Firm-level Database: Variable Overview

Roberto Jara

Survey of R&D Expenditure and Personnel

This dataset provides detailed information at the **firm level** regarding research and development (R&D) activities. Firms are further classified according to **size** and **economic activity**, allowing for disaggregated analyses across different sectors and firm scales.

Structure and Coverage

A notable feature of this dataset is its **panel dimension**, which tracks firms from 2009 to 2023. This longitudinal structure facilitates the study of R&D evolution over time, as well as potential links with firm performance and internationalization trends.

The dataset includes a wide range of variables related to R&D, including personnel, expenditures, and other innovation indicators. In particular, the variable **B2001** reports firms' export values in thousands of pesos, providing a measure of international engagement at the firm level.

Limitations

While the dataset offers extensive information on R&D activities, it presents certain limitations:

- **Lack of product-level export information:** There is no breakdown of the specific products that are exported by each firm.
- **No export destination data:** The countries or regions to which the exports are sent are not recorded.

Notes

- The combination of firm-level panel data and rich R&D variables makes this dataset a valuable resource for studying innovation dynamics across sectors and over time.
- Linking this dataset with trade registries or customs data remains a challenge, as it would allow richer information on exported products and destinations.

National Customs Service: Export Data

The **National Customs Service (Servicio Nacional de Aduanas)** provides a complementary dataset that records **export operations by firms at the product level**, classified according to the **Harmonized System (HS)** of goods. This database contains detailed information on the **type of exported products**, **export destinations**, and **trade flows** aggregated by customs offices and operations.

Structure and Coverage

The **Customs dataset** is structured at the **firm level**, where each record corresponds to an exporting firm identified by its **export operations** over time. However, the available information focuses almost exclusively on **transactional data**, with limited details about firm characteristics beyond their export activity.

Each observation includes variables such as:

- **FECHAACEPT:** acceptance date of the export declaration.
- **PAISDESTINO:** destination country of the exported product.
- **CODIGOARANCEL:** product classification according to the international Harmonized System (HS).
- **FOBUS:** export item value (in USD).

Although firms can be identified within this dataset, **the available indicators are only valid internally**, meaning that they cannot be directly merged with external firm-level sources such as the R&D survey. Moreover, **firm-specific information is scarce**, restricting the analysis mainly to export behavior and patterns rather than broader firm characteristics.

This dataset provides **broad coverage across sectors and destinations**, offering a comprehensive picture of Chile's export activity over time.

Potential Integration with R&D Data

Although the customs data do not include explicit firm-level identifiers compatible with the R&D survey, an **indirect linkage** can be established through **industry codes**. Specifically, by applying a **correspondence between the Harmonized System (HS)** —which classifies products— **and the ISIC/CIIU Rev.4** —which classifies economic activities—, it is possible to map exported products to their corresponding industries.

It is important to note that **this correspondence relies on a strong assumption**: that the exported products are representative of the firm's **main economic activity**. In practice, a firm may belong to a particular industry (for example, food manufacturing) while also exporting products from different sectors (such as packaging materials or processed inputs).

By applying this mapping, we therefore **associate each export with the primary industry of the firm**, under the assumption that export activity reflects its main line of production. This methodological choice has **relevant implications**, as it affects how we interpret sectoral export intensity and innovation linkages — particularly in cases where firms diversify their product portfolio beyond their principal activity.

This mapping enables analyses such as:

- Estimating **export intensity by industry** and comparing it with firms' R&D investment patterns.
- Studying **sectoral specialization in innovation and exports**, identifying industries with high R&D activity and high export participation.
- Exploring **heterogeneity across destinations**, such as whether more technologically intensive sectors export disproportionately to developed economies.

Complementarity with the R&D Survey

The integration of the **Customs dataset** and the **R&D survey** allows for a richer understanding of the relationship between **innovation and internationalization**. While the R&D survey captures firm-level efforts and resources devoted to innovation, the customs data provide an external validation of market outcomes — namely, **export performance** and **product specialization**.

In this sense, the two datasets are **complementary**:

Aspect	R&D Survey	Customs Data
Unit of observation	Firm	Export transaction
Temporal coverage	2009–2023	Continuous (annual)

Aspect	R&D Survey	Customs Data
Focus	R&D personnel, expenditure, innovation	Products, destinations, trade value
Classification system	ISIC Rev.4 (CIIU4.CL 2012)	Harmonized System (HS)
Linkage potential	Via HS–CIIU4 correspondence	Via HS–CIIU4 correspondence

Aggregation and Sector-Level Indicators

Given that the customs data are disaggregated by **product and destination**, whereas the R&D survey operates at the **firm level**, the integration requires shifting the analytical focus toward the **sectoral level**.
This approach involves constructing **aggregated indicators of innovation and productivity** by sector and year, derived from the microdata of the R&D survey.

This alignment allows for the creation of a **sector–year panel** combining:

- **Export outcomes:** total export value, product diversification, destination markets.
- **Innovation outcomes:** average R&D intensity, share of R&D-performing firms, productivity metrics.

Such an integrated database enables the analysis of how **sectoral R&D investment relates to export performance**, product composition, and destination patterns.