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 the 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 the Ibicaba plantation (1888-1958)

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

Abstract:

The beginning of mass migration to the coffee plantations of Southeastern Brazil, particularly to the province of São Paulo, was concomitant with the intensification of the abolitionist movement and of the slave revolts that led to Abolition in 1888. Despite the centrality of the coffee economy to the socioeconomic prospects of immigrants and freedmen, still little is known about the functioning of agricultural labor markets in the immediate post-Abolition era and in subsequent decades. This paper sheds new light on the economic performance of non-whites, European immigrants, and white Brazilians in the coffee economy. It draws on microdata from the *Ibicaba Collection* – a newly created repository dedicated to preserving and digitizing the primary sources of the homonymous plantation, at once the first to experiment with non-captive European laborers and a major slaveholder until 1888. Contrary to the classical hypothesis that Europeans replaced freedmen and colored Brazilians in *colonato* contracts – the labor arrangement considered the most advantageous to laborers –, the evidence compiled in this paper shows that non-whites were the second to third most common ethnolinguistic group working as *colonos*, even if Italians and their descendants far outnumbered all others. Results further show that, among *colonos*, plantation managers did not discriminate in remuneration against non-whites or in favor of Europeans for the execution of unskilled agricultural tasks. Nonetheless, non-whites still showed the lowest annual earnings profiles of any ethnolinguistic group. The number of economically active members in a household was the most consistent determinant of *colonos*' earnings and of the earnings differentials between non-whites, on the one hand, and Europeans and their descendants, on the other. This paper thus concludes that a major burden of slavery for non-whites was the size of their households in the post-Abolition period. European immigrants and their descendants, in turn, showed distinct aggregate earnings profiles. Whereas the Portuguese and German-speakers closely tracked white Brazilians – a result attributable to these groups' longer migratory history to the region –, Italians and their descendants split between higher and lower performers, suggesting the existence of a subgroup that failed to achieve its "(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 the Ibicaba plantation (1888-1958)

Author

Bruno Gabriel Witzel de Souza

Assistant Professor in Economic History at the Faculty of Economics and Business Administration - University of São Paulo (FEA-USP)

E-mail

bwitzelsouza@usp.br

Address

Av. Professor Luciano Gualberto, 908
Cidade Universitária, São Paulo - SP, 05508-210, Brazil
FEA-2, Room 249

Abstract - Extended

The beginning of mass migration to the coffee plantations of Southeastern Brazil, particularly to the province of São Paulo, was concomitant with the intensification of the abolitionist movement and of the slave revolts that led to Abolition in 1888. Despite the centrality of the coffee economy to the socioeconomic prospects of immigrants and freedmen, still little is known about the functioning of agricultural labor markets in the immediate post-Abolition era and in subsequent decades. This paper sheds new light on the economic performance of non-whites, European immigrants, and white Brazilians in the coffee economy. It draws on microdata from the *Ibicaba Collection* – a newly created repository dedicated to preserving and digitizing the primary sources of the homonymous plantation, at once the first to experiment with non-captive European laborers and a major slaveholder until 1888. Contrary to the classical hypothesis that Europeans replaced freedmen and colored Brazilians in *colonato* contracts – the labor arrangement considered the most advantageous to laborers –, the evidence compiled in this paper shows that non-whites were the second to third most common ethnolinguistic group working as *colonos*, even if Italians and their descendants far outnumbered all others. Results further show that, among *colonos*, plantation managers did not discriminate in remuneration against non-whites or in favor of Europeans for the execution of unskilled agricultural tasks. Nonetheless, non-whites still showed the lowest annual earnings profiles of any ethnolinguistic group. The number of economically active members in a household was the most consistent determinant of *colonos*' earnings and of the earnings differentials between non-whites, on the one hand, and Europeans and their descendants, on the other. This paper thus concludes that a major burden of slavery for non-whites was the size of their households in the post-Abolition period. European immigrants and their descendants, in turn, showed distinct aggregate earnings profiles. Whereas the Portuguese and German-speakers closely tracked white Brazilians – a result attributable to these groups' longer migratory history to the region –, Italians and their descendants split between higher and lower performers, suggesting the existence of a subgroup that failed to achieve its "(South-)American dreams".

Abstract - Short

Brazil's post-Abolition agricultural labor markets remain poorly understood for lack of microdata. Benefiting from the newly digitized *Ibicaba Collection*, this paper compares earnings among non-whites, European immigrants, and white Brazilians on a plantation that pioneered non-captive European labor while remaining a major slaveholder. Contrary to the replacement hypothesis, non-whites remained among the most common *colonos*, and managers did not discriminate in pay for unskilled agricultural tasks. Earnings hinged on the number of economically active household members – a burden of slavery leaving non-whites with the lowest earnings. Among Europeans, earnings profiles differed by migratory history.

Acknowledgments

The current version of this working paper is based on five preliminary drafts (May 2023, August 2023, July 2024, May 2025, and May 2026). This version supersedes the structure, findings, and discussions of all previous versions, which should be cited appropriately. Comments on earlier versions, presented at the following academic events, are gratefully acknowledged: UC Davis' *Conference on resource-use & specialization in rural change in Europe and the Americas* (Davis, USA – May 26, 2023); UC Davis' *History Colloquium and Latin American History Workshop* (Davis, USA – April 26, 2023); UCLA' *Seminar in Latin American History, Advanced Topics' Seminar* (William Summerhill) (Los Angeles, USA – May 16, 2023); UCLA's *Library Council Internal Meeting* (Los Angeles, USA - online – May 23, 2023); UCLA's *Proseminar in Economic History* (Los Angeles, USA – May 30, 2023); *Economic History Workshop*, FEA-USP (São Paulo, Brazil – August 04, 2023); UCLA-MEAP's *Online Workshop: MEAP in South America: Digital Archives and Contemporary History* (September 12, 2023); *Agricliometrics V* Conference (Montevideo, Uruguay – December 7-8 2023); Internal seminars at the Chair of Economic History, University of Tuebingen (Tübingen, Germany – November 2023 - August 2024); *Political Economy and Economic History Seminar*, EESP/FGV (São Paulo, Brazil – September 12, 2024); Seminar Series of the Economics Department at *Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto* (FEARP-USP) (Ribeirão Preto, Brazil – September 26, 2024); Minicourse at the *Brazilian Association of Researchers in Economic History*'s Graduate Workshop (Ouro Preto, Brazil – October 1-2, 2024); Federal University of São Carlos' Graduate Workshop (Sorocaba, Brazil - online – November 22, 2024); *Institute for Economic and Social History*' Graduate Colloquium, University of Göttingen (Göttingen, Germany – April 23, 2025); Outreach Program *Nacht des Wissens*, University of Göttingen (Göttingen, Germany – June 21, 2025); *Hermes & Clio Seminar Series* (São Paulo, Brazil – October 01, 2025). I thank all team members of the *Ibicaba Project* and of MEAP-UCLA, as well as José Theodoro Hayden Carvalhaes, Theodoro Hayden Carvalhaes, and Paulo Masuti Levy for their invaluable support in creating the physical archives on which this paper draws. I am indebted to Jörg Baten, Renato Colistete, Sarah Ferber, Pedro Américo Ferreira, Leonardo Gardenal, Michela Giorcelli, Guilherme Grandi, André Lanza, Jann Logemann, Marcos Nakaguma, Sabrina Sigel, Manuel Santos Silva, William Summerhill, and Thales Zamberlan for their comments on previous versions of this paper. All errors and omissions are solely my own.

Disclaimer

Most of this paper was written under the 2022-2023 DAAD-PRIME Fellowship, while the author was a Visiting Scholar at the History Department, University of California, Los Angeles (UCLA), and Research Associate at the Chair of Economic History, University of Tübingen. Funding for the creation of the *Ibicaba Collection*, on which this paper is based, is from Arcadia Fund – MEAP's Project Grant No. MEAP-0104. Funding for the creation of the *Accounting for the Countryside Collection* is from Arcadia Fund – MEAP's Regional Grant No. MEAP-6154.

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

Bruno Gabriel Witzel de Souza

I. Introduction

The persistently high levels of inequality in Brazil and their blatant racial dimensions have made it a commonplace to attribute the socioeconomic disadvantages of the Non-White population to the legacies of slavery. Simultaneously, the above-average income levels of Brazilians of Asian and European descent correlate strongly, on average, with particular ancestries (Monasterio 2017; Monasterio and Lopes 2018). Historical explanations for such diverging performances include immigrants' educational levels (Carvalho Filho and Colistete 2010; Carvalho Filho and Monasterio 2012; Rocha, Ferraz, and Soares 2017; Witzel de Souza 2018); upward occupational mobility based on their on-the-job skills and employment in industries founded by compatriots (*cf.* Andrews 1988, pp. 503-6); and, more controversially, immigrants' access to landownership (Holloway 1980, Chap. 6; Costa 2015; Lanza 2021).

Another obvious mechanism relate to the relative success of these groups as agricultural laborers, as the coffee economy of Southeastern Brazil remained the main pillar sustaining the institution of slavery until 1888 and was, at the same time, the very reason most Europeans and Asians immigrated to the region in the first place. Yet, with important exceptions (Bassanezi 1973, 2019; Monsma 2010; Bucifferro 2017, 2021; Cavalcanti Rodrigues 2021; Colistete 2024; Monti and Oliveira 2025),[1] quantitative research on labor markets in post-Abolition Brazil remains scant.[2] This paper contributes to that debate by evaluating the economic performance of the various ethnolinguistic groups who worked on one of the leading plantations of nineteenth-century Brazil, the *Ibicaba* plantation, from the year of Abolition to the 1950s, the period of final decline of the coffee economy in the region. The demographic characteristics and earnings profiles of households headed by Non-Whites, White Brazilians, and European immigrants and their descendants are used to test classical hypotheses on the functioning of agricultural labor markets and on discrimination in them.[3]

Such a gap in the literature is largely attributable to the scarcity of microdata for nineteenth- and early twentieth-century Brazil, which, in turn, results from the purposeful destruction of sources prior to 1888 (Slenes 1983; Monsma 2010, p. 511; Colistete 2024); and constraints

[1] For *Ibicaba* itself, *cf.* also Kuntz Busch (1967) and Witter (1971; 1982). Berrance de Castro and Scarano (1971) remains a valuable early quantitative comparison of *causa mortis* between immigrants and Non-Whites.
[2] This does not diminish the depth of the historiography on slavery in Brazil. My work dialogues directly with the classical literature on the economics of slavery put in place by Carvalho de Mello (1978), Graham (1981), Carvalho de Mello and Slenes (1980), Eisenberg (1983), Andrews (1988), and Versiani (1994).
[3] In the same direction as Theodoridis, Rönbäck, and Galli's (2025) study for St. Croix Island, this paper provides another point of comparison to the classic cliometric literature on the Postbellum U.S., and to recent quantitative research on the determinants of wealth- and earning-gaps in post-Abolition U.S. (Canaday and Jeremsky 2012; Olmstead and Rhode 2018; Collins, Holtkamp, and Wanamaker 2024; Derenoncourt et al. 2024; Karger and Wray 2024; Collins and Holtkamp 2026).

sources prior to 1888 (Slenes 1983; Monsma 2010, p. 511; Colistete 2024); and constraints for the preservation of public and private archives in low- and middle-income countries (Berrance de Castro 1965; Reber 1985; Palma 2019; Witzel de Souza and Santin Gardenal 2026). Accordingly, this paper was only made possible by the creation of a new *in situ* archive at *Ibicaba* and by the digitalization of its contents. Under the *Modern Endangered Archives Program* (MEAP-UCLA), the *Ibicaba Project* has safeguarded all currently-known private historical records related to that plantation, including its managerial and labor accounting ledgers from 1847 to 1977 (with gaps) (Witzel de Souza and Santin Gardenal 2026). This paper's sources and methods thus add to the resurgence of interest in having accounting records as a source of microdata in economic history (Rosenthal 2018; Palma 2019; Rhode 2020; Jones and Gibbon 2025; Palma and Lambais 2026); including for Brazil, in general (Saraiva and Freire 2016; Fornereto Mariano *et al.* 2022; Monti and Oliveira 2025);[4] and with data from the *Ibicaba Collection* itself (Rodrigues and McBride 2026; Silva Santos 2026; Witzel de Souza and Santin Gardenal 2026). In addition, the methods developed in this paper to transcribe, clean, and cross-validate sixteen accounting ledgers from *Ibicaba* aim at providing a benchmark for similar research with other plantation records, in Brazil and elsewhere (*cf.* the concluding remarks of this paper).

Ibicaba ranks among the most important plantations of Brazilian history. Beginning in the 1840s, it captained institutional experiments with European indentured labor that had major consequences for positioning São Paulo as a leading destination for European immigrants during the Age of Mass Migration *ca.* forty years later. At the same time, *Ibicaba*'s first proprietary firm, *Vergueiro & Co.*, held the largest number of enslaved Africans and African-Brazilians in Limeira, the county where *Ibicaba* was located. *Vergueiro & Co.* went bankrupt soon after Abolition and *Ibicaba* was bought by German-speaking immigrants first arrived in Brazil as indentured sharecroppers in that very plantation. Albeit historically prominent, *Ibicaba* became a typical plantation of São Paulo's leading coffee zones. Studying post-Abolition labor relations in it is justifiable not only by data availability, but most prominently for that plantation's archetypical reflection of the history of São Paulo's coffee economy. On the eve of Abolition, Limeira was among the five largest coffee counties of São Paulo, while the province was on the brink of becoming the leading coffee exporter in the world (Simonsen 1978, Chapter 3; Colistete 2024, p. 5). *Ibicaba*'s productivity level, measured by volume of coffee harvested per bush, was then comparable to its regional average (Colistete 2024, pp. 12-3).

The classic historiography on post-slavery Brazil posits that agricultural labor markets fragmented along lines of skin color and race in São Paulo as a consequence of the massive inflow of European laborers since 1886 (Bassanezi 1973, 2019; Dean 1976; Andrews 1988). According to this view, after a short attrition caused by Abolition, freedmen returned to plantations under more precarious and worse-paying arrangements than recently-arrived immigrants. Trapped in worse occupational conditions, most freedmen would have been

[4] The usage of plantation records for quantitative assessments in the economic history of Brazil can be traced back to Jeanne Berrance de Castro's pioneering initiatives in cultural heritage preservation (1965; *cf.* Witzel de Souza and Santin Gardenal 2026). The usage of such data for scientific inquiry and economic analyses by contemporaneous, in turn, can be traced back to Louis Couty and C. H. van Delden Laërne (*cf.* Colistete 2024).

relegated to "[...] a nearly nomadic existence", implying the "[...] near impossibility of forming nuclear families" (Dean 1976, p. 153).[5] In contrast, Europeans, particularly Italians, would have accessed the most advantageous contracts, the so-called *colonato* arrangements (Bassanezi 2019; *cf.* Monsma 2010 for a critical review).

This view is embedded in three hypotheses about the functioning of agricultural labor markets that will be tested in this paper. First, based on racism, landowners would have preferred Europeans to Non-Whites. Taste-based discrimination would have barred Non-Whites from signing *colonato* contracts, leading to a *replacement* of Black fieldhands by Southern Europeans in São Paulo's coffee zones (Klein and Luna 2010, p. 131; Monsma 2010, p. 511; for a critical review, *cf.* Carvalho de Mello and Slenes 1980, pp. 119-22; and Colistete 2024, pp. 31-2). Second, landowners would have held positive priors about the overall productivity of Europeans *vis-à-vis* Non-Whites, which translated into lower remunerations to the latter (for a critical review, *cf.* Lamounier 2007).[6] Thus, even if hired, Non-Whites (freedmen or previously free) would have earned lower remunerations than Europeans for the same type of agricultural task. Third, landowners would have preferred to hire large households for *colonato* arrangements to incur in fewer transaction costs and benefit from more efficient intrahousehold divisions of labor. Smaller Black families, which further excluded women and children from fieldwork as a response to previous captivity, would have faced a disadvantage *vis-à-vis* larger European households (Andrews 1988, pp. 516-8; Monsma 2010, p. 527).

While properly describing the growth of aggregate labor supply caused by the gross immigration of more than two million foreigners to São Paulo in 1886-1930, these hypotheses require some qualifications. First, explanations embedded in taste-based discrimination presume an elastic supply of immigrants across the province. In reality, however, São Paulo had diverse regions of agricultural specialization, with labor demands that varied over time and by crops (Monbeig 1984; Colistete 2024, p. 19). Moreover, labor demand is estimated to have grown tenfold from 1886 to 1932 due to successive westward expansion of the agricultural frontiers (Holloway 1978, p. 192). Second, by the turn of the twentieth century, labor turnover and mobility between plantations and counties were features of the rural economy of São Paulo as a whole; planters recurrently used the very same word that Warren Dean employed to describe mobility among freedmen to complain about immigrants' turnover: 'nomadism' (Holloway 1978, p. 203; Andrews 1988; Bassanezi *et al.* 2008). Third, while it is unquestionable that racist premises embedded public discourses on immigration, it is much more doubtful that coffee planters were preoccupied with their fieldhands' skin shades. This is revealed, for instance, by their immediate acceptance of migrants from Northeastern Brazil after the end of São Paulo's immigration subsidization

[5] Dean's (1976) conclusion is in line with the so-called "anomy" hypothesis, which posited that slavery stripped the Black population of its ability to form stable families, with major negative consequences for their kinship ties and social relations. The anomy hypothesis has been categorically rejected by the findings of the demographic, familial, and cultural history of slavery (*cf.* Monsma 2010; Motta 1988; and Slenes 2011, Chap. 1, for exhaustive literature reviews; and Klein and Luna 2010, pp. 220-38, for a summary). Among that rich historiography, Slenes (2011), Fragoso and Florentino (1987), and Rios (1990) deserve special attention.

[6] As put by Dean (1976, p. 154): "planters evidently thought the work of the freedmen was less satisfactory than that of the Europeans [...] a false assumption, hence mere prejudice".

program (Holloway 1978, p. 206; Colistete 2024, pp. 11-2). The most efficacious pro-immigration policymakers in São Paulo aimed at creating a unified labor market, in which immigrants pushed down wages for all workers, freedmen included (Holloway 1978, pp. 203-4; Carvalho de Mello and Slenes 1980, pp. 119-22; Andrews 1988, p. 494; Versiani 1994; Monsma 2010, p. 520).[7] In such markets, freedmen could even have been among the most experienced laborers available. Illustratively, among the eight *colono* households considered by *Ibicaba*'s managers as having an "exemplary" behavior in 1890, only four were foreigners – all Portuguese –, while three of the others were Non-Whites.[8] Finally, societal perceptions of immigrants shifted profoundly from 1827, when the first state-sponsored rural settlement was founded in São Paulo, to 1927, when subsidized European immigration was terminated (Andrews 1988, p. 521; Sánchez-Alonso 2019). Arriving under different contracts, macroeconomic conditions, and geopolitical contexts, European laborers were hardly ever evaluated *a posteriori* as "superior" fieldhands, irrespective of racial motivations (if any) behind their hirings. Since the 1850s, disgruntled landowners complained about the moral qualities of various European nationalities, conflating these laborers' responsiveness to economic incentives with their cultural and national origins – in a fashion strikingly similar to the complaints about the vicious nature of Brazilian laborers (Graham 1973, pp. 11-4; Lamounier 2007, pp. 357-8; Monsma 2010, p. 517).

The evidence newly compiled in this paper leads to a number of revisions concerning the functioning of these agricultural labor markets. First, the hypothesis of labor replacement is rejected. In line with the findings of Monsma (2010), Cavalcanti Rodrigues (2021), and Colistete (2024), this research shows that a large number of *colono* households in *Ibicaba* were headed by Non-Whites. Although Italians and their descendants far outnumbered other *colonos*, households headed by Non-Whites ranked third in the pooled sample, above all other foreigners and below only Brazilians who could not be identified as Non-Whites. Second, after controlling for households' quantity-based TFPs, ethnolinguistic origins were not a significant determinant of remunerations earned by *colonos*. Among *colonos*, plantation managers did not discriminate in remuneration against Non-Whites or in favor of Europeans for the execution of homogeneous, unskilled agricultural tasks. Relatedly, laborers responded strongly to price incentives as a determinant of their annual earnings; during the lean season, price elasticities were even larger than unity. These results indicate a competitive labor market in which the "luxury" of taste-based discrimination would be mostly untenable. At the same time, results support the hypothesis of statistical discrimination. The number of economically active household members was the most consistently estimated determinant of *colonos*' annual earnings and of the earning differentials between Non-Whites and Europeans and their descendants. While Europeans, Italians in particular, had an average of seven people per household, the average size of a Non-White household was four people only.[9] Under

[7] Andrews (1988, pp. 502-3) unfortunately did not push forward this idea; rather, he relied on the literature that interpreted *colonato* arrangements as essentially dependent on taste-based discrimination.
[8] Data derived from BPML #0004 (*cf. Online Appendix* III).
[9] Results are consistent with those for São Carlos' County, *ca.* 90km to the North of *Ibicaba*. For that population, Monsma (2010, p. 528) found that families headed by Non-Whites varied between 4.3 and 4.5

labor arrangements highly dependent on intrahousehold divisions of labor, larger households were hired more frequently to cultivate coffee groves during the lean season and worked for longer during the harvest season. A major burden from slavery to Non-Whites was, therefore, the size of their households in the immediate post-Abolition.

These conclusions do not call into question the fact that racism is an inherent component of Brazilian history.[10] Even among *colonos* who were not discriminated against in remuneration, Non-Whites still had the lowest annual earning profiles of any ethnolinguistic group and carried major burdens from captivity, including demographic ones. Instead, the contribution of this research is to show that simplistic explanations of socioeconomic mobility miss important mechanisms of persistence and change. Along these lines, the predicted earning distributions of German-speakers and Portuguese are shown to be the closest to that of white Brazilians. Since these were the nationalities with the longest immigration histories as indentured laborers to São Paulo's coffee economy, results suggest a successful process of economic integration. In contrast, Italians and their descendants were the only group with a bimodal earnings distribution, suggesting the existence of a subgroup that failed to achieve their "(South-)American dreams" and about whom we know little regarding subsequent migratory patterns (return, circular etc.) and, perhaps, downward socioeconomic mobility *cf.* Sánchez-Alonso 2007, 2019; and Pérez 2017 for a contrast to the Argentinian case).

The paper is organized as follows. Section II sketches the history of the Brazilian coffee economy, linking it to the history of slavery, European indenturership, and mass immigration. Section III presents the data, focusing on the data-generating process of historical rural accounting and its consequences for the creation of datasets of interest to economic history. Sections IV present the descriptive statistics. Section V discusses the results, estimating the determinants of earnings and remunerations in coffee production for each ethnolinguistic group in *Ibicaba*'s labor force. Section VIII concludes.

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

From 1790 to 1890, coffee turned from an exquisite curiosity cultivated in courtyards of the Brazilian capital to the country's dominant export product (Simonsen 1978, pp. 175-7; Stein 1986, Chap. II; Absell 2020). Growth was essentially extensive, with plantations expanding in successive regional waves as planters responded to continuous constraints posed by transportation and labor costs (James 1932; Holloway 1978; Monbeig 1984). From the outskirts of Rio de Janeiro, coffee plantations spread along the *Paraíba River Valley*, benefiting from excellent agroclimatic conditions and proximity to the market and ports of the capital. By the 1820s, coffee already amounted to a fifth of all Brazilian exports (Simonsen 1978, pp. 192 ff.) The crossing of provincial borders, into São Paulo's portions of the *Paraíba River*, was a natural process of extensive expansion. By the late 1830s, the

members, while those headed by Italians had 5.6 members, on average. Spaniard and Portuguese households had 5.16 and 5.04 members, respectively, close to the 4.98 average of white Brazilians.

[10] For an economic reflection, *cf.* especially Andrews 1988, pp. 491-2; and Klein and Luna (2010, pp. 248-9).

volume of coffee harvested in São Paulo surpassed that of sugarcane, the province's leading crop in the late colonial era. Throughout this period, expanding plantations were fueled by the mass importation of slaves (Stein 1986, pp. 62-7; Fragoso and Florentino 1987, p. 158; Salles and Muaze 2022, pp. 83-5).

After long experimentation, coffee production took root in São Paulo's central-western plateau, the "Old West" agricultural frontier, in the 1840s (Dean 1976, pp. 30-49). Given its marginal economic importance up to that point, the stock of slaves in the region was comparatively small and the enforcement of the Brazilian ban on the transatlantic slave traffic as of 1850 curtailed the country's main supply of labor. The average price of a male prime-fieldhand in the county neighboring *Ibicaba* increased from 505 to 1,222 *milréis* from the 1840s to the 1850s, plateauing at 1,800 *milréis* in the 1860s-1870s (Dean 1976, p. 55).[11]

It was in this context that *Ibicaba* became one of Brazil's most prominent plantations.[12] Its founder, Senator Nicolau Pereira de Campos Vergueiro, and his firm, *Vergueiro & Co.*, started employing non-captive laborers from the German States, Switzerland, and Portugal as of 1840 (Buarque de Holanda 1941; Witter 1982; Heflinger Jr. 2007). These immigrants were hired under the so-called *Vergueiro system*. That contract's main dimension was a sharecropping-based labor arrangement. The net profit of a household's annual produce was shared between the household, who supplied labor for coffee cultivation and harvesting, and *Vergueiro & Co.*, which provided the agricultural capital and marketed the produce. Another dimension of that contract was an indentureship mechanism, by which *Vergueiro & Co.* advanced loans to cover the transportation and settlement costs of impoverished immigrants in Brazilian plantations; in return, immigrants agreed to tie their labor force to the repayment of outstanding debts (Davatz [1858] 1941; *cf.* Witzel de Souza 2024 and references therein).

The number of indentured sharecroppers thus immigrated remained a small fraction of the Black population trafficked as slaves into São Paulo's Old West coffee zones (Klein and Luna 2010, pp. 76-8; Colistete 2024, pp. 18, 23-4). Planters who could afford the prices of prime-fieldhands kept relying primarily on slave labor (Carvalho de Mello 1978; Dean 1976, pp. 57-61; Colistete 2024). Given the profitability of these plantations, slaves were shifted from other crops and tasks into coffee production and imported from other Brazilian provinces mainly into Rio de Janeiro's and São Paulo's coffee zones (Carvalho de Mello 1978; Carvalho de Mello and Slenes 1980; Andrews 1988; Klein and Luna 2010, Chapter 10).

Institutionally, however, these early migratory waves laid the ground for the Age of Mass Migration in Brazil. The later subsidization program to stimulate the immigration of agricultural laborers built upon the concept of credit interlocking. That mechanism allowed for mass immigration without the implementation of deep institutional reforms to make Brazil more attractive *vis-à-vis* alternative destinations, especially the U.S. and Argentina (Graham 1973; Holloway 1978; Andrews 1988, pp. 492-4; Sánchez-Alonso 2019).

[11] For similar trends in the provinces of Rio de Janeiro, Minas Gerais, and Bahia and Pernambuco (jointly), *cf.* Carvalho de Mello (1978), Klein and Luna (2010, pp. 298-9), and Carvalho de Mello and Slenes (1980).
[12] *Cf.* Witzel de Souza and Santin Gardenal (2021, 2026) for thorough historiographical reviews on *Ibicaba*.

