Estimating an Interregional Input-Output Table for Brazil Using NF-e Data: A Data-Driven Approach to Regional Economic Analysis

Joao Maria de Oliveira - IPEA

Patieene Alves Passoni – UFAL/IPEA

Abstract

This study estimates an Interregional Input-Output Table (IIOT) for Brazil for the year 2018 using a data-driven methodology based on the Brazilian Electronic Invoice System (Nota Fiscal Eletrônica - NF-e). By leveraging detailed administrative records of business-to-business transactions, this research overcomes historical limitations related to outdated or aggregated data traditionally used in regional economic modeling. The resulting IIOT covers all 27 Brazilian states and 68 economic sectors, capturing interregional trade flows with greater precision. This study contributes to both the methodological advancement of regional input-output analysis and to policy-relevant assessments of regional economic interdependencies.

Key words: Input-Output Analysis, Electronic Invoices, Brazil, Regional Analysis, Big Data.

Estimating an Interregional Input-Output Table for Brazil Using NF-e Data: A Data-Driven Approach to Regional Economic Analysis

1. Introduction

Input-Output (IO) Models are essential tools in economic analysis, supporting policymaking, forecasting, and the study of intersectoral and interregional linkages. However, national input-output tables treat the entire country as a single homogeneous region and therefore fail to capture the economic interactions among states, especially those related to trade flows embedded in the productive structure and final demand between regions. In the Brazilian case, regional analysis becomes particularly important. Given the country's pronounced economic disparities and the diverse productive structures across its states, identifying how inter-state economic linkages are established—and how growth in one state can impact others—is key to designing more precise impact assessments for development policies. This is especially true when considering strategies aimed at reducing regional inequalities.

In this context, estimating an interregional input-output system is essential for providing accurate information on regional economies and their mutual interactions—offering a more realistic depiction of economic structures. Nevertheless, since the early contributions of Isard (1951) and Leontief (1986), the estimation of interregional input-output tables has faced major challenges. Chief among them are the scarcity of detailed regional data and the high cost of collecting census-based information, particularly concerning interregional trade flows. These constraints have led to the development of alternative methods that approximate real trade flows by combining official survey data with estimates derived from non-census-based techniques.

As noted by Miller and Blair (2009), the variety of available estimation techniques and the ways in which they can be combined mean that multiple interregional input-output systems can be constructed for the same set of regions. Among the most prominent methods are:
(i) the Interregional Input-Output Adjustment System (IIOAS), proposed by Haddad, Júnior, and Nascimento (2017); and (ii) the Supply and Use Interregional Tables (SUIT) method, developed by Guilhoto et al. (forthcoming).

The construction of interregional input-output tables (IIOTs) has traditionally been constrained by the lack of disaggregated and reliable data. However, recent advances in administrative data availability—such as fiscal records and electronic transaction databases—have opened new opportunities to improve the granularity and precision of these estimations. This paper explores the use of data from the Brazilian Electronic Invoice System (Nota Fiscal Eletrônica – NF-e) to estimate a comprehensive IIOT for Brazil at the state level, thereby strengthening the empirical foundation for regional economic research.

The IIOT used in this study was estimated through an adaptation of the SUIT method. By incorporating NF-e records from the year 2018, the estimation differs from traditional approaches, as it enables the direct capture of inter-state transactions. The inclusion of NF-e data allows for a much more detailed mapping of purchases and sales of goods and services between Brazilian states, enhancing the analysis of regional economic interactions. As such, the method provides key information on interregional trade linkages and supports more accurate spatial economic modeling.

2. Literature Review

Although input-output models were originally designed for national applications, they have been applied to subnational geographical units since the second half of the 20th century. According to Miller and Blair (2009), there are two specific regional characteristics that make the distinction between national and regional input-output models both evident and necessary. First, each region's production technology is distinct and may differ substantially from the national input-output structure. Factors such as the age of regional industries, the characteristics of input markets, and the educational level of the workforce can influence regional technologies and lead to significant deviations from national averages. Second, smaller regional economies tend to be more open and dependent on external markets, which increases the importance of accounting for both imported and exported components of supply and demand. Notably, such flows involve not only international trade but also interregional trade within the country.

