Do more open economies lead to export concentration? The case of selected agribusiness goods in Brazilian municipalities

Alan Leal[1]
Michelle Marcia Viana Martins[2]

Área ENABER: 3- Localização e Concentração das Atividades Econômicas

Abstract: This paper aims to characterize an important question: does a more export-open economy for a municipality imply a larger export concentration in terms of goods exported? Recent developments in export imputation for Brazilian municipalities allow for an analysis of the concentration profile of each municipality and its relationship with export openness. Our research makes three major contributions: we develop a novel Export-Intensive Gini Index using Monte Carlo simulations that isolates commodity-specific concentration patterns while controlling for the composition of remaining export portfolios; we apply double machine learning methodology to establish causal relationships while controlling for hundreds of municipal characteristics including geographic, infrastructural, demographic, and institutional factors; we provide the first comprehensive analysis of subnational export concentration patterns in a major developing economy using production-based rather than customs-based trade data. Our results reveal heterogeneity across commodities: coffee exhibits the strongest positive relationship between export openness and concentration, followed by corn, soybeans and sugar, while beef demonstrates negative coefficient. These findings indicate that trade integration affects agricultural specialization differently across commodities, with implications for regional economic vulnerability and policy design. For policymakers, our results suggest that coffee and corn-producing municipalities face increasing exposure to price volatility as they integrate into global markets, requiring targeted risk mitigation strategies. Conversely, the weak relationships for soybeans and beef suggest these sectors may offer more stable foundations for export-oriented development.
Key-words: exports; Gini index; machine learning.

Resumo: Este artigo visa caracterizar uma questão importante: uma economia mais aberta à exportação para um município implica uma maior concentração de exportações em termos de bens exportados? Desenvolvimentos recentes na imputação de exportações para municípios brasileiros permitem uma análise do perfil de concentração de cada município e sua relação com a abertura à exportação. Nossa pesquisa traz três contribuições principais: desenvolvemos um novo Índice de Gini Intensivo à Exportação usando simulações de Monte Carlo que isola padrões de concentração específicos de commodities enquanto controla a composição dos portfólios de exportação restantes; aplicamos a metodologia de aprendizado de máquina duplo para estabelecer relações causais enquanto controlamos centenas de características municipais, incluindo fatores geográficos, infraestruturais, demográficos e institucionais; fornecemos a primeira análise abrangente dos padrões de concentração de exportações subnacionais em uma grande economia em desenvolvimento usando dados comerciais baseados na produção em vez de dados alfandegários. Nossos resultados revelam heterogeneidade entre commodities: o café exibe a relação positiva mais forte entre abertura à exportação e concentração, seguido por milho, soja e açúcar, enquanto a carne bovina demonstra coeficiente negativo. Essas descobertas indicam que a integração comercial afeta a especialização agrícola de forma diferente entre as commodities, com implicações para a vulnerabilidade econômica regional e a formulação de políticas. Para os formuladores de políticas, nossos resultados sugerem que os

[1] Ph.D. candidate in Economics at University of São Paulo. E-mail: leal@thetalab.com.br.
[2] Professor of Economics at Federal University of Viçosa. E-mail: michelle.viana@ufv.br.

municípios produtores de café e milho enfrentam uma exposição crescente à volatilidade dos preços à medida que se integram aos mercados globais, exigindo estratégias direcionadas de mitigação de riscos. Por outro lado, as fracas relações entre soja e carne bovina sugerem que esses setores podem oferecer bases mais estáveis para um desenvolvimento voltado para a exportação.
Palavras-Chaves: exportações; índice de Gini; machine learning.
JEL Codes: F14; O13; C55

1 Introduction

The relationship between trade openness and export specialization represents a relevant question in international economics, with particular relevance for understanding how local economies respond to global market integration. While the classical international trade theory posits that deeper exposure to world markets steers economies toward sectors in which they have comparative advantages (Ricardo, 1817; Ohlin, 1933), new-trade and economic-geography models go further: with increasing returns and home-market effects, falling trade costs can concentrate activity in a subset of goods, raising export specialization (Krugman, 1980; Krugman, 1991; Wuri, 2024).

Although recent work has investigated this mechanism for European regions (Ricordel, 2024; Pardy & Rodríguez-Pose, 2025), the spatial economics of export concentration within developing countries remains understudied. Brazil offers a compelling case: since the turn of the century, its external trade flows have shifted toward primary commodities, while the share of manufactured goods has waned (Nassif et al., 2020; Oreiro and Feijó, 2010). While the country consolidates its position as a powerhouse in global agriculture (USDA, 2025), an important question arises: have Brazilian municipalities, increasingly integrated into world markets, concentrated their cross-border shipment on a narrow set of agricultural products instead of diversifying across a broader mix?

For many years, that question was unanswerable because Brazilian trade statistics attribute traded volumes to the municipality where customs clearance occurs rather than to the place of production, masking the geography of value creation. Leal and Martins (2025) addressed this limitation by reallocating export data at the six-digit level of the Harmonized System (HS-6)[3] from bureaucratic hubs to actual production sites. This was achieved using a spatial downscaling algorithm informed by micro-level data from municipalities. Their dataset reveals, for the first time, the true export footprint of each of Brazil's 5,570 municipalities.

Building on that advance, the present study delivers three innovations. First, it develops an Export-Intensive Gini Index that quantifies, commodity, how export revenue is distributed across municipalities. The analysis includes both traditional export goods, such as coffee and sugar, and more recent export expansions, like soybeans, corn, and beef. All those products that define Brazil's comparative advantage yet differ markedly in agronomic needs and commercial networks. This multi-commodity approach enables an understanding of how the concentration varies across agricultural sectors with different production characteristics, market conditions, and historical backgrounds. Second, it employs the export-to-revenue ratio as a measure of municipal openness, sidestepping the absence of reliable import data at that scale.[4] Third, it applies the double machine learning framework of Chernozhukov et al. (2018) to estimate the

[3] Refers to an international nomenclature used to classify traded products. This level of disaggregation allows for more detailed identification of goods than broader classifications, such as HS-2 or HS-4, and is commonly used in trade data analysis to capture product-level patterns.

[4] A conventional trade openness index, defined as the sum of imports and exports divided by municipal revenue, is not applied; instead, openness is captured by the ratio of exports to municipal revenue. To date, no robust methodology exists to estimate municipal-level imports, and Comexstat data suffer from the limitations identified by Leal and Martins (2025).

causal effect of commodity-specific concentration while controlling for a high-dimensional set of geographic, infrastructural, demographic and institutional attributes.

The Export-Intensive Gini Index is generated through a Monte Carlo routine that simulates alternative allocations of the residual export basket while holding the observed value of the target commodity fixed; the average Lorenz-curve area across simulations yields a bounded index between zero and one. This procedure isolates each commodity's spatial concentration without conflating it with the diversity of the remaining export bundle. Cross-fitting within the double machine learning estimator ensures that nuisance functions are learned on data segments excluded from the final regression, preserving asymptotic properties even amid hundreds of covariates.

The results carry tangible policy relevance. If municipalities with high concentrations are also those whose exports are tightly channeled into a single commodity, regional exposure to price volatility may intensify, calling for risk-mitigation strategies. Conversely, evidence of dispersion would suggest that trade integration can broaden local income sources and temper vulnerability.

Studies examining regional aspects and export concentration at subnational levels provide limited but growing evidence on municipal patterns. Pardy and Rodríguez-Pose (2025) analyze trade ties and income inequality across European regions, demonstrating how regional trade patterns affect economic outcomes. Ricordel (2024) examines domestic versus export orientation at the regional level, also in the EU, applying spatial endogenous growth models to understand regional specialization patterns. These studies establish that subnational trade patterns differ from national aggregates and require distinct analytical approaches.

Riaño (2024) reproduces stylized facts motivating heterogeneous models, showing that export participation correlates with higher productivity, larger size, and greater innovation capabilities. Giroud et al. (2024) show that productivity improvements in one location affect distant areas through multi-region firm networks. Bilgin, Fala and Ottaviano (2024) demonstrate technology diffusion along supply chains, where upstream robotization improves downstream productivity even without direct input-output linkages.