Furthermore, the most enduring labor arrangement in the Brazilian coffee economy as of the 1880s, the *colonato*, had its origins in the unfolding of clauses first proposed by the *Vergueiro system* (Stolcke and Hall 1983; Stolcke 1986, pp. 35-48, 62-9; Lamounier 2007, p. 357; Witzel de Souza, 2024).

The typical *colonato* contract was a labor arrangement with two sources of monetary income (Bassanezi 1973; 2019; Dean 1976, Chapter 6; Holloway 1980).[13] First, households earned risk-free income during the lean season, for which they contracted a fixed number of coffee groves to cultivate during the year in exchange for a prespecified sum, paid to them in fixed installments. Plantation managers assigned households to *colônias* (literally, "colonies"), where they worked and resided. Within each *colônia*, households cultivated the coffee groves of specific *estacas*, literally "pickets" on the ground corresponding to the smallest unit of rural planning. These coffee groves had to be hoed, cleaned, weeded, and pruned.[14] In the period covered by this paper, five to seven prunings were executed per agricultural season in *Ibicaba*. Second, households earned a varying income during the harvest season, determined by the volume of coffee harvested and prevailing market prices. This source of income was highly volatile due to variations in the bushes' natural productivity and in market prices (*cf.* Section IV). According to the historiography, in the harvest season, the available labor force was allocated to harvest the most mature berries. Under this view, *colono* households would not necessarily harvest the same coffee groves they had cultivated in the lean season, in contrast to the *Vergueiro system*.[15]

Between 1886 and 1891, Brazil underwent some of the most significant institutional and economic changes of its history. After almost a half-century of resistance from the enslaved population and legislation gradually curbing the institutional survival of slavery, the abolitionist movement gained political and social momentum in the 1880s. Mass flees, violent strikes, and recurrence to legal manumissions finally led to unconditional Abolition in 1888. Concomitantly, the country joined decisively the circuits of the Age of Mass Migration (Graham 1973; Levy 1974; Dean 1976, Chapters 5-6). After passing clearer pro-immigration legislation in 1884, the economic potentates and, later, the government of São Paulo put in place a policy of subsidized immigration, which lasted from 1886 to 1927. Beyond subsidizing immigration costs, that program included organizations to hire, transport, and allocate immigrants to the coffee plantations (Holloway 1978, 1980). Between 1881 and 1921 *ca.* 3.2 million individuals (gross) immigrated to Brazil, 57% of whom to the province of São Paulo (Levy 1974; Bassanezi *et al.* 2008; Witzel de Souza 2023, pp. 309).

[13] *Colonos* were also allowed to cultivate subsistence goods either between coffee rows, or in separate fields, depending on the bushes' ages and productivity. Laborers strongly preferred the former and, hence, favored employment in agricultural frontiers, as new lands were capable of sustaining subsistence agriculture in-between coffee rows (Holloway 1980; Stolcke 1986, pp. 53-69; Colistete 2024, pp. 12-3).
[14] This contractual component was named "carpa", the Portuguese noun to "pruning".
[15] Paysheets in the *Ibicaba Collection* support this interpretation, since laborers were pooled together, separated only by labor categories, but not by *estacas*. In contrast, it is implicit in some ex-*colonos*' memories that they harvested the same coffee groves they had cultivated in the lean season (*e.g.* Tonelotto in Witzel de Souza and Santin Gardenal 2021, pp. 355-6; *cf.* also Fragoso 1983, pp. 51-2).

Paradoxical as it may seem for the inaugurators of the scheme of indentured immigration, the *Vergueiros*' era at *Ibicaba* collapsed with Abolition. Highly indebted since the 1860s and holding the largest slave chattel in Limeira County by the 1870s, *Vergueiro & Co.* faced its bankruptcy right after Abolition (Marcondes 2021; Heflinger Jr. 2021). *Ibicaba* was publicly auctioned in 1890 to a consortium led by Simão and José Levy, German-speaking immigrants who had worked indentured sharecroppers in the very *Ibicaba* from 1857 to 1871 (Tamiazo 2021; Witzel de Souza and Santin Gardenal 2026).

Coffee production in São Paulo surpassed Rio de Janeiro's in 1886 (Simonsen 1978, Chapter 3; Stein 1986, Chaps. IX-X). By the turn of the twentieth century, agricultural frontiers in São Paulo's "New West", which had expanded to agronomically evermore propitious regions for coffee cultivation since the 1870s, further benefitted from mass immigration and rapidly growing railroad networks (James 1932; Berrance de Castro and Scarano 1971, pp. 717-20; Monbeig 1984). In the first three decades of the twentieth century, São Paulo's westernmost frontiers gradually overshadowed the Old West in total output and productivity. The volume of coffee harvested in Rio Claro (SP) County peaked in the 1900s (Dean 1976, p. 38). That of *Ibicaba* did so in 1912 (*cf.* Online Appendix III). Consequently, some counties in the Old West started diversifying agricultural production, with Limeira County relying on sugarcane and citrics. During the *Levys*' era, while maintaining its historical fame and regional importance, *Ibicaba* growingly became a typical plantation of São Paulo's Old West.

Regional variations notwithstanding, "Coffee was King" in the Brazilian First Republic (1889-1930), never amounting to less than 50% of total national exports (Delfim Netto 1979, p. 16 ff.; Stolcke 1986, pp. 53-6). Planters, however, grew increasingly preoccupied with the capacity of global markets to absorb increasing supplies at stable prices. Consequently, as Antonio Delfim Netto (1979) puts it, the era of "coffee markets under state intervention" began in 1906. Coalitions of the leading coffee producing states, spearheaded by São Paulo, captained market interventions in 1906, 1917, and 1921. The strategy was to hold stocks to manipulate international price levels. The short-term and fortuitous successes of these policies, the scale of financial resources required, and their macroeconomic implications led to the federalization of the program in 1924. In spite of legislation to the contrary, for almost a quarter century these policies further stimulated the expansion of coffee production, geographically and within plantations (Defilm Netto 1976, pp. 143-58). Due to plummeting international demand, contracting financial markets, and severe domestic constraints, the "coffee-defense policies" became untenable as of October 1929.

The political coalitions supporting the First Republic were shaken and fell to a *coup d'état* in 1930. Nonetheless, emergency measures were kept in place to support coffee planters throughout the Great Depression, including programs to buy and burn coffee sacks (Delfim Netto 1979, pp. 101-8). New federal and state programs attempted to increase the value-added of Brazilian coffee since then (Delfim Netto 1979, pp. 236-9; Stolcke 1986, pp. 188-95). In the 1950s, producers were even paid to eradicate older, less productive coffee groves (Stolcke 1986, pp. 152-70, 200-4; *cf.* Olivatto in Witzel de Souza and Santin Gardenal 2021). *Colonato* contracts endured in the plantations as long as coffee remained São Paulo's main cash crop. As of the 1950s, however, profound structural changes led to the demise of

that labor arrangement. Crop diversification implied varying demands for labor in different periods of the year. The flexibility of hiring daily fieldhands and the ease to monitor them per task surpassed the benefits of maintaining a large residing labor force devoted to a single crop. In addition, sugarcane was supplied to monopsonistic sugarmills that either provided labor gangs, or supported the daily gathering of harvests. Wage laborers, including informal fieldhands hired per day or task, finally replaced the old *colonos* (Stolcke 1986, pp. 179-81).

III. Data: From historical rural accounting to a cleaned dataset

The historical ledgers from the *Ibicaba Collection* can be classified as containing predominantly plantation management data or labor accounting.[16] This paper uses all ledgers from the latter category referring to "Coffee harvest registries" (seven ledgers) and "Registries of coffee trees assigned to laborers" (nine ledgers).

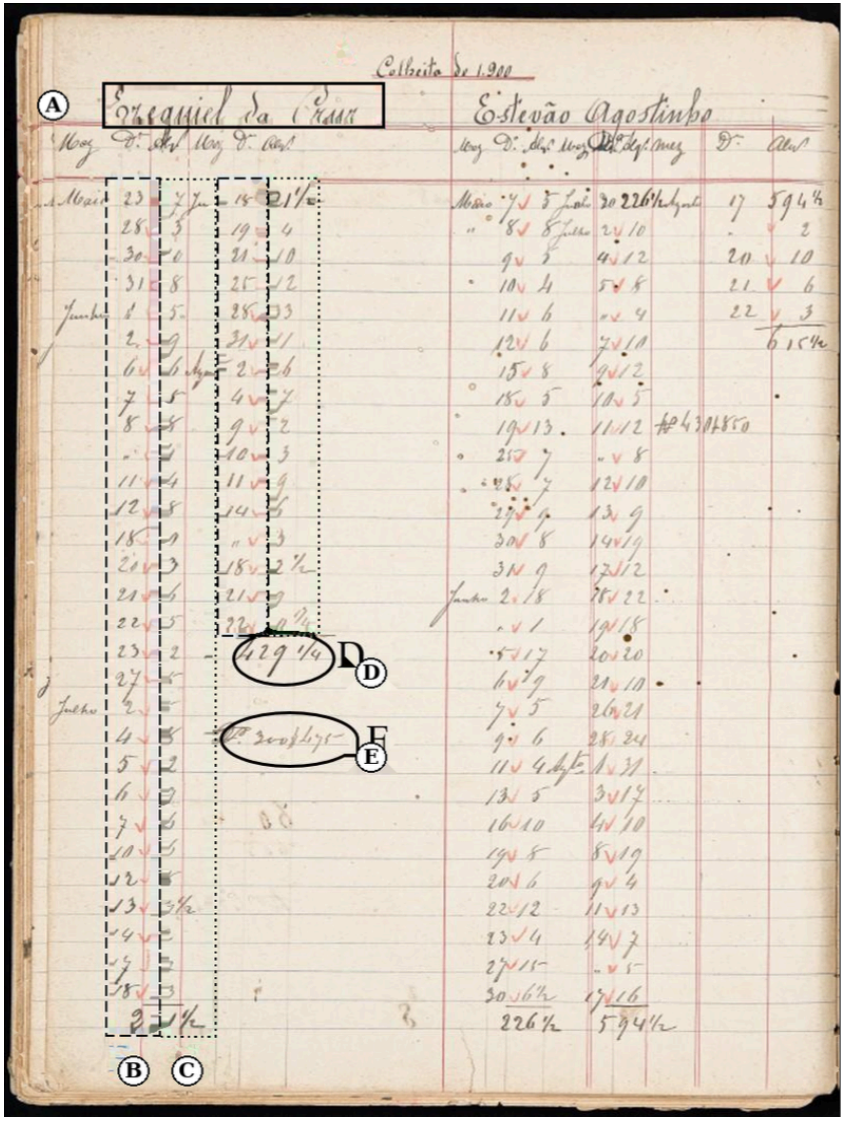
"Coffee harvest registries" (henceforth *Harvest ledgers*) recorded the volume of coffee harvested by each household per day during the harvest season. These ledgers were usually organized as current accounts under the name of a household head, to whom the volume of coffee harvested per day was assigned. Figure 1-A illustrates a typical entry in such a ledger. Under the heading of a household head (A), it registered the exact days when the household worked in harvesting (B), the amount of coffee harvested per day (C) and in total (D); the corresponding prices were either explicitly recorded, or, as in this example, can be retrieved from total output (D) and the ensuing income (E). For this paper, I aggregated the data per harvest season, computing average prices and nominal earnings for all households.

"Registries of coffee trees assigned to laborers" (henceforth *Registries*), in turn, were population counts of residing laborers that recorded the agricultural capital assigned to workers. Figure 1-B illustrates a typical entry in such a ledger. Under a household head (A) – usually patriarchs, oldest economically active men, or widowed matriarchs –, these ledgers most frequently listed all household members (B) and their demographic characteristics (C), including marital status, sex, age, and nationality. 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 or cultivated coffee (D) and the number of coffee groves per *estaca* or *talhão* therein (E). Since *estacas* were the smallest unit of agricultural planning, these entries provide a very precise proxy for the quality of the agricultural capital available to each household. *Registries* further included prices paid for the cultivation of each *estaca* and the number of prunings contracted per year, which were either explicitly recorded (F), or can be derived from total output (G) and the ensuing income (H). Finally, most *Registries* included information on residences. Accuracy and detail of such data vary across ledgers, but the most complete entries include the numbering of the house within the colony, number of rooms, windows (with leaves), doors (with leaves), locks in the front door, and electric lightbulb sockets (as of 1926), if any.

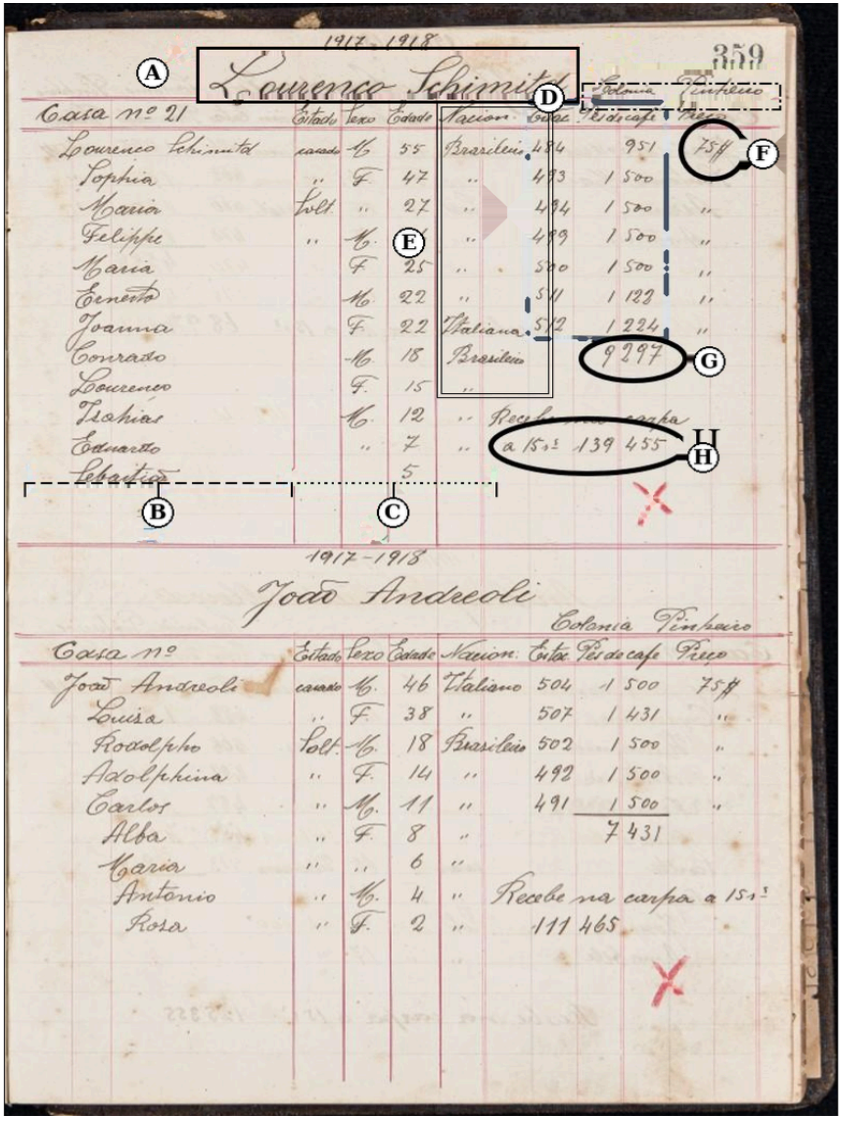
[16] This is obviously only a crude distinction, since ledgers frequently included more than one type of accounting and all managerial ledgers necessarily made references to labor costs. For details on the collection, *cf.* Witzel de Souza and Santin Gardenal (2026). For a thorough discussion of the data compilation, *cf.* Online Appendix I.

Figure 1 – Sources: Annotated *Harvest Ledgers & Registries*

Panel A – Ledger #0013 (Digital Image 144)



Panel B – Ledger #0043 (Digital Image 181)



The layout of the *Registries* and the various loose sheets annexed to them reveal that those ledgers were compiled as annual population counts conducted at each *colony* as a survey. As an illustration, the slips of paper annexed to ledger #0043 (DIs 258-277) contain data on households' demographics, coffee cultivation, and housing, separated by colonies; their dates and various handwritings indicate that they were written by different sub-directors at each colony.[17] Given the errors identified in the households' series (especially in ages and age orderings of the youngest children) and references to family ties therein ("mother", "brother", "sister", "in-laws" etc.), it is likely that only household heads were surveyed about the characteristics of individual household members. Such information was passed on to general managers and accountants, who altered preexisting entries in the *Registries* if household had changed between years (by birth, marriage, death, in- or out-migration etc.) Moreover, managers allocated newly arrived households to colonies (and *estacas* therein) where labor was most demanded, and reallocated them accordingly once contracts were renegotiated (*cf.* Section IV on labor distribution among *Ibicaba*'s five largest colonies).

For the 1,205 uniquely identified households in the pooled sample, the average household was listed for 5 years (median: 3 years). These households frequently renegotiated contracts, as evidenced by the changing IDs of the *estacas* they cultivated. This implied that *Registries* had to be updated frequently. Such operations led to a plethora of corrections, glosses, crossing-outs, and overwritings, with great variation of clarity and consistency within and between ledgers (*e.g.* #0043, Digital Image 14 – detail here).[18] The first decision in compiling the dataset was to choose what information best corresponded to the actual economic and demographic characteristics of households at any point in time. Data quality also varied in the availability of demographic characteristics, exact location of *estacas*, and details on housing. The compiled dataset adheres as strictly as possible to accuracy in data transcription and to minimizing inconsistencies across complete household series. Missing data or inconsistencies in naming, sex, marital status, and age have been interpolated or corrected using households' entire series, *i.e.* after matching all households within and between ledgers.[19] The same applies to the identifier of *estacas* and the number of coffee groves therein. Whenever inconsistencies arose from corrections and glosses present in the sources, the baselines used the data that most closely matched the total number of coffee groves reported for that year, with alternative interpretations assessed in the robustness checks.

Households were matched within and between ledgers based on the names and demographic characteristics of the household head and of their members in mono- or multinuclear families. The same unique identifier was assigned to a household if its head remained in the sample for the entire series or if a widowed matriarch took over the headship after the passing away of her husband, independently of what happened to other members. However, I distinguished between two cases if a son took over the headship of a household from a deceased patriarch.

[17] Numerous annexes have been contemporaneously added to the ledgers. They are described in the "Abstract" of each object in the *Ibicaba Collection*. A complete list is in Witzel de Souza and Santin Gardenal (2026).
[18] "Digital Image" henceforth referred to as DI, for brevity.
[19] The few inconsistencies or missing data on marital status were corrected assuming no divorce and no illegitimate births. Of the 19,616 entries compiled with all household members, only one person was ever registered as "Apartado", *i.e.* divorced. There were cases in which one partner was missing for one or more years; however, I did not impute these individuals to the series, nor made any change to their marital statuses.

If that son was already married before the passing away of the previous household head, the same identifier was assigned to that household, since it continued to be a multinuclear household as it had been before. However, if the new household head was a single son, I assigned the same identifier to that household only until that son got married, because the newly married couple would lead to substantial alterations in the composition of the original household (by childbearing, splits of siblings who married out etc.) Another minor complication is that a few households made private arrangements among themselves, *e.g.* household *i* working on behalf of household *j* in year *t*. Bearing no connection to their private businesses, *Ibicaba* managers remunerated household *j*, who must have paid household *i* according to their own arrangements, obviously not recorded in the plantation's ledgers. In these exceptional cases, I attributed physical yields to the household that actually produced them, as physical output is the outcome of interest for computing quantity-based TFPs; earnings, however, were assigned to each household as recorded in the sources.

The historical managerial and accounting practices of a plantation were very complex. During the lean season, its daily operations required allocating labor to the cultivation of coffee and other cash crops; to subsistence agriculture and orchards; to animal husbandry; and to workshops (*e.g.* carpentry and blacksmithing) and mills (*e.g.* corn flour and cane brandy). Most of these tasks were carefully tracked, as evidenced by the variety of accounting objects preserved in the *Ibicaba Collection*. Since most entries cut across various managerial operations, the same information had to be accurately copied and updated between various ledgers. For instance, overseers and assistant managers made quick notes in *Daybooks* and *Laborers' Timesheets*. Once the workday or piecerate was completed, that information had to be copied into *Ibicaba*'s *Current Accounts* and *Cash Flows*, subsequently used to compute the plantation's annual yields, revenues, costs, and profits. That same information had to be disaggregated for each household and copied into *Laborers' Current Accounts*. Finally, legal mandates and good managerial practices implied that data on output and labor remuneration had to be recorded in a booklet (*caderneta*) held by each household head.

During the harvest season, households exchanged the coffee berries they harvested for tokens or receipts (*bilhetes*) corresponding to the volume handed over.[20] Overseers and assistant managers who received the berries in the field recorded those yields into *Harvest Ledgers*, *Daybooks*, or slips of paper that were then passed on to general managers. By the end of the harvest season or subperiods thereof, managers and accountants checked the tokens or receipts that laborers presented back to them and compared those to the yields registered in the ledgers. Mismatches were taken very seriously. *Harvest ledgers* #0068 and #0077 have detailed notes on tokens lost by laborers during the harvest season, with strict monitoring in the 1920s. The mismatch in the account of José Braz on August 11, 1927, is illustrative (#0068, DI 34). By noting that "there is a difference of 1 extra *alquere* [SIC] in the slip from the harvesters' list", the accountant reveals to us that he was comparing the daily registries given to him, probably by an overseer, with the number of tokens held by the laborer.[21]

[20] Also referred to in the sources as *cartão* (card), *alqueire* (unit of volume), *ficha* (chip), and *talão* (coupon). In a 1946 note, an accountant mentioned "13 [*alqueires*] sobra taboleiro", probably referring to an abacus.
[21] Similarly, a 1929 note informs that "One card has not been delivered on the 1st [/] We received 403, but harvested [=] 402" (#0077, DI 10). A detailed and curious case is that of Martim Zanetti's household (*Idem*, DIs

Such complex accounting practices and contemporaneous corrections, as well as historical allegations of accounting fraud against immigrants raise concerns about data quality overall. Online Appendix I evaluates accounting errors in earnings reported for coffee cultivation. No evidence was found for a significant correlation between accounting errors and the sources from which the data was transcribed, or for time trends. The mean overall accounting error computed for the corresponding 2,905 entries equaled 0.066 *milréis* per pruning, *i.e.* minorly positive in favor of the plantation;[22] the median, however, was actually zero.[23] Excluding approximation errors of 0.001 *milréis* per pruning, the mean positive errors equaled 1.125 *milréis* (median: 0.033 *milréis*).[24] The latter was, however, driven by an outlier of 120.694 *milréis* – an error too large to have gone unnoticed at the time.[25] There were also 58 negative errors, *i.e.* favoring laborers, with an absolute mean of 2.043 *milréis* (|median|: 0.300 *milréis*).

Ethnolinguistic origins: Definitions

National origin was the *Registries*' most inconsistently recorded variable. While most entries followed the *jus-soli* legislation and assigned Brazilian nationality to any Brazilian-born child, loose *jus-sanguinis* criteria were also adopted at times – and yet other entries indiscriminately classified every household member as foreigner or Brazilian, even where previous entries had categorized some as foreigners (usually the elderly) and others as Brazilians (usually young children). By listing all origins recorded in the sample, I first consistently classified every individual into seven ethnolinguistic groups based on skin colors and/or national origins: *Non-Whites*, *Italians, Germanics, Portuguese, Spaniards, Eastern Europeans*, and a group that will be referred to as *White Brazilians*.[26] In addition, my classification of ethnolinguistic origins relied on the creation of a list of surnames found in the *Registries* to identify direct descendants of Non-Whites and of Europeans within *Ibicaba*. Within the sample, surnames were identified as having ever-belonged to *Non-Whites*, *Italians*, *Germanics*, *Portuguese*, *Spaniards*, or *Eastern Europeans*.

Non-Whites were persons ever classified as "non-whites" in any consulted sources, including *Harvest ledgers* that recorded household heads as "First Name – Preto" and ledger BPML #0004, which classified some individuals by skin color. Except for two African men in 1888-1890, all other *Non-Whites* were Brazilians. *Non-Whites* also include individuals that,

---

111, 113). In August 1937, they 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. The accountants did add that extra *alqueire* to the household's credit, but made it clear that he was very attentive to any deviations between tokens and field notes.

[22] For reference, the 1904-1942 mean pooled price for cultivating one coffee grove was 0.025 *milréis*.

[23] The exercise obviously depends on the assumption that the sources correctly reported *prices* and *quantities*.

[24] 892 observations had errors larger than zero. However, 617 were smaller than 0.001 *milréis*, *i.e.* approximation errors in millesimal units that neither accountants nor laborers would have been able to compute.

[25] Account of Benedicto Cunha, 1929-1930 (#0062, DI 205). Strangely, the accountant copied to Cunha's entry the exact same earnings registered on that same page for another household (João Franco Silveira).

[26] *Italians* include individuals classified in the sources as Italians, Italian-Brazilians, Austrian-Italians and the unidentified *Chare/Charecia*. *Germanics* include Germans, German-Brazilians, Austrians, Swiss, Danish, and German-Danish, *i.e.* individuals inconsistently classified under these nationalities. *Portuguese* include Portuguese and Portuguese-Brazilians. *Spaniards* include Spaniards, Spaniard-Brazilians and one Argentinian descending from an Italian-Spaniard couple. *Eastern Europeans* include Russians, Polish, German-Russians, and German-Russian-Polish-Lithuanians, *i.e.* individuals inconsistently classified under these nationalities.

within the sample, could be assigned as descendants of *Non-Whites*. Among the latter, 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, especially those listed in ledger BPML #0004. *Italians*, *Germanics*, *Portuguese*, *Spaniards*, and *Eastern Europeans* were individuals whose birthplace belonged to these nationalities, or, within the sample, could be assigned as their descendants. This classification emphasizes the intergenerational transmission of cultural traits among Brazilian-born individuals of European descent within *Ibicaba*. Finally, *White Brazilians* were Brazilians (or individuals without recorded origins) netted out from all individuals identifiable as *Non-Whites*, or who did not have a surname traceable *within* the sample to non-Iberian Europeans.

($Origins_h$) corresponds to the ethnolinguistic origin of the household head, *i.e.* the person who actually bargained contractual conditions with *Ibicaba*'s managers. For the few cases in which household heads varied in origin over time, I assigned them to the ethnolinguistic group that was culturally most distant from Brazilians (*Eastern Europeans* $\succ$ *Germanic* $\succ$ *Italians* $\succ$ *Spaniards* $\succ$ *Portuguese*).[27]

Table 1 presents *Ibicaba*'s labor force composition by ethnolinguistic origin. Panel A pools all households for 1888-1958 – the vast majority of them *colonos*. Panel B tabulates origins and occupations in the immediate post-Abolition using data from BPML #0004 (1888-1890).

