Department of Economics - FEA/USP

Legacies of Slavery or a Brave New World? Labor productivity and remuneration in the Brazilian coffee economy – new microdata evidence from Ibicaba plantation (1888-1958)

BRUNO GABRIEL WITZEL DE SOUZA

WORKING PAPER SERIES Nº 2026-16

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2026-16

Legacies of Slavery or a Brave New World? Labor productivity and remuneration in the Brazilian coffee economy – new microdata evidence from Ibicaba plantation (1888-1958)

Bruno Gabriel Witzel de Souza (bwitzelsouza@usp.br)

Abstract:

Studying the historical origins of income differentials in the past is crucial for explaining the socioeconomic mobility of distinct ethnolinguistic groups, in the short- and long-run. This issue is particularly pressing in Brazil, a country characterized by extreme levels of income inequality that correlate with racial components traceable to the country's historic slave-based economy. Yet rural labor markets in post-Abolition Brazil remain poorly understood, from a quantitative viewpoint, despite those emerging non-captive markets being also the main channel for the mass immigration of European fieldhands from the late 1880s to the 1920s. Such an important research gap reflects the scarcity of adequate microdata. The *Ibicaba Project* has begun to address this problem by creating a new archive in that plantation. *Ibicaba* is one of Brazil's most important historical plantations precisely for being the first rural unit to employ indentured European laborers during the early transition from slavery in the country, while, simultaneously, remaining the largest slaveholder in São Paulo's Old Western coffee zone. Drawing upon sixteen ledgers that record Ibicaba's entire resident labor force between 1888 and 1958 (with gaps), this paper studies the emergence of labor markets in the post-Abolition era and tests competing hypotheses about the role of ethnolinguistic origins in determining labor remunerations. The empirical analysis does not support the hypothesis of taste-based discrimination. After controlling for households' quantity-based TFP, ethnolinguistic origins did not lead to higher (positive discrimination in favor of European immigrants and their descendants) or lower (negative discrimination against Non-Whites) prices earned for the same type of homogenous agricultural task. However, results support the hypothesis of statistical discrimination. Based on labor arrangements highly dependent on intrahousehold divisions of labor, larger households were hired more frequently to cultivate coffee in the lean season and engaged for longer in harvesting. A major burden from slavery to the Black population in *Ibicaba* was the small number of economically-active household members. Finally, the paper traces the income profiles of various European immigrants and their descendants. German-speakers and Portuguese had labor income profiles closer to that of White Brazilians. The income profile of Italian immigrants and their descendants, however, shows evidence of households who performed rather poorly and probably failed to reach their "South American dreams".

Keywords: Age of Mass Migration; Slavery; Post-Abolition; Labor History; Ibicaba.

JEL Codes: N56; N36; N86.

TITLE:
Legacies of Slavery or a Brave New World? Labor productivity and remuneration in the Brazilian coffee economy – new microdata evidence from Ibicaba plantation (1888-1958)

AUTHOR:
Bruno Gabriel Witzel de Souza
Affiliation: Assistant Professor, Faculty of Economics and Business Management - University of São Paulo, Brazil
Email: bwitzelsouza@usp.br
Address: Av. Prof. Luciano Gualberto 908 - Cidade Universitária. FEA 2, Room 249. São Paulo – SP, Brazil. ZIP-Code: 05508-210

ABSTRACT
Studying the historical origins of income differentials in the past is crucial for explaining the socioeconomic mobility of distinct ethnolinguistic groups, in the short- and long-run. This issue is particularly pressing in Brazil, a country characterized by extreme levels of income inequality that correlate with racial components traceable to the country's historic slave-based economy. Yet rural labor markets in post-Abolition Brazil remain poorly understood, from a quantitative viewpoint, despite those emerging non-captive markets being also the main channel for the mass immigration of European fieldhands from the late 1880s to the 1920s. Such an important research gap reflects the scarcity of adequate microdata. The *Ibicaba Project* has begun to address this problem by creating a new archive in that plantation. *Ibicaba* is one of Brazil's most important historical plantations precisely for being the first rural unit to employ indentured European laborers during the early transition from slavery in the country, while, simultaneously, remaining the largest slaveholder in São Paulo's Old Western coffee zone. Drawing upon sixteen ledgers that record Ibicaba's entire resident labor force between 1888 and 1958 (with gaps), this paper studies the emergence of labor markets in the post-Abolition era and tests competing hypotheses about the role of ethnolinguistic origins in determining labor remunerations. The empirical analysis does not support the hypothesis of taste-based discrimination. After controlling for households' quantity-based TFP, ethnolinguistic origins did not lead to higher (positive discrimination in favor of European immigrants and their descendants) or lower (negative discrimination against Non-Whites) prices earned for the same type of homogenous agricultural task. However, results support the hypothesis of statistical discrimination. Based on labor arrangements highly dependent on intrahousehold divisions of labor, larger households were hired more frequently to cultivate coffee in the lean season and engaged for longer in harvesting. A major burden from slavery to the Black population in *Ibicaba* was the small number of economically-active household members. Finally, the paper traces the income profiles of various European immigrants and their descendants. German-speakers and Portuguese had labor income profiles closer to that of White Brazilians. The income profile of Italian immigrants and their descendants, however, shows evidence of households who performed rather poorly and probably failed to reach their "South American dreams".

ACKNOWLEDGMENTS

The current version of this working paper is based on four preliminary drafts (May 2023, August 2023, July 2024, and May 2025). The structure and results of previous versions are subsided by those of the current version (May 2026). Comments on these versions, presented at the following academic events, are gratefully acknowledged: Conference: "Resource-use & specialization in rural change in Europe and the Americas. Politics, reform & revolution" – UC Davis, May 26, 2023 (Davis, USA); UC Davis History Colloquium and Latin American History Workshop, April 26, 2023 (Davis, USA); UCLA Seminar in Latin American History, Advanced Topics Seminar (Prof. William Summerhill), May 16, 2023 (Los Angeles, USA); UCLA Library Council Internal Meeting, May 23, 2023 (Los Angeles, USA - online); UCLA's Proseminar in Economic History, May 30, 2023 (Los Angeles, USA); Workshop in Economic History of Brazil, FEA-USP (Organized by Prof. Renato Colistete), August 04, 2023 (São Paulo, Brazil); MEAP's Online Workshop: MEAP in South America: Digital Archives and Contemporary History, September 12, 2023 (online); International Conference Agricliometrics V, December 7-8 2023 (Montevideo, Uruguay); Internal seminar - University of Tuebingen, November 2023 - August 2024 (Tübingen, Germany); Seminar Series in Political Economy and Economic History, EESP/FGV, September 12, 2024 (São Paulo, Brazil); Seminar Series of the Economics Department, *Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto* (FEARP-USP), September 26, 2024 (Ribeirão Preto, Brazil); Graduate Minicourse at the Graduate Workshop of the *Brazilian Association of Researchers in Economic History* (ABPHE), October 1-2, 2024 (Ouro Preto, Brazil); Graduate Program Workshop from the Federal University of São Carlos (UFSCar), November 22, 2024 (Sorocaba campus - Brazil, online presentation); Graduate Colloquium of the Institute for Economic and Social History, University of Göttingen, April 23, 2025 (Göttingen, Germany); Outreach Program *Nacht des Wissens*, University of Göttingen, June 21, 2025 (Göttingen, Germany); Seminar Series *Hermes & Clio*, FEA-USP, October 01, 2025 (São Paulo, Brazil). I would like to specially thank all team members of the *Ibicaba Project*, the *Ibicaba farm*, *Paulo Masuti Levy Library*, and MEAP-UCLA for invaluable support in creating the physical archive (and digitizing it) that embeds the novelty of this paper. I am particularly indebted to comments by Jörg Baten; Renato Colistete; Sarah Ferber; Pedro Américo Ferreira; Leonardo Gardenal; Michela Giorcelli; Guilherme Grandi; André Lanza; and William Summerhill on previous versions of this paper.

DISCLAIMER

The core of this paper was written under a DAAD-PRIME Fellowship 2022-2023, while the author was a visiting scholar at the History Department, University of California, Los Angeles (UCLA) and Research Associate to the Chair of Economic History, University of Tübingen. Funding for the creation of the *Ibicaba Collection*, on which this paper is based: Arcadia Fund - Grant MEAP0104. Funding for the creation of the *Accounting for the Countryside Collection*: Arcadia Fund - Grant MEAP6154. All errors and ommissions are solely my own.

Legacies of Slavery or a Brave New World?
Labor productivity and remuneration in the Brazilian coffee economy – new microdata evidence from *Ibicaba* plantation (1888-1958)

Bruno Gabriel Witzel de Souza

I. Introduction

This paper studies the economic characteristics of emerging non-captive labor markets in the Brazilian coffee economy in the immediate post-Abolition era (1888), concomitantly to the consolidation of the state of São Paulo as a major destination to European mass immigration, mostly of subsidized Southern European households hired as agricultural field laborers to coffee plantations (1888-1927). This study investigates the short- and long-term performance of the various ethnolinguistic groups that composed the agricultural labor force of *Ibicaba*, one of the most important historical plantations of Brazil, located at the heart of São Paulo's coffee zones by the end of the nineteenth century. The empirical analysis shows the centrality of labor as a production factor, which strongly benefitted households with many economically-active members, as well as laborers' strong responsiveness to price incentives, as expected in that competitive agricultural labor market. I then proceed to test specific mechanisms of racial discrimination in those newly emerging non-captive labor markets, as hypothesized by the historical literature. Results do not support the hypothesis of purely taste-based discrimination. After controlling for households' quantity-based total factor productivity, their ethnolinguistic origins did not lead to higher (positive discrimination in favor of European immigrants and their descendants) or lower (negative discrimination against Non-Whites) prices earned for the same type of homogenous and unskilled labor task. In contrast, results support the hypothesis of statistical discrimination. Based on labor arrangements highly dependent on intrahousehold divisions of labor, larger households were hired to cultivate coffee in the lean season and engaged for longer in harvesting during the harvest season. A major burden from slavery to the Black population in *Ibicaba*, as representative of São Paulo's Old West, was the small size of their households (4 people per household in the pooled sample), in comparison to European immigrants, Italians in particular (7 people per household in the pooled sample).

It should be highlighted from the beginning that these conclusions are not revisionist at all. Discrimination and racism are integral components of Brazilian history, embedded as it was on slavery for almost four centuries. Notwithstanding, simplistic explanations in "black and white" of historical socioeconomic mobility will necessarily miss the actual processes of exclusion, especially in the post-Abolition era, when new mechanisms might have substituted for the previous explicit violence of captivity. Over the long-run, Asian and White Brazilians have reached higher average wages and access to high-paying jobs. Their positions in the labor market can even be ranked by specific ancestries (Balderas and Greenwood 2010; Monasterio 2018; Monasterio and Lopes 2018). The literature in economic history and on long-run development has solidly shown the effects of human capital accumulation for such an outcome (Carvalho Filho and Colistete 2010; Carvalho Filho and Monasterio 2012; Stolz,

Baten, and Botelho 2013; Monasterio 2018; Rocha, Ferraz, and Soares 2019; Witzel de Souza 2019). Other well-hypothesized but empirically not yet fully fleshed-out mechanism includes the scarcity of on-the-job skills brought in by Europeans and their upward occupational mobility upon quitting their first agricultural jobs and opening specialized workshops in urban centers, particularly around the most booming plantations (Dean 1976; Argollo Ferrão 1999; Bezerra 2001; Rocha, Ferraz, and Soares 2019; Witzel de Souza 2021). A third rich historiographical branch has explored historic landownership as a channel of upward mobility among European immigrants in Brazil. The evidence is that a substantial number of immigrants and their descendants did become small landowners, but these smaller properties remained economically in the fringes of the large plantations (Holloway 1980; Colistete and Lamounier 2014; Lanza 2021).[1]

These burgeoning research lines notwithstanding, few quantitative works have explored the economics of the first insertion channel of European immigrants into the Southeastern, coffee-producing, provinces of Brazil, namely as unskilled agricultural fieldhands.[2] First hired by the mid-nineteenth century as indentured laborers and then *en masse* during the death rattles of slavery in the late 1880s, these immigrants' initial contact with Brazilian labor markets was as low-ranks who feared for their future, given their economic (but not socio-juridical) proximity to the enslaved population; or had to compete in the post-Abolition era with freedmen, other Brazilians, and among themselves; or fell into a quite unexplored poverty trap of low mobility in geographic, occupational, and income terms over the twentieth century.

The scarcity of more empirical work on a problem that is economic in nature is most likely explained by the lack of adequate data, particularly micro-level, mostly high-frequency, evidence required for assessing labor arrangements and their economic outcomes. Similarly to other low- and middle-income countries, Brazil faces severe economic constraints and institutional hurdles to preserve its public historical heritage, including unique written sources (Palma 2019; Jenkins and Rubin 2024). The history of slavery at the micro-level was further hampered by the historic purposeful destruction of sources (Slenes 1983). Moreover, lack of adequate incentives and funding have severely limited the expansion of business archives that could have contributed precisely with such micro-evidence (*cf.* the rather contrasting positions of Reber 1985; and Frost, Verhoef, and Durepos 2026).

The new approaches to that question advanced in this paper were only made possible because of the creation of the new physical archive at *Ibicaba farm* and the digitalization of its historical records by the *Ibicaba Project* (Witzel de Souza and Santin Gardenal 2026).[3] That

[1] *Cf.* Lanza's (2021) exhaustive review of the historiography (classical and current) on the matter.
[2] This assertion is not to diminish the importance of the vast and rich historiography of Brazilian labor history, upon which this very paper is based (*cf.* Section II). The problem is rather the lack of new quantitative works testing well-embedded hypotheses put forward by that literature, some of which are mutually exclusive.
[3] It is certainly not coincidental that similar, but independent initiatives to collect Brazilian plantation records in the 1960s by Jeanne Berrance de Castro and José Sebastião Witter have led to a flourishing, but unfortunately discontinued, empirical literature in the 1970s, using precisely plantation accounting records (Bassanezi 1973; Faro Leal 1973). For the efforts of the former: Berrance de Castro and Witter (1964), Berrance de Castro (1965), and Berrance de Castro and Baiocco (1969); for a discussion of those temporally distant initiatives and the goals of the *Ibicaba Project*, *cf.* Witzel de Souza and Santin Gardenal (2026).

project was an initiative to safeguard and preserve *in situ Ibicaba*'s endangered historical records and to make them available in an open-source platform, in line with the project's umbrella program, the *Modern Endangered Archives* (MEAP), from the Library of the University of California, Los Angeles (UCLA).[4]

In addition to the empirical results discussed above, the very data compilation from the *Ibicaba Collection* has revealed important facets frequently ignored by the historiography on labor in Brazil. First, the labor arrangement that embedded the mass immigration of Italians (and, to a lesser degree, of Portuguese and Spaniards) to São Paulo's coffee economy was also largely employed with White Brazilians and Non-Whites, who were, in fact, the second and third most common ethnolinguistic groups with whom the so-called *colonato contracts* were signed for the pooled sample. Second, immigrant nationalities with older migratory histories to Brazil, the Germans and Portuguese in particular, showed income profiles very similar to the group classified as White Brazilians. This is indicative of completed acclimatization of the first generation of immigrants and a longer process of socioeconomic and cultural integration, including intergenerational. Finally, Italians (plus their Brazilian-born descendants) were the only group whose pooled income distribution showed evidence of bimodality. This indicates the existence of a subgroup that failed to reach the "South-American dream", about whom we know close to nothing in terms of subsequent migratory patterns (return, circular, local, or between Latin American countries) or, perhaps, downward economic mobility in the long-run.

This paper is a contribution to the overlapping fields of immigration and slavery history, under the broader umbrella of labor history. In the broadest sense, this is an empirical contribution on the emergence and consolidation of new agricultural labor markets in post-Abolition Brazil. By quantifying the demographic and economic characteristics of the labor force of a representative plantation of the Brazilian coffee economy at the critical historical juncture of Abolition, this paper dialogue directly with classical studies on the economics of slavery and post-slavery in Brazil (Carvalho de Mello 1978; Eisenberg 1983; Andrews 1988; Viotti da Costa 1998; *cf.* also the review in Monsma 2017).[5] Moreover, results dialogue here, in an international perspective, with the classical empirical literature on slavery, Abolition, and labor relations during Reconstruction in the U.S. In terms of slavery and post-slavery history, the data compiled for this paper and its analysis further advances recent empirical assessments of rural labor markets in Brazil, focused in explaining differential performances, either according to ethnolinguistic origins, or of labor arrangements (Cavalcanti Rodrigues 2021; Colistete 2024; Monsma 2010; Witzel de Souza 2012, 2024). In addition to the previous discussion on the distinct income profiles of German-speaking, Portuguese, and Italian immigrants in Brazil, this paper has the specific contribution of assessing the long-run economic performance of various ethnolinguistic

[4] Kuntz Busch (1967) and Witter (1971; 1982) conducted systematic studies of *Ibicaba*'s quantitative records, including sources used in this paper. In a distinct, but related initiative, Heflinger Jr. (2007, 2009, 2012) pioneered the publication of historical sources related to *Ibicaba* research not only in Brazil, but in European archives, as well.

[5] The paper also marginally contributed to the literature on plantation history, particularly via its compilation of *Ibicaba's* agricultural capital, including the location of all its *estacas* and the distribution of all coffee groves planted in the period 1888-1958 between all colonies of residence within the plantation.

groups. These results are an attempt at a more in-depth testing of hypotheses derived from a classical historiography on immigration and labor arrangements in the coffee economy. Finally, by tracking households over time and assessing the economic profile across generations, this paper dialogues directly with similar empirical efforts conducted for Argentina (in particular Sánchez-Alonso 2007, 2019; Pérez 2017).

Methodologically, this paper adds to the literature conducting new empirical research with historically sparse microdata. Accounting records have been amply recognized in the history of the discipline as an invaluable source of microdata. Recent studies have revived that tradition, particularly in using plantation records to understand the functioning of credit and personal indebtedness in markets with limited monetary supplies (Faleiros 2010; Saraiva and Freire 2016; Fornereto Mariano *et al.* 2022; Monti and Oliveira 2025). For Brazil, Palma and Lambais (2026) have recently used other types of historical accounting to provide new estimates of the country's GDP over the long-run. Beyond this very paper, the *Ibicaba Project* itself has spurred new research in the field (Rodrigues and McBride 2026; Witzel de Souza and Santin Gardenal 2026). To this effervescent literature, the methodological discussions of the data generating processes, in the plantation itself and for the compilation of the dataset that embeds my analysis, is of particular interest. It should be highlighted at this point that the *Ibicaba Collection* has been followed by the *Accounting for the Countryside Project*. This new initiative is currently scaling up the efforts to promote *in situ* patrimonial preservation and digitalization to five other collections speaking to the rural history Brazil: *Santa Gertrudes plantation* (divided in two different repositories, in Santa Gertrudes and Rio Claro - SP Counties); the historic small landholding and general storehouse *Casa Feltrin* (Itaqueri da Serra County); the historic government-sponsored rural settlement of Cascalho (Cordeirópolis County); and the archives of the old Agroforestry Station of Cia. Paulista Railroad, current *Floresta Estadual Edmundo Navarro de Andrade* (Rio Claro - SP County). In the future, the protocols on data transcription, cleaning, harmonization, and matching developed for this paper should apply without much cost to the records of these new collections.

The paper is organized as follows. Section II sketches the history of the coffee economy in Brazil, linking it to the types of labor relations most prevalent in the coffee zones. The section highlights the labor market relationships between slave labor and the experiments conducted with indentured European laborers from the 1850s until the consolidation of *colonato* contracts in the 1880s - the most prevalent arrangement during the time covered by this paper. The section concludes with a review of hypotheses on forms of discrimination and the economic mechanisms behind them. Section III presents the data compiled for this paper. It details the processes of transcription, cleaning, and matching, focusing in particular on the nature of historical accounting data and its usage in economic history. Beyond its immediate goal for the rest of the paper, this section aims at providing protocols on data curation for future research using rural historical accounting. Section IV presents the descriptive statistics. It discusses in detail the demographic and economic characteristics of *Ibicaba*'s laborers, as well as the agricultural capital made available to them by plantation managers. Section IV speaks directly to Section II, as the macroeconomic policies discussed in the former section

can be clearly visualized in the microeconomic responses given at the plantation level. Section V discusses the empirical results. It first estimates the determinants of physical output in coffee production, separately for coffee cultivation in the lean and for harvesting in the harvest seasons. Based on these estimates, the income profiles of every ethnolinguistic group in *Ibicaba*'s labor force is estimated, which is then decomposed by the characteristics of each group using the Oaxaca-Blinder twofold decomposition and an analysis of simple differences-in-means tests (which do not require the more strict identification hypotheses of the OB decomposition). Finally, the determinants of price differentials among ethnolinguistic groups is estimated after controlling for households' TFPs. Section VI looks into some specific labor market dynamics in the immediate post-Abolition (1888-1890) by exploring the data of Ledger *Biblioteca Paulo Masuti Levy #0004*,[6] the last accounting registry conducted by the Vergueiro era in *Ibicaba*, before its foreclosure and public auctioning. This section studies the demographic, occupational, and economic characteristics of *Ibicaba*'s labor force from 1888 to 1890, estimating the likelihood of different ethnolinguistic groups to sign a *colonato* contract, be allocated to different colonies of work, and receive coffee trees of varying agronomical conditions. Section VII conducts a battery of robustness checks. I check, in particular, the sensitivity of the results to any choice made during the process of data compilation, as well as to distinct manners to classify the ethnolinguistic origins of households. Section VIII concludes.

II. Historical Context: Coffee Production and Labor Markets in Southeastern Brazil

II.I. From Slavery to Abolition, from European Indenturership to Mass Migration

Smuggled into Northern Brazil by the end of the colonial era, coffee was soon thereafter transplanted to Rio de Janeiro, the capital of the United Kingdom of Portugal, Brazil, and Algarve, in 1815. First cultivated as an exquisite beverage in villas and courtyards in the cities of Rio de Janeiro and São Paulo, the plant adapted astonishingly well to the climate and hilly topography of Rio de Janeiro's outskirts (Simonsen 1978, pp. 175-7; Stein 1986). What started as a curious stimulant appreciated by few courtesans and local elites grew to represent *ca.* 20% of all Brazilian exports already in the 1820s (Simonsen 1978, pp. 192 ff.; Alden 2004, p. 330). Stimulated by a growing international demand, coffee production spread rapidly in the province of Rio de Janeiro until the third quarter of the nineteenth century. Production grew extensively. The incorporation of agricultural land in newly opened agricultural frontiers, especially alongside the southern axis of *Paraíba River Valley*, was worked by the mass importation of slaves, either directly from Africa or from economically declining regions (Preston 1932; Simonsen 1978; Stein 1986). In the province of São Paulo, coffee production commenced with the natural extension of those plantations growing into the southern portions of the *Paraíba River Valley*. By the late 1830s, the volume of coffee exported had already surpassed that of sugarcane, São Paulo's leading agricultural cash crop since the late colonial era (Simonsen 1978, Chap. 3).

[6] For brevity, ledgers identified by prefix *Biblioteca Paulo Masuti Levy* are henceforth referred to as "BPML".

Coffee was introduced experimentally to the vast hinterlands of São Paulo's central-western plateau in the late 1820s, but sugarcane continued to dominate agricultural production in the region that later became known as São Paulo's Old West (Simonsen 1978, Chap. 3). By the mid-nineteenth century, however, growing international demand and excellent agroecological conditions for coffee cultivation in those agricultural frontiers stimulated a boom in coffee production (Preston 1932; Dean 1976).

The expansion of agricultural frontiers in São Paulo faced two main constraints, which became a constant problem deep into the twentieth century: transportation infrastructure and a secure and elastic supply of labor. A region of only marginal economic relevance to the Portuguese Crown at the end of the colonial era, São Paulo's countryside had a comparatively small stock of slaves by the mid-nineteenth century (*cf.* Witzel de Souza 2024, p. 617 and references therein). In addition, the transatlantic slave traffic had been abolished in 1850, after a half-century of entrenched domestic and international disputes, including diplomatic and military tensions with Great Britain and the Admiralty in the recently independent Brazilian Empire. In Rio Claro (SP) Country, which neighbored *Ibicaba*, the average price of a prime-field hand (men between 15 and 29 years old) increased from 505 *milréis* in 1843-9 to 1,222 *milréis* in 1850-9, before plateauing at the high level of 1,800+ *milréis* in the 1860s and 1870s.[7]

Within this context of increasing demand for labor in expanding agricultural frontiers and the complete curtailing of what had been the main supply of laborers to Brazilian agriculture for more than three centuries, *Ibicaba* rose to domestic and international spotlights. As of 1840, Senator Nicolau Pereira de Campos Vergueiro started experimenting with the employment of non-captive laborers in that plantation of his (Buarque de Holanda 1940; Witter 1982; Heflinger Jr. 2007: Mendes 2017). Building upon his vast experience in policymaking related to immigration, which dated back to the 1820s, Senator Vergueiro promoted the private hiring of immigrant laborers to coffee plantations via *Vergueiro & Co.*, a family corporation in the business of international coffee factoring, exports, and immigration (Witter 1982).

Vergueiro & Co.'s contractual arrangement with European indentured immigrants was contemporaneously known as the *Vergueiro system*, being commonly referred to as a "sharecropping contract" by the historiography. The latter nomenclature is somewhat illusive, since the *Vergueiro system* interlocked three arrangements into one single contract, as I have thoroughly discussed elsewhere (Witzel de Souza 2011, 2024). The *Vergueiro system* is better described as a sharecropping-*cum*-debt arrangement. Its main dimension was a labor contract based on shares: from a household's annual produce, the net profit was divided between the household, who provided labor services in coffee cultivation and harvesting, and *Vergueiro & Co.*, which provided agricultural capital and marketed the produce. The second dimension was a land rental agreement that allowed households to cultivate subsistence goods in the plantation. Self-consumption was exempted from sharing, but households were expected to

[7] Complementary data for Rio de Janeiro in Carvalho de Mello (1978); and for Minas Gerais in Luna and Klein (2010).

distribute fifty percent of the profits from crops sold to third parties.[8] The third component of the arrangement was a credit mechanism, by which *Vergueiro & Co.* advanced loans to impoverished immigrants for them to cover the costs of transportation and settling in the Brazilian plantations. Immigrants, in turn, agreed to tie the labor of all their household members to the repayment of their outstanding debts with *Vergueiro & Co*. Alternative contracts at the time and in most of the second-half of the nineteenth century substituted that mechanism by a fixed-term indentureship.

On the one hand, the *Vergueiro system* inserted São Paulo into the map of European immigration by attracting poor households who did not have the means to migrate to economically, politically, and religiously more promising destinations, particularly the U.S., the Southern Cone, or even settlement colonies in Brazil's southernmost provinces (Levy 1974; Holloway 1980; Sánchez-Alonso 2019; Witzel de Souza 2024). Variations of this indentureship scheme allowed were responsible for a large share of the *ca.* 7,000 Europeans who immigrated to São Paulo between 1840 and 1880. On the other hand, immigration schemes based on indenturership are prone to conflict. The potential for escalation was severe in a country where labor relations were still dominated by slavery and where non-captive labor markets were mostly characterized by patronage. Unsurprisingly, a number of labor strikes occurred in the 1850s and 1860s, the most prominent of which in *Ibicaba* itself, in what became known as *Sharecroppers' Rebellion* (Davatz [1858] 1941; Tschudi [1866] 1980; Buarque de Holanda 1940; Heflinger 2009).

The effects of such labor movements and their repercussion in Europe on immigration flows to Brazil is contested in the historiography (Kamphoefner 2000; Pérez Meléndez 2024).[9] Their impact on the willingness of planters to employ European immigrants tends to be based on exaggerated accounts of landowners who reported dissatisfaction with laborers demanding rights beyond the farmgate, such as to consular authorities. What is certain is that planters who were able to afford the high prices of prime-field hands kept relying on slave labor (Carvalho de Mello 1978; Dean 1976; Viotti da Costa 1998). Given the profitability of southeastern plantations, slave labor was first shifted from other crops and tasks into coffee production, and then increasingly imported from other Brazilian provinces into Rio de Janeiro's and São Paulo's coffee zones (Carvalho de Mello 1978; Klein and Luna 2010; Dean 1976, pp. 127, 135). As of the 1850s, São Paulo's coffee zones concentrated the largest shares of European immigrants and of black slaves in the province (Dean 1976, p. 51; *cf.* Witzel de Souza 2024, p. 617).

Immigration figures in the 1840s-1860s are very small compared to European inflows to destinations even in the Southern Cone, to Brazil itself during the Age of Mass Migration, or, most dramatically, to the enslaved Black population trafficked into the coffee zones (Levy

[8] Similar clauses can be found in many contracts by the mid-nineteenth century, but were probably not enforced due to very high monitoring costs. In a non-systematic reading of records from the period in the *Ibicaba Collection*, I found no evidence that plantation managers ever enforced such a clause.

[9] Prussia (and in 1871, the German Empire), Italy, and Spain passed legislation curbing or limiting emigration to Brazilian coffee zones in 1859, 1902, and 1911, respectively. In the context of the *Age of Mass Migration*, these countries prohibited subsidized immigration to Brazil, having the subsidized immigration program from the state of São Paulo as its main target (*cf.* Dean 1976, pp. 181-4; Pérez Meléndez 2024).

1974; Luna and Klein 2010; Sánchez-Alonso 2019). Institutionally, however, these early migratory waves prepared the grounds for the Age of Mass Migration in Brazil. Institutionally, however, these early migratory waves prepared the grounds for the Age of Mass Migration in Brazil. The genesis of the most prevalent and enduring labor arrangement in the Brazilian coffee economy, the *colonato contract*, can be traced back to the historical unfoldings of clauses first elaborated by the *Vergueiro system.*[10] Moreover, the subsidization of immigrant laborers was a mechanism to promote immigration without dealing with institutional reforms to make the country more attractive to the pool of European immigrants during the Age of Mass Migration (Witzel de Souza 2024). Finally, by the 1880s, São Paulo's economic and political elites were very experienced in dealing with diverse organizations and designing public contracts to hire, transport, allocate, and manage immigrant laborers.[11]

In the 1860s and 1870s, three trends emerged in rural, non-captive labor markets. First, there was a spread in the arrangements based on daily tasks (*jornais* and *diaristas*) and piecerates or hires for a fixed period (*camaradas*) (Dean 1976, pp. 20, 45; Lamounier 1986; Viotti da Costa 1998). Second, contracts based on crop-sharing clauses kept being proposed, frequently by economic and political potentates who experimented with the *Vergueiro system*. Third, contracts became better specified and substituted crop-sharing clauses with predetermined payments for specific tasks, particularly in the lean season, thus reducing laborers' uncertainties throughout the agricultural year (*cf.* Stolcke and Hall 1983; Stolcke 1988; Faleiros 2010; Witzel de Souza 2024). All these arrangements frequently included labor tying mechanisms, as private credit was still advanced for the hiring of European immigrants, particularly in the German States, Imperial Germany, Switzerland, and Portugal (Witzel de Souza 2024).

Brazil underwent some of the most significant institutional and economic changes of its history between 1886 and 1891. The final Abolition of slavery was promulgated in 1888 after mounting pressure from the abolitionist movement and of the enslaved population, whose threats to the country's economic and political order became pressing as of 1886, with mass flees, more violent strikes, and recurrence to manumission legislation (Dean 1976, pp. 144-7). Concomitantly, Brazil entered the circuits of the Age of Mass Migration in a decisive manner (Dean 1976, pp. 156-8). After passing clearer pro-immigration legislation as of 1884, in 1886 the provincial government finally put in place a centralized organizational framework for hiring, transporting, and allocating Europeans, particularly Italians, to coffee plantations that kept expanding rapidly into western agricultural frontiers (James 1932; Monbeig 1984; *cf.* further the literature review in Holloway 1978, 1980). The official subsidization of immigration to coffee plantations spanned from 1886 to 1927. Between 1881 and 1921 *ca.*

[10] The *Vergueiro system*, in turn, had precedents in Portuguese sharecropping arrangements, very much aligned with similar experiences other Southern European countries (Câmara 2006; *cf.* Carmona and Simpson 1999; Garrido and Calatayud 2011; Garrido 2017 for some parallels).

[11] This does not imply a linear transition towards most efficient labor arrangements, as defended by a classical literature on post-abolition modernization. Rather, similar to other economies under pressure for institutional changes due to abolitionism, non-captive labor markets in São Paulo were equally characterized by a complex mix of labor arrangements throughout the second-half of the nineteenth century, including the earliest proposals for the usage of sharecropping arrangements that go as far back as the 1830s (*cf.* Witzel de Souza 2024).

3.2 million (gross immigration) individuals immigrated to Brazil, 57% of whom went to the province/state of São Paulo (Levy 1974; data summarized in Witzel de Souza 2023, pp. 309).

II.II. Rural labor markets in the post-abolition era: Some hypotheses from the literature

The literature on the economic history of post-Abolition Brazil hypothesizes that rural labor markets fragmented within lines of skin color or "race". In the province of São Paulo, mass immigration of Southern Europeans would have substituted Black field-hands (Dean 1976, p. 195). The freedmen would have been relegated to inferior contracts, while immigrant families, particularly Italians, would have signed the economically most advantageous *colonato contracts*.

After a short period of attrition in the newly created non-captive rural labor markets, the freedmen would have returned to the plantations as rural laborers, but relegated to more precarious positions in terms of job security and pay. Dean (1976, p. 152) argues that the hiring of *camaradas*, for instance, was explicitly conditional on the supply of immigrants. Moreover, the literature posits that Black rural laborers were discriminated against in pay for their skin color and origins. Dean (1976, p. 154) even argues that "planters evidently thought the work of the freedmen was less satisfactory than that of the Europeans [...] a false assumption, hence mere prejudice". Black laborers were then relegated to lower paying jobs and to shorter-term and more precarious positions, which implied "[...] a nearly nomadic existence for most males and the near impossibility of forming nuclear families" (Dean 1976, p. 153).

According to the literature, immigrants hired under *colonato contracts* would be on the opposite side of the precariousness faced by Black freedmen (*cf.* Bassanezi 2019). Under this arrangement, entire households were hired for a predetermined period, which, in the case of immigrants, was expected to compensate landowners privately for the public subsidies that immigrants had received. Similarly to the *Vergueiro system,* a typical *colonato* contract had three main components. It was primarily a labor arrangement with two sources of labor income and, for those immigrants arrived with public subsidies, it was further a labor-tying mechanism for a prespecified period, *i.e.* an implicit market for credit received abroad. During the lean season, households contracted the cultivation of a fixed number of coffee groves, for which they were paid a prespecified sum. During the harvest season, households were summoned to harvest the berries and earned a varying income according to their productivity. Throughout the year, *colonos* were allowed to cultivate subsistence goods (Dean 1976; Holloway 1980; Basanezi 2019).

Income earned in coffee cultivation during the lean season was risk-free to laborers. At the beginning of the contract, each household negotiated a fixed number of coffee groves to cultivate, usually with the option of renegotiating that amount in subsequent agricultural seasons. Plantation managers assigned households to so-called *colônias* (literally, "colonies"), where they worked and resided.[12] Within each *colônia*, households cultivated coffee groves in

[12] Housing and rent conditions varied between plantations.

specific *estacas* (literally, "pickets"), corresponding to the smallest unit of agricultural planning, usually containing multiples of 500 coffee bushes. These coffee groves needed to be appropriately taken care of via hoeing, cleaning, weeding, and pruning, which was literally how this contractual component was called (*carpa*). The frequency of these tasks was determined by plantation managers and stated clearly in the contracts. In *Ibicaba* during the period considered in this paper, five to six prunings were executed per agricultural season.[13]

Income from harvesting was a function of the volume harvested by the household. This type of income was very volatile. The volume harvested in consecutive years could vary by a factor larger than 100% due to agroecological vagrancies, oscillations in the natural productivity of the bushes, the type and quality of the coffee bushes, and the location of the groves between *colônias* and *estacas* therein. We do not know exactly how the harvest was managed on the ground. According to classical historiography, overseers brought the available labor force in a plantation, including *colonos* to those colonies in the plantation with the most mature berries (Holloway 1980; Bassanezi 2019). Under this interpretation, *colono* households did not necessarily harvest the same groves they had cultivated in the lean season. Sources from the *Ibicaba Collection* support this interpretation because of the exact notes on the days when the harvest season had started and ended; and the pooling of all workers, separated only by labor categories, but not by *estacas* in Paysheets referring to the harvests . If harvests were indeed managed that way, then the *colonato* differed from the *Vergueiro system* in that regard. Under the *Vergueiro system,* households harvested the same coffee groves they had cultivated in the lean season (Davatz [1858] 1941; Witzel de Souza 2012, 2024). If laborers expected a positive relationship between higher yields and better cultivated plants, then the *Vergueiro system* would have been superior in terms of labor effort during the lean season. Qualitative evidence from interviews with *Ibicaba*'s ex-*colonos*, dating to the 1950s, seem to support more such interpretation (*cf.* Claudio Pott and Luiz Tonelotto in Witzel de Souza and Gardenal 2021).[14]

The *colonato* contract also allowed for the production of subsistence goods and the maintenance of private cattle for meat and dairy. Depending on the age and productivity of the coffee bushes, *colono* households were allowed to cultivate subsistence goods between the coffee rows, or in separate fields. Laborers strongly preferred the former, as it optimized the intra-household allocation of labor. The classical historiography has argued that these non-monetary incentives played a central role in determining labor allocation between older coffee zones and those located at the agricultural frontiers, where soil quality had not yet

[13] At the 95th percentile, some households executed 7 prunings per year, with 7.67 (averaged within household) as the maximum. Some of these are upper bound estimates, derived as the ratio between total labor income during the lean season and the number of income installments paid to laborers during one agricultural season.

[14] Luis Tonelotto worked in *Ibicaba* only in the later period covered in this paper, but he recalls that his family had worked as coffee (and cotton) *colonos*: "[… We] harvested and earned per coffee bag. We spoke of '*alqueire*' back then. One *alqueire* equaled 50 liters. I can't remember anymore how much we earned, but the plantation owners paid a certain amount for each *alqueire*. So we harvested coffee; [the yield] was then weighted and written down". It is implicit that the family harvested those coffee groves they had cultivated in the lean season: "[…]. My father earned a certain amount for each thousand coffee groves we cultivated. Every month we earned some little money to take care of the groves. But, as I said, there was no such thing as a fixed wage like in [an urban] firm. That was the custom of the time, right? Later on, in the harvest season, we harvested and the owners paid a certain amount for each bag we harvested" (Tonelotto in Witzel de Souza and Santin Gardenal 2021, pp. 355-6).

been depleted and could, therefore, accommodate non-cash crops between newly planted coffee rows. Thomas Holloway (1980) has estimated that up to a third of the total income of *colono* households accrued to the self-consumption or commercialization of such crops.

II.III. Coffee markets with state intervention & the long-term decline of *Colonato*

In 1889, a military coup supplanted the Brazilian centralist Monarchy by a federalist Republic that gave much autonomy to the states, including the prerogative to contract international debt. Among other causes, the Abolition had removed the last pillars of support to the Monarchy. In this new constellation, São Paulo took up the most prominent economic position in the federation. Coffee production in São Paulo had surpassed that of Rio de Janeiro for the first time in 1886 and grew steadily thereafter (Simonsen 1978, Chapter 3).[15] Meanwhile, the coffee economy in Rio de Janeiro, which was even more reliant on slave labor than São Paulo's, saw its demise with Abolition. The loss of labor and capital was aggravated by high indebtedness among landowners and by declining productivity in lands exhausted by decades-long exploitation without investments in technology or agronomical improvements (Simonsen 1978; Delfim Netto 1979; Stein 1986).

Paradoxically as it may seem for the plantation that had inaugurated the scheme of European indenturership in the Brazilian coffee economy, the *Vergueiros*' era in *Ibicaba* did not survive Abolition. Highly indebted since the mid-1860s, *Vergueiro & Co.* faced a series of legal cases as of 1888, which finally led to the public seizure and auctioning of *Ibicaba* in 1889/90 (Heflinger Jr. 2021; Tamiazo 2021; Witzel de Souza and Santin Gardenal 2026). *Vergueiro & Co.* had, in fact, invested very heavily in slaves. The firm was the largest proprietor of captives in Limeira County by the late 1870s, whose total value was almost equivalent to the amount of capital that the firm had invested in farmlands (Marcondes 2021). The winning bid for *Ibicaba* was placed by a consortium led by brothers Simão and José Levy, who became its sole proprietors in 1896. As one of the most astonishing cases of upward mobility among immigrants in Brazilian history, Simão and José had arrived with their siblings and parents as indentured laborers in *Ibicaba* in 1857. That household, headed by Jacob Levy, worked for fourteen years in *Ibicaba*, before settling down in the neighboring town of Limeira (Freitas Levy [1993] 2023; Tamiazo 2021; Witzel de Souza and Santin Gardenal 2026).

During the Brazilian First Republic (1889-1930), Coffee was King. As of the 1890s, that product alone was always beyond 50% of all Brazilian exports, such that macroeconomic policies became highly dependent on public revenues from coffee exports (Delfim Netto 1979). Growth in coffee production was extensive in nature, with the continuous expansion of agricultural frontiers to the westernmost regions of the state, made possible by the expansion of railroads and mass immigration, whose increasing costs were compensated by agroecological conditions that were ideal for coffee cultivation, especially the type of virgin soil found in the frontier regions (James 1932; Monbeig 1984).

[15] In Rio Claro (SP) County, coffee remained the dominant crop from the 1850s to the 1930s, reaching its peak in volume harvested in 1901 (Dean 1976, p. 38).

As the twentieth century unfolded, the New and Newest Westen frontiers of São Paulo gradually shadowed the Old West in terms of agricultural productivity (Monbeig 1984). The volume of coffee harvested in Rio Claro (SP) County, for instance, peaked in the 1910s (Dean 1976). Coffee remained, by far, São Paulo's main cash crop deep into the mid-twentieth century. However, some counties started diversifying agricultural production. Most turned back to sugarcane, São Paulo's late colonial cash crop *par excellence*. Over the long-run, Limeira County, where *Ibicaba* was located, further invested heavily in citrics (Pissinato and Vian 2021). As argued in this paper, in spite of its historical fame, *Ibicaba* had become a typical plantation of São Paulo's Old West as the Levys' era commenced in 1890.

Since the 1890s, coffee planters grew increasingly preoccupied with the ability of international markets to absorb, at stable prices in Brazilian domestic currency, the increasing supply of coffee. The era of "coffee markets under state intervention" began in 1906, when a coalition of the main coffee producing states of Southeastern Brazil - São Paulo, Minas Gerais, and Rio de Janeiro - captained the first intervention in the market to keep prices in domestic currency artificially high for Brazilian producers (Delfim Netto 1979). The strategy was to maintain stocks that were controllably supplied without disturbing price levels considered adequate by domestic producers (Delfim Netto 1979, pp. 47-82). Similar operations were conducted in 1917 and 1921, under the more direct leadership of São Paulo. Their short-term successes, helped by fortuitous drops in domestic production in the years following the operations, the scale of financial resources required, and their political and macroeconomic implications led to the federalization of these efforts in 1924, becoming a permanent policy. In spite of legislation to the contrary, for almost a quarter century these policies had the side effect of stimulating further the expansion of coffee production – geographically and within plantations –, thus worsening the long-term effects of oversupply (Defilm Netto 1979, pp. 143-58).