Ranran and Jingsuo (2024) examine how agricultural production agglomeration strengthens economic resilience through spatial spillover effects in China. Their findings indicate that agricultural clusters generate benefits extending beyond immediate participants to influence broader regional development patterns. Phillips (2024) examines subnational import exposure, highlighting data limitations that constrain analysis of local trade patterns. Transportation costs, storage facilities, and port access determine which municipalities can effectively participate in export markets and may moderate the export concentration by affecting the costs of market participation for different commodities.

Methodological advances, such as the contribution of Leal and Martins (2025), enable analysis of Brazilian municipal trade patterns. Our study confronts two interrelated research questions: To what extent does a higher export-to-revenue ratio lead a municipality to channel its foreign sales into a few agricultural commodities rather than spread them across a broader set of products? Does the strength of that association vary across individual commodities, such as soybeans, beef, coffee, sugar, and corn, whose agronomic demands, supply networks, and market structures differ markedly? Answering these points sheds light on whether deeper global integration narrowed or widened the productive base of Brazilian rural economies and under what circumstances diversification could be fostered. The hypothesis that a municipality more engaged with exporting selected primary goods tends to have a more concentrated export profile.

The paper proceeds as follows. Beyond this introduction, a section detailing the theoretical framework about export orientation and concentration at the subnational level. A third section introduces the methods employed in this paper, displaying the flexibility of double machine learning and why it is a good method for the task at hand. The results section displays descriptive statistics and the main results of this research, emphasizing our findings. A final concluding remarks details the implications of our findings and suggests new directions for future research.

2 Conceptual foundations: trade openness, commodity concentration, and the municipal scale

A growing body of literature recognizes that the forces shaping export structures operate with marked heterogeneity at finer territorial scales. Building on firm-heterogeneity models (Melitz, 2003) and new economic geography, recent studies observe that the thresholds for export participation, the costs of market access, and the distribution of productivity advantages vary between countries and also across municipalities within a single economy. This subnational lens is particularly relevant for Brazil, where municipalities differ widely in factor endowments, infrastructure, and institutional capacity.

Municipal trade specialization can be interpreted as the joint outcome of comparative advantage and spatial frictions. Fally and Hillberry (2018) show that even modest variations in transport costs reconfigure local export patterns, while Bustos, Garber and Ponticelli (2022) document how access to better logistics nodes shifts firms' export entry decisions within states. These findings resonate with the Brazilian context, in which inland municipalities face longer distances to ports and greater reliance on road transport, raising the fixed and variable costs of serving foreign buyers. Where such frictions are lower, due, for example, to proximity to river terminals or inter-modal corridors, municipalities more readily exploit their comparative advantages, often in crops suited to local agro-climatic conditions.

Agglomeration mechanisms magnify these initial advantages. Henderson, Kriticos and Venables (2021) demonstrate that specialized clusters emerge when scale economies in processing, storage, and certification reduce unit costs for proximate producers. In Brazilian agriculture, soybean complexes in Mato Grosso and coffee networks in Minas Gerais illustrate how local supply-chain density reinforces commodity orientation, concentrating both primary output and related services within a limited radius. These clusters foster knowledge spillovers, encouraging incremental innovation in seed varieties, planting techniques, and logistics management that sustain competitiveness.

Path-dependence further anchors specialization. Historical investments in transport infrastructure, warehouse capacity, and technical training tilt later choices toward the prevailing commodity mix. Muendler, Rocha and Ramey (2020) show that municipalities with early export booms maintain export intensity decades later, even after controlling for contemporary factor prices. Such persistence implies that trade liberalization is likely to amplify, not reset, existing patterns, deepening the concentration of export revenue in municipalities that were already oriented toward a given commodity.

Learning-by-exporting also matters at the municipal scale. Using firm-level data from Brazil, Ferraz and Rossi-Hansberg (2023) find that knowledge acquired in foreign markets disseminates through local labor mobility, raising average productivity within the municipality. When local firms learn to meet international standards for soy traceability or beef sanitary protocols, this knowledge elevates entry barriers for unrelated products, indirectly reinforcing commodity concentration.

Finally, environmental and regulatory constraints can interact with concentration. The European Union Deforestation Regulation and similar initiatives reward municipalities able to demonstrate compliance with land-use restrictions. Sauer and Lima-Ribeiro (2024) document how Brazilian municipalities with lower recent deforestation secure faster growth of certified coffee exports, whereas soy-dominated areas facing compliance gaps risk market exclusion.

These regulatory dynamics can reroute investment within and across municipalities, shaping where export growth and concentration materialize.

Collectively, these strands of theory suggest that municipal-level specialization and commodity concentration are entwined through a combination of comparative advantage, spatial frictions, agglomeration, learning, and regulatory pressure. The present study positions the Export-Intensive Gini Index within this conceptual framework, enabling an empirical test of whether municipalities with higher export-to-revenue ratios display sharper commodity concentration once these mediating factors are taken into account.

3 Methods

The study applies to the double machine learning estimator of Chernozhukov et al. (2018) to uncover a causal relationship between municipal export openness, named export-intensive Gini index and commodity-specific concentration. The procedure unfolds in two stages. Flexible algorithms such as gradient-boosted trees show how a wide array of local attributes, such as geographic position, infrastructure, demographics, governance, affect the export-to-revenue ratio and our concentration index. Second, the residual components of concentration that remain unexplained by those attributes are related to one another through cross-fitting, which separates training from testing folds and limits overfitting. This combination of non-parametric prediction with orthogonal regression captures complex interactions, filters out spurious correlations, and still provides standard errors that permit statistical inference. Consequently, the estimated coefficient for each commodity reflects the expected change in its concentration index that accompanies a one-unit increase in municipal export openness, after controlling for hundreds of observable characteristics.

The empirical strategy couples three building blocks: a two-stage spatial downscaling algorithm that assigns customs-recorded exports to origin municipalities without altering state totals or product composition; a Monte-Carlo-based Export-Intensive Gini Index that fixes one commodity, randomly redistributes the residual basket 5 000 times and averages Lorenz areas to gauge concentration; and a double machine learning estimator that relates this index to the export-to-GDP ratio while controlling for hundreds of geographic, climatic, infrastructural, demographic and institutional covariates through cross-fitting and orthogonalization.

3.1. Double Machine Learning (DML)

The empirical strategy centers on the DML estimator introduced by Chernozhukov et al. (2018). Assessing how export concentration shapes the commodity-specific Export-Intensive Gini Index demands an econometric tool that can sift through a high-dimensional list of potential confounders. The index itself is engineered to isolate a single product's spatial concentration by holding the remainder of the export basket constant through Monte Carlo simulation; DML supplies the parallel step on the statistical side, flexibly modeling the multifaceted influence of geography, infrastructure, human capital, and governance on both openness and concentration. By combining modern machine-learning algorithms with orthogonalization and cross-fitting, the method extracts the residual, causal link between the export-to-revenue ratio and the index while guarding against bias that would arise if nonlinear interactions or omitted combinations of municipal attributes were ignored.

Consider the following specification for a linear model:

$$Y = D\theta + g(X) + \epsilon, \qquad E(\epsilon | D, X) = 0 \tag{1}$$

$$D = m(X) + V, \qquad E(V | X) = 0 \tag{2}$$

θ is the treatment parameter, associated with the treatment variable D. Moreover, θ measures the average change in municipalities' concentration that accompanies a marginal rise in openness after all observable municipal attributes have been removed from both sides of the relationship. This specification, accounts for a vector of covariates X affecting Y and the treatment variable D in different ways. The functions $m(\cdot)$ and $g(\cdot)$ are not necessarily the same. The proposed solution by Chernozhukov et al. (2018) is to leverage machine learning methods to estimate a partialled out D and X and the partialled out effect of X on Y and then use the residuals of this regression to estimate the effect of D on Y, capturing the treatment effect given by θ.