Ibicaba's pooled population was representative of the Age of Mass Migration to São Paulo. *Italians* constituted the bulk of laborers, followed by *Portuguese* and *Spaniards*, while *Germanics* remained an important minority, probably as a consequence of older chain migration to the region (Graham 1973, pp. 20-2; Levy 1974; Bassanezi *et al.* 2008). In line with new findings by Monsma (2010) and Colistete (2024), the pooled data further reveal a large number of *Non-White* and *White Brazilian colonos*, who were the third and second most common group in the sample, respectively. Further countering the labor replacement hypothesis, from the 174 uniquely-identified *colono* households in *Ibicaba* in 1888-1890, *Non-Whites* were the (distant) second most numerous, trailing only the *Italians*. While it is equally true that *Non-Whites* also predominated in the most precarious labor categories, the evidence shows that *Non-Whites* were not barred from *colonato* in São Paulo's Old West.[28]

[27] Three alternative classifications of ethnolinguistic origins increasingly augment the likelihood of households being classified as *Non-White* or of non-Iberian origins. These classifications are based on extended surname-nationality pairs for surnames not listed as belonging to foreigners in the *Ibicaba Collection*; and on categorizing entire households as *Non-whites* or as belonging to the culturally farthest European origins if that household ever had a member who was either *Non-White* or whose ancestry was traceable to that farthest European nationality. Reassuringly, these alternative categorizations do not alter the empirical conclusions (Online Appendices V). Consequently, for being solely anchored in the sources, being attuned to Brazilian perceptions of ethnolinguistic origins, and reflecting the origins of the person who bargained with plantation managers, *Origins_h* will be our main variable of interest in this paper. A thorough discussion of such criteria and lists of all surname-nationality pairs can be found in Online Appendix II.

[28] *cf.* Online Appendix VI for the correlates of being a *colono* household in the immediate post-Abolition era.

Table 1 – *Ibicaba*'s labor force composition by ethnolinguistic origin

	Wht. Brz.	Non-Wht.	Italians	Grms.	Portug.	Span.	East. Eurp.	Total
			Panel A - Pooled data for 1888-1958 (with gaps)					
Origins$_h$								
n	193	65	379	30	101	42	16	826
N	739	434	2,659	214	658	179	70	4,953
			Panel B - Immediate post-Abolition, only (1888-1890)					
Colonos								
n	14	23	85	4	19	15	14	174
N	17	26	160	5	23	25	14	270
Other agricult.								
n	1	12	0	1	1	1	0	16
N	1	12	0	1	1	1	0	16
Managers								
n	1	0	2	1	0	0	0	4
N	1	0	2	1	0	0	0	4
Skilled								
n	2	2	1	2	1	0	0	8
N	2	2	1	2	1	0	0	8
Unskilled								
n	3	5	0	1	2	0	0	11
N	3	5	0	1	2	0	0	11

Notes: (1) *N* refers to total number of observations; *n* to uniquely identified households. (2) *Other agricultural laborers* include "camarada", "trabalhador", "camarada & colono", "agregado", and "empreiteiro". (3) *Managers* include "administrador", "director", "escrivão", and "inspector". (4) *Skilled* include "marceneiro", "farinheira", "ferreiro", "foguista & lavrador", "machinista", and "pedreiro". (5) *Unskilled* include "ajudante", "campeiro", "carregador de comida", "carreiro", "carroceiro", "conductor de serviço", "cozinheiro", "jardineiro", "puxador", "servente", and "vigia".

IV. Descriptive Statistics

Table 2 provides the demographic characteristics of *Ibicaba*'s labor force, the characteristics of houses, and laborers' mobility between colonies in the plantation. The pooled sample contains a maximum of 5,595 observations (*N*) corresponding to 1,205 (*n*) uniquely identified households, pooled from 1888 to 1958 (with gaps). Most observations are concentrated around the year 1914 and remain, on average, 4.98 years in the sample.

Households had an overall sample mean of 6.7 people, distributed in 1.3 families.[29] Men were 54.90% of the labor force, reflecting the presence of single male laborers in addition to *colonato*'s typical family-based structure. Out of the 27 uniquely identified households (30 observations) that had only singles (marital status) in them, 13 had just one male: one *White Brazilian*, one *Italian*, one *Spaniard*, two *Germanics*, and eight *Non-Whites*.[30] 38.90% and 2.99% of the overall sample mean were married or widowed, respectively. The residing population was very young, with an overall sample mean of 23.4 years (median: 11.5).[31] Households had an overall mean of *ca.* 5 members in the age range 7-60y, *i.e.* this paper's definition of economically active individuals.[32] Consequently, the consumer-to-producer ratio at the household level was only 0.45, indicating that most household members were in position to contribute in coffee production, subsistence agriculture, or family care. Households were thus well-equipped to labor-intensive tasks, but, as discussed thoroughly in Section V, such endowments varied largely by ethnolinguistic origins, especially the number of economically active members in a household.

The average household in *Ibicaba* resided in one single house with 5-6 rooms and one lock in its front door. Excluding common rooms for a kitchen and (perhaps) a living room, the average 6.7 household members shared their private spheres in 3-4 rooms. Electric lightbulb sockets were registered for the first time in 1926 and remained a luxury. They were reported only in *Grama Larga* and *Morro Alto* colonies, as well as in *Iracema farm*: 114 uniquely identified households (317 observations) had one lightbulb socket, and 12 others (25 observations) had two.[33] Once in *Ibicaba*, most households stayed put in the first colony of allocation. With few exceptions (*cf.* maxima),[34] the average household changed the colony where it cultivated coffee less than once, and most worked in less than two colonies (1.51).

---

[29] Number of families is defined as (#widows + (#married/2)). The computation of multinuclear households takes into account only those households with all required data on civil status. The median shows that the majority of households was mononuclear, with the average pushed up by large multinuclear households.

[30] If additional data from summary tables and from household-heads-only are considered, the number of uniquely identified households with just one single (marital status) individual increases to 18 (20 observations), adding two additional *White Brazilians*, two additional *Italians*, and one additional *Spaniard*.

[31] In line with demographic data of Southern European immigrants estimated by Levy *apud* Clevelario Jr. (1997, pp. 52-3). Similarly, Holloway (1978, p. 198) and Colistete (2024, p. 10) show that, during the Age of Mass Migration to São Paulo, *ca.* 70% of immigrant *colonos* were 12y of age or older.

[32] Maximum ages averaged across all households reveal that, for the sample as a whole, the oldest household member was 48.04 years old (S.D. = 14.55).

[33] Online Appendix III reports figures on housing per capita and changes over time. For most households, housing conditions did not change while they lived in *Ibicaba*.

[34] Four households cultivated coffee in as many as four different colonies and one even moved seven (known) times between colonies.

Table 2 – Descriptive statistics, pooled sample 1888-1958 (with gaps)

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

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

Table 3 turns to *Ibicaba*' productive capacity, disaggregated by its various colonies. Data are pooled for all years with available information (1888-1958, with gaps) at the level of colonies, *i.e.* summing up all people and coffee groves ever recorded in them. *Colônia Grande* (literally, "Large Colony") lived up to its name, followed by *Grama Larga* and *Morro Alto* – and, after a considerable gap, by *Saltinho & Pinheiro* and *Pão-de-Ló* colonies –, each with more than one million coffee groves across all available data. These major colonies rank in the same order in regard to total population, indicating the absence of labor-saving technologies in any of them (*cf.* Holloway 1978, p. 189).

The last two columns of Table 3, in turn, report the average number of coffee groves and *estacas* cultivated per household. These averages were computed with actual cultivation data, *i.e.* not as the ratio of aggregate groves per total population. The five largest colonies listed above are here at the center of the distribution, with households cultivating an average of 6,500 coffee groves across 6 *estacas* in them. These figures suggest that managers allocated most laborers to these large colonies, while either demanding higher effort or accepting worse outcomes in colonies such as *Mariquita*, *Paineira*, and *Colônia Nova*, where households cultivated, on average, more than 8,000 groves across 7 *estacas*. *Linha* and *Boa Esperança*, notably colonies where only *Non-Whites* resided, had the smallest number of cultivated groves per household.[35] This figure alone should not be taken as evidence of discrimination. First, the data show no evidence supporting the hypothesis of spatial segregation (*cf.* Online Appendix III). Second, most freedmen wished to live far away from old slave quarters as a condition for returning to rural employment in the post-Abolition era (Andrews 1988, p. 515).

The average household in the pooled sample cultivated 6,377 coffee groves across 6 *estacas*, which were pruned 5-6 times per year (Table 2). At an average nominal price of 0.025 *milréis* per grove cultivated, the average household earned 963.900 *milréis* during the lean season. Further, the average household harvested coffee for nearly four months, between mid-May and mid-August, corresponding to the peak harvest season in the 1910s (*cf.* Kuhlmann *apud* Holloway 1980, p. 130). At an average nominal price of 0.890 *milréis* per *alqueire*, the overall mean volume of 491.10 *alqueires* harvested per year led to an income of 364.900 *milréis* during the harvest season. The mean annual nominal income of a *colono* household was thus 1,328.800 *milréis* for the pooled sample. With an overall mean of 6.7 people per household, the average annual earning per capita from coffee cultivation was about 199.910 *milréis*, or 269.260 *milréis* per person of working age. These figures align well with other estimates of agricultural income for early-twentieth-century São Paulo. Annual income in government-sponsored rural settlements were estimated at 219.000 (per laborer) and 225.000 (per capita) *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, an unskilled agricultural laborer with add-ons for subsistence (*i.e.* similar to a *colono* who was allowed to cultivate subsistence goods) earned 1.196 *milréis* per day (Zamberlan Pereira 2020, p. 213) – equivalent to 225 days of work for a person of working age in *Ibicaba*.[36]

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

[36] For additional benchmarks, *cf.* 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 total of 800.000 *milréis* per annum. In 1901, the Italian owner of a plantation in Taquaritinga County estimated that 3 economically active people could cultivate 4,000 coffee groves (earning 320.000 *milréis* in the lean season) and harvest 480 *alqueires* (earning 240.000 *milréis* in the harvest season). In 1903, the Italian Vice-Consul in Ribeirão Preto County estimated that

Table 3 – 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	Total population	Total groves	# of cult. Groves	# of cult. Estacas
Col. Grande	1025	217	5,921	6,042,748	6159.78	6.26
Grama Larga	883	240	5,311	5,789,359	6723.99	6.32
Morro Alto	509	124	2,996	3,343,457	6740.84	6.30
Saltinho-Pinheiro	237	59	1,367	1,496,144	6679.21	5.83
Pão Ló	146	47	1,021	1,027,281	7390.51	5.78
Col. Ibicaba	185	120	816	404,045	3848.05	3.08
Lage	89	35	451	649,738	7555.09	6.76
Teteia	102	48	435	572,992	5907.13	5.13
Abundância	17	17	105	25,569	3652.71	.
Mariquita	9	2	69	95,342	10593.56	10.44
Port. Isabel	3	1	12	18,574	6191.33	6.67
Boa Esperança	15	15	60	18,439	1317.07	.
Paineira	1	1	1	10,157	10157.00	10.00
Linha	4	2	7	9,702	2425.50	2.75
Col. Nova	1	1	.	8,022	8022.00	7.00
Ipiranga	3	3	27	.	.	.
Multiple colonies	2	2	18	15,237	7618.50	.
Laborers	70	70	260	.	.	.
Iracema farm	53	39	333	294,001	5547.19	6.30

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

2 workers and one housewife could cultivate 5,000 coffee groves (300.000 *milréis*) and harvest 450 *alqueires* (180.000 *milréis*). Finally, the 1903 estimates of the Department of Agriculture in the *Mogiana Zone* expected four workers to cultivate 10,000 coffee groves (800.000 *milréis*) and harvest 1,200 *alqueires* (600.000 *milréis*). For comparability, I excluded estimates from subsistence agriculture from these figures.

Finally, Figures 2 and 3 plot, respectively, the number of coffee groves cultivated in the lean season and the volume of coffee harvested in the harvest season (per household and per person of working age) for 1904-1942, the period in which the bulk of the data for both agricultural seasons overlap.

Except for short-term fluctuations, the average number of coffee groves annually cultivated per household almost doubled from 1906 (5,054) to 1929 (9,098), before rapidly declining in the 1930s.[37] The period of growth coincides with the 1906, 1917, and 1921 "coffee-defense" policies, with the expected lags of 3-5 years for the blossoming of newly-planted bushes. In contrast, starting in 1929, the average number of coffee groves cultivated per household declined steadily as the Great Depression unrolled (Holloway 1978, p. 206). Trends per person of working age (7-60y) are similar, except for 1906-1921. In that interval, the number of coffee groves cultivated per household increased, but not per person of working age, implying that *Ibicaba*'s managers were either hiring households with more people in working age, and/or that residing households benefited from a demographic gift. Thereafter, from 1921 to 1929, the average number of coffee groves cultivated by person of working age rose from 1,250 to a maximum of 1,900. The subsequent decline 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 3 illustrates, at the micro-level, the long-term decline of the coffee economy in São Paulo's Old West. Since we do not know the ages of the coffee groves, it is not possible to attribute this decline to decreasing labor productivity or to the deterioration of *Ibicaba*'s agricultural capital (Dean 1976, pp. 175, 180). However, Figures 2 and 3 combined show that the mean volume of coffee harvested per household was falling simultaneously to substantial 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 bushes or exogenous shocks (*e.g.* the 1906 supercrop or the 1918 Great Frost), which can be visualized in the series, but do not explain the trend. In fact, the highly significant linear trend implies that, in every agricultural season from 1904 to 1942, the average household was harvesting 15 *alqueires* less than in the preceding year. This trend cannot be solely attributable to the Great Depression, either: if the series is split in 1929, the trends for the sub-periods 1904-1929 and 1930-1943 are equivalent. The same conclusions are reached with volumes harvested per person of working age.[38]

The harvest series also reveals the short-term volatility that was typical of coffee production. Beyond exogenous shocks, the agronomy of trees led to natural cycles of high and low yields in 1-2y intervals (Dean 1976, p. 106; Holloway 1980). The within-S.D. of volume harvested per household in 1904-1942 corresponded to 50.44% of that variable's overall mean; in contrast, the within-S.D. of the number of coffee groves cultivated per household was, in that same period, 24.51% of the latter's overall mean.

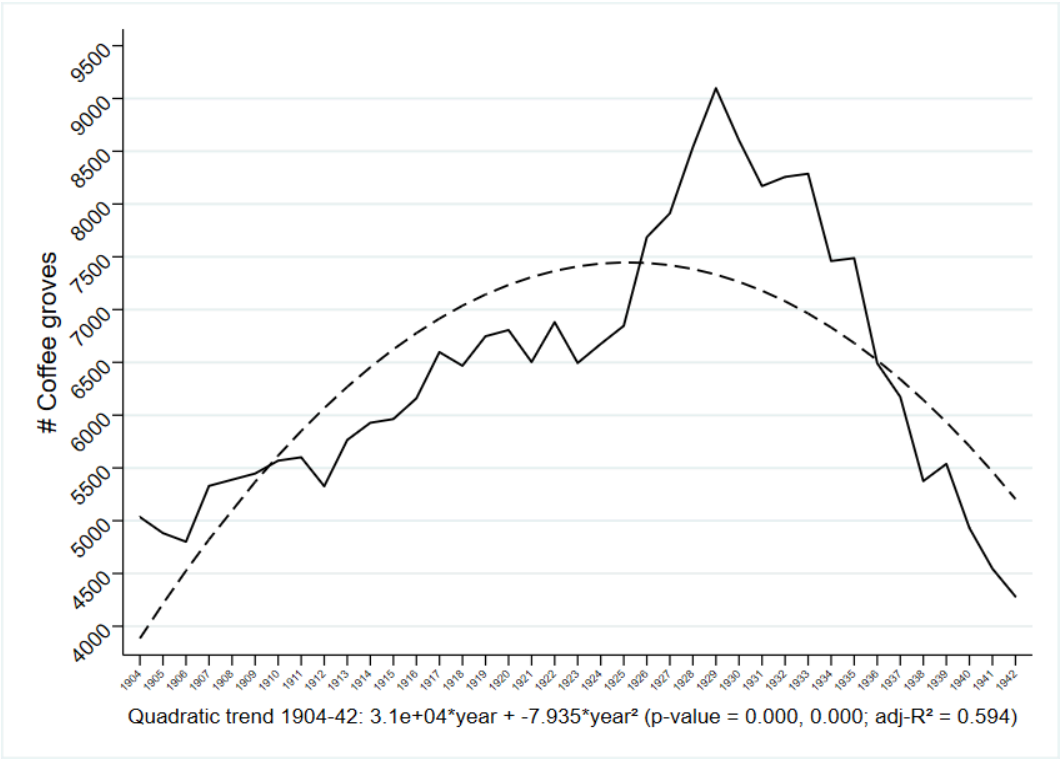
---

[37] Interestingly, the 1942 observation was comparable to the last figures of the *Vergueiro*'s era. In 1890, households cultivated, on average 3,730 coffee groves, which had declined from 4,996 in 1888.

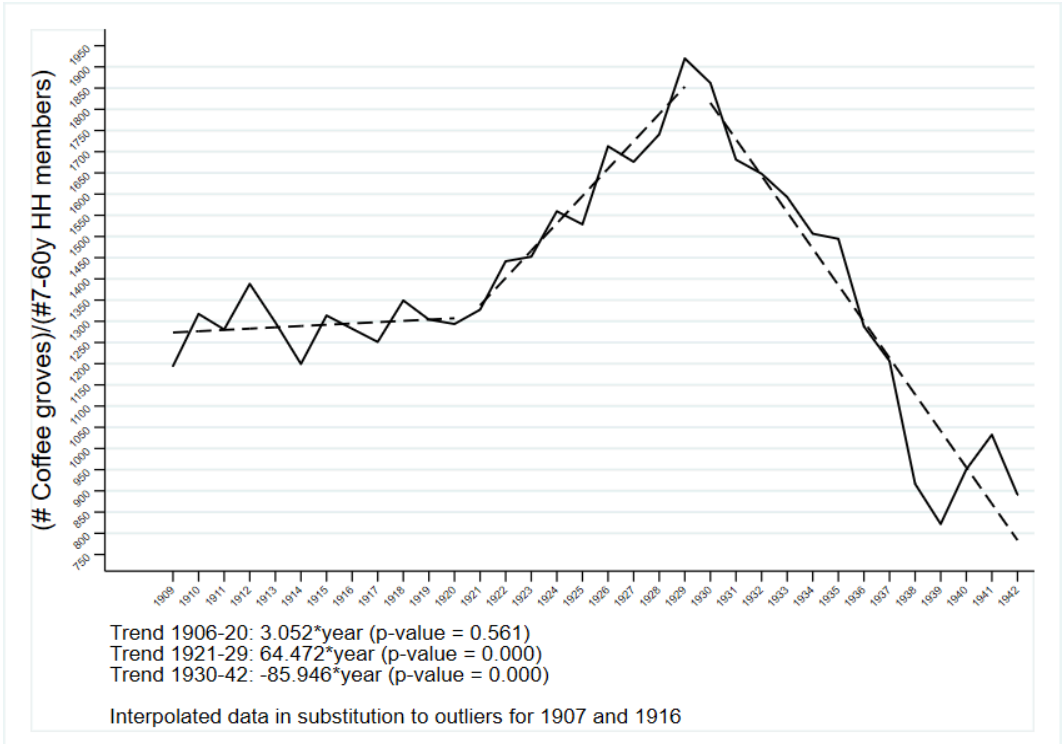
[38] For the period 1888-1958, the quadratic trend for volume harvested *in Ibicaba* peaks in 1912, in line with Dean's (1976) estimates for the neighboring county of Rio Claro (SP) (*cf.* Online Appendix III).

Figure 2 – Mean number of coffee groves cultivated in the lean season (1904-1942)

Panel A – per household



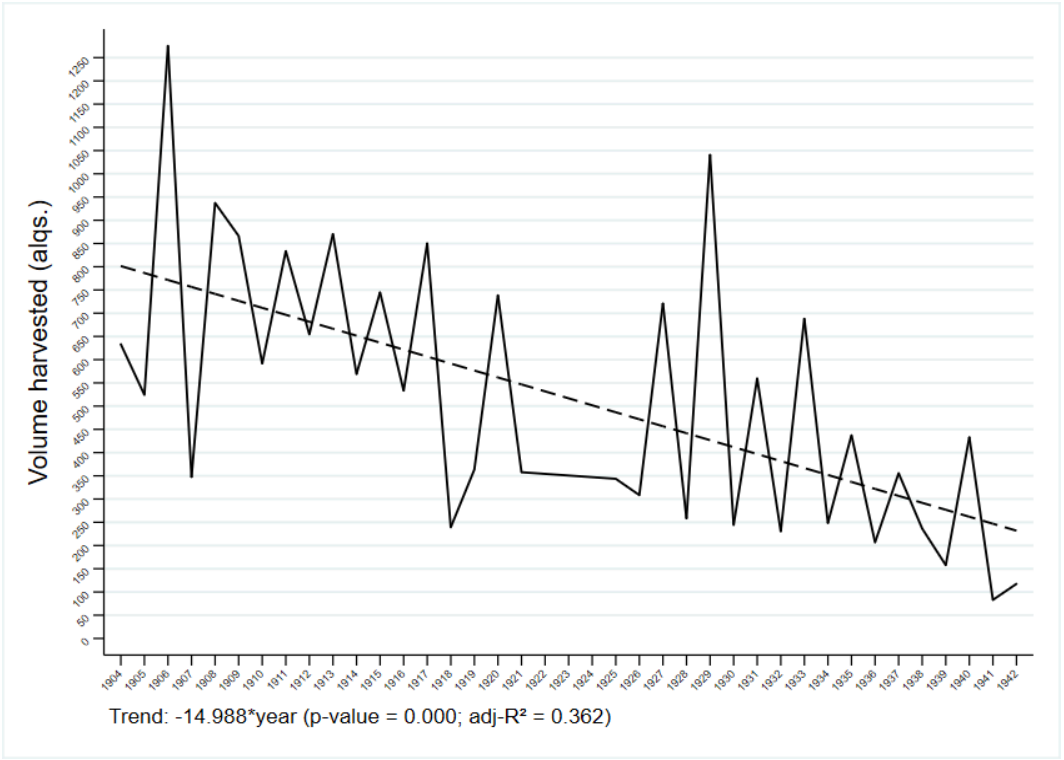
Panel B – per person of working age (7-60y)



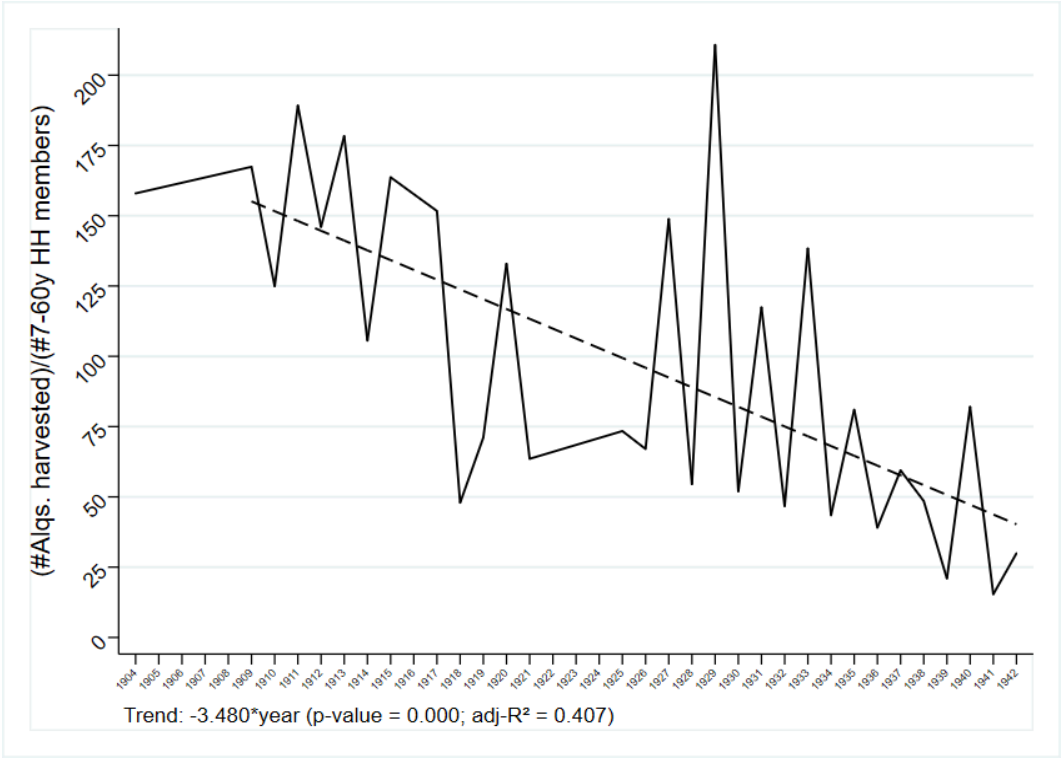
Note: Data for production per person of working age are interpolated for 1906 and 1917, as the original data show unreasonable spikes, probably due to undercounting of the working population in those years.

Figure 3 – Mean volume of coffee harvested (*alqueires*) in the harvest seasons (1904-1942)

Panel A – per household



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



Note: *Cf.* Figure 2.

V. Earnings and Discrimination: Empirical results

This section starts by disentangling the partial effects of prices, frequency of executed tasks, agricultural capital, and households' characteristics on their earnings. These estimates are then used to predict the earning profiles of each ethnolinguistic group and to decompose their earning differentials according to each group's observable characteristics.

The baselines estimate the determinants of households' earnings in coffee production netted out of prices and frequencies of tasks executed in the lean and harvest seasons (Foster, Haltiwanger, and Syverson 2008), according to the following specification:[39]

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

in which $Y_{ht}^{output} = (Y_{ht}^{cult}, Y_{ht}^{harv})$ refers to the nominal earnings of household h in year t.[40] Y_{ht}^{cult} and Y_{ht}^{harv} are estimated separately and refer to household's earnings from coffee cultivation and harvesting, respectively.[41] ϵ_{ht} is an i.i.d. disturbance.

p_{ht} is the price (remuneration) negotiated by the household for executing the task and N_{ht} is the frequency in which that task is executed. 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 a household's labor inputs. The baseline controls for the number of household members of any age (household size) and the number of individuals in the age range 7-60y (*i.e.* the economically active), independent of sex.[42]

$\left(K_{ht}\right)'$ refers to the agricultural capital available to the household. For the lean season, $\left(K_{ht}\right)'$ controls for colony of work and for the mean, S.D., and mean deviation from the mean (MDEV) of *estacas*' identifiers.[43] These measures of central tendency and dispersion control for the coffee groves that households cultivated in the lean season, thus accommodating for all characteristics and shocks shared by coffee groves in the smallest unit of rural planning.[44] For the harvest season, $\left(K_{ht}\right)'$ controls for the same variables and, additionally, for the timing

[39] Baselines without controls for prices and frequency of executed tasks (*i.e.* total effect of inputs on earnings); and without frequency only (*i.e.* net of task-assignments only) are reported in Online Appendix IV.
[40] Given the unit of observation in the *Harvest ledgers* and the relevant unit for *colonato* arrangements, all analyses are conducted at the household level. Individual workers are treated as single-person households.
[41] Beyond their differing time trends (*cf.* Section IV), coffee cultivation and harvesting can be seen as effort- and care-intensive activities, with distinct implications for laborers' responsiveness to coercion and incentives (Versiani 1994, p. 471, nonetheless, considers both tasks effort-intensive).
[42] Results are robust to different age windows and controls for sex and family structures (Online Appendix V).
[43] Results are robust to alternative measures and definitions of *estaca* identifiers (Online Appendix V).
[44] Soil quality, topography, and distance from colonies influenced remunerations in coffee cultivation and harvesting (Dean 1976, p. 168).

of the harvest, based on the assumption that households were given priority to harvest the same coffee groves they had themselves cultivated during the lean season.