According to Isard et al. (1998), the main advantage of interregional input-output (IRIO) models is their ability to capture **spillover effects**, whereby a demand shock in one region influences other regions and feeds back into the original region through interregional economic linkages.

The core challenge in a multi-regional input-output system lies in estimating **interregional trade flows** (Isard et al., 1998). Interregional trade matrices are rarely available at the required level of disaggregation. As a result, the estimation of intermediate trade flows across regions and sectors has become a central obstacle, especially given that the data requirements grow exponentially with the number of regions (Miller and Blair, 2009). For example, a system with three regions requires six interregional tables, whereas one with four regions requires twelve. In the Brazilian case—with 27 states—there are 702 potential interregional trade matrices.

According to the typology proposed by Round (1983), when an interregional system is built entirely from direct census-type information, it is referred to as a **census-based method**. Conversely, when the system is constructed using estimates of the required direct information, it is classified as a **non-census method**. However, Round (1983) also points out that, in practice, these two approaches are not mutually exclusive; most interregional input-output tables are developed through **hybrid methods**, combining different techniques depending on the availability and quality of primary data.

Among the most widely used non-census techniques for IRIO estimation are: (i) **Location Quotient (LQ)** methods and their variations; (ii) **Gravity models**; and

(iii) **Iterative procedures**, such as the RAS method and its extensions. According to Montoya (1999), however, in practice it is common to combine these approaches—and possibly others—when developing an interregional input-output system.

Several studies describe hybrid methods that combine non-census techniques for estimating IRIO tables (Roy and Thill, 2004; Sargento, 2009; Riddington, Gibson, and Anderson, 2006). The latter also offer a comparative analysis of the gravity model and the traditional LQ approach. They argue that using national tables or adjusted national tables to reflect regional specialization entails significant limitations. This motivated the development of gravity-based methods to estimate regional input-output and trade tables.

More recently, Boero, Edwards, and Rivera (2018) proposed a method that simultaneously estimates both regional input-output tables and interregional trade flows, reflecting a growing trend in the literature. Similarly, the SUIT method (Supply and Use Input-Output Tables), developed by Guilhoto et al. (forthcoming), does not directly use trade matrices derived from origin–destination impedance models, although it does incorporate them in the final calibration phase. A key feature of SUIT is that it combines **bottom-up** and **top-down** approaches. As a result, the method ensures consistency with national Use, Make, and Supply-Use tables as well as with the Regional Accounts, while preserving the unique economic structure of each region.

In the Brazilian context, there are several studies estimating interregional systems. However, initiatives that produce a **complete interregional input-output table**—as does the present study—are relatively scarce and more recent. Among the notable examples are Haddad, Júnior, and Nascimento (2017), who applied the IIOAS method to construct an IRIO system for all 27 Brazilian states, with 68 sectors and 128 products; and Guilhoto et al. (forthcoming), whose SUIT method is adapted in this study through the use of administrative records from the **Brazilian Electronic Invoice System (NF-e)**, in what the authors refer to as the **SUITnf** approach.

3. Methodology

This study introduces a pioneering approach by integrating administrative microdata from Brazil's Electronic Invoice System (Nota Fiscal Eletrônica – NF-e) with the conceptual framework of the System of National Accounts (SNA) and the Regional Accounts System (RAS), both maintained by the Brazilian Institute of Geography and Statistics (IBGE). The estimation procedure transforms the National Supply and Use Tables (SUT_n) into their interregional version (SUT_i^r), capturing the flows of goods and services across Brazil's 27 states and 68 economic sectors.

Implemented in 2006 by the Brazilian Federal Revenue Service (Receita Federal do Brasil – RFB), the NF-e system replaced printed invoices and became mandatory for all commercial transactions involving industrialized goods (IPI), wholesale and distribution operations, interstate or foreign trade, and transactions with public administration entities.

In 2018, access to the NF-e database was granted to researchers through a cooperation agreement between the RFB and the Institute of Applied Economic Research (IPEA), enabling unprecedented analytical potential.