However, that does not explain exactly why machine learning is a good procedure to carry out this exercise. A simple ordinary least squares (OLS) regression could, in principle, deliver an unbiased estimate of θ only if $g(\cdot)$ and $m(\cdot)$ were linear and if the covariate space were small. Nevertheless, in this study, the data set contains hundreds of variables describing geography, agronomic conditions, climate, logistics, urbanization, education, credit, and local institutions. Such breadth introduces myriad interactions and nonlinear thresholds; for instance, road density influences export intensity very differently in soy-producing frontiers than in traditional coffee zones. Enumerating every plausible functional form within an OLS framework would exhaust degrees of freedom and risk omitted-variable bias. Modern machine-learning algorithms overcome this challenge by flexibly approximating $g(\cdot)$ and $m(\cdot)$ with data-driven ensembles, such as gradient-boosted trees or random forests, that capture complex relationships automatically avoiding over-fitting.

At the same time, a two-stage procedure would contaminate the second-stage estimate of θ with the first-stage prediction error because the same observations would influence both stages. DML resolves this issue through cross-fitting. The sample of N municipalities is partitioned into K mutually exclusive folds of comparable size. For a given fold k, $g(\cdot)$ and $m(\cdot)$ are trained on the remaining $K - 1$ folds, and the fitted models generate out-of-sample predictions $\hat{g}_{-k}(X)$ and $\widehat{m}_{-k}(X)$ for the observations withheld. Residuals $\tilde{Y} = Y - \hat{g}_{-k}(X)$ and $\widetilde{D} = D - \widehat{m}_{-k}(X)$ are orthogonal to X in expectation, so regressing $\tilde{Y}$ ~ on $\widetilde{D}$ within fold k yields an unbiased estimate $\hat{\theta}_k$. Repeating this procedure across all folds and averaging the $\hat{\theta}_k$ values produce a final estimator that remains $\sqrt{N}$-consistent and asymptotically normal, even though the nuisance functions are learned with flexible machine-learning algorithms. Because every residual is based on out-of-sample predictions, cross-fitting also curbs overfitting and preserves efficiency across diverse data configurations. To further control bias-variance trade-offs, hyperparameters are tuned through nested cross-validation applied within each training subset.

Diagnostic checks include: (i) overlap assessment to confirm that municipalities with similar X display a wide support of D; (ii) placebo regressions that replace Y with pre-treatment outcomes to gauge spurious correlation; and (iii) sensitivity analysis to alternative fold partitions. Together, these steps ensure that the estimated θ reflects a causal link from export concentration rather than artifacts of model selection or sample idiosyncrasies. By merging flexible, high-dimensional prediction with rigorous orthogonalization, the DML strategy supplies a consistent approach between modern machine learning and classical econometric inference. In the context of Brazilian municipal exports, this strategy allows the research to quantify how export-to-revenue ratios reshapes commodity concentration while honoring the intricate web of local characteristics that might otherwise confound the relationship.

Over-fitting can be a problem in both ordinary least squares estimation and double machine learning, when specification is specialized in the training data, but it does not generalize well to out-of-sample data. The K-folding (or cross-fitting) explained previously decreases the possibility of over-fitting, while increasing the efficiency of the estimator of θ in different data scenarios.

3.2 Exports Imputation

Brazil's official trade database (Comex Stat) links each shipment to the municipality where the exporting firm files customs paperwork, not to the place where the good is produced. For bulk agribusiness commodities, this impedes the connection between local production conditions and recorded exports. Leal and Martins (2025) overcome the problem with a two-step regional-economics procedure (spatial matrix and downscaling), reallocating state-level trade flows to producing municipalities. We follow their method completely.

Let $E_{i,HS6}$ be the imputed value of exports of a specific HS-6 product originating in municipality i, and $E_{i,HS4}$ the officially reported exports of the corresponding HS-4 aggregate. The reallocation combines a spatial diffusion term with a downscaling adjustment:

$$E_{i,HS6} = \alpha W E_{i,HS4} + rE *_{i,HS6} \quad (3)$$

Where:

- $i = 1, \dots, I$, is a single municipality.
- $E *_{i,HS6}$ is defined as $E *_{i,HS6} = \sum_{i=1}^{I} \alpha(E_{i*,HS4} - WE_{i*,HS5})$, a residual term that guarantees the state total for every HS-6 code is matched exactly after spatial redistribution.
- $\alpha = \frac{E_{state,HS6}}{E_{sate,SH4}}$ rescales the HS-4 value to the HS-6 level, preserving each state's observed product mix.
- r is an $n \times 1$ vector describing the weights applied to the downscaling exercise. In this case, as in Leal and Martins (2025), r comes from Brazilian labor data, RAIS, from the Ministry of Labor.

The first component, $WE_{i,HS4}$, diffuses recorded exports across space in proportion to distance, mirroring the cost advantage of shipping through the nearest customs post. The second component, $rE *_{i,HS6}$, reallocates any residual so that municipalities with greater employment in a commodity's processing or logistics receive a larger fraction of the remaining volume.

Equation (3) is applied to the five commodities that dominate Brazil's agribusiness trade: soybeans, corn, beef, coffee, and sugar, yielding municipality-level HS-6 for 2023. These imputed series underpin both the Export-Intensive Gini Index and the measure of municipal export-to-revenue ratios used in the causal analysis.

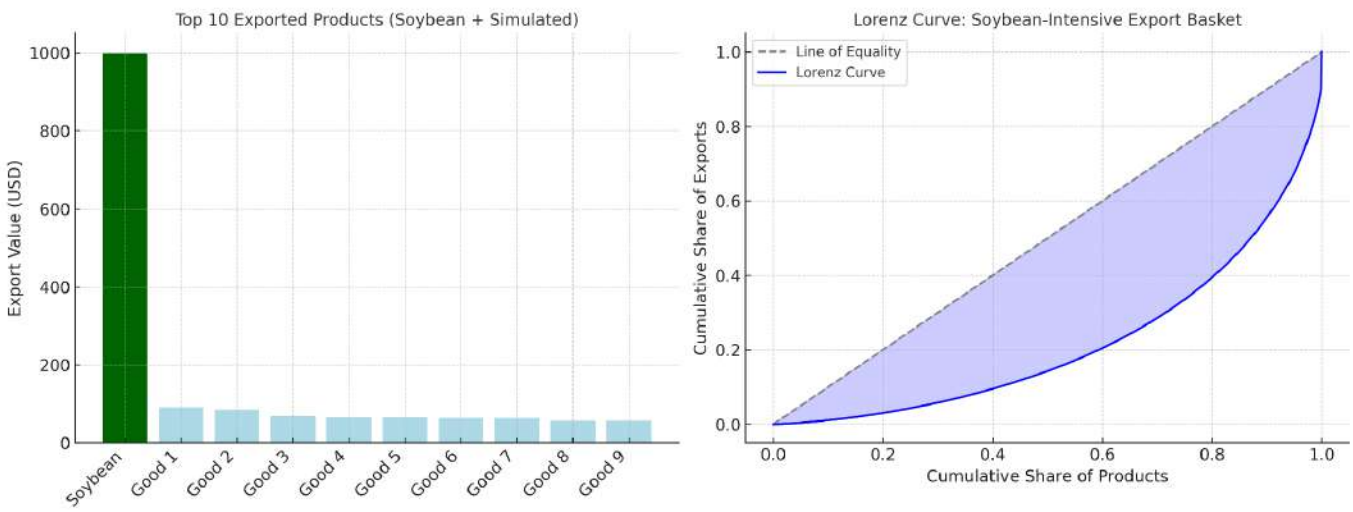
3.3 Good-Intensive Gini Index

Our dependent variable is an export-intensive Gini index. For each commodity considered, we have built a different measure of export concentration. This measure is a Gini-type index, meaning that there is an area of discrepancy between the ideal distribution of export values in a municipality and its realized values. Conceptually, the index is the area between the Lorenz curve generated by a municipality's simulated export distribution and the forty-five-degree line that represents perfect equality. The larger the area, the sharper the concentration. The full distribution of exports, occurring for each Brazilian municipality for the five goods considered (soybean, beef, coffee, sugar, and corn[5]), is not known, given its high dimensionality aspect. In order to overcome this difficulty, we create a good-intensive Gini index for each commodity. This index is created by following the next steps:

Because the complete joint distribution of municipal exports is unobserved at six-digit HS resolution, the index is obtained with a Monte Carlo procedure. First, the imputed export value of the target commodity in 2023 is fixed for each municipality (section 3.2). Second, five thousand random reallocations of the remaining export basket are drawn, each time keeping the municipality's total exports constant while varying the shares of the other one thousand HS-6 goods. Third, the Gini coefficient is calculated for every draw, and the mean across simulations becomes the export-intensive Gini for that municipality–commodity pair. The procedure is repeated for all municipalities; for example, a soybean-intensive Gini vector, a corn-intensive vector, and so forth, for municipality A. The index ranges from zero (complete spatial equality) to one (all exports of the commodity originate in a single municipality); the same value can arise from different underlying distributions, so caution is required when comparing municipalities with identical scores.