Due to plummeting international demand, contracting global finance, and severe domestic macroeconomic constraints, the "coffee-defense policies" became untenable as the effects of the 1929 Crash hit Brazil. The political coalitions supporting the First Republic were shaken and, in 1930, a *coup d'état* installed a populist regime turned into a dictatorship from 1937 to 1945. In spite of policies to address new concerns of emerging industrialists, urban middle-classes, and low-rank military officers, the new government could not afford ignoring the "coffee problem". As such, emergency measures were kept in place to support coffee planters. In Brazil the economic contraction and loss of wealth implied by the Great Depression were literally visible with the federal program to buy and burn coffee sacks in attempts to keep producer prices from falling further.

New governmental programs at the federal and state level attempted further to increase the value added of Brazilian coffee, particularly as of the 1950s. Agronomical programs proposed improvements in cultivation techniques, the recuperation of exhausted soils, and increasing the quality of Brazilian coffee in order to compete in different segments of the global market (*cf.* Delfim Netto 1979). Simultaneously, producers were literally paid to eradicate older, less productive, and inferior coffee groves (Stolcke 1988). One of *Ibicaba*'s

last residing workers recalls watching as a child, in the 1950s, how his father went out to the field to cut-off coffee bushes (Olivatto in Witzel de Souza and Santin Gardenal 2021).

Colonato contracts persisted as long as coffee remained the plantations' main cash crop. Its long-term demise had already begun, but variations of the original arrangement subsisted throughout the crisis years of the World Wars and the Great Depression. Verena Stolcke (1988) has argued that *colonato* arrangements were able to buffer falling nominal wages and inflationary pressures with clauses regulating the self-consumption of subsistence crops. However, as of the 1950s, profound structural changes led to the final demise of *colonato*. The diversification of cash crops implied varying demands for labor in different periods of the year. The flexibility to hire daily field hands and the ease to monitor them per task surpassed the benefits of maintaining a residing labor force devoted to a single crop. In addition, sugarcane was supplied by plantations to newly installed, frequently monopsonistic, sugarmills. The latter either provided labor gangs, or supported the gathering of harvests completed on a daily basis. Wage laborers, including informal daily field hands, substituted the *colonos* (Stolcke 1988).[16] To a large degree, an elastic supply of unskilled rural laborers was maintained with migrants from other Brazilian states, particularly from impoverished Northeastern counties.

The historiographical description of rural labor markets in São Paulo's coffee economy since Abolition aligns well with abundant historical and sociological evidence demonstrating the covert racism that has prevailed in Brazilian history up to present days. That fact notwithstanding, we still know very little about the paths of socioeconomic mobility and the economic mechanisms underlying differential economic performance between various ethnolinguistic groups that have worked in one and the same rural environment since the 1850s, albeit under abysmal distinctions in their legal-political statuses and numbers.

Even if convoluted under the general umbrella of racial discrimination, three distinct mechanisms have been hypothesized by the literature in regard to the insertion of (or lack thereof) freedmen and Black workers in São Paulo's rural labor markets. The first would be explicit racial barriers to entry. Implicitly assuming an elastic supply of immigrant laborers for plantations across the state of São Paulo since the 1880s, this hypothesis posits a classical case of taste-based discrimination, à Becker (1957). The second would be discrimination in pay for equivalent levels of productivity in the same occupation. Based on racist premises, landowners would have had negative priors about productivity levels of the Black population *vis-à-vis* European immigrants, a clear case of statistical discrimination à Phelps (1972). Finally, in the immediate post-Abolition era, barriers to entry and low pay would have forced young freedmen to migrate constantly between insecure jobs between plantations. This deleterious "nomadic lifestyle" seriously impacted the constitution of nuclear families and, consequently, negatively affected the demographics of the Black population. Because landowners preferred to hire large households for *colonato contracts* (due to fewer

[16] "Bóia-frias" (literally, "cold lunchboxes") were agricultural field hands, usually hired informally and from the poorest social rankings, a large number of whom were migrants from Brazilian North-eastern states.

transaction costs and more efficient intrahousehold division of labor), smaller Black families would have faced a recurring dynamic disadvantage à Arrow (1973).

In parallel, European immigrants would have benefited from positive racial discrimination. This would have allowed them to substitute the Black population in the post-Abolition era as agricultural laborers and to have access to economically advantageous contractual arrangements, characterized by high incomes in booming years and buffering mechanisms against downswings. To reiterate, this paper is far from questioning the many legal-political advantages of European immigrants in Brazilian history. That fact notwithstanding, we still have very limited knowledge about the actual paths of socioeconomic mobility among immigrants, particularly those who arrived in the country under different forms of indentureship and in distinct geopolitical periods. Paralleling similar trends in North and South America, the historiography on immigration to Brazil has shown how the tightening of immigration policies correlate with the hardening of nationalistic identitarian movements and increasing xenophobia since the 1890s, exacerbated by World War I and, in Brazil, reaching its apex with a nationalistic dictatorship in the 1930s. In short, the welcomeness to foreigners changed over time and might be reflected in discrimination in labor markets.

The historical evidence on the priors of landowners in Brazil about the productivity of Black laborers, enslaved or free, is also more ambiguous than usually posited by the literature. Racist raptures abound in public discussions (in the press and political) about the productivity of Black slaves and European immigrants. Nonetheless, the revealed preference of most landowners was for slavery until its very last bitter end. Particularly in the 1850s, in official communications with the provincial government, landowners positively ranked various European nationalities as hardworking and orderly, but also complained about their far too demanding and recalcitrant behavior, features in fact most likely related to the immigrants's socioeconomic statuses than to their origins. Moreover, discontented landowners complained about the moral qualities of immigrants, although we have no evidence that European indentured laborers in the coffee zones differed in any substantial way to their counterparts in other countries of the Americas. Finally, with the turmoil of Abolition subsiding, statistical discrimination could even have favored ex-enslaved workers, given their experience in cultivating coffee in contrast to recently arrived Europeans, who either did not have rural backgrounds, or were experienced with crops and agricultural techniques very distinct to those of Brazilian plantations.

III. Data: Transforming Historical Labor Accounting

III.I. Selected sources from the *Ibicaba Collection*: Harvest Ledgers & Registries

The accounting and managerial ledgers of the *Ibicaba Collection* can be generally classified into two main categories: those referring primarily to the plantation's managerial practices, including other farms and private businesses of the Levy family, and those primarily concerned with laborers' accounting. Obviously, this is a crude categorization, since most ledgers included more than one type of accounting and all managerial ledgers necessarily referred to labor costs (Witzel de Souza and Santin Gardenal 2026).[1] In fact, some managerial operations were indistinguishable from labor management, *e.g.* the allocation of various labor categories to daily piece-rates executed in the plantation's mills, as registered in *Daybooks* and *Timesheets*. Notwithstanding, such a categorization of historical ledgers allows us to distinguish those documents most directly concerned with laborers' output and remuneration.

This paper selected 16 ledgers of the *Ibicaba Collection* directly related to labor accounting, namely 7 "Coffee harvest registries" (henceforth *Harvest ledgers*) and 9 "Registries of coffee trees assigned to laborers" (henceforth *Registries*). This section documents the processes of data transcription, cleaning, and harmonization, as well as of matching different sources and accounting ledgers of varying contents and quality. Beyond thoroughly explaining the origins and meaning of the variables used in the empirical exercises, this section aims at providing a functioning and well-tested protocol on how to transform historic accounting and managerial data from Brazilian plantations into innovative sources for scholarship in economic history, thus complementing the protocols on material and digital preservation I discuss elsewhere (jointly with Santin Gardenal 2026).

Harvest ledgers recorded the amount of coffee that each household harvested per day during the harvest season. They were usually organized as current accounts under the name of a household head, to whom the total volume of coffee harvested per day was registered.[2] Panel A of Figure 01 illustrates a typical entry in such a ledger. Here formatted as current accounts under the heading of a household head (A), these ledgers registered the days a household worked in harvesting (B), the amount of coffee harvested per day (C) and in total (D), and corresponding prices, which were either explicitly recorded, or, as in this example, can be retrieved from total output (D) and the ensuing labor income (E). In this paper, I aggregated the data per harvest season, computing the average price and nominal labor income earned by each household in harvesting coffee.

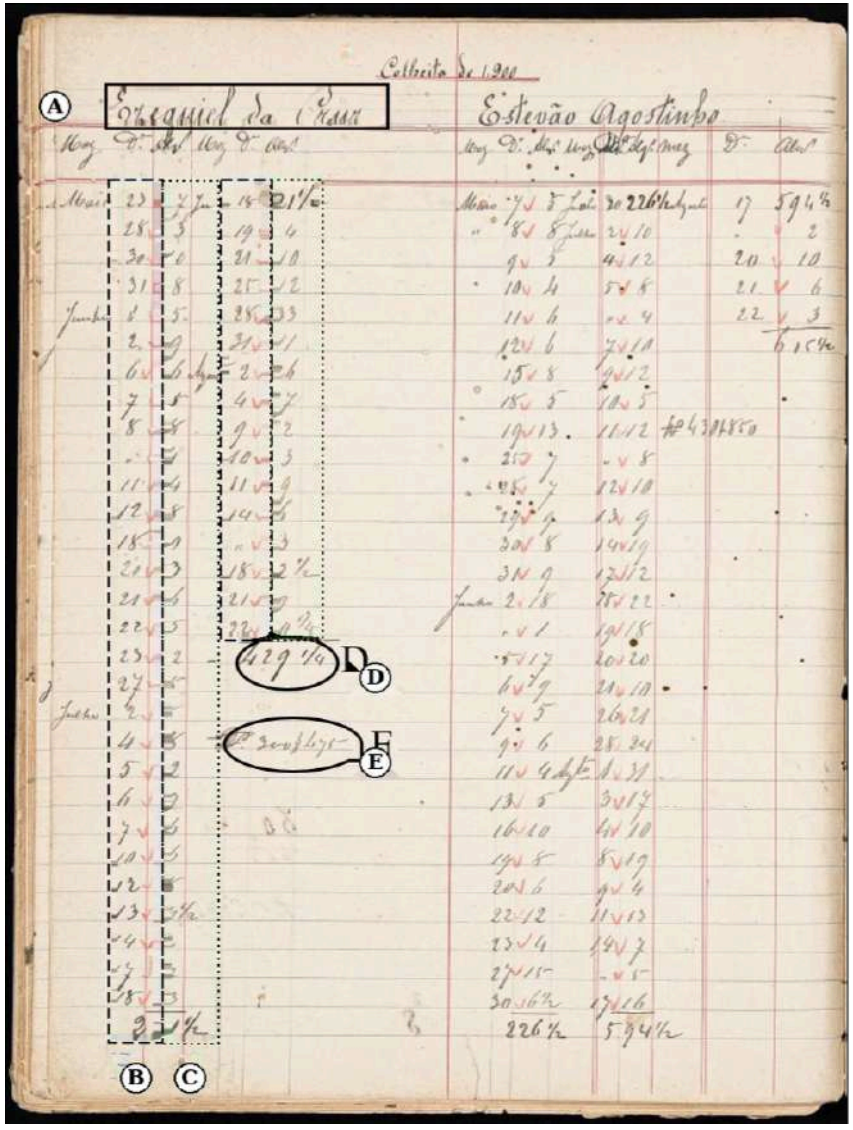
---

[1] These ledgers can be readily identified by the double categories used for their titles in the *Ibicaba Collection*. Some refer to the same type of accounting, but for different subjects (*e.g. Paulo M. Levy's Library #0001 – Checking account from Vergueiro & Co. & Workers' checking accounts, Ibicaba farm (1846-1850)*); others refer to different types of accounting, but for the same subject (*e.g. Ledger #0001 – Workers' checking accounts & Laborers' payroll, Ibicaba farm (1889-1892)*); and yet others refer to different types of accounting for different subjects (*e.g. Ledger #0019 – **Workers'** checking accounts & Current account (Accounting book of customers and commercial partners), **Ibicaba farm** (1899-1912)*). For a thorough discussion of types of ledgers and accountings in the collection, *cf.* Witzel de Souza and Santin Gardenal (2026).

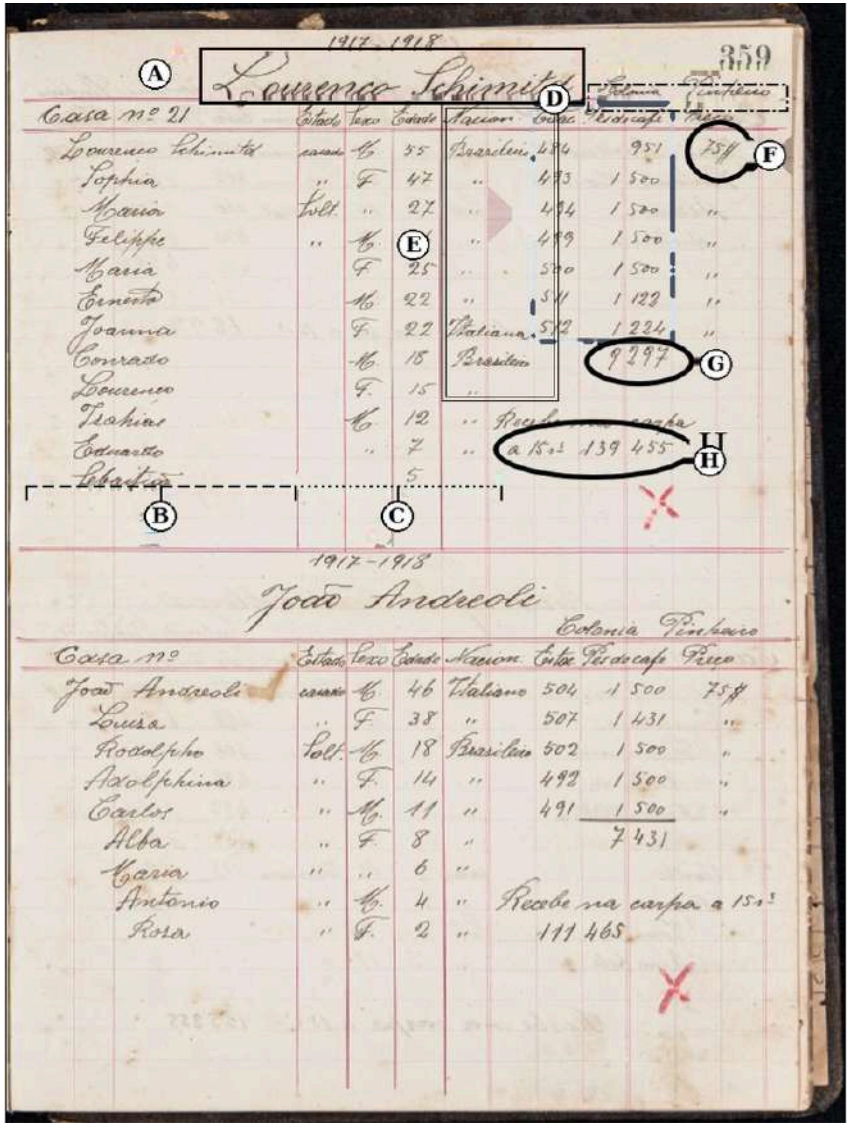
[2] Almost all entries are in *alqueires*, which here refer to a unit of volume corresponding to *ca.* 40L of dry goods. All other units of volume were converted into *alqueires* whenever necessary.

Figure 01 – Sources: Annotated *Harvest Ledger* & *Registries*

Panel A – Ledger #0013 (Digital Image 144)



Panel B – Ledger #0043 (Digital Image 181)



Registries, in turn, are population counts of residing laborers. They further contain detailed data on the agricultural capital assigned by plantation managers to *colono* households and somewhat more erratic data on housing. Panel B of Figure 01 illustrates a typical entry in such a ledger. Formatted as current accounts under the heading of a household head (A) – usually the patriarch or the oldest economically-active man, with exceptions including widowed matriarchs –, these ledgers most frequently listed all household members (B) and their demographic characteristics (C), including marital status, sex, age, and nationality (mostly according to *jus-soli* principles). Since its managerial purpose was to keep track of coffee groves cultivated during the lean season, most accounts recorded the colony within the plantation where laborers resided and/or cultivated coffee (D)[3] and the number of coffee groves per *estaca* or *talhão* (E).[4] Since *estacas* were the smallest unit of agricultural planning within a plantation, these data provide the exact location of the groves cultivated by the households, an uncommonly precise historical proxy for the quality of agricultural capital. *Registries* also included the prices paid for sub-totals in each *estaca*, which were either explicitly recorded (F), or can be derived from total output (G) and labor income (H). The number of prunings the household executed per agricultural season were also frequently recorded. Finally, most *Registries* included information on households' residences. Accuracy and detail of such data vary across ledgers, but the most complete entries include numbering of the house within the colony, total number of rooms, windows (with leafs), doors (with leafs), lockers in the front door, and of electric lightbulb sockets, if any, as of 1926.

III.II. Data generating process: Accounting & managerial practices in a coffee plantation

The exact managerial and accounting practices by which data were collected and processed by plantation managers and accountants have now been lost for the most part. Nonetheless, hints onto the complex functioning of a Brazilian plantation and their accounting practices are provided by contemporaneous observers (Perret-Gentil 1851; Davatz [1858] 1941); rural manuals, including of agricultural accounting;[5] current academic research (Heflinger Jr. 2007; Bassanezi 2019; Pissinato and Vian 2021); and, now, by the multiple layers of information available in the ledgers of the *Ibicaba Collection*.

During the lean season, the daily functioning of a plantation required allocating labor to the cultivation of coffee and other cash crops, to the production of subsistence goods and fruit orchards, to livestock care and husbandry, and to various specialized workshops (*e.g.* pottery, carpentry, and blacksmithing), as well as to construction and maintenance of infrastructure (*e.g.* pathways, fences, residences, yards for drying coffee berries etc.), and to the operation

[3] Colonies of residence and of work were almost always the same. For a thorough discussion on the matter, *cf.* Section IV.

[4] In theory, *estacas* and *talhões* refer to different spatial units. However, these terms have been used interchangeably by *Ibicaba*'s accountants.

[5] Interviews with plantation managers and local accountants are particularly fruitful, even if referring to the second half of the twentieth century. One such interview was with Luiz Atílio Bragotto de Castro, the last operating manager ("administrador") of *Ibicaba* itself (in Witzel de Souza and Gardenal 2021).

of various mills (*e.g.* corn flour and sugarcane brandy) (Holloway 1980; Stein 1986; Bassanezi 2019; Vian and Pissinato, 2021).

Most of these tasks had to be carefully tracked in accounting, as evidenced by the variety of objects preserved in the *Ibicaba Collection* (Witzel de Souza and Gardenal 2026). Since most accounting entries cut across various managerial operations, the same information had to be accurately transferred and verified across various ledgers. For instance, as an ordinary workday went on, overseers and assistant managers made quick notes on *Daybooks* and *Laborers' Timesheets*. Once the workday or piecerates were completed, the corresponding information had to be correctly transferred to *Ibicaba*'s *Current Accounts* and *Cash Flows*, subsequently used to compute the plantation's annual yields, revenues, costs, and net profit. That same information had to be disaggregated for each household and copied onto *Laborers' Current accounts*. Finally, legal mandate and good managerial practices implied that data on output and labor remuneration had to be recorded with precision in a booklet (*caderneta*) held by each *colono* household, for them to keep track of their own Current Accounts.[6]

During the harvest season, households exchanged the produce of the harvest for tokens or receipts (*bilhetes*) corresponding to the volume of coffee berries they had delivered.[7] Overseers and assistant managers who received the berries in the field must have immediately recorded these yields on *Harvest Ledgers*, *Daybooks*, or slips of paper that were then passed on to the plantation's managers . By the end of the harvest season or sub-periods thereof, managers and accountants counted the tokens or receipts that laborers presented back to them and compared those to the yields registered for each household in the ledgers. The sources reveal that mismatches were taken very seriously. *Harvest ledgers* #0068 and #0077 have detailed notes on tokens lost by laborers during the harvest season, with stricter monitoring as of the 1920s. The mismatch noted in the account of José Braz on August 11, 1927, is illustrative of the accounting procedures (#0068, Digital Image 0034).[8] By noting that "[…] there is a difference of 1 extra *alquere* [SIC] in the slip from the harvesters' list", the accountant made it clear that he was comparing the daily registries delivered to him by an overseer with the number of tokens held by the laborer.[9] A more detailed, and somewhat curious case is that of Martim Zanetti's household (#0077, DIs 0111, 0113). In August 1937, they had lost one token, which the accountant noted down. One month later, the accountant remarked that Martim Zanetti had "h[arvested] 1 extra alq.", suggesting that his household had found the lost token and added it to their September's yield. In the end, the sum of daily yields recorded for Martin Zanetti's household corresponds exactly to the total computed in the source for these two months, showing that the accountants did add that extra *alqueire* in their credit. Notwithstanding, the notes made it crystal-clear that the managers were very attentive to any deviation between tokens, overseers' field notes, and accounting records.

---

[6] There are no *cadernetas* left in *Ibicaba*, as these were kept by the laborers.

[7] Also referred to in the consulted ledgers as "*cartão*" (card), "*alqueire*" (unit of volume), "*ficha*" (chip or token), and "*talão*" (coupon or receipt). In a hurried note for the 1946 harvest, an accountant wrote: "13 [*alqueires*] sobra taboleiro", *i.e.* 13 *alqueires* were missing; the word "Taboleiro" was not found elsewhere, but most likely refer to an abacus or a similar device.

[8] "Digital Image" henceforth referred to DI, for brevity.

[9] Similarly, the general note for *Grama Larga* colony's harvest in August 1929 informs that "One card has not been delivered on the 1st [/] We received 403, but harvested [=] 402" (Ledger #0077, DI 0010).

In regard to *Registries*, I believe that managers allocated newly arrived households to colonies (and *estacas* therein) where labor was most demanded, and reallocated them accordingly once contracts were renegotiated. This view would be in line with the allocation of most laborers to *Ibicaba*'s five largest colonies, as discussed in Section IV. Moreover, various slips of paper annexed to Ledger #0043 (DIs 0258-0277) contain data on households' demographics, coffee cultivation, and housing, separated by colonies. Their dates and various handwritings suggest that the resident population was surveyed once a year by sub-directors in each colony.[10] Given the errors identified in the households' series (especially in ages and age orderings of the youngest children) and reference to family ties ("mother", "brother", "sister", in-laws etc.), it is likely that sub-directors surveyed only household heads about the characteristics of individual household members. That information was passed on to general managers and accountants, who altered preexisting entries in the *Registries* if household characteristics had changed between years by birth, marriage, death, in- or out-migration etc.

Four other ledgers support this interpretation of the data generating process for the *Registries*. Ledgers #0099 and #0100 recorded data from *Colônia Grande* and *Pão de Lot* (overwritten to read *Grama Larga*) colonies in *Ibicaba*, while Ledgers #0098 and #0101 refer to *Iracema* and *São Francisco & Ipiranga* farms. All four started in the 1936-1937 agricultural season and were written in booklets originally printed to serve as laborers' booklets (*cadernetas*). The different handwritings and accounting styles show that these ledgers were written by multiple agents over a short period of time. The orthography and calligraphy of some suggests that the enumerators had only basic levels of instruction, with frequent phonetic spellings of names and surnames.[11] Because all demographic data found in Ledger #0100 overlap perfectly with those of Ledger #0083 for the 1936-1937 and 1941-1942 agricultural seasons, it is very likely that the former was the draft produced at the colony and then transferred to *Ibicaba*'s head managers and accountants. The same is true of Ledger #0099, which, in addition, overlaps perfectly with Ledger #0083 for the 1937-1938 and 1942-1943 agricultural seasons.

III.III. Transcription, Matching, and Data Wrangling to the Household level

In addition to keeping track of daily operations, *Ibicaba*'s accountants had to document numerous changes in its labor force caused by the intense labor mobility that characterized the economy of São Paulo at the turn of the twentieth century. The high turnover documented by the ledgers is a testimony of that phenomenon: For 1,205 uniquely identified households, the average household remained in the sample for 4.97 years, with a median of only 3 years.

[10] Numerous annexes have been added by contemporaneous to the historical ledgers. They are described in the "Abstract" of each object in the *Ibicaba Collection* and have been fully listed as an appendix to Witzel de Souza and Santin Gardenal (2026). The single-page annex to Ledger #0083 (DIs 270-1) is particularly interesting for the purposes of this paper: It is a drawn map of what was probably a section of *Colônia Grande*, tracing the location of the houses of some *colonos*, and facilities such as the colony's chapel, school, and the house of the colony's inspector (*fiscal*).

[11] This issue was very prominent in the 1940-1941 entries for *Pão de Lot – Grama Larga* colonies. The phonetic spelling reflects the dialect of São Paulo's countryside, especially with the usage of "r" for the phonetic "l" ("Sirva" instead of the correct "Silva") and corruptions of non-Iberian surnames (such as "Bergue" for "Berg", "Manhais" for "Magnaes", and "Croce" for "Kloss").

These households were also frequently renegotiating contracts, as perceived by the varying IDs of their *estacas* and colonies of residence, with 147 uniquely identified households moving out at least once between colonies in *Ibicaba*. Households' composition was frequently changing as well, be it via immigration of new family members or by high birth rates among very young households (for details, *cf.* Section IV). This implied that managers and accountants had to update ledgers frequently, including the *Registries*. Such operations led to a plethora of data corrections, glosses, crossing-outs, and overwritings in the original sources, with great variation of clarity and consistency within and between ledgers. Some annotations were methodically documented, or are as intuitive as the "*L[ançado]*" notes (accounting postings) thickly penciled in blue or red on top of an account. Others have unfortunately lost their meanings to the modern reader, including symbols of various shapes and colors, and numbers that must have been used as keys to match households and accounts within and between ledgers.[12]

The first decision in compiling the dataset was, therefore, to choose what information best corresponded to the actual economic and demographic characteristics of households at any point in time. In transcribing the data, I opted to extract the maximum amount of information available per page, creating various possible combinations with entries that had been crossed out, corrected, or added *a posteriori* over an original account. This extra information is used in the robustness checks to evaluate the sensitivity of results to choices I made in compiling the dataset.

The transcription of *Harvest ledgers* was straightforward. For the purposes of this paper, total yields per month were aggregated per agricultural season for each household.[13] Information on prices and labor income were added whenever directly available or derivable from the sources themselves. The transcription of the *Registries* was more challenging. Beyond many crossing-outs and additions of new household members and *estacas*, it was not uncommon to find entire accounts crossed-out or attributed to more than one household head. The sources were sometimes clear about what exactly had been swapped between households: Coffee groves, houses, credit, debt obligations etc. However, on some occasions it was not obvious which piece of information belonged to which household. In transcribing the data, I took into consideration solely the evidence available in the source, including annotations; in a later stage, with all data transcribed and matched, I returned to those cases and checked for inconsistencies within and between ledgers. In addition, the quality of the data varied within and between *Registries*, including the availability of demographic characteristics, the exact location of *estacas*, and housing amenities. In compiling the dataset, I opted to minimize such inconsistencies within the household series. The compiled dataset thus sticks as strictly as possible to historical accuracy in transcription and minimizes internal inconsistencies at the

[12] I did not transcribe such keys because this paper does not require cross-ledger referencing beyond matching households by name and demographic characteristics. Nonetheless, the application of new LLMs to them might shed new light onto the complex managerial functioning of plantations.

[13] I have exceptionally transcribed daily data on harvesting only if monthly data were not available or if there was any reason to doubt the correctedness of the values reported per month. The transcription of daily harvests is of great interest, but still depends on future improvements in automated transcription of handwritten data.

household level after matching the households. All decisions in attributing *estacas* and houses have been thoroughly documented in the transcription files.

With the data transcribed, the next step was to match households within and between ledgers. At first, this had to be accomplished within and between *Registries* only. The second stage then required wrangling individual-level data from the *Registries* onto the level of households, for matching them with the *Harvest ledgers*.

Difficulties in matching increased proportionally to the time dimension of the panel and number of household members. The longer the series and the larger the household, the more likely it became that first names would be repeated for different people, a problem particularly common in households with more than one nuclear family, whose naming practices included that of paying homage to family members still living in the same household, especially grandparents and godparents. The *Registries* sometimes recorded all household members by decreasing age (starting with all married and/or widows), but sometimes separated them by nuclear families, which facilitated the matching of those more complex and larger households once the data was pooled together across all *Registries*.

A minor complication worth mentioning for the matching is that some few households made private arrangements between themselves. It could thus happen that household i worked on behalf of household j. Bearing no connection to their private businesses, *Ibicaba* managers remunerated household j, who must have paid household i according to their own private arrangements, which were obviously not recorded in *Ibicaba's* ledgers.[14] In these exceptional cases, I attributed the physical yields to the household that actually produced it, as physical output is my main outcome of interest to compute quantity-based TFPs later on; however, I assigned to each household the level of labor income exactly as recorded in the sources.[15]

In the remainder of this section, I discuss the process of data cleaning for households' demographics, with particular focus on refining the sources' categories for ethnolinguistic origins; the identification of *estacas* for coffee cultivation; and information on housing.

III.III.I. *REGISTRIES*'S DATA: DEMOGRAPHICS

The first inconsistency identified in the transcribed data was with the first names of some household members. This issue can be a problem for the matching households if the inconsistencies referred to the household head directly or if additional information on

[14] *e.g.*: "N.B. Fica creditado O Cafe que Joao Ciavollela colhem a metade em c/ de José Giovanelle partes eguaes", *i.e.* a 50%-sharing of total yields and labor income between these households, totaling 364.000 *mil-réis* for each] (Ledger #0040, DI 0106).

[15] There were also some special agreements made by the plantation managers. For instance, in 1924/5, Jorge Küller cultivated 6,678 coffee groves as sharecropper in *Pinheiro* colony, where other households worked as usual *colonos* (Ledger #0062, DI 0128). Similarly, by the 1920s, a number of households started cultivating coffee and subsistence crops "A meia", *i.e.* leasing-in land under a crop-sharing mechanism. Such arrangements became common in the 1930s, with various old *colonos* working as "Meeiros" and "Terceiros", *i.e.* leasing-in land under ½ and ⅓ shares (Ledger #0083). During the Great Depression, this was likely a strategy to keep the labor force around, while retaining landownership, instead of parcelling out and selling plots of land to laborers.

household's characteristics were necessary to guarantee the match. Some errors were obvious misspellings, poor translations of foreign names, diminutives etc. of no bearing consequence. In other cases, however, different first names were used to refer to one and the same person (identifiable by demographic characteristics). In those cases, all available information in the household series – including data on sex, age, civil status, and nationality – was used to decide whether entries referred to the same individual or not. These cases are thoroughly documented in the transcription and matching files. As we cannot positively know what happened to these individuals in that interval, I decided not to impute any household member to the households' series if they were not explicitly listed as a household member in year t. Some mistakes further implied the attribution of different sexes for an observation, given that most first names are declined by gender in Portuguese. Besides inconsistencies in naming, the series also had some few mismatches in recorded sex, especially of infants and toddlers. The same procedure as before was applied to these cases and these discrepancies were corrected with the most likely sex based on the entire household series.

A more common inconsistency in the first transcribed and matched series referred to the recorded ages. Changes in individual age orderings of household members were common, reinforcing the hypothesis that demographic data were collected from a survey-like enumeration made with household heads, who reported the age of each household member. Except for those cases in which errors would have serious implications for the structure of the household (*e.g.* swapping siblings for couples while ordering household members by age), I kept these inconsistencies as in the sources, as they might provide information on the household head's numeracy. For the few cases in which individual ages were not recorded, I imputed them as a linear function of the most reliable neighboring year with available data.[16]

Data on marital status were much more consistent, with only sparse swaps between couples and other household members, which could be corrected with their age series. More curious were the listings of new household members, usually single young-adult women, for whom little additional information was recorded. These were most likely brides who had joined their new households. I recorded them as singles if no explicit marital status had been recorded in the sources, except if the woman: (i) had been recorded as married in $(t - i)$, *i.e.* I assumed that no divorce took place, especially if both people remained in the household;[17] (ii) was recorded as married in $(t + i)$ and the couple had a child at least as old as the interval i, *i.e.* I assumed that there were no illegitimate births; or (iii) her marital status could be unequivocally inferred from the household's structure, especially in mononuclear households.

[16] These errors are averaged out once the demographic data are wrangled to the household level, but they might be used in future research as proxies for numeracy. All imputations are marked in the transcription files and should obviously be excluded from estimations of numeracy.

[17] Of the 19,616 entries in the series with all household members, only one person was ever registered as "Apartado", *i.e.* divorced, "separated". There were many cases in which one partner was missing for one or more years; however, as explained, I did not impute these people to the series, nor made any change to their marital statuses.

III.III.I.I Defining ethnolinguistic origins & groups

The most problematic demographic variable in the *Registries* was the nationality recorded for each household member. Most entries in the sources followed the Brazilian *jus-soli* legislation and assigned the Brazilian nationality to any Brazilian-born child. However, loose *jus-sanguinis* criteria were also adopted at times, while yet others entries pell-mell classified everyone in a household as foreigner or Brazilian, even if previous entries had categorized some household members as foreigners (usually the elderly) or Brazilians (usually young children). After identifying all ethnolinguistic origins in the sample, I first standardized the nationality assigned to each household member according to the household's entire demographic series.[18]

Seven ethnolinguistic groups were defined from the complete list of individuals' origins and nationalities found in the sources: *Non-Whites*, *Italians, Germanics, Portuguese, Spaniards, Eastern Europeans*, and a group that can be loosely defined as *White Brazilians* (Table 01). As discussed in detail below, *White Brazilians* should in fact be strictly defined as *Brazilians* who have not been identified as Non-Whites and who did not have a surname in the sample traceable to non-Iberian European immigrants.

Non-Whites include individuals for whom the skin color Black (*Preto*) was recorded in the *Registries* in addition to any birthplace origin (African or Brazilian). Their classification was complemented with any note on skin colors found in the consulted source from the *Ibicaba Collection* (*e.g. Harvest ledgers* that recorded household heads as "First Name - Preto") and registries of Non-White skin colors from ledger BPML,[19] which eventually classified individuals by shades of color: Blacks (*Pretos*), Browns or Mixed (*Pardos*), and lighter-skinned Non-Whites (*Trigueiros*). Except for two men registered as Africans in 1888-1890, all other *Non-Whites* were Brazilians.

After harmonizing the data on birthplace origins from the *Registries*, the first step to categorize the ethnolinguistic origins of individuals in the sample was to create a list of surnames uniquely identified in the *Registries*, including variations in their spellings. The second step was to check whether these intrasample surnames had ever been registered in any transcribed source as belonging to *Non-Whites*, *Italians*, *Germanics*, *Portuguese*, *Spaniards*, or *Eastern Europeans*.[20]

[18] "*nation_source_i*" is the corresponding variable in the files used for the creation of the dataset. The subscript "i" refers to individual household members. Notice that the longer a household worked and resided in *Ibicaba*, the more likely it was for inconsistencies in their recorded nationalities to appear. Substantially, we cannot know whether some European immigrants even naturalized, as such information was never stated in the sources.
[19] For brevity, ledgers identified by prefix *Biblioteca Paulo Masuti Levy* are henceforth referred to as "BPML".
[20] This classification was particularly fruitful to identify the ethnolinguistic origins of women who married while working in *Ibicaba*, but whose maiden name had been recorded in the sources.

Table 01 – Classification of individuals by ethnolinguistic origins based on origins listed in the *Ibicaba Collection*

Ethnoling. category	Nationalities listed in the source
White Brazilian	Brazilian not identified as non-white (see below)
Non-white	African (non-white); Brazilian non-white; any nationality (including unidentified) categorized as "non-white" in ledger *PMLEVY0004* or notes in the source (see below)
Italian	Italian; Italian-Brazilian; Austrian-Italian; Unidentified locations: "Chare/Charecia"
Germanic	German; German-Brazilian; Austrian; Swiss; Danish & German-Danish (individuals inconsistently classified as Germans and Danish)
Portuguese	Portuguese
Spaniard	Spaniard; Argentinian (one single individual from HH headed by an Italian father and a Spaniard mother)
Eastern European	Russian; Polish; German-Russian & German-Russian-Polish-Lithuanian (HH inconsistently classified under these various nationalities)

These data on birthplace origins and surnames were used to classify the ethnolinguistic origins of each individual according to the following rules: *Non-Whites* were all individuals ever classified as Non-Whites by their skin color or being intrasample traceable descendants of those individuals; particular attention was paid to not classify as *Non-Whites* those Portuguese immigrants and their descendants who shared a common surname with the Black population, particularly among those listed in ledger BPML #0004. This classification highlights the potential discrimination by skin color faced by the Black population and the intergenerational transmission of cultural traits between them, including potential descendants of ex-enslaved people. *Italians*, *Germanics*, *Portuguese*, *Spaniards*, and *Eastern Europeans* were individuals whose birthplace origins had been harmonized as belonging to these nationalities or their intrasample traceable descendants. This classification emphasizes the potential intergenerational transmission of cultural traits among Brazilian-born individuals of European descent in the sample. Finally, *White Brazilians* were Brazilians (or individuals without recorded origins) netted out from all individuals identifiable as Non-Whites or of intrasample European descent.[21]

$(Origins_h)'$ corresponds simply to the ethnolinguistic origin of the household head. Since the bargaining for contractual conditions were mostly carried out between a manager and the household head, the latter's origin was the information that the former had on that household's origin. For few cases in which more than one ethnolinguistic group headed the same household, I assigned to them the ethnolinguistic group that was culturally most distant from Brazilians, *i.e. Eastern Europeans* $\succ$ *Germanic* $\succ$ *Italians* $\succ$ *Spaniards* $\succ$ *Portuguese*.

Three alternative classifications are used to check for the sensitivity of the baseline results: $(Origins\ imposed_h)'$, $(One\ drop_h)'$, and $(One\ drop\ imposed_h)'$ increasingly augment the likelihood of households being classified as of *Non-White* or non-Iberian European origins.

$Origins\ imposed_i$ is based on extended surname-nationality pairs that classify surnames not listed in the *Ibicaba Collection* as of *Italian* and *Germanic* origins. This classification also attributes *Portuguese* and *Spaniard* origins to rarer Iberian surnames. While there is little reason to doubt that *Andrietas* and *Benettis* or *Helds* and *Scherers* – whom the sources classified as "Brazilians" – had Italian and German-speaking forefathers, we do not know how far removed these were. The classifications of $Origins\ imposed$ accommodate these cases.[22] $Onedrop_h$ and $Onedrop\ imposed_h$, in turn, classify the entire household as *Non-white* or of European origins with the farthest cultural distance to Brazilians if that household ever had any single member who was Non-White or whose ancestry was traceable to any of the listed European nationalities.[23] In other words, if household h ever had a

[21] An unavoidable caveat of this approach is the impossibility of identifying matrilineal ascendancy. However, whenever available, the surnames and nationalities of the wives of household heads were also included in the list of uniquely identified surnames.
[22] The complete list of surnames is available upon request..
[23] The nomenclature "One drop" is evocative of the American system of racial classification prior to the Civil Rights Movement. This principle is at odds with perceptions of ethnolinguistic origins in Brazil (and Latin America more broadly) and has been used in this paper as an extreme bound classification. Racial classifications in Brazil were frequently entangled with other variables related to socioeconomic status, including occupations.

Non-White member, the whole household was classified as *Non-White* for its entire series; if household h had more than one member whose origin was traceable to more than one European nationality, the entire household was assigned to the ethnolinguistic group most distant from Brazilians for its entire series (*i.e. Eastern Europeans* $\succ$ *Germanic* $\succ$ *Italians* $\succ$ *Spaniards* $\succ$ *Portuguese* $\succ$ *Brazilians*).

Table 02 reports the number of uniquely identified households per classification criterion. By construction, $Origins_h$, $Onedrop_h$, $Origins\ imposed_h$, and $Onedrop\ imposed_h$ decrease the number of households classified as Brazilians and increase the number of Non-Whites and of European descendants, except for Portuguese and Spaniards. *Vis-à-vis* $Origins_h$, $Origins\ imposed_h$ increases the proportion of *Germanic* households by 30.00%, and of *Non-Whites* by 27.69%.[24] The proportions of re-categorized *Italians* (13.19%) and *Portuguese* (12.87%) are not negligible, either.[25]

Reassuringly, however, such variations do not alter the conclusions of the empirical analyses (*cf.* Section VII). Consequently, for being solely anchored in the sources themselves, having a clear *jus-sanguinis* interpretation attuned to Brazilian perceptions of ethnolinguistic origins, and reflecting the origins of the person who most likely negotiated the contractual conditions with plantation managers, $Origins_h$ remains our main variable of interest in this paper.

This implied a fluidity of racial categorizations that is common across Latin America, more broadly, as extensively recognized and discussed in Sociology. The dataset of this paper includes many examples of uncertainties related to historical "racial" classifications. For instance, in 1888, Maria Josepha das Dores, 27 years old, was registered as the white Brazilian wife of a white Brazilian. In 1890, Rosalinda Anna das Dores, 21 years old, was registered as the black Brazilian wife of a black Brazilian. Given their surnames, ages, and listings in consecutive years, it is likely that Josefa and Rosalinda related by blood, with the skin-colors of their partners attributed to them upon marriage. For similar perceptions on the fluidity of skin colors in Black families during the twentieth century in the region of *Ibicaba*, *cf.* the interviews with Black persons compiled by Callari and Baptista (1989), in particular the interviews with Aparecida Marques (especially p. 47).

[24] The increase in the *Germanic* group is explained by the longer history of German immigration to Brazil, in general, and to the region, in particular.

[25] The relatively high proportion of re-categorized *Eastern Europeans* is due to the small denominator (2 re-categorized households in a universe of 16 households).

Table 02 – Number of households per ethnolinguistic classification

Panel A: Uniquely identified HHs by ethnolinguistic group, according to classification criteria

	Brz.	**Non-wht.**	**Ital.**	**Grm.**	**Port.**	**Span.**	**East. Erp.**
Origins$_h$	193	65	379	30	101	42	16
Onedrop$_h$	177	71	392	32	98	39	17
Origins_imp$_h$	101	83	426	40	115	43	18
Onedrop_imp$_h$	92	95	433	42	105	40	19

Panel B: Distributing Origins$_h$ between different classification criteria

	Brz.	**Non-wht.**	**Ital.**	**Grm.**	**Port.**	**Span.**	**East. Erp.**
Brazilians = 193							
Onedrop$_h$	177	3	6	3	4	0	0
Origins_imp$_h$	101	18	50	9	13	0	2
Onedrop_imp$_h$	92	23	51	11	14	0	2
Non-Whites = 65							
Onedrop$_h$	0	65	0	0	0	0	0
Origins_imp$_h$	0	65	0	0	0	0	0
Onedrop_imp$_h$	0	65	0	0	0	0	0
Italians = 379							
Onedrop$_h$	0	0	379	0	0	0	0
Origins_imp$_h$	0	0	376	1	1	1	0
Onedrop_imp$_h$	0	2	375	1	0	1	0
Germanic = 30							
Onedrop$_h$	0	0	0	29	0	0	1
Origins_imp$_h$	0	0	0	30	0	0	0
Onedrop_imp$_h$	0	0	0	29	0	0	1
Portuguese = 101							
Onedrop$_h$	0	3	4	0	94	0	0
Origins_imp$_h$	0	0	0	0	101	0	0
Onedrop_imp$_h$	0	5	4	1	91	0	0
Spaniards = 42							
Onedrop$_h$	0	0	3	0	0	39	0
Origins_imp$_h$	0	0	0	0	0	42	0
Onedrop_imp$_h$	0	0	3	0	0	39	0
East. Europ.= 16							
Onedrop$_h$	0	0	0	0	0	0	16
Origins_imp$_h$	0	0	0	0	0	0	16
Onedrop_imp$_h$	0	0	0	0	0	0	16

III.III.II. REGISTRIES'S DATA: *ESTACAS* & NUMBER OF COFFEE GROVES CULTIVATED IN THE LEAN SEASON

The difficulties generated by the multiple layers of information from contemporaneous corrections, annotations, and overwritings are most prominent in the *Registries*' data on coffee cultivation . Because households could renegotiate yearly the number of coffee groves they cultivated, all ledgers with the number of coffee groves and *estacas* had many entries crossed-out, added up, annotated, and/or corrected upon. In most cases, accountants were careful enough to correct the total number of coffee groves cultivated and/or the corresponding labor income accordingly. Such corrections allow us to trace back all changes and to determine the correct quantities and *estacas* to be transcribed. However, it was not uncommon to find uncorrected sums in the sources, in spite of annotations over individual entries. It becomes much less obvious what should be transcribed in such cases and to which year those changes should be attributed. The problem is that we do not know when exactly the accountant altered each individual entry, as it could well be the case that households renegotiated their contracts for agricultural season $(t + 1)$ at the end of year t and that accountants wrote down those changes over the accounts of year t, before starting a new account in year $(t + 1)$.