Household's fixed effects (α_h) capture its time-invariant characteristics, including cultural beliefs (labor ethics, thriftiness, norms etc.) and ethnolinguistic origins. Year fixed effects (λ_t) accommodate common shocks (agroecological and economic) across households.

Results reported in Table 4 show that price elasticities were highly significant, even with year-FEs. During the lean season, the point estimate (1.285) is statistically larger than unity. Because the specification controls for the frequency of executed tasks, the larger-than-unit estimate implies that households responded more than proportionally in their labor-efforts to changes in remuneration. For the harvest season, the highly significant point estimate (1.313), however, is not statistically different from one. Moreover, the frequency of executed tasks was statistically and economically very significant. A household hired to execute one more pruning than that of the overall sample mean would have increased its earnings by 21% in the lean season.[45] It should be recalled, however, that only five prunings per year were registered prior to 1921.[46] For subsequent years, results show that a household's capacity to contract more prunings had a major impact on its annual earnings. The effect of being engaged for longer in harvesting was even more substantial. A household working one month longer in harvesting than the overall sample mean would have increased its earnings in the harvest season by 174%.[47] At the 95th percentile, households were engaged for up to five months in harvesting per year, with only the top-1% engaged for six months. Working for so long in harvesting was exceptional, but it really paid off economically.[48]

Results show the importance of having a large number of economically active members in the household. Significant at the 1% level, *ln(#members 7-60y)* implies labor elasticities of 0.34 and 0.40 in the lean and harvest seasons, respectively – point estimates that are very robust across alternative specifications and age ranges (*cf.* Online Appendix V). Household sizes, however, had no effect beyond that of members in the age range 7-60y. As noted by Luiz Tonelotto, a Brazilian-born *colono* of Italian descent 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 […]" (in Witzel de Souza and Gardenal, p. 357; *cf.* also Holloway 1978, pp. 189-90; Monsma 2010, pp. 527-8).

[45] Evaluated at the sample mean = ((1 + 1/5.42418)^1.139 - 1)*100.
[46] In the nineteenth century, 3-4 prunings were executed per year in the region (Dean 1976, p. 36). In Rio de Janeiro's *Paraíba Valley*, only 3 *carpas* were executed annually (Fragoso 1983, p. 51).
[47] Evaluated at the sample mean: ((1+ 1/3.937352)^4.447 - 1)*100.
[48] The timing of harvest was relevant (*cf.* Fragoso 1983, pp. 51, 133). *C.p.*, households that started to harvest earlier had a disadvantageous position; however, the partial effect of increasing the mean month of engagement in harvesting is negative and significant, suggesting competition for the best bushes between latecomers.

Table 4 – Determinants of earnings from coffee production in *Ibicaba*, Lean and Harvest seasons

	Dep. Var.:	
	$\ln(Y_{ht}^{cult})$	$\ln(Y_{ht}^{harv})$
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)
Controls for location of *Estacas*		
Mean	Yes	Yes
MDEV	Yes	Yes
SD	Yes	Yes
HH-FE	Yes	Yes
Year-FE	Yes	Yes
Colony: Lean season	Yes	Yes
Observations	2,358	1,470
Number of id	512	324

Notes: (1) Robust standard errors in parentheses. (2) *** p<0.01, ** p<0.05, * p<0.1.

Based on these estimates and each ethnolinguistic group's characteristics, their average annual earnings can be estimated as the sum of predicted earnings in the lean and harvest season, *i.e.* $\hat{Y}_{ht}^{total} = \hat{Y}_{ht}^{cult} + \hat{Y}_{ht}^{harv}$. Figure 4 plots the corresponding Kernel Density Estimates, contrasting each group to the pooled sample (solid line) and *White Brazilians* (dashed line).[49]

The mean predicted annual earnings of a household in the pooled sample was 1,471.839 *milréis*, slightly above the overall sample mean (1,328.800 *milréis*) reported in the descriptive statistics.[50] *Non-White* households were by far the poorest, with predicted annual earnings equivalent to 45% of the sample's mean and 47% of the *Italians* – the lowest among foreigners, except for a single *Eastern European* household. A few *Non-White* households escaped this trap, as shown by distribution's right tail, but they were exceptions.[51] *White Brazilians*, in turn, outperformed all groups at the mean and median, while households of different European origins showed rather distinct profiles. *Italians* were the only group with a non-unimodal distribution.[52] While some fared far better than the sample mean, others fell short of average – contrary to the classic interpretation of the benefits of *colonato* arrangements for Italian immigrants, irrespective of any other characteristic.[53] The earning profiles of the *Germanic* and *Portuguese*, in turn, were much more spread, with the former reaching the largest predicted mean among all groups and the latter attaining the third and second highest mean and median predicted earnings. These groups' closeness to *White Brazilians* can most likely be attributed to their longer migratory histories to São Paulo, including under the *Vergueiro system* (similar conclusions in Monsma 2010, pp. 515-6; Berrance de Castro and Scarano 1971, p. 720).

Households headed by *Non-Whites* had, on average, only 4 members, while those headed by *White Brazilians* and Europeans had 6-7 members. To account for such differences, Table 5 ranks all groups by their mean and median predicted earnings per capita, separately for the lean and harvest seasons. Surprisingly, while having the lowest mean and median earnings per capita in the lean season, *Non-White* households reached the 3rd highest mean earnings per capita during the harvest season – and the very top of the median ranking. In contrast, given the sizes of their households, *Italians* ranked third to last and last in their mean and median earnings per capita during the harvest season. These results raise the possibility that *Non-White* households could have rationally allocated labor between seasons, using their comparative advantage in harvesting to accumulate a level of income per capita that they considered sufficient for the rest of the year. The rationale for *Italians* was the opposite: the household as a *unit* had to compensate in the harvest season for their comparatively poor performance in the lean season; their large sizes also implied that, in per capita terms, *Italians* could not specialize in one season at the expense of the other (*cf.* Andrews 1988, p. 518).

[49] Online Appendix V reports the descriptive statistics of the corresponding predictions.
[50] Table 5 estimates the earnings of households that worked in both seasons, while the descriptive statistics report data for households that could have worked in one season, only.
[51] *cf.* also the outliers among *Non-Whites* in the boxplots of earnings reported in Online Appendix V.
[52] Unimodality was rejected at the 10%-significance level.
[53] However, in line with Andrews's (1988, p. 493), Carvalho de Mello and Slenes's (1980, pp. 119-22), and Versiani's (1994, p. 465) views on labor market competitiveness in the coffee economy of post-Abolition Brazil.

Figure 4 – Kernel Density Estimates of predicted earnings among *Ibicaba*'s laborers, by ethnolinguistic groups (pooled for 1888-1958, w/ gaps

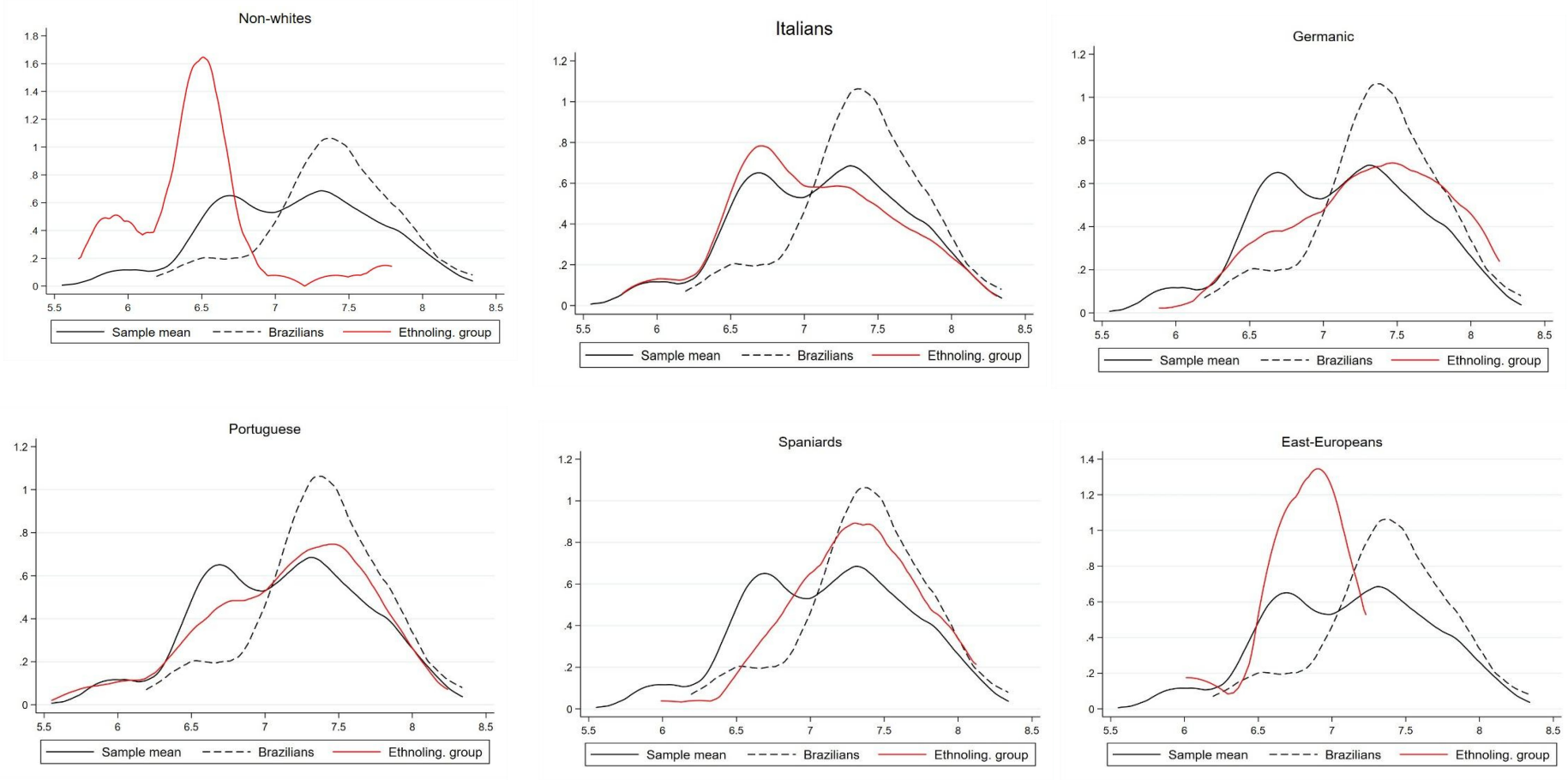


Table 5 – Rankings of ethnolinguistic groups by mean and median total predicted earnings per capita (in *milréis*)

Mean		Median	
Earnings (*milréis*): Lean season			
White Brz.	234.223	Germanic	202.310
Germanic	225.123	White Brz.	188.783
Spaniards	185.065	Spaniards	162.168
Portuguese	178.694	Portuguese	155.965
Italians	156.062	Italians	121.804
East. Europ.	136.689	East. Europ.	102.206
Non-Whites	126.633	Non-Whites	75.046
S.D.	38.278	S.D.	42.883
Earnings (*milréis*): Harvest season			
Germanic	67.870	Non-Whites	56.061
White Brz.	67.066	Germanic	54.538
Non-Whites	66.875	White Brz.	53.159
Portuguese	60.326	East. Europ.	48.614
Italians	54.562	Portuguese	47.455
East. Europ.	52.477	Spaniards	44.374
Spaniards	51.261	Italians	44.213
S.D.	6.780	S.D.	4.480

Notes: (1) S.D. computed with data from all ethnolinguistic groups. (2) Number of observations for each ethnolinguistic group = (N Coffee_cult; N Harv): *White Brazilians* = (382; 200); *Non-Whites* = (69; 47); *Italians* = (1376; 886); *Germanics* = (114; 78); *Portuguese* = (312; 195); *Spaniards* = (86; 53); East. Europeans = (19; 11).

Oaxaca-Blinder decompositions of earning differentials

Household demographics, rational choice, and racial discrimination are non-exclusive potential explanations for differences in earnings across ethnolinguistic groups. These explanations align well with the baseline results, which showed the centrality of households' labor force composition (an endowment), frequency of engagement in agricultural tasks (a choice variable for managers and laborers), and remuneration (a potential source of taste-based discrimination) in determining earnings.

Before testing for taste-based discrimination, this section first decomposes the earnings' differentials between groups with the twofold Oaxaca-Blinder (OB) decomposition.[54] These decompositions assume the existence of non-discriminatory parameters, estimated with the pooled sample, and refer solely to cross-sectional differences between groups (Jann 2008, p. 5).[55] In other words, they do not take into account dynamic changes within or between groups, for two reasons. Econometrically, the model identification is so heavily dependent on categorical variables that re-scaling them for dynamic comparisons would generate meaninglessly complex estimates (Heitmüller 2005). Historically, beyond Abolition, I see no reason to set specific thresholds after which differences in endowments and coefficients should vary systematically between the groups considered.

Table 6 coalesce the twofold OB decompositions for the lean and harvest seasons.

During the lean season, *Non-Whites* were the only households with mean earnings significantly smaller than that of the pooled sample. Households headed by *White Brazilians*, in turn, had higher earnings during the lean season *vis-à-vis* the sample mean, with a marginally significant difference.

White Brazilians had an advantage over the sample mean in their explained component, which stemmed from higher remunerations earned for cultivating coffee and being more frequently engaged in pruning. Because *Non-White* and *Italian* households suffered a penalty in these two variables, the positive effect for *White Brazilians* cannot be attributed to racial discrimination. Their explained advantage in remunerations cannot be attributed to language skills (*e.g.* bargaining capacity) either, considering the significant negative effect for *Non-Whites* (essentially all Brazilians) and the non-significant effect for the *Portuguese*.

Non-Whites, in turn, had significant and large overall disadvantages in the explained components of remuneration, number of economically active members, and number of prunings. It would be incorrect to attribute these endowment disadvantages to discrimination, since *Italians* faced an equally significant disadvantage in their explained components related to remuneration and number of prunings, even if the absolute disadvantage of *Non-Whites* was larger. In regard to the unexplained components, *Non-White* households had a major

[54] *cf.* Online Appendix V for (i) diff-in-means tests that do not require the assumptions of OB decompositions; and (ii) alternative reference groups for the OB decompositions. Results remain qualitatively unaltered.

[55] The OB decompositions follow from LSDV re-estimations of the baseline, controlling for colony, year, and household categorical identifiers, with first categories omitted and errors clustered at the household level. Coefficients thus estimated are identical to those of Table 4. Given the panel's unbalanced nature, the option *relax* was required. Twofold decompositions were preferred to threefold, since the latter's coefficient differences depend on intercept differences – which in the LSDV model simply pool all omitted categories.

disadvantage in the number of prunings, implying that they faced an endowment *and* an unexplained penalization in the frequency with which they were engaged in that task. If we interpret the explained difference as a choice variable (à Andrews 1988, pp. 514, 523), then the additional difference in the unexplained component can be attributable to discrimination.

Finally, *Italian* households enjoyed a demographic advantage, especially relative to the endowments of *Non-Whites* and to the unexplained component of *White Brazilians*. Large, well-functioning, and young households provided an advantage to Italians in the Brazilian coffee economy (Andrews 1988, p. 515-6; Clevelario Jr. 1997, pp. 52-3; Monsma 2010, p. 515). Once again in the words of Mr. Tonelotto: "You asked me whether it was hard to save money? Well… It depended. My father sometimes made some money, because the family was numerous [and] worked hard [...]" (in Witzel de Souza and Santin Gardenal 2021, pp. 356).[56]

With respect to harvesting, *Italian* households had the highest predicted mean earnings, significantly above the sample mean. Their main advantages lay in the length of engagement in harvesting and the number of economically active members. In contrast, *Non-White* households had the lowest mean predicted earnings, significantly below the sample mean. Their main disadvantages stemmed from the number of economically active members and the length of engagement in harvesting. However, the disadvantage of *Non-Whites* in the unexplained component of the number of months engaged in harvesting cannot be attributed to racial discrimination, as *White Brazilians* also had a significant disadvantage in it. Finally, households headed by *White Brazilians* had an explained advantage in remuneration that adds to the effect found for the lean season. The hypothesis of taste-based discrimination in remuneration, however, is again rejected, since both *Non-White* and *Italian* households faced an explained disadvantage in that same variable.

The OB decompositions thus point to a strong demographic disadvantage among *Non-White* households in *Ibicaba*, operating directly through the number of economically active members in their households, and indirectly through the frequency of their engagement in agricultural tasks. These results show a demographic burden from slavery carried on by *Non-Whites* in the post-Abolition era. Such a conclusion is in no way a revival of the anomy hypothesis, as *Ibicaba*'s labor force included a large number of well-structured *Non-White* households under *colonato* arrangements (*cf.* Table 1). Notwithstanding, throughout the nineteenth century, *Non-Whites* suffered with severe sex imbalances, particularly in the booming coffee zones that concentrated prime fieldhands (Fragoso and Florentino 1987, pp. 157-60; Motta 2006; Slenes 2011, pp. 80-1), *Ibicaba* included (Marcondes 2021). Similarly, there was an overall decline in the share of slaves ever-married as of the 1860s and the natural growth among the enslaved population of Brazil was most likely negative in 1870s, resulting from high mortality rates and low nativity rates (Slenes 2011, p. 100; Bercovich 1988; Clevelario Jr. 1997, p. 55).

[56] Berrance de Castro and Scarano (1971, p. 725), Stolcke (1986), and Andrews (1988, pp. 515-6) notice that the acceptance of harsh labor conditions was part of an immigrant's calculations in deciding to emigrate.

Table 6 – Twofold Oaxaca-Blinder decompositions: Earnings from the lean and harvest seasons (Reference group: Pooled sample)

Dep. Var.: $ln(Y_{ht}^{cult})$

	White Brazilians		Non-Whites		Italians		Germanics		Portuguese		Spaniards	
Pooled 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***	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	-0.231***	-0.761	0.475***	2.555	0.147***	0.972	-0.157	-0.0937	-0.0680	-0.321	-0.128	-6.704**
ln(N)	-0.0702***	-1.265	0.0893***	1.850***	0.0439***	-0.169	-0.0292	3.276	-0.0129	0.722	-0.0390	0.438
ln(HH size)	0.00215	-0.307	0.00826	0.0898	-0.00265	0.0672	0.00105	-0.290	0.00105	-0.161	-0.00150	0.0829
ln(# members 7-60y)	0.0400	0.429**	0.183***	-0.0206	-0.0396*	-0.181	-0.00530	0.0107	0.00121	-0.0146	-0.0136	0.619***

Dep. Var.: $ln(Y_{ht}^{harv})$

	White Brazilians		Non-Whites		Italians		Germanics		Portuguese		Spaniards	
Pooled sample	5.700***		5.696***		5.554***		5.681***		5.695***		5.690***	
Ethn. 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***	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	-0.292***	0.0312	0.428**	-3.321	0.185**	-0.300	-0.148	-0.538	-0.115	0.786*	-0.168	5.225
ln(N)	0.105	9.663***	-0.0934	3.423**	-0.213**	-0.263	-0.119	-28.58**	0.294*	-6.156***	0.467	616.6***
Mean month harv.	0.0706	-32.71**	0.0750	-12.49**	-0.0455	-11.88	0.0846	-12.17**	0.0159	-7.850*	-0.131	-14.18**
1st month harv.	-0.180	20.21**	0.0205	8.288**	0.209*	4.511	-0.0372	-16.39*	-0.236*	-5.049	-0.101	515.3***
Last month harv.	-0.00106	1.971	0.00657	4.094	0.00420	7.636	0.00517	4.969	-0.00760	16.84***	-0.0140	-842.5***
ln(HH size)	-0.00635	0.0814	-0.0554	0.695	0.0138	0.132	-0.00563	0.0292	-0.00677	-0.486	0.00499	-1.375
ln(# members 7-60y)	0.0202	-0.0621	0.239***	-0.276	-0.0354	-0.394	-0.0110	-0.182	0.000115	0.519**	0.00138	0.945

Notes: (1) N = (2358,1470) for the lean and harvest seasons, respectively. (2) Estimates based on LSDV with categories for colony of work, year, and household-identifier; the full set of controls follow the baseline specification (*cf.* Online Appendix V for a discrimination of every colony). (3) Standard errors clustered at the household-level; (4) Reference group: Pooled sample, for all estimations. (5) Option "relax" required by the panel's unbalanced nature.

Testing for taste-based discrimination in remuneration

This section tests explicitly for taste-based discrimination in remuneration, defined as different pay by ethnolinguistic origins for the execution of the same task, after controlling for households' quantity-based TFPs and year-FEs:

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

in which p_{ht}^{output} is estimated separately for prices (remunerations) earned by household h in year t during the lean (p_{ht}^{cult}) and harvest (p_{ht}^{harv}) seasons. E_{ht} is an i.i.d. disturbance.

$Origin'_{ht}$ is the set of indicators for *Non-White*, *Italian*, *Germanic*, *Portuguese*, *Spaniard*, or *Eastern European* households, with *White Brazilians* as the reference group. $\delta_g \lesseqgtr 0$ captures discrimination against or in favor of group g.

$\hat{A}_{ht} = exp(\hat{e}_{ht})$ is the household's quantity-based TFP, which provides a metric of intra-household "technologies" in coffee cultivation and harvesting netted out of inputs and time-invariant characteristics (Atkin, Khandelwal, and Osman 2019, pp. 444-6; Foster, Haltiwanger, and Syverson 2008), obtained as the residual from the following estimate:

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

in which $Output_{ht}$ is estimated separately for physical output in the lean ($Cult_{ht}$) and harvest ($Harv_{ht}$) seasons. ($Cult_{ht}$) refers to the total number of coffee groves cultivated by household h in year t and ($Harv_{ht}$) to the volume of coffee harvested (in *alqueires*).[57]

Table 7 shows no evidence of taste-based discrimination in remunerations earned by different ethnolinguistic groups for the execution of homogeneous, unskilled agricultural tasks, in the lean and harvest seasons, once their TFPs and common shocks are controlled for.[58]

[57] Historians of Brazil might question the usefulness of studying the determinants of *Cult_ht*, as it is commonly assumed that the number of cultivated coffee groves by *colono* households was given by their age composition. This assumption is based on custom and contract. Since the *Vergueiro system*, contracts stipulated a fixed number of coffee groves to be cultivated per person of working age. However, such interpretation ignores that households could contract the cultivation of more or fewer groves over time. Besides changes in size and composition, households' productive capacity, labor effort also depended on 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 (*cf.* Sections III and IV).

[58] Misspecified models missing controls for households' quantity-based TFPs and/or year-FEs would lead to the incorrect conclusion that *Non-White*, *Italian*, *Eastern European*, and, less robustly, *Portuguese* households were all discriminated against. These incorrect results are available upon request.

Table 7 – Testing for taste-based discrimination in prices earned by ethnolinguistic groups: Lean & harvest seasons

	Dep. var.: Average price earned for:	
	Cultivating one coffee grove	Harvesting one *alqueire*
TFP agr. task	0.00144***	0.00731***
	(0.000279)	(0.00268)
Non-Whites	0.000321	0.00500
	(0.000207)	(0.00372)
Italians	-8.63e-05	0.00443
	(0.000230)	(0.00396)
Germanics	0.000473	0.000883
	(0.000446)	(0.00426)
Portuguese	-0.000270	0.00601
	(0.000254)	(0.00470)
Spaniards	-0.000184	-0.0136
	(0.000424)	(0.00930)
East-Europs.	-0.000557	0.00342
	(0.000733)	(0.00378)
HH controls	Yes	Yes
Year-FE	Yes	Yes
Observations	2,360	1,470
Number of id	513	324

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

VI. Concluding remarks

This paper studied the functioning of agricultural labor markets in the most important coffee zone of the immediate post-Abolition Brazil. The newly collected micro-evidence qualifies three leading hypotheses of the literature. First, *Non-White* households were not barred from *colonato* contracts. Their position as second to third most common ethnolinguistic group of *colonos* rejects the replacement hypothesis, which posits that European *colonos* substituted Black laborers in the post-Abolition era. Such conclusion validates, at the micro-level, similar evidence for the economic agency of freedmen found at the regional- and state-level by Monsma (2010) and Colistete (2024). Second, results showed no evidence of discrimination in remuneration against *Non-Whites* or in favor of Europeans once their quantity-based TFPs and common temporal shocks are controlled for. While the literature usually emphasizes the growth in labor supply implied by mass immigration,[59] this paper showed that labor demand remained inelastic in regions competing with agronomically more productive agricultural frontiers. In that tight labor market, plantation managers were not able to discriminate by ethnolinguistic origins in their payments for unskilled labor. Third, households of various European origins showed distinct aggregate earning profiles. Households of *Portuguese* and *Germanic* origin performed similarly to *White Brazilians* – a result explainable by those groups' long migratory history to São Paulo's coffee economy. Even more importantly, this paper rejects that the *colonato* was a safe path to the "(South) American dream" of *Italians*. While the average *Italian* household did not perform as poorly as *Non-Whites* in *Ibicaba*, it is certainly true that an important parcel of *Italians* did fail economically. The geographic and socioeconomic mobility of the poorest among them and their descendants is worth pursuing (à Andrews 1988, p. 493).

At the same time, most *Non-Whites* in *Ibicaba* remained poverty-stricken at the household level, even among those who signed *colonato* contracts. Households' demographic composition and the length and timing of their engagement in coffee cultivation and harvesting were entangled in complex ways. Statistical discrimination by the number of economically active members in a household was certainly at play. *Ceteris paribus*, plantation managers favored households with more economically active members; the smaller size of *Non-White* households was a burden they carried from centennial captivity. Discriminated against in the frequency of engagement in rural tasks, *Non-Whites* seem to have specialized in harvesting, reaching levels of annual income that were high at the per capita level, but that remained amongst the lowest at the household-level.

In conclusion, the new evidence compiled by this paper does not at all imply that racism did not pervade most dimensions of post-Abolition Brazil or the decades that followed. What it does is to reject simplistic black-on-white explanations on the functioning of labor markets, as well as to revive the understanding that contractual relations, formal and informal, are always multidimensional and complex.