The NF-e system records a vast volume of business-to-business transactions with exceptional sectoral and geographic granularity. Each invoice contains detailed information about buyers and sellers (including sector classification, legal nature, and state location), transaction characteristics (such as scope, type, and tax profile), and product-level data including quantity, unit, price, freight costs, discounts, and tax components. Products are coded using the Mercosur Common Nomenclature (NCM), which is fully compatible with the Harmonized System (HS).

The availability of the complete 2018 NF-e dataset allowed for the construction of a detailed Interregional Input-Output Table (IOTir) for Brazil. Although the data are anonymized, they preserve key transaction attributes necessary to estimate capital flows and analyze the supply and demand structure for investment goods across the country. To ensure alignment with the SNA, the study adopts the most disaggregated level of the Brazilian national accounting system: 68 activities and 128 products. The Broad Economic Categories (BEC), version 5.0, was applied to classify goods according to their economic function and to identify those corresponding to gross fixed capital formation (GFCF).

Despite its standardization, NF-e data required extensive preprocessing to harmonize product and sector classifications and ensure compatibility with national and regional accounts. This methodological effort resulted in a coherent and granular IOTir, offering critical insights into interregional trade patterns and capital distribution in Brazil. The resulting table provides a robust empirical foundation for regional economic analysis and policy design aimed at fostering investment-led development.

3.1 Initial Adjustments and Generation of SUT_n and SUT_i^r

The first step involves generating national (SUT_n) and interregional (SUT_i^r) supply and use tables. This requires extensive processing and classification of transaction-level data extracted from the NF-e database, comprising the following stages:

- Legal classification of agents (public administration, corporations, or households);
- Conversion of the National Classification of Economic Activities (CNAE) into activity groups compatible with the SNA's 68-sector structure;
- Conversion of the Mercosur Common Nomenclature (NCM) into product codes aligned with the national accounts classification;
- Assignment of product use categories according to version 5.0 of the Broad Economic Categories (BEC) classification, distinguishing between intermediate

consumption, final consumption (by households or government), and gross fixed capital formation (GFCF);

- Calibration to national totals, ensuring consistency with IBGE official aggregates on resources, supply, uses, and final demand.

3.2 Transformation from SUT_n to SUT^s and SUT_i^r

Once the national tables are processed, they are disaggregated into two dimensions:

- SUT^s: tables organized by economic sector;
- SUT_i^r: interregional tables disaggregated by sector and federative unit (state).

These tables contain:

- A supply matrix, including domestic production, imports, trade margins, transport margins, and product taxes;
- A use matrix, comprising intermediate consumption and the main components of final demand;
- Total supply, computed as the sum of domestic production and adjusted imports.

The procedure adopted for estimating the Brazilian interregional input-output table aims to cover all **27 Brazilian states**, encompassing 68 activitives (n) and 128 commodities (m). Figure 1 presents the general schematic of the interregional model, including both the input-output tables used and those estimated. At the end of the process, the table resulting from the interregional model is contained within the area outlined by the red border.

Figure 1 – General schematic of the interregional input-output model.

Commodities	ST 1					U_{1_1}	U_{1_2}	...	U_{1_27}	E_{1_1}	E_{1_2}	...	E_{1_27}	Q_1
	ST 2					U_{2_1}	U_{2_2}	...	U_{2_27}	E_{2_1}	E_{2_2}	...	E_{2_27}	Q_2
	...					...	...	...	...	...	...	...	...	...
	ST 27					U_{27_1}	U_{27_2}	...	U_{27_27}	E_{27_1}	E_{27_2}	...	E_{27_27}	Q_{27}
Activities	ST 1	V'_{1_1}	V'_{1_2}	...	V'_{1_27}	Z_{1_1}	Z_{1_2}	...	Z_{1_27}	Y_{1_1}	Y_{1_2}	...	Y_{1_27}	X_1
	ST 2	V'_{2_1}	V'_{2_2}	...	V'_{2_27}	Z_{2_1}	Z_{2_2}	...	Z_{2_27}	Y_{2_1}	Y_{2_2}	...	Y_{2_27}	X_2
	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	ST 27	V'_{27_1}	V'_{27_2}	...	V'_{27_27}	Z_{27_1}	Z_{27_2}	...	Z_{27_27}	Y_{27_1}	Y_{27_2}	...	Y_{27_27}	X_{27}
Imports						M_{IC_1}	M_{IC_2}	...	M_{IC_27}	M_{DF_1}	M_{DF_2}	...	M_{FD_27}	M
Tazxes						T_{IC_1}	T_{IC_2}	...	T_{IC_2}	T_{FD_1}	T_{FD_2}	...	T_{FD_27}	T
Value added						W_1	W_2	...	W_{27}			...		W
Gross output		Q'_1	Q'_2	...	Q'_{27}	X'_1	X'_2	...	X'_{27}			...		PT