Figure 1 illustrates the formation of the good-intensive Gini index. In the left panel, the export shares of a representative municipality are displayed, with soybeans singled out as the target product and goods 1 through 9 aggregated as "non-soy" for modeling purposes so that only soybean concentration is measured. The right panel shows one of the 5,000 Monte Carlo–generated Lorenz curves for soybean exports, where the area between that curve and the 45-degree line captures spatial concentration. Averaging these areas over all simulations yields the soybean-intensive Gini index for each municipality.

Figure 1. Good-Intensive Gini Index creation.



Note: Authors' elaboration.

A Gini index possesses the following properties:

- It ranges from 0 to 1, indicating an equanimous distribution of exports values for all goods and perfect concentration of exports value in a single good, respectively.
- The same Gini index, α can support several export distributions in values, that is, there is not a 1:1 mapping from a Gini index value to a single distribution of exports value. This also implies that two municipalities may have the same Gini value, generated by two different distributions of exports.

[5] You can download the Good-intensive Gini Index for all five goods considered in this analysis here.

- The Gini index captures values dispersion in a distribution, ignoring which values produce this distribution.

3.4 Explicative variables

The study relies on an extensive set of control variables. For clarity, **Table 1** organizes them into thematic groups. Most observations correspond to 2023, and when a specific series is unavailable for that year, the nearest recent data point is used instead.

Table 1. Explicative variables.

Theme	Variables/Source	Justification
Added Values	Added Values for agribusiness, industry, services, and public administration from IBGE	Economic diversification variables
Rain	Mean monthly rain from INMET	Meteorological variables relevant to primary goods
Distance to infrastructure	Distance to nearest ports and airports from Ministry of Infrastructure	Ease of exports
Distance to capitals	Distance to all 26+DF capitals	Ease of access to economic centers
PIX variables	Payments, number of payments of person and businesses using PIX, any combination from Brazilian Central Bank. The variables are values and quantities transferred from and to physical persons and corporations.	Precise measure of economic and monetary activity
IBGE basic demographics	Number of households, population, area, literacy, age, ratio sex, old age ratio, etc., from Instituto Brasileiro de Geografia e Estatística (IBGE)	Main qualifiers of a municipality in terms of demographics
ESTBAN	Number of bank branches from Brazilian Central Bank	Access to credit and financial products
Anatel's Brazilian connectivity index	Variables used to create the Brazilian connectivity index, such as presence of backhaul of optic fiber, percentage of use of 4G, concentration indexes of mobile and household ITC measures, from Anatel.	Access to information
Sistema Nacional de Cadastro Rural	Mean and total size of rural properties from Ministry of Agrarian Development	Size of the primary economy in the municipality in terms of area
PRODES	Municipal deforestation from PRODES	Pressure on natural resources for economic production
CAPAG	Payment capacity in fiscal terms Ministry of Finance	Municipalities better positioned in terms of fiscal capacity
CAR	Mean fiscal module of properties in CRA, Mean Area, Sum of fiscal module, sum of area, proportion of validated properties in CAR from Ministry of Environment	Relationship between production and environmental compliance
RAIS	Income and education of formal labor force from Ministry of Labor	Labor force in a given municipality

Note: Authors' elaboration.

3.5 Treatment

The treatment variable is the export-to-GDP ratio in 2023. Imports are unavailable at the municipal scale, so the conventional $(exports + imports)/GDP$ measure cannot be used[6]. The outcome variable is the commodity-specific Export-Intensive Gini Index described in Section 3.3. Throughout the paper we refer to the treatment as export intensity.

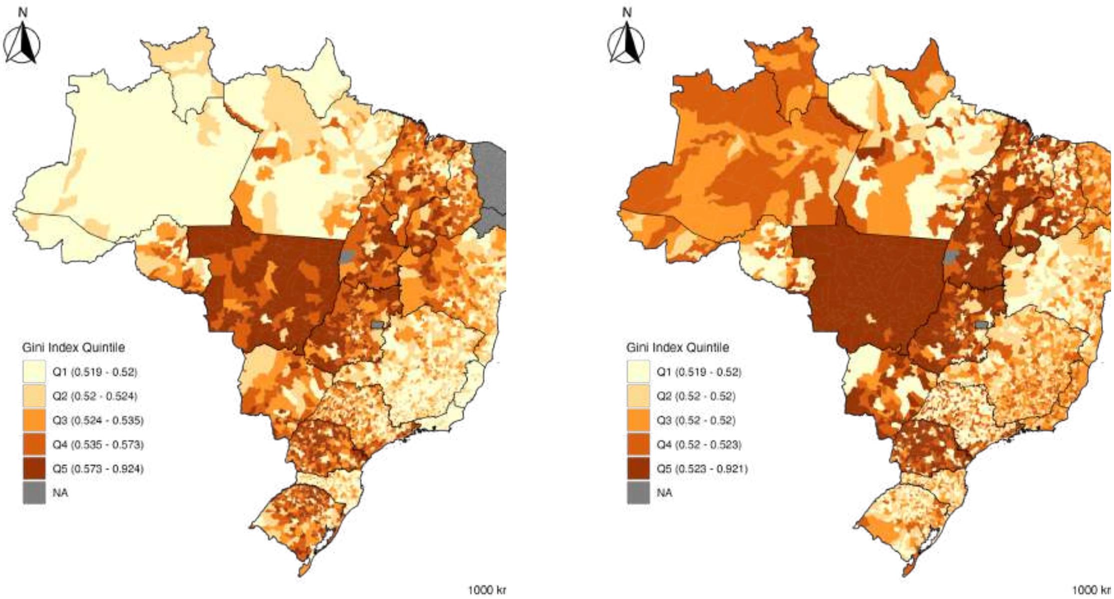
4 Results

4.1 Descriptive statistics

The analysis reveals distinct patterns of export concentration across Brazil's five major agricultural commodities, with interesting implications for understanding how municipal trade openness affects agricultural specialization. The spatial distribution of commodity-intensive Gini indices demonstrates marked heterogeneity across Brazilian territory, while econometric results provide evidence of varying relationships between export openness and concentration depending on the specific commodity examined. The spatial distribution of export concentration varies across commodities, reflecting Brazil's diverse agricultural geography and regional specialization patterns. The results are presented in Figure 2, which displays in map format the distribution of the good-intensive Gini index for the 5 goods analyzed in this paper (soybeans, corn, beef, coffee, and sugar).

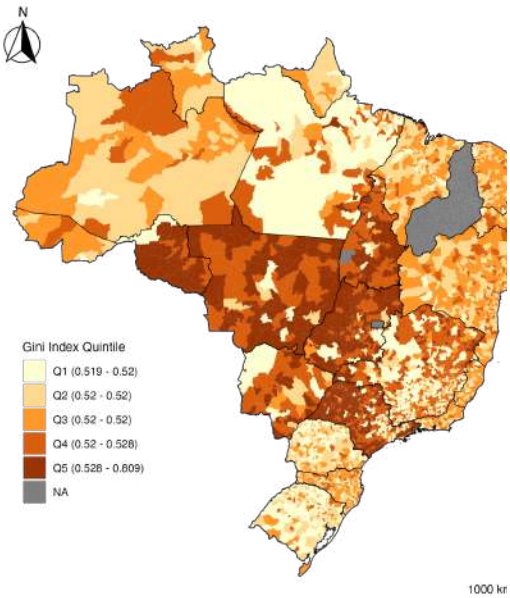
Figure 2. Good-intensive Gini index municipality profile.