This paper explored only a few facets of those relations. Albeit exhaustively covering *Ibicaba*'s agricultural labor force from 1888 to 1958, it used only a fraction of the material

[59] An important exception is Holloway (1978).

available at the *Ibicaba Collection*. Numerous other angles, based on that collection itself, remain unexplored (Witzel de Souza and Gardenal 2026).[60] Most prominently, we know close to nothing, in quantitative terms, about the functioning of *colonato* mechanisms regulating the production of subsistence goods, even though contemporaneous observers and qualitative research have argued that this was a major incentive for labor turnover and mobility (Dean 1976; Holloway 1980; Andrews 1988, p. 514; Bassanezi 2019). Combined with models of time usage, data on foodstuff sold by *colonos* (available in *Daybooks* and *General Accounting Ledgers*) and their consumption patterns (available in *Laborers' Current Accounts* and *Stock Registries*) can shed completely new light onto that question. Similarly, we still know little, in quantitative terms, about non-agricultural employment in plantations. Even basic metrics on the length of the working year, including in non-agricultural tasks, remains unknown.

The general interest of this paper goes beyond that of a case study of one single plantation with a *suis generis* history. Rather, it is embedded in *Ibicaba*'s representativeness of São Paulo's Old West on the eve of Abolition, when the region was the most booming coffee economy in the world. Although São Paulo's Old West was historically central to Brazil's economic positioning in the nineteenth and early twentieth centuries, it was not representative of the country's coffee economy as a whole (Fragoso 1983, pp. 137-144; Stein 1986); and even less so of the economics of slavery in other crops and regions. The very demographic history of São Paulo's Old West, with the historical inflows of indentured laborers started in *Ibicaba* itself, was distinct even from that of other regions within the province (Holloway 1978, p. 188; Colistete 2024, p. 23).

Collecting new microdata from other rural units is therefore crucial to start piecing together a more representative picture of Brazil's agricultural and labor history. Finding and preserving physical records similar to those of *Ibicaba* – mostly private and endangered – is the necessary condition for such an endeavor. The *Accounting for the Countryside Project* is a first step in that direction. Funded by MEAP-UCLA, that project is currently (2026) working to extend the efforts of material preservation and digitalization commenced in *Ibicaba* to five other collections related to São Paulo's coffee economy, as well as into Rio de Janeiro. New collections include those of *Santa Gertrudes*, a plantation neighboring *Ibicaba* but with a distinctive migratory and business history; *Casa Feltrin*, a smallholding with a storehouse founded by Italian immigrants in Itaqueri da Serra County; the old government-sponsored settlement of *Cascalho*, in Cordeirópolis County; and the old forestry station from the railroad company *Cia. Paulista de Estradas de Ferro* (later *FEPASA*), housed at *Floresta Estadual Edmundo Navarro de Andrade* (Rio Claro - SP County).

[60] The same questions addressed in this paper can be replicated with other ledgers, *e.g.* laborers' *Checking Accounts*. Beyond earnings, prices, number of groves cultivated, and volume harvested, those ledgers contain data on consumption and loans. The sources chosen for this paper had a lower cost in terms of close-reading transcriptions, but advances in automated transcriptions are promising for replication exercises of that type.

VII. References

Absell, C. D. (2020). "The rise of coffee in the Brazilian South-East: Tariffs and foreign market potential, 1827–40". *The Economic History Review*, Vol. 73 - No. 4.

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

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.

Bercovich, A. M. (1988). "Considerações sobre a fecundidade da população negra no Brasil". *Revista Brasileira de Estudos de População*, Vol. 6 - No. 1.

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 Scarano, J. M. L. (1971). "A mão-de-obra escrava e estrangeira numa região de economia cafeeira". *Anais do VI Simpósio Nacional dos Professores Universitários de História,* Vol. I - XLIII (published in 1973).

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

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

Bucciferro, J. R. (2017). "Racial inequality in Brazil from independence to the present". In Bértola, L.; and Williamson, J. (eds.): *Has Latin American inequality changed direction? Looking over the long run*. Cham: Springer Open.

Bucciferro, J. R. (2021). "A lucrative end: Abolition, immigration, and the new occupational hierarchy in southeast Brazil". *Cliometrica*, Vol. 15.

Canaday, N.; and Jeremsky, M. (2012). "Legacy, location, and labor: Accounting for racial differences in postbellum cotton production". *Explorations in Economic History*, Vol. 49.

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 de Mello, P.; and Slenes, R. W. (1980). "Análise econômica da escravidão no Brasil". In Neuhaus, P. (Coord.): *Economia brasileira: uma visão histórica.* Rio de Janeiro: Editora Campus Ltda.

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).

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.

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).

Clevelario Jr., J. (1997). "A participação da imigração na formação da população brasileira". *Revista Brasileira de Estudos de População*, Vol. 14 - No. 1/2.

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).

Collins, W. J.; Holtkamp, N. C.; and Wanamaker, M. H. (2024). "Black Americans' landholdings and economic mobility after emancipation: Evidence from the census of agriculture and linked records". *The Journal of Economic History*, Vol. 84 - No. 4.

Collins, W. J.; and Holtkamp, N. C. (2026). "The geography of opportunity after the Civil War: Black and white Americans' intra- and intergenerational mobility into property ownership". *Explorations in Economic History*, Vol. 100.

Costa, C. E. C. da (2015). "Revisitando 'Família e Transição': Família, terra e mobilidade social no pós-Abolição: Rio de Janeiro (1888-1940)". *Revista Brasileira de História*, Vol. 35 - No. 69.

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. (1979). *O Problema do café no Brasil*. Rio de Janeiro: Editora da Fundação Getúlio Vargas.

Derenoncourt, E.; Kim, C. H.; Kuhn, M.; and Schularick, M. (2024). "Wealth of two nations: The U.S. racial wealth gap, 1860–2020". *The Quarterly Journal of Economics*, Vol. 139 - No. 2.

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

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.

Fragoso, J. L. R. (1983). *Sistemas agrários em Paraíba do Sul (1850-1920) – um estudo de relações não-capitalistas de produção.* Rio de Janeiro: Universidade Federal do Rio de Janeiro (M.A. thesis).

Fragoso, J. L. R.; and Florentino, M. G. (1987). "Marcelino, filho de Inocência Crioula, neto de Joana Cabinda: Um estudo sobre famílias escravas em Paraíba do Sul (1835-1872)". *Estudos Econômicos*, Vol. 17 - No. 2.

Graham, D. H. (1973). "Migração estrangeira e a questão da oferta de mão-de-obra no crescimento econômico brasileiro - 1880-1930". *Estudos Econômicos*, Vol. 3 - No. 1.

Graham, R. (1981). "Slavery and economic development: Brazil and the United States South in the nineteenth century." *Comparative Studies in Society and History*, Vol. 23 - No. 4.

Heflinger Jr., J. E. (2007). *Ibicaba: O berço da imigração européia de cunho particular*. Limeira: 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.
Heitmüller, A. (2005). “A note on decompositions in fixed effects models in the presence of time-invariant characteristics”. *IZA DP* - No. 1886. Available here (last accessed ast accessed on April 28, 2026).
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.
James, P. E. (1932). “The coffee lands of southeastern Brazil”. *Geographical Review*, Vol. 22 - No. 2.
Jann, B. (2008). “The Blinder-Oaxaca decomposition for linear regression models”. *The Stata Journal*, Vol. 8 - No. 4.
Jones, S.; and Gibbon, P. (2025). “Firm profitability and forced wage labour in Portuguese Africa: Evidence from the Sena Sugar Estates, 1920–74”. *The Economic History Review*, Vol. 78 - No. 1.
Karge, E.; and Wray, A. (2024). “The Black-White lifetime earnings gap”. *Explorations in Economic History*, Vol. 94.
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.
Lamounier, M. L. (2007). “Agricultura e mercado de trabalho”. *Estudos Econômicos*, Vol. 37 - No. 2.
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.
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).
Monbeig, P. (1984). *Pioneiros e fazendeiros de São Paulo*. São Paulo: Editora Hucitec / Editora Polis (2nd edition).
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.
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.
Motta, J. F. (1988). “Família escrava: Uma incursão pela historiografia”. *História: Questões & Debates*, Vol. 9 - No. 16.
Motta, J. F. (2006). “Escravos daqui, dali e de mais além: O tráfico interno de cativos em Constituição (Piracicaba), 1861-1880”. *Revista Brasileira de História*, Vol. 26 - No. 52.
Olmstead, A. L.; and Rhode, P. W. (2018). “Cotton, slavery, and the new history of capitalism”. *Explorations in Economic History*, Vol. 67.

Palma, N. (2019). “Historical account books as a source for quantitative history”. *Maddison-Project Working Paper WP-13*.
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.
Reber, V. B. (1985). “Archival sources for Latin American Business History”. *The Business History Review*, Vol. 59 - No. 4.
Rhode P. (2020). “Review of ‘*Accounting for slavery: Masters and management*’. By Caitlin Rosenthal. Cambridge, MA: Harvard University Press, 2018. Pp. 320”. *The Journal of Economic History*, Vol. 80 - No. 1.
Rios, A. M. L. (1990). “Famílias e transição. Famílias negras em Paraíba do Sul, 1872-1920”. *Revista Brasileira de Estudos de População*, Vol. 7 - No. 2.
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).
Rosenthal, C; (2018). *Accounting for slavery: Masters and management*. Cambridge, MA: Harvard University Press.
Salles Ricardo; and Muaze, M. (2022). “Agrarian empires, plantation communities, and slave families in a nineteenth-century Brazilian coffee zone”. In Fischer, B.; and Grinberg, K. (Eds.): *The boundaries of freedom. Slavery, abolition, and the making of modern Brazil.* Cambridge, MA: Cambridge University Press.
Sánchez-Alonso, B. S. (2007). “The other Europeans: Immigration into Latin America and the international labour market (1870–1930)”. *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.
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.
Simonsen, R. C. (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*, 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. (1986). *Vassouras: A Brazilian coffee county, 1850–1900*. Princeton: Princeton University Press (2nd edition).
Stolcke, V. (1986). *Cafeicultura. Homens, mulheres e capital*. São Paulo: Editora Brasiliense.
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.
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.

Theodoridis, D.; Rönbäck, K.; and Galli, S. (2025). “The failed promise of freedom - Emancipation and wealth inequality in the Caribbean”. *The Economic History Review*, Vol. 78 - No. 3.

Versiani, F, R. (1994). “Brazilian slavery: Toward an economic analysis”. *Revista Brasileira de Economia*, Vol. 48 - No. 4.

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 (2nd edition).

Witzel de Souza, B. G. (2018). “Immigration and the path dependence of education: The case of German-speakers in São Paulo, Brazil” (2018). *The Economic History Review*, Vol. 71 - No. 2.

Witzel de Souza, B. G. (2023). “Like the swing of the pendulum: The history of government-sponsored rural settlements (1820-1920)”. *Economic History of Developing Regions*, Vol. 38 - No. 3.

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)”. *The Economic History Review*, Vol. 77 - No. 2.

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).

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

Online Appendix I
Data from historical accounting and managerial ledgers

Contents

I. Historical accounting & Data-generating process: Some additional notes

This section presents additional evidence on how historical accounting was conducted in practice, and its implications for the data-generating process of the sources themselves.

Particular attention is paid to the historical methods used for the compilation of *Registries*. As argued in the main text, such ledgers were population counts, compiled as an annual survey-like listing of household members by household heads. Their compilers, in a function similar to that of enumerators, were sub-directors or assistant managers who responded back to *Ibicaba*'s general managers (for further evidence on the verticalization of managerial practices in that plantation, *cf.* Witzel de Souza and Santin Gardenal 2026).

Ledgers #0098, #0099, #0100, and #0101 provide the most compelling evidence on how information was historically compiled for the *Registries*. Commenced in the 1936/7, these *Registries* were compiled in booklets of easy handling; ledger #0101 was even written directly onto an official laborer's booklet (*caderneta*), its fifteen first pages containing printed labor legislation from the early twentieth century (DIs 03-10). Ledgers #0099 and #0100 contain data from *Colônia Grande* and *Pão de Lot* (overwritten to read *Grama Larga*) colonies. Ledgers #0098 and #0101, in turn, refer to *Iracema* and *São Francisco & Ipiranga* farms. 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 suggest that their compilers had only a basic level of education, with frequent phonetic spellings of names and surnames.[1]

The final dataset has revealed that all demographic data contained in ledger #0100 overlap perfectly with those of ledger #0083 for the 1936/7 and 1941/2 agricultural seasons. This is clear evidence that the former was the draft produced in the colonies and then transferred to *Ibicaba*'s general managers and accountants, who used it to update the latter. The same holds for ledger #0099, whose contents, in addition, overlap perfectly with the data from ledger #0083 for the 1937/8 and 1942/3 agricultural seasons. Moreover, two nominal lists were annexed to ledgers #0099 (DI 65-5) and #0101 (DI 28-9). The former includes the number of coffee groves, annual earnings, and bimonthly instalments paid to fourteen household heads (DI 65-5) in 1936/7;[2] the latter contains a demographic summary of the residing labor force in *São Francisco* and *Ipiranga* farms in 1937.

It is also worth noting that practically all ledgers in the *Ibicaba Collection* contain numerous annotations whose contents go beyond the focus of this paper, but that might nonetheless be of interest to economic and business historians, in general. Some of those annotations were methodically documented by historical accountants themselves, or are as intuitive as "L[ançado]" notes for accounting postings and correction checks (*e.g.* ledgers #0025, DI 50

[1] This issue was very prominent in the 1940/1 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" (*e.g.* "Sirva" instead of the correct "Silva") and corruptions of non-Iberian surnames (*e.g.* "Bergue" for "Berg", "Manhais" for "Magnaes", and "Croce" for "Kloss").

[2] The source informs that 34,355 coffee bushes had been uprooted in that year, a result of federal policies to curtail the supply of coffee to international markets during the Great Depression (*cf.* Section II of the main text).

& 332 - details here & here; #0040, DI 190 - detail here; #0043, DI 08 - detail here; #0062, DI 32 - detail here). Others, however, have lost their meanings to the modern reader (*e.g.* ##0029, DI 13 - detail here; #0043, DI 06 - detail here), including numbers probably used as keys to match accounts within and between ledgers (*e.g.* ##0029, DI 103 - detail here; or the listings of *São Francisco* & *Ipiranga* referred to above in ledger #0099, DI 65 - detail here). I did not transcribe these symbols and keys, as the marching of households in this paper did not require cross-ledger referencing beyond the names and demographic characteristics of household members. I am confident, however, that the application of LLMs to them will shed new light onto the complex managerial functioning and accounting practices of historical plantations.

Some annotations reveal the complexity of contractual relations within the plantation, which went beyond the annual (re-)negotiations of *colonato* contracts. As mentioned in the main text, some accounting entries include notes on private arrangements made between *colonos* themselves.[3] Others registered special agreements they had negotiated with *Ibicaba*'s managers. As an illustration, in 1924/5, Jorge Küller cultivated 6,678 coffee groves as sharecropper in *Pinheiro* colony, where other households worked as usual *colonos* (#0062, DI 128). Similarly, by the 1920s, various households started cultivating coffee and subsistence crops "A meia" [SIC], *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 (*e.g.* #0083, DIs 146-7 - details here & here). During the Great Depression, this was likely a strategy to keep the labor force on the plantation while retaining full landownership of all its lands, instead of parceling out and selling plots of land to laborers.

II. *Registries*: Transcription, corrections & data wrangling

ii.i. Demographic data

The most readily apparent inconsistency found in the complete household series compiled from the *Registries* concerns some members' first names. Some such errors were obvious misspellings, poor translations of foreign first names, diminutives etc. of no consequence. In other cases, however, different first names were used to refer to the same person, otherwise identifiable by their demographic characteristics. In those cases, all available information in the household series – including sex, age, civil status, and nationality – was used to decide whether entries referred to the same individual or not.[4] Some errors further implied the attribution of different sexes for the same observation, given that most first names are declined by gender in Portuguese.

[3] An interesting example is that of a crop-sharing arrangement between two *colonos* in 1914: "N.B. Fica creditado O Cafe que João Ciavollela colhem a metade em c/ de José Giovanelle partes eguaes". This note implied a 50%-sharing of yields and earnings between these two households, which was credited by *Ibicaba*'s manager in the account of Mr. José Giovanelle (Ledger #0040, DI 106).
[4] These cases are thoroughly documented in the transcription and matching files.

Besides inconsistencies in naming, the series also had some mismatches in recorded sex, especially of infants and toddlers. The same correction procedure was applied to these discrepancies, with the most likely names and sexes attributed according to the complete household series.[5] Finally, a few household members went missing in the enumerations of specific intervals, *i.e.* they were listed in years $(t \pm i)$, but not in t. For that reason, we know that they did not pass away, outmigrate, or marry out of the household; however, since we cannot positively determine what happened to them in the corresponding interval, I did not impute these missing individuals into the household series when they were absent.

A common inconsistency in the initially transcribed dataset concerns the ages recorded in the sources. Changes in the age ordering of individual members were common between years – reinforcing the hypothesis that the demographic data were collected through a survey-like enumeration conducted with household heads, who reported the age of each household member. Except for those cases in which errors would have had serious implications for the household's structure (*e.g.* swapping siblings for couples when ordering household members by age), I kept these inconsistencies as they appeared in the sources, under the premise that such mistakes 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.[6]

Data on marital statuses were much more consistent in the original sources, with only sparse swaps between couples and other household members, which were easily corrected using their age series. More curious were listings of new household members – usually single, young adult women – for whom little additional information had been recorded in the sources. These were most likely brides who had joined their new households, perhaps having newly arrived as immigrants themselves. I attributed the status of "single" to these women if no explicit marital status had been registered in the sources, except if the woman: (i) had been recorded as married in $(t - i)$, *i.e.* assuming that no divorce had taken place, especially if both spouses remained in the household; (ii) was recorded as married in $(t + i)$ and the couple had a child at least as old as that interval, *i.e.* assuming no illegitimate births; or (iii) her marital status could be inferred from the household's structure, especially in mononuclear households.

[5] Difficulties in matching households increase with the time dimension of the panel and number of household members. The longer the series and the larger the household, the more likely it becomes that first names were 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 series was pooled across all *Registries*.

[6] 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.

ii.ii. *Estacas* & number of coffee groves cultivated

The multiple layers of information stemming from corrections, glosses, annotations, and overwritings made by contemporaneous managers and accountants in the *Registries* are nowhere more prominent than in the data on coffee cultivation. Because households could renegotiate yearly the number of coffee groves they cultivated, all ledgers registering 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 earnings 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, despite 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 cannot determine when exactly the accountants altered each individual entry. In fact, it could well be 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 the actual new account in $(t + 1)$.

The baselines control for quantities and *estacas* that are clearly traceable to individual entries, specific corrections, and totals reported in the sources. The corresponding variables are the closest to the sources, as they take into account all crossed-out and added-up *estacas* that most closely match the sums reported in the original sources. Those variables also include extra information contained in a few summary tables available in the *Registries*.[7]

The robustness checks, in turn, take into account other possible, but less likely scenarios, according to the following definitions:

1. Variable *tree_tot_estaca* provides an upper-bound estimate: it sums the coffee groves from all *estacas* registered in an account, irrespective of whether they were ever crossed out.[8]
2. Variable *tree_tot_estaca_check* sums the coffee groves of all *estacas* ever registered in an account but excludes those crossed out in the account, irrespective of the sums reported in the sources.
3. Variable *tree_tot_check02* mimics the baseline but excludes *all* groves from crossed-out *estacas*, again irrespective of the sums reported in the sources.[9]

[7] These include extra information from ledger BPML #0004 on the number of coffee groves cultivated in 1890.
[8] In principle, this variable would exclude ledger BPML #0004, as the latter reports only the sum of coffee groves cultivated. However, for comparability with the baseline, that data have been added to *tree_tot_estaca*.
[9] Two additional indicators are used as conditionals to exclude certain observations in the robustness checks. The first indicates whether total earnings from coffee cultivation from household *h* had been crossed out in year *t*. The second indicates whether the entire account of household *h* had been crossed out in year *t*.

ii.iii. HOUSING

Despite their invaluable nature as proxies for basic housing characteristics, the variables on housing are the most inconsistently defined and recorded in the *Registries*.

First, house numbers were not always consistently recorded over time, making it impossible to determine whether changes in their numbering for the same colony capture households moving between houses, or simply resulted from accountants relabeling the number of residences therein. Ledger #0083 (DI 270-1) contains a unique attachment that helps elucidate these accounting practices, namely a map with the distribution of houses, probably in a section of *Colônia Grande.* The map indicates the location of the colony's chapel and school, the house of the colony's inspector, and squares probably referring to common washtubs and water tanks. The blocks of residences contain an irregular number of houses, suggesting different sizes between them. Noticeably, only the front houses are numbered in a growing order, from 99 to 108; all the others are numbered irregularly.

In addition, accountants were not always clear about the nature of the amenities they recorded in each year. This problem was salient if those features changed for the same house in consecutive years. While such changes could indicate actual structural renovations in the houses (*e.g.* installing new windows, building up or demolishing rooms etc.), they could also reflect changes in conventions only. For example, ledger #0062 records the number of *quartos* in 1921-1924 (with gaps), but uses the label *commodos* for 1929-1931 (with gaps). Although colloquial synonyms, more strictly speaking these words could refer to different types of rooms in a house: the former denoting bedrooms, only, and the latter, any type of room. Similarly, it is not always clear whether the *Registries* recorded the number of doors and windows in a house or the number of leaves in those doors and windows, or whether locks in the front doors had been counted among all "doors" (if not, the latter referred only to internal doors and could, therefore, proxy for the number of rooms in the house).

I corrected these inconsistencies in the series according to the following criteria. First – and analogously to the (many fewer) inconsistencies in the demographic data – if more than one entry referred to the same house in a given year, only the entries most consistent with the rest of the series were retained in the dataset. Second, conditional on the previous criterion, entries containing more information were preferred. Third, the series also allowed interpolating some of the housing data between years. However, colonies of residence and house numbers were interpolated only if the gap between the actually known datapoints was one year at most, as otherwise households would have had enough time to move in and out more than once. Moreover, data on housing characteristics were interpolated only if the house number was actually known, or if there was clear evidence that the household resided in the same house in that interval.[10] In these cases, if a missing value in year t was neighbored by

[10] It was possible to derive some information on colonies and house numbers from the complete household series. 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.

conflicting information in years ($t \pm 1$), the smallest number of amenities was used. Finally, data on locks and electric lightbulb sockets were never imputed retroactively.

ii.iv. Wrangling the data from individual- to household-level

Harvest ledgers were originally compiled at the household level. Conversely, except for information on housing, the *Registries* compiled the data at the level of individual household members and of individual *estacas*. This implied the necessity of wrangling all variables related to demographics and *estacas* into household-level aggregates.

By definition, demographic data were collected at the level of household members. Wrangling them from individual- to household-level required only the compilation of count variables or measures of central tendency and dispersion that aggregated individual members to their own households.

The procedures to wrangle data from individual *estacas* (*i.e.* their identifiers and number of coffee groves in them) into household-level data were somewhat more complex. Such information was originally registered in the sources at the level of individual *estacas* (*cf.* entries (E) and (F) of Figure 1 - Panel B, in the main text). The purpose of these entries was to determine how many coffee groves each household cultivated in each *estaca.* The total number of cultivated coffee groves per household can be obtained by summing the values of each individual *estaca*. However, the number attributed to each individual *estaca* is simply an identifier, without substantive meaning; in addition, some *estaca* identifiers were strings that I had to convert into numeric labels, which, again, had no substantive quantitative meaning.

One possibility to deal with this data structure was to wrangle each *estaca* identifier from the long (per household over time) to the wide (per *estaca*, for every household over time) format 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.[11] The more parsimonious option was to collapse *estaca* identifiers to the level of households and to summarize them with measures of central tendency and dispersion. For every household in each year, the mean, median, standard deviation, mean deviation from the mean (MDEV), and median absolute deviation (MAD) were computed for the *estaca* identifiers. This approach has the drawback of compressing information that is historically much richer. Nonetheless, this option is statistically efficient and, combined with colony identifiers, provides solid controls for the agricultural capital available to each household.

[11] 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*.

III. Assessing data quality: Auditing errors in historic accounting

Contemporaneous observers and historical research have long recognized the importance of good accounting practices for enhancing economic decision-making and increasing productivity in agricultural businesses (*e.g.* Carneiro and Carneiro 1933; Brandão 1945; Lampe and Sharp 2019; Sangster 2025).[12] As Section III of the main text and this Appendix make clear, however, the ledgers from the *Ibicaba Collection* are far from homogeneous in their managerial and accounting qualities, which may raise concerns about the quality of the data in general.

The Levy Brothers undoubtedly had the skills required to properly manage a plantation by the 1890s. Having immigrated to Brazil in 1857 as indentured sharecroppers, José and Simão Levy had a comparatively high level of human capital and built a solid entrepreneurial career: in the 1860s, the Levys managed the storehouse of *Ibicaba*'s *Colônia Grande*; in the 1870s, they built a solid reputation as large merchants; and in the 1880s, they consolidated their financial position by founding a local banking house in the town of Limeira (Freitas Levy ([1993] 2023; Heflinger 2021; Tamiazzo 2021; Witzel de Souza and Santin Gardenal 2026).

Nonetheless, two objects in the *Ibicaba Collection* reveal their remarkable learning curve as landowners and plantation managers. Ledger #0001, the first managerial record of *Ibicaba* written under the Levys' ownership, compressed two types of account (the plantation's *Current Accounts* and laborers' *Payrolls*) into a single document. This mixing of different accounts into one book probably resulted from inexperience, before practice proved the benefits of separating these records into different ledgers. Even more telling is ledger #0005, the first *Harvest ledger* produced under their administration, in 1890. The authors of that ledger (the Levy Brothers themselves or *Ibicaba*'s general managers) first adopted a *Daybook* notation to register the harvests. This implied that, every single day, the accountant wrote down the full name of the household head and the corresponding amount of coffee harvested that day (*e.g.* DI 05 - detail here). This extremely laborious procedure changed in 1891, when the accountant adopted a *Current Accounts*' notation. This implied that, under one single heading for each household, the accountant noted down all days worked and the volume of coffee harvested by that household only, *i.e.* the same format as that of Figure 1 in the main text (*e.g.* DI 47 - detail here). The improvement in efficiency and transparency is staggering and can be grasped by the number of rows whose transcriptions are necessary to compute total annual harvests per household:[13] under the 1890 *Daybook*'s notation, 2,612 individual rows had to be transcribed in order to compute the volume harvested by 110 households in that year; under the 1891-1895 *Current Accounts*' notation, only 131 individual rows had to

[12] References not listed in the main text: (1) Brandão, E. D. (1945). *Administração de fazenda. Contabilidade agrícola.* Rio de Janeiro (RJ): Serviço de Documentação, Ministério da Agricultura. (2) Carneiro, J.; and Carneiro, E. (1933). *Tratado de contabilidade. Vol. III - Contabilidade rural.* Rio de Janeiro: Calvino Filho Editor (2nd ed.) (3) 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. (4) Sangster, A., (2025). "The emergence of double entry bookkeeping". *The Economic History Review*, Vol. 78 - No. 2.