Consequently, the number and location of coffee groves can be compiled by (i) simply copying the sums of groves reported in the sources, irrespective of changes in individual entries; (ii) summing up all groves listed in each *estaca*, including those crossed out in the sources; (iii) netting out (ii) from crossed-out groves in the sources.

In the descriptive statistics and baselines, I use quantities and *estacas* that are clearly traceable to individual entries, corrections, and reported totals. These variables reconstitute the most likely historical scenario for any year t, taking into account all crossed-out and added up *estacas* that most closely match totals reported in the sources (*cf.* Sections IV and V). Those variables also include extra information contained in few summary tables available in the *Registries*.[26] In the robustness checks, I take into account the two other criteria (*cf.* Section VII). Variable *tree_tot_estaca* uses criterion (ii): It sums up the coffee groves from all *estacas* registered in an account, irrespective of ever being crossed out.[27] Variable *tree_tot_estaca_check* uses criterion (iii): It sums up the coffee groves in all *estacas* ever registered in an account, but deletes those crossed out in the account. The variable *tree_tot_check02* mimics the baseline, but deletes *all* groves from crossed-out *estacas*.[28]

[26] These include extra information from ledger BPML #0004 on the number of coffee groves cultivated in 1890.
[27] In principle, this variable would exclude ledger BPML #0004, as it reports only the sum of coffee groves cultivated. However, for comparability with the baseline, that data have been added to *tree_tot_estaca*.
[28] In addition, the dummy variables *Crossed_out_Income_ht* and *Nullified_ht* are used as conditionals to create sub-samples without certain observations. *Crossed_out_Income_ht* indicates whether the total labor income from coffee cultivation from a certain household *h* had been crossed out in year *t*. *Nullified_ht* indicates whether the entire account of a certain household *h* had been crossed out in year *t*, *i.e.* it excludes observations that the accountants deemed null and void for any unspecified reason.

III.III.III. Registries's data: Housing

In spite of their uniqueness and invaluable nature to proxy for basic housing conditions, data on housing are the most inconsistently recorded in the *Registries*. The numbering of the houses was not always consistent, making it impossible to determine whether changes in numbering for the same colony capture households moving between houses, or resulted simply from accountants relabeling the number of the residences therein. Moreover, accountants were not very clear about the characteristics of the houses they recorded in each year. This problem was salient if those characteristics changed for the same house in consecutive years. While those changes could indicate actual structural renovations in the house (*e.g.* installing new windows, building up or demolishing rooms etc.), they could also just reflect changes in the conventions used by the accountants. For example, Ledger #0062 records the number of "quartos" in 1921-1924 (with gaps), but uses the label "commodos" for 1929-1931 (with gaps). Albeit colloquially synonyms, more strictly speaking, these words could refer to different spaces in a house, with the former defining bedrooms, only, and the latter, any room. Similarly, it is not always clear whether the *Registries* recorded the number of doors and windows in a house or the number of leafs in those doors and windows; or whether lockers in front doors had been counted or not among all "doors" - if not, the latter referred only to internal doors and could therefore proxy for the number of rooms.

The series on housing was harmonized over time as much as possible. Analogously to (the many fewer) inconsistencies in the demographic data, if more than one entry referred to the same house in one single year, only those entries that were most consistent with the rest of the series were retained in the dataset. Conditional on that criterion, entries containing more information were preferred. The series also permitted interpolating some of the housing data between years. However, I interpolated colonies of residence and house numberings only if the gap between the actually known data was of one year maximum, given that households would have had enough time to move between houses and/or colonies more than once if the gap was larger than one year. Moreover, I interpolated data on housing amenities only if the house was actually known, or if there was clear evidence that the household resided in the same house in that interval.[29] In these cases, if a missing value in year t was surrounded by conflicting information in years $(t\pm1)$, the smallest number of amenities was used. Finally, data on lockers and electric lightbulb sockets were never imputed retroactively.

III.III.IV. Data Wrangling & Matching households across sources

With data harmonized within and between *Registries* and *Harvest ledgers*, matching these two types of sources required wrangling the data to the level of households. Beyond a methodological requirement, as households were the unit of data collection in *Harvest*

[29] It was possible to derive data on colonies and house numberings from the complete transcribed data. For instance, if the house number was missing in the source, but the numbering of all neighbors had been listed, then the missing number could be derived by exclusion (*e.g.* if houses number 01, 02, and 04 had been registered and only one entry was left for that colony, I assigned that entry to be house number 03). These decisions are fully documented in the transcription and matching files.

ledgers, that was also a historically solid choice, as the household was the relevant unit of production in the coffee economy, in particular under *colonato* labor arrangements (*cf.* Sections II and V).

Data from *Harvest ledgers* had originally been collected at the household level. As such, the only required intervention, if any, was to convert eventual daily and monthly entries into aggregate yields per household per annum.[30] Conversely, except for housing data, *Registries* collected all data at the individual level. Per definition, demographic data had been collected at the level of household members and wrangling them from individual- to household-level required only counting them or obtaining measures of central tendency or dispersion over household members. Somewhat more complicated were the variables referring to the location of coffee groves and the distribution of *estacas*. These had been collected at the level of individual *estacas* (see Items "E" and "F" of Figure 01 – Panel B), precisely because one of the purposes of the *Registries* was to determine how many coffee groves the household cultivated in each *estaca.* The total number of cultivated coffee groves per household can be obtained by simply averaging its value for each household per annum. However, the numbering attributed to each individual *estaca* was just an identifier, without any intrinsic quantitative meaning. One possibility would have been to wrangle each *estaca* numbering from long to wide and to construct *estaca-fixed effects* for each *estaca* ever listed in the sources. I refrained from this option because it would have unreasonably increased the number of controls in all regressions up to 20 *estaca-FEs* to accommodate the maximum number of cultivated *estacas* in the sample.[31] The more parsimonious option was to collapse the numbering of *estacas* to the level of households with metrics of central tendency and dispersion. For each household in each year, I computed the mean, median, standard deviation, mean deviation from the mean (MDEV), and median absolute deviation (MAD) of the numberings attributed to the *estacas*. This approach has the drawbacks of compressing a historically much richer variable and the fact that some *estacas* were actually strings that I had to convert into numeric labels without any quantitative meaning. Nonetheless, this option is statistically more efficient and, combined with colony identifiers, still provides solid controls for the agricultural capital available to households.

With all data wrangled to the household level, the final step in compiling the dataset was to assign unique identifiers to each household and to match all sources over those identifiers.

I attributed the same unique identifier to a household if the head(s) of that household listed within and between ledgers were the same person, as identified by: (i) Identical names and surnames, or variations thereof that could be unequivocally attributed to misspellings; or (ii) sufficiently similar names and surnames (orthographically or phonetically) and household series with a demonstrably continuity in their demographic characteristics, and/or cultivated *estacas*, and/or housing data, and/or harvest series.

[30] I have nonetheless forced the Stata© command "collapse" over the *Harvest* data to be sure that all outcomes could be matched with data from *Registries*, which were collapsed to the household-level.
[31] There were 2,902 observations (overall) for *estaca* numberings, but only four observations (corresponding to only two uniquely identified households) reached that maximum number of 20 cultivated *estacas*.

An important issue in uniquely identifying households was to define whether and when an original household broke down into distinct units. *Ibicaba*'s labor force was composed of a large share of young married adults; and although the sample median indicates that most households were mononuclear, the sample abounds with multinuclear households. In this demographically dynamic scenario, if the household head remained in the sample for the entirety of the available series, I attributed the same unique identifier to that household. I applied the same rule if a widowed matriarch took over the household head after the passing away of her husband, independently of what happened to other household members. However, I distinguished between two cases if a son took over the household head of a deceased patriarch. If that son was already married before the passing away of the original household head, I kept the same identifier to that household because it continued to be a multinuclear household as it had been before. However, if the new household head was a single son, I attributed the same identifier to that household only until that son got married, because the newly married couple would (or could) lead to substantial alterations in the composition and structure of the original household (by childbearing, splits of siblings who married out of the household etc.)

III.IV. Assessing Data Quality: Auditing errors in historic accounting

Contemporaneous commentators and historical research have long recognized and consistently demonstrated the importance of good accounting practices for enhancing economic decision-making and increasing productivity (Carneiro and Carneiro 1933; Brandão 1945; Yamey 1949; Pollard 1983; Matthews, Anderson, and Edwards 2003; Lampe and Sharp 2019; Sangster 2024). As this section made clear, the ledgers from the *Ibicaba Collection* are far from homogenous, which may raise concerns about data quality, in general.

Although first arrived in Brazil as indentured sharecroppers in *Ibicaba*, José and Simão Levy had a comparatively high level of human capital and long entrepreneurial experience, having managed the storehouse of *Ibicaba*'s *Colônia Grande* in the 1860s and worked as merchants and local bakers in the neighboring town of Limeira (Freitas Levy ([1993] 2023); Heflinger 2021; Tamiazzo 2021; Witzel de Souza and Santin Gardenal 2026). Consequently, the Levy Brothers undoubtedly had the skills required to properly manage a plantation by the 1890s. Nonetheless, two objects in the *Ibicaba Collection* reveal their remarkable learning curve as landowners and plantation head managers. Ledger #0001, the first managerial record produced under the Levys' ownership of *Ibicaba*, squeezed two types of account (the plantation's *Current Accounts* and laborers' *Payrolls*) into one single document. This mixing up of different accounts into one book was probably due to inexperience, before practice proved the benefits of separating these records in different ledgers. Even more telling is Ledger #0005, the first *Harvest ledger* produced under their administration, in 1890. The authors of that ledger (Levy Brothers themselves or *Ibicaba*'s central managers) first adopted a *Daybook* notation to record the harvest: Every single day, the accountant(s) wrote down the full name of the household head and the corresponding amount of coffee harvested that day. This extremely laborious procedure changed in 1891, when the accountant(s) adopted a

Current Accounts' notation: Under one single heading for each household, the accountant(s) noted down all days worked and the volume of coffee harvested by that household only (*i.e.* the same format as that of Figure III.1 - Panel A). The improvement in accounting efficiency and transparency is staggering and can be grasped by the number of rows whose transcriptions were necessary to compute the total annual harvest per household:[32] With the 1890 *Daybook*'s notation, 2,612 rows had to be transcribed in order to compute the volume harvested per year of 110 households; the *Current Accounts*' notation used thereafter (1891-1895) required transcribing only 131 rows, on average, to compute the same information for 108 households per year, on average.[33]

Accounting also played a central role in labor disputes in an economy characterized by low monetization and high reliance on credit. It is worth recalling that the *Sharecropper's Rebellion* of 1856/7 took place in *Ibicaba* itself (*cf.* Section II). At the core of that labor strike were laborers' complaints about fraudulent accounting: At least two out (§§ 9 and 15) of the 18 complaints raised by the sharecroppers related to intransparencies in accounting practices (Davatz [1858] 1941, pp. 249-55; see also Vergueiro 1874, p. 6).

To the best of my knowledge, no study has ever attempted to check the validity of laborers' allegations about fraudulent accounting, in *Ibicaba* or elsewhere. Such study would be as relevant as hazardous. Moreover, even for *Ibicaba*, conclusions would be strictly circumscribed to the ledgers studied, first because there is no reason to extrapolate to the *Vergueiros' era* the same types of errors committed by the Levys (or vice-versa), and second because improvements in accounting might have been a *consequence* of previous labor strikes. Finally, as revealed by the data transcription for this paper, various types of error could be committed in different stages of the production process, with liability falling upon different agents. Errors could have been committed by sub-directors who incorrectly registered daily production; head directors who failed to spot these mistakes as the agricultural season progressed; accountants who mistakenly updated entries within and between ledgers etc.[34] Inferring *purpose* from such errors would be even more daunting.[35]

A much simpler question of direct interest to this paper is whether accounting errors introduced any bias in the labor income compiled in this particular dataset. To assess this question, I computed the difference between the labor income per pruning recorded in the *Registries* and the value that the household should have perceived, assuming that individual prices and quantities registered in the sources were correct. I computed this error only for the

[32] The accountant(2) restarted using the *Daybook*'s notation for the 1902 harvest registries (#0005 – DI 78), but quickly reverted back to the more efficient notation of current accounts on the next page.
[33] (130, 225, 95, 106, and 99) rows for (129, 115, 95, 104, and 98) household, respectively.
[34] The sheer scale of the plantation posited challenges to accounting. Slips of paper annexed to Ledger #0040 (DIs 301-2 ff.), probably from the late 1910s, exemplify the hazardous task of computing total yields: Operations were computed by hand and checked multiple times (see ticks in different colors).
[35] An indirect solution to this issue would be to compare the number and magnitude of mistakes in laborers' accounts to those found in purely managerial ledgers. Analogously, it would be useful to show the production phases in which errors were more common and quantitatively relevant: Were they mainly caused by sub-directors, whose levels of human capital were expectedly lower? Or by head directors? If so, were the errors in output more common (against good managerial practices, more generally, and of direct interest to landowners), or were the problems mainly with labor income?

lean season since prunings represented almost ¾ of the average nominal income annually earned by *colono* households in our sample (*cf.* Section IV) and because auditing the harvest season would have required the transcription of all daily individual entries on coffee picking.

The mean overall accounting error computed for the corresponding 2,905 entries was 0.066 *mil-réis* per pruning, *i.e.* minorly positive in favor of the plantation, but with median zero.[36] Excluding the approximation error of 0.001 *mil-réis* per pruning, the mean positive error was 1.125 *mil-réis* (median: 0.033 *mil-réis*),[37] with a maximum of 120.694 *mil-réis*.[38] There were also 58 negative errors, *i.e.* favoring the laborers, with |mean| = 2.043 *mil-réis* (|median| = 0.300 *mil-réis*; |minimum| = 19.600 *mil-réis*; |maximum| = 0.002 *mil-réis*). For reference, the mean pooled price for cultivating one additional coffee grove in the sample (from 1904 to 1942) was 0.025 mil-réis.

These figures do not indicate any serious problem with accounting errors. In addition, Table 03 shows no evidence that such sporadic errors were correlated either with the sources from which labor income in the lean season was derived (variable $Source$), or with different time trends (linear or quadratic) to capture both the learning curve of the Levy's administration or the posterior economic decline of *Ibicaba*. These results increase our confidence in the quality of the data, suggesting that the contemporaneous data generation process was not subjected to systematic biases against the laborers.[39]

Independently of having these econometric issues resolved, such errors certainly affected daily lives on the ground. As an illustration, in 1923/4 the accountants overwrote the decimal and centesimal units of Antonio Zambusi's income in the lean season (Ledger #0062, DI 58). If individual prices and quantities are correct, the accountants rounded up (or missed) 0.008 *mil-réis* per pruning; since Zambusi's household pruned 6 times that year (as inferred from their annual income), that household perceived 0.048 *mil-réis* less than the contracted value, which corresponds to little more than two *vinténs* (0.020 *mil-réis*), the smallest coin denomination of the Brazilian *mil-réis*. Although that value is derisory and the difference almost certainly resulted from a harmless rounding up of cents, the difference might have dissatisfied the Zambusis if they ever noticed the difference. After all, as the legend in the *vintém* copper coin reminds us to this day: "A saved *vintém* is an earned *vintém*".[40]

---

[36] The entire exercise obviously depends on the assumption that p and Q were correctly registered in the sources.
[37] Strictly speaking, there were 892 observations with errors larger than zero; however, 617 were smaller than 0.001 *mil-réis*. Those are approximation errors in millesimal units that neither accountants nor laborers would have been able to compute.
[38] Corresponding to the account of Benedicto Cunha in 1929-30 (#0062, DI 205). Strangely, the accountant copied to Cunha's account the exact same income registered on that same page for another household (João Franco Silveira). This error is so large that it must have been noticed by the involved parties. I nonetheless transcribed the data as in the source.
[39] Results remain qualitatively the same if year-FEs substitute for the time trends. With non-clustered standard errors, the only statistically significant year-FE is 1929, corresponding to the outlier error of Benedicto Cunha. Independently of specification, if standard errors are clustered at the *Source* level, other year-FEs become statistically significant, but without any clear pattern or trend. To eliminate any concern that common shocks accrued to accounting practices could bias any estimate, the baseline specifications will control for year-FEs. All previous results are available upon request.
[40] The rounding up of centesimals (for *mil-réis*) became more prominent in the 1930s and meddled as of 1943 with the monetary reform that led to the issuance of *Cruzeiro* (based on decimals) as the new Brazilian currency.

Table 03 - Accounting errors in laborers' income during the lean season

	Dep. Var.: Accounting errors in income from coffee cultivation (lean season)			
	(1)	(2)	(3)	(4)
Source	-0.00872	-0.00863		
	(0.00795)	(0.00819)		
Year	0.0183	0.0993	0.00200	0.538
	(0.0156)	(1.811)	(0.00457)	(1.762)
$Year^2$		-2.11e-05		-0.000140
		(0.000472)		(0.000459)
Constant	-34.64	-112.3	-3.773	-518.8
	(29.50)	(1,736)	(8.778)	(1,693)
Observations	2,905	2,905	2,905	2,905
R^2	0.000	0.000	0.000	0.000
Adj-R^2	-0.000209	-0.000553	-0.000279	-0.000591

Notes**:** (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) *Source* is the categorical indicator for the historical source (account ledgers) from which the observation was extracted. *Year* is the calendar year of the observation, and $Year^2$ its quadratic term.

IV. Descriptive Statistics

iv.i. Laborers' demographic characteristics: Pooled Sample

The productive and demographic characteristics, as well as basic housing conditions of *Ibicaba*'s labor force, pooled for the entire period 1888-1958 (with gaps, according to data availability), is reported in Table 04.

A total of 5,595 observations have been compiled for 1,205 uniquely identified households. Most of these observations are concentrated around the year 1914 and remain, on average, 4.98 years in the sample.[1] Variations across the number of observations (*N*), panel identifiers for uniquely identified households (*n*), and average number of years in which the panel identifier remained in the sample (*T-bar*) occur due to attrition, with households entering and quitting *Ibicaba*.

Households had, on average, 6.65 people and, at the mean, were multinuclear, with 1.3 families per household.[2] However, households that had all data required for such estimation, the majority (689 uniquely identified households, corresponding to 2,143 observations) was mononuclear, in contraposition to the 197 uniquely identified households (754 observations) that were multinuclear.[3] Therefore, the impact of multinuclear households on the sample's overall mean is a consequence of the size and number of families within such households, not of the overall prevalence of multinuclear households.

Men were 54.90% of the overall sample mean, indicating the presence of seasonally employed male labor in addition to the household structure that prevailed in *colonato* arrangements. Out of the 27 uniquely identified households (corresponding to 30 observations) that had only singles (marital status) in them, 13 were households with just one male: One Brazilian, one Italian, one Spaniard, two Germanic, and eight non-whites. If additional data from summary tables and those with information only for household heads are considered instead, the number of uniquely identified households formed by just one single (marital status) individual increases to 18 (corresponding to 20 observations). The extra uniquely identified households were two additional Brazilians, two additional Italians, and one newly identified Spaniard.

[1] The same statistic could be obtained by summarizing the panel identifier; however, because the IDs bear no substantive information, I preferred to summarize the time dimension of the panel. These statistics are in line with those computed for the neighboring *Santa Gertrudes* plantation: According to Dean (1976, p. 171), in the post-Abolition era, "(…) two thirds of the colonos stayed more than three years" in that plantation.

[2] (#widows + (#married/2)).

[3] The computation of multinuclear households takes into account only those households that had all required data, *i.e.* non-missing civil status (married and widowed). It also excludes entries with data only for household heads and summary tables without clear assessments of households' structures.

Table 04 – Descriptive statistics, pooled sample 1888-1958 (with gaps)

VARIABLES	N	n	T-bar	Mean	S.D.	Min	Max
Year	5,995	1205	4.98	1,914	16.01	1,888	1,958
Production in the coffee economy: cultivation in the lean season							
Income cult.	2,953	618	4.78	963.900	853.300	32.220	8472.000
Price cult.	2,959	620	4.77	0.025	0.011	0.010	0.058
# of cult. groves	3,147	755	4.17	6,377	3,005	537	28,096
# of cult. *estacas*	2,942	620	4.75	6.18	2.83	1.00	20.00
# pruning per season	2,952	617	4.78	5.42	0.50	4.26	7.67
Production in the coffee economy: harvesting							
Income harv.	4,029	915	4.40	364.900	419.800	0.500	9713.000
Price harv.	4,032	915	4.41	0.890	1.128	0.307	15.000
Total harv. (in *alqs*.)	4,604	973	4.73	491.10	375.80	1.00	3323.00
# months harv.	4,661	990	4.71	3.94	1.16	0.00	6.00
1st month harv.	4,622	978	4.73	5.59	0.76	4.00	9.00
Last month harv.	4,622	978	4.73	8.56	0.78	5.00	10.00
Mean month harv.	4,622	978	4.73	7.08	0.54	4.50	9.00
Household characteristics							
HH size	2,972	826	3.60	6.65	3.41	1	30
# families per HH	2,897	760	3.81	1.30	0.69	0	6
Married (total)	2,897	760	3.81	2.24	1.30	0	12
Married (share)	2,897	760	3.81	0.3890	0.2400	0	1
Widowed (total)	2,897	760	3.81	0.18	0.42	0	3
Widowed (share)	2,897	760	3.81	0.0299	0.0839	0	1
Males (total)	2,972	826	3.60	3.52	2.05	0	16
Males (share)	2,972	826	3.60	0.5490	0.1840	0	1
Age (mean)	2,883	740	3.90	23.36	9.86	4	74
Age (median)	2,883	740	3.90	19.33	11.55	3	74
Age (max.)	2,883	740	3.90	48.04	14.55	4	112
7-60y (total)	2,972	826	3.60	4.94	2.79	0	19
7-60y (share)	2,972	826	3.60	0.7320	0.2360	0	1
Cons./Prod. ratio	2,972	826	3.60	0.4510	0.4640	0	3
Known amenities in the houses (figures per household)							
# houses	3,045	723	4.21	1.10	0.32	1	4
# rooms	882	199	4.43	5.84	1.92	2	17
# windows	1,664	365	4.56	5.11	1.91	1	16
# doors	1,651	361	4.57	5.71	2.21	2	17
# lockers	1,041	253	4.11	1.06	0.25	1	3
# lightbulb sockets	3,416	855	4.00	0.11	0.33	0	2
HHs mobility between colonies where coffee was cultivated							
Δ colony N	3,001	494	6.07	0.71	1.08	0	7
(Δ colony N)/T	3,001	494	6.07	0.07	0.11	0	0.67
Δ colonies j	3,405	859	3.96	1.51	0.75	1	4
(Δ colonies j)/T	3,405	859	3.96	0.31	0.28	0.03	1

Notes: (1) *N* refers to total number of observations; *n*, to number of panel identifiers, which in this sample corresponds to uniquely identified households; T-bar, to the average number of years *n* remained in the sample. (2) *Price cult.* is the weighted price for coffee cultivation. (3) The numerical values of *1st month harv.*, *Last month harv.*, and *Mean month harv.* correspond to months scaled from 01 (January) to 12 (December). (4) *Cons./Prod. ratio* = ratio of HH members in economically non-active ages (*i.e.* consumers of HH services) to those in economically active ages; "economically active" defined as 7-60y. (5) *Δ colony N* = Number of known times a HH changed colony of coffee cultivation; *(Δ colony N)/T = Δ colony N* divided by the number of years a HH remained in the sample; *Δ colonies j* = Number of different (unique) colonies where a HH cultivated coffee; *(Δ colonies j)/T = Δ colonies j* divided by the number of years a HH remained in the sample. (6) Figures are practically unaltered if we consider HHs changing residence between colonies.

The working population residing in *Ibicaba* was very young. The mean pooled age was 23.36 years (S.D. = 9.86) and the median was at striking 19.33 years (S.D. = 11.55). Even the maximum ages pooled mean across households was 48.04 years (S.D. = 14.55).

These households had a pooled mean of 4.94 individuals who were between 7 and 60 years old, *i.e.* in the age range used in the baselines of this paper to define the economically-active population of a household. Mirroring this statistic, the ratio of consumers to producers at the household level (*i.e.* the share of the very young – under 7 years old – and the elderly over 60) was at only 0.4510. In short, almost ¾ of household members were in a position to contribute economically to its internal division of labor, either in coffee production, subsistence agriculture, or house- and family-care.

In addition, 38.90% of the pooled sample were married and another 2.99% widowed. In an economy where numerous, unskilled laborers were highly demanded, these households were, on average, well-equipped to reap the fruits of economic success. However, as it will be discussed thoroughly in the rest of this paper, the endowment of working hands per household varied significantly by ethnolinguistic groups. A major burden of slavery was the small households it had generated for a period when a large household was a key asset (*cf.* the diff-in-means tests of Section V).

IV.II. Colonies of Work and Housing Amenities

The average household in *Ibicaba* resided in one single house with one locker in its front door. These numbers were pushed up by the maxima of a household distributed between nothing less than four houses; however, cases of one household residing in more than one house were exceptional, as indicated by the S.D.s of the number of houses (0.32) and lockers (0.25).[4] The small sample with the recorded number of rooms and the larger sample with the recorded number of windows and doors show that the average house had between 5 and 6 rooms.[5] Excluding common rooms for a kitchen and a living room, the average 6.65 household members had to share their private spheres in 3 to 4 rooms. Electric lightbulb sockets were registered for the first time in 1926, but remained a luxury for the entire period considered in this paper. They were reported only in *Grama Larga* and *Morro Alto* colonies, as well as in *Iracema farm*, where 114 uniquely identified households (corresponding to 317 observations) had one electric lightbulb socket, and 12 others (corresponding to 25 observations) had two.

The average household changed less than once (0.71) between colonies where they cultivated coffee and most households worked in less than two colonies (1.51). There were exceptions,

[4] With these distributions and by common sense, it is unsurprising that the median number of houses and lockers per household were equal to one.

[5] Figures per capita are available upon request. For most households, these conditions did not change at all while they lived in *Ibicaba*: the average household experienced a mean maximum variation of only 0.48 rooms, 0.48 windows, and 0.55 doors (and the quantitatively irrelevant variation of 0.06 lockers and 0.03 lightbulb sockets). However, there were exceptions. The maxima of variations over time indicate that some households could experience a sharp increase or decline in housing space of up to 7.5 rooms or, more realistically, 4.67-5.63, based on the number of windows and doors, respectively.

as shown by the maxima of these variables: four uniquely identified households cultivated coffee in as many as four different colonies and one even moved seven (known) times between colonies. Nonetheless, at the median, households did not move much between colonies where they cultivated coffee.[6] This is relevant because, in an economy characterized by high geographic mobility, once in *Ibicaba* most households stayed put with the agricultural capital first made available to them.

Table 05 now turns to each colony's productive capacity and size. Data on aggregate factors of production are pooled at the level of colonies for all years with available information (1888-1958, with gaps), summing up all people who ever lived in these colonies and the number of coffee groves ever registered in them. As for the rest of the paper, except for a few robustness checks, Table 05 does not distinguish between colonies where *colonos* resided and colonies where they worked.[7]

In terms of coffee trees ever planted in them, *Colônia Grande* (literally, "Large Colony") lived up to its name, followed by *Grama Larga*, and, in a distant third place by *Morro Alto*, and, again after a large gap, by *Saltinho & Pinheiro* and *Pão-de-Ló* colonies. All these colonies had more than one million coffee groves summed up for all years with available data. These five major colonies rank in that exact same order with regard to the total population ever registered in them, indicating that no labor-saving technology was used in any. Rather, the marginal increase in the number of groves between them is accompanied by a proportional increase in the number of laborers.

The last two columns of Table 05 report the average number of coffee groves and *estacas* cultivated per household. These averages were computed for each colony with households' actual average cultivation, *i.e.* not simply as the ratio of aggregate groves per total population.

[6] I did not consider changes in housing within colonies because it is impossible to determine whether a change in house numbering was caused by a household moving between them or by different conventions for the numbering of the exact same houses, as recorded by the plantation managers.

[7] In the most extreme scenario, colonies of residence and of work would have differed for 200 observations maximum, if we assigned all data that are missing in that regard to correspond to colonies of residence differing from those of work. Actual data show that colonies of residence and of work differed for only 25 observations with all required data for such an assessment. The maximum number of actual differences occurred for the household of Mathias Perez (ID = 0151), who in 1925 was curiously living in the "house of a neighboring property" ("Casa de propriedade annexa" – Ledger 0062 - Digital Image 128). Similar information about workers living outside of *Ibicaba* appears another two times on that same ledger (Digital Images 76 and 101).

Table 05 – Descriptive statistics: Labor and agricultural capital per colony (pooled sample 1888-1958, w/ gaps)

	Obs. & uniquely identified HHs			Sum of totals ever recorded		Averages per HH (with original data)	
	N	n	T-bar	Total population	Total groves	# of cult. Groves	# of cult. Estacas
Col. Grande	1025	217	4.72	5,921	6,042,748	6159.78	6.26
Grama Larga	883	240	3.68	5,311	5,789,359	6723.99	6.32
Morro Alto	509	124	4.10	2,996	3,343,457	6740.84	6.30
Salt.-Pinh.	237	59	4.02	1,367	1,496,144	6679.21	5.83
Pão Ló	146	47	3.11	1,021	1,027,281	7390.51	5.78
Col. Ibicaba	185	120	1.54	816	404,045	3848.05	3.08
Lage	89	35	2.54	451	649,738	7555.09	6.76
Teteia	102	48	2.13	435	572,992	5907.13	5.13
Abundância	17	17	1.00	105	25,569	3652.71	.
Mariquita	9	2	4.50	69	95,342	10593.56	10.44
Port. Isabel	3	1	3.00	12	18,574	6191.33	6.67
Boa Esperança	15	15	1.00	60	18,439	1317.07	.
Paineira	1	1	1.00	1	10,157	10157.00	10.00
Linha	4	2	2.00	7	9,702	2425.50	2.75
Col. Nova	1	1	1.00	.	8,022	8022.00	7.00
Ipiranga	3	3	1.00	27	.	.	.
Multiple colonies	2	2	1.00	18	15,237	7618.50	.
Laborers	70	70	1.00	260	.	.	.
Iracema farm	53	39	1.36	333	294,001	5547.19	6.30

Notes: (1) For each colony, the "Sum of totals ever recorded" sums up all observations ever available in the sample. (2) "Averages per HH" compute the mean number of coffee groves and *estacas* cultivated per household in each colony, according to the original data, *i.e.* prior to the aggregation over time for each colony.

The five largest colonies rank here at the center of the distribution, with households cultivating an average of 6,576 coffee groves and 6 *estacas*, numbers similar to the much smaller colonies of *Portão Isabel* (in average coffee groves) and *Lage* (in average *estacas*). Due to their weight on the sample, these figures are very close to the 6,377 coffee groves cultivated in 6.18 *estacas* that correspond to the overall sample mean reported in Table 04. These numbers suggest that managers focused the allocation of labor to the larger colonies, while, at the same time, either demanded higher labor effort or accepted worse outcomes from colonies such as *Mariquita*, *Paineira*, and *Colônia Nova*, where households cultivated, on average, more than 8 thousand groves and 7 *estacas*. Finally, the smallest number of cultivated groves per household occurred in *Linha* and *Boa Esperança*, notably the colonies where only Non-white households resided (*cf.* discussion below).[8] Similarly, the fourth to last colony in terms of average number groves per household was *Colônia Ibicaba*, where Non-whites had their largest number of uniquely identified households.

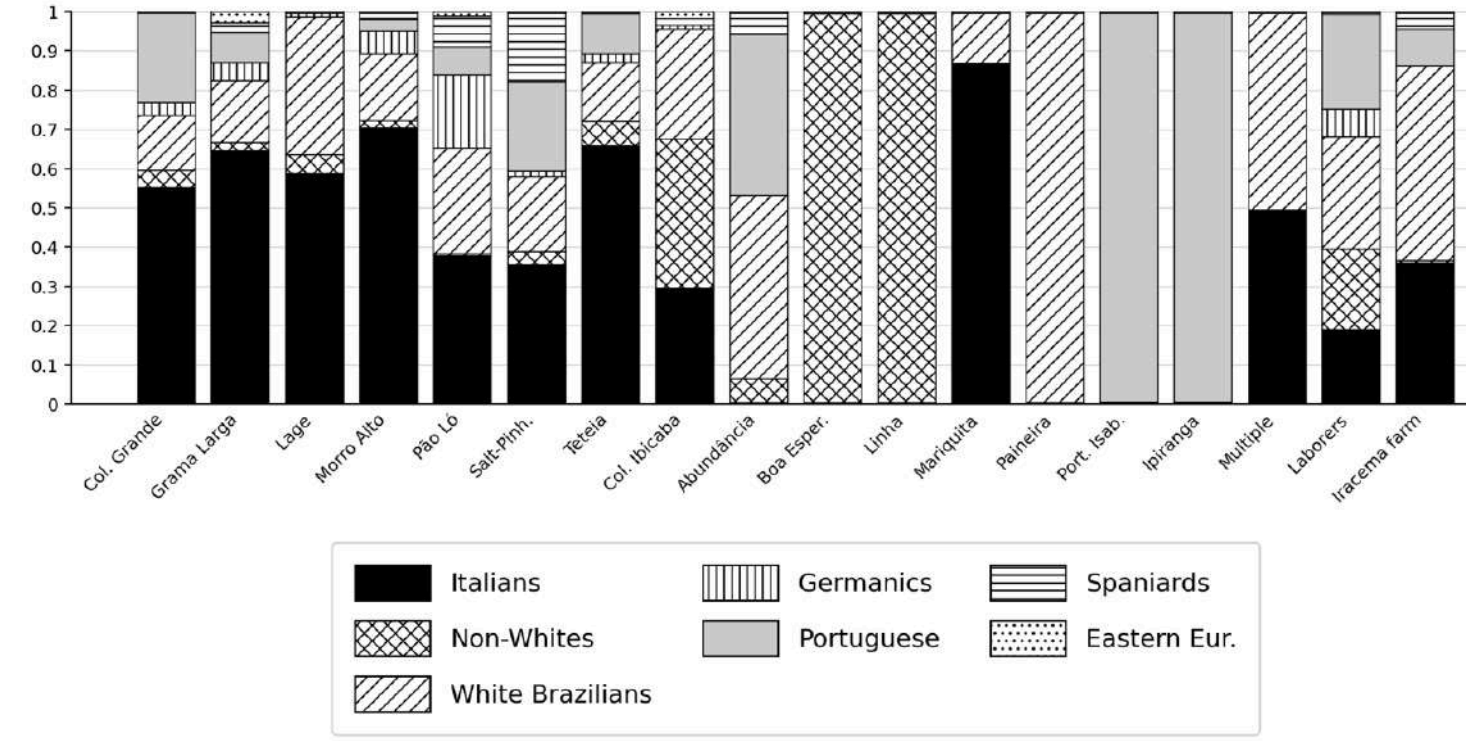
Given a number of confounding factors, these numbers should not be seen as indicative of racial discrimination (*cf.* Sections V and VI). In fact, there is no evidence supporting the hypothesis that the distribution of households within *Ibicaba* was based on ethnolinguistic criteria: Table 06 lists the number of observations and of uniquely identified households per colony according to ethnolinguistic origins; Figure 02 plots the share of each ethnolinguistic group per colony.[9]

The ethnolinguistic composition of *Ibicaba*, pooled for the period 1888-1958, is quite representative of the migratory history of the Age of Mass Migration in São Paulo from 1886 to 1927. Italians (and their identified descendants) constituted the overall majority of rural laborers, with 452 uniquely identified households, corresponding to 1,630 observations. Italians were the majority in *Colônia Grande*, *Grama Larga*, *Lage*, *Morro Alto*, *Teteia*, and *Mariquita* and were close to or above ⅓ of the identified households in *Pão-de-Ló*, *Saltinho & Pinheiro*, *Colônia Ibicaba*, and *Iracema farm*. The number of uniquely identified Portuguese (116 households, corresponding to 366 observations) and, in a distant third place, Spaniards (42 households, corresponding to 110 observations) are also aligned with the migratory history to the region (Levy 1974; Bassanezi *et al.* 2008; Bassanezi 2019). Finally, Germanic immigrants remained an important minority (36 uniquely identified households, corresponding to 137 observations), probably as a reflex of the migratory flows of German-speakers to São Paulo's central-western plateau since the mid-19th century, as pioneered by the German and Swiss sharecroppers of *Ibicaba* itself (*cf.* Section II).

[8] *Linha* had, however, just one uniquely identified household.

[9] Excluding observations with missing data for colony of work. As such, these figures are similar, but not equivalent to the ethnolinguistic composition of the dataset as a whole.

Figure 02 – Share of ethnolinguistic groups per colony in *Ibicaba*



Note: Shares reported in this table exclude observations with missing values for nationalities or colonies of work.

Table 06 – Colonies of coffee cultivation, by ethnolinguistic groups, pooled sample (1888-1942, w/ gaps)

	N	n	T-bar	N	n	T-bar	N	n	T-bar	N	n	T-bar	N	n	T-bar	N	n	T-bar	N	n	T-bar	N	n	T-bar
	White Brazilians			Non-Whites			Italians			Germanics			Portuguese			Spaniards			Eastern Europeans			Total		
Col. Grande	107	38	2.82	42	11	3.82	449	92	4.9	35	4	8.75	198	36	5.50	4	3	1.33	1	1	1.00	836	185	4.52
Grama Larga	119	56	2.13	12	6	2.00	495	120	4.1	37	7	5.29	57	22	2.59	22	6	3.67	22	5	4.40	764	222	3.44
Lage	24	13	1.85	3	1	3.00	46	17	2.7	1	1	1.00	0	0	.	0	0	.	0	0	.	74	32	2.31
Morro Alto	57	23	2.48	5	2	2.50	321	70	4.6	19	6	3.17	15	9	1.67	9	5	1.80	0	0	.	426	115	3.70
Pão Ló	43	15	2.87	0	0	.	54	20	2.7	27	4	6.75	10	5	2.00	11	2	5.50	1	1	1.00	146	47	3.11
Salt.-Pinh.	35	14	2.50	4	2	2.00	81	21	3.9	5	3	1.67	39	5	7.80	44	9	4.89	0	0	.	208	54	3.85
Teteia	15	10	1.50	4	3	1.33	42	19	2.2	2	1	2.00	11	7	1.57	1	1	1.00	0	0	.	75	41	1.83
Col. Ibicaba	10	7	1.43	16	10	1.60	106	69	1.5	5	4	1.25	3	3	1.00	14	11	1.27	10	10	1.00	164	114	1.44
Abundância	8	8	1.00	1	1	1.00	0	0	.	0	0	.	7	7	1.00	1	1	1.00	0		.	17	17	1.00
Boa Esperança	0	0	.	15	15	1.00	0	0	.	0	0	.	0	0	.	0	0	.	0	0	.	15	15	1.00
Col. Nova	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Linha	0	0	.	3	1	3.00	0	0	.	0	0	.	0	0	.	0	0	.	0	0	.	3	1	3.00
Mariquita	1	1	1.00	0	0	.	7	1	7.0	0	0	.	0	0	.	0	0	.	0	0	.	8	2	4.00
Paineira	1	1	1.00	0	0	.	0	0	.	0	0	.	0	0	.	0	0	.	0	0	.	1	1	1.00
Port. Isabel	0	0	.	0	0	.	0	0	.	0	0	.	3	1	3.00	0	0	.	0	0	.	3	1	3.00
Ipiranga	0	0	.	0	0	.	0	0	.	0	0	.	3	3	1.00	0	0	.	0	0	.	3	3	1.00
Multiple colonies	1	1	1.00	0	0	.	1	1	1.0	0	0	.	0	0	.	0	0	.	0	0	.	2	2	1.00
Laborers	19	19	1.00	17	17	1.00	10	10	1.0	6	6	1.00	15	15	1.00	1	1	1.00	0	0	.	68	68	1.00
Iracema farm	27	21	1.29	0	0	.	18	12	1.5	0	0	.	5	3	1.67	3	3	1.00	0	0	.	53	39	1.36
Total	467	227	2.06	122	69	1.77	1630	452	3.6	137	36	3.81	366	116	3.16	110	42	2.62	34	17	2.00	2866	959	2.99

Notwithstanding the prevalence of European immigrants and their identified descendants as *Ibicaba*'s major laborforce in the post-Abolition era, Table IV.5 further shows the large number of Non-Whites and White Brazilians in the sample (recall that White Brazilians are those not identified as non-Whites and those having a surname not attributable to other European ancestry than Iberian). With 227 uniquely identified households (corresponding to 467 observations), White Brazilians were the second most common ethnolinguistic group across colonies. Their numbers are not even disproportionally affected by the category "laborers", who sum up only 19 uniquely identified households. The data thus provide clear evidence that *colonato* arrangements were frequently signed by non-foreigners (for similar results, *cf.* Monsma 2010; Colistete 2024).

Non-whites, in turn, were the fourth most prevalent ethnolinguistic group across colonies, with 69 uniquely identified households.[10] As for White Brazilians, this ranking remains unaltered even if we discount the category "laborers", which, however, is more relevant to their numbers than to any other ethnolinguistic group. These numbers demonstrate that the *colonato* was not exclusively based upon immigrant labor and of their descendants. The Abolition in 1888 and São Paulo's expanding agricultural frontiers continuously pressured labor demand in the coffee economy. Therefore, it was natural that the non-white population would be employed as *colonos* in São Paulo, in the same way as they did as sharecroppers in Rio de Janeiro (Holloway 1980, p. 97; Stein 1986). At the same, Non-white households were only a fraction (7.19%) of the uniquely identified households. Even for the overall sample, in which their shares as uniquely identified households reach a potential maximum of 11.50%

Figure 02 and Table 06 confirm that colonies in *Ibicaba* were non-homogenous in their ethnolinguistic composition, even in regard to Italians. The apparent homogeneity displayed by some colonies is only an artifact, given that *Paineira* and *Portão Isabel* had only one uniquely identified household each (Brazilian and Portuguese, respectively); *Linha* had only two (Non-whites); and *Ipiranga*, only three (Portuguese). The only actual exception was *Boa Esperança* colony, where all 15 uniquely identified households in 1890, corresponding to 60 people, were Non-whites. Literally meaning *Good Hope*, that colony was founded in the post-Abolition era. Its productive characteristics will be explored in detail in Section VI.