Source: Authors' elaboration.

- **U (m × n): Use table at purchaser's prices**
- **V (n × m): Make table**
- **Z (n × n): Intermediate consumption table**
- **E (m × 6): Final demand table at purchaser's prices**
- **Y (n × 6): Final demand table at basic prices**
- **X (n × 1): Total output vector**
- **M_{IC} (1 × n): Vector of intermediate consumption imports**
- **M_{FD} (1 × n): Vector of final demand imports**
- **T_{IC} (t × n): Tax matrix for intermediate consumption**
- **T_{FD} (t × n): Tax matrix for final demand**
- **W (w × n): Value added table**

Where:

- ***m* = number of commodities**
- ***n* = number of activities**

- ***t* = number of tax components**
- ***w* = number of value added components**

3.3 Final Demand Estimation and Regional Allocation

Final demand is disaggregated into six components, each estimated using specific sources and allocation methods. The interregional breakdown is guided by proportional shares derived from either NF-e data or complementary surveys, as shown in the table below:

Demand Component	Source
Exports	NF-e
Government Consumption	NF-e
Household Consumption	Household Budget Survey (POF)
Investment (GFCF)	NF-e + Investment Absorption Matrix
Inventory Variation	Residual estimate

Here, rdr_drd represents the regional distribution share, and mmm and nnn refer to the origin and destination of regional flows, respectively.

Gross fixed capital formation is broken down by institutional sector (government, corporations, households), with allocation matrices developed using transaction-level investment data and national absorption coefficients.

3.4 Regional Adjustment and Matrix Balancing

The estimated IIOT is subsequently reconciled with the Regional Accounts System (RAS) published by IBGE, using the following macroeconomic aggregates for each federative unit:

- Gross Output (GO);
- Value Added (VA);
- Intermediate Consumption (IC);
- Inventory Changes (as residuals).

To ensure consistency, the following procedures are applied:

- Regional reweighting using RAS values;
- RAS adjustment technique (Row and Column Scaling) adapted for interregional dimensions;
- Manual corrections for underrepresented sectors in NF-e data, especially public services such as education, health, and administration. These adjustments use

complementary sources such as the Annual Survey of Services (PAS), the Annual Industrial Survey (PIA), and the Annual Trade Survey (PAC).

3.5 Outcomes and Analytical Potential

The resulting matrix offers an unprecedented level of regional and sectoral disaggregation for Brazil:

- Coverage of all 27 states;
- 68 economic sectors, consistent with the national accounts;
- Detailed interregional flows of final demand components, including government, households, exports, investment, and inventory variation.

Unlike previous approaches that relied on approximations from national tables or synthetic surveys, this method leverages observed fiscal microdata from NF-e to capture actual interregional trade flows. This enhances the accuracy and empirical reliability of the IIOM, making it a valuable tool for regional economic modeling, policy simulations, and structural diagnostics.

The compatibility with the SNA and RAS ensures the matrix's integration with official statistics and its suitability for various analytical applications, including impact assessments, value chain studies, and spatially differentiated policy design.

3.6 From SUT to IIOT: Estimation Procedure

The transition from the Supply and Use Tables (SUT) to the Interregional Input-Output Table (IIOT) was carried out using the methodology proposed **by Guilhoto and Sesso Filho (2005; 2010).** This approach allows for the estimation of key components—such as the National Use Table at basic prices, the Imported Use Table, trade and transport margins, and a comprehensive set of taxes (including Import Duties, Tax on Industrialized Products – IPI, State VAT – ICMS, and total net taxes excluding subsidies)—starting from the Total Use Table at purchaser's prices, as published in the official SUTs by IBGE.