[13] In 1892, the accountant started again using the *Daybook*'s notation for harvest registries (#0005, DI 78 - detail here), but quickly reverted, on the ledger's next page, to the current accounts' more efficient notation.

be transcribed, on average, to compute the same information for an average of 108 households per year.[14]

Accounting also played a central role in labor disputes in an economy characterized by low monetization and high reliance on credit (Saraiva and Freire 2016; Fornereto Mariano *et al.* 2022; Monti and Oliveira 2025). It is worth recalling that a laborers' strike, the so-called *Sharecropper's Rebellion* took place in *Ibicaba* in 1856/7. A core complaint of those laborers related to alleged fraudulent accounting. At least two of their 18 complaints (§§ 9 and 15) related directly to lack of transparency in the plantation's accounting practices (Davatz [1858] 1941, pp. 249-55; *cf.* also Vergueiro 1874, p. 6). Similar complaints were continuously raised across plantations in São Paulo throughout the nineteenth- and early-twentieth centuries.

To the best of my knowledge, however, no study has ever attempted to check the validity of laborers' allegations about fraudulent accounting, in *Ibicaba* or elsewhere. Such a study would be as relevant as it would be hazardous. Its conclusions would be strictly circumscribed to the ledgers studied without any external validity, even for *Ibicaba* itself. There is no reason to extrapolate to the *Vergueiros*' era the same types of error committed by the Levys (or vice-versa); moreover, improvements in accounting practices might have been a *consequence* of previous labor strikes. Also, as revealed by the process of data transcription, different types of error could have been committed at different stages of agricultural production, with liability falling upon different agents. Errors could have been committed by sub-directors who incorrectly recorded daily production; by general directors who failed to spot these mistakes as the agricultural season progressed; and by accountants who mistakenly updated entries within and between ledgers.[15] Inferring *purpose* from such errors would be even more daunting.[16]

A much simpler and more restricted question of direct interest to this paper is whether accounting errors introduced biases in the earnings compiled in this particular dataset. To address this question, I calculated the difference between earnings per pruning recorded in the *Registries* and the income that households should have earned if one assumes that individual prices and quantities originally registered in the sources were correct. I computed this error only for the lean season, since prunings represented almost ¾ of the average annual earnings of *colono* households (*cf.* Section IV), and because auditing the harvest season would have required the transcription of all individual daily entries on coffee picking.

[14] (130, 225, 95, 106, and 99) rows for (129, 115, 95, 104, and 98) households, respectively.
[15] The sheer scale of a plantation posed challenges to accounting. Slips of paper annexed to ledger #0040 (DIs 301-2 ff. - detail here & here), probably from the late 1910s, exemplify the hazardous task of computing total yields. Computations were obviously made by hand and checked multiple times (*cf.* ticks in different colors).
[16] An indirect solution to this problem would be to compare the frequency and magnitude of errors in laborers' accounts to those found in purely managerial ledgers. Analogously, it would be useful to show the stages of agricultural production in which errors were most frequent and quantitatively relevant. Were they caused mainly by sub-directors, whose levels of human capital were expectedly lower? Or by general managers? If the latter, were errors in physical output more common (*i.e.* against good managerial practices, in general, and of direct interest to the landowners' profits), or were the problems mainly with laborers' earnings?

As reported in the main text and recalled in this paragraph, the mean overall accounting error computed for the corresponding 2,905 entries equaled 0.066 *milréis* per pruning, *i.e.* minorly positive in favor of the plantation;[17] the median, however, was actually zero.[18] Excluding approximation errors of 0.001 *milréis* per pruning, the mean positive errors equaled 1.125 *milréis* (median: 0.033 *milréis*).[19] The latter was, however, driven by an outlier of 120.694 *milréis* – an error too large to have gone unnoticed at the time.[20] There were also 58 negative errors, *i.e.* favoring laborers, with an absolute mean of 2.043 *milréis* (|median|: 0.300 *milréis*).

These figures do not point to any serious problem with accounting errors. In addition, Online Appendix I - Table 1 shows no evidence that such errors were correlated either with the sources from which the earnings in the lean season were derived (variable $Source$) or with different time trends (linear or quadratic), which aim to capture an eventual learning curve of the Levys' managerial era and *Ibicaba*'s late economic decline. These results increase our confidence in the quality of the data, suggesting that the contemporaneous data-generation processes were not subject to systematic biases against laborers.[21]

Independently of easing the econometric concerns of our own time, such errors certainly affected the daily lives of laborers in the past. As an illustration, in 1923/4 the accountants overwrote the decimal and centesimal units of Antonio Zambusi's earnings during the lean season (Ledger #0062, DI 58). If the individual prices and quantities are correct in his account, then the entry rounded up or misrecorded 0.008 *milréis* per pruning; since Zambusi's household pruned six times that year (as inferred from their annual earnings), that household received 0.048 *milréis* less than the contracted value, which corresponds to little more than two *vinténs* (0.020 *milréis*), the second smallest coin denomination of the Brazilian *milréis* at the time. Although that value is derisory and almost certainly resulted from a harmless rounding up of cents, such a difference might have dissatisfied the Zambusis had they ever noticed it. After all, as the legend of the *vintém* copper coin reminds us to this day: "A saved *vintém* is an earned *vintém*".[22]

[17] For reference, the 1904-1942 mean pooled price for cultivating one coffee grove was 0.025 *milréis*.
[18] The exercise obviously depends on the assumption that the sources correctly reported *prices* and *quantities*.
[19] 892 observations had errors larger than zero. However, 617 were smaller than 0.001 *milréis*, *i.e.* approximation errors in millesimal units that neither accountants nor laborers would have been able to compute.
[20] Account of Benedicto Cunha, 1929-1930 (#0062, DI 205). Strangely, the accountant copied to Cunha's entry the exact same earnings registered on that same page for another household (João Franco Silveira).
[21] 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. To eliminate any remaining concern that common shocks related to accounting practices could bias any estimate, the baseline specifications control for year-FEs.
[22] The rounding up of centesimals (for *milréis*) 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.

Online Appendix I - Table 1 – Accounting errors in earnings during the lean season

	Dep. Var.: Accounting errors in earnings 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
R2	0.000	0.000	0.000	0.000
Adj-R2	-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 (ledgers) from which the observation was transcribed. *Year* is the calendar year of the observation, and $Year^2$ its quadratic term.

Online Appendix II
Alternative classifications of Ethnolinguistic origins

The cliometric exercises in the main text used a classification of ethnolinguistic origins based on *jus sanguinis* criteria and on a list of surname-nationality pairs created solely from data in the *Ibicaba Collection*. Based on the origins of the households' heads, each household was classified as *Non-White*, *Italian*, *Germanic*, *Portuguese*, *Spaniard*, or *Eastern European*; the group *White Brazilian* was residual, *i.e.* comprising all households that did not fall into any of the previous categories.

This appendix proposes alternative classifications of ethnolinguistic origins that increasingly augment the likelihood of households being classified as *Non-Whites* or of European origins: $Origins\ imposed_h$, $One\ drop_h$, and $One\ drop\ imposed_h$.

These classifications use extended surname-nationality pairs for surnames not listed in the *Ibicaba Collection* and categorize entire households as *Non-Whites* or as descending from the culturally farthest European origins if that household ever had a member who was Non-White or whose ancestry was traceable to those European nationalities. Reassuringly, these different categorizations do not alter this paper's empirical conclusions (*cf.* Online Appendix V).

$Origins\ imposed_h$ builds upon an *ad hoc* extended list of surname-nationality pairs that classify surnames not listed in the *Ibicaba Collection* as *Italian* or *Germanic* in origin, and attributes *Portuguese* and *Spaniard* origins to certain identifiable Iberian surnames (*cf.* the lists below). While there is no reason to doubt that the Brazilian *Andrietas* and *Benettis* or *Helds* and *Scherers* had Italian and German-speaking forefathers, we cannot determine how far removed these ancestors actually were using only the data from the *Ibicaba Collection*. The indicators of $Origins\ imposed_h$ accommodate these cases and classify them in the groups of foreigners or of Non-Whites, depending on the traceable origins of surnames (*cf.* Levy 1974 and Clevelario Jr. 1997, pp. 51-4).

$Onedrop_h$, in turn, classifies the entire household as *Non-White* or of European origins with the farthest cultural distance to Brazilians if that household ever had a member who was Non-White or whose ancestry was traceable to any of the listed European nationalities. In other words, if household h ever had a 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*).

The nomenclature "One drop" is used here to evoke the extremism of the American system of racial discrimination prior to the Civil Rights Movement.[1] That principle of racial classification is at odds with perceptions of ethnolinguistic origins in Brazil, but it can nonetheless be useful for extreme-bound analyses, and it reflects the Brazilian political counterpart to "state-imposed racial controls used, for example, in segregationist South Africa or the U.S. South during the same period" (Andrews 1988, pp. 492-3).[2]

Finally, $Onedrop\ imposed_h$ is based upon the same principle of $Onedrop_h$ and uses the extended list of surnames created for $Origins\ imposed_h$.

Online Appendix II - Table 1 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%. The increase in the *Germanic* group is explained by their long migratory history to Brazil, in general, and to the region, in particular, while the increase in the number of *Non-Whites* can be explained by intermarriages between Whites and Non-Whites (*cf.* Colistete 2024, pp. 22, 27-9). The proportions of re-categorized *Italians* (13.19%) and *Portuguese* (12.87%) are not negligible either, which is indicative of Brazilian-born children of Italians and Portuguese whose parents had not worked in *Ibicaba*.[3]

[1] *cf.* Monsma (2010, footnote 3) for a potentially similar racial categorization among planters in São Carlos County at the beginning of the twentieth century. A critical sociological reflection on the matter can be found in Monsma's footnote 12 and in Costa (2015). Finally, Colistete (2024) provides estimates of the Brazilian labor force in São Paulo's coffee economy prior to Abolition, discussing their potential multiethnic origins (pp. 10-11).

[2] Racial classifications in Brazil were further entangled with socioeconomic status, implying a fluidity of racial categorization shared with other Latin American countries (*cf.* Monsma 2010, pp. 512-4 for a summary of particular interest to the region considered in this paper). *Ibicaba*'s data include many examples of uncertainties related to such "racial" classifications in the past. As an illustration, 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 were related by blood, with the skin-colors of their partners attributed to them upon marriage. On a similar matter, *cf.* the historical interviews with Black persons by Callari and Baptista (1989, especially p. 47). Reference: 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.

[3] 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).

Online Appendix II - Table 1 – Number of households by ethnolinguistic origins, according to different categorization criteria

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

	White 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

	White Brz.	Non-Wht.	Ital.	Grm.	Port.	Span.	East. Erp.
White-Brz. = 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
Germanics = 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

Online Appendix III
Additional descriptive statistics: Demographics, Housing & Earnings

Contents

I. Mono- and Multinuclear households among *colonos* in *Ibicaba*

Online Appendix III - Table 1 tabulates the number of families per household with the pooled data (1888-1958, with gaps). Figures includes only those observations for whom the marital status was available to compute the number of families, defined as (#widows + (#married/2)).

Online Appendix III - Table 1 – Tabulation of the number of nuclear families in *Ibicaba*'s households, pooled sample (1888-1958, with gaps)

# Nuclear families	Freq.	Percent	Uniquely-ident. HH
0	38	1.31	33
0.5	115	3.97	100
1	1,990	68.69	648
1.5	45	1.55	28
2	542	18.71	162
2.5	5	0.17	5
3	109	3.76	40
3.5	7	0.24	4
4	35	1.21	6
4.5	2	0.07	2
5	5	0.17	3
6	4	0.14	2
Total	2,897	100.00	1,033

Notes: (1) The number of nuclear families is defined by the number of married and widowed household members as (#widows + (#married/2)). (2) "Freq." reports the number of observations in the pooled sample, while "Uniquely-identified HH" refers to uniquely-identified households with the corresponding number of nuclear families in them.

The overall sample mean of 1.3 families per household reported in Table 2 of the main text is attributable to the existence of large multinuclear households in the sample and to the fact that 57 uniquely-identified households had at least three nuclear families each. The majority of households residing in *Ibicaba* were, however, mononuclear (648 uniquely-identified households, corresponding to 1,990 observations).

The distribution of families further reveals a substantial number of households with one widowed person and with either two widows or an additional couple in them, *i.e.* the 1.5-2 nuclear households. This result is in line with the expectation of families being newly formed among the young labor force of *Ibicaba* – *i.e.* young couples who initially remained with their parents (usually the groom's family), before constituting their own households.

II. Additional characteristics of housing & colonies

ii.i. Housing characteristics: Figures per capita and changes over time

Online Appendix III - Tables 2-4 complement the descriptive statistics on housing presented in Table 2 of the main text with figures per capita, variations in the number of amenities between neighboring years with available data, and the maximum variation in those amenities across any two years for the same household. These are general metrics of living conditions and of their changes over time for uniquely identified households.

Figures per capita are well-aligned with the back-of-the-envelope calculations presented in the main text using the sample's overall mean for the number of household members. The average household shared one house with *ca.* five people; more than two individuals usually shared any single room. Also, the overall mean and median of rooms and doors per capita are very similar. These variables most likely identify the same information, *i.e.* the total number of rooms within a house available to the entire household.

Living conditions were generally poor. As Luiz Tonelotto, an ex-*colono* in *Ibicaba* in the 1950s-1960s, recalled: "[...] houses were extremely simple, right? [...] There were houses with two bedrooms, kitchen, and a veranda. There was no yard. And there were larger ones, with ten rooms. There were others with four. And the house had its floor tiled with bricks. But we also lived in earth-floored houses. Earth, red earth.[1] Earth, indeed. Once it rained, puddles formed in the house. But that was a long time ago, no?! 80 years ago. The house was poor, right? There was no electric light, no. It was an oil lamp. Kerosene lamp. I will tell you one thing about kerosene lamps: you wake up with your nose black with soot [laughter]. [...] One breathed that smoke, no? That was it. Times were different, oh Our Dear Lady!" (in Witzel de Souza and Santin Gardenal, 2021, pp. 358-9).

Most households never had a lightbulb socket in the period covered by this paper. Access to electricity remained a luxury identified in only three colonies as of 1926. Even among households with access to electricity, most had only a single lightbulb socket for the entire house. This statistical figure is made vivid by the memories of *Ibicaba*'s last operating manager, Luiz Bragotto de Castro: "In the beginning, there was only one lightbulb per house. The person chose the room where she wanted it and we installed it.[2] But we installed a long wire. Are you going to cook dinner? The woman brought the lightbulb to the kitchen. Did you finish your dinner and are you going to the living room? Then you brought the lightbulb to the living room. Later [...] electricity was installed in all houses. [...] But then the electricity meter was also installed. [...]. Because then there was already the famous radio, right? The ones that were plugged into the electric socket. And then came the television [...]. When the houses had just one lightbulb, at 06.30AM it was turned off. The supervisor went to the light pole and turned it off [...]" (Bragotto de Castro in Witzel de Souza and Santin Gardenal 2021, p. 349).[3]

---

[1] The region of *Ibicaba* is characterized by the distinctively red purple latosol soil, with an intense reddish color.
[2] Bragotto de Castro used here the feminine pronoun, most likely indicating that the housewife was the person responsible to determine where the lightbulb was to be installed.
[3] Felisberto Pott, another laborer, similarly recalls the lack of electric light during evening gatherings and childrens' plays (Witzel de Souza and Santin Gardenal 2021, p. 381).

Online Appendix III - Table 2 – Estimates of housing characteristics per capita, pooled sample (1888-1958, with gaps)

Variable	N	n	Mean	Median	S.D.	Min	Max
# houses p.c.	2,552	658	0.2249	0.1667	0.2166	0.0333	2
# houses (imp.) p.c.	2,657	658	0.2280	0.1667	0.2203	0.0333	2
# rooms p.c.	786	158	0.9333	0.8000	0.6094	0.2222	8
# windows p.c.	1,654	365	0.8205	0.7143	0.4426	0.1739	6
# doors p.c.	1,641	361	0.9122	0.8000	0.5040	0.1818	6
# locks p.c.	1,039	253	0.1733	0.1429	0.0883	0.0333	1
# lightbulb sockets p.c.	2,867	783	0.0245	0	0.0887	0	2
For houses with at least one lightbulb socket							
(# lightbulb sockets p.c.) > 0	342	120	0.2056	0.1667	0.1697	0.0667	2

Notes: (1) Figures per capita defined by the number of household members living in a residence, identified by their IDs. (2) "# houses (imp.)" include imputed data.

Online Appendix III - Table 3 – Distribution of lightbulb sockets per capita, by colony of residence, pooled sample (1888-1958, with gaps)

# lightbulb sockets p.c.	Colonies of residence			
	Grama Larga	Morro Alto	Iracema	Total
0.07-0.22	259	8	4	271
0.25-0.40	58	0	3	61
0.5	1	0	0	1
1	8	0	0	8
2	1	0	0	1
Total	327	8	7	342

Online Appendix III - Table 4 – Estimates of housing characteristics per capita, pooled sample (1888-1958, with gaps)

Variable	N	n	Mean	Median	S.D.	Min	Max
Panel A: Variation between neighboring years in number of characteristics p.c.							
Δ(# houses p.c.)	1,606	486	-0.0233	0	0.1790	-1.91	1.95
Δ(# houses (imp.) p.c.)	1,662	504	-0.0248	0	0.1879	-1.91	1.95
Δ(# rooms p.c.)	545	114	0.0127	0	0.1847	-2.0	1.17
Δ(# windows p.c.)	1,112	296	0.0126	0	0.2540	-1.0	4.67
Δ(# doors p.c.)	1,106	293	0.0082	0	0.2812	-1.20	5.58
Δ(# locks p.c.)	676	191	-0.0009	0	0.0373	-0.25	0.20
Δ(# lightbulb sockets p.c.)	1,768	594	0.0043	0	0.0512	-0.25	1.00
Panel B: Maximum variation in the number of characteristics p.c.							
maxΔ(# houses p.c.)	3,100	658	0.3210	0.1	0.4213	0	1.95
maxΔ(# houses (imp.) p.c.)	3,100	658	0.3596	0.1104	0.4537	0	1.95
maxΔ(# rooms p.c.)	1,543	158	0.5118	0.2064	1.118	0	7.50
maxΔ(# windows p.c.)	2,101	365	0.5008	0.3333	0.6598	0	5.00
maxΔ(# doors p.c.)	2,086	361	0.5742	0.4286	0.7665	0	5.62
maxΔ(# locks p.c.)	1,762	253	0.0644	0.0520	0.0696	0	0.67
maxΔ(# lightbulb sockets p.c.)	3,287	783	0.0386	0	0.1191	0	2.00

For most households, housing conditions did not change while they resided in *Ibicaba*. The median variation per capita between adjacent years is zero for all variables considered; even the maximum variation was never larger than unity for any of these variables. In extreme cases (minima and maxima), households could significantly improve their livelihoods in terms of housing amenities per capita (with the worsening of the minima being asymmetric to such improvements); however, these were the exceptions that prove the rule: while residing in *Ibicaba*, most households did not improve or worsen their housing conditions much.

ii.ii. The ethnolinguistic composition of colonies in *Ibicaba*

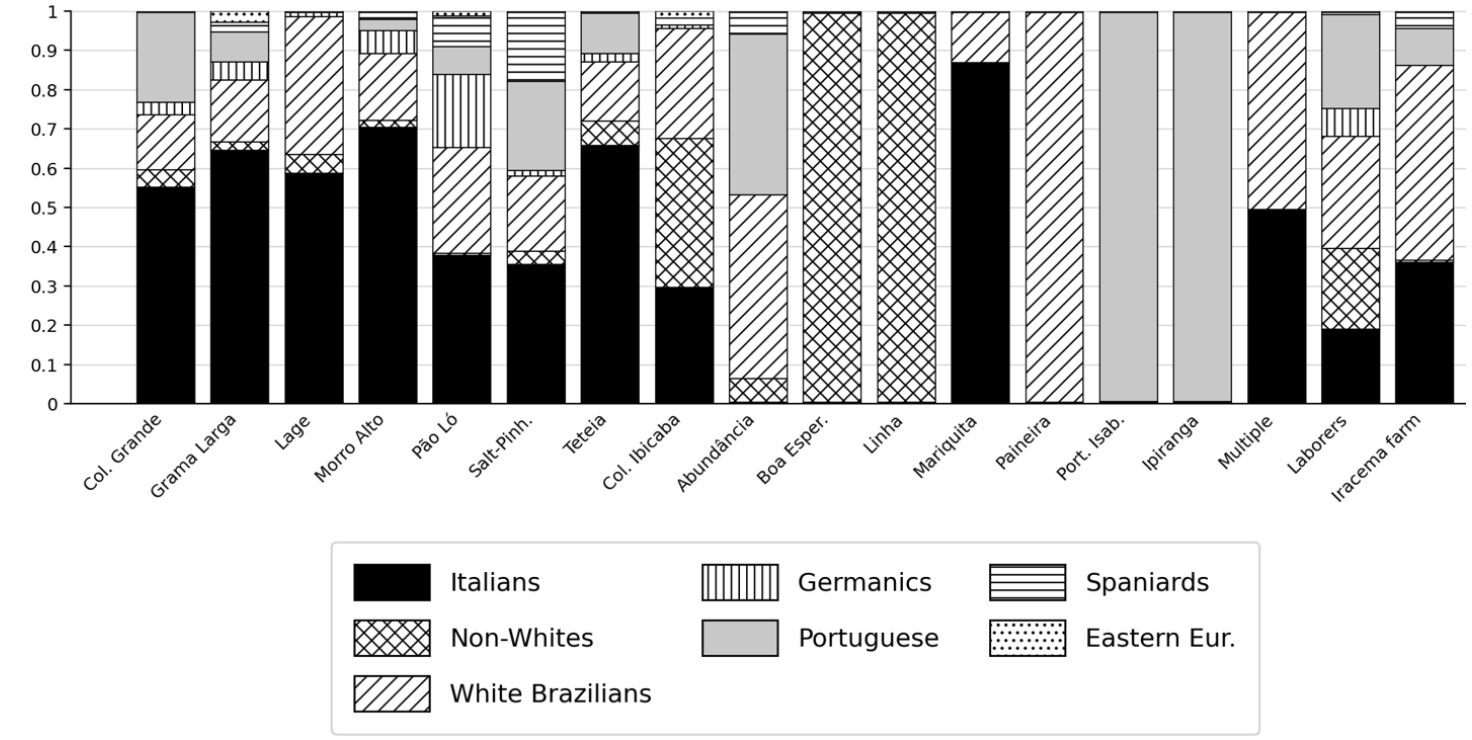
This section provides additional data on the distribution of ethnolinguistic groups across colonies in *Ibicaba* for the pooled sample. As discussed in Section IV of the main text, the hypothesis that colonies were formed on principles of racial segregation has no support from the data, even if some colonies had a majority of certain European nationalities, *Italians* in particular, and others had a majority – or exclusivity – of *Non-Whites*. In general, the distribution of residing households per colony followed the overall ethnolinguistic distribution of the sample as a whole, which, as argued in the main text, is representative of the European migratory inflows to São Paulo during the Age of Mass Migration.

Online Appendix III - Figure 1 plots the share of each ethnolinguistic group per colony. Online Appendix III - Table 5, in turn, lists the number of observations and of uniquely identified households per colony according to their ethnolinguistic origins. Both exclude observations with missing data on colonies of work/residence, implying a number of uniquely identified households smaller than the overall sample. Moreover, because households could move between colonies, the total number of observations reported here is larger than the overall sample. Therefore, these figures are comparable, but not equivalent to the ethnolinguistic composition of the dataset discussed in Sections III and IV of the main text.

The additional data confirm that colonies in *Ibicaba* were not homogeneous in their ethnolinguistic composition, even with regard to *Italians*.[4] The apparent homogeneity displayed by some colonies in Online Appendix III - Figure 1 is only an artifact, given that *Paineira* and *Portão Isabel* had only one uniquely identified household each with the required information to determine origins (a *White Brazilian* and a *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 (not reported, available from households' composition), were *Non-Whites* (*cf.* Section VI for a thorough study of that particular colony).

[4] Given their weight in gross inflows to São Paulo at the time, Italians and their descendants were nonetheless the most prevalent nationality in any colony where they resided in *Ibicaba* (except for *Iracema farm*). They also constituted the majority of uniquely-identified households and observations in *Colônia Grande*, *Grama Larga*, *Lage*, *Morro Alto*, *Teteia*, and *Mariquita*; and were close to or above ⅓ of uniquely-identified households in *Pão-de-Ló*, *Saltinho* & *Pinheiro*, *Colônia Ibicaba*, and *Iracema farm*. For a broader discussion on the nature of immigration to São Paulo compared to Southern Brazil, *cf.* Buarque de Holanda (1941, pp. 23-4).

Online Appendix III - Figure 1 – Share of ethnolinguistic groups per colony in *Ibicaba*, pooled sample



Note: The shares reported in this table exclude observations with missing values for nationalities or colonies of work/residence.