IV.III. Sources of labor income in coffee production: Cultivation (Lean Season) & Harvesting aaa

The average household in the pooled sample cultivated 6,377 coffee groves, distributed in *ca.* 6 *estacas*, which were pruned between 5 and 6 times per agricultural season (mean = 5.42). At an average nominal price of 0.025 *mil-réis* per coffee grove cultivated, the overall mean nominal income during the lean season was 963.900 *mil-réis* per household. The average household also harvested coffee for almost four months, usually between mid-May and mid-August, which corresponded exactly to the peak harvest season, as noted in 1911 by the

[10] Uniquely identified Non-White households corresponded to 122 observations. This latter metric ranks them as the 5th most numerous group, instead. Non-Whites also had the smallest T-bar among all ethnolinguistic groups.

director of government-sponsored settlement colonies in the city of São Paulo.[11] At an average nominal price of 0.890 *mil-réis* per *alqueire*, the overall mean yield of 491.10 *alqueires* harvested per year earned households an overall mean nominal income of 364.900 *mil-réis* during the harvest season (Table 04). The pooled mean annual nominal income was thus 1,328.800 *mil-réis* per household.[12]

Per capita figures were much dimmer, however. With an overall mean of 6.65 people per household, the pooled annual labor income per capita falls to 199.910 *mil-réis*, or 269.260 *mil-réis* per person of working age (7-60y). These figures align well with other estimates of labor income in São Paulo during the first two decades of the twentieth century. Annual per labor and per capita production in government-sponsored rural settlements were between *ca.* 219.000 and 225.000 *milréis* in the 1910s-1920s (Rocha, Ferraz, and Soares 2017, p. 115; Witzel de Souza 2023, pp. 325-8). For the state of São Paulo as a whole in 1920, the wage of an unskilled agricultural laborer with add-ons for subsistence (*i.e.* similar to a *colono* who was allowed to cultivate a plot of land for subsistence) earned 1.196 milréis per day (Zamberlan Pereira 2020, p. 213 - Table A4); a value corresponding to 225 days of work from a person of working age in *Ibicaba.*

Figures 03 and 04 further explore households' sources of income in *Ibicaba* by plotting the number of coffee groves cultivated in the lean season and the volume of coffee harvested in the harvest season, respectively, for 1904-1942, when the bulk of the data for both agricultural seasons overlap. Both figures are presented per household and per person of working age.

Panel A of Figure 03 shows the number of coffee groves cultivated per household, averaged across all households with available data in year t. Except for minor fluctuations in the short-run, the average number of coffee groves annually cultivated per household increased from 5,054 in 1906 (having fallen from 5,137 and 5,188 in 1904 and 1905, respectively) to a maximum of 9,098 in 1929, before steadily declining to 4,263 in 1942.[13] Increases in the average number of coffee groves cultivated per household coincided with the coffee defense programs of 1906, 1917, and 1921, frequently with lags of 3-5 years, as further demonstrated by Delfim Netto (1979) with aggregate national series. By contrast, the 1929 Crash and the Great Depression gave slow, but continuous deadly blows to coffee production in *Ibicaba*. Starting in 1929, the average number of coffee groves cultivated per household declined

[11] The director Ernesto Kuhlmann was particularly interested in seasonal migration between subsistence agriculture and coffee production in that quadrimester (*apud* Holloway 1980, p. 130).

[12] For comparison, consider the following estimates compiled by Holloway (1980, pp. 84-6). In 1897, the State of São Paulo's Secretary of Agriculture estimated that a representative *colono* household with two economically-active people could cultivate 5,000 coffee groves and harvest 500 *alqueires*, earning a nominal annual income of 800.000 *mil-réis*. In 1901, the Italian owner of a plantation in Taquaritinga County estimated that 3 economically-active people could cultivate 4,000 coffee trees (earning them 320.000 *mil-réis* in the lean season) and harvest 480 *alqueires* (earning them 240.000 *mil-réis* in the harvest season). In 1903, the Italian Vice-Consul in Ribeirão Preto County estimated that 2 workers and one housewife could cultivate 5,000 trees (300.000 *mil-réis*) and harvest 450 *alqueires* (180.000 *mil-réis*). Finally, the 1903 estimates of the Department of Agriculture in the railroad *Mogiana Zone* expected four workers to cultivate 10,000 trees (800.000 *mil-réis*) and harvest 1,200 *alqueires* (600.000 *mil-réis*). For comparability, I excluded from these figures any earnings from subsistence agriculture, which I do not estimate in this paper.

[13] Interestingly, that observation for 1942 was comparable to the last figures of the Vergueiro's era. In 1890, households cultivated, on average 3,730 groves, having declined from 4,996 in 1888.

steadily throughout the 1930s. However, that number had been so high in 1929 that in 1936 the number of coffee groves cultivated per household, on average, had declined to the early 1920s levels. For the period 1904-42 as a whole, the ascension and decline in the number of coffee groves cultivated per household can be fairly described by a quadratic trend (adj-R2 = 0.594).[14]

Panel B plots the number of coffee groves cultivated by people of working age (7-60y). Trends are very similar to those of Panel A, except for the period 1906-1921. In this interval, the number of coffee groves cultivated per household increased, but the number figures per person of working age did not. In fact, differently from the highly significant positive and negative trends estimated for 1921-1929 and 1930-1942, respectively, there is no trend for the period 1906-1920. This implies that *Ibicaba*'s managers were hiring households with more people in working age, and/or that households working in the plantation were benefitting from a demographic gift in this period. From 1921 to 1929, the average number of coffee groves cultivated by people of working age rose from *ca.* 1,250 to a maximum of *ca.* 1,900. The decline thereafter was as steady as for the figures per household. In 1937, the number of coffee groves cultivated per person of working age had returned to its 1900s level, before reaching the depth of little more than 800 in 1939.

Figure 04 plots the mean annual volume of coffee harvested. Volume is measured in the contemporary *alqueires* metric, which corresponds to *ca.* 40L-volume of dry goods.

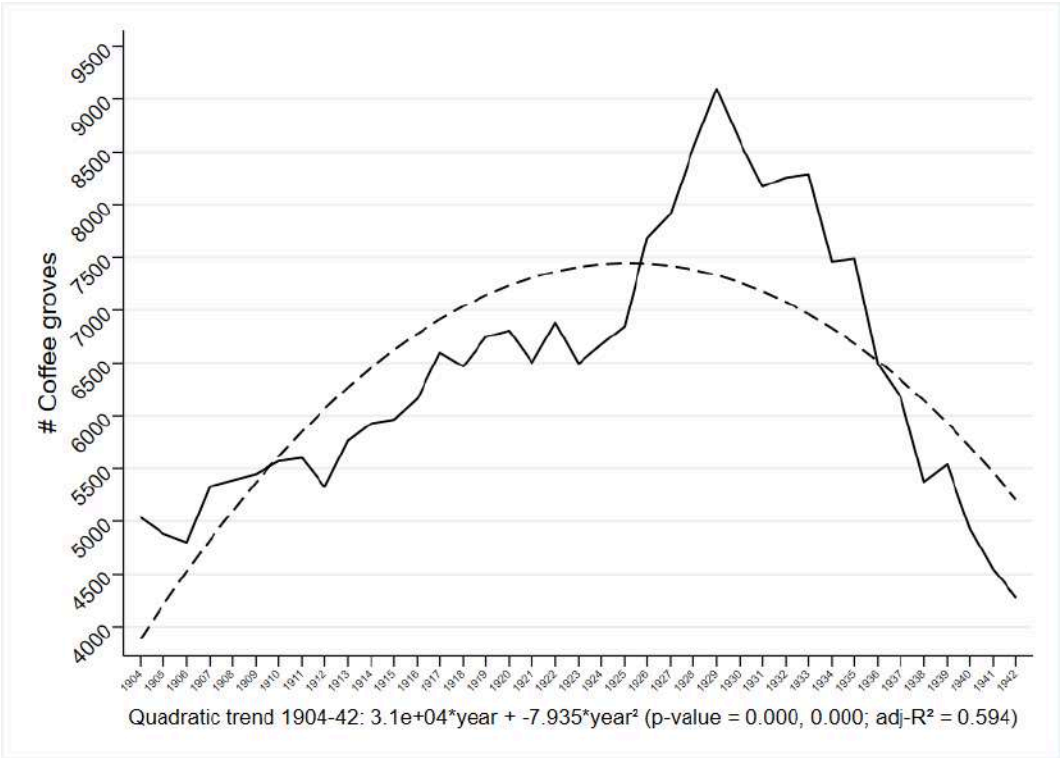
Panel A shows the average annual volume of coffee harvested per household in 1904-42.[15] The figure reveals the long-term decline of the coffee economy in São Paulo's Old West, with *Ibicaba* as a representative plantation. Combined with Panel A of Figure 03, this series indicates an overall productivity decline in *Ibicaba's* coffee production. Since we do not know the ages of the coffee groves, it is not possible to attribute such a decline to laborers' lower productivity or to the worsening of the plantation's agricultural capital (*cf.* Dean 1976, pp. 175, 180). However, Figures 03 and 04 combined show that the mean amount of coffee harvested per household was falling simultaneously to increases in the number of coffee groves cultivated per household in 1906-1929. This was not a transient phenomenon attributable to the planting of new trees (with lower productivity in the first years) or to exogenous shocks, such as the super-crop of 1906 or the Great Frost of 1918, which can all be visualized in the series, but do not explain the trend. In fact, a highly significant linear trend shows that, in every year between 1904 and 1942, the average household was harvesting *ca.* 15 *alqueires* less than in the preceding year (adjusted-R2 = 0.36). This trend cannot be solely attributable to the Great Depression. If the series is split in 1929, the trends for the sub-periods 1904-1929 and 1930-1943 are statistically equivalent. Exactly the same conclusions are reached with the volume harvested per person of working age (Panel B).

[14] The model that predicts 1929 as the year with maximum coffee cultivation has the sample restricted to the period 1917-1942. Available upon request.

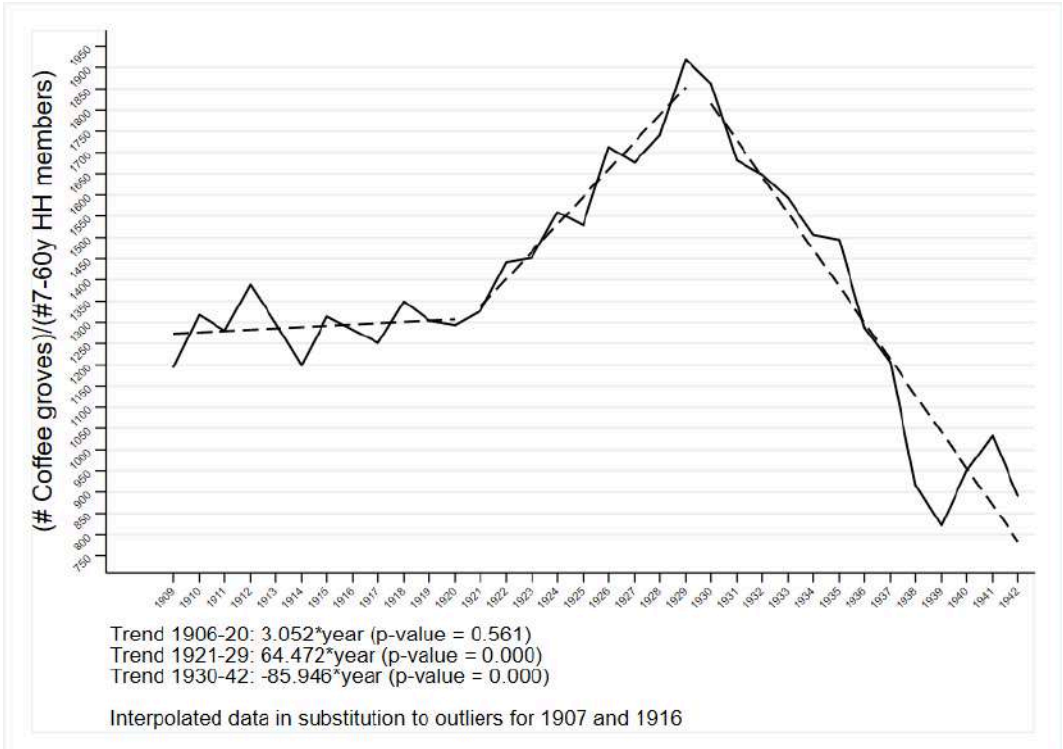
[15] The same series for the period 1889-1958 is available upon request. The full sample can be well approximated by a quadratic trend, with 1912 as the estimated peak year of harvest per household, very much in line with Dean's (1976) estimates for Rio Claro (SP) County.

Figure 03 – Number of coffee groves cultivated in the lean season (1904-42)

Panel A – per household

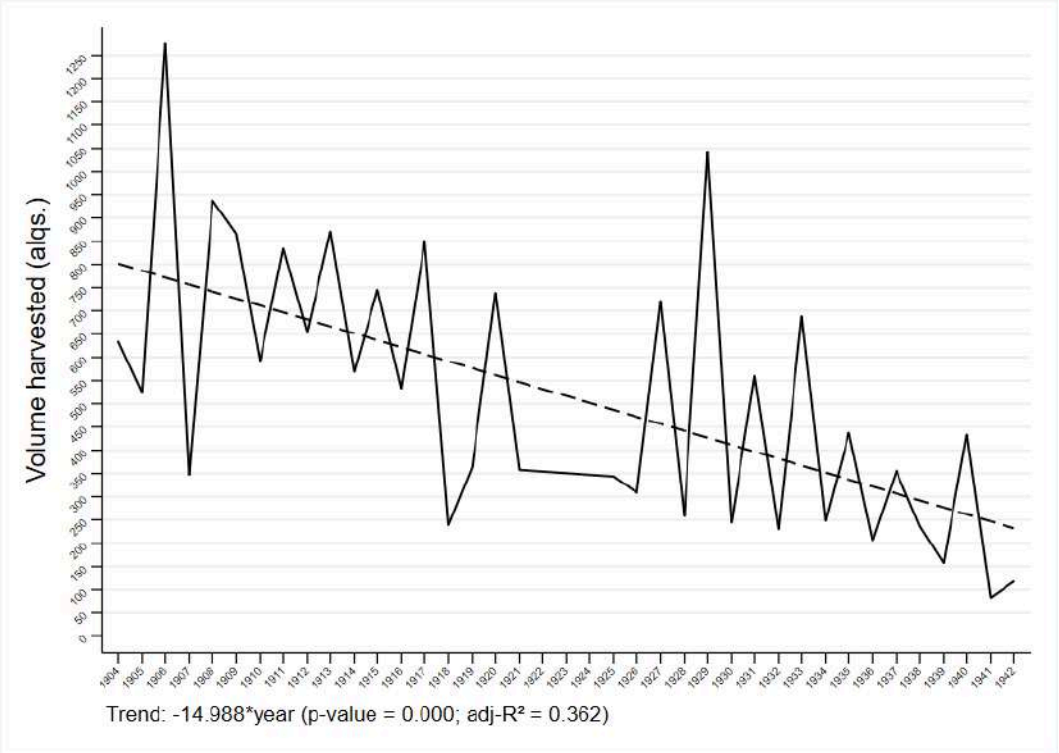


Panel B – per person of working age (7-60y)



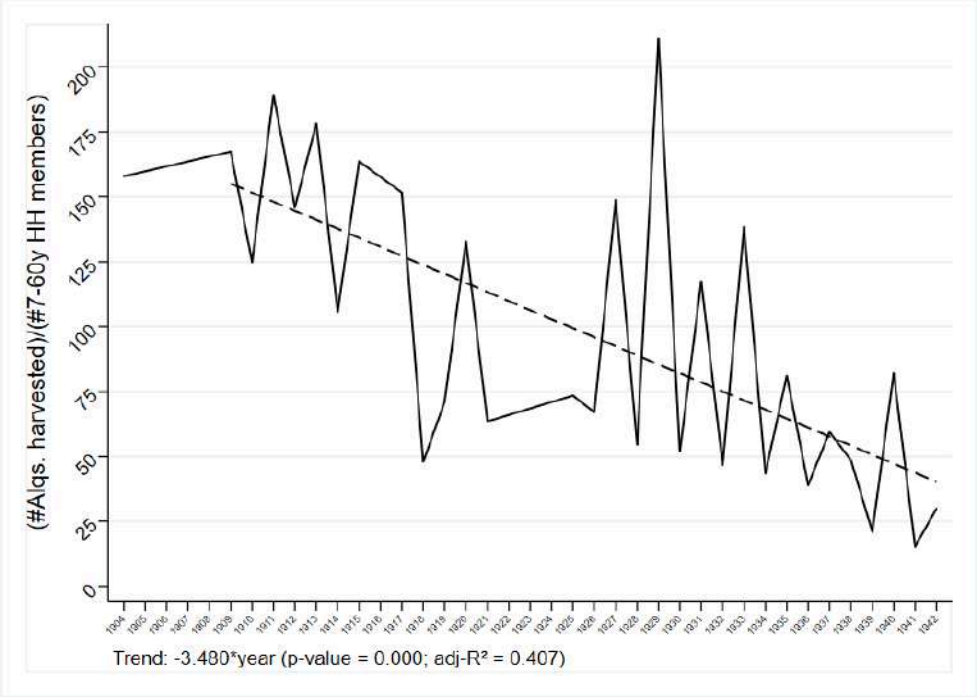
Note: Data for production per person of working age are interpolated for years 1906 and 1917, as the original data show unreasonable spikes, probably due to undercounting of the working population in those years.

Figure 04 – Volume of coffee harvested (*alqueires*) in the harvest season (1904-42)



Panel A – per household

Panel B – per person of working age (7-60y)



Notes: (1) 1 *alqueire* (volume) = 40L of dry goods. (2) Data for harvest per person of working age are interpolated for years 1906 and 1917, as the original data show unreasonable spikes, probably due to undercounting of the working population in those years.

V. Labor income and Discrimination in the Brazilian coffee economy

This section tests the hypothesis of labor discrimination (positive or negative) between White Brazilians, Non-Whites, Italians, Germanic, Portuguese, Spaniards, and Eastern Europeans (including their descendants) with microdata from *Ibicaba* from the immediate post-Abolition to the 1950s. It starts by estimating the determinants of physical output in coffee production. The goal is to decompose the price, quantity, and factor elements of households' nominal income. An estimation of interest to the agricultural history of the coffee economy, this first step disentangles the partial effects of price incentives, frequency of executed tasks, agricultural capital, and households' demographic composition on households' labor income in the lean and harvest seasons. In the sequence, the income profiles of the various ethnolinguistic groups working in *Ibicaba* are computed and broken down by season (lean and harvest) and estimated at per capita levels. Finally, previous estimates are used to decompose the income differentials by households's observable characteristics. Conclusions are robust to the usage of twofold Oaxaca-Blinder decompositions and simple difference-in-means tests that do not require the estimation assumptions of the former. The section concludes by explicitly testing the hypothesis of taste-based discrimination in prices earned by different ethnolinguistic groups. The physical productivity of households is obtained from an estimation of their physical outputs in the lean and harvest seasons and then controlled for in regressions estimating prices earned in the lean and harvest seasons.

v.i. Determinants of Physical Output in coffee production (Lean and Harvest Seasons)

Households were the historical unit upon which *colonato* labor relations were set (*cf.* Section II). As such, all econometric analyses that follow have the household as the unit of analysis. Single individuals eventually working in agricultural tasks per piece rate or as daily laborers will be treated as single-person households. The temporal dimension of the panel, in turn, will be the calendar year. Moreover, the labor income of *colono* households was composed of a fixed parcel from coffee cultivation in the lean season and a varying parcel from the volume of coffee harvested during the harvest season (*cf.* Section IV).

As such, let us thus define Y_{ht}^{output} as the nominal labor income of household h in year t: $Y_{ht}^{output} = (Y_{ht}^{cult}, Y_{ht}^{harv})$, in which Y_{ht}^{cult} and Y_{ht}^{harv} refer to the household's income from coffee cultivation and harvesting, respectively. In all estimations that follow, Y_{ht}^{output} will be estimated twice: Separately as income from coffee cultivation during the lean season (Y_{ht}^{cult}) and as income from coffee harvesting (Y_{ht}^{harv}) in the harvest season. Y_{ht}^{total} refers to household's total nominal income, which is simply the sum of these components: $Y_{ht}^{total} = Y_{ht}^{cult} + Y_{ht}^{harv}$.

Nominal labor income is equivalent to $Y_{ht}^{output} \equiv p_{ht} * Q_{ht}$, in which Q_{ht} is the physical output produced by household h under price p_{ht}, negotiated with the household for the corresponding output in in year t. Physical output (Q_{ht}), in turn, depends on the frequency (N_{ht}) of executed agricultural tasks and the household's production function, $f(.)_{ht}$, for that specific agricultural task: $Y_{ht}^{output} \equiv p_{ht} * N_{ht} * f(.)_{ht}$ Considering the production of a single, homogenous crop with agricultural capital provided by the plantation, $f(.)_{ht}$ will be modeled by a standard Cobb-Douglas production function. The following baselined specification thus estimates the physical output in coffee production, netting out prices and frequencies of tasks from labor income (Foster, Haltiwanger, and Syverson 2008):

$$ln\left(Y_{ht}^{output}\right) = \beta_0 + \beta_1 ln\left(p_{ht}\right) + \beta_2 ln\left(N_{ht}\right) + \left(ln\left(L_{ht}\right)\right)'\delta + K_{ht}'\Gamma + \alpha_h + \lambda_t + \epsilon_{ht},$$

in which ϵ_{ht} is an i.i.d. disturbance and Y_{ht}^{output} is always estimated twice, separately for Y_{ht}^{cult} and Y_{ht}^{harv}.

N_{ht} refers to the frequency of executed agricultural tasks. For Y_{ht}^{cult}, N_{ht} is the number of times the household was engaged (and paid for) to cultivate the coffee groves during the lean season. For Y_{ht}^{harv}, N_{ht} is the number of months a household was engaged in (and paid for) harvesting.

$\left(L_{ht}\right)'$ refers to household's labor. The baseline controls for the number of household members of any age (household size) and the number of individuals in the age range 7-60 years old (economically-active household members), independent of sex .[1]

$\left(K_{ht}\right)'$ refers to the agricultural capital made available by plantation managers to the households, which is proxied by a set of categorical indicators for the colony and *estacas* therein worked by the household. Each *estaca* indicates the exact location of groves within the plantation, and, therefore, controls for common characteristics shared by all coffee groves in that smallest unit of rural planning.[2] Beyond controls for colony of work, the baseline further controls for *estacas* identifiers' mean, standard deviation (SD), and mean deviation from the mean (MDEV).[3] These metrics of central tendency and dispersion provide a uniquely precise proxy for the specific coffee groves that households cultivated in the lean season and income thereupon derived (Y_{ht}^{cult}). Controls for agricultural capital in the harvest

[1] *Cf.* robustness checks for sensitivity analyses of age windows and sex. Gender norms have been hypothesized as relevant to differentiate the economically-active population of black and Italian families in the post-Abolition era. Stein (1986) and Bassanezi (2019) show anecdotal evidence that, in the immediate post-Abolition, black men abrogated the employment of women in the fields.

[2] Varying soil quality, topography, distance from colonies etc. might all have influenced the price paid for attending to and harvesting different *estacas* (Dean 1976, p. 168).

[3] *Cf.* robustness checks for alternative metrics of central tendency and dispersion for the location of *estacas*.

season (Y_{ht}^{harv}), however, are somewhat less obvious. As discussed in section IV, there are two opposite historical scenarios that should be taken into account. In one extreme, households could have harvested exactly the same coffee groves they cultivated in the lean season (à *Vergueiro system*), in which case the controls for capital would be identical to those in Y_{ht}^{cult}. In the other extreme, plantation managers could have identified the most mature berries across the plantation as a whole and allocated laborers to harvest those specific coffee bushes. All coffee groves would have been pooled together within the plantation and laborers would have been randomly allocated to those bushes that could be optimally harvested. In this scenario, the agricultural capital of the harvest season would be the pool of all coffee groves in *Ibicaba* as a whole. From the household's viewpoint, $K_{ht}^{'}$ would only correspond to the timing of that household's engagement in harvesting: If brought in too early or too late, the household would have harvested a suboptimal volume of coffee. This scenario can be proxied by controlling for the first, last, and mean month when the household was employed in harvesting.[4] Finally, an intermediary, likely more plausible, scenario is that households were given priority to harvest the same coffee groves they had themselves cultivated in the lean season, with extra laborers brought in only if necessary. This is the scenario modeled in the baseline for Y_{ht}^{harv}, with the specification controlling for colony of work, metrics of central tendency and dispersion of *estacas*' identifiers, and additional variables for the timing of the harvest.

Household fixed effects (α_h) capture all their time-invariant characteristics, including cultural beliefs, which might be related to their origins, such labor ethics, thriftiness, gender norms etc. Obviously, α_h also captures ethnolinguistic origins and skin colors themselves. Finally, the baselines include year fixed effects (λ_t) to account for common shocks (agroecological and economic) affecting all households in *Ibicaba.*

Baseline results are reported in Table 07.

Price elasticities were highly significant, even with controls for year-FEs. During the lean season, the point estimate (1.285) is statistically larger than one. Because the baseline specification controls for the frequency of executed tasks - in the lean season, number of prunings -, the larger-than-unit estimate implies that households responded more than proportionally in their labor efforts to changes in prices earned. For the harvest season, the highly significant point estimate (1.313), however, is not statistically different from one.

The frequency of executed tasks was also significant at the 1% level and economically meaningful. A household hired to execute one more pruning than that of the overall sample

[4] In 1890-1958 (with gaps) the average household spent approximately 4 months harvesting coffee (s.d. ≈ 1 month). A typical harvest season started in May and ended in September, with April and October being the earliest and latest months ever recorded. For households with available data, 95.91% worked for at least two months in harvesting; 91.10%, 77.81%, 54.50%, and 17.38% did so for at least three to six months, respectively.

mean would have increased its nominal income by 21% in the lean season.[5] It should be recalled, however, that only five prunings per year were registered or estimated in the sources prior to 1921.[6] For subsequent years, however, the results show that a household's capacity to contract and execute more prunings played a major role in their annual income. The effect of being engaged for longer in harvesting was even larger. A household working one month longer in harvesting than the overall sample mean would have increased its nominal labor income in the harvest season by 174%.[7] At the 95th percentile, households were engaged for up to five months in harvesting per year. The very top of the distribution, in turn, was engaged for six months in total. In other words, being engaged for longer in harvesting was exceptional, but it really paid-off economically.

Estimates show the statistical and economic significance of the share of economically active members as a determinant of output per task executed in each agricultural season. Significant at the 1% level, *ln(#members 7-60y)* implies an elasticity of physical output to household's labor input of 0.34 during the lean season and of 0.40 during the harvest season. These point estimates are very robust across alternative specifications and refinements in households' composition (*cf.* Section VI; *cf.* further Fogel and Engerman 1971; Schmitz and Schaefer 1978; Olmstead and Rhode 2008). By contrast, household sizes had no significant effect beyond that of members in the age range 7-60. Very young and relatively old individuals thus neither hampered the ability of households to allocate their economically-active members to market-oriented tasks, nor contributed to the elasticity of physical output derived in rural markets.[8] As noticed by Luiz Tonelotto, a Brazilian-born *colono* of Italian origins who worked in *Ibicaba* in the 1950s-1960s, "[…] landowners wanted to hire those families that had lots of people to work, [because] at that time there was no machinery: It was all manual labor, so they needed lots of people" (Tonelotto in Witzel de Souza and Gardenal, p. 357).

Finally, the timing of engagement in harvesting was a relevant determinant during the harvest season. *Ceteris paribus*, households that started to harvest earlier in the season had an economic disadvantage, most likely due to more labor effort required to pick between mature and green coffee berries. However, the partial effect of increasing the mean month of engagement in harvesting is negative and significant at the 1% level, suggesting competition for best bushes between the latecomers.[9]

[5] Evaluated at the sample mean = ((1 + 1/5.42418)^1.139 - 1)*100. Notice that this effect is not driven by inflation, as the specifications control for prices earned per pruning and year-FEs.
[6] In the nineteenth century, 3-4 prunnings were executed per agricultural season in that region (Dean 1976, p. 36).
[7] Evaluated at the sample mean: ((1+ 1/3.937352)^4.447 - 1)*100.
[8] In the pooled sample, maximum ages recorded at the 75th, 90th, 95th, and 99th percentiles were, respectively, 59, 68, 73, and 82 years old.
[9] All regressions control for the number of months a household was engaged in harvesting. Therefore, controls for the first, last, and mean months worked in harvesting capture solely the timing of the task, not its frequency.

Table 07 – Determinants of Physical Output in coffee production (Lean and Harvest Seasons)

	(1)	(2)
	$\ln(Y_{ht}^{cult})$	$\ln(Y_{ht}^{harv})$
ln(price)	1.285***	1.313***
	(0.0836)	(0.350)
ln(N)	1.139***	4.447***
	(0.252)	(0.634)
ln(HH size)	0.0174	-0.0996
	(0.0485)	(0.0846)
ln(# members 7-60y)	0.336***	0.401***
	(0.0398)	(0.0704)
Mean month harv.		-1.668***
		(0.642)
1st month harv.		1.740***
		(0.481)
Last month harv.		-0.0649
		(0.174)
Observations	2,358	1,470
Number of id	512	324
HH-FE	Yes	Yes
Year-FE	Yes	Yes
Colony: Lean season	Yes	Yes
Proxies for location of *estacas*		
Mean	Yes	Yes
MDEV	Yes	Yes
SD	Yes	Yes

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "ln(price)" corresponds to price earned for cultivating one coffee tree during the lean season weighted by number of coffee trees cultivated under each price (= "Price cult." in the summary statistics) and price earned by *alqueire* harvested (= "Price harv.") (4) ln(N) refers to the frequency of executed agricultural tasks, *i.e.* number of times per year a HH pruned the coffee groves (= "#pruning per season") in the lean season, and number of months the HH was engaged in harvesting (= "'#months harv.") during the harvest season.

v.ii. Estimated income profiles by ethnolinguistic origins: Household's annual monetary income

The total annual labor income of a household in a certain year can be estimated from the predicted values of the baseline specification: $\hat{Y}^{total}_{ht} = \hat{Y}^{cult}_{ht} + \hat{Y}^{harv}_{ht}$.[10] Averaged across observations for each ethnolinguistic group, these estimates provide the income profiles of Non-Whites, White Brazilians, and Europeans (jointly with their identified descendants) working in *Ibicaba*'s coffee economy. Table 08 reports such estimates for the pooled sample and for each ethnolinguistic group.[11] Figure 05 plots the Kernel Density Estimates (KDEs) for $\hat{Y}^{total}_{ht}$ for each ethnolinguistic group, always contrasting them to the distribution of the pooled sample (solid line) and of White Brazilians (dashed line).

The annual income of a mean household in the pooled sample was 1,471.839 *mil-réis*, only slightly above the mean income of 1,328.900 *mil-réis* reported in the descriptive statistics for the pooled sample. This difference is explainable by the fact that Table 08 reports the results for households that worked in both seasons, while the descriptive statistics report data for households that could have worked only in the lean or harvest seasons.

Income profiles by ethnolinguistic groups, in turn, show that Non-white households were the poorest by far and by any metric. White Brazilians performed above all other groups, both at the mean and median and at the right tail of the distribution, except for the highest predicted income, which belonged to an Italian household. Households from different European origins showed rather distinct profiles. Portuguese and Germanics had metrics of central tendency and KDEs closer to those of White Brazilians, while Italians and Spaniards were comparatively worse-off.

The mean annual income predicted for Non-White households was 664.254 *mil-réis* (median = 599.201 *mil-réis*), which corresponded to 45% of the mean income predicted for the sample and to 47% of Italian households, *i.e.* immigrants with the lowest predicted mean (excluding the single Eastern European household). The KDE for non-whites was strongly concentrated around the median, leading to a high kurtosis.[12] Some few Non-White households escaped this trap of low income, as shown at the right tail of the distribution, but these were exceptional. In contrast, White Brazilian households were the best-performing, with a predicted annual income of 1,786.994 *mil-réis* (median = 1,597.017 *mil-réis*). Their KDE is the closest to a normal distribution, but showed a heavy right tail and a mode much more to the right than that of the pooled sample. In short, White Brazilian households were doing well, both at the center of their distribution *vis-à-vis* that of other ethnolinguistic groups, as well as among best performers.

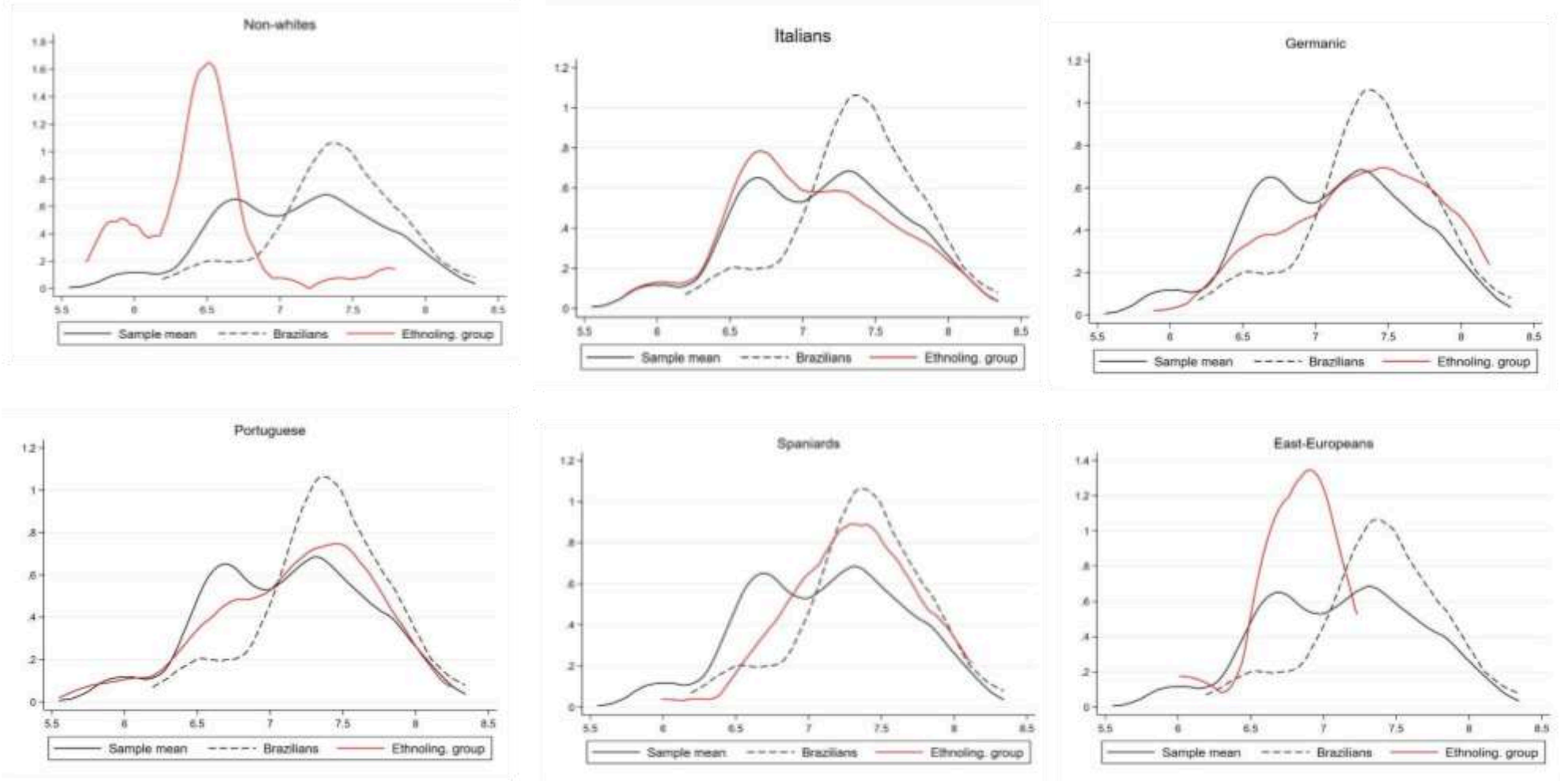
[10] The number of observations for (Ytotal)_hat is limited to that of (Yharv)_hat. Estimates of the determinants of labor income and discrimination in the lean season, however, use the full set of observations allowed for by (Ycult)_hat.
[11] The statistics and KDE for Eastern Europeans are reported, but they refer to one single household.
[12] This result cannot be attributed to the number of uniquely identified Non-white households, given the correspondingly smaller numbers for the Germanic and Portuguese households.

Table 08 – Predicted total income (in *mil-réis*), by ethnolinguistic group

	N	n	Mean	S.D.	Min	Max	Median	Skew.	Kurt.
White Brzs.	200	63	1786.994	882.091	360.123	5382.275	1597.017	1.499	5.884
Non-whites	46	18	664.254	428.596	230.564	2473.867	599.201	3.174	13.36
Italians	886	180	1401.688	875.686	296.905	5430.222	1159.930	1.601	5.905
Germanic	78	14	1763.865	958.203	389.058	4657.936	1567.096	1.049	3.94
Portuguese	195	38	1515.916	853.569	246.917	4928.570	1426.474	1.334	5.329
Spaniards	53	10	1669.917	854.917	434.073	3987.827	1421.802	1.145	3.713
Eastern-Eurp.	11	1	962.728	294.634	439.424	1552.611	934.569	0.339	3.073
Pooled	1469	324	1471.839	888.486	230.564	5430.222	1299.580	1.459	5.475

Notes: (1) All predicted values from the baseline specification; (2) Total labor income (in *mil-réis*) is the sum of labor income from the lean and harvest seasons; (3) N = observations; n = number of uniquely identified HHs.

Figure 05 – Kernel Density Estimates (KDEs) of total predicted labor income (lean & harvest seasons), by ethnolinguistic groups



Total income distribution among Italian households was very distinct from any other group. Its peak close to the sample mean, followed by a slowly declining slope to the right, indicates the existence of different socioeconomic subgroups among them. In fact, Italian households are the only ones to show statistical evidence of a non-unimodal distribution (at the 10% significance-level).[13] The predicted annual incomes thus reveal that some Italian households fared better than the sample mean, but others did fall short of the average household in the sample. Rather surprisingly *vis-à-vis* the historiography, Italian households as a whole performed worse than White Brazilians in *Ibicaba*'s coffee economy.[14] Similar conclusions stem from the fact that Italian households had the lowest mean predicted income among all immigrants (1,401.688 *mil-réis*; median = 1,159.930 *mil-réis*);[15] at the same time, however, the maximum income predicted for any group accrued also to an Italian household.

The much smaller group of Spaniards fared comparatively better. Among immigrants, their mean predicted annual income of 1,669.917 *mil-réis* was second only to the Germanics', while their median predicted income (1,421.802 *mil-réis*) was only slightly below that of the Portuguese. The right tail of their KDE is very heavy, showing that some few Spaniard households earned high annual incomes, even beyond those of White Brazilians. Interestingly, however, their maximum income was much smaller than those of all other immigrant groups, leading to a more homogenous distribution than other immigrants in the sample. In contrast, the predicted annual income distributions for Germanic and Portuguese households were much more spread than those of any other ethnolinguistic group, showing a larger variation in their income profiles. Furthermore, Germanic households had the largest predicted mean annual income (1,763.865 *mil-réis*) among all immigrant groups and tracked White Brazilians very closely. The Portuguese, in turn, had the third highest mean income (1,515.916 *mil-réis*) and the second highest median (1,426.474 *mil-réis*).

The income profiles of the Germanic and Portuguese households, on the one hand, and of Italian and Spaniard households, on the other, align well with the migratory histories of these nationalities, more broadly. Their more spread KDEs and the convergence of Germanic and Portuguese households to the mean of White Brazilians can be attributed to their longer migratory histories to São Paulo's coffee economy. German-speakers and Portuguese were in fact the main nationalities working as sharecroppers in the labor arrangements proposed by *Vergueiro & Co.* since the 1840s. In addition, they were two of the oldest nationalities immigrating to Brazil since its political independence – the Portuguese for obvious reasons related to the Brazilian colonial past and the Germans for various migratory experiments before the Age of Mass Migration. The more compact distribution for Spaniard households is suggestive of greater homogeneity within a more recent immigrant group. Finally, the income profiles of the Italians reflect their numbers as the most numerous nationality working under *colonato* contracts. Their KDE's pronounced peak, large kurtosis, and the fact that Italian households had simultaneously the lowest mean, but the highest maximum predicted income of any group strongly suggest a gradual splitting of that ethnolinguistic group between

[13] Results available upon request and in the reproduction files that accompany this paper.
[14] The skewness and kurtosis of the Italian KDE was larger than that of any other group, except for Non-whites.
[15] Excluding the single Eastern European household.

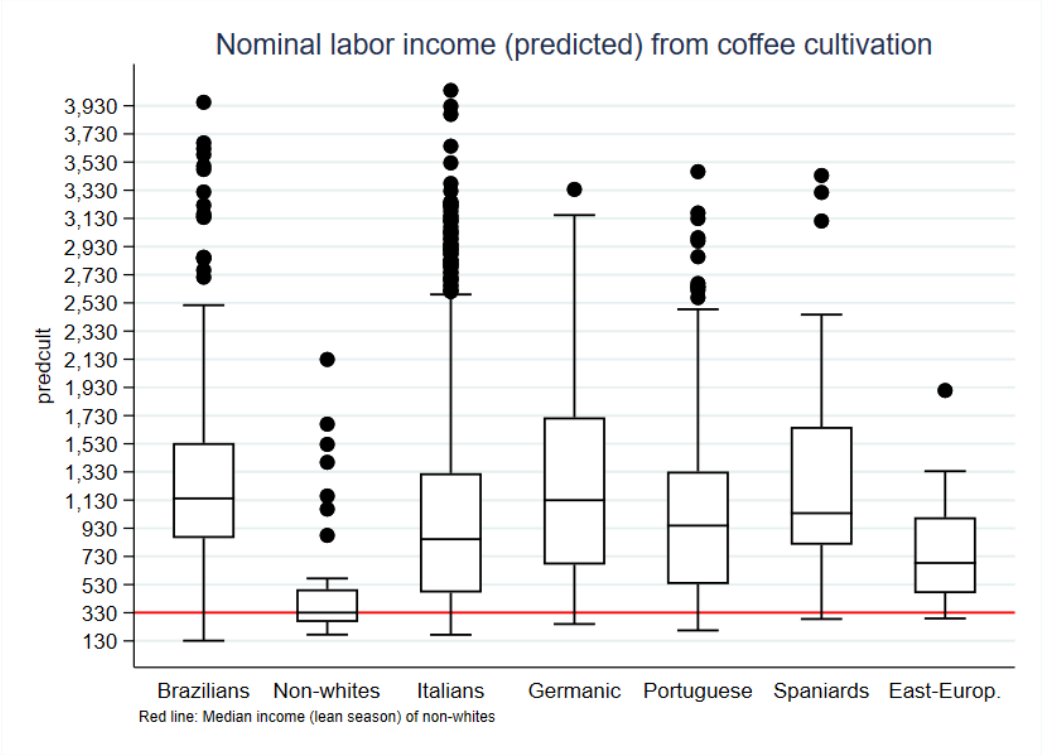
high-earning households and a subgroup the historiography knows much less about, namely those who failed to make it to the (South-)American dream.

