to 2007. The USGS also refers to employment data from the BLS in their report, suggesting that the discrepancy may arise from the timing of data retrieval, as the data is regularly updated. Furthermore, employment estimates in 2022 from GEM dataset are lower than those from USGS and BLS by approximately 28%; this is because GEM only reports employment related to facilities, primarily focusing on larger facilities and excludes 27 smaller EAF facilities.

A graph of a line

Description automatically generated with medium confidence

Figure 1. Comparison of annual employment from 1998 to 2022 reported by USGS and BLS. In panel a), each point represents reported employment data, and the blue line and shaded region are the mean and standard deviation based on simple linear regression. The dotted line is the identity line. In panel b), the red and blue lines represent BLS and USGS data, respectively.

Figure 2provides a comparison of state-level employment datasets for the iron and steel industry for 2022. There are notable differences in state-level employment between the BLS and GEM datasets. For example, the GEM dataset reports that there is no employment in California, where there are no steel mills, whereas BLS estimates a substantial amount of employment in the state. NAICS code 3311 in BLS includes establishments primarily engaged in manufacturing iron and steel products from iron ore, pig iron, and scrap. This encompasses a variety of activities such as marketing, business operations, and recycling steel, which can occur in facilities other than traditional iron and steel mills. As a result, employment in this sector can still be reported in California despite the absence of steel mills. In Nebraska, where a steel facility is located, the GEM dataset reports employment, whereas as the BLS reports none.

A map of the united states

Description automatically generated

Figure 2. Comparison of 2022 employment reported by GEM (a) and BLS (b). In panel c), each point represents reported state-level employment, and the blue line and shaded region are the mean and standard deviation based on simple linear regression. The dotted line is the identity line.

## Employment factors

We derive employment factors, in units of jobs per unit of activity, using simple linear regression. We specify the regression models based on state-level data for 2022 and national-level data for 1988 to 2022.

Table 1 provides summary statistics for the state- and national-level datasets. As illustrated in Figure 3, steel production capacity and employment are positively correlated based on state- and national-level data. Steel production capacity and employment are declining over time. In addition, labor productivity (in units of jobs per MMT of capacity) is increasing over time, as BF-BOF plants have retired and with increased automation.

Table 1. Summary statistics for national-level (USGS and BLS) and state-level (BLS) datasets.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | National | | | State | | |
| Mean | Std. Dev | Obs. | Mean | Std. Dev | Obs. |
| Employment | 99,116.7 | 18,305.3 | 24 | 2,565.1 | 3,686.5 | 32 |
| Steel Capacity (MMT) | 114.7 | 6.1 | 25 | 4.1 | 6.0 | 30 |
| BOF (MMT) | - | - | - | 7.4 | 7.3 | 5 |
| EAF (MMT) | - | - | - | 2.9 | 3.0 | 30 |

A group of graphs showing different values

Description automatically generated with medium confidence

Figure 3. Relationship between employment and capacity. For panels a) – c), each point represents annual, national estimates from 1998 to 2022, and for panel d), each point represents state estimates for 2022. The blue lines and shaded areas are based on simple linear regression.

For the regression analysis based on national-level data, we specify models based on employment alternatively regressed on steel capacity, iron and steel production, and the proportion of steel production using BOF and EAF technologies. For the regression analysis based on state-level data, we specify models based on employment regressed on steel capacities by technology.

Results from the regression analysis are provide in Table 2. The estimated employment factors are 611 and 865 jobs per MMT of steel production based on the state- and national-level datasets, respectively; the national-level employment factor is notably higher than the state-level employment factor, reflecting increasing labor productivity over time. Based on the state-level regression, we further show differences in employment factors by technology; the employment factors for BOF is slightly higher than EAF.

The employment factors based on the state-level data likely best reflect the current industry. These estimates provide valuable insights into the current state of employment dynamics within the iron and steel industry, highlighting the need for continued monitoring and analysis to understand the evolving landscape.

Table 2 Regression model results of coefficients for national-level (USGS and BLS) and state-level (BLS) datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | National | | State | |
| Steel Capacity | 865.5\*\*\*  (25.5) | 2,537.2\*\*\*  (371.3) | 611.1\*\*\*  (25.8) | - |
| BOF | - | - | - | 619.6\*\*\*  (56.5) |
| EAF | - | - | - | 602.9\*\*\*  (54.9) |
| Constant | - | -192,874.0\*\*\*  (42,781.2) | - | - |
| Observations | 24 | 24 | 50 | 50 |
| Adjusted R2 | 1.0 | 0.7 | 0.9 | 0.9 |

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# References

1. BLS. Quarterly Census of Employment and Wages. *U.S. BUREAU OF LABOR STATISTICS* https://data.bls.gov/cew/apps/data\_views/data\_views.htm#tab=Tables (2022).

2. USGS. *Mineral Commodity Summaries*. (2024).

3. GEM. Tracker Map. *Global Energy Monitor* https://globalenergymonitor.org/projects/global-steel-plant-tracker/tracker-map/ (2024).