Online Appendix III - Table 5 – Distribution of ethnolinguistic groups by colonies of work (coffee cultivation), pooled sample

	White Brazilians		Non - Whites		Italians		Germanic.		Portg.		Spaniards		East. Europe.		Total	
	N	n	N	n	N	n	N	n	N	n	N	n	N	n	N	n
Col. Grande	119	39	65	12	532	95	42	5	221	36	5	3	1	1	985	191
Grama Larga	133	58	13	6	562	123	40	8	57	22	23	6	26	5	854	228
Lage	26	14	4	1	55	18	1	1	0	0	0	0	0	0	86	34
Morro Alto	57	23	10	3	371	71	19	6	26	11	13	5	0	0	496	119
Pão Ló	44	16	0	0	54	20	27	4	10	5	11	2	1	1	147	48
Saltinho & Pinheiro	38	15	7	2	86	21	5	3	47	5	45	9	0	0	228	55
Teteia	17	10	6	3	54	21	2	1	12	7	1	1	0	0	92	43
Col. Ibicaba	10	7	31	13	106	69	5	4	7	4	14	11	10	10	183	118
Abundância	8	8	1	1	0	0	0	0	7	7	1	1	0	0	17	17
Boa Esperança	0	0	15	15	0	0	0	0	0	0	0	0	0	0	15	15
Col. Nova	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1
Linha	0	0	4	2	0	0	0	0	0	0	0	0	0	0	4	2
Mariquita	1	1	0	0	8	1	0	0	0	0	0	0	0	0	9	2
Paineira	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Port. Isabel	0	0	0	0	0	0	0	0	3	1	0	0	0	0	3	1
Ipiranga	0	0	0	0	0	0	0	0	3	3	0	0	0	0	3	3
Multiple colonies	1	1	0	0	1	1	0	0	0	0	0	0	0	0	2	2
Laborers	19	19	17	17	10	10	6	6	15	15	2	2	0	0	69	69
Iracema farm	27	21	0	0	18	12	0	0	5	3	3	3	0	0	53	39
Total	501	233	173	75	1,857	462	148	39	413	119	118	43	38	17	3,248	988

III. Additional data on Harvests & Earnings

Online Appendix III - Figure 2 plots the series of the volume of coffee harvested per household for the entire period 1888-1958, *i.e.* including all years with available data.[5]

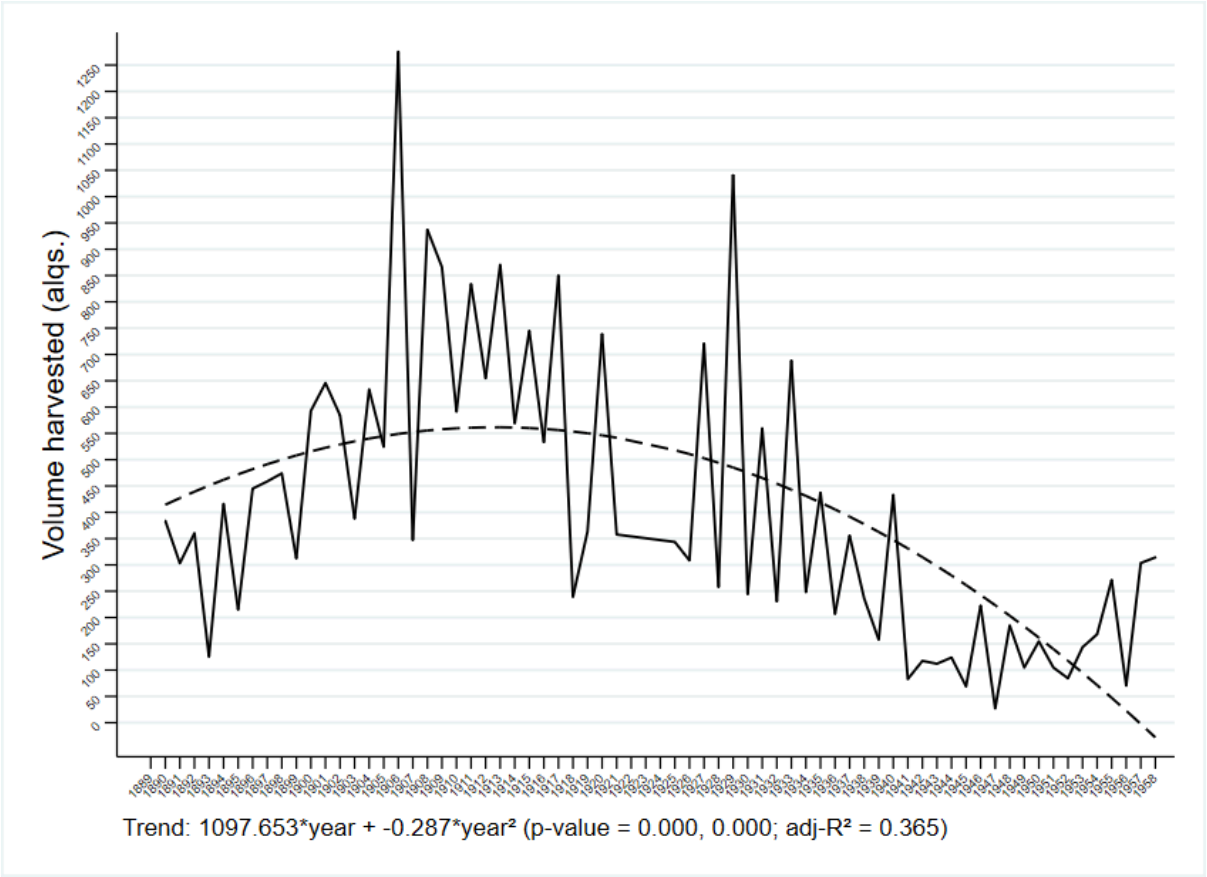
The same volatility in yearly yields observed in the restricted sample (1904-1942) is apparent again for the series in its entirety. Unlike the restricted sample, however, the complete series can be approximated by a quadratic trend. In line with the literature on the expansion of agricultural frontiers in São Paulo, Online Appendix III - Figure 2 shows that harvests per household in *Ibicaba* peaked in the early twentieth century (*cf.* Monbeig 1984, pp. 165-72, 176-9). In fact, the year 1912 was the turning point of the quadratic curve, very much in line with the estimates of Warren Dean (1976, p. 38) for Rio Claro (SP), a neighboring county of *Ibicaba*. This result increases our confidence that, for the period considered in this paper, *Ibicaba* was a representative plantation of São Paulo's Old West.

Online Appendix III - Figure 3, in turn, plots the series of the average nominal earnings (in *milréis*) per household from 1904 to 1942, distinguishing between the lean and harvest seasons. The volatility in physical output discussed in the main text and in the previous figure is clearly reflected in the nominal earnings from harvesting. Its troughs, however, were compensated, at least partially, by the less volatile earnings from coffee cultivation during the lean season. Moreover, the series decoupled during World War I, with earnings from coffee cultivation taking up the pace since the 1917 "coffee-defense" policy. The data thus suggest that *colono* households in *Ibicaba* were able to reap at least some benefits of the artificially high prices to producers attained by State interventions on the market during the late 1910s. The long-term trends in earnings also reflected the trends in physical output discussed in the main text. In 1904-1942, households' annual earnings in the lean season can be well-approximated by a quadratic trend (adj-R2 = 0.459), with a predicted maximum in 1929-30. At the same time, their earnings during the harvest season showed no trend at all.

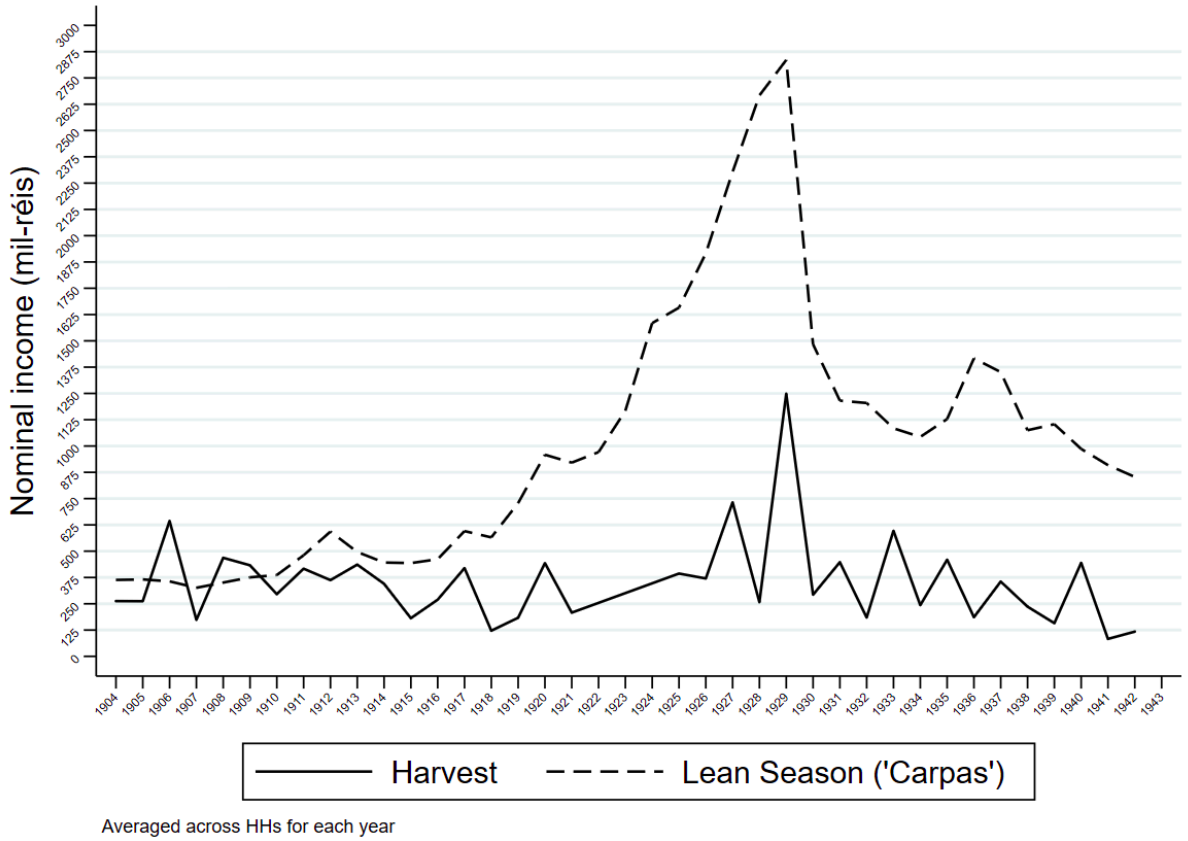
Finally, Online Appendix III - Figure 4 plots the first, last, and mean month in which harvesting took place in *Ibicaba* over the complete sample (1888-1958), averaged by year across all households for which data were available. The extended series is very consistent with the results reported in the main text, though the gap between the first and last harvest months is wider over the long run. In 1888-1958, 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 as the earliest and latest months ever recorded. This again accords with Dean (1976, p. 36), who found that hoeing in nineteenth-century Rio Claro (SP) was carried out from spring to early autumn (*i.e.* from September to March-April), with the harvest season beginning in April. For the 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.

[5] An analysis of coffee cultivation in the lean season cannot be accomplished for that entire period due to lack of data. Hence the limitation of the sample to 1904-1942.

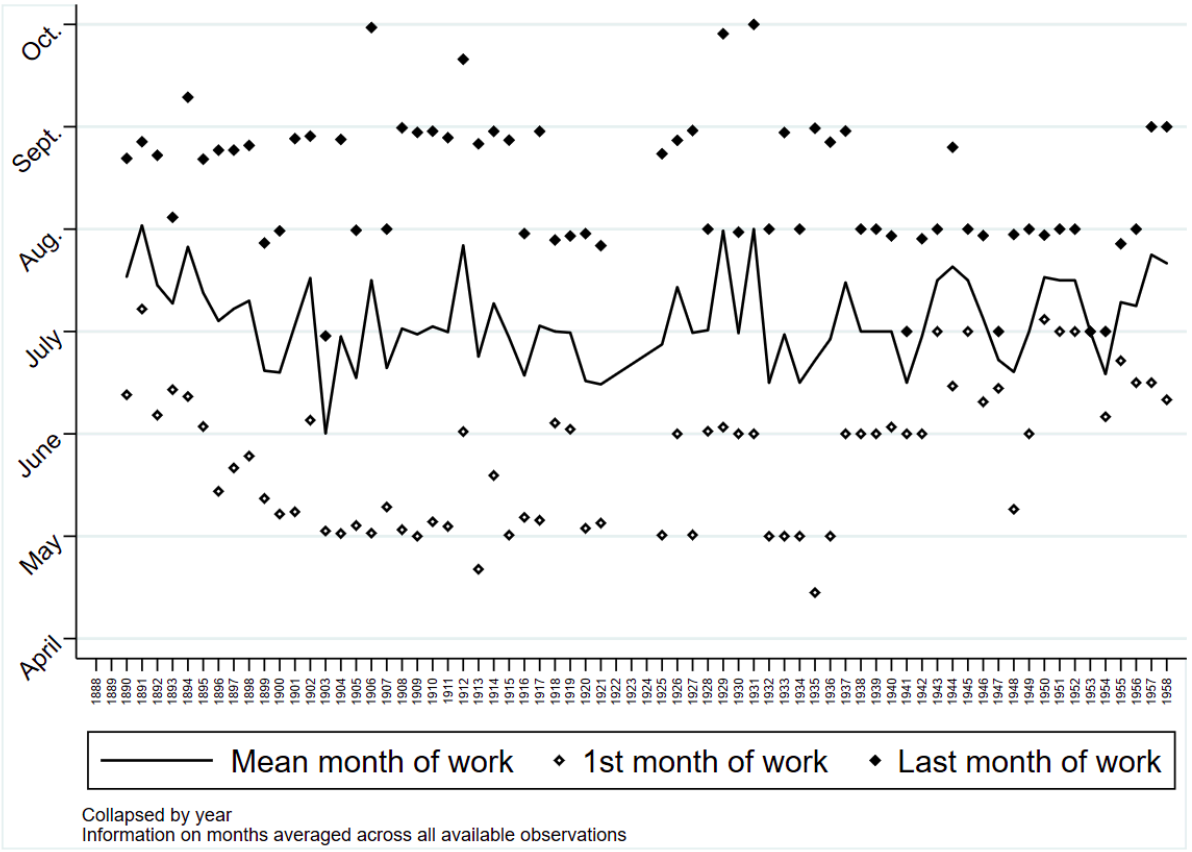
Online Appendix III - Figure 2 – Volume of coffee harvested (in *alqueires*) (1888-1958, with gaps)



Online Appendix III - Figure 3 – Nominal earnings (in *milréis*) from coffee production: coffee cultivation and harvesting (1904-1942, with gaps)



Online Appendix III - Figure 4 – Timing of the harvest season, averaged across households (1888-1958, with gaps)



Online Appendix IV

Additional analyses of the Baseline & Decompositions of earning differentials

Contents

I. Baseline specifications: Total effect of inputs on earnings & Net of prices and of task-assignments only

The baselines used the partial decomposition of households' nominal earnings ($Y_{ht}^{output} \equiv p_{ht} * Q_{ht}$) to estimate the statistical effect of each component and, based on them, to compute differences in the earning profiles of each ethnolinguistic group. As discussed in the main text, Q_{ht} is the physical output produced by household h under price (remuneration) p_{ht}, which was negotiated with households for the corresponding output in in year t. Physical output (Q_{ht}) is a function of the frequency (N_{ht}) of executed agricultural tasks and the households' production function $f(.)_{ht}$ for that ask, implying that $Y_{ht}^{output} \equiv p_{ht} * N_{ht} * f(.)_{ht}$.

Considering the production of a single, homogenous crop with agricultural capital supplied by the plantation, $f(.)_{ht}$ was modeled by a standard Cobb-Douglas production function, leading to the baseline specification reproduced in the first columns of Online Appendix IV - Tables 1 and 2:

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

By definition, the baseline thus estimates the per-task physical productivity in the harvest and lean seasons, *i.e.* net of the partial effects of remunerations and frequencies of executed tasks. Online Appendix IV - Tables 1 and 2 re-estimate the baseline netting out the effects of prices and frequencies separately. In each table, column (2) controls for the prices of executed tasks, but not for their frequencies (*i.e.* net of negotiated prices only); column (3) does the opposite, controlling for the frequencies, but not for prices (*i.e.* net of task-assignment only); column (4) controls for the inputs only, thus providing a revenues-based estimation of earnings.

The partial effects of the number of economically active members (in both the lean and harvest seasons) and the larger-than-unity responsiveness of households to prices (in the lean season) remain essentially unaffected by these modifications. Changes reLast month harv.d to the timing of harvest, by contrast, are explainable by omitted variable bias: absent controls for frequencies (columns 2 and 4), the signs of *Mean month harv.* and *1st month harv.* reverse relative to the baseline, while *Last month harv.* becomes statistically significant. Therefore, specifications that do not control for the number of months in which the household was engaged in harvesting load that effect onto the timing of the harvest.

In short, Online Appendix IV - Tables 1 and 2 reinforce the centrality of household composition, in general, and of price elasticities during the lean season. The baseline results are therefore not an artifact of earnings' partial decomposition; instead, they hold for revenue-based specifications and for specifications netted separately of prices (remunerations) and frequencies of executed tasks.

Online Appendix IV - Table 1 – Alternative specifications to the baseline's partial decomposition: Lean season

	Dep. Var.: $ln(Y_{ht}^{cult})$			
	(1)	(2)	(3)	(4)
ln(price)	1.285***	1.284***		
	(0.0836)	(0.0844)		
ln(N)	1.139***		1.094***	
	(0.252)		(0.258)	
ln(HH size)	0.0174	0.0181	0.0388	0.0395
	(0.0485)	(0.0487)	(0.0519)	(0.0521)
ln(# members 7-60y)	0.336***	0.336***	0.342***	0.342***
	(0.0398)	(0.0400)	(0.0452)	(0.0453)
Controls for location of *Estacas*				
Mean	Yes	Yes	Yes	Yes
MDEV	Yes	Yes	Yes	Yes
SD	Yes	Yes	Yes	Yes
HH-FE	Yes	Yes	Yes	Yes
Year-FE	Yes	Yes	Yes	Yes
Colony: Lean season	Yes	Yes	Yes	Yes
R-squared	0.862	0.861	0.833	0.833
r2_a	0.859	0.858	0.830	0.829
Observations	2,358	2,358	2,358	2,358
Number of id	512	512	512	512

Online Appendix IV - Table 2 – Alternative specifications to the baseline's partial decomposition: Harvest season

	Dep. Var.: $ln(Y_{ht}^{harv})$			
	(1)	(2)	(3)	(4)
ln(price)	1.313***	0.141		
	(0.350)	(0.439)		
ln(N)	4.447***		-0.197**	
	(0.634)		(0.0912)	
ln(HH size)	-0.0996	-0.0824	0.0123	-0.0797
	(0.0846)	(0.0871)	(0.0597)	(0.0873)
ln(# members 7-60y)	0.401***	0.434***	0.370***	0.432***
	(0.0704)	(0.0769)	(0.0530)	(0.0771)
Mean month harv.	-1.668***	2.335***	0.429**	2.335***
	(0.642)	(0.200)	(0.174)	(0.200)
1st month harv.	1.740***	-1.641***	-0.253**	-1.643***
	(0.481)	(0.123)	(0.105)	(0.121)
Last month harv.	-0.0649	-0.446**	-0.143*	-0.449**
	(0.174)	(0.177)	(0.0822)	(0.176)
Controls for location of *Estacas*				
Mean	Yes	Yes	Yes	Yes
MDEV	Yes	Yes	Yes	Yes
SD	Yes	Yes	Yes	Yes
HH-FE	Yes	Yes	Yes	Yes
Year-FE	Yes	Yes	Yes	Yes
Colony: Lean season	Yes	Yes	Yes	Yes
R-squared	0.844	0.745	0.855	0.745
r2_a	0.839	0.737	0.851	0.737
Observations	1,470	1,470	1,648	1,470
Number of id	324	324	337	324

II. Predicted earnings (*in milréis*) in coffee production: Aggregate descriptive statistics & Boxplots by agricultural season

Descriptive statistics for the predicted earnings of each ethnolinguistic group in the pooled sample are reported in Online Appendix IV - Table 3. These are the same estimates underlying the Kernel Density Estimates (KDEs) discussed in the main text. As a reminder, these figures were derived from the baseline estimates and computed using each ethnolinguistic group's observable characteristics, summing their average predicted earnings across the lean and harvest seasons, *i.e.* $\hat{Y}_{ht}^{total} = \hat{Y}_{ht}^{cult} + \hat{Y}_{ht}^{harv}$.

Online Appendix IV - Figures 1 and 2, in turn, presents the boxplots of each group's nominal earnings per household and per capita, respectively. The figures further distinguish between earnings from the lean and harvest seasons.

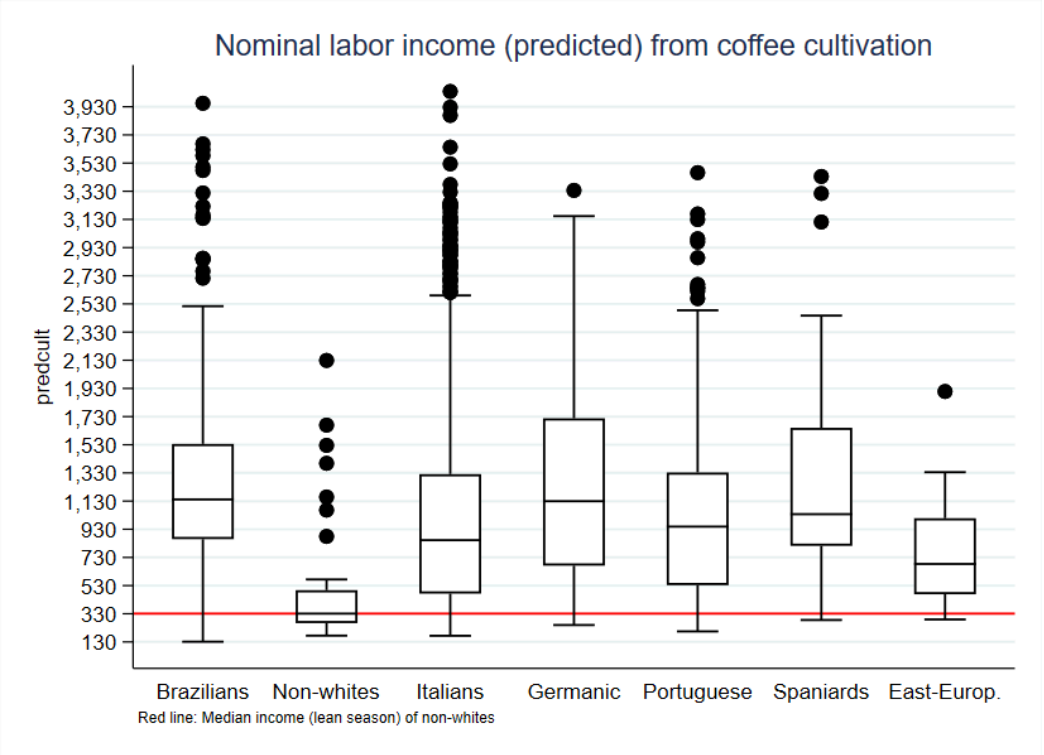
The evidence presented in the boxplots reinforces the conclusions drawn from the rankings of mean and median earnings across ethnolinguistic groups discussed in the main text. The precarious economic position of *Non-White* households is again apparent, with the most acute results occurring during the harvest season. Consistent with the rankings presented in the main text, *Non-Whites* recorded a high median relative to other groups during the harvest season once figures per capita are considered instead. Households of European origins, by contrast, show a worsened relative position with respect to this metric – a result particularly strong among the *Italians*.

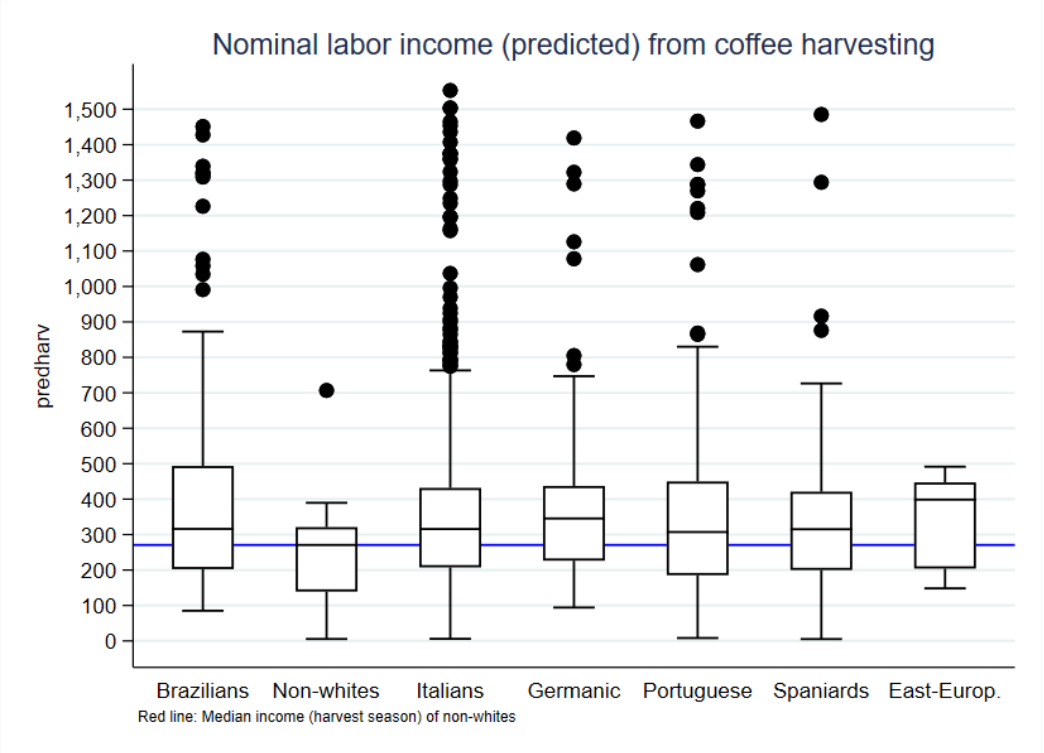
Online Appendix IV - Table 3 – Total predicted earnings (lean + harvest seasons) of *Ibicaba*'s laborers, by ethnolinguistic groups (pooled data for 1888-1958, w/ gaps)

	N	n	Mean	S.D.	Min	Max	Median	Skew.	Kurt.
White Brzs.	200	63	1,786.994	882.091	360.123	5,382.275	1,597.017	1.499	5.884
Non-whites	46	18	664.254	428.596	230.564	2,473.867	599.201	3.174	13.36
Italians	886	180	1,401.688	875.686	296.905	5,430.222	1,159.930	1.601	5.905
Germanic	78	14	1,763.865	958.203	389.058	4,657.936	1,567.096	1.049	3.94
Portuguese	195	38	1,515.916	853.569	246.917	4,928.570	1,426.474	1.334	5.329
Spaniards	53	10	1,669.917	854.917	434.073	3,987.827	1,421.802	1.145	3.713
Eastern-Eurp.	11	1	962.728	294.634	439.424	1,552.611	934.569	0.339	3.073
Pooled	1,469	324	1,471.839	888.486	230.564	5,430.222	1,299.580	1.459	5.475

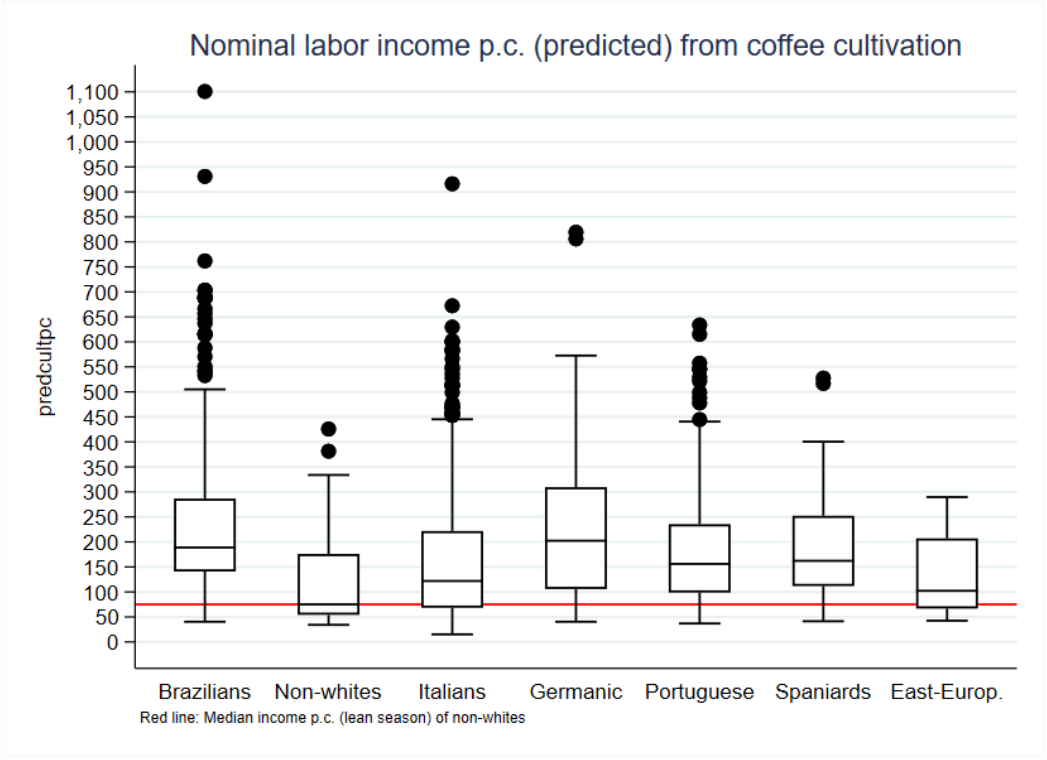
Notes: (1) N = observations; n = number of uniquely-identified HHs. (2) All predicted values derived from the baseline specification. (3) Total earnings (in *milréis*) corresponds to the sum of earnings from the lean and harvest seasons.