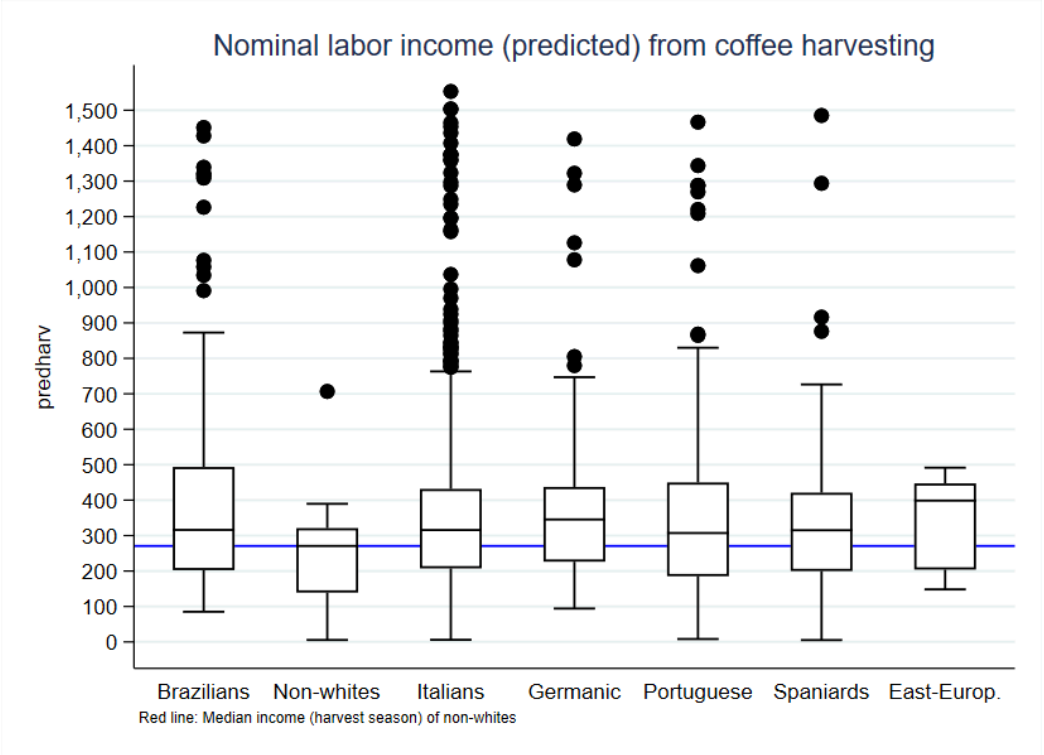
v.ii.i. Heterogeneities per agricultural season & figures per capita

In order to disentangle households' performance in the distinct agricultural tasks of pruning (and weeding, hoeing, and cleaning) from harvesting, I now separate the labor income earned by each ethnolinguistic group per agricultural season. Figure 06 provides the corresponding boxplots, with horizontal lines reflecting the median of Non-White households.

In general, the income profiles predicted in each season align with the previous conclusions. Non-white households had the lowest median income of all groups in both seasons. Some outliers far outperformed the group during the lean season, but even these outliers were beneath the maximum predicted for all other ethnolinguistic groups. In the harvest season, Non-white households had only one outlier and that, again, was beneath all other groups' maximum predicted incomes. Also in line with previous conclusions, Italian households showed very dispersed predicted income in both seasons, with comparatively low medians, but many outliers. Figure 06 also shows some nuances that could reflect households' decisions in allocating labor to different agricultural tasks. The most astonishing finding is the extremely low income predicted for Non-white households during the lean season, when their 3rd quartile was only as high as that of Italian households'1st quartile Non-white households were also the poorest group during the harvest season, but their position was comparatively less bad. In short, Figure 06 reveals that Non-white households had the lowest predicted income for all ethnolinguistic groups in any season, but their worst outcomes were in the lean season. By contrast, Italian households performed much better comparatively to other groups in the harvest season than in the lean season. While Italian households had a median predicted income lower than that of any other group except for Non-white households during the lean season, their median predicted income during the harvest season was second only to Germanic households. In harvesting, Italians at the median outperformed even White Brazilian households, *i.e.* the group with the highest total annual income.

Figure 06 – Boxplots of predicted income per household by agricultural task, per ethnolinguistic group



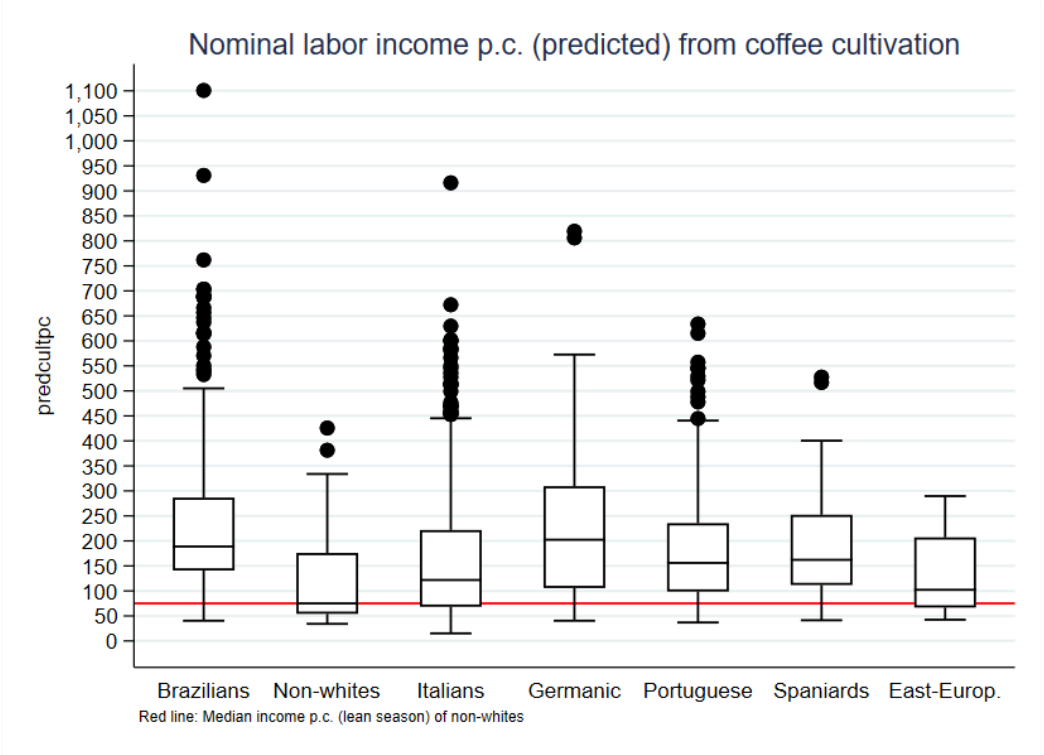


As thoroughly discussed below, households' size and age composition were among the main burdens of slavery for Non-whites in *Ibicaba*. Households headed by Non-whites had 4.11 members, on average, compared to 5.87 and 6.94 members in households headed by White Brazilians and Europeans (and their descendants), respectively; households headed by Italians reached the impressive mean of 7.19 members. To account for such large differences, Figure 07. presents the previous boxplots, now with figures per capita, while Table 09 ranks the groups by their mean and median income per capita during the lean and harvest seasons.

The income per capita profiles observed during the lean season tell a very consistent story with the figures per household discussed before. Irrespective of household sizes, the median income per capita of Non-whites during the lean season was only as high as the Italian's 3rd quartile per capita. However, a surprising picture emerges for the harvest season. Albeit at a very low level, the median income per capita predicted for Non-whites here superseded that of any other ethnolinguistic group. Such results are confirmed by Table 09. While Non-whites had the lowest mean and median income per capita of any group in the lean season, they reached the 3rd highest mean income per capita during the harvest season – and the very top of the ranking with the median. Unsurprisingly, given the sizes of their households, Italians ranked third to last and last in their mean and median income per capita during the harvest season. In other words, while household size did not alter the relative positions of Non-whites and Italians during the lean season, it did so during the harvest season. Finally, during the lean season, white Brazilians and Germanics showed the highest income per capita (with their position to one another reversed at the mean and median). Spaniards, Portuguese, and Italians followed with a stable ranking.

These results raise the possibility that the low-income profile of Non-whites could have been a rational response in their allocation of labor between agricultural seasons. Non-white households could have made use of their comparative advantage in harvesting to accumulate an amount of income per capita that they considered sufficient for the entire year. The rationale for Italians was exactly the opposite. The household as a *unit* had to compensate in the harvest season for the household's comparatively lower performance in the lean season. Their large households also implied that, in per capita terms, Italians could not specialize in the tasks of one season at the expense of the other.

Figure 07 – Boxplots of predicted income per capita by agricultural task, per ethnolinguistic group



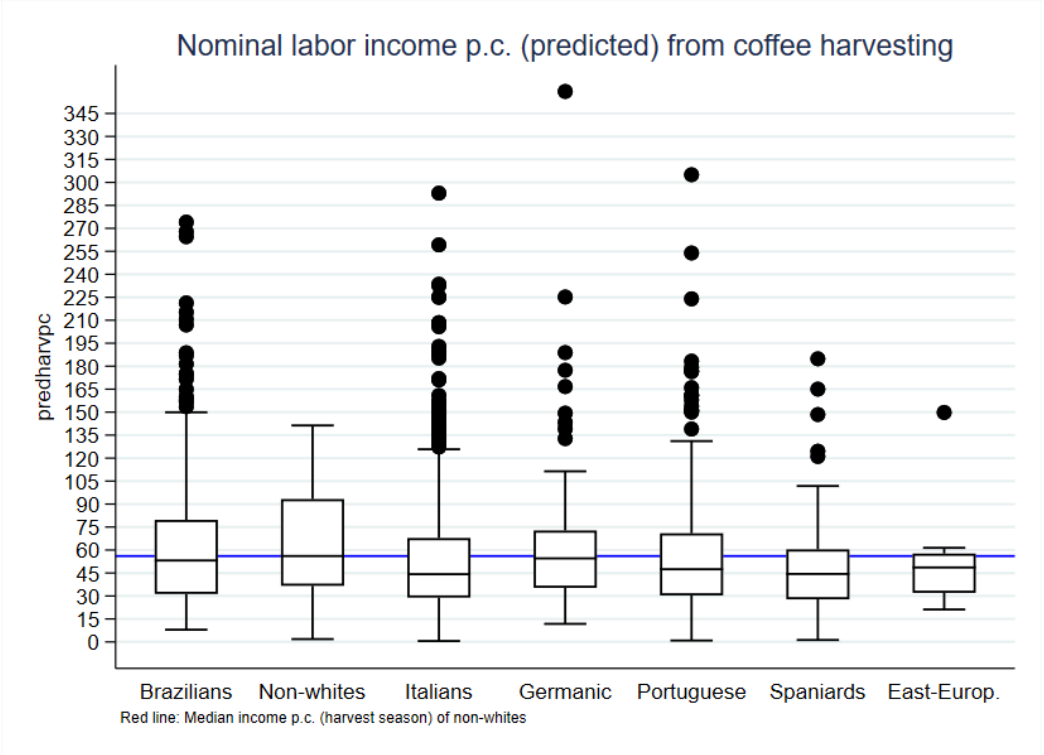


Table 09 – Ranking of ethnolinguistic groups by mean and median income per capita (in *mil-réis*)

Income per capita – Coffee cultivation					
Mean				Median	
1	White Brz.	234.223	1	Germanic	202.310
2	Germanic	225.123	2	White Brz.	188.783
3	Spaniards	185.065	3	Spaniards	162.168
4	Portuguese	178.694	4	Portuguese	155.965
5	Italians	156.062	5	Italians	121.804
6	East. Europ.	136.689	6	East. Europ.	102.206
7	Non-whites	126.633	7	Non-whites	75.046
	S.D.	38.278		S.D.	42.883
Income per capita – Harvest					
Mean				Median	
1	Germanic	67.870	1	Non-whites	56.061
2	White Brz.	67.066	2	Germanic	54.538
3	Non-whites	66.875	3	White Brz.	53.159
4	Portuguese	60.326	4	East. Europ.	48.614
5	Italians	54.562	5	Portuguese	47.455
6	East. Europ.	52.477	6	Spaniards	44.374
7	Spaniards	51.261	7	Italians	44.213
	S.D.	6.780		S.D.	4.480

Notes: (1) S.D. computed with all data from each ethnolinguistic group. (2) Number of observations for each ethnolinguistic category = [N_Coffee cult; N_Harv]: Brazilians = [382; 200]; Non-whites = [69; 47]; Italians = [1376; 886]; Germanic = [114; 78]; Portuguese = [312; 195]; Spaniards = [86; 53]; East. Europ. = [19; 11].

V.III. DECOMPOSITION OF THE SOURCES OF INCOME DIFFERENTIALS

The evidence discussed so far showed that Non-whites had the lowest overall mean income of all ethnolinguistic groups in both agricultural seasons. By contrast, households headed by White Brazilians had the highest overall predicted income, followed closely by the Germanic and Portuguese. Italian households, in turn, had the most widespread income distributions, which is attributable to their numbers and, possibly, more recent migratory histories. Finally, differences between figures per capita and per household highlight not only the relevance of households' size and age composition in determining their productive capacity, but also suggest the possibility of some groups choosing to specialize in different agricultural tasks. Hence, demographics, rational choice, and discrimination are the main, non-mutually exclusive explanations for differences in these groups' income profiles. These explanations are well-aligned with the baseline results, which showed the centrality of households' labor force composition (an endowment), of the frequency of households' engagement in agricultural tasks (a choice variable of managers *and* laborers), and of corresponding prices (a potential source of discrimination) in determining households' labor income. Before testing for discrimination in prices in the next section, the current section uses two methods to decompose the sources of income differentials between groups. The first is a standard Oaxaca-Blinder (OB) decomposition. Given the strictness of assumptions required to interpret the OB decompositions with a large number of categorical variables, I then report simple differences-in-means tests for the independent variables used in the baselines, in order to investigate how each ethnolinguistic group differed from the rest of the sample.

V.III.I OAXACA-BLINDER (OB) DECOMPOSITION

The OB decompositions follow from an LSDV model analogous to the baseline, which controls for categorical variables for colony of work, and for year- and household-identifiers. The first categories were omitted for each categorical variable and the errors were clustered at the household level. The coefficients thus estimated are identical to those of Table 07. and preferable to the inconsistent RE estimates, as shown by the Hausman test.[16] Given the unbalanced nature of the panel, the option *relax* was required for all estimates that follow.[17]

All decompositions reported below refer solely to cross-sectional differences between groups pooled for the entire sample. They do not take into account dynamic changes within each group or between groups over time, for two reasons. Econometrically, the identification of the model is so heavily dependent on categorical variables, including FEs, that re-scaling them for dynamic comparisons would generate estimates so complex to the point of meaninglessness. Historically, beyond Abolition itself, I see no substantial reason to set *a priori* specific thresholds after which differences in endowments and coefficients should vary in any systematic manner between groups.

[16] Available upon request.

[17] This option is necessary if coefficients in the reference or in the comparison model are dropped due to collinearity or if there are coefficients with zero variance, both of which occur in the estimates below due to year- and household-identifiers.

The twofold OB decompositions reported below assume the existence of a set of non-discriminatory parameters, which are estimated using the pooled sample (Neumark 1988 apud Jann 2008, p. 5). These decompositions use a single set of coefficients for all groups and compare each group to the sample as a whole, *i.e.* including that specific group in the estimation (Jann 2008, pp. 5, 12, 16-7).[18] For the purposes of this paper, twofold decompositions were preferred to the threefold because differences in coefficients computed by the latter must take into account differences in the intercept, which in the LSDV model just pools together all omitted categories.[19] Tables 10 and 11 report the twofold OB decompositions for the lean and harvest seasons, respectively. Results with different reference groups are also discussed below.

OB decompositions: Lean season

Table 10 shows that non-whites were the only ethnolinguistic group with a mean income significantly smaller than that of the pooled sample. In line with previous findings, the difference was not only statistically significant, but sizeable: 335.963 *mil-réis* for non-white households against 882.713 *mil-réis* for the pooled average. Households headed by white Brazilians, in turn, had a higher income during the lean season *vis-à-vis* the sample mean, but the difference was only marginally significant at the 10% level. The other ethnolinguistic groups did not show statistically significant overall differences with respect to the sample mean. Italian, Germanic, and Spaniard households had mean incomes in the lean season that were above the corresponding sample means, while Portuguese households had a lower mean. However, the magnitudes of the explained and unexplained components are either not sufficiently large to render the overall difference significant, or the different signs of the components cancel each other out.

White Brazilians had an advantage over the sample mean in their explained component, which roughly corresponds to their endowments. This advantage stemmed from higher prices earned for cultivating coffee groves and a larger number of prunings. Because non-white and Italian households suffered a penalty in their endowments of these two variables, the positive effect for white Brazilians cannot be attributed to explicit racism. The explained advantage of Brazilians in prices earned cannot be attributed to language skills (which could have allowed them to negotiate better contracts), either: first, because non-whites, who were all Brazilians at this point, had a significantly negative impact stemming from this explained component; second, because the Portuguese did not perceive any endowment advantage in these variables. The unexplained component, in turn, played against white Brazilians, with the share of individuals in the age range 7-60y as the main contributor to such negative effect.

[18] Stata 17©, command *oaxaca*, option *pooled* (Jann 2008, pp. 5, 12, 16-7).
[19] In fact, the threefold OB decompositions for the harvest season estimates "endowment" components that are mostly non-significant, favoring instead the so-called "coefficient" components. This is at odds with results found for the explained components in the twofold OB decompositions.

Table 10 - Twofold Oaxaca-Blinder decomposition: Labor income from coffee cultivation - LSDV model w/ pooled data (Reference: Pooled sample)

	$\ln(Y_{ht}^{cult})$											
	White Brazilians		Non-Whites		Italians		Germanic		Portuguese		Spaniards	
Pool Sample	6.731***		6.783***		6.746***		6.749***		6.761***		6.748***	
Ethn. group	6.876***		5.817***		6.760***		6.866***		6.711***		6.927***	
Difference	-0.145*		0.966***		-0.0141		-0.117		0.0499		-0.178	
Explained	-1.373***		1.577***		-0.0907		0.833***		0.126		0.452**	
Unexplained	1.228***		-0.611***		0.0766		-0.950***		-0.0766		-0.631***	
	explained	unexplained	explained	unexplained	explained	unexplained	explained	unexplained	explained	unexplained	explained	unexplained
lpricewcult_tree	-0.231***	-0.761	0.475***	2.555	0.147***	0.972	-0.157	-0.0937	-0.0680	-0.321	-0.128	-6.704**
lncarpa	-0.0702***	-1.265	0.0893***	1.850***	0.0439***	-0.169	-0.0292	3.276	-0.0129	0.722	-0.0390	0.438
lhhsize	0.00215	-0.307	0.00826	0.0898	-0.00265	0.0672	0.00105	-0.290	0.00105	-0.161	-0.00150	0.0829
lenxadan07	0.0400	0.429**	0.183***	-0.0206	-0.0396*	-0.181	-0.00530	0.0107	0.00121	-0.0146	-0.0136	0.619***
estaca_mean	-0.000141	0.0112	0.0000288	0.860	0.000271	0.0224*	-0.000300	-0.00447	-0.000312	-0.0900	-0.0000573	6.228***
estaca_mdev	-0.00169	-0.106	-0.000571	-0.257	0.000633	-0.0575	0.00122	0.00667	0.00179	0.0667	-0.00384	2.320
estaca_sd	0.00167	0.00267	0.000521	0.0197	-0.000564	0.0547	-0.00129	-0.0173	-0.00188	-0.0465	0.00386	-2.940
coltree_cat_2	0.00510	-0.00152	0.0256	0.190	-0.0140	0.0268	0.0000486	0.0867	0.0219	-0.0474	0.00656	-2.575
coltree_cat_3	-0.00731	-0.00816	-0.00349	0.0106	-0.00155	0.00284	0.00513	0.00684	0.00803*	-0.000478	0.00723*	-0.000115
coltree_cat_4	0.00406	-0.00522	0.0134	0.0502	-0.0156	0.00971	0.00160	0.0119	0.0198*	-0.0150	0.00903	-0.685
coltree_cat_5	-0.0113	-0.0211	0.00828*	-0.000200	0.00961	0.00159	-0.0251	0.0723	0.00502	0.000411	-0.0148	0.0225
coltree_cat_6	0.000310	-0.00870	0.00846	-0.00336	0.00988	-0.0105	0.00612	0.0204	-0.00711	0.00222	-0.0641	-7.047
coltree_cat_7	-0.000179	-0.0912	-0.00685	0.0228	-0.00142	-0.000576	0.00170	0.00511	0.00222	-0.00115	0.00540**	-0.000220
coltree_cat_8	0.00206	0.00303	0.0106	-0.000514	-0.00147	-0.000174	-0.00104	-0.0000066	-0.00114	-0.000107	-0.00103	-0.0000376
coltree_cat_12	-0.00126	0.00125	0.00934	0.0108	-0.000156	-0.000159	0.000530	-0.0000099	0.0000375	-0.0000260	0.000524	0.00000407
coltree_cat_13	0.000263	-0.000433	0.000995	0.0000105	-0.00116	0.0000237	0.00101	0.0000278	0.00111	-0.0000168	0.00100	-0.0000167
coltree_cat_14	0.000361	-0.000461	-0.0000602	-0.0000012	-0.000140	0.0000743	-0.0000614	-0.0000043	-0.0000674	-0.0000044	-0.0000607	0.00000822
coltree_cat_15	0.000657	-0.0000082	0.000567	0.00000089	0.00132	0.0000962	0.000579	-0.0000038	-0.00416	-0.000807	0.000572	0.00000077
coltree_cat_19	0.00614	-0.00168	-0.00247	0.0000888	-0.00241	-0.00126	-0.00252	0.000204	-0.000793	-0.0387	0.00137	-0.281

Notes: (1) N = 2358; (2) Estimates based on LSDV with categories for colony of work, year, and panel identifier (household); (3) All categories have been included; household IDs, time dummies, and omitted colonies not reported to save space (available upon request and retrievable with the publicly-available .do file; (4) Standard errors clustered at the panel identifier (household ID); (5) Reference group is always the pooled sample, for all estimations; (6) Option "relax" required due to the unbalanced nature of households over time.

Table 11 - Twofold Oaxaca-Blinder decomposition: Labor income from harvesting - LSDV model w/ pooled data (Reference: Pooled sample)

	$\ln(Y_{ht}^{harv})$											
	White Brazilians		Non-Whites		Italians		Germanic		Portuguese		Spaniards	
Pooled sample	5.700***		5.696***		5.554***		5.681***		5.695***		5.690***	
Ethnol. group	5.557***		5.223***		5.765***		5.684***		5.585***		5.443***	
difference	0.144		0.473***		-0.211**		-0.00293		0.111		0.247	
explained	-0.250		0.657**		0.0458		-0.959**		-0.543*		-0.935**	
unexplained	0.393		-0.184		-0.257		0.956***		0.653*		1.182***	
	explained	unexplained	explained	unexplained	explained	unexplained	explained	unexplained	explained	unexplained	explained	unexplained
lpriceharv	-0.292***	0.0312	0.428**	-3.321	0.185**	-0.300	-0.148	-0.538	-0.115	0.786*	-0.168	5.225
lnmonth	0.105	9.663***	-0.0934	3.423**	-0.213**	-0.263	-0.119	-28.58**	0.294*	-6.156***	0.467	616.6***
monthmean	0.0706	-32.71**	0.0750	-12.49**	-0.0455	-11.88	0.0846	-12.17**	0.0159	-7.850*	-0.131	-14.18**
month01	-0.180	20.21**	0.0205	8.288**	0.209*	4.511	-0.0372	-16.39*	-0.236*	-5.049	-0.101	515.3***
late	-0.00106	1.971	0.00657	4.094	0.00420	7.636	0.00517	4.969	-0.00760	16.84***	-0.0140	-842.5***
lhhsize	-0.00635	0.0814	-0.0554	0.695	0.0138	0.132	-0.00563	0.0292	-0.00677	-0.486	0.00499	-1.375
lenxadan07	0.0202	-0.0621	0.239***	-0.276	-0.0354	-0.394	-0.0110	-0.182	0.000115	0.519**	0.00138	0.945
estaca_mean	0.00206	-0.637	0.000411	1.088	-0.00282	0.0186	0.00445	0.00612	0.00146	-0.310	0.00180	180.1***
estaca_mdev	0.00131	0.686	0.0000674	1.338	-0.000521	-0.233	-0.00316	2.770	-0.000779	0.520	0.00540	-39.91
estaca_sd	-0.00178	-0.911	-0.000160	-1.529	0.000602	0.219	0.00454	-2.765	0.00122	-0.455	-0.00738	41.97*
coltree_cat_2	0.00377	0.000377	0.0142	0.130	-0.00659	-0.0592	-0.00237	0.0615	0.00779	-0.0300	0.00821	0.0152
coltree_cat_3	-0.000345	-0.0176	-0.000141	0.00152	-0.0000604	-0.00714	0.000274	-0.000124	0.000299	0.00103	0.000269	-0.000211
coltree_cat_4	0.000664	-0.0576	0.00541	0.00681	-0.00765	-0.0120	0.00287	-0.0242	0.0100	0.0309	0.00602	28.11
coltree_cat_5	-0.00404	0.0725	0.00272	0.000274	0.00301	-0.0179	-0.00915	0.167	0.00245	0.00214	-0.00494	28.93
coltree_cat_6	0.00117	-0.528	0.00842	0.0128	0.0178	-0.0165	0.0185	0.000206	-0.0149	0.00939	-0.115	-69.95**
coltree_cat_7	-0.000152	-0.0914	-0.00115	-0.0000556	0.000546	-0.000977	-0.000290	-0.00231	-0.000123	0.0111	-0.00116	-0.000196
coltree_cat_8	-0.00113	0.00261	-0.0102	-0.0566	0.00123	0.00108	0.000713	-0.0000458	0.000778	-0.0000004	0.000700	-0.000120
coltree_cat_12	0.00238	-0.00411	-0.0526	0.0526	-0.00359	0.0105	-0.000163	-0.000137	0	-0.00171	-0.000160	-0.000368
coltree_cat_13	0.000691	0.0000396	0.000616	0.00000709	-0.000990	0.00000931	0.000630	0.0000511	0.000688	-0.0000921	0.000619	-0.0000293
coltree_cat_15	-0.000253	0.00000357	-0.000225	0.0000377	-0.000549	0.000242	-0.000230	-0.0000583	0.00165	-0.00831	-0.000226	0.0000522
coltree_cat_19	-0.000376	-0.0000230	-0.000335	-0.0000026	0.000539	-0.000157	-0.000343	0.00000604	-0.000374	0.0000537	-0.000337	0.0000188

Notes: (1) N = 1470. All other notes identical to those in Table V.03.A.

Non-whites had statistically significant and large overall explained disadvantages, which were only partially compensated by overall unexplained components. In regard to the explained components of the decompositions, Non-white households had significantly worse endowments in prices earned for cultivating coffee, in the share of 7-60y individuals in the household, and in the number of prunings (in this order of magnitude). It would be incorrect to attribute all these endowment disadvantages to racial discrimination, since Italians faced an equally significant disadvantage in their explained components related to prices earned and number of prunings, although the absolute disadvantage of Non-whites was larger. In regard to the unexplained components of the decompositions, Non-whites had a major disadvantage in the number of prunings. Table 10 shows that Non-whites faced an endowment *and* an unexplained penalization in the frequency with which their households were engaged in pruning coffee in the lean season. If we interpret the explained difference as a choice variable (the number of prunings that Non-whites were willing to negotiate with plantation managers), then the additional difference in unexplained component might be attributable to discrimination, *i.e.* negative variation non-attributable to an endowment chosen by that ethnolinguistic group.

In line with previous evidence, Table 10 shows no statistically significant individual component that differentiated Germanic and Portuguese households from corresponding sample means during the lean season, which further suggests their longer integration into São Paulo's rural economy.

Additional analyses use White Brazilians (only), all Brazilians (Whites and Non-whites), and Italians as the reference group, respectively (available upon request). The first aims to compare White Brazilians to their Non-white compatriots and to European immigrants and their descendants, thus distinguishing more clearly between skin color and nationality. The second focuses on differences between Brazilians of all skin colors and various immigrants. The third compares White Brazilians, Non-white Brazilians, and other immigrants to Italians, the most numerous group of *colonos*.

Having White Brazilians (only) as the reference group, the significance of the overall explained and unexplained components vanishes for the Non-whites. The individual variables contributing to the lower mean income of Non-white households *vis-à-vis* White Brazilians, however, remain mostly unaltered, with one important exception: the negative explained contribution of the share of individuals 7-60y becomes statistically non-significant. In other words, the demographic endowment disadvantage of Non-white households was not due to difference with respect to White Brazilians. At the same time, Italian households showed a highly significant advantage in their unexplained component referring to members 7-60y. This suggests that the internal organization of Italian households, beyond sheer quantity of economically active members, gave them a positive edge over households headed by White Brazilians. If all Brazilians (Whites and Non-whites) are the reference group, this demographic advantage of Italian households becomes obviously even more evident. Both the explained and unexplained components referring to members 7-60y are highly significant in favor of the Italians. Finally, if Italians are the reference group, Non-whites become the only households with a significantly smaller difference in mean income during the lean

season. *Vis-à-vis* Italians, Non-white households had endowment disadvantages in prices earned and in the number of household members 7-60y. However, and confirming results from table 10, households headed by White Brazilians showed explained advantages over Italians in price earned and number of prunings.

In conclusion, during the lean season, households headed by White Brazilians had the highest average income. Their main advantages accrued to endowments in prices earned and number of prunings. In contrast, Non-whites were the only group with a statistically significant lower mean income in the lean season, independent of the comparison group. Their disadvantage in prices earned raises a red flag about discrimination, but Italians faced similar disadvantages, even if at lower magnitudes. More conclusively, the number of prunings imposed a heavier toll on the mean income of Non-white households than on any other ethnolinguistic group; the significant explained *and* unexplained components suggest discriminatory practices against them. Finally, Italian households enjoyed a demographic advantage that was especially strong in comparison to the endowments of Non-white households and to the unexplained component of households headed by White Brazilians. Large, well-functioning (in their internal division of labor), and young households provided Italians with an advantageous head-start in the Brazilian coffee economy. The same *colono* mentioned above summarized precisely, in practice, the empirical results found here: "You asked me whether it was hard to save money? Well… It depended. My father sometimes made some money ["ganhar um dinheirinho"], because the family was numerous [and] worked a lot [...]" He reported that the entire family was able to earn a good sum, but then highlighted: "But make no mistake: One entire year… [and] the entire family working in 10 people [...]", further reflecting on the dilution of the total earned in per capita terms (Tonelotto in Witzel de Souza and Santin Gardenal 2021, pp. 356).

OB decompositions: Harvest season

Table 11 reports the twofold OB decompositions for the harvest season.[20] In it, Italian households had the largest mean income (866.966 *mils-réis*), with a positive difference to the sample's mean significant at the 1% level. Non-white, in turn, had the lowest estimated income per household. Their mean of only 185.490 *mil-réis* was below the corresponding sample mean at the 1% significance level. No other ethnolinguistic group differed, in a statistically significant manner, from the overall sample means.

Households headed by White Brazilians showed an explained advantage in prices earned during the harvest season, which adds to the same effect found for the lean season. The evidence for racial discrimination in pay, however, is again weak, since both Non-white and Italian households faced an explained disadvantage in that same variable, although the absolute disadvantage of Non-whites was the largest of any group.

[20] It should be noticed that some point estimates for the frequency and timing of harvest among the Spaniards are unrealistically large in the decompositions, at times implying contributions far larger than the biggest outliers in the sample. These cases must be considered with caution.

Similarly to the lean season, the negative explained effect of household members 7-60y was significant at the 1% level. This demographic burden was not observed for any other group and consistently impaired Non-white households in both agricultural seasons. The unexplained component of the number of months engaged in harvesting added yet another disadvantage to the Non-whites. Differently from the lean season, however, it is difficult to attribute this effect to discrimination, as White Brazilians also had a statistically significant (at the 1% level) disadvantage in the unexplained component of that same variable.[21]

Finally, Italians enjoyed mostly an explained advantage in the number of months engaged in harvesting. *Vis-à-vis* the sample mean, however, they did not enjoy the same demographic advantage as during the lean season.

Having White Brazilians (only) as the reference group, overall mean differences drop in statistical significance and become marginally significant (10% level) only for Non-whites (negative) and Italians (positive). Non-white households faced an explained disadvantage in prices *vis-à-vis* White Brazilians even more severe than with respect to the sample mean, in both cases significant at the 5% level.[22] Non-white households also had an explained disadvantage in the number of members 7-60y, but with more weakly so (10% level). If all Brazilians are the reference group, only Italian households show a statistically significant higher mean income. The main individual contributors for differences between groups are the unexplained components related to the timing of harvest. All other differences are significant only at the 10% level. Among the latter, Italian households showed a marginally significant disadvantage in the explained component of prices earned. As expected, this result is statistically weaker and economically much smaller than if only White Brazilians are in the reference group. Nonetheless, these results reinforce that Europeans did not benefit from positive discrimination in prices. Italian households also showed a marginally significant explained advantage in the number of members 7-60y in comparison to all Brazilians; because such difference was not observed in relation to White Brazilians only, this marginal explained difference can be attributed to the relative demographic advantage of Italian households *vis-à-vis* Non-whites. This result is confirmed if Italians are the reference group, as Non-white households were the only ones facing an explained disadvantage in the number of members 7-60y. This demographic component is, in fact, the only variable for which the explained difference is statistically significant between these groups. Similarly to previous analyses, most variation between groups is attributable to the unexplained components of the timing of harvest. Finally, White Brazilians had a marginally significant explained advantage in prices *vis-à-vis* Italians, while the Portuguese had a marginally significant disadvantage – again suggesting no systematic positive discrimination in prices by ethnolinguistic origins.[23]

In conclusion, during the harvest season, Italian households had the highest average income. Their main advantages were in the length of their harvest seasons (relative to the sample mean) and the number of economically-active members (particularly relative to Non-White

[21] The magnitude of that effect on the mean income, |9.663| log-points, is unrealistically high: it would correspond to an effect as large as 162% of the highest income ever registered in the sample.
[22] Italian and Portuguese households faced similar disadvantages, but their estimates were significant only at the 10% level and corresponded to only 31% and 41% of the effect estimated for the Non-whites, respectively.
[23] All further estimations are available upon request.

households), variables that might have been historically correlated. The opposing signs for the decomposition of prices earned by different European nationalities do not indicate any form of positive discrimination in their favor. Similarly to the lean season, Non-Whites had the lowest level of income per household during the harvest season. Reverse mirroring the Italians, Non-White households' main disadvantages accrued to the number of household members 7-60y and timing engaged in harvesting. This result further strengthens an explanation based on the complementarity between the number of economically active members in a household and hiring for the harvest season, in line with the hypothesis of statistical discrimination.

v.iii.ii. A simpler approach: Differences-in-means tests

A potential limitation of OB decompositions is their sensitivity to the choice of reference groups for categorical variables (Jann 2008, p. 9). This could explain the unreasonably large estimates for the timing of the harvest, for instance. To circumvent such an issue, this section takes one step back and computes simple differences-in-means tests (Table 12) for all independent variables used in the baselines, assuming unequal variances between groups. By comparing the mean of these variables for each ethnolinguistic group to the mean of the pooled sample without that group, all differences between them are attributed to their endowments in the variables considered.

Results are reassuringly similar to the OB decompositions for the explained components, especially for prices and demographic characteristics. Both Non-white and Italian households earned lower prices than the corresponding sample means in the lean and harvest seasons (significant at the 1% level), further indicating that variations in prices earned were not caused (at least not exclusively) by racial discrimination, although the absolute disadvantage of the Non-whites was larger. Moreover, households headed by White Brazilians, Germanics, Portuguese, and Spaniards earned higher prices than the corresponding sample's means in both seasons. This suggests that variations in prices earned were not caused (at least not exclusively) by xenophobia. The fact that Spaniards earned a higher-than-average price in the harvest season further excludes the possibility that these results reflect a longer integration process of Germanic and Portuguese immigrants.[24] In addition, households headed by White Brazilians and Non-whites had significantly smaller households, in general, and fewer members in the age range 7-60y. Italian households, in turn, had both larger households and more members in the age range 7-60y than the corresponding sample's means.[25]

[24] Differently from non-significant results for the lean season, Eastern Europeans earned lower-than-average prices than the remaining sample's means in the harvest season.
[25] At the 10% level, Germanic households also had more members in the age range 7-60y, while Eastern Europeans had fewer than the corresponding sample means.

Table 12 – Differences-in-means tests of observable characteristics by ethnolinguistic groups (pooled sample)

	obs1	obs2	Mean1	Mean2	Diff.	S.D.	p value
			lpricewcult				
White Brz.	2513	446	-3.821	-3.626	-0.195	0.021	0.000
Non-whites	2830	129	-3.776	-4.14	0.364	0.025	0.000
Italians	1299	1660	-3.764	-4.14	0.049	0.016	0.002
Germanic	2825	134	-3.799	-4.14	-0.172	0.037	0.000
Portuguese	2591	368	-3.8	-4.14	-0.073	0.023	0.002
Spaniards	2866	93	-3.796	-4.14	-0.161	0.04	0.000
East Europ.	2935	24	-3.79	-4.14	0.117	0.073	0.121
			lpriceharv				
White Brz.	3601	431	-0.316	-0.168	-0.148	0.022	0.000
Non-whites	3735	297	-0.286	-0.472	0.185	0.018	0.000
Italians	2340	1692	-0.26	-0.354	0.094	0.015	0.000
Germanic	3884	148	-0.303	-0.207	-0.097	0.03	0.002
Portuguese	3601	431	-0.309	-0.218	-0.09	0.024	0.000
Spaniards	3935	97	-0.302	-0.204	-0.099	0.034	0.005
East Europ.	3997	35	-0.298	-0.47	0.172	0.041	0.000
			lncarpa				
White Brz.	2507	445	1.677	1.738	-0.06	0.005	0.000
Non-whites	2828	124	1.69	1.619	0.07	0.004	0.000
Italians	1292	1660	1.701	1.676	0.024	0.004	0.000
Germanic	2818	134	1.685	1.716	-0.031	0.008	0.000
Portuguese	2585	367	1.685	1.699	-0.014	0.005	0.006
Spaniards	2859	93	1.685	1.728	-0.042	0.01	0.000
East Europ.	2928	24	1.687	1.64	0.048	0.015	0.003
			lnmonth				
White Brz.	4160	462	1.329	1.313	0.017	0.017	0.323
Non-whites	4256	366	1.329	1.315	0.014	0.018	0.449
Italians	2642	1980	1.274	1.398	-0.123	0.01	0.000
Germanic	4461	161	1.326	1.375	-0.05	0.022	0.028
Portuguese	4122	500	1.337	1.248	0.089	0.018	0.000
Spaniards	4506	116	1.33	1.244	0.086	0.036	0.021
East Europ.	4572	50	1.328	1.304	0.024	0.047	0.612

Table 12 (ctd.)

	obs1	obs2	Mean1	Mean2	Diff.	S.D.	p value
			monthmean				
White Brz.	4160	462	7.084	7.006	0.078	0.026	0.004
Non-whites	4256	366	7.064	7.215	-0.15	0.03	0.000
Italians	2642	1980	7.078	7.074	0.004	0.015	0.804
Germanic	4461	161	7.075	7.099	-0.024	0.037	0.526
Portuguese	4122	500	7.083	7.027	0.056	0.025	0.026
Spaniards	4506	116	7.076	7.082	-0.005	0.045	0.9
East Europ.	4572	50	7.075	7.17	-0.095	0.077	0.224
			month01				
White Brz.	4160	462	5.593	5.546	0.048	0.035	0.169
Non-whites	4256	366	5.574	5.754	-0.18	0.043	0.000
Italians	2642	1980	5.675	5.473	0.202	0.022	0.000
Germanic	4461	161	5.59	5.553	0.037	0.052	0.485
Portuguese	4122	500	5.58	5.662	-0.083	0.036	0.024
Spaniards	4506	116	5.585	5.716	-0.131	0.07	0.065
East Europ.	4572	50	5.587	5.74	-0.153	0.124	0.224
			late				
White Brz.	4160	462	8.575	8.467	0.107	0.041	0.009
Non-whites	4256	366	8.555	8.675	-0.12	0.04	0.003
Italians	2642	1980	8.481	8.675	-0.194	0.022	0.000
Germanic	4461	161	8.562	8.646	-0.085	0.056	0.132
Portuguese	4122	500	8.585	8.392	0.193	0.038	0.000
Spaniards	4506	116	8.567	8.449	0.119	0.071	0.095
East Europ.	4572	50	8.564	8.6	-0.036	0.091	0.694

Table 12 (ctd.)

lhhsize							
White Brz.	2496	476	1.762	1.608	0.154	0.032	0.000
Non-whites	2842	130	1.764	1.179	0.585	0.065	0.000
Italians	1276	1696	1.631	1.819	-0.188	0.022	0.000
Germanic	2834	138	1.738	1.732	0.006	0.046	0.898
Portuguese	2597	375	1.739	1.729	0.01	0.029	0.731
Spaniards	2849	123	1.735	1.806	-0.071	0.051	0.162
East Europ.	2938	34	1.74	1.542	0.198	0.104	0.067
lenxadan07							
White Brz.	2419	441	1.493	1.391	0.102	0.03	0.001
Non-whites	2755	105	1.495	1.017	0.478	0.057	0.000
Italians	1195	1665	1.411	1.526	-0.115	0.022	0.000
Germanic	2727	133	1.474	1.551	-0.077	0.042	0.064
Portuguese	2499	361	1.476	1.491	-0.015	0.03	0.613
Spaniards	2738	122	1.48	1.435	0.045	0.057	0.429
East Europ.	2827	33	1.478	1.383	0.095	0.114	0.41
year							
White Brz.	5256	739	1912.53	1922.56	-10.032	0.569	0.000
Non-whites	5561	434	1914.69	1901.95	12.742	0.512	0.000
Italians	3336	2659	1912.46	1915.4	-2.938	0.408	0.000
Germanic	5781	214	1913.57	1918.91	-5.332	1.036	0.000
Portuguese	5337	658	1913.3	1917.54	-4.24	0.652	0.000
Spaniards	5816	179	1913.66	1917.22	-3.566	1.381	0.011
East Europ.	5925	70	1913.9	1902.09	11.817	1.53	0.000
estaca mean							
White Brz.	2466	436	373.967	398.155	-24.188	53.453	0.651
Non-whites	2775	127	375.981	412.988	-37.006	48.711	0.449
Italians	1274	1628	427.575	338.493	89.082	30.148	0.003
Germanic	2776	126	372.725	485.032	-112.31	94.223	0.236
Portuguese	2538	364	361.377	490.724	-129.35	42.73	0.003
Spaniards	2809	93	378.137	361.403	16.735	29.546	0.572
East Europ.	2879	23	379.026	199.16	179.867	16.629	0.000
estaca mdev							
White Brz.	2466	436	189.233	133.859	55.373	30.439	0.07
Non-whites	2775	127	178.315	237.681	-59.367	69.37	0.394
Italians	1274	1628	178.118	183.101	-4.983	24.202	0.837
Germanic	2776	126	178.279	238.95	-60.672	64.844	0.351
Portuguese	2538	364	169.764	258.651	-88.887	42.088	0.036
Spaniards	2809	93	184.944	59.166	125.778	31.259	0.000
East Europ.	2879	23	181.969	48.824	133.144	12.26	0.000
estaca sd							
White Brz.	2463	430	263.82	188.044	75.776	43.227	0.08
Non-whites	2766	127	248.798	334.427	-85.629	98.091	0.384
Italians	1267	1626	250.167	254.419	-4.252	34.131	0.901
Germanic	2767	126	248.581	339.872	-91.291	93.737	0.332
Portuguese	2529	364	236.728	362.532	-125.81	58.404	0.032
Spaniards	2800	93	258.276	80.368	177.908	45.657	0.000
East Europ.	2870	23	254.085	61.872	192.213	17.241	0.000

The indicators for central tendency and dispersion of the numbering of *estacas* reveal statistically significant differences between groups. Although these numeric differences do not have a substantive meaning per se, they nonetheless indicate that different ethnolinguistic groups clustered in various parts of the plantation and within colonies of work. Finally, the mean-year when each group was most commonly registered in the sample shows the dynamic changes in the average ethnolinguistic composition of *Ibicaba's* labor force.