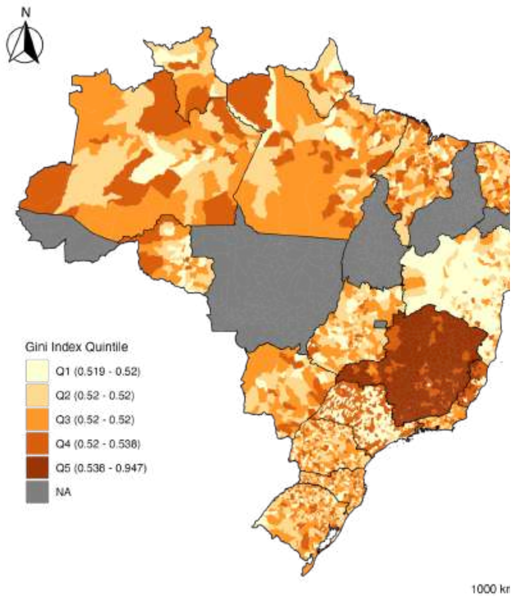
(a) Soybean-intensive Gini index (b) Corn-intensive Gini index



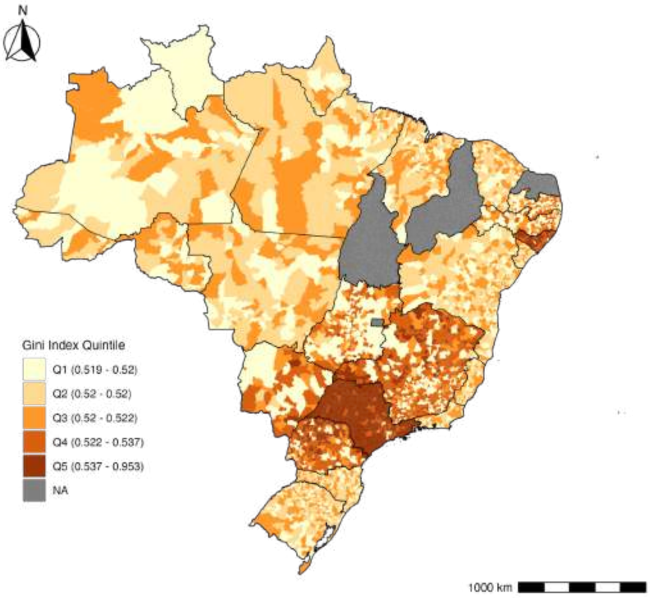
(c) Beef-intensive Gini index (d) Coffee-intensive Gini index

[6] Gross Domestic Product.





(e) Sugar-intensive Gini index



Note: Authors' elaboration.

Soybean concentration (figure 2.a) exhibits the highest values in the Center-West region, particularly in Mato Grosso and western Bahia, where the Gini index reaches values above 0.9 in many municipalities. This pattern aligns with Brazil's agricultural frontier expansion documented by Du et al. (2024), who demonstrate how export-oriented agricultural development concentrates in areas with favorable agro-climatic conditions and infrastructure access. The concentration in these regions reflects the emergence of specialized soybean clusters that benefit from economies of scale in production, processing, and logistics.

Corn concentration (figure 2.b) displays a similar but less pronounced spatial pattern, with high concentration values in the Center-West and parts of the South region. The overlap with soybean areas reflects the complementary nature of these crops in Brazilian farming systems, where corn often serves as a second crop (safrinha) following soybean harvest. This dual-crop system creates synergies that reinforce regional specialization, consistent with Henderson et al. (2021) findings on how agglomeration economies strengthen commodity clusters.

Beef concentration (figure 2.c) shows a distinct pattern, with the highest values in traditional cattle-raising regions of Mato Grosso do Sul, Goiás, and parts of Minas Gerais. The spatial distribution reflects both natural comparative advantages in pasture-based systems and historical investments in cattle infrastructure. Unlike crop production, beef concentration appears less dependent on proximity to ports, reflecting different logistics requirements and the ability to transport live cattle over longer distances.

Coffee concentration (figure 2.d) demonstrates the most geographically constrained pattern, with high values concentrated in traditional coffee regions of Minas Gerais, Espírito Santo, and parts of São Paulo. This spatial concentration reflects both agro-climatic requirements specific to coffee cultivation and path-dependent investments in processing infrastructure and quality certification systems. The geographic specificity of coffee production creates natural barriers to diversification, supporting theoretical predictions about how environmental constraints shape agricultural specialization patterns.

Sugar concentration (figure 2.e) exhibits moderate spatial clustering, primarily in São Paulo's traditional sugarcane regions and expanding areas in Goiás and Minas Gerais. The pattern reflects both historical investments in sugar mills and ethanol plants and the crop's specific agro-industrial requirements. The relatively lower concentration values compared to other commodities may reflect sugar's dual market orientation toward both export and domestic consumption.

In the Appendix, Figure A.1, we display the relationship between our Good-Intensive Gini index and the treatment variable.

4.2 Econometric results and causal relationships

The double machine learning estimates reveal heterogeneous effects of export openness on commodity concentration, with statistically significant positive relationships for most commodities but varying magnitudes. Coffee exhibits the strongest relationship (coefficient = 0.0667), followed by corn (0.0421), soybeans (0.0331) and sugar (0.0145), all significant at the 1 % level. For beef the effect is –0.0180 and significant at 5 %, signaling mild deconcentration. Ordinary-least-squares coefficients are uniformly larger, reinforcing the argument that neglecting non-linearities inflates estimated effects. The results are presented in the following **Table 2**.

Table 2. Impact of municipal-export opening on exports concentration.

Good	OLS	Double Machine Learning	Number of observations
Soybeans	0.00907*** (0.00485)	0.03306*** (0.00994)	4632
Beef	-0.00456 (0.00606)	-0.01798** (0.00573)	4879
Corn	0.11880*** (0.00280)	0.04206*** (0.00938)	5178
Coffee	0,38050*** (0.00815)	0.06667*** (0.01128)	4291
Sugar	0.03060*** (0.00874)	0.01454** (0.00669)	5039

Note: Authors' elaboration. Standard deviation in parenthesis. *** Significant at 1%; ** Significant at 5%; * Significant at 10%.

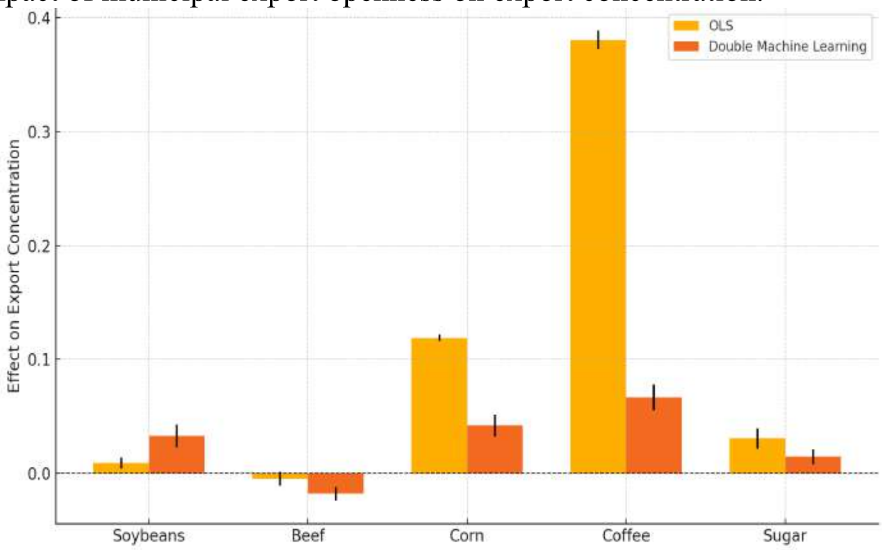
However, before diving into the possible causes of this non-concentration, we analyze the result for each good, presenting the accompanying OLS estimate. In the case of soybeans, the DML estimator indicates that an increase of one unit in the exports share in the municipality

GDP leads to 3.31% in the soybeans-intensive Gini index, indicating that a more intense relationship with foreign market concentrates the production directed abroad. The conclusion from OLS is similar, although the effect is much smaller (about 0,9%).