The method works by estimating the distribution of each product across intermediate consumption (by sectors) and final demand components. This is done through proportioning the total demand for each product based on its use across different sectors and final demand categories. The resulting shares are then applied to vectors extracted from the Resource Table—such as domestic supply, margins, taxes, and imports—thus enabling the disaggregation of the original purchaser's price data into its underlying components.

This procedure results in a consistent estimation of:

- The Use Table at basic prices
- The Imported Use Table

- Trade and Transport Margins
- Disaggregated tax tables (Import Duties, IPI, ICMS, and Total Net Taxes)

An important advantage of this method is that it allows for a complete and separate estimation of all tax components, unlike other approaches that aggregate taxes or apply adjustments post-estimation. Additionally, because product-level proportions are used to guide the allocation, no balancing adjustments are necessary—the total of the estimated components automatically aligns with the original SUT resource totals.

This step is fundamental in producing a coherent national table that can be further disaggregated spatially to construct the IIOT. The method ensures consistency with the national accounts while enabling a more detailed representation of taxes and margins than what is typically provided in official publications.

3.7 Construction of the Basic Input-Output Model for the 27 States

Following the construction of the national-level Supply and Use Tables (SUT), a basic Input-Output (I-O) model was developed for each of Brazil's 27 federative units. This model serves as the analytical foundation for capturing the structural relationships between activities, products, and final demand within and across states. The approach mirrors the procedures applied in traditional I-O modeling at the national level, adapting them to the regional dimension through detailed disaggregation.

The basic I-O model is built upon the data structure of the regionalized SUTs and their corresponding transformation matrices—commonly referred to as "tabelas de passagem" in the Brazilian System of National Accounts. These transformation tables enable the breakdown of transaction values from purchaser prices into basic prices, trade margins, transportation margins, and taxes on products. The model is constructed to align with the valuation principles established by the Brazilian Institute of Geography and Statistics (IBGE) and is consistent with the treatment of taxes, trade, and transport across the entire production and consumption system.

The database comprises:

- Local production flows by state and activity, at basic prices;
- Intermediate and final demand matrices by state, including imports from other states (Rest of Brazil) and from abroad (Rest of World);
- Trade and transport margins, and product-level tax components, disaggregated by source (local, domestic imports, international imports);
- Production vectors and value added by economic activity and federative unit.

For each state, the matrices of intermediate consumption (local, domestic, and international) were derived by dividing the value of each input by the total output of the consuming activity. This allowed the construction of technical coefficients that indicate the input requirements per unit of output, reflecting both local production structures and import dependencies.

The next step involved estimating a market-share matrix that reflects the share of each activity in the production of a given product. This was done under the "technology by activity" assumption, following the IBGE convention. The resulting matrix was then used to convert product-level demand (intermediate and final) into activity-level demand. This conversion enabled the construction of the intersectoral technical coefficients matrix, which captures the value of inputs required from one activity to another within each state.

From this, the Leontief inverse matrix was derived, enabling the computation of total (direct and indirect) production requirements in response to changes in final demand. This matrix—often referred to as the Leontief multiplier—provides a comprehensive view of the economic interdependencies across sectors and serves as the backbone for impact analysis and policy simulations.

By replicating this process across all 27 federative units, the study produces a harmonized set of basic I-O models, each tailored to the specific production and demand structure of the corresponding state. These models not only provide a basis for understanding local economic dynamics but also enable their integration into a unified interregional framework—allowing for the analysis of inter-state trade, spillover effects, and regional development strategies.

4. Preliminary Results and Ongoing Work

This study is currently in progress. The methodological framework, as described in previous sections, has been largely established, but refinements and calibrations are still underway. The estimation of Brazil's Interregional Input-Output Table (IIOT) using NF-e data is ongoing, and the generation of final results is being conducted in parallel with the validation and quality control of the processed data.