As for timing of harvest, the most problematic estimates of the OB decompositions, the differences-in-means tests show that Non-white households were employed systematically later in the harvest season: They started and ended their season later than the corresponding sample means. Italian households, in turn, started earlier in the first month and continued for longer than the corresponding sample mean. In the opposite direction, Portuguese and Spaniard households started later in their first month and ended earlier in the last month of harvesting. Finally, households headed by White Brazilians ended earlier in the last month of harvesting, leading to the smallest mean month of work among all groups.

v.iii. Testing for Taste-Based Discrimination in pay

Results thus far showed that differences in income profiles across ethnolinguistic groups are attributable to households's demographic composition, frequency of engagement in agricultural tasks, and prices earned. The demographic determinant was an endowment that cannot be attributed to discrimination. The explained and unexplained components of the OB decomposition during the lean season suggest that the frequency of engagement in agricultural tasks of hired Blacks, *i.e.* who had already surpassed a potential barrier to entry, was a complex mix between households' choice and negative discrimination. In addition, the complementarity between the number of economically-active members in a household and the length of their harvest seasons is indicative of statistical discrimination. In regard to prices earned, there is no evidence of discrimination in favor of Europeans, given the various signs attributed by the OB decompositions to different European origins and the disadvantages of both Black and Italian households in that dimension.

Since p_{ht} was negotiated between plantation managers and laborers, and could vary over time and across households, that variable can be used to test for taste-based discrimination directly, *i.e.* different pay across households for homogenous agricultural tasks, after controlling for household characteristics (Hellerstein, Neumark, and Troske 1999).

The determinants of physical output (instead of income, as in the baseline) are estimated in order to obtain households' TFPs in the lean and harvest seasons. With that estimate, we can check whether ethnolinguistic origins influenced the price paid to laborers after having controlled for households' characteristics, including their TFPs. The TFPs are obtained as the quantity-based productivity (Atkin, Khandelwal, and Osman 2017, pp. 444-6):

$$ln\left(Output_{ht}\right) = \beta_0 + \left(ln\left(L_{ht}\right)\right)'\delta + K_{ht}'\Gamma + \alpha_h + \lambda_t + e_{ht},$$

in which $Output_{ht}$ is always estimated twice, separately for physical output in the lean season ($Cult_{ht}$) and physical output in the harvest season ($Harv_{ht}$). The former corresponds to the total number of coffee groves cultivated by household h in year t, while the latter refers to the volume of coffee harvested (in *alqueires*).

Historians of Brazil might question the usefulness of estimating $Cult_{ht}$, as it is commonly assumed that the number of cultivated coffee groves by *colono* households was fully determined by their size and age composition. This assumption is based on custom and contract. Going all the way back to the *Vergueiro system*, various labor arrangements stipulated a fixed number of coffee groves to be cultivated per person of working age, usually multiples of 500.[26] However, such interpretation ignores that households could contract the cultivation of more or fewer groves over time. Besides changes in their size and composition, households' productive capacity and labor effort also depended on varying health conditions, motivation, and outside options. Finally, the number of coffee groves a household was willing and able to cultivate depended on natural conditions determined by the location of groves and distance from the colony, all of which changed between agricultural seasons (*cf.* Section IV).

Household's total factor productivity is obtained as the residual, $\hat{A}_{ht} = exp\,(\hat{e}_{ht})$ (Foster, Haltinwanger, and Syverson 2008). $\hat{A}_{ht}$ captures all unexplained variation in physical output after netting out household's composition and time-invariant characteristics. It gives a clean metric of intra-household "technologies" in coffee cultivation and harvesting, which reflect the household's adaptation to *Ibicaba*'s coffee economy, *e.g.* acclimatization of immigrants to their new working conditions or the adapting expertise of households who had previously worked in the coffee economy, eventually as ex-enslaved laborers).

The final step is then to evaluate whether prices paid for the same agricultural tasks varied by households' ethnolinguistic origins even after controlling for their TFPs and year-FEs:

$$p_{ht}^{output} = \beta_0 + \beta_1 \hat{A}_{ht} + Origin'_{ht}\delta_g + \lambda_t + E_{ht},$$

in which E_{ht} is an i.i.d. disturbance and p_{ht}^{output} is always estimated twice, separately for the average price earned in the lean season (p_{ht}^{cult}) and in the harvest season (p_{ht}^{harv}).

$Origin'_{ht}$ is the set of indicators on whether the household was classified as Non-white, Germanic, Italian, Portuguese, Spaniard, or Eastern European. White Brazilians are the reference group. $\delta_g \lessgtr 0$ will capture the prevalence of taste-based discrimination against or in favor of group g in comparison to the households of White Brazilians.

[26] See, for instance, the discussion on labor contracts in Santa Gertrudes plantation in Dean (1976, p. 168).

Table 13 – Testing taste-based discrimination in prices earned by different ethnolinguistic groups: Lean & harvest seasons

	(1)	(2)	(3)	(4)	(5)	(6)
	Avg. price: cultivating one coffee grove (lean season)			Avg. price: harvest of one *alqueire* (harv. season)		
TFP agr. task		0.000441	0.00144***		-0.0291*	0.00731***
		(0.00156)	(0.000279)		(0.0162)	(0.00268)
Non-whites	-0.0108***	-0.0114***	0.000321	-0.251***	-0.342***	0.00500
	(0.00149)	(0.00181)	(0.000207)	(0.0744)	(0.0447)	(0.00372)
Italians	-0.00338***	-0.00455***	-8.63e-05	-0.165**	-0.160***	0.00443
	(0.000995)	(0.00102)	(0.000230)	(0.0708)	(0.0313)	(0.00396)
Germanic	0.00121	0.00105	0.000473	-0.0490	0.00542	0.000883
	(0.00224)	(0.00226)	(0.000446)	(0.0846)	(0.0573)	(0.00426)
Portuguese	-0.00170	-0.00286**	-0.000270	0.00222	-0.0828*	0.00601
	(0.00137)	(0.00136)	(0.000254)	(0.111)	(0.0442)	(0.00470)
Spaniards	0.000639	-0.000908	-0.000184	-0.0900	0.0386	-0.0136
	(0.00213)	(0.00209)	(0.000424)	(0.0830)	(0.0603)	(0.00930)
East-Europ.	-0.00262	-0.00435	-0.000557	-0.249***	-0.355***	0.00342
	(0.00328)	(0.00300)	(0.000733)	(0.0725)	(0.0266)	(0.00378)
Observations	2,854	2,360	2,360	3,131	1,470	1,470
Number of id	565	513	513	579	324	324
HH controls	Yes	Yes	Yes	Yes	Yes	Yes
Year-FE	No	No	Yes	No	No	Yes

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "TFP agr. task" refers to HH's TFPs in producing physical output, *i.e.* number of cultivated coffee groves in the lean season and quantity (in *alqueire*) of coffee berries picked in the harvest season.

Columns (3) and (6) of Table 13 report the correct specifications for the lean and harvest seasons, respectively. These correct specifications show no evidence of taste-based discrimination by ethnolinguistic origins in the prices earned by households for the execution of homogenous, unskilled agricultural tasks in the lean and harvest seasons once those households' TFPs and common time trends and shocks are controlled for. In contrast, misspecified models that do not control for households' TFPs nor year-FEs (columns 1 and 4), or do not control for year-FEs (columns 2 and 5) would lead to the wrong conclusion that Non-white, Italian, Eastern European, and, less robustly, Portuguese households were all discriminated against.

VI. The immediate post-abolition era (1888-1890): Testing other potential mechanisms of discrimination

This section zooms into the critical historical juncture of the 1888 Abolition and its immediate aftermath to address the questions of whether ethnolinguistic origin influenced labor income and remuneration, the likelihood of a household signing a *colonato* contract, and the quality of coffee groves that households were assigned to. More than a sub-sample analysis, this section uses a unique object that has been preserved in the *Ibicaba Collection*. Written in 1888-1890, ledger *Biblioteca Paulo Masuti Levy #0004* is the last *Registries* produced under *Vergueiro & Co*., before the public seizure and auctioning of *Ibicaba* in 1889/90. Beyond more usual data on agricultural production and demographics, this ledger contains ingrained detail on household characteristics. These include gradations of skin color, literacy, religion, occupations (including non-agricultural) and behavior (as assessed by plantation managers). The contents of ledger *BPML#0004* thus differ from the more homogenous data of subsequent years. This implies that the baseline specifications cannot be replicated here. Also, given the short time span and missing values, all estimates that follow are either Pooled OLS for 1888-1890, or cross-sectional for 1890. Notwithstanding, the unique nature of this source and of the historical period it covers are worth being explored in conjunction with the remaining, econometrically more robust, analyses of this paper.

VI.I. The effects of skin color and ethnolinguistic origins on labor income and remuneration

The following analysis adapts baseline specification to the cross-sectional nature of the available data for income from harvesting, Y_h^{harv}.[27] Fortunately, some of the time-invariant characteristics that would otherwise be captured by household fixed effects can be controlled for with the extra information collected by the accountants of *BPML#0004*:

[27] Data on coffee cultivation during the lean season registered in *BPML#0004* differ from that of subsequent ledgers. *Ibicaba*'s managers recorded the number of coffee groves cultivated by some households, including assessments of the quality of the groves, but not their location in *estacas*. This information is used below to assess whether the quality of the coffee groves varied according to households' ethnolinguistic origins.

$$ln\left(Y_h^{harv}\right) = \beta_0 + \beta_1 ln\left(p_h\right) + \beta_2 ln\left(N_h\right) + \left(ln\left(L_h\right)\right)'\delta + K_h'\Gamma + H_h'\theta + \epsilon_h$$

The definition of variables is the same as in the baselines, except that they are now a cross-section for 1890. K_h' controls for the number of coffee groves cultivated in the lean season and adds an indicator for the corresponding colony (but not for *estacas* therein). Time-invariant characteristics $(H_h)'$ added to this specification include: Literacy (= 1 if literate); the accountants' assessment of household's behavior; and indicators for the ethnolinguistic origins of the household.[1] Moreover, particular attention is paid to the added controls for ethnolinguistic origins, proxied by a set of different indicators. Column (1) uses the same indicators as in the baselines. Column (2) uses the indicator $Nonwhite_h$ (= 1 if household was listed as *Preto*, *Pardo*, or *Trigueiro*, and = 0 if listed as *White*). Column (4) uses the categorical variable $Skin\ color_h$ for gradations of skin color (= 1 if White; = 2 if *Trigueiro*; = 3 if *Pardo*; = 4 if *Preto*). Based on skin colors recorded in the source itself, the last two variables are direct metrics of contemporary racial characterizations.

Results suggest that the normal economic functioning of the plantation was disrupted two years after Abolition. The otherwise very robust indicator for the number of economically-active members in a household becomes statistically insignificant, while the total number of coffee trees cultivated in lean season becomes highly significant, with a coefficient much higher than those estimated for the agricultural capital in any other model. In addition, the partial effect of price remains statistically significant, but turns negative. In clear contrast with the pooled sample, this result suggests that the average household was substituting quantity harvested for price.

In line with the broader conclusions of this paper, Table 14 shows no evidence that the skin color of household heads played a role in determining the labor income of the 62 households for whom all information was available in 1890.[2] Included in this limited sample were seven households classified as *Non-Whites*.[3] If anything, the only statistically significant category was that of the *Germanic*, with a large negative partial effect. The problem with this result is that its counterfactual is historically irrelevant, even if it mattered for that individual family: The only *Germanic* household that remained in the 1890 cross-sectional sample was that of Jorge Somar (sic), a literate *colono* family, classified as having "good" behavior.

[1] The variable *Religion* (= 0 if Catholic and = 1 if "Protestant", according to the source) was not included because it leads to the omission of the category *Germanics* from the set of indicators for ethnolinguistic origins. The variable *Entry year*, defined as the year when a household started working in *Ibicaba*, was not included because it leads to a very small sample (N = 36 for the regressions on labor income during the harvest season).
[2] Results are identical for the variables that control for the darkest skin color of any household member.
[3] All proxies for ethnolinguistic origins used in this section lead to the same number of *Non-White* households.

Table 14 - Conditional correlations of the determinants of labor income (cross-section, 1890)

	$\ln(Y_h^{cult})$		
ln(price)	-2.356**	-2.871***	-2.810***
	(1.049)	(1.018)	(1.007)
ln(HH size)	0.158	0.115	0.118
	(0.158)	(0.157)	(0.155)
ln(# members 7-60y)	-0.240	-0.226	-0.237
	(0.156)	(0.156)	(0.152)
ln(# coffee trees *BPML04*)	0.785***	0.763***	0.773***
	(0.146)	(0.142)	(0.141)
Grama Larga (colony)	-0.0113	-0.0189	-0.0137
	(0.0918)	(0.0907)	(0.0899)
Teteia (colony)	0.888***	0.753***	0.774***
	(0.315)	(0.256)	(0.252)
Abundancia (colony)	-0.252	0.182	0.248
	(0.449)	(0.276)	(0.280)
Boa Esperança (colony)	-0.0619	0.000292	-0.0968
	(0.323)	(0.324)	(0.316)
Literate = 1	0.634*	0.357	0.330
	(0.317)	(0.221)	(0.218)
Behavior = Good	0.750**	0.678**	0.720***
	(0.293)	(0.257)	(0.234)
Non-Whites	-0.111		
	(0.408)		
Italians	-0.355		
	(0.361)		
Germanics	-1.045**		
	(0.497)		
Portuguese	-0.419		
	(0.378)		
Spaniards	-0.231		
	(0.387)		
Non-white HH head		0.185	
		(0.229)	
Skin color HH head			0.0983
			(0.0766)
Observations	62	62	62
R-squared	0.804	0.782	0.786
r2_a	0.722	0.716	0.722

Notes: (1) All models include a constant and the complete set of variables for the timing of harvest, as in the baselines; (2) The colony of reference is "Colonia Ibicaba", the largest one in terms of residing households in 1888-90.

Table 15 - Conditional correlations of price paid per *alqueire* of coffee harvested (cross-section, 1890)

VARIABLES	$\ln(p_h^{harv})$		
ln(HH size)	0.0132	0.0214	0.0220
	(0.0226)	(0.0220)	(0.0220)
ln(# members 7-60y)	0.00779	0.00338	0.00226
	(0.0224)	(0.0221)	(0.0218)
ln(# coffee trees *BPML04*)	-0.0177	-0.0123	-0.0123
	(0.0208)	(0.0201)	(0.0201)
Grama Larga (colony)	0.00453	0.00417	0.00401
	(0.0132)	(0.0128)	(0.0129)
Teteia (colony)	-0.0128	0.0223	0.0237
	(0.0452)	(0.0362)	(0.0359)
Abundância (colony)	-0.0354	-0.0384	-0.0409
	(0.0643)	(0.0387)	(0.0397)
Boa Esperança (colony)	0.127***	0.132***	0.130***
	(0.0423)	(0.0419)	(0.0412)
Literate = 1	0.0111	0.0677**	0.0719**
	(0.0456)	(0.0298)	(0.0294)
Behavior = Good	-0.0100	0.0165	0.0217
	(0.0420)	(0.0364)	(0.0333)
Non-Whites	-0.0464		
	(0.0583)		
Italians	-0.0141		
	(0.0518)		
Germanics	0.0913		
	(0.0701)		
Portuguese	0.0112		
	(0.0543)		
Spaniards	-0.0255		
	(0.0555)		
Non-white HH head		-0.0319	
		(0.0321)	
Color HH head			-0.0104
			(0.0109)
Observations	62	62	62
R-squared	0.430	0.384	0.383
r2 a	0.210	0.217	0.216

Notes: (1) All models include a constant and the complete set of variables for the timing of harvest, as in the baselines; (2) The colony of reference is "Colonia Ibicaba", the largest one in terms of residing households in 1888-90.

In addition, Table 15. presents regress the price paid for the *alqueire* of coffee harvested on the same set of controls. Differently from the baseline specifications, the cross-sectional results reported below do not control for households' time-varying TFPs. Again in line with the broader conclusions of this paper, Table Table shows no evidence that ethnolinguistic origins or contemporaneous perceptions about skin color influenced the price paid for households in the same labor category. Literate individuals had a higher probability of earning higher prices (between 7% and 7.5%), which could be a consequence of better access to information (*e.g.* salary listings or job advertisements in newspapers), or ability to monitor conditions written in *colonato* contracts. Although the table suggests that racial discrimination was not a statistically significant correlate of pay for the execution of homogenous, unskilled agricultural tasks, only one Non-White household in the sample of Table 15 was literate. Finally, *colonos* working in *Boa Esperança* colony earned between 13.54% and 14.11% more per *alqueire* of coffee harvested, *ceteris paribus*, *vis-à-vis* the largest colony at the time, the so-called *Colônia Ibicaba*. It is noteworthy that this is the only significant variable in specifications 1-3 and that its inhabitants were all *Non-Whites*. I will study in the sequence the implications of working on that colony for the quality of the coffee groves cultivated in the lean season.

VI.II. The effects of skin color and ethnolinguistic origins on the likelihood of signing a *colonato* contract & quality of cultivated coffee groves

Table 16 reports the occupations of all laborers for whom this information was registered in ledger *BPML#0004*. Labor categories are grouped by ethnolinguistic origin for uniquely identified households (HH) across all available years. Occupational categories include households under a *colonato* arrangement (*colonos*); other agricultural laborers, including daily laborers and workers per piece rate;[4] households with at least one member in managerial positions; households with at least one member listed in a skilled, non-agricultural, occupation; *idem* for unskilled, non-agricultural.

While it is obvious that Italian immigrants were the core of the *colonato* system, the prominent position of *Non-Whites* shows that signing a *colonato* contract was not barred to them, including in the immediate post-Abolition era. Notwithstanding, the table also shows that *Non-Whites* were underrepresented as *colonos* vis-à-vis their demographic weight and that no other ethnolinguistic group had as many households in unskilled occupations and in other agricultural labor categories.

In addition, ledger *BPML#0004* provides a classification for the "behavior" of households in 1890 according to the perception of the plantation manager(s). Households' behavior are classified as "bad/poor (máo)", "good (bom)", and "exemplary (exemplar)", categorized by ethnolinguistic origins in Table 17.

[4] Data for *Other agricultural laborers* should be treated with caution. I am certain that not all non-resident laborers had their characteristics recorded in the source, not least due to the small number of observations. This variable is a lower-bound estimate of the total number of agricultural laborers who were not *colonos*.

Table 16 – Occupations by ethnolinguistic origin (1888-90) – data from ledger *BPML#0004* only

	Colonos		Other agric.-L		Managers		Skilled		Unskilled	
	N	HHs	N	HHs	N	HHs	N	HHs	N	HHs
Brazilians	17	14	1	1	1	1	2	2	3	3
Non-whites	26	23	12	12	0	0	2	2	5	5
Italians	160	85	0	0	2	2	1	1	0	0
Germanic	5	4	1	1	1	1	2	2	1	1
Portuguese	23	19	1	1	0	0	1	1	2	2
Spaniards	25	15	1	1	0	0	0	0	0	0
East. Eur.	14	14	0	0	0	0	0	0	0	0
Total	270	174	16	16	4	4	8	8	11	11

Notes: (1) *Other agricultural laborers* include “camaradas”, “trabalhadores”, “camarada & colono”, “agregado”, and “empreiteiro”. (2) *Managers* include “administrador”, “director”, “escrivão”, and “inspector”. (3) *Skilled* include “marceneiro”, “farinheira”, “ferreiro”, “foguista & lavrador”, “machinista”, and “pedreiro”. (4) *Unskilled* include “ajudante”, “campeiro", “carregador de comida”, “carreiro”, “carroceiro”, “conductor de serviço”, “cozinheiro”, “jardineiro”, “puxador”, “servente”, and “vigia”.

Table 17 – Household's behavior according to *Ibicaba's* manager(s) (1890)

	Bad/Poor	Good	Exemplary	Total
Brazilians	0	8	1	9
Non-Whites	0	22	3	25
Italians	0	45	0	45
Germanics	0	2	0	2
Portuguese	0	10	4	14
Spaniards	0	6	0	6
East. Eurs.	1	11	0	12
Total	1	104	8	113

Noticeably, among foreigners, the status *exemplary* was attributed only to Portuguese; all other four were Brazilians, three of whom, Non-Whites. Although subjective, the rankings of Non-Whites certainly reflect the plantation manager(s)' satisfaction with them (and, by implication, with their labor efforts).

Using this extra information collected by the plantation manager(s), Table 18 estimates the likelihood of of household becoming *colono*, with focus on their ethnolinguistic origins and skin colors. The model is augmented stepwise with controls for time-invariant characteristics in the short-run, available in ledger *BPML#0004* (literacy, religion, and behavior), household size and number of economically-active individuals, as well as a time-trend:

$$logit\left(Colono_{h}\right) = \beta_{0} + \beta_{1} Ethnoling_{h}^{'}\delta + H_{h}^{'}\theta + (L_{h})^{'}\mu + T + \mathrm{E}_{h}$$

Tables 18-20 report the corresponding odd-ratios, controlling for households's $Nonwhite_{h}$, $Skin\ color_{h}$, and ethnolinguistic origins, respectively. In addition, margins for $Ethnoling_{h}$ are computed at four set of values for the independent variables: (1) Sample means; and at the means of (2) *Non-Whites*; (3) All whites (including Brazilians and foreigners); and (4) White Brazilians, only.

Table 18 shows that the negative effect of being *Non-White* on the probability of having signed a *colonato* arrangement disappears if we control for the labor force available in each household. In fact, increasing the number of household members in the age range 7-60y increases the odds of being a *colono* by a factor between 2.6 and 2.9. These point estimates are statistically robust to any further controls (and to changes in sample sizes they imply).[5] The margins tell a similar story, but with a crucial difference. The largest difference between *Non-Whites* and *Whites* is indeed attributable to the values over which the margins are computed. For *Non-Whites*, differences with respect to *All Whites* that are due to demographic and time-invariant characteristics are always larger than differences between *Non-Whites* and *All Whites* if they had shared exactly the same "endowments". However, the probability of *Non-Whites* being a *colono* is always smaller than that of a white household, even if they had shared all the same characteristics. Results in Table 19, which use gradations of skin color of any household member, are qualitatively identical, not least because only individuals classified in the sample as Whites and Blacks remain in the final sample.

Finally, estimates using ethnolinguistic origins do not reveal any disadvantage of Non-Whites to engage in *colonato* arrangements (Table 20). However, mirroring previous results, Italians had a higher likelihood of being *colonos*. Nonetheless, the effect of being Italian disappears if household size and number of individuals in the age range 7-60y are controlled for. The margins reveal that Italians had the highest likelihood of being *colonos*, independent of endowments; in fact, computed at the mean values for *All Whites*, the likelihood of an Italian household being a *colono* is practically 100%. Similar to previous estimates, differences in endowments are the main reason for the lower likelihood of *Non-Whites* to be *colonos*. As such, the subcategorization by ethnolinguistic origins reveals that the main difference between *Whites* and *Non-Whites* stems from the preeminence of Italian *colonos*.

[5] Households with literate individuals had a lower probability of being *colonos*. These households might have been more frequently recruited to managerial or to non-agricultural rural tasks; for the whites, beyond general illiteracy, some recently-arrived foreign *colonos* might have been illiterate in Portuguese.

Table 18 – Conditional correlations of being a *colono* household, including binary identifier for Non-Whites (1888-90): Odd-ratios

	colono						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1.nonwhitehh	0.319***	0.215***	0.328***	0.0697**	0.571	0.365	0.478
	(0.124)	(0.114)	(0.134)	(0.0863)	(0.212)	(0.561)	(0.255)
Literate		0.123***				-	0.104***
		(0.0879)					(0.0886)
Religion			0.789			-	0.885
			(0.537)				(0.697)
Behavior				7.133**		1.726	
				(5.685)		(1.367)	
HH size					0.701**	0.989	0.658*
					(0.108)	(0.269)	(0.145)
# members 7-60y					2.607***	2.305**	2.952***
					(0.623)	(0.893)	(1.023)
Year trend	Yes	Omitted	Yes	Omitted	Yes	Omitted	Omitted
Observations	314	138	307	112	314	90	137
r2_p	0.126	0.133	0.118	0.179	0.259	0.153	0.283

Notes: (1) Robust standard errors in parentheses; (2) All models include a constant; (3) "-": omitted due to collinearity.

Margins: From model reported in column (7) at values of				
	Sample means	Non-Whites	All Whites	White Brazilians
Non-whites	0.804***	0.662***	0.851***	0.717***
	(0.0809)	(0.0888)	(0.0731)	(0.135)
Whites	0.896***	0.804***	0.923***	0.841***
	(0.0392)	(0.0560)	(0.0358)	(0.0695)
Diff. (in perc. points)	-9.2	-14.2	-7.2	-12.4

Table 19 – Conditional corrrelates of being a colono household, including gradations of skin color (1888-90): Odd-ratios

	colono						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
colorany	0.679***	0.591***	0.687***	0.412**	0.825	0.711	0.774
	(0.0885)	(0.107)	(0.0943)	(0.168)	(0.104)	(0.366)	(0.142)
Literate		0.107***				-	0.100***
		(0.0752)					(0.0834)
Religion			0.820			-	0.888
			(0.558)				(0.700)
Behavior				6.151**		1.641	
				(4.728)		(1.248)	
HH size					0.693**	0.988	0.652*
					(0.107)	(0.271)	(0.146)
# members 7-60y					2.662***	2.293**	2.991***
					(0.633)	(0.886)	(1.035)
Year trend	Yes	Omitted	Yes	Omitted	Yes	Omitted	Omitted
Observations	315	139	308	112	315	90	138
r2_p	0.128	0.146	0.121	0.180	0.267	0.153	0.299

Notes: (1) Robust standard errors in parentheses; (2) All models include a constant; (3) "-": omitted due to collinearity.

Margins: From model reported in column (7) at values of				
	Sample means	Non-Whites	All Whites	White Brazilians
Blacks	0.792***	0.654***	0.846***	0.706***
	(0.0864)	(0.0928)	(0.0768)	(0.142)
Whites	0.892***	0.804***	0.923***	0.838***
	(0.0398)	(0.0561)	(0.0358)	(0.0698)
Diff. (in perc. points)	-10	-15	-7.7	-13.2

Table 20 – Conditional correlations of being a *colono* household, including ethnolinguistic groups (1888-90): Odd-ratios

	colono				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Non-Whites	0.677	0.763	0.905	0.952	1.634
	(0.399)	(0.542)	(0.547)	(0.580)	(1.397)
Italians	2.935*	6.705**	3.329**	2.134	5.497
	(1.645)	(5.938)	(1.909)	(1.281)	(6.226)
Germanics	0.355	0.510	1.29e-06***	0.187	4.28e-08***
	(0.273)	(0.513)	(1.45e-06)	(0.192)	(7.11e-08)
Portuguese	3.008	2.179	3.132	2.087	1.453
	(2.349)	(1.892)	(2.464)	(1.962)	(1.846)
Spaniards	1.999	2.111	2.066	1.727	1.510
	(1.628)	(2.667)	(1.701)	(1.380)	(2.233)
Literate		0.339			0.366
		(0.275)			(0.364)
Religion			334,086***		3.070e+06***
			(420,956)		(4.501e+06)
HH size				0.760	0.708
				(0.127)	(0.178)
# members 7-60y				2.296***	3.043***
				(0.571)	(1.234)
Year trend	Yes	Omitted	Yes	Yes	Omitted
Observations	306	127	295	306	126
r2_p	0.175	0.171	0.167	0.292	0.360

Notes: (1) Robust standard errors in parentheses; (2) All models include a constant; (3) Ethnolinguistic "Eastern European" not reported for being omitted due to lack of observations; (4) Specifications with "behav" not reported as most variables are omitted due to sample size (N = [23, 28], depending on specification); (5) "-": omitted due to collinearity.

	Margins: From model reported in column (4) at values of			
	Sample means	Non-Whites	All Whites	White Brazilians
Brazilians	0.861***	0.598***	0.924***	0.778***
	(0.114)	(0.197)	(0.0737)	(0.162)
Non-whites	0.910***	0.708***	0.952***	0.851***
	(0.0459)	(0.0778)	(0.0305)	(0.0879)
Italians	0.971***	0.891***	0.985***	0.951***
	(0.0177)	(0.0581)	(0.00996)	(0.0392)
Germanics	2.65e-07	6.35e-08	5.17e-07	1.50e-07
	(3.59e-07)	(9.67e-08)	(6.80e-07)	(1.83e-07)
Portuguese	0.900***	0.683***	0.946***	0.836***
	(0.0982)	(0.223)	(0.0595)	(0.148)
Spaniard	0.903***	0.692***	0.948***	0.841***
	(0.103)	(0.250)	(0.0593)	(0.169)

The conditional correlations of labor income from the pooled 1888-1890 harvests showed that being a *colono* was not all that mattered for households' economic performance in the immediate post-Abolition era. Where exactly they worked within that large plantation also influenced their labor income. Table 21 shows the spatial distribution of households by ethnolinguistic groups in *Ibicaba*'s various colonies in 1888-1890.[6] *Non-White colonos* worked side-by-side with Europeans and white Brazilians in most colonies. Nonetheless, the vast majority of Non-White *colonos* were concentrated in *Boa Esperança* colony.

Ledger *BPML#0004* further classified the cultivated coffee bushes as "*new* (novos)", *i.e.* recently planted; "old (*velhos*)"; and "bearing fruits (*dando fructos*)". In addition, from the total number of coffee bushes that can be derived from these categories, some have been further classified as "with weed (*com gramma*)" and "failed (*falhos*)", *i.e.* likely with gaps and dead trees in the rows of coffee groves. These categories need to be considered with caution, as 1888/9 was the last agricultural season in which *Ibicaba* was under *Vergueiro & Co.*'s proprietorship. The firm thus had all incentives to downplay the valuation of the agricultural capital available in *Ibicaba*, both to justify potentially small harvests that would not cover its mortgages and to lower the interest of creditors to foreclose it. Exactly the same incentive persisted in 1889/90, now from the demand side, once the plantation was publicly auctioned. Moreover, these categorizations of coffee trees were not carried out in a systematic manner. Notwithstanding, this information allows us to study the contours of the agricultural capital available to laborers in the immediate post-Abolition, as there is no reason to believe that managers would bias the accounting of the quality of coffee groves attributed to any particular household.

Analogously to all specifications in this section, what follows is a study of conditional correlations between the quality of coffee trees and the ethnolinguistic origins of households:

$$share(trees_h^{quality}) = \beta_0 + \beta_1 Ethnoling_h' \delta + H_h' \theta + (L_h)' \mu + Colony_h' \sigma + T + E_h$$

The dependent variable is the share of coffee trees of $quality = \{bearing\ fruits,\ new,\ w/\ weed,\ failed\}$ cultivated by household h, with data pooled for 1888-90.[7]

[6] For *Non-Whites*, the number of observations in Table 21 corresponds also to the number of uniquely identified households.
[7] I did not analyze the category "old" given the corresponding small sample size.

Table 21 – Colonies of coffee cultivation, by ethnolinguistic group (1888-90)

	Braz	Non-wht.	Ital	Grm	Port	Span.	East. Eur	Total
Grama Larga	0	2	17	0	2	0	4	25
Saltinho & Pinheiro	0	1	0	0	1	2	0	4
Teteia	5	0	1	0	6	1	0	13
Col. Ibicaba	4	7	104	5	3	14	10	147
Abundância	8	1	0	0	7	1	0	17
Boa Esperança	0	15	0	0	0	0	0	15
Ypiranga	0	0	0	0	3	0	0	3
Laborers	5	13	1	4	3	0	0	26
Total	22	39	123	9	25	18	14	250

Table 22 shows that *Ibicaba* managers in 1888-1890 did not discriminate against or in favor of any skin color or ethnolinguistic groups. *Non-Whites* and *Skin Color* were not a significant correlate with any type of coffee tree cultivated, except for trees with weed; however, even in this case, all other ethnolinguistic groups also had a statistically significant probability of getting worse trees than *White Brazilians*. Moreover, while *Germanics* and *Eastern Europeans* had a higher conditional probability of being allocated to blossoming trees and a lower probability of receiving new trees, they also had a higher probability of getting trees with weed.[8] Even more interestingly, working in *Boa Esperança* colony increased the probability of a household (all of them *Non-Whites*) to be allocated to coffee trees of higher quality than the reference group. In other words, plantation managers were willing to allocate those better coffee trees to a group formed entirely by *Non-Whites*, perhaps as an incentive to keep their labor force in the plantation.

Grama Larga was chosen as the reference group for being a rather ordinary colony in 1888-1890, with the usual high share of Italians and a minority of Non-Whites, Portuguese, and Eastern Europeans (Table 21). Results are robust if *Boa Esperança* is taken as the reference group instead (Table 23). The only difference is that all ethnolinguistic groups other than the *Non-Whites* have statistically non-significant coefficients on the probability of being allocated to trees with weed.

[8] For the *Germanics*, the net effect of being Protestant and ethnolinguistic origin is almost zero on the probability of receiving trees that were new.

Table 22 – LPM correlates for the quality of coffee trees cultivated (1888-90) – Reference group for the colonies: *Grama Larga*

	Share of coffee trees											
VARIABLES	Blossoming			New			With Weed			Failed		
1.nonwhitehh	0.0276			-0.0325			0.341***			0.0387		
	(0.0474)			(0.0498)			(0.0849)			(0.0484)		
colorany		0.0106			-0.0116			0.114***			0.00547	
		(0.0181)			(0.0188)			(0.0283)			(0.0150)	
Non-Whites			0.0788			-0.0524			0.536***			-0.00246
			(0.0580)			(0.0398)			(0.0897)			(0.123)
Italians			0.0486			-0.0141			0.188***			-0.0505
			(0.0648)			(0.0633)			(0.0391)			(0.116)
Germanic			0.208***			-0.184***			0.339***			-0.0827
			(0.0610)			(0.0541)			(0.0243)			(0.111)
Portuguese			0.0350			0.00730			0.496**			-0.0280
			(0.0548)			(0.0777)			(0.187)			(0.108)
Spaniards			0.0549			-0.0199			0.266***			-0.0393
			(0.0724)			(0.0733)			(0.0885)			(0.124)
Eastern Europeans			0.242***			-0.208***			0.337***			-0.0917
			(0.0660)			(0.0632)			(0.0473)			(0.118)
Literate	0.0448*	0.0430	0.0608*	-0.0347	-0.0337	-0.0406	0.121**	0.121**	0.0862**	-0.0817**	-0.0780**	-0.0952**
	(0.0255)	(0.0276)	(0.0358)	(0.0319)	(0.0336)	(0.0502)	(0.0536)	(0.0536)	(0.0408)	(0.0316)	(0.0302)	(0.0372)
Religion	0.0509**	0.0515**	-0.136***	-0.0518**	-0.0522**	0.137***	-0.0358	-0.0358	-0.135**	0.0607**	0.0579**	0.104**
	(0.0224)	(0.0227)	(0.0291)	(0.0222)	(0.0225)	(0.0291)	(0.0240)	(0.0240)	(0.0534)	(0.0265)	(0.0265)	(0.0447)
Saltinho-Pinheiro	-0.831***	-0.831***	-0.812***	0.830***	0.830***	0.809***				0.0770*	0.0738*	0.0609
	(0.0216)	(0.0215)	(0.0491)	(0.0220)	(0.0220)	(0.0513)				(0.0431)	(0.0427)	(0.0606)
Teteia	-0.117***	-0.116***	-0.0694	0.115***	0.114***	0.0694				-0.102***	-0.104***	-0.126***
	(0.0293)	(0.0297)	(0.0486)	(0.0293)	(0.0298)	(0.0488)				(0.0288)	(0.0279)	(0.0431)
Colonia Ibicaba	0.107***	0.107***	0.135***	-0.107***	-0.107***	-0.135***	-0.166***	-0.166***	-0.164***	0.00266	0.00302	-0.00388
	(0.0298)	(0.0298)	(0.0295)	(0.0299)	(0.0299)	(0.0295)	(0.0463)	(0.0463)	(0.0501)	(0.0263)	(0.0265)	(0.0285)
Abundância	0.0916***	0.0945***	0.151**	-0.0742**	-0.0784**	-0.110*	-0.218***	-0.218***		-0.0291	-0.0266	-0.0705
	(0.0329)	(0.0329)	(0.0583)	(0.0370)	(0.0360)	(0.0604)	(0.0405)	(0.0405)		(0.0466)	(0.0506)	(0.120)
Boa Esperaça	0.127**	0.123**	0.158***	-0.122**	-0.120**	-0.150***	-0.400***	-0.400***	-0.351***	-0.0904*	-0.0694	-0.105**
	(0.0488)	(0.0546)	(0.0524)	(0.0510)	(0.0565)	(0.0565)	(0.107)	(0.107)	(0.102)	(0.0502)	(0.0481)	(0.0452)
Observations	101	101	102	99	99	100	67	67	68	99	99	100
R-squared	0.735	0.735	0.775	0.735	0.735	0.775	0.374	0.374	0.516	0.175	0.167	0.188
r2_a	0.702	0.702	0.733	0.701	0.701	0.732	0.275	0.275	0.400	0.0702	0.0618	0.0314

Notes: (1) Robust standard errors in parentheses; (2) All models follow specification (5) of this section and include the full set of independent variables, including Enxada and HHSize (not reported to save space and non-significant throughout) and a time trend; (3) All models include a constant.

Table 23 – LPM correlates for the quality of coffee trees cultivated (1888-90) – Reference group for the colonies: *Boa Esperança*

	Share of coffee trees											
VARIABLES	Blossoming			New			With Weed			Failed		
1.nonwhitehh	0.0276			-0.0325			0.341***			0.0387		
	(0.0474)			(0.0498)			(0.0849)			(0.0484)		
colorany		0.0106			-0.0116			0.114***			0.00547	
		(0.0181)			(0.0188)			(0.0283)			(0.0150)	
Non-Whites			0.0788			-0.0524			0.184***			-0.00246
			(0.0580)			(0.0398)			(0.0628)			(0.123)
Italians			0.0486			-0.0141			-0.163			-0.0505
			(0.0648)			(0.0633)			(0.102)			(0.116)
Germanic			0.208***			-0.184***			-0.0128			-0.0827
			(0.0610)			(0.0541)			(0.107)			(0.111)
Portuguese			0.0350			0.00730			0.145			-0.0280
			(0.0548)			(0.0777)			(0.186)			(0.108)
Spaniards			0.0549			-0.0199			-0.0856			-0.0393
			(0.0724)			(0.0733)			(0.117)			(0.124)
Easter Europeans			0.242***			-0.208***			-0.0143			-0.0917
			(0.0660)			(0.0632)			(0.123)			(0.118)
Literate	0.0448*	0.0430	0.0608*	-0.0347	-0.0337	-0.0406	0.121**	0.121**	0.0862**	-0.0817**	-0.0780**	-0.0952**
	(0.0255)	(0.0276)	(0.0358)	(0.0319)	(0.0336)	(0.0502)	(0.0536)	(0.0536)	(0.0408)	(0.0316)	(0.0302)	(0.0372)
Religion	0.0509**	0.0515**	-0.136***	-0.0518**	-0.0522**	0.137***	-0.0358	-0.0358	-0.135**	0.0607**	0.0579**	0.104**
	(0.0224)	(0.0227)	(0.0291)	(0.0222)	(0.0225)	(0.0291)	(0.0240)	(0.0240)	(0.0534)	(0.0265)	(0.0265)	(0.0447)
Grama Larga	-0.127**	-0.123**	-0.158***	0.122**	0.120**	0.150***	0.400***	0.400***	0.351***	0.0904*	0.0694	0.105**
	(0.0488)	(0.0546)	(0.0524)	(0.0510)	(0.0565)	(0.0565)	(0.107)	(0.107)	(0.102)	(0.0502)	(0.0481)	(0.0452)
Saltinho-Pinheiro	-0.958***	-0.954***	-0.970***	0.953***	0.950***	0.959***				0.167**	0.143**	0.166**
	(0.0482)	(0.0548)	(0.0590)	(0.0522)	(0.0583)	(0.0683)				(0.0648)	(0.0616)	(0.0727)
Teteia	-0.243***	-0.239***	-0.227***	0.237***	0.234***	0.219***				-0.0111	-0.0349	-0.0215
	(0.0566)	(0.0644)	(0.0628)	(0.0599)	(0.0673)	(0.0690)				(0.0594)	(0.0556)	(0.0676)
Col. Ibicaba	-0.0201	-0.0164	-0.0231	0.0153	0.0131	0.0153	0.234**	0.234**	0.188**	0.0931*	0.0724	0.101**
	(0.0485)	(0.0543)	(0.0495)	(0.0518)	(0.0572)	(0.0550)	(0.0935)	(0.0935)	(0.0833)	(0.0510)	(0.0471)	(0.0443)
Abundância	-0.0353	-0.0286	-0.00672	0.0482	0.0417	0.0401	0.183*	0.183*		0.0613	0.0428	0.0344
	(0.0490)	(0.0598)	(0.0543)	(0.0481)	(0.0599)	(0.0330)	(0.105)	(0.105)		(0.0731)	(0.0736)	(0.130)
Observations	101	101	102	99	99	100	67	67	68	99	99	100
R-squared	0.735	0.735	0.775	0.735	0.735	0.775	0.374	0.374	0.516	0.175	0.167	0.188
r2_a	0.702	0.702	0.733	0.701	0.701	0.732	0.275	0.275	0.400	0.0702	0.0618	0.0314

Notes: Same as Table IV.08.

VIII. Concluding remarks

This paper studied the emergence and consolidation of non-captive labor markets in the coffee economy of Southeastern Brazil. It used the newly founded archive from the *Ibicaba Project* to compile a dataset on the plantation's labor composition from 1888 to 1958 (with gaps). Exceptional for inaugurating the employment of indentured Europeans as of the 1850s, *Ibicaba* remained a major slave-based plantation until the very eve of Abolition. It kept its economic importance, until the economic decline experienced in the 1960s-1970s, *i.e.* beyond the period covered in this paper. From 1888 to 1958, *Ibicaba* took an intermediary position between its heyday and subsequent economic decline. In fact, in the period covered by this paper, *Ibicaba* can be considered as a representative plantation of São Paulo's Old West, *i.e.* the first agricultural frontier of the province by the mid-nineteenth century.

The new empirical evidence provided by this paper rejects the hypothesis of pure taste-based discrimination in *Ibicaba*'s agricultural labor market for fieldhands. Results showed no evidence of negative (against the Black population) or positive (in favor of European immigrants and their descendants) discrimination in prices negotiated with household heads once the quantity-based TFP of those households was controlled for. This is not at all to argue that racism was not present in most dimensions of Brazilian society in the immediate post-Abolition or in the decades that followed. The argument is purely economic: In a tight labor market with an inelastic demand for laborers and in strong competition with new, agronomically more productive agricultural frontiers, plantation managers could (or were not able) to discriminate by ethnolinguistic origins in their payments for homogenous, unskilled labor. In fact, some descriptive evidence from the immediate post-Abolition in *Ibicaba* even shows a positive view on Black families, probably due to their longer experience in coffee cultivation - and certainly for not requiring a period of acclimatization, as the newly arrived European immigrants did.

The analysis, however, showed strong evidence of statistical discrimination. The *colonato* contracts established a complex relationship between households' demographic composition and the length and timing of their engagement in cultivating and harvesting coffee. *Ceteris paribus*, plantation managers favored the hiring of households with more economically-active members, given the efficient allocation of labor within those households in a contract like the *colonato* and, most likely, for lowering transaction costs. The distinction between the number of coffee groves cultivated per household and per capita during the first two decades of the twentieth is clear evidence of this managerial strategy. Households headed by Black persons had the smallest number of economically-active individuals in them, a burden of slavery that majorly impacted their performance in the immediate post-Abolition.