Regarding the beef results, we find that the OLS estimates are negative and non-significant, whereas the DML estimates are significant at the 5% level and also negative. Considering an increase of 1 in the exports share with regard to GDP, then the concentration reduces by 1.8%. This result might seem counterintuitive, but we explore it in the Discussion section.

Corn results are all positive and significant, indicating that a more open municipality to exports leads to an increase in the exports concentration of about 4.20% (DML estimator), whereas this increase amounts to 11.88% for the ordinary least squares (OLS) estimator. Coffee estimates are also positive and significant, with a more open municipality (one increase in the ratio exports-GDP) amounting to a change in concentration of 6.66%, while this estimate using OLS equals 38.50%. Sugar's estimates are also positive and significant, with a more open to exports municipality seeing an increase in concentration of 1.45% (DML) and 3.06% (OLS). Figure 3 shows these coefficients in a graph format, which allows for easy comparison.

Figure 3. Impact of municipal export openness on export concentration.



Note: Authors' elaboration.

Coffee and corn have the largest increase in export concentration whenever there is an increase in the exports-GDP municipalities ratio. One methodological feature is how conservative (i.e., tending to 0) the DML estimator is when compared to the OLS estimator. This happens among all goods because OLS does not consider as many non-linearities in the relationship between the dependent variable, treatment, and explanatory variables as the DML estimator.

These results provide empirical support for theoretical predictions from firm heterogeneity models (Melitz, 2003) while revealing important commodity-specific variations. The strong positive relationship for coffee aligns with the crop's geographic specificity and high-quality requirements, which create natural barriers to diversification. Municipalities with established coffee production capabilities and export infrastructure face strong incentives to specialize further as export opportunities expand, consistent with learning-by-exporting mechanisms documented by Juergensen et al. (2024).

The moderate positive relationship for corn reflects the crop's integration into Brazil's agricultural modernization process. As Ranran and Jingsuo (2024) note, agricultural production agglomeration generates spillover effects that strengthen regional specialization. Corn's role as a complementary crop in soybean systems creates synergies that reinforce concentration as export markets expand.

The weak positive relationship for soybeans, despite the crop's prominence in Brazilian exports, suggests that soybean production may have reached a level of geographic dispersion where additional export openness does not significantly increase concentration. This finding aligns with the spatial patterns observed in the maps, where soybean production has expanded across multiple regions, potentially reducing the marginal effect of export openness on concentration.

Beef's negative coefficient likely reflects the sector's distinct production logic. Cattle fattening, slaughter and export clearance often occur in different municipalities; long biological cycles and stricter sanitary requirements encourage a more dispersed supply chain than row-crop agriculture. These features weaken the tight link between local export intensity and spatial concentration observed for annual crops.

3.3. Discussion

The previous results are illuminating for many reasons. First, 4 out of 5 goods considered indeed have a good-intensive Gini index more concentrated when the municipality exports more, considering its GDP. Second, DML estimates are more conservative than OLS estimates, indicating that, unless we have employed this method, our estimates would be inflated.

With regards to the beef-intensive Gini index being disconcentrated when considering a more open to exports municipality, its productive nature is different from the remaining 4 products considered. Soybeans, corn, coffee, and sugar are agricultural goods, produced directly from the plantation to harvest. Beef production follows a different pattern. First, it appears that problems plaguing export data in Brazil are more acutely present in the beef exports. For instance, both livestock food and cattle slaughter probably do not happen in the same municipality where the export happens. Moreover, beef production differs completely in nature from the other 4 goods considered here. Beef has more added value, uses more land per output, is a long-term product, and it is up to more phytosanitary and regulatory concerns when compared to the other four goods (soybeans, coffee, corn, and sugar).

This might justify why the export concentration does not happen with the increase of the municipal exports-GDP ratio. Moreover, the low, albeit significant and positive, correlation between the beef-intensive Gini index and the export opening is revealing, reinforcing the argument that beef production directed to foreign markets indeed happens differently from the other four goods considered. Even considering these explanations, this beef puzzle is informative, given that, in environmental and climate terms, beef production is more aggressive than any of the other four goods considered. It usually leads to more deforestation and more emissions from production, not only land use change.

A further interesting analysis pertains to the size of the coefficients found in the previous estimation. Considering the four goods with increased concentration given more export-opening, we have that coffee, corn, soybeans, and sugar are in descending order of a larger coefficient. Moreover, they tend to have their exports concentrated when there is more export-oriented activity in a given municipality. Coffee is a perennial good, meaning that once planted, it takes up to 5 years to reach full production, and when properly managed, tends to produce a large harvest for many years. This is not the case for corn, soybeans, and sugar. The three crops need to be planted every year to produce annually. In the case of corn, production might take place twice a year. Hence, these three productions tend to be more elastic in the short term to market movements, including market concentration. Coffee, however, displays a lagged response to market movements; hence, one might say there is even hysteresis in how coffee concentration happens. This hysteresis does not necessarily imply lower concentration; it probably implies cumulative concentration, given the inability of the coffee culture to answer as quickly to market movements as the other three crops (soybeans, corn, and sugar).

The results carry important implications for regional development policy and risk management strategies. The strong positive relationship between export openness and concentration for coffee and corn suggests that municipalities specializing in these commodities face increasing exposure to international price volatility as they become more integrated into global markets. This finding supports concerns raised by Pardy and Rodríguez-Pose (2025) about how trade integration can exacerbate regional economic inequalities.

For coffee-producing municipalities, the high concentration levels combined with strong sensitivity to export openness indicate particular vulnerability to external shocks. Policy interventions might focus on developing complementary activities within coffee value chains, such as specialty processing or agritourism, rather than attempting to diversify into unrelated agricultural activities where municipalities lack comparative advantages. The moderate effects for corn and sugar suggest opportunities for managed diversification strategies that leverage existing agricultural capabilities while reducing concentration risks. These commodities' integration into broader agricultural systems provides natural pathways for diversification that maintain synergies with existing production capabilities.

The spatial patterns revealed in this study also inform infrastructure investment priorities. The concentration of high-value agricultural exports in specific regions highlights the importance of targeted investments in transportation, storage, and processing infrastructure to support continued competitiveness while managing concentration risks.

These findings contribute to several strands of economic literature while revealing new patterns specific to agricultural trade in developing countries. The commodity-specific variations in concentration responses support theoretical predictions from new trade theory about how product characteristics interact with trade costs to shape specialization patterns. The strong effects for coffee align with literature on quality differentiation and geographic branding, while the weaker effects for soybeans reflect the commodity's more standardized nature.

The spatial patterns documented here complement recent work on agricultural transformation in Brazil while providing new insights into municipal-level dynamics. The concentration of export-oriented agriculture in specific regions supports findings by Du et al. (2024) on how agricultural expansion responds to global demand, while revealing the heterogeneous impacts across different commodities and regions.

The results also contribute to literature on spatial spillovers and agglomeration economies in agriculture. The clustering patterns observed for different commodities provide evidence of how specialized agricultural systems generate self-reinforcing advantages through infrastructure development, knowledge spillovers, and supply chain integration. These findings support theoretical predictions about how initial comparative advantages become amplified through agglomeration processes.

5 Concluding remarks

This paper carried out an ambitious empirical exercise to test the hypothesis that a municipality more engaged with exporting selected primary goods tends to have a more concentrated exports profile. The study advances subnational trade research in three ways. First, the Export-Intensive Gini Index measures commodity-specific concentration while holding the rest of the export basket constant, allowing precise comparisons across municipalities and products. Second, the double machine learning estimator isolates causal effects in the presence of hundreds of geographic, demographic and institutional covariates, overcoming the omitted-variable concerns that limit traditional regressions. Third, the spatial downscaling algorithm of

Leal and Martins (2025) reallocates customs data to production sites, revealing patterns concealed in aggregate statistics and offering a template for other countries with similar data constraints.