4.1 Interregional Trade Patterns (in progress)

Preliminary outputs suggest that the resulting IIOT will reveal detailed interregional trade flows, capturing the heterogeneity of regional production structures and trade dependencies. Spillover and feedback effects are expected to be explicitly modeled, allowing for a more nuanced understanding of economic interdependence among Brazilian states.

4.2 Comparison with Previous Estimates (forthcoming)

While a full comparative analysis is still being developed, early indications point to significant improvements in resolution and accuracy over prior estimations based on national-level data. The use of NF-e is expected to unveil regional variations in input structures and trade patterns that are typically masked in aggregated models.

4.3 Macro-Regional Aggregation (under development)

A macro-regional version of the IIOT is also under construction. This will allow for an aggregated view of interregional dynamics at the level of Brazil's five major regions, supporting broader policy assessments related to infrastructure, regional development, and fiscal planning.

5. Contributions and Expected Policy Implications

Despite being in progress, the research already illustrates the potential of administrative microdata for advancing regional economic modeling in Brazil. The final IIOT is expected to provide:

- A robust empirical base for subnational impact analysis;
- Inputs for simulations of public investment, fiscal policy, and regional planning;
- A foundation for further studies on regional inequality, structural change, and productive transformation.

6. Conclusion and Next Steps

This work-in-progress represents the first attempt to estimate a full IIOT for Brazil based entirely on NF-e microdata. The research aims to enhance our understanding of regional structures and interlinkages with a high degree of granularity. Next steps include:

- Finalizing the estimation and validation of the IIOT;
- Expanding the methodology for time series analysis;
- Incorporating household consumption, labor income, and investment data for broader modeling applications.

The final results will support evidence-based policymaking and enable more precise evaluations of the spatial effects of economic policies across Brazilian states.

References

Boero, R.; Edwards, B. K.; Rivera, M. K. (2018). *Regional input–output tables and trade flows: an integrated and interregional non-survey approach*. *Regional Studies*, 52(2), 225–238.

Guilhoto, J. J. M. et al. (no prelo). *Sistema interestadual de insumo-produto do Brasil, estimação em condições de informação limitada para o ano de 2011: uma aplicação do método SUIT*. *Economia Aplicada*.

Guilhoto, J. J. M., & Sesso Filho, U. A. (2005). Estimação da matriz insumo-produto a partir de dados preliminares das contas nacionais. *Economia Aplicada*, *9*(2), 277–299.

Guilhoto, J. J. M., & Sesso Filho, U. A. (2010). Estimação da matriz insumo-produto utilizando dados preliminares das contas nacionais: Aplicação e análise de indicadores econômicos para o Brasil em 2005. *Economia & Tecnologia*, *6*(23), outubro/dezembro.

Haddad, E. A.; Júnior, C. A. G.; Nascimento, T. O. (2017). *Matriz Interestadual De Insumo-Produto Para o Brasil: Uma Aplicação do Método IIOAS*. *Revista Brasileira de Estudos Regionais e Urbanos*, 11(4), 424–446.

Isard, W. (1951). *Interregional and regional input-output analysis: a model of a space-economy*. *The Review of Economics and Statistics*, 318–328.

Isard, W. (1998). *Methods of interregional and regional analysis*. Ashgate Publishing.

Miller, R. E.; Blair, P. D. (2009). *Input-Output Analysis: Foundations and Extensions* (2nd ed.). Cambridge University Press.

Montoya, M. A. (1999). *Análise insumo-produto internacional no Mercosul: desenvolvimento econômico regional e interdependência estrutural*. Universidade de Passo Fundo.

Riddington, G.; Gibson, H.; Anderson, J. (2006). *Comparison of gravity model, survey and location quotient-based local area tables and multipliers*. *Regional Studies*, 40(9), 1069–1081.

Round, J. I. (1983). *Nonsurvey techniques: a critical review of the theory and the evidence*. *International Regional Science Review*, 8(3), 189–212.

Roy, J. R.; Thill, J.-C. (2004). *Spatial interaction modelling*. *Papers in Regional Science*, 83(1), 339–361.

Sargento, A. L. M. (2009). *Regional input-output tables and models: interregional trade estimation and input-output modelling based on total use rectangular tables*. Universidade de Coimbra.