Beyond discussions on labor market differentials, this paper also provided evidence of interest to agricultural history. The descriptive statistics of the factors of production in *Ibicaba*'s largest colonies and the analysis of the determinants of physical output in coffee production highlighted the importance of labor throughout the period considered in this paper. Technological innovations, if any, did not influence the labor per coffee grove in the

period considered. In fact, the long-term trends observed in total production in the coffee economy of *Ibicaba* had a quadratic format, with its peak around the 1910s-1920s, precisely the timing when São Paulo's Old West lost decisively its position as the world's main coffee producer to the state's newer and more fertile (*i.e.* less agronomically exhausted) agricultural frontiers. The estimates of the determinants of physical output in coffee cultivation and harvesting further highlighted the strong responsiveness of laborers to prices earned, as well as the highly significant and large economic impacts of the frequency of executed tasks and timing of harvest.

In short, this paper is a rejection of clear-cut "black and white" explanations about labor relations in Brazil, particularly in the post-Abolition era. Once again, this is not a revisionist perspective, as this paper thoroughly documented the economic hardships of the Black population throughout the period considered. These included poverty-stricken levels of income for a population who had just left centennial captivity and whose burdens were directly felt, including by the very structure of the households they inherited. The contribution of this paper is to better understand the specific mechanisms by which those Black laborers did not reap all the potential benefits that other ethnolinguistic groups attained, particularly over the long-run.

In a similar vein, the paper documented the varying paths of different European immigrants and their descendants. Portuguese and German-speaking immigrants showed a labor market performance similar to that of White Brazilians netted out from those origins. This result is explainable by the longer migratory history of these two nationalities to Brazil, in general, and to the region, in particular. Socioeconomic and cultural integration were clearly at play. Italians, on the other hand, were the only group to show evidence of a bimoldal distribution of their pooled (across all observations) income profiles. While more evidence is now due to understand the actual dynamics involved, this result is suggestive of an immigrant group that failed in their "South American dream". In complement to the classical sociological argument that, in Brazil, the richer, the whiter (by actual income and by historic sociological "whitening" of non-White high income people), this paper raises the hypothesis that the contrary could also be true, not simply by inverse logic, but also by downward socioeconomic mobility: Poor European immigrants and their impoverished descendants (in relative or absolute terms) participating in marriages markets where Non-Whites were a majority.

This paper explored only one facet of the complex economic relationships that characterized labor markets in the post-Abolition era, both in the short- and in the long-run. My approach was based on a sample that, albeit large in its coverage of the plantation's labor force, is still only a fraction from the material available in the *Ibicaba Collection*. Numerous different angles, based on that collection itself, remain completely unexplored.[1] As highlighted by the

[1] Even the same research question addressed in this paper could be replicated with other types of accounting, namely laborers' *Checking Accounts*. Beyond data on labor income, prices, number of coffee groves cultivated, and volume harvested, those ledgers contain information on consumption patterns, including intertemporal decisions, with loans advanced by the plantation. The methodology chosen for the current paper was less costly in terms of human (close-reading) transcriptions, but advances in automated transcription are promising for replication exercises of this type.

classical historiography, non-marketed income provided major incentives for rural laborers to sign *colonato* contracts. Yet, we know close to nothing (in quantitative terms) about the actual functioning of that contractual mechanism. More elaborate economic models of time usage, combined with the newly available data could shed some urgently needed light onto this issue. *Daybooks* and *General Accounting Ledgers* contain sporadic data on foodstuffs that households sold to *Ibicaba*'s central managers. *Laborers' Current Accounts* and *Stock Registries* in turn, provide information on the quantity of foodstuff that laborers bought in the plantation's storehouse and some of the monetary payments and loans they demanded. Together, these data on demand and supply of foodstuff and demand for currency can be used to compute the net demand for foodstuff of each household. Combined with a labor-leisure choice model and agronomical-technical estimates of the time required to produce certain subsistence crops in the early twentieth century, it would be possible to estimate the maximum yields of subsistence goods produced by each household In addition, practically all ledgers consulted hint upon the economic relevance of producing other other crops, including subsistence crops (maize, in particular) and cash crops (with sporadic references to cotton, sugarcane, and eucalyptus trees). Similarly, we know little to nothing (in quantitative terms) about non-agricultural rural employment in the plantations, although *Ibicaba* had a number of mills and workshops that constantly demanded labor and paid workers, usually hired from among its residing labor force, per piece rates or time worked. Even basic metrics on the length of the working year in a plantation, including in non-agricultural tasks, is unknown and could be derived from the previous analysis.

The expansion of the number of rural units with the *Accounting for the Countryside Project* will magnify these research possibilities and open completely new research areas, which can be only guessed currently. Among potential new fields are the environmental dimension of the coffee economy (via records from the *Floresta Estadual Edmundo Navarro de Andrade*); the quantification of agricultural history between plantations and comparatively to small rural units (via records from *Santa Gertrudes* plantation, *Casa Feltrin*, and the rural settlement of *Cascalho*); and, eventually, historical auditing of actual laborers' booklets. Moreover, a team of researchers is further surveying the historical records of plantations in Rio de Janeiro's *Paraíba Valley region*, *i.e.* the first zone of coffee expansion in Brazil, where slave labor was even more prevalent, in relative terms to other laborers, than in São Paulo's multiple agricultural frontiers.[2] Together with the already-available sources from the *Ibicaba Collection*, these sources will be made publicly available until the end of 2027 with basically zero consultation costs.

[2] The survey is led by Mariana Muaze, as part of the *Accounting for the Countryside Project*.

VII. Bibliography

Alden, D. (2004). "Late Colonial Brazil, 1750–1808". In Bethell, L. (ed.): *Colonial Brazil*. Cambridge: Cambridge University Press.

Andrews, G. R. (1988). "Black and white workers: São Paulo, Brazil, 1888–1928". *Hispanic American Historical Review*, Vol. 68 - No. 3.

Argollo Ferrão, A. M. (1999). "Colonos na fazenda Ibicaba, empresários em Piracicaba: A evolução sócio-econômica de um grupo de alemães (1850–1880)". *Proceedings of the 3rd Brazilian Congress of Economic History and 4th International Conference on Business History*.

Arrow, K. J. (1973). "The theory of discrimination". in Ashenfelter , O; and Rees, A. (eds.): *Discrimination in Labor Markets*. Princeton: Princeton University Press.

Balderas, J. U.; and Greenwood, M. J. (2010). "From Europe to the Americas: A comparative panel-data analysis of migration to Argentina, Brazil, and the United States, 1870–1910". *Journal of Population Economics*, Vol. 23 - No. 4.

Atkin, D.; Khandelwal, A. K.; and Osman A. (2017). "Exporting and firm performance: Evidence from a randomized experiment". *The Quarterly Journal of Economics*, Vol. 132 - No. 2.

Bassanezi, M. S. C. B. (1973). *Fazenda de Santa Gertrudes: Uma abordagem quantitativa das relações de trabalho, em uma propriedade rural paulista (1895-1930).* Rio Claro (SP): Faculdade de Filosofia, Ciências e Letras de Rio Claro (PhD thesis).

Bassanezi, M. S. C. B. (2019). *Colonos do café.* São Paulo: Editora Contexto.

Bassanezi, M. S. C. B.; Scott, A. N. S.; Bacellar, C. A. P.; and Truzzi, O. M. S. (2008). *Atlas da imigração internacional de São Paulo*. São Paulo: Editora da UNESP.

Becker, G. S. (1957). *The Economics of discrimination*. Chicago: University of Chicago Press.

Berrance de Castro, J. (1965). "A organização de um arquivo local: uma experiência". *Revista de História*, Vol. 30 – No. 62.

Berrance de Castro, J.; and Baiocco, S. (1969). "Inventário analítico de uma fazenda paulista: Santa Gertrudes (1885-1961)". *Anais do IV Simpósio Nacional dos Professores Universitários de História*.

Berrance de Castro, J.; and Witter, J. S. (1964). "Arrolamento das fontes históricas de Rio Claro". *Revista de História*, Vol. 28 – No. 58.

Bezerra, M. C. dos S. (2001). *Imigração, educação e religião: Um estudo histórico-sociológico do Bairro dos Pires de Limeira, uma comunidade rural de maioria teuto-brasileira.* Campinas: Universidade Estadual de Campinas (M.A. thesis).

Brandão, E. D. (1945). *Administração de fazenda. Contabilidade agrícola.* Rio de Janeiro (RJ): Serviço de Documentação, Ministério da Agricultura.

Buarque de Holanda, S. (1941). "Prefácio do tradutor". In Davatz, Thomas ([1858] 1941). *Memórias de um colono no Brasil (1850)*. São Paulo: Livraria Martins.

Callari, C. R.; and Baptista, M. R. de B. (1989). "Histórias de quem viveu a História. I - Da Escravidão". Rio Claro (SP): Arquivo do Município de Rio Claro (SP).

Carneiro, J.; and Carneiro, E. (1933). *Tratado de contabilidade. Vol. III - Contabilidade rural.* Rio de Janeiro: Calvino Filho Editor (2nd ed.)

Carvalho de Mello. P. (1978). "Aspectos econômicos da organização do trabalho da economia cafeeira do Rio de Janeiro, 1850-88". *Revista Brasileira de Economia*, Vol. 32 - No. 1.

Carvalho Filho, I.; and L. M. Monasterio (2012). “Immigration and the origins of regional inequality: Government-sponsored European migration to Southern Brazil before World War I”. *Regional Science and Urban Economics*, Vol. 42 - No. 5.

Carvalho Filho, I.; and Colistete, R. P. (2010). “Education performance: Was it all determined 100 years ago? Evidence from São Paulo, Brazil”. MPRA Working Paper No. 24494. Available here (last accessed on April 28, 2026).

Cavalcanti Rodrigues, L. (2021). *Sobre as cinzas do cativeiro: Educação, trabalho e terra entre ex-escravos e descendentes após a Abolição no Oeste Paulista*. São Paulo: Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo (PhD thesis).

Colistete, R. P. (2024). “Quem eram os trabalhadores do café nas vésperas da Abolição? Escravizados, imigrantes e nacionais em São Paulo em 1886-1887”. *FEA-USP Working Paper Series No 2024-27*. Available here (last accessed on April 28, 2026).

Colistete, R. P.; and Lamounier, M. L. (2014). “Land inequality in a coffee economy: São Paulo during the early twentieth century”. *FEA-USP Working Paper Series No 2024-27*. Available here (last accessed on April 28, 2026).

Davatz, T. ([1858] 1941). *Memórias de um colono no Brasil (1850)*. São Paulo: Livraria Martins Editora.

Dean, W. (1976). *Rio Claro: A Brazilian plantation system, 1820–1920*. Stanford: Stanford University Press.

Delfim Netto, A. ([1959] 1979). *O Problema do Café no Brasil*. Rio de Janeiro: Editora da Fundação Getúlio Vargas.

Eisenberg, P. L. (1983). “Escravo e proletário na história do Brasil”. *Estudos Econômicos (São Paulo)*, Vol. 13 - No. 1.

Faleiros, R. N. (2010). *Fronteiras do Café: fazendeiros e colonos no interior paulista (1917–1937)*. Bauru: Edusc/Fapesp.

Faro Leal, D. (1973). *Rio Claro e o café: Desenvolvimento, apogeu e crise (1850-1900)*. Rio Claro (SP): Faculdade de Filosofia, Ciências e Letras de Rio Claro (PhD thesis).

Fogel, R. W.; and Engerman, S. L. (1971). “The relative efficiency of slavery: A comparison of Northern and Southern agriculture in 1860”. *Explorations in Economic History*, Vol. 8 - No. 3.

Fornereto Mariano, F.; Nascimento, J. O. do; Nunes Pereira, A.; Moreira Zittei, M. V.; and Takao Ozaki, M. (2022). “Evidências de uma contabilidade rural histórica brasileira: Reconstituição, levantamento e correção monetária de ativos de fazendas cafeeiras paulistas nos idos de 1879”. In *Pesquisas e Inovações Multidisciplinares em Ciências Humanas e Sociais no Século XXI*. Goiânia: Instituto Scientia.

Foster, L.; Haltiwanger, J.; and Syverson, C. (2008). “Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?”. *American Economic Review*, Vol. 98 - No. 1.

Freitas Levy, C. de ([1993] 2023). “Crônica anedótica sentimental para a família”. In Masuti Levy, P. (org.): *Família Levy. Álbum de família*. Limeira: Ed. Biblioteca Paulo Masuti Levy.

Frost, A. K.; Verhoef, G.; and Durepos, G. (2026). “In the absence of archives: Epistemic repair and the remaking of business history”. *Business History*, preprint.

Heflinger Jr., J. E. (2007). *Ibicaba: O berço da imigração européia de cunho particular / Ibicaba: Die Wiege der privat organisierten europäischen Einwanderung*. Limeira: Unigráfica.

Heflinger Jr., J. E. (2009). *A Revolta dos Parceiros na Ibicaba / The Rebellion of the Sharecroppers in Ibicaba*. Limeira: Editora Unigráfica.

Heflinger Jr., J. E. (2021). “A decadência e a falência da Vergueiro & Cia. após a revolta dos colonos europeus (1856-1890)”. In Witzel de Souza, Bruno G.; and Santin Gardenal, Leonardo A. (Eds.) *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Editora Pontes.

Hellerstein, J. K.; Neumark, D.; and Troske, K. R. (1999). “Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations”. *Journal of Labor Economics*, Vol. 17 - No. 3.

Holloway, T. H. (1978). “Creating the reserve army? The immigration program of São Paulo, 1886–1930”. *International Migration Review*, Vol. 12 - No. 2.

Holloway, T. H. (1980). *Immigrants on the land: Coffee and society in São Paulo, 1886–1934*. Chapel Hill: University of North Carolina Press.

Jann, B. (2008). “The Blinder-Oaxaca decomposition for linear regression models”. *The Stata Journal*, Vol. 8 - No. 4.

Jenkins, J. A.; and Rubin, J. (2024). “Historical political economy: What is it?”. In Jenkins, J. A.; and Rubin, J. (eds.): *The Oxford Handbook of Historical Political Economy*. New York: Oxford University Press.

Kamphoefner, W. (2000). “Südamerika als Alternative? Bestimmungsfaktoren der deutschen Überseewanderung im 19. Jahrhundert. *Jahrbuch für Wirtschaftsgeschichte / Economic History Yearbook*, Vol. 41 - No. 1.

Klein, H. S. (1995). “European and Asian migration to Brazil”. In Cohen, R. (ed.): *The Cambridge Survey of World Migration*. Cambridge: Cambridge University Press.

Klein, H. S.; and Luna, F. V. (2010). *Slavery in Brazil*. New York: Cambridge University Press.

Kuntz Busch, R. (1967). *História de Limeira, Vol. I*. São Paulo: Edição da Prefeitura Municipal de Limeira.

Lampe, M.; and Sharp, P. (2019). “Accounting for the wealth of Denmark: a case study of Smithian growth using the emergence of modern accounting in Danish dairying”. *The European Journal of the History of Economic Thought*, Vol. 26 - No. 4.

Lamounier, M. L. (1986). *Formas de transição da escravidão ao trabalho livre: a Lei de Locação de Serviços de 1879*. Campinas (SP): Universidade Estadual de Campinas (M.A. thesis).

Lanza, A. (2021). *De braços para a lavoura à proprietários rurais: imigrantes e acesso a terras em São Paulo, 1886-1920*. São Paulo: Universidade de São Paulo (PhD thesis).

Levy, M. S. F. (1974). “O papel da imigração internacional na evolução da população brasileira (1872 a 1972)”. *Revista de Saúde Pública*, Vol. 8 - Suplemento.

Marcondes, R. L. (2021). “A escravidão persiste no Oeste Paulista: Limeira, 1875”. In Witzel de Souza, Bruno G.; and Santin Gardenal, Leonardo A. (Eds.) *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Editora Pontes.

Matthews, D.; Anderson, M.; and Edwards, J. R. (1997). “The rise of the professional accountant in British management”. *The Economic History Review*, Vol. 50 - No. 3.

Mendes, F. L. R. (2017). “Ibicaba revisitada outra vez: espaço, escravidão e trabalho livre no oeste paulista”. *Anais do Museu Paulista*, Vol. 25 - No. 1.

Monasterio, L. (2017). “Surnames and ancestry in Brazil”. *PLoS ONE*, Vol. 12 - No. 5.

Monasterio, L.; and Lopes, D. (2018). “Brazil without immigrants: microdata long run estimates”. *MPRA Paper* No. 88170. Available here (last accessed on April 28, 2026).

Monsma, K. (2010). “Vantagens de imigrantes e desvantagens de negros: Emprego, propriedade, estrutura familiar e alfabetização depois da abolição no Oeste Paulista”. *Dados*, Vol. 53 - No. 3.

Monsma, K. (2017). “Como pensar o racismo: o paradigma colonial e a abordagem da sociologia histórica”. *Revista de Ciências Sociais*, Vol. 48 - No. 2.

Monti, C. G.; and Oliveira, L. L. de (2025). “Colonato e imigração: padrões de produção e diversificação dos ganhos na fazenda Boa Vista (Ribeirão Preto, 1915)”. *Revista Mundos do Trabalho*, Vol. 17.
Olivatto, F. F. (2021). “Francisco Frederico Olivatto. Eternizando o Morro Alto em suas telas” [Interview]. In Witzel de Souza, B. G.; and Santin Gardenal, L. A. (orgs.): *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Pontes Editores.
Olmstead, A. L.; and Rhode, P. W. (2008). “Biological innovation and productivity growth in the Antebellum cotton economy”. *Journal of Economic History*, Vol. 68 - No. 4.
Palma, N.; and Lambais, G. (2026). “How a nation was born: Brazilian economic growth, 1574-1920”. *Lewis Lab Working Paper* No. 2026-03. Available here (last accessed on April 28, 2026)
Pérez, S. (2017). “The (South) American Dream: Mobility and economic outcomes of first and second-generation immigrants in nineteenth-century Argentina”. *The Journal of Economic History*, Vol. 77 - No. 4.
Pérez Meléndez, J. J. (2024). “The meanings of restriction: the *Heydt Rescript* of 1859 and German migrations to Brazil”. *História Econômica & História de Empresas*, Vol. 27 - No. 3.
Perret-Gentil, C. (1851). *A Colônia Senador Vergueiro*. Santos: Typographia Imparcial.
Phelps, E. S. (1972). “The statistical theory of racism and sexism”. *American Economic Review*, Vol. 62 - No. 4.
Pissinato, B.; and Vian, C- E. de F. (2021). “Um estudo da evolução da fazenda Ibicaba: Contribuição para a história econômica”. In Witzel de Souza, Bruno G.; and Santin Gardenal, Leonardo A. (Eds.) *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Editora Pontes.
Pollard, S. (1983). “Capitalism and rationality: A study of measurements in British coal mining, ca. 1750–1850”. *Explorations in Economic History*, Vol. 20 - No. 1.
Pott, C.; and Pott, F. (2021). “Claudio Pott & Felisberto Pott. Trabalho e memórias em cada canto do Ibicaba” [Interview]. In Witzel de Souza, B. G.; and Santin Gardenal, L. A. (orgs.): *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Pontes Editores.
Preston James, E. (1932). “The coffee lands of southeastern Brazil”. *Geographical Review*, Vol. 22 - No. 2.
Reber, V. B. (1985). “Archival sources for Latin American Business History”. *The Business History Review*, Vol. 59 - No. 4.
Rocha, R.; Ferraz, C.; and Soares, R. R. (2017). “Human capital persistence and development”. *American Economic Journal: Applied Economics*, Vol. 9 - No. 4.
Rodrigues, A.; and McBride, K. (2026). “Narratives of exploitation and resistance: European immigration, business, and accounting on Ibicaba farm in the 19th century”. *Business History* (pre-print).
Sánchez-Alonso, B. S. (2007). “The other Europeans: Immigration into Latin America and the international labour market (1870–1930)”. *Revista de História Económica, Journal of Iberian and Latin American Economic History*, Vol. 25 - No. 3.
Sánchez-Alonso, B. S. (2019). “The Age of Mass Migration in Latin America”. *The Economic History Review*, Vol. 72 - No. 1.
Sangster, A. (2025). “The emergence of double entry bookkeeping”. *Economic History Review*, Vol. 78.
Saraiva, L. F.; and Freire, J. (2016). “‘Pertence à liberta Firmina, 09 de maio de 1888’: Relações de crédito na transição do trabalho escravo para o livre em uma região cafeeira - Zona da Mata mineira”. In Guimarães, C. G.; and Saraiva, L. F. (eds.):

Crédito e descrédito: Relações sociais de empréstimos na América - Séculos XVIII ao XX. Niterói (RJ): Editora da Universidade Federal Fluminense.
Schmitz, M. D.; and Schaefer, D. F. (1978). "Slavery, freedom, and the elasticity of substitution between slave and free labor in the Antebellum South". *Explorations in Economic History*, Vol. 15 - No. 3.
Simonsen, R. C. ([1937] 1978). *História Econômica do Brasil (1500–1820)*. São Paulo: Companhia Editora Nacional.
Slenes, R. W. (1983). "O que Rui Barbosa não queimou: novas fontes para o estudo da escravidão no século XIX". *Estudos Econômicos (São Paulo)*, Vol. 13 - No. 1.
Slenes, R. W. (2011). *Na senzala, uma Flor: Esperanças e recordações da família escrava - Brasil, Sudeste - século XIX*. Rio de Janeiro: Nova Fronteira.
Stein, S. J. ([1957] 1986). *Vassouras: A Brazilian Coffee County, 1850–1900*. Princeton: Princeton University Press (2nd edition).
Stolcke, V. (1988). *Coffee Planters, Workers and Wives: Class Conflict and Gender Relations on São Paulo Plantations, 1850–1980*. London: Macmillan / New York: St. Martin's Press.
Stolcke, V.; and Hall, M. M. (1983). "The introduction of free labour on São Paulo coffee plantations". *The Journal of Peasant Studies*, Vol. 10 - No. 1.
Stolz, Y.; Baten, J.; and Botelho, T. (2013). "Growth effects of nineteenth-century mass migrations: 'Fome Zero' for Brazil?". *European Review of Economic History*, Vol. 17 - No. 1.
Tamiazo, P. C. (2021). "A fazenda Ibicaba, a família Levy e a comunidade de Cordeirópolis". In Witzel de Souza, Bruno G.; and Santin Gardenal, Leonardo A. (Eds.) *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Editora Pontes.
Tonelotto, L. (2021). "Luiz Tonelotto. Vide e trabalho entre o Ibicaba e o Conde" [Interview]. In Witzel de Souza, B. G.; and Santin Gardenal, L. A. (orgs.): *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Pontes Editores.
Tschudi, J. J. von ([1866] 1980). *Viagem às províncias do Rio de Janeiro e S. Paulo*. São Paulo: Martins / Edusp.
Viotti da Costa, E. ([1966] 1998). *Da Senzala à Colônia*. São Paulo: Editora UNESP (4th edition).
Witter, J. S. (1971). "Um documento para a história dos preços no Brasil: Descrição e avaliação da Fazenda de Ibicaba (1890)". *Revista de História*, Vol. 43 - No. 87.
Witter, J. S. (1982). *Ibicaba, uma experiência pioneira*. São Paulo: Arquivo do Estado (2ndedition).
Witzel de Souza, B. G. (2011). *Liberdade ou grilhões? Um estudo dos contratos de parceria à luz da imigração germânica em São Paulo, 1840-1870*. São Paulo (SP): Universidade de São Paulo (BA dissertation).
Witzel de Souza, B. G. (2012). "Imigração alemã e mercado de trabalho na cafeicultura paulista – Um estudo quantitativo dos contratos de parceria". *História Econômica & História de Empresas*, Vol. 15 - No. 2.
Witzel de Souza, B. G. (2019). *From bonded laborers to educated citizens? Immigration, labor markets, and human capital in São Paulo, Brazil (1820-2010)*. Göttingen: University of Göttingen (PhD thesis).
Witzel de Souza, B. G. (2022). "Precedents of mass migration: Policies, occupations, and the sorting of foreigners in São Paulo, Brazil (1872)". *Economic History of Developing Regions*, Vol. 37 - No. 1.
Witzel de Souza, B. G. (2024). "'The same contract that is suitable for your Excellency': Immigration and emulation in the adoption of sharecropping-cum-debt arrangements in Brazil (1835-80)". *Economic History Review*.

Witzel de Souza, B. G.; and Santin Gardenal, L. A. (orgs.) (2021). *Ibicaba (1817-2017). Entendendo, vivendo e construindo futuros*. Campinas: Pontes Editores.

Witzel de Souza, B. G.; and Santin Gardenal, L. A. (2026). "The archives of *Ibicaba* plantation, Brazil (1846-1977): New sources of microdata & guidelines on creating physical and digital collections from rural archives". FEA-USP Working Paper Series - No. 2026-14. Available here (last accessed on May 07, 2026).

Yamey, B. S. (1949). "Scientific bookkeeping and the rise of capitalism". *The Economic History Review*, Vol. 1 - Nos. 2-3.

Zamberlan Pereira, T. (2020). "The North–South divide: Real wages and welfare in Brazil during the early 20th century". *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, Vol. 38 - No. 1.

Table Appendix 01 – Robustness checks: Alternative specifications of household demographic composition (Lean and Harvest Seasons)

	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
			$\ln(Y_{ht}^{cult})$					$\ln(Y_{ht}^{harv})$		
ln(HH size)	0.0337	0.255**	-0.112*	-0.115*	0.297*	-0.0912	-0.101	-0.0987	-0.0912	-0.0534
	(0.0544)	(0.124)	(0.0595)	(0.0687)	(0.174)	(0.0916)	(0.190)	(0.102)	(0.114)	(0.310)
ln(# members 7-60y)	0.334***	0.269***	0.327***	0.331***	0.266***	0.404***	0.496***	0.399***	0.401***	0.495***
	(0.0408)	(0.0830)	(0.0372)	(0.0388)	(0.0825)	(0.0721)	(0.180)	(0.0709)	(0.0729)	(0.183)
ln(# families)	-0.0163			0.00285		0.00430			0.00596	
	(0.0318)			(0.0326)		(0.0292)			(0.0304)	
ln(# married)		-0.00709			-0.0117		-0.0113			-0.0187
		(0.0674)			(0.0700)		(0.152)			(0.148)
ln(# widows)		-0.0926			-0.0942		-0.0649			-0.0671
		(0.0770)			(0.0767)		(0.179)			(0.180)
ln(# males)			0.150***	0.153***	-0.0405			0.000360	0.000468	-0.0421
			(0.0393)	(0.0407)	(0.0869)			(0.0596)	(0.0612)	(0.179)
Observations	1,464	196	1,466	1,460	195	1,464	196	1,466	1,460	195
R-squared	0.844	0.891	0.845	0.844	0.891	0.844	0.891	0.845	0.844	0.891
Number of id	322	57	324	322	56	322	57	324	322	56

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 02 – Robustness checks: Alternative specifications of household demographic composition (Lean and Harvest Seasons)

	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
			$\ln(Y_{ht}^{cult})$					$\ln(Y_{ht}^{harv})$		
ln(HH size)	0.354***	0.367***			0.175***	0.351***	0.445***			0.260***
	(0.0307)	(0.0559)			(0.0503)	(0.0411)	(0.147)			(0.0735)
ln(# members 7-60y)	-0.00836			-0.00283		-0.0163			0.00199	
	(0.0288)			(0.0299)		(0.0299)			(0.0321)	
ln(# families)		0.0562			0.129		-0.0292			0.0971
		(0.0729)			(0.0893)		(0.152)			(0.160)
ln(# married)		-0.0882			0.00728		-0.0658			0.0533
		(0.0800)			(0.0859)		(0.181)			(0.187)
ln(# widows)										
ln(# males)			0.161***	0.280***	0.208***			0.0128	0.189***	0.149
			(0.0450)	(0.0305)	(0.0785)			(0.0660)	(0.0454)	(0.125)
Observations	2,338	293	2,400	2,350	292	1,627	214	1,664	1,639	213
R-squared	0.495	0.639	0.437	0.428	0.530	0.815	0.877	0.802	0.800	0.864
Number of id	508	79	539	508	78	332	59	346	334	58

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 03 – Robustness checks: Alternative age ranges for economically-active household members (Lean and Harvest Seasons)

	$\ln(Y_{ht}^{cult})$						$\ln(Y_{ht}^{harv})$					
	Age ranges for *lenxada*						Age ranges for *lenxada*					
	17-45y	17+y	10-45y	10+y	7-45y	7+y	17-45y	17+y	10-45y	10+y	7-45y	7+y
ln(HH size)	0.305***	0.241***	0.0687*	0.0126	0.0549	-0.0375	0.289***	0.193***	-0.0360	-0.123**	-0.0195	-0.173**
	(0.0395)	(0.0422)	(0.0409)	(0.0394)	(0.0473)	(0.0517)	(0.0528)	(0.0582)	(0.0671)	(0.0617)	(0.0832)	(0.0861)
ln(# members)	0.131***	0.172***	0.356***	0.440***	0.316***	0.436***	0.0807**	0.207***	0.408***	0.576***	0.325***	0.545***
	(0.0263)	(0.0360)	(0.0310)	(0.0393)	(0.0381)	(0.0524)	(0.0406)	(0.0530)	(0.0543)	(0.0631)	(0.0677)	(0.0909)
Observations	2,268	2,376	2,307	2,378	2,309	2,378	1,416	1,485	1,441	1,486	1,443	1,486
R-squared	0.854	0.847	0.872	0.867	0.862	0.859	0.838	0.838	0.850	0.852	0.843	0.845
Number of id	501	512	507	512	507	512	313	326	319	326	319	326

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 04 – Robustness checks: Alternative age ranges for economically-active household members, incl. 7-60y (Lean and Harvest Seasons)

	$\ln(Y_{ht}^{cult})$							$\ln(Y_{ht}^{harv})$						
	Age ranges for *lenxada*							Age ranges for *lenxada*						
	17-45y	17+y	10-45y	10+y	7-45y	7+y	7-60y	17-45y	17+y	10-45y	10+y	7-45y	7+y	7-60y
ln(HH size)	0.304***	0.240***	0.0674*	0.0125	0.0534	-0.0381	0.0175	0.287***	0.192***	-0.0366	-0.121*	-0.0211	-0.173**	-0.103
	(0.0393)	(0.0421)	(0.0409)	(0.0394)	(0.0474)	(0.0518)	(0.0484)	(0.0534)	(0.0586)	(0.0670)	(0.0619)	(0.0833)	(0.0861)	(0.0847)
ln(# members)	0.129***	0.171***	0.356***	0.439***	0.316***	0.436***	0.336***	0.0759*	0.202***	0.405***	0.570***	0.322***	0.542***	0.402***
	(0.0264)	(0.0361)	(0.0310)	(0.0393)	(0.0382)	(0.0525)	(0.0396)	(0.0407)	(0.0532)	(0.0542)	(0.0630)	(0.0673)	(0.0905)	(0.0700)
Observations	2,266	2,374	2,305	2,376	2,307	2,376	2,356	1,414	1,483	1,439	1,484	1,441	1,484	1,468
R-squared	0.854	0.847	0.872	0.867	0.862	0.859	0.861	0.837	0.837	0.850	0.852	0.843	0.845	0.844
Number of id	501	512	507	512	507	512	512	313	326	319	326	319	326	324

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 05 – Pairwise correlations between alternative age ranges for economically-active household members

	17-45y	17+y	10-45y	10+y	7-45y	7+y	7-60y
17-45y	1.0000						
17+y	0.8803*	1.0000					
10-45y	0.8143*	0.7371*	1.0000				
10+y	0.7617*	0.8495*	0.9331*	1.0000			
7-45y	0.7645*	0.6828*	0.9626*	0.8893*	1.0000		
7+y	0.7409*	0.8055*	0.9271*	0.9596*	0.9525*	1.0000	
7-60y	0.7325*	0.7483*	0.9398*	0.9350*	0.9654*	0.9796*	1.0000

Notes: (1) Pairwise correlation coefficients reported below the diagonal. (2) * denotes statistical significance at $p<0.05$.

Table Appendix 06 – Robustness checks: Alternative dependent variables for physical output in coffee cultivation (Panel A – Physical output and labor income)

Dep. var.	ltree_tot_ check	ltree_tot_ check02	ltree_tot_ estaca	ltree_tot_ estaca_check
ln(HH size)	0.0278	0.0377	0.0313	0.0474
	(0.0483)	(0.0526)	(0.0476)	(0.0504)
ln(# members 7-60y)	0.332***	0.345***	0.318***	0.316***
	(0.0403)	(0.0435)	(0.0409)	(0.0431)
Mean (estaca)	2.45e-06	3.67e-06	-3.48e-06	-8.37e-06
	(2.64e-05)	(2.72e-05)	(2.68e-05)	(2.74e-05)
MDEV (estaca)	-4.84e-05	4.34e-05	-0.000154*	-0.000132
	(8.69e-05)	(0.000115)	(9.15e-05)	(9.69e-05)
SD (estaca)	3.53e-05	-3.09e-05	0.000121*	0.000115
	(6.17e-05)	(8.32e-05)	(6.66e-05)	(7.13e-05)
Grama Larga	0.133**	0.143**	0.127**	0.140**
	(0.0565)	(0.0595)	(0.0572)	(0.0596)
Lage	0.215***	0.226***	0.211***	0.155*
	(0.0721)	(0.0764)	(0.0745)	(0.0870)
Morro Alto	0.121*	0.139**	0.121*	0.144**
	(0.0620)	(0.0645)	(0.0622)	(0.0640)
Pão de Ló	0.185**	0.198**	0.183**	0.202***
	(0.0771)	(0.0771)	(0.0781)	(0.0780)
Salt. & Pinh.	0.145*	0.218**	0.127	0.194**
	(0.0797)	(0.0995)	(0.0798)	(0.0914)
Teteia	0.202***	0.211***	0.199***	0.210***
	(0.0653)	(0.0661)	(0.0645)	(0.0663)
Col. Ibicaba	-0.294**	-0.256*	-0.299**	-0.285**
	(0.140)	(0.150)	(0.142)	(0.142)
Linha	–	–	–	–
Mariquita	0.295***	0.275***	0.309***	0.270***
	(0.0292)	(0.0308)	(0.0297)	(0.0352)
Paineira	0.128***	0.124***	-0.121***	-0.216***
	(0.0253)	(0.0364)	(0.0234)	(0.0302)
Portão Isabel	0.465***	0.351***	0.499***	0.369***
	(0.0265)	(0.0368)	(0.0271)	(0.0347)
Iracema farm	-0.105	-0.0982	-0.107	-0.108
	(0.0651)	(0.0672)	(0.0659)	(0.0694)
Observations	2,360	2,316	2,359	2,317
R-squared	0.494	0.473	0.471	0.451
Number of id	513	506	512	505

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 07 – Robustness checks: Alternative dependent variables – estaca dispersion measures (Panel A – Physical output and labor income)

Dep. var.	ltree_ estacamad	ltree_ estacamad_ check	ltree_ estacamdev	ltree_ estacamdev_ check	ltree_ estacasd	ltree_ estacasd_ check
ln(HH size)	0.0944	0.106	0.00340	0.0389	0.00946	0.0470
	(0.183)	(0.190)	(0.0891)	(0.0988)	(0.0848)	(0.0955)
ln(# members 7-60y)	-0.0792	-0.0736	0.0329	0.0206	0.0405	0.0235
	(0.154)	(0.156)	(0.0861)	(0.0862)	(0.0834)	(0.0840)
Mean (estaca)	0.000258***	0.000267***	0.000248***	0.000253***	0.000250***	0.000253***
	(6.40e-05)	(6.68e-05)	(3.37e-05)	(4.09e-05)	(3.37e-05)	(4.11e-05)
MDEV (estaca)	0.000766	0.000549	-0.000162	-0.000209	-0.000217	-0.000316
	(0.000556)	(0.000558)	(0.000255)	(0.000275)	(0.000248)	(0.000283)
SD (estaca)	-0.000543	-0.000409	5.04e-05	7.32e-05	9.35e-05	0.000158
	(0.000349)	(0.000351)	(0.000154)	(0.000161)	(0.000152)	(0.000164)
Grama Larga	-0.336	-0.369	-0.136	-0.152	-0.129	-0.144
	(0.232)	(0.227)	(0.126)	(0.126)	(0.118)	(0.120)
Lage	-0.427*	-0.405*	-0.100	-0.0814	-0.0372	-0.0267
	(0.239)	(0.232)	(0.132)	(0.127)	(0.118)	(0.118)
Morro Alto	-0.198	-0.224	-0.118	-0.130	-0.107	-0.117
	(0.243)	(0.241)	(0.121)	(0.122)	(0.117)	(0.119)
Pão de Ló	-0.949**	-0.977**	-0.889***	-0.909***	-0.916***	-0.933***
	(0.454)	(0.450)	(0.196)	(0.200)	(0.190)	(0.195)
Salt. & Pinh.	0.549*	0.551*	0.323	0.304	0.240	0.212
	(0.295)	(0.324)	(0.236)	(0.259)	(0.236)	(0.260)
Teteia	-0.852	-0.910	-0.0351	-0.0565	-0.0867	-0.0921
	(0.596)	(0.632)	(0.193)	(0.204)	(0.162)	(0.168)
Col. Ibicaba	-0.340	-0.390	0.256	0.213	0.268	0.229
	(0.660)	(0.694)	(0.208)	(0.221)	(0.208)	(0.220)
Linha	–	–	–	–	–	–
Mariquita	-0.0534	-0.0491	-0.0503	-0.0476	-0.0132	-0.0141
	(0.123)	(0.122)	(0.0606)	(0.0614)	(0.0529)	(0.0543)
Paineira	0.431***	0.856***	-0.674***	0.920***	-0.761***	0.936***
	(0.0968)	(0.0874)	(0.0518)	(0.0455)	(0.0477)	(0.0417)
Portão Isabel	-0.590***	0.0297	-0.299***	-0.192***	-0.251***	-0.173**
	(0.106)	(0.138)	(0.0541)	(0.0718)	(0.0505)	(0.0673)
Iracema farm	-0.849**	-0.867**	-0.325*	-0.335*	-0.323**	-0.333**
	(0.366)	(0.371)	(0.185)	(0.187)	(0.162)	(0.164)
Observations	2,336	2,271	2,359	2,294	2,359	2,294
R-squared	0.098	0.094	0.100	0.099	0.104	0.104
Number of id	510	502	512	504	512	504

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 08 – Robustness checks: Alternative dependent variables – estaca-based means and medians (Panel A – Physical output and labor income)

Dep. var.	ltree_ estacamean	ltree_ estacamean_ check	ltree_ estacamed	ltree_ estacamed_ check
ln(HH size)	0.0100	0.00948	0.0173	0.0189
	(0.0302)	(0.0310)	(0.0295)	(0.0300)
ln(# members 7-60y)	0.0388	0.0390	0.0273	0.0289
	(0.0294)	(0.0296)	(0.0286)	(0.0288)
Mean (estaca)	0.000179***	0.000178***	0.000159***	0.000158***
	(8.77e-06)	(8.96e-06)	(1.02e-05)	(1.02e-05)
MDEV (estaca)	-0.000180***	-0.000223***	-0.000130***	-0.000160***
	(4.56e-05)	(5.09e-05)	(3.26e-05)	(2.68e-05)
SD (estaca)	2.90e-05	6.17e-05	1.17e-05	3.49e-05*
	(3.54e-05)	(4.03e-05)	(2.40e-05)	(2.11e-05)
Grama Larga	0.159***	0.159***	0.115***	0.118***
	(0.0427)	(0.0433)	(0.0418)	(0.0422)
Lage	0.185***	0.182***	0.109**	0.100**
	(0.0390)	(0.0394)	(0.0462)	(0.0428)
Morro Alto	0.143***	0.144***	0.105**	0.106**
	(0.0447)	(0.0454)	(0.0431)	(0.0436)
Pão de Ló	0.0936	0.0932	0.171**	0.171**
	(0.0650)	(0.0651)	(0.0732)	(0.0730)
Salt. & Pinh.	0.136***	0.126**	0.150**	0.135*
	(0.0492)	(0.0536)	(0.0650)	(0.0712)
Teteia	0.249***	0.247***	0.168***	0.167***
	(0.0442)	(0.0465)	(0.0428)	(0.0450)
Col. Ibicaba	0.183	0.186	0.197	0.197
	(0.113)	(0.117)	(0.134)	(0.138)
Linha	–	–	–	–
Mariquita	0.0418	0.0422	-0.0288*	-0.0315*
	(0.0281)	(0.0289)	(0.0174)	(0.0168)
Paineira	-0.106***	0.198***	0.0246	0.0254*
	(0.0119)	(0.0103)	(0.0183)	(0.0149)
Portão Isabel	-0.0553***	-0.0546**	-0.149***	-0.135***
	(0.0178)	(0.0233)	(0.0175)	(0.0232)
Iracema farm	-0.198**	-0.198**	-0.274**	-0.274**
	(0.0816)	(0.0819)	(0.137)	(0.137)
Observations	2,359	2,294	2,359	2,294
R-squared	0.269	0.269	0.224	0.230
Number of id	512	504	512	504

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 09 – Robustness checks: Alternative dependent variables – estaca counts (Panel A – Physical output and labor income)

Dep. var.	ltree_tot_check	lestaca_count	lestaca_count_check
ln(HH size)	0.0278	0.0213	0.0165
	(0.0483)	(0.0488)	(0.0502)
ln(# members 7-60y)	0.332***	0.279***	0.286***
	(0.0403)	(0.0415)	(0.0409)
Mean (estaca)	2.45e-06	-0.000183***	-0.000182***
	(2.64e-05)	(2.27e-05)	(2.34e-05)
MDEV (estaca)	-4.84e-05	2.63e-05	0.000142*
	(8.69e-05)	(7.93e-05)	(8.02e-05)
SD (estaca)	3.53e-05	9.19e-05	8.99e-06
	(6.17e-05)	(5.78e-05)	(5.61e-05)
Grama Larga	0.133**	-0.0319	-0.0370
	(0.0565)	(0.0551)	(0.0564)
Lage	0.215***	0.0262	0.0339
	(0.0721)	(0.0866)	(0.0887)
Morro Alto	0.121*	-0.0220	-0.0132
	(0.0620)	(0.0666)	(0.0683)
Pão de Ló	0.185**	0.0899	0.0912
	(0.0771)	(0.0826)	(0.0826)
Salt. & Pinh.	0.145*	-0.00826	0.0288
	(0.0797)	(0.0666)	(0.0671)
Teteia	0.202***	-0.0498	-0.0429
	(0.0653)	(0.0650)	(0.0675)
Col. Ibicaba	-0.294**	-0.482***	-0.459***
	(0.140)	(0.154)	(0.165)
Linha	–	–	–
Mariquita	0.295***	0.267***	0.255***
	(0.0292)	(0.0376)	(0.0401)
Paineira	0.128***	-0.0155	0.0578**
	(0.0253)	(0.0238)	(0.0247)
Portão Isabel	0.465***	0.555***	0.440***
	(0.0265)	(0.0305)	(0.0371)
Iracema farm	-0.105	0.0904	0.0903
	(0.0651)	(0.111)	(0.111)
Observations	2,360	2,359	2,294
R-squared	0.494	0.448	0.471
Number of id	513	512	504