The empirical exercise shows a clear pattern: greater export-to-GDP ratios are associated with sharper commodity concentration in soybeans, corn, sugar and coffee, whereas beef displays a modest move toward diversification. The estimated coefficients are largest for coffee and corn, which means that municipalities focused on these two crops become progressively more exposed to international price swings as trade integration deepens. Coffee-growing areas, already characterized by high concentration, appear especially vulnerable; policy could concentrate on upgrading activities within the coffee value chain, such as specialty roasting, quality certification, and agritourism, rather than forcing diversification into unrelated crops in which the locality lacks comparative advantage. Corn and sugar present more moderate effects, suggesting scope for "managed diversification" that builds on existing agronomic capabilities while diluting exposure to any single product. The spatial distribution of concentration also guides infrastructure priorities: targeted investment in transport corridors, storage, and processing facilities can support competitiveness and at the same time help manage the risks that stem from dependence on a narrow export base.

The contrasting result for beef hints at an avenue for risk mitigation. Livestock production involves longer production cycles, higher per-unit value, and stricter sanitary standards, elements that appear to foster a broader export mix at the municipal level. Promoting more complex or higher value-added activities, whether in meat processing, agro-industrial inputs, or other sophisticated goods, may therefore offer a path toward export diversification without abandoning comparative advantages in agriculture.

This question matters in several ways. First, primary commodities tend to suffer more price volatility from international markets, allowing at times an influx of export income, whereas at other times, export income is decreased severely. Hence, less export concentration means more diversification in exports and reduces a municipality's exposure to a single commodity market price variation. In a second moment, labor employed in a single commodity production also becomes a more fragile type of labor, hunting workers' income whenever prices are low and hurting their employability due to their super-specialized and non-convertible expertise.

Several limitations suggest directions for future work. The analysis covers five major commodities; extending the coverage would clarify whether the observed relationships prevail for other agricultural or industrial products and whether substitution across goods tempers concentration. The cross-sectional design prevents examination of how concentration evolves; panel data across multiple years could uncover dynamic processes of diversification or path dependence. Linking export specialization to wider development outcomes, income growth, employment stability, and resilience to shocks would inform the debate on whether trade integration magnifies or reduces regional inequalities. Finally, replicating the methodology in other developing economies would test its generality and identify context-specific factors that shape the openness-concentration nexus.

References

Bilgin, N. M., Fala, E., & Ottaviano, G. (2024). Technology Spillovers, Diffusion and Rivalry in Firm Networks. *CEPR Discussion Paper* DP19804. https://cepr.org/publications/dp19804

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Giroud, X., Lenzu, S., Maingi, Q., & Mueller, H. (2024). Propagation and Amplification of Local Productivity Spillovers. *Econometrica*, 92(5), 1589-1619. https://doi.org/10.3982/ECTA20029

Juergensen, J., Love, J. H., Surdu, I., & Narula, R. (2024). Learning-by-exporting: The strategic role of organizational innovation. International Business Review, 33(6), 102339. https://doi.org/10.1016/j.ibusrev.2024.102339

Krugman, P. R. (1980). Scale economies, product differentiation, and the pattern of trade. American Economic Review, 70(5), 950-959.

Krugman, P. R. (1991). Increasing returns and economic geography. Journal of Political Economy, 99(3), 483-499.

Leal, A., & Martins, M. M. V. (2025). Brazilian exports imputation: A new algorithm for estimating the municipal production directed at the foreign market. *Revista Brasileira de Estudos Regionais e Urbanos*, 19(2), 266-288. https://doi.org/10.54766/rberu.v19i2.1156

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica, 71(6), 1695-1725. https://doi.org/10.1111/1468-0262.00467

Nassif, A., Bresser-Pereira, L. C., & Feijó, C. (2020). The case for reindustrialisation in developing countries: Towards the connection between the macroeconomic regime and the industrial policy in Brazil. *Cambridge Journal of Economics*, 44(2), 355-381.

Ohlin, B. (1933). Interregional and international trade. Harvard University Press.

Oreiro, J. L., & Feijó, C. A. (2010). Desindustrialização: conceituação, causas, efeitos e o caso brasileiro. *Revista de Economia Política*, 30(2), 219-232.

Pardy, M., & Rodríguez-Pose, A. (2025). Trade Ties and Economic Divides: Trade and Income Inequality in the Regions of Europe. *Papers in Evolutionary Economic Geography*, 25.11. Utrecht University. http://econ.geo.uu.nl/peeg/peeg2511.pdf

Pardy, M., & Rodríguez-Pose, A. (2025). Trade Ties and Economic Divides: Trade and Income Inequality in the Regions of Europe. *Papers in Evolutionary Economic Geography*, 25.11. Utrecht University. http://econ.geo.uu.nl/peeg/peeg2511.pdf

Phillips, K. R. (2024). An Estimation of Import Exposure at the Subnational Level. *USITC Working Paper*. https://www.usitc.gov/sites/default/files/publications/332/working_papers/phillips_import_exposure.pdf

Ranran, C., & Jingsuo, L. (2024). The impact of agricultural production agglomeration on agricultural economic resilience: based on spatial spillover and threshold effect test. *Frontiers in Sustainable Food Systems*, 8, 1464732. https://doi.org/10.3389/fsufs.2024.1464732

Riaño, A. (2024). Reproducing the stylized facts that motivate models of international trade with heterogeneous firms. *The Journal of Economic Education*, 56(2), 139-153. https://doi.org/10.1080/00220485.2024.2438048

Ricardo, D. (1817). On the principles of political economy and taxation. John Murray.

Ricordel, P. (2024). Domestic or export: What is basic at the NUTS 2 regional level? A spatial endogenous regional growth model applied in the EU. *Growth and Change*, 55(1), 1-29. https://doi.org/10.1111/grow.12707

Ricordel, P. (2024). Domestic or export: What is basic at the NUTS 2 regional level? A spatial endogenous regional growth model applied in the EU. *Growth and Change*, 55(1), 1-29. https://doi.org/10.1111/grow.12707

Tabe-Ojong Jr., M. P., Nshakira-Rukundo, E., Zimmermann, A., & Jafta, Y. (2024). Trends and evolution of global value chains in food and agriculture: Implications for food security and nutrition. Food Policy, 127, 102679. https://doi.org/10.1016/j.foodpol.2024.102679

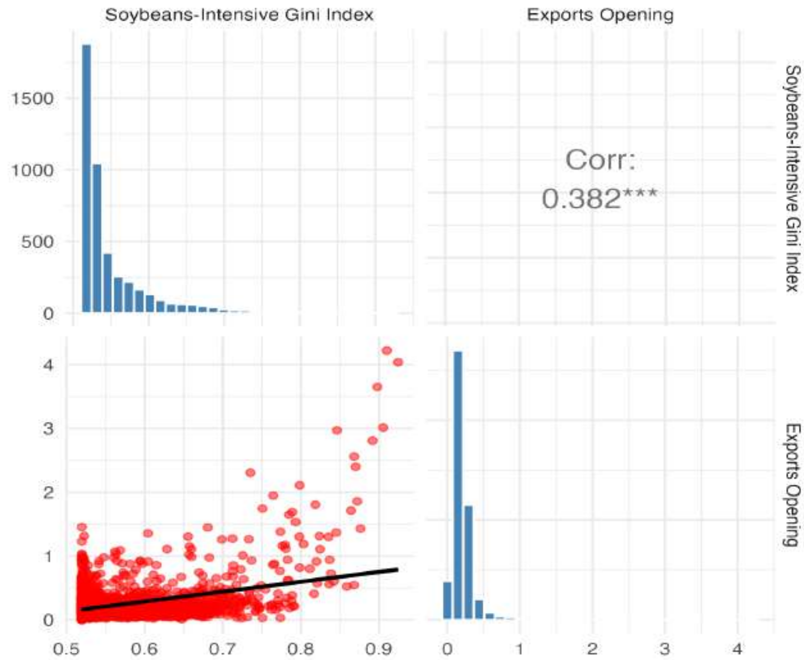
USDA Economic Research Service. (2025). Brazil. https://www.ers.usda.gov/topics/international-markets-us-trade/countries-regions/brazil

Wuri, J. (2024). The Role of Comparative Advantage in Enhancing Trade in Value-Added Using a Dynamic GMM Model. *Economies*, 12(7), 187. https://doi.org/10.3390/economies12070187

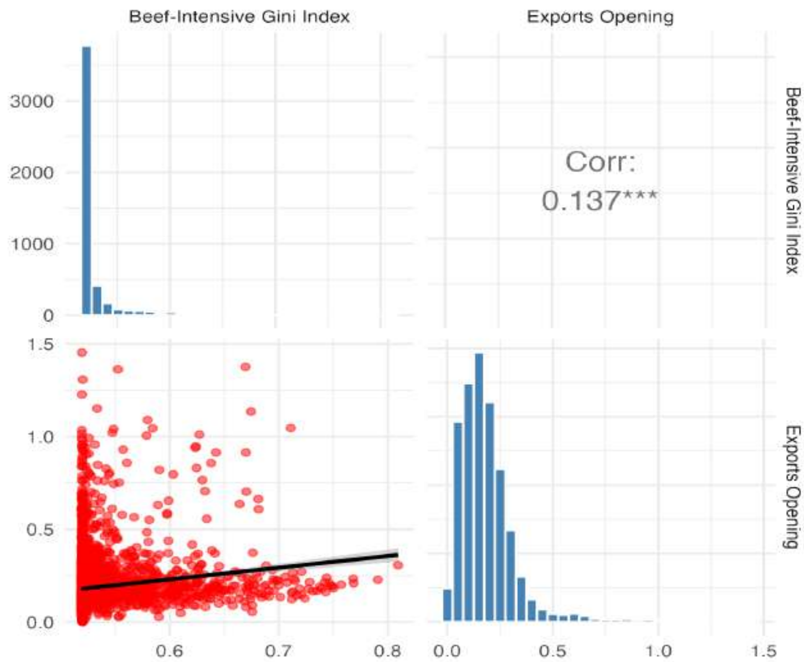
Appendix

Figure A.1. Good-Intensive Gini Index vs exports opening.

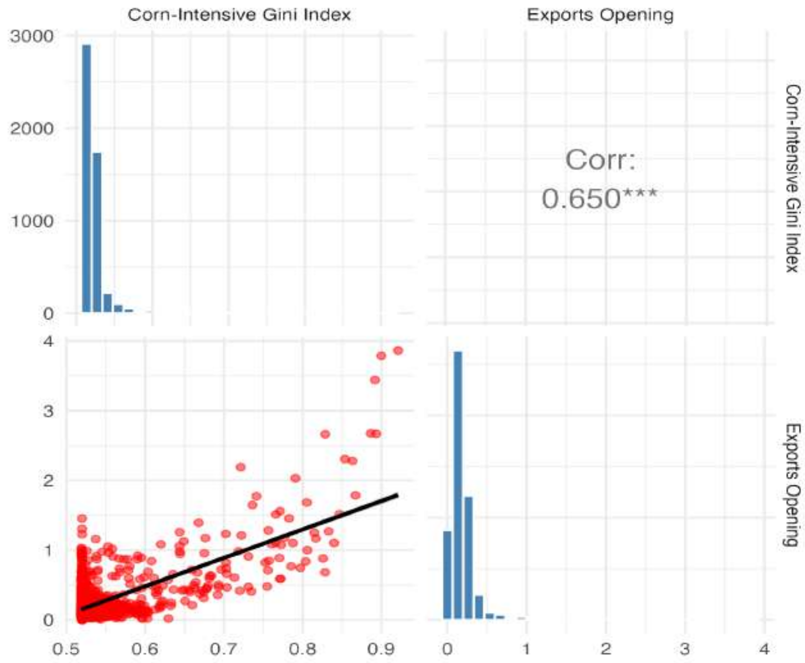
(a) Soybean-Intensive Gini Index vs exports opening.



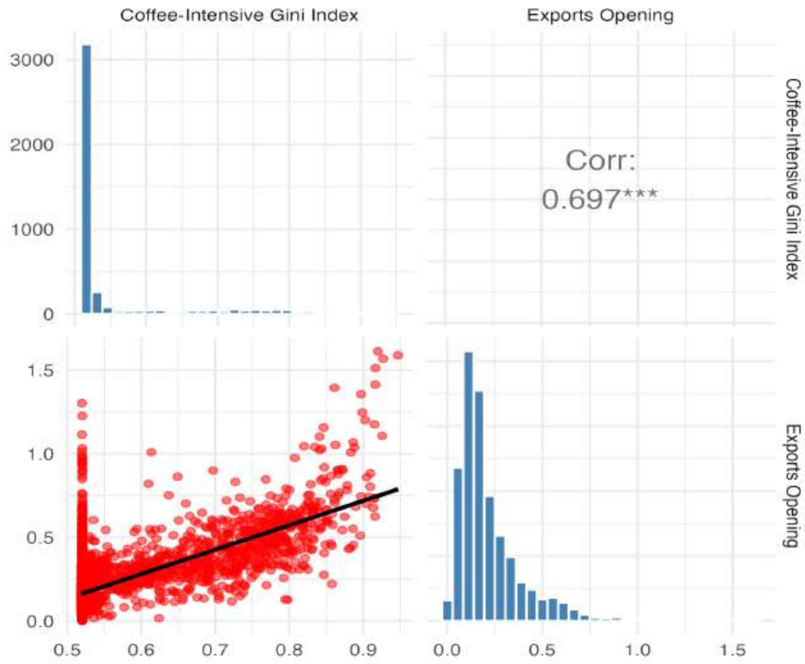
(b) Beef-Intensive Gini Index vs exports opening.



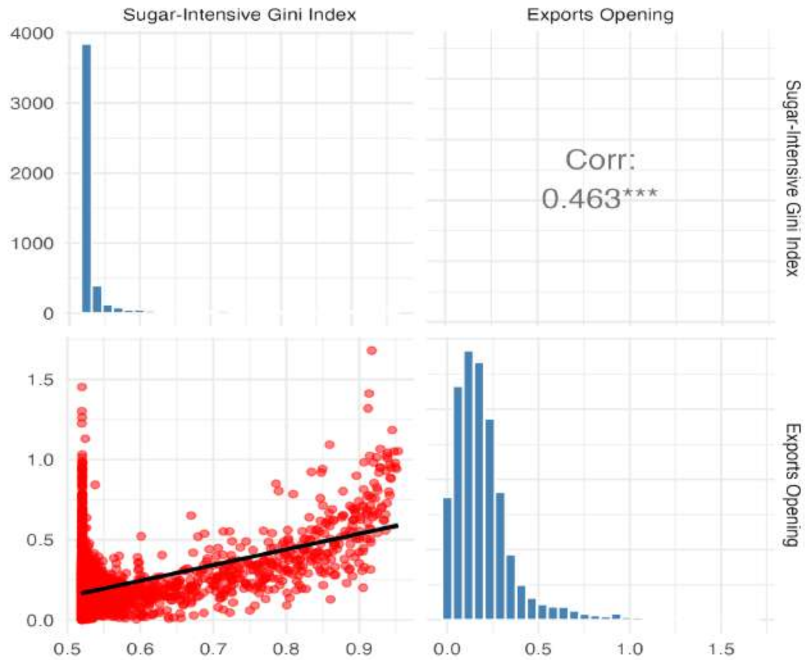
(c) Corn-Intensive Gini Index vs exports opening.



(d) Coffee-Intensive Gini Index vs exports opening.



(e) Sugar-Intensive Gini Index vs exports opening.



Note: Authors’ elaboration.