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 10 – Robustness checks: Alternative estaca-based control specifications – Dependent variable: lincometree_annual

	lincometree_annual									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(HH size)	0.0186	0.0192	0.0187	0.0128	0.0140	0.0127	0.0158	0.00999	0.0231	0.0182
	(0.0484)	(0.0477)	(0.0485)	(0.0491)	(0.0483)	(0.0492)	(0.0474)	(0.0479)	(0.0469)	(0.0475)
ln(# members 7-60y)	0.335***	0.338***	0.335***	0.339***	0.341***	0.339***	0.339***	0.342***	0.329***	0.332***
	(0.0397)	(0.0398)	(0.0397)	(0.0403)	(0.0403)	(0.0402)	(0.0395)	(0.0401)	(0.0390)	(0.0398)
estaca_mean	1.92e-06	1.52e-07	1.33e-07						-1.98e-05	-0.000281***
	(2.60e-05)	(2.62e-05)	(2.56e-05)						(2.22e-05)	(8.50e-05)
estaca_mdev	-4.53e-05		3.39e-06							
	(8.50e-05)		(2.51e-05)							
estaca_sd	3.34e-05		2.55e-06							0.000127***
	(6.03e-05)		(1.77e-05)							(3.70e-05)
estaca_mean_check				3.54e-06	2.01e-06	2.61e-06		-2.60e-05		-0.000298***
				(2.64e-05)	(2.66e-05)	(2.63e-05)		(2.31e-05)		(9.05e-05)
estaca_mdev_check				-4.09e-05		-1.22e-06				
				(0.000115)		(2.62e-05)				
estaca_sd_check				2.72e-05		-1.52e-06				0.000137***
				(8.47e-05)		(1.91e-05)				(4.16e-05)
estaca_mad							9.63e-05		0.000588***	
							(7.66e-05)		(0.000175)	
estaca_mad_check								0.000113		0.000618***
								(7.85e-05)		(0.000182)
Grama Larga	0.143**	0.140**	0.142**	0.148**	0.145**	0.147**	0.134**	0.137**	0.0736	0.0730
	(0.0568)	(0.0567)	(0.0568)	(0.0573)	(0.0571)	(0.0573)	(0.0568)	(0.0570)	(0.0583)	(0.0591)
Lage	0.213***	0.214***	0.212***	0.212***	0.212***	0.212***	0.211***	0.208***	0.155**	0.148**
	(0.0712)	(0.0705)	(0.0711)	(0.0720)	(0.0714)	(0.0720)	(0.0702)	(0.0710)	(0.0729)	(0.0739)
Morro Alto	0.129**	0.128**	0.128**	0.136**	0.135**	0.135**	0.121*	0.126**	0.0496	0.0501
	(0.0620)	(0.0619)	(0.0621)	(0.0626)	(0.0625)	(0.0626)	(0.0624)	(0.0629)	(0.0650)	(0.0658)
Pão de Ló	0.189**	0.189**	0.190**	0.194**	0.194**	0.195***	0.180**	0.188**	0.248***	0.257***
	(0.0751)	(0.0750)	(0.0750)	(0.0752)	(0.0751)	(0.0752)	(0.0733)	(0.0734)	(0.0743)	(0.0747)
Salt. & Pinh.	0.157**	0.157**	0.157**	0.171**	0.171**	0.172**	0.159**	0.177**	0.207***	0.217**
	(0.0790)	(0.0792)	(0.0790)	(0.0853)	(0.0851)	(0.0848)	(0.0794)	(0.0849)	(0.0780)	(0.0841)
Teteia	0.200***	0.205***	0.200***	0.214***	0.219***	0.214***	0.193***	0.205***	0.122*	0.134*
	(0.0644)	(0.0633)	(0.0642)	(0.0658)	(0.0648)	(0.0657)	(0.0635)	(0.0652)	(0.0674)	(0.0689)
Col. Ibicaba	-0.269**	-0.212*	-0.269**	-0.242*	-0.188	-0.242*	-0.215*	-0.192	-0.316**	-0.294*
	(0.134)	(0.124)	(0.134)	(0.146)	(0.132)	(0.146)	(0.126)	(0.134)	(0.154)	(0.167)
Linha	–	–	–	–	–	–	–	–	–	–
Mariquita	0.277***	0.277***	0.276***	0.274***	0.274***	0.274***	0.282***	0.280***	0.297***	0.294***
	(0.0303)	(0.0299)	(0.0302)	(0.0306)	(0.0303)	(0.0305)	(0.0297)	(0.0301)	(0.0292)	(0.0298)
Paineira	0.0876***	0.0887***	0.0854***	0.199***	0.208***	0.206***	0.0873***	0.214***	0.0978***	0.167***
	(0.0282)	(0.0277)	(0.0282)	(0.0374)	(0.0334)	(0.0343)	(0.0275)	(0.0326)	(0.0278)	(0.0357)
Portão Isabel	0.433***	0.420***	0.433***	0.376***	0.360***	0.375***	0.418***	0.358***	0.432***	0.375***
	(0.0261)	(0.0291)	(0.0260)	(0.0335)	(0.0370)	(0.0334)	(0.0291)	(0.0370)	(0.0259)	(0.0333)
Iracema farm	-0.115*	-0.115*	-0.116*	-0.111*	-0.111*	-0.111*	-0.122*	-0.119*	-0.189***	-0.191***
	(0.0644)	(0.0641)	(0.0645)	(0.0649)	(0.0646)	(0.0650)	(0.0641)	(0.0648)	(0.0672)	(0.0675)
Observations	2,358	2,364	2,358	2,293	2,299	2,293	2,364	2,299	2,358	2,293
R-squared	0.861	0.854	0.861	0.863	0.855	0.863	0.854	0.855	0.863	0.864
Number of id	512	512	512	504	504	504	512	504	512	504

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 11 – Robustness checks: Alternative estaca-based control specifications – Dependent variable: lincomeharv

	lincomeharv									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(HH size)	-0.100	-0.0996	-0.100	-0.115	-0.115	-0.115	-0.100	-0.117	-0.0933	-0.109
	(0.0847)	(0.0844)	(0.0846)	(0.0831)	(0.0831)	(0.0834)	(0.0833)	(0.0816)	(0.0855)	(0.0822)
ln(# members 7-60y)	0.402***	0.401***	0.402***	0.407***	0.407***	0.407***	0.401***	0.406***	0.393***	0.397***
	(0.0705)	(0.0700)	(0.0702)	(0.0697)	(0.0692)	(0.0694)	(0.0686)	(0.0676)	(0.0723)	(0.0706)
estaca_mean	-2.87e-05	-2.84e-05	-2.74e-05						-7.82e-05*	-0.000232
	(1.99e-05)	(2.03e-05)	(2.08e-05)						(4.21e-05)	(0.000168)
estaca_mdev	5.14e-05		-2.04e-05							
	(0.000120)		(4.11e-05)							
estaca_sd	-5.06e-05		-1.55e-05							7.33e-05
	(0.000101)		(3.01e-05)							(8.41e-05)
estaca_mean_check				-2.34e-05	-2.33e-05	-2.41e-05		-0.000105**		-0.000332**
				(2.16e-05)	(2.11e-05)	(2.24e-05)		(4.69e-05)		(0.000168)
estaca_mdev_check				-5.65e-05		-3.88e-05				
				(0.000207)		(4.46e-05)				
estaca_sd_check				1.29e-05		-2.71e-05				0.000113
				(0.000159)		(3.35e-05)				(8.55e-05)
estaca_mad							0.000125		0.000419	
							(0.000111)		(0.000327)	
estaca_mad_check								0.000196		0.000623*
								(0.000123)		(0.000328)
Grama Larga	0.0735	0.0737	0.0739	0.0635	0.0635	0.0632	0.0617	0.0439	0.0232	-0.0118
	(0.0757)	(0.0758)	(0.0760)	(0.0752)	(0.0753)	(0.0756)	(0.0721)	(0.0723)	(0.0893)	(0.0873)
Lage	0.0239	0.0241	0.0243	0.0233	0.0233	0.0230	0.0151	0.00817	-0.0164	-0.0387
	(0.0749)	(0.0750)	(0.0752)	(0.0770)	(0.0769)	(0.0771)	(0.0732)	(0.0754)	(0.0860)	(0.0883)
Morro Alto	0.0618	0.0621	0.0623	0.0428	0.0428	0.0423	0.0482	0.0197	0.00538	-0.0431
	(0.0877)	(0.0878)	(0.0879)	(0.0911)	(0.0913)	(0.0915)	(0.0856)	(0.0898)	(0.101)	(0.103)
Pão de Ló	0.0501	0.0498	0.0491	0.0417	0.0416	0.0423	0.0535	0.0504	0.0904	0.106
	(0.0699)	(0.0699)	(0.0690)	(0.0718)	(0.0722)	(0.0709)	(0.0736)	(0.0757)	(0.0694)	(0.0684)
Salt. & Pinh.	0.221	0.220	0.220	0.189	0.190	0.191	0.230	0.206	0.256	0.242
	(0.184)	(0.182)	(0.182)	(0.192)	(0.189)	(0.189)	(0.185)	(0.191)	(0.171)	(0.178)
Teteia	-0.0628	-0.0632	-0.0631	-0.0745	-0.0745	-0.0747	-0.0788	-0.0987	-0.118	-0.152
	(0.159)	(0.158)	(0.158)	(0.156)	(0.156)	(0.156)	(0.156)	(0.153)	(0.163)	(0.159)
Col. Ibicaba	0.172	0.172	0.172	0.192	0.193	0.192	0.166	0.182	0.141	0.143
	(0.258)	(0.258)	(0.258)	(0.303)	(0.302)	(0.302)	(0.262)	(0.308)	(0.266)	(0.313)
Linha	–	–	–	–	–	–	–	–	–	–
Mariquita	0.222***	0.222***	0.222***	0.224***	0.224***	0.224***	0.228***	0.233***	0.236***	0.246***
	(0.0412)	(0.0411)	(0.0410)	(0.0406)	(0.0404)	(0.0402)	(0.0409)	(0.0404)	(0.0456)	(0.0452)
Paineira										
Portão Isabel	-0.107**	-0.106*	-0.107*	-0.279***	-0.279***	-0.280***	-0.108**	-0.280***	-0.105*	-0.278***
	(0.0543)	(0.0545)	(0.0544)	(0.0547)	(0.0549)	(0.0548)	(0.0542)	(0.0546)	(0.0542)	(0.0541)
Iracema farm	-0.235**	-0.236**	-0.235**	-0.252**	-0.252**	-0.252**	-0.249**	-0.273***	-0.288***	-0.331***
	(0.0979)	(0.0978)	(0.0979)	(0.102)	(0.102)	(0.102)	(0.0976)	(0.102)	(0.105)	(0.109)
Observations	1,470	1,470	1,470	1,447	1,447	1,447	1,470	1,447	1,470	1,447
R-squared	0.844	0.844	0.844	0.846	0.846	0.846	0.844	0.846	0.845	0.846
Number of id	324	324	324	322	322	322	324	322	324	322

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 12– Robustness checks: Household amenities – Dependent variable: lincomeharv

	lincomeharv							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(HH size)	-0.100	-0.0611	0.0810	-0.125	-0.122	-0.0770	-0.101	-0.129*
	(0.0847)	(0.0662)	(0.108)	(0.0756)	(0.0767)	(0.0890)	(0.0851)	(0.0765)
ln(# members 7-60y)	0.402***	0.405***	0.357***	0.521***	0.525***	0.546***	0.405***	0.532***
	(0.0705)	(0.0585)	(0.0912)	(0.0795)	(0.0794)	(0.104)	(0.0712)	(0.0807)
Mean (estaca)	-2.87e-05	-0.000361	0.000112	3.29e-05	2.81e-05	0.000343	-2.81e-05	7.88e-05
	(1.99e-05)	(0.000283)	(0.000760)	(0.000452)	(0.000440)	(0.000557)	(1.98e-05)	(0.000457)
MDEV (estaca)	5.14e-05	0.000477	0.00821***	0.00399	0.00384	0.00320	7.19e-05	0.00382
	(0.000120)	(0.000303)	(0.00282)	(0.00277)	(0.00277)	(0.00231)	(0.000134)	(0.00275)
SD (estaca)	-5.06e-05	-0.000218**	-0.00614***	-0.00252	-0.00238	-0.00237	-6.66e-05	-0.00243
	(0.000101)	(0.000107)	(0.00184)	(0.00218)	(0.00220)	(0.00183)	(0.000111)	(0.00217)
Grama Larga	0.0735	0.0274	0.235	0.0496	0.0521	0.110	0.0485	-0.0347
	(0.0757)	(0.0883)	(0.165)	(0.122)	(0.117)	(0.136)	(0.0730)	(0.127)
Lage	0.0239	0.00291		0.0197	0.0250	0.166	0.0262	0.0254
	(0.0749)	(0.0908)		(0.0935)	(0.0912)	(0.107)	(0.0743)	(0.0964)
Morro Alto	0.0618	-0.00469		0.155	0.162	0.253*	0.0540	0.167
	(0.0877)	(0.103)		(0.139)	(0.134)	(0.149)	(0.0867)	(0.140)
Pão de Ló	0.0501	0.0927	0.0988	-0.0325	-0.0125	-0.167	0.0562	-0.0135
	(0.0699)	(0.0908)	(0.135)	(0.0983)	(0.1000)	(0.129)	(0.0713)	(0.0983)
Salt. & Pinh.	0.221	0.360*		0.198	0.220	-0.188**	0.223	0.194
	(0.184)	(0.198)		(0.202)	(0.207)	(0.0848)	(0.185)	(0.205)
Teteia	-0.0628	-0.191		-0.317	-0.302	-0.239	-0.0590	-0.237
	(0.159)	(0.172)		(0.228)	(0.225)	(0.247)	(0.152)	(0.228)
Col. Ibicaba	0.172	-0.360**					0.166	
	(0.258)	(0.140)					(0.262)	
Mariquita	0.222***	0.192***	0.0122	0.129*	0.111	0.301***	0.217***	0.113
	(0.0412)	(0.0371)	(0.0350)	(0.0679)	(0.0706)	(0.0767)	(0.0432)	(0.0694)
Paineira								
Portão Isabel	-0.107**	-0.0621					-0.109**	
	(0.0543)	(0.0681)					(0.0546)	
Iracema farm	-0.235**	-0.310***		-0.113	-0.104		-0.278***	-0.214
	(0.0979)	(0.100)		(0.133)	(0.128)		(0.0970)	(0.141)
nhouse		0.00524						-0.0674
		(0.0294)						(0.0564)
room			-0.00623					
			(0.0230)					
window				0.00985				0.0113
				(0.00849)				(0.0109)
door					-0.000284			
					(0.00629)			
locker						0.0193		
						(0.0319)		
light							0.0357	0.104*
							(0.0435)	(0.0600)
Observations	1,470	1,424	443	996	987	673	1,462	996
R-squared	0.844	0.867	0.891	0.876	0.876	0.866	0.844	0.877
Number of id	324	316	96	212	207	149	322	212

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 13 – Robustness checks: Colony of work vs. colony of residence (Panel A – Physical output and labor income)

	lincometree_annual		ltree_tot_check		lharv_tot		lincometree_annual	
	work	residence	work	residence	work	residence	work	residence
ln(HH size)	0.0186	0.0157	0.0278	0.0248	-0.113	-0.110	0.0186	0.0157
	(0.0484)	(0.0485)	(0.0483)	(0.0484)	(0.0735)	(0.0741)	(0.0484)	(0.0485)
ln(# members 7-60y)	0.335***	0.337***	0.332***	0.334***	0.393***	0.392***	0.335***	0.337***
	(0.0397)	(0.0398)	(0.0403)	(0.0405)	(0.0616)	(0.0621)	(0.0397)	(0.0398)
Mean (estaca)	1.92e-06	1.82e-06	2.45e-06	2.35e-06	-3.10e-05*	-3.13e-05*	1.92e-06	1.82e-06
	(2.60e-05)	(2.61e-05)	(2.64e-05)	(2.66e-05)	(1.74e-05)	(1.78e-05)	(2.60e-05)	(2.61e-05)
MDEV (estaca)	-4.53e-05	-5.18e-05	-4.84e-05	-5.39e-05	-8.01e-05	-7.35e-05	-4.53e-05	-5.18e-05
	(8.50e-05)	(8.55e-05)	(8.69e-05)	(8.77e-05)	(0.000107)	(0.000112)	(8.50e-05)	(8.55e-05)
SD (estaca)	3.34e-05	3.68e-05	3.53e-05	3.80e-05	4.95e-05	4.50e-05	3.34e-05	3.68e-05
	(6.03e-05)	(6.07e-05)	(6.17e-05)	(6.23e-05)	(8.32e-05)	(8.73e-05)	(6.03e-05)	(6.07e-05)
Grama Larga	0.143**	0.142**	0.133**	0.132**	0.0706	0.0687	0.143**	0.142**
	(0.0568)	(0.0559)	(0.0565)	(0.0555)	(0.0716)	(0.0714)	(0.0568)	(0.0559)
Lage	0.213***	0.182***	0.215***	0.182***	0.0310	0.0151	0.213***	0.182***
	(0.0712)	(0.0659)	(0.0721)	(0.0676)	(0.0724)	(0.0793)	(0.0712)	(0.0659)
Morro Alto	0.129**	0.128**	0.121*	0.119**	0.0523	0.0471	0.129**	0.128**
	(0.0620)	(0.0593)	(0.0620)	(0.0593)	(0.0843)	(0.0819)	(0.0620)	(0.0593)
Pão de Ló	0.189**	0.189**	0.185**	0.184**	0.0316	0.0313	0.189**	0.189**
	(0.0751)	(0.0748)	(0.0771)	(0.0768)	(0.0759)	(0.0763)	(0.0751)	(0.0748)
Salt. & Pinh.	0.157**	0.158**	0.145*	0.144*	0.223	0.223	0.157**	0.158**
	(0.0790)	(0.0785)	(0.0797)	(0.0792)	(0.159)	(0.159)	(0.0790)	(0.0785)
Teteia	0.200***	0.198***	0.202***	0.200***	-0.0459	-0.0500	0.200***	0.198***
	(0.0644)	(0.0632)	(0.0653)	(0.0642)	(0.160)	(0.160)	(0.0644)	(0.0632)
Col. Ibicaba	-0.269**	-0.177	-0.294**	-0.206	0.151	-0.0609	-0.269**	-0.177
	(0.134)	(0.139)	(0.140)	(0.141)	(0.274)	(0.378)	(0.134)	(0.139)
Linha	–	–	–	–	–	–	–	–
Mariquita	0.277***	0.197***	0.295***	0.204***	0.287***	0.230***	0.277***	0.197***
	(0.0303)	(0.0277)	(0.0292)	(0.0293)	(0.0465)	(0.0441)	(0.0303)	(0.0277)
Paineira	0.0876***	-0.240***	0.128***	-0.304***	–	–	0.0876***	-0.240***
	(0.0282)	(0.0414)	(0.0253)	(0.0379)			(0.0282)	(0.0414)
Portão Isabel	0.433***	0.443***	0.465***	0.479***	-0.136***	-0.134***	0.433***	0.443***
	(0.0261)	(0.0282)	(0.0265)	(0.0321)	(0.0377)	(0.0382)	(0.0261)	(0.0282)
Laborers	–	0.0512		0.0751		0.0191	–	0.0512
		(0.0805)		(0.115)		(0.0469)		(0.0805)
Iracema farm	-0.115*	-0.120*		-0.111*	-0.267***	-0.272***	-0.115*	-0.120*
	(0.0644)	(0.0639)		(0.0646)	(0.0967)	(0.0945)	(0.0644)	(0.0639)
Observations	2,358	2,342	2,360	2,344	1,641	1,632	2,358	2,342
R-squared	0.861	0.860	0.494	0.491	0.816	0.815	0.861	0.860
Number of id	512	508	513	509	335	333	512	508

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted category or category not applicable to the regression. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 14 – Robustness checks: Households moving between colonies (Panel A – Physical output and labor income)

	ltree_tot_ check		lharv_tot		lincometree_ annual		lincomeharv	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
ln(HH size)	0.0127	0.0122	-0.148**	-0.116	0.00235	0.00317	-0.132	-0.105
	(0.0536)	(0.0529)	(0.0719)	(0.0755)	(0.0534)	(0.0528)	(0.0827)	(0.0855)
ln(# members 7-60y)	0.355***	0.354***	0.410***	0.397***	0.357***	0.356***	0.424***	0.414***
	(0.0439)	(0.0437)	(0.0615)	(0.0629)	(0.0434)	(0.0432)	(0.0701)	(0.0712)
Mean (estaca)	8.18e-06	8.24e-06	-3.92e-05**	-3.45e-05*	7.16e-06	7.49e-06	-3.61e-05*	-2.94e-05
	(2.62e-05)	(2.61e-05)	(1.74e-05)	(1.83e-05)	(2.55e-05)	(2.55e-05)	(1.88e-05)	(2.04e-05)
MDEV (estaca)	-4.12e-05	-5.93e-05	-2.25e-05	-5.51e-05	-3.79e-05	-5.76e-05	9.55e-05	9.20e-05
	(8.56e-05)	(8.19e-05)	(9.81e-05)	(0.000109)	(8.40e-05)	(7.95e-05)	(0.000123)	(0.000136)
SD (estaca)	2.50e-05	3.79e-05	1.82e-05	3.32e-05	2.35e-05	3.72e-05	-7.19e-05	-8.09e-05
	(6.13e-05)	(5.91e-05)	(8.03e-05)	(8.60e-05)	(6.02e-05)	(5.74e-05)	(0.000104)	(0.000113)
Grama Larga	0.117**	0.116*	0.0783	0.0795	0.129**	0.129**	0.0870	0.0841
	(0.0592)	(0.0595)	(0.0728)	(0.0735)	(0.0595)	(0.0598)	(0.0781)	(0.0776)
Lage	0.216**	0.214**	0.0528	0.0623	0.215**	0.214**	0.0452	0.0500
	(0.0835)	(0.0858)	(0.0705)	(0.0700)	(0.0831)	(0.0854)	(0.0730)	(0.0725)
Morro Alto	0.114*	0.114*	0.0527	0.0537	0.124*	0.125*	0.0648	0.0639
	(0.0678)	(0.0680)	(0.0845)	(0.0841)	(0.0681)	(0.0683)	(0.0883)	(0.0878)
Pão de Ló	0.155*	0.154*	0.0726	0.0688	0.157*	0.156*	0.0844	0.0782
	(0.0829)	(0.0828)	(0.0823)	(0.0857)	(0.0800)	(0.0802)	(0.0795)	(0.0788)
Salt. & Pinh.	0.0947	0.0968	0.243	0.227	0.107	0.108	0.189	0.190
	(0.0868)	(0.0863)	(0.173)	(0.168)	(0.0846)	(0.0842)	(0.196)	(0.191)
Teteia	0.155***	0.157***	-0.0448	-0.0582	0.153***	0.154***	-0.0660	-0.0777
	(0.0599)	(0.0596)	(0.163)	(0.164)	(0.0587)	(0.0585)	(0.162)	(0.161)
Col. Ibicaba	-0.333**	-0.335**	0.217	0.238	-0.312**	-0.312**	0.225	0.242
	(0.162)	(0.164)	(0.265)	(0.254)	(0.150)	(0.151)	(0.251)	(0.241)
Linha	–	–	–	–	–	–	–	–
Mariquita	0.303***	0.303***	0.298***	0.270***	0.287***	0.288***	0.226***	0.198***
	(0.0281)	(0.0278)	(0.0469)	(0.0482)	(0.0280)	(0.0277)	(0.0415)	(0.0441)
Paineira	0.284***				0.205***			
	(0.0330)				(0.0383)			
Portão Isabel	0.459***	0.459***	-0.131***	-0.163***	0.430***	0.429***	-0.0900	-0.109**
	(0.0264)	(0.0268)	(0.0375)	(0.0398)	(0.0257)	(0.0264)	(0.0560)	(0.0550)
Iracema farm	-0.164**	-0.168**	-0.167	-0.154	-0.171**	-0.171**	-0.140	-0.139
	(0.0766)	(0.0796)	(0.104)	(0.106)	(0.0749)	(0.0780)	(0.107)	(0.107)
dcolony_ind	-0.00282		-0.0800**		-0.00197		-0.0846**	
	(0.0206)		(0.0372)		(0.0205)		(0.0397)	
movedout_ind		5.50e-05		-0.103***		-0.00211		-0.0900**
		(0.0234)		(0.0379)		(0.0232)		(0.0390)
Observations	1,972	1,961	1,617	1,608	1,971	1,960	1,450	1,442
R-squared	0.533	0.536	0.818	0.818	0.876	0.876	0.844	0.845
Number of id	380	377	328	325	380	377	317	313

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) Except for dcolony_ind and movedout_ind, all other indicators for households' moving out or between colonies have been omitted in the estimations. (5) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 15 – Robustness checks: Household amenities – Dependent variable: lincometree_annual

	lincometree_annual							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(HH size)	0.0186	0.0458	0.201*	0.138***	0.140***	0.202***	0.0171	0.135***
	(0.0484)	(0.0473)	(0.103)	(0.0520)	(0.0540)	(0.0676)	(0.0484)	(0.0509)
ln(# members 7-60y)	0.335***	0.317***	0.261***	0.337***	0.341***	0.319***	0.335***	0.330***
	(0.0397)	(0.0402)	(0.0602)	(0.0505)	(0.0495)	(0.0799)	(0.0399)	(0.0502)
Mean (estaca)	1.92e-06	-0.000386**	-0.000763*	-0.000544	-0.000540	-0.000172	1.81e-06	-0.000563*
	(2.60e-05)	(0.000194)	(0.000437)	(0.000342)	(0.000334)	(0.000549)	(2.73e-05)	(0.000337)
MDEV (estaca)	-4.53e-05	0.000342	0.00265***	0.00185**	0.00181**	0.00401**	-5.76e-05	0.00190**
	(8.50e-05)	(0.000226)	(0.000550)	(0.000930)	(0.000908)	(0.00161)	(8.50e-05)	(0.000952)
SD (estaca)	3.34e-05	-6.66e-05	-0.00115***	-0.000759	-0.000741	-0.00210**	4.17e-05	-0.000783
	(6.03e-05)	(9.00e-05)	(0.000304)	(0.000512)	(0.000500)	(0.000971)	(6.01e-05)	(0.000522)
Grama Larga	0.143**	0.0490	0.0983	0.00132	0.00752	0.0398	0.171***	0.0159
	(0.0568)	(0.0727)	(0.123)	(0.103)	(0.100)	(0.134)	(0.0580)	(0.101)
Lage	0.213***	0.121		0.0798	0.0860	0.157	0.216***	0.0807
	(0.0712)	(0.0914)		(0.110)	(0.109)	(0.148)	(0.0715)	(0.107)
Morro Alto	0.129**	0.0278		0.0579	0.0623	0.115	0.137**	0.0467
	(0.0620)	(0.0825)		(0.115)	(0.114)	(0.144)	(0.0613)	(0.114)
Pão de Ló	0.189**	0.251***	0.235**	0.286***	0.302***	0.117	0.186**	0.279***
	(0.0751)	(0.0773)	(0.0943)	(0.0967)	(0.0973)	(0.118)	(0.0749)	(0.0989)
Salt. & Pinh.	0.157**	0.156*		0.142	0.155	0.0828	0.161**	0.161
	(0.0790)	(0.0868)		(0.134)	(0.128)	(0.0686)	(0.0789)	(0.130)
Teteia	0.200***	0.0699		0.144	0.162	0.143	0.200***	0.115
	(0.0644)	(0.0948)		(0.162)	(0.158)	(0.228)	(0.0639)	(0.165)
Col. Ibicaba	-0.269**	-0.445**					-0.259**	
	(0.134)	(0.225)					(0.128)	
Mariquita	0.277***	0.291***	0.239***	0.202***	0.191***	0.279***	0.283***	0.199***
	(0.0303)	(0.0306)	(0.0237)	(0.0280)	(0.0268)	(0.0578)	(0.0311)	(0.0306)
Paineira	0.0876***							
	(0.0282)							
Portão Isabel	0.433***	0.354***					0.433***	
	(0.0261)	(0.0316)					(0.0264)	
Iracema farm	-0.115*	-0.217***		-0.219**	-0.216**		-0.101	-0.220**
	(0.0644)	(0.0790)		(0.104)	(0.103)		(0.0656)	(0.103)
nhouse		0.0646***						0.0912**
		(0.0199)						(0.0375)
room			0.0258*					
			(0.0133)					
window				0.00914				0.00284
				(0.00735)				(0.00895)
door					-0.000673			
					(0.00732)			
locker						0.110***		
						(0.0374)		
light							-0.0473	-0.0198
							(0.0392)	(0.0337)
Observations	2,358	2,265	766	1,536	1,523	948	2,342	1,536
R-squared	0.861	0.865	0.904	0.842	0.842	0.803	0.861	0.844
Number of id	512	494	152	336	332	227	508	336

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 16 – Robustness checks: Alternative timing controls for harvest season – Dependent variable: lincomeharv

	lincomeharv					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(#months harv.)	4.447***	4.447***	2.502***	2.500***	4.174***	4.174***
	(0.634)	(0.634)	(0.161)	(0.160)	(0.633)	(0.633)
Mean month harv.	-1.669***		0.0792		-1.301***	
	(0.642)		(0.0826)		(0.358)	
1st month harv.	1.740***	1.323***			1.574***	1.249***
	(0.481)	(0.323)			(0.460)	(0.381)
Last month harv.	-0.0646	-0.482***			-0.245	-0.570**
	(0.174)	(0.127)			(0.231)	(0.277)
Median month harv.		-0.834***		0.0774		-0.651***
		(0.321)		(0.0821)		(0.179)
SD month harv.			-1.175***	-1.175***	0.374	0.374
			(0.136)	(0.136)	(0.512)	(0.512)
ln(HH size)	-0.100	-0.100	-0.0733	-0.0733	-0.0775	-0.0775
	(0.0847)	(0.0847)	(0.0700)	(0.0700)	(0.0678)	(0.0678)
ln(# members 7-60y)	0.402***	0.402***	0.427***	0.427***	0.424***	0.424***
	(0.0705)	(0.0705)	(0.0575)	(0.0575)	(0.0574)	(0.0574)
Mean (estaca)	-2.87e-05	-2.87e-05	-2.05e-05	-2.05e-05	-2.09e-05	-2.09e-05
	(1.99e-05)	(1.99e-05)	(1.58e-05)	(1.58e-05)	(1.48e-05)	(1.48e-05)
MDEV (estaca)	5.14e-05	5.14e-05	6.28e-05	6.29e-05	8.67e-05	8.67e-05
	(0.000120)	(0.000120)	(9.93e-05)	(9.93e-05)	(9.45e-05)	(9.45e-05)
SD (estaca)	-5.06e-05	-5.06e-05	-6.77e-05	-6.78e-05	-8.31e-05	-8.31e-05
	(0.000101)	(0.000101)	(7.76e-05)	(7.76e-05)	(7.59e-05)	(7.59e-05)
Observations	1,470	1,470	1,455	1,455	1,455	1,455
R-squared	0.844	0.844	0.840	0.840	0.843	0.843
Number of id	324	324	320	320	320	320

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 17 – Robustness checks: Determinants of labor income from coffee cultivation – Comparison with check sample

	lincometree_ annual (1)	lincometree_ annual_ check (2)
ln(price cult.)	1.286***	1.286***
	(0.0838)	(0.0838)
ln(N pruning)	1.138***	1.138***
	(0.252)	(0.251)
ln(HH size)	0.0186	0.0179
	(0.0484)	(0.0488)
ln(# members 7-60y)	0.335***	0.338***
	(0.0397)	(0.0403)
Mean (estaca)	1.92e-06	2.18e-06
	(2.60e-05)	(2.60e-05)
MDEV (estaca)	-4.53e-05	-4.30e-05
	(8.50e-05)	(8.74e-05)
SD (estaca)	3.34e-05	3.09e-05
	(6.03e-05)	(6.23e-05)
Observations	2,358	2,354
R-squared	0.861	0.861
Number of id	512	512

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 18 – Robustness checks: Alternative price indices for coffee cultivation (Panel A – Physical output and labor income)

	ltree_tot_check			lincometree_annual		
	(2)	(6)	(10)	(1)	(5)	(9)
Indep var for prices				wcult_tree	cult_mean	cult_med
ln(price cult.)				1.286***	0.987***	0.975***
				(0.0838)	(0.0824)	(0.0816)
ln(N pruning)				1.138***	1.168***	1.176***
				(0.252)	(0.222)	(0.222)
ln(HH size)	0.0278	0.0278	0.0278	0.0186	0.0269	0.0272
	(0.0483)	(0.0483)	(0.0483)	(0.0484)	(0.0494)	(0.0493)
ln(# members 7-60y)	0.332***	0.332***	0.332***	0.335***	0.339***	0.339***
	(0.0403)	(0.0403)	(0.0403)	(0.0397)	(0.0416)	(0.0415)
Mean (estaca)	2.45e-06	2.45e-06	2.45e-06	1.92e-06	4.43e-06	4.73e-06
	(2.64e-05)	(2.64e-05)	(2.64e-05)	(2.60e-05)	(2.46e-05)	(2.44e-05)
MDEV (estaca)	-4.84e-05	-4.84e-05	-4.84e-05	-4.53e-05	-4.78e-05	-4.81e-05
	(8.69e-05)	(8.69e-05)	(8.69e-05)	(8.50e-05)	(8.82e-05)	(8.82e-05)
SD (estaca)	3.53e-05	3.53e-05	3.53e-05	3.34e-05	3.27e-05	3.28e-05
	(6.17e-05)	(6.17e-05)	(6.17e-05)	(6.03e-05)	(6.30e-05)	(6.30e-05)
Observations	2,360	2,360	2,360	2,358	2,358	2,358
R-squared	0.494	0.494	0.494	0.861	0.849	0.849
Number of id	513	513	513	512	512	512

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 19 – Robustness checks: Accounting errors and sample restrictions – Dependent variable: lincometree_annual

	lincometree_annual				
	(1)	(2)	(3)	(4)	(5)
ln(price cult.)	1.287***	1.287***	1.162***	1.239***	1.287***
	(0.0885)	(0.0885)	(0.0883)	(0.112)	(0.0884)
ln(N pruning)	1.147***	1.147***	0.594***	1.750***	1.139***
	(0.251)	(0.251)	(0.138)	(0.278)	(0.250)
ln(HH size)	0.0178	0.0178	-0.0399	0.177***	0.0149
	(0.0496)	(0.0496)	(0.0577)	(0.0665)	(0.0496)
ln(# members 7-60y)	0.338***	0.338***	0.318***	0.289***	0.340***
	(0.0403)	(0.0403)	(0.0425)	(0.0807)	(0.0404)
Mean (estaca)	2.57e-06	2.57e-06	2.87e-05	-0.000392	3.14e-06
	(2.57e-05)	(2.57e-05)	(2.00e-05)	(0.000371)	(2.59e-05)
MDEV (estaca)	-4.08e-05	-4.08e-05	-3.38e-05	0.00122**	-3.80e-05
	(8.83e-05)	(8.83e-05)	(7.42e-05)	(0.000533)	(8.87e-05)
SD (estaca)	2.97e-05	2.97e-05	6.51e-06	-0.000600**	2.73e-05
	(6.28e-05)	(6.28e-05)	(5.17e-05)	(0.000273)	(6.31e-05)
error_account	-0.00763*	-0.00763*			
	(0.00401)	(0.00401)			
Grama Larga	0.143**	0.143**	0.186***	0.0230	0.143**
	(0.0569)	(0.0569)	(0.0639)	(0.0979)	(0.0568)
Lage	0.235***	0.235***	0.300***	0.0818	0.235***
	(0.0744)	(0.0744)	(0.0920)	(0.120)	(0.0742)
Morro Alto	0.127**	0.127**	0.142**	0.106	0.127**
	(0.0607)	(0.0607)	(0.0679)	(0.122)	(0.0609)
Pão de Ló	0.191**	0.191**	0.150*	0.287***	0.189**
	(0.0747)	(0.0747)	(0.0881)	(0.103)	(0.0747)
Salt. & Pinh.	0.159**	0.159**	0.302**	0.0165	0.159**
	(0.0782)	(0.0782)	(0.127)	(0.158)	(0.0791)
Teteia	0.216***	0.216***	0.150**	0.202	0.216***
	(0.0680)	(0.0680)	(0.0745)	(0.170)	(0.0680)
Col. Ibicaba	-0.264*	-0.264*	-0.0919		-0.263*
	(0.135)	(0.135)	(0.101)		(0.136)
Mariquita	0.283***	0.283***	0.245***	0.459***	0.282***
	(0.0305)	(0.0305)	(0.0886)	(0.0287)	(0.0305)
Paineira	-0.133***	-0.133***	-0.210***		-0.136***
	(0.0375)	(0.0375)	(0.0390)		(0.0375)
Portão Isabel	0.429***	0.429***	0.432***		0.431***
	(0.0262)	(0.0262)	(0.0275)		(0.0263)
Iracema farm	-0.0877	-0.0877	-0.0363	-0.189*	-0.0872
	(0.0622)	(0.0622)	(0.110)	(0.113)	(0.0622)
Accounting errors	Yes	< 120 mil-réis	No	Yes	< 120 mil-réis
Observations	2,339	2,338	1,457	901	2,338
R-squared	0.862	0.862	0.860	0.837	0.862
Number of id	511	510	416	298	510

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) Column (1) adds "error_account" variable to the baseline specification, without any changes in the sample. (4) "< 120 mil-réis" refers to the sample dropping the accounting error outlier of 120.649 réis. (5) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 20 – Robustness checks: Excluding observations by source / availability of household information

	lincometree_annual						lincomeharv					
	Demogr.	Coffee groves	Housing	All	Known HH	Unknown HH	Demogr.	Coffee groves	Housing	All	Known HH	Unknown HH
ln(HH size)	0.0390	0.0174	0.0207	0.0425	0.100**	0.0176	-0.0924	-0.0996	-0.0996	-0.0924	-0.0202	-0.0996
	(0.0517)	(0.0485)	(0.0497)	(0.0531)	(0.0494)	(0.0486)	(0.0882)	(0.0846)	(0.0846)	(0.0882)	(0.0804)	(0.0846)
ln(# members 7-60y)	0.348***	0.336***	0.331***	0.344***	0.335***	0.336***	0.405***	0.401***	0.401***	0.405***	0.384***	0.401***
	(0.0408)	(0.0398)	(0.0409)	(0.0420)	(0.0417)	(0.0399)	(0.0705)	(0.0704)	(0.0704)	(0.0705)	(0.0717)	(0.0704)
Mean (estaca)	3.12e-06	3.13e-06	5.09e-06	5.16e-06	4.04e-06	2.99e-06	-2.61e-05	-2.91e-05	-2.91e-05	-2.61e-05	-2.89e-05	-2.91e-05
	(2.54e-05)	(2.58e-05)	(2.59e-05)	(2.55e-05)	(2.51e-05)	(2.58e-05)	(1.90e-05)	(2.00e-05)	(2.00e-05)	(1.90e-05)	(1.85e-05)	(2.00e-05)
MDEV (estaca)	-3.07e-05	-4.35e-05	-5.49e-05	-4.23e-05	8.44e-05	-4.62e-05	6.52e-05	5.21e-05	5.21e-05	6.52e-05	7.76e-05	5.21e-05
	(8.51e-05)	(8.53e-05)	(8.62e-05)	(8.58e-05)	(8.16e-05)	(8.68e-05)	(0.000113)	(0.000120)	(0.000120)	(0.000113)	(0.000113)	(0.000120)
SD (estaca)	2.17e-05	3.12e-05	3.61e-05	2.66e-05	-6.27e-05	3.34e-05	-6.56e-05	-5.08e-05	-5.08e-05	-6.56e-05	-7.01e-05	-5.08e-05
	(6.08e-05)	(6.06e-05)	(6.15e-05)	(6.17e-05)	(5.86e-05)	(6.18e-05)	(9.33e-05)	(0.000101)	(0.000101)	(9.33e-05)	(9.30e-05)	(0.000101)
Observations	2,290	2,358	2,314	2,246	2,286	2,352	1,427	1,470	1,470	1,427	1,426	1,469
R-squared	0.860	0.862	0.861	0.859	0.864	0.862	0.841	0.844	0.844	0.841	0.843	0.844
Number of id	508	512	511	507	510	507	320	324	324	320	321	323

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 21 – Robustness checks: Controls for data sources and PDFs (Panel A – Physical output and labor income)

	lincometree_annual		lincomeharv	
	(1)	(2)	(1)	(2)
ln(HH size)	0.0151	0.0160	-0.0945	-0.0970
	(0.0486)	(0.0488)	(0.0840)	(0.0849)
ln(# members 7-60y)	0.337***	0.338***	0.400***	0.399***
	(0.0395)	(0.0394)	(0.0689)	(0.0698)
Mean (estaca)	4.41e-06	5.16e-06	-2.91e-05	-2.71e-05
	(2.57e-05)	(2.56e-05)	(2.01e-05)	(1.96e-05)
MDEV (estaca)	-5.45e-05	-5.95e-05	7.44e-05	6.17e-05
	(8.51e-05)	(8.54e-05)	(0.000135)	(0.000137)
SD (estaca)	3.73e-05	4.01e-05	-6.71e-05	-5.79e-05
	(6.05e-05)	(6.08e-05)	(0.000112)	(0.000113)
source01	0.00280	0.0121	0.0122*	-0.0380
	(0.00471)	(0.0226)	(0.00666)	(0.0375)
source02	0.00671	-0.00898	-0.00196	-0.116*
	(0.00533)	(0.0234)	(0.00764)	(0.0644)
source03	-0.00542	-0.00458	-0.0186***	0.129*
	(0.00418)	(0.00583)	(0.00618)	(0.0698)
source04	–	–		
pdf01		0.00139		-0.00744
		(0.00314)		(0.00540)
pdf02		-0.00225		-0.0159*
		(0.00328)		(0.00909)
pdf03		0.000118		0.0212**
		(0.000548)		(0.0102)
pdf04_multiple				0.0859
				(0.0919)
Observations	2,342	2,342	1,462	1,462
R-squared	0.862	0.862	0.846	0.846
Number of id	508	508	322	322

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) Specifications for the lean season (physical output and income) do not control for "source04", as this refers to the harvest season only; however, specifications for the harvest season control for "source02" (sources for info on coffee groves) because the agricultural capital in the lean season is controlled for. (5) "source04" is omitted due to collinearity in the specifications referring to the harvest season. (6) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.

Table Appendix 22 – Robustness checks: Excluding observations with nullified records, crossed-out income, or quit (Panel A – Physical output and labor income)

	lincometree_annual			lincomeharv		
	Nullified	Crossedout	Quit	Nullified	Crossedout	Quit
ln(HH size)	0.0225	0.0222	0.0174	-0.108	-0.100	-0.0996
	(0.0489)	(0.0488)	(0.0485)	(0.0856)	(0.0851)	(0.0846)
ln(# members 7-60y)	0.335***	0.334***	0.336***	0.408***	0.408***	0.401***
	(0.0400)	(0.0405)	(0.0398)	(0.0692)	(0.0708)	(0.0704)
Mean (estaca)	1.67e-06	4.32e-06	3.13e-06	-2.75e-05	-3.01e-05	-2.91e-05
	(2.61e-05)	(2.58e-05)	(2.58e-05)	(2.02e-05)	(1.99e-05)	(2.00e-05)
MDEV (estaca)	-4.43e-05	-4.54e-05	-4.35e-05	5.88e-05	5.19e-05	5.21e-05
	(8.86e-05)	(8.82e-05)	(8.53e-05)	(0.000119)	(0.000120)	(0.000120)
SD (estaca)	3.32e-05	3.04e-05	3.12e-05	-5.99e-05	-4.95e-05	-5.08e-05
	(6.35e-05)	(6.28e-05)	(6.06e-05)	(0.000101)	(0.000101)	(0.000101)
Observations	2,330	2,345	2,358	1,458	1,465	1,470
R-squared	0.861	0.863	0.862	0.848	0.845	0.844
Number of id	507	511	512	320	324	324

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1. (3) "–" denotes omitted base category. (4) All specifications follow the baseline regressions, except for changes reported in this table. Please refer to the main text for the full set of controls.