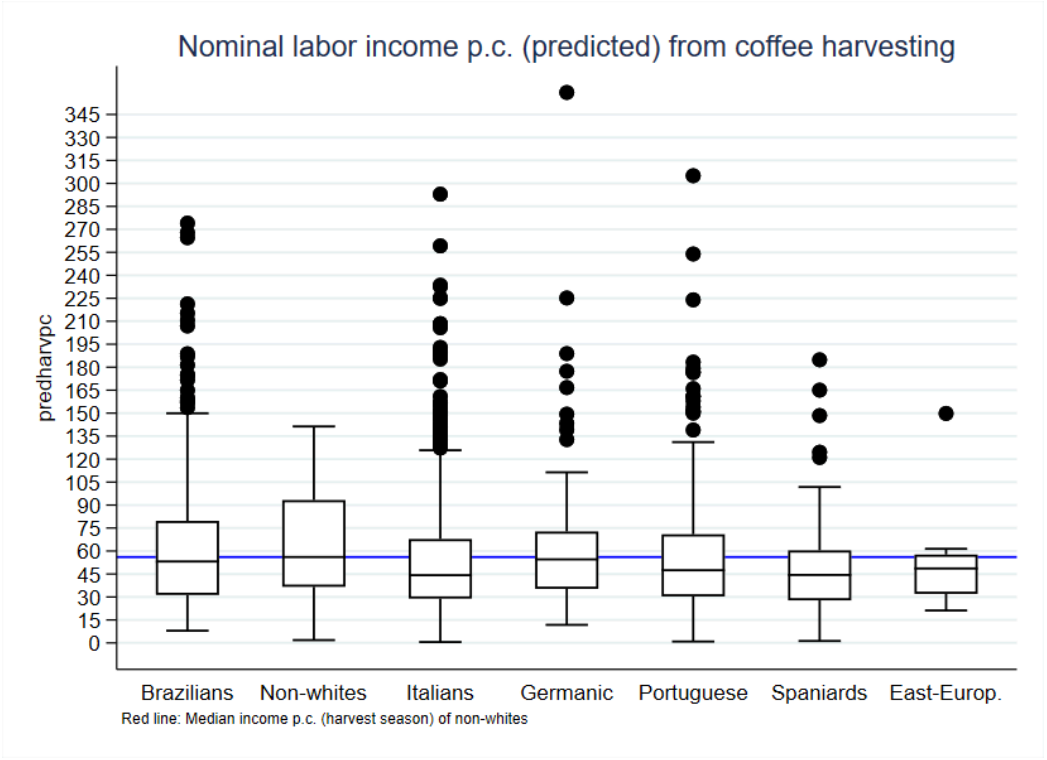
Online Appendix IV - Figure 1 – Boxplots of total predicted earnings (in *milréis*) per household by agricultural task, per ethnolinguistic group





Online Appendix IV - Figure 2 – Boxplots of total predicted earnings (in *milréis*) per capita by agricultural task, per ethnolinguistic group





III. Oaxaca-Blinder (OB) twofold decompositions: Information on colonies & Alternative reference groups

For reference, Online Appendix IV - Tables 4 and 5 report the full set of results summarized in Table 6 of the main text, for the lean and harvest seasons. In addition to Table 6 of the main text, they further report the explained and unexplained components for all colonies in *Ibicaba*.

Online Appendix IV - Tables 6-8 vary the reference groups for the twofold OB decompositions of household earnings during the lean season. Instead of using the pooled sample as the reference with which to estimate the parameters under the hypothesis of non-discrimination, these tables use *White Brazilians* (only), all Brazilians (*Whites* and *Non-Whites*), and *Italians* as the reference groups, respectively.

With *White Brazilians* as the reference group, the significance of the overall explained and unexplained components vanishes for *Non-White* households. In particular, the negative explained contribution of the number of economically active members becomes statistically non-significant. The demographic endowment disadvantage of *Non-White* households was, therefore, not driven by differences relative to *White Brazilians*. At the same time, *Italian* households showed a highly significant advantage in the explained *and* unexplained components of the number of economically active members. This suggests that the internal organization of *Italian* households, beyond the sheer number of economically active members, gave them a positive edge over households headed by *White Brazilians*. Finally, with *Italians* as the reference group, *Non-Whites* become the only households with a significantly smaller difference in mean earnings during the lean season. Relative to *Italians*, *Non-White* households faced endowment disadvantages in prices and in the number of economically active members. Consistent with results from the main text, however, households headed by *White Brazilians* showed explained advantages over *Italians* in prices (remuneration) and number of prunings (frequency of executed tasks during the lean season).

Online Appendix IV - Tables 9-11 repeat the exercise for the harvest season.

With *White Brazilians* as the reference group, the overall mean differences drop in statistical significance, remaining marginally significant only for *Non-Whites* (negative) and *Italians* (positive). *Non-White* households faced an even more severe explained disadvantage in prices relative to *White Brazilians* than relative to the sample mean. If *All Brazilians* are taken as the reference group instead, only *Italian* households show significantly higher mean earnings, with the main contributors being the unexplained components of the timing of the harvest. All other differences are significant only at the 10% level, including a marginally significant disadvantage of *Italians* in the explained component of prices (remunerations). *Italian* households also showed a marginally significant explained advantage in their number of economically active members relative to *All Brazilians*; because such difference was not observed in comparison to *White Brazilians*, this difference can only be attributed to the demographic advantage of *Italian* households *vis-à-vis Non-Whites*. That conclusion is in fact confirmed if *Italians* are used as the reference group, as *Non-White* households are then the only ones facing an explained disadvantage in their number of economically active members.

Online Appendix IV - Table 4 – Twofold Oaxaca-Blinder decomposition: Earnings from the lean season. Reference group: Pooled sample

	Dep. Var.: $ln(Y_{ht}^{cult})$											
	White Brazilians		Non-Whites		Italians		Germanics		Portuguese		Spaniards	
Pooled 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***	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	-0.231***	-0.761	0.475***	2.555	0.147***	0.972	-0.157	-0.0937	-0.0680	-0.321	-0.128	-6.704**
ln(N)	-0.0702***	-1.265	0.0893***	1.850***	0.0439***	-0.169	-0.0292	3.276	-0.0129	0.722	-0.0390	0.438
ln(HH size)	0.00215	-0.307	0.00826	0.0898	-0.00265	0.0672	0.00105	-0.290	0.00105	-0.161	-0.00150	0.0829
ln(# members 7-60y)	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 ID 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 ID 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 ID S.D.	0.00167	0.00267	0.000521	0.0197	-0.000564	0.0547	-0.00129	-0.0173	-0.00188	-0.0465	0.00386	-2.940
Colonies (Ref.: Col. Grande)												
Grama Larga	0.00510	-0.00152	0.0256	0.190	-0.0140	0.0268	0.0000486	0.0867	0.0219	-0.0474	0.00656	-2.575
Lage	-0.00731	-0.00816	-0.00349	0.0106	-0.00155	0.00284	0.00513	0.00684	0.00803*	-0.000478	0.00723*	-0.000115
Morro Alto	0.00406	-0.00522	0.0134	0.0502	-0.0156	0.00971	0.00160	0.0119	0.0198*	-0.0150	0.00903	-0.685
Pão de Ló	-0.0113	-0.0211	0.00828*	-0.000200	0.00961	0.00159	-0.0251	0.0723	0.00502	0.000411	-0.0148	0.0225
Saltinho & Pinheiro	0.000310	-0.00870	0.00846	-0.00336	0.00988	-0.0105	0.00612	0.0204	-0.00711	0.00222	-0.0641	-7.047
Teteia	-0.000179	-0.0912	-0.00685	0.0228	-0.00142	-0.000576	0.00170	0.00511	0.00222	-0.00115	0.00540**	-0.000220
Col. Ibicaba	0.00206	0.00303	0.0106	-0.000514	-0.00147	-0.000174	-0.00104	-0.0000066	-0.00114	-0.000107	-0.00103	-0.0000376
Linha	-0.00126	0.00125	0.00934	0.0108	-0.000156	-0.000159	0.000530	-0.0000099	0.0000375	-0.0000260	0.000524	0.00000407
Mariquita	0.000263	-0.000433	0.000995	0.0000105	-0.00116	0.0000237	0.00101	0.0000278	0.00111	-0.0000168	0.00100	-0.0000167
Paineira	0.000361	-0.000461	-0.0000602	-0.0000012	-0.000140	0.0000743	-0.0000614	-0.0000043	-0.0000674	-0.0000044	-0.0000607	0.00000822
Portão Isabel	0.000657	-0.0000082	0.000567	0.00000089	0.00132	0.0000962	0.000579	-0.0000038	-0.00416	-0.000807	0.000572	0.00000077
Laborers	0.00614	-0.00168	-0.00247	0.0000888	-0.00241	-0.00126	-0.00252	0.000204	-0.000793	-0.0387	0.00137	-0.281

Note: *Cf.* Table 6 of the main text.

Online Appendix IV - Table 5 – Twofold Oaxaca-Blinder decomposition: Earnings from the harvest season. Reference group: Pooled sample

	Dep. Var.: $ln(Y_{ht}^{harv})$											
	White Brazilians		Non-Whites		Italians		Germanics		Portuguese		Spaniards	
Pooled sample	5.700***		5.696***		5.554***		5.681***		5.695***		5.690***	
Ethn. 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***	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	-0.292***	0.0312	0.428**	-3.321	0.185**	-0.300	-0.148	-0.538	-0.115	0.786*	-0.168	5.225
ln(N)	0.105	9.663***	-0.0934	3.423**	-0.213**	-0.263	-0.119	-28.58**	0.294*	-6.156***	0.467	616.6***
Mean month harv.	0.0706	-32.71**	0.0750	-12.49**	-0.0455	-11.88	0.0846	-12.17**	0.0159	-7.850*	-0.131	-14.18**
1st month harv.	-0.180	20.21**	0.0205	8.288**	0.209*	4.511	-0.0372	-16.39*	-0.236*	-5.049	-0.101	515.3***
Last month harv.	-0.00106	1.971	0.00657	4.094	0.00420	7.636	0.00517	4.969	-0.00760	16.84***	-0.0140	-842.5***
ln(HH size)	-0.00635	0.0814	-0.0554	0.695	0.0138	0.132	-0.00563	0.0292	-0.00677	-0.486	0.00499	-1.375
ln(# members 7-60y)	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 ID 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 ID 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 ID S.D.	-0.00178	-0.911	-0.000160	-1.529	0.000602	0.219	0.00454	-2.765	0.00122	-0.455	-0.00738	41.97*
Colonies (Ref. Col. Grande)												
Grama Larga	0.00377	0.000377	0.0142	0.130	-0.00659	-0.0592	-0.00237	0.0615	0.00779	-0.0300	0.00821	0.0152
Lage	-0.000345	-0.0176	-0.000141	0.00152	-0.0000604	-0.00714	0.000274	-0.000124	0.000299	0.00103	0.000269	-0.000211
Morro Alto	0.000664	-0.0576	0.00541	0.00681	-0.00765	-0.0120	0.00287	-0.0242	0.0100	0.0309	0.00602	28.11
Pão de Ló	-0.00404	0.0725	0.00272	0.000274	0.00301	-0.0179	-0.00915	0.167	0.00245	0.00214	-0.00494	28.93
Saltinho & Pinheiro	0.00117	-0.528	0.00842	0.0128	0.0178	-0.0165	0.0185	0.000206	-0.0149	0.00939	-0.115	-69.95**
Teteia	-0.000152	-0.0914	-0.00115	-0.0000556	0.000546	-0.000977	-0.000290	-0.00231	-0.000123	0.0111	-0.00116	-0.000196
Col. Ibicaba	-0.00113	0.00261	-0.0102	-0.0566	0.00123	0.00108	0.000713	-0.0000458	0.000778	-0.0000004	0.000700	-0.000120
Linha	0.00238	-0.00411	-0.0526	0.0526	-0.00359	0.0105	-0.000163	-0.000137	0	-0.00171	-0.000160	-0.000368
Mariquita	0.000691	0.0000396	0.000616	0.00000709	-0.000990	0.00000931	0.000630	0.0000511	0.000688	-0.0000921	0.000619	-0.0000293
Portão Isabel	-0.000253	0.00000357	-0.000225	0.0000377	-0.000549	0.000242	-0.000230	-0.0000583	0.00165	-0.00831	-0.000226	0.0000522
Laborers	-0.000376	-0.0000230	-0.000335	-0.0000026	0.000539	-0.000157	-0.000343	0.00000604	-0.000374	0.0000537	-0.000337	0.0000188

Note: *Cf.* Table 6 of the main text.

Online Appendix IV - Table 6 – Twofold Oaxaca-Blinder decomposition: Earnings from the lean season. Reference group: *White Brazilians*, only

	Dep. Var.: $ln(Y_{ht}^{cult})$									
	Non-Whites		Italians		Germanics		Portuguese		Spaniards	
Group 1: White Brazilians	6.876***		6.876***		6.876***		6.876***		6.876***	
Group 2: Ethnoling. group	5.817***		6.760***		6.866***		6.711***		6.927***	
Difference	1.059***		0.116		0.00981		0.165*		-0.0505	
Explained	1.108		0.584*		-0.803*		0.208*		-0.0474	
Unexplained	-0.0493		-0.469		0.813**		-0.0434		-0.00317	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	0.594***	3.235	0.269***	1.220	0.0403	0.569	0.123*	0.431	0.0519	-5.853**
ln(N)	0.218***	2.793***	0.109***	0.909	0.0415	4.246	0.0463*	1.276	0.0262	1.468**
ln(HH size)	0.0409	0.337	-0.00591	0.282	-0.00728	0.00825	-0.00458	0.126	-0.0109	0.340
ln(# members 7-60y)	0.0529	-0.342*	-0.0518*	-0.463***	-0.0191	-0.414*	-0.0250	-0.424*	-0.0111	0.182
Estaca ID Mean	-0.00206	0.851	-0.00253	0.00421	-0.00217	0.000619	-0.00860	-0.0924	-0.00695	6.224***
Estaca ID MDEV	-0.0122	-0.139	0.000554	0.103	0.0287	0.0822	0.0455	0.128	0.142	2.282
Estaca ID S.D.	0.0110	0.00597	-0.00123	-0.00224	-0.0267	0.00422	-0.0390	-0.0118	-0.0410	-2.900
Colonies (Ref. Col. Grande)										
Grama Larga	0.0314	0.177	-0.00869	0.00854	-0.00292	0.0822	0.0286	-0.0560	0.000884	-2.575
Lage	0.00394	0.0170	0.00518	0.00927	0.0169	0.00886	0.0236	-0.00269	0.0139	0.00696
Morro Alto	0.0158	0.0487	-0.00836	0.00881	-0.00178	0.0167	0.0258	-0.0191	0.00217	-0.676
Pão de Ló	0.0404	-0.00346	0.0180	0.0143	-0.0141	0.0889	0.0231	0.0111	-0.0112	0.0482
Saltinho & Pinheiro	0.0132	-0.00220	0.00631	-0.00298	0.00512	0.0251	-0.00967	0.0109	-0.0715	-7.035
Teteia	-0.0101	0.118	-0.000477	0.0908	0.00122	0.0955	0.00288	0.0892	0.171	-0.0747
Col. Ibicaba	0.00638	-0.000835	-0.00322	-0.00190	-0.00601	0.000654	-0.00427	-0.00108	-0.00673	0.00138
Linha	0.00893	0.0112	0	0	0	0	0	0	0	0
Mariquita	0.000823	0.000361	-0.000695	0.000427	0.000864	0.000320	0.00102	0.000165	0.00147	-0.000287
Paineira	-0.000792	0.000892	-0.000340	0.000440	-0.0000927	0.000193	-0.0000662	0.000167	0.000452	-0.000352
Portão Isabel	0	0	0	0	0	0	-0.00499	0.0000185	0	0
Laborers	-0.000599	-0.00508	-0.00483	-0.000178	-0.0130	0.00734	-0.00321	-0.0394	-0.00667	-0.276

Online Appendix IV - Table 7 – Twofold Oaxaca-Blinder decomposition: Earnings from the lean season. Reference group: *All Brazilians*: *White Brazilians + Non-Whites*

	Dep. Var.: $ln(Y_{ht}^{cult})$							
	Italians		Germanics		Portuguese		Spaniards	
Group 1: All Brazilians	6.714***		6.714***		6.714***		6.714***	
Group 2: Foreigners	6.760***		6.866***		6.711***		6.927***	
Difference	-0.0465		-0.152		0.00261		-0.213	
Explained	-0.0686		-0.890***		-1.813		-1.377	
Unexplained	0.0221		0.738***		1.815		1.165	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	0.165**	0.985	-0.0553	0.326	0.0321	0.183	-0.0246	-6.116**
ln(N)	0.0775***	0.893	0.0118	4.229	0.0246	1.251	-0.00126	1.448**
ln(HH size)	-0.00722	0.161	-0.0134	-0.108	-0.00856	0.00813	-0.0105	0.218
ln(# members 7-60y)	-0.0729**	-0.349***	-0.0331	-0.307	-0.0414	-0.314	-0.0258	0.289
Estaca ID Mean	-0.00252	-0.00350	-0.000975	-0.00828	-0.00856	-0.100	-0.00288	6.212***
Estaca ID MDEV	0.000500	0.0728	0.0118	0.0687	0.0467	0.0967	0.0511	2.343
Estaca ID S.D.	-0.00138	-0.0650	-0.0132	-0.0722	-0.0428	-0.0709	-0.0401	-2.964
Colonies (Ref. Col. Grande)								
Grama Larga	-0.0114	0.0294	-0.00687	0.104	0.0224	-0.0316	-0.00149	-2.555
Lage	0.00460	0.00718	0.0140	0.00911	0.0195	-0.00124	0.0137	0.00456
Morro Alto	-0.00950	0.0151	-0.00382	0.0239	0.0220	-0.0102	0.00461	-0.673
Pão de Ló	0.0148	0.0148	-0.0209	0.0931	0.0189	0.0126	-0.0178	0.0521
Saltinho & Pinheiro	0.00435	-0.000883	0.00465	0.0258	-0.0114	0.0128	-0.0734	-7.033
Teteia	0.000647	0.00324	0.00276	0.00747	0.00537	0.000259	0.00868	0.00120
Col. Ibicaba	-0.00419	0.00142	-0.00430	0.00130	-0.00306	0.0000562	-0.00398	0.000979
Linha	-0.00193	0.00193	-0.000766	0.000766	-0.00255	0.00255	-0.0103	0.0103
Mariquita	-0.000797	0.0000429	0.000599	0.0000982	0.000721	-0.0000233	0.000768	-0.0000706
Paineira	-0.000282	-0.000388	-0.000178	-0.000493	-0.0000416	-0.000629	-0.000841	0.000170
Portão Isabel	0	0	0	0	-0.00474	-0.000230	0	0
Laborers	-0.00413	0.00429	-0.0102	0.00966	-0.00300	-0.0344	-0.000823	-0.277

Online Appendix IV - Table 8 – Twofold Oaxaca-Blinder decomposition: Earnings from the lean season. Reference group: *Italians*

	Dep. Var.: $ln(Y_{ht}^{cult})$									
	White Brazilians		Non-Whites		Germanics		Portuguese		Spaniards	
Group 1: Italians	6.760***		6.760***		6.760***		6.760***		6.760***	
Group 2: Ethnoling. group	6.876***		5.817***		6.866***		6.711***		6.927***	
Difference	-0.116		0.944***		-0.106		0.0491		-0.166	
Explained	-0.584*		0.473		0.0372		-0.0818		-0.714*	
Unexplained	0.469		0.470*		-0.143		0.131		0.548	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	-0.269***	-1.220	0.453***	1.888	-0.233	-0.646	-0.127*	-0.808	-0.189	-7.101***
ln(N)	-0.109***	-0.909	0.0717*	1.922*	-0.0473	3.317	-0.0235*	0.328	-0.0581	0.534
ln(HH size)	0.00591	-0.282	0.00832	0.0937	0.00299	-0.278	0.00156	-0.156	0.000232	0.0526
ln(# members 7-60y)	0.0518*	0.463***	0.211***	0.0153	0.0126	0.0690	0.0192	0.0470	0.00375	0.682***
Estaca ID Mean	0.00253	-0.00421	0.00105	0.846	-0.000901	-0.00233	0.00437	-0.107	0.00180	6.214***
Estaca ID MDEV	-0.000554	-0.103	-0.000319	-0.254	0.000919	0.00628	0.000176	0.0700	0.000126	2.320
Estaca ID S.D.	0.00123	0.00224	0.000718	0.0197	-0.000798	-0.0182	-0.000860	-0.0465	0.00132	-2.939
Colonies (Ref. Col. Grande)										
Grama Larga	0.00869	-0.00854	0.0229	0.185	0.00428	0.0751	0.0227	-0.0499	0.00930	-2.583
Lage	-0.00518	-0.00927	-0.00228	0.00873	0.00474	0.00657	0.00691*	-0.000461	0.00675*	-0.000301
Morro Alto	0.00836	-0.00881	0.0155	0.0486	0.00648	0.00796	0.0221	-0.0158	0.0132	-0.687
Pão de Ló	-0.0180	-0.0143	0.00491	-0.000262	-0.0162	0.0588	0.000437	0.00149	-0.0210	0.0257
Saltinho & Pinheiro	-0.00631	0.00298	0.00726	0.000435	0.00280	0.0241	-0.0137	0.0116	-0.0669	-7.043
Teteia	0.000477	-0.0908	-0.00697	0.0245	0.00155	0.00480	0.00266	-0.000916	0.00646*	-0.000466
Col. Ibicaba	0.00322	0.00190	0.00659	0.00409	-0.000226	-0.00000769	-0.000185	-0.0000488	-0.000215	-0.0000189
Linha	0	0	0.0121	0.00803	0	0	0	0	0	0
Mariquita	0.000695	-0.000427	0.00145	0.00000593	0.00141	0.0000400	0.00146	-0.00000402	0.00149	-0.0000339
Paineira	0.000340	-0.000440	0	0	0	0	0	0	0	0
Portão Isabel	0	0	0	0	0	0	-0.00410	-0.000868	0	0
Laborers	0.00483	0.000178	-0.000743	0.0000722	-0.000939	0.000268	0.000205	-0.0378	0.00156	-0.279

Online Appendix IV - Table 9 – Twofold Oaxaca-Blinder decomposition: Earnings from the harvest season. Reference group: *White Brazilians*, only

	Dep. Var.: $ln(Y_{ht}^{harv})$									
	Non-Whites		Italians		Germanic		Portuguese		Spaniards	
group_1: White Brazilians	5.557***		5.557***		5.557***		5.557***		5.557***	
group_2: Ethonling. group	5.223***		5.765***		5.684***		5.585***		5.443***	
difference	0.334*		-0.208*		-0.127		-0.0283		0.114	
explained	-2.304*		-1.422*		-1.852***		-0.706		0.484	
unexplained	2.637**		1.214		1.725***		0.678		-0.371	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	0.731**	-3.390	0.230*	-0.110	0.130	-0.580	0.299*	0.539*	0.118	5.144
ln(N)	-0.0119	-6.381*	-0.174	-9.767**	0.101	-38.42***	0.244	-14.98***	0.268	606.6***
Mean month harv.	-0.0139	32.65**	-0.0767	32.72**	-0.0498	32.69**	-0.110	32.75**	-0.0665	32.71**
1st month harv.	-0.109	-16.39*	0.235*	-20.39**	-0.180	-40.82***	-0.0987	-27.39***	0.0376	489.3***
Last month harv.	0.0677	-6.826	0.000607	-2.134	0.0953	-4.972	0.0413	7.753	-0.0293	-852.3***
ln(HH size)	-0.0894	0.661	0.0128	-0.0480	0.000231	-0.0173	0.000411	-0.448	0.0302	-1.607
ln(# members 7-60y)	0.262*	-0.269	-0.0379	-0.0738	-0.0339	-0.145	-0.00745	0.443	-0.0203	1.066
Estaca ID Mean	0.0487	1.675	-0.00369	0.640	-0.00245	0.655	0.00670	0.316	0.00209	180.8***
Estaca ID MDEV	0.0635	0.591	-0.00247	-0.689	0.135	1.949	0.126	-0.305	-0.498	-40.10
Estaca ID S.D.	-0.0414	-0.578	0.00276	0.913	-0.127	-1.727	-0.126	0.598	0.807	42.06*
Colonies (Ref. Col. Grande)										
Grama Larga	0.0255	0.113	-0.00993	-0.0195	0.0110	0.0443	0.00298	-0.0395	-0.000599	0.0184
Lage	0.00147	0.0172	0.00224	0.0130	0.00207	0.0166	0.00414	0.0145	0.00571	0.0129
Morro Alto	0.0255	0.0404	-0.00518	0.0498	0.00238	0.0337	0.0202	0.0707	0.0203	28.15
Pão de Ló	-0.0425	-0.0235	0.00730	-0.0759	0.0185	0.0697	-0.0190	-0.0449	0.0108	28.85
Saltinho & Pinheiro	0.0300	0.514	0.0114	0.511	-0.0844	0.629	-0.00754	0.525	-0.0143	-69.51**
Teteia	-0.0134	0.103	0.000324	0.0913	0.000457	0.0877	-0.000160	0.102	0.273	-0.183
Col. Ibicaba	-0.0231	-0.0440	-0.00182	0.00151	-0.00544	0.00494	0.000480	-0.000983	-0.00142	0.000914
Linha	-0.0539	0.0539	0	0	0	0	0	0	0	0
Mariquita	0	0	-0.000907	-0.0000736	0	0	0	0	0	0
Portão Isabel	0	0	0	0	0	0	-0.00383	-0.00284	0	0
Laborers	0	0	0.000397	-0.0000146	0	0	0	0	0	0

Online Appendix IV - Table 10 – Twofold Oaxaca-Blinder decomposition: Earnings from the harvest season. Reference group: *All Brazilians*: *White Brazilians + Non-Whites*

	Dep. Var.: $ln(Y_{ht}^{harv})$							
	Italians		Germanics		Portuguese		Spaniards	
Group 1: All Brazilians	5.493***		5.493***		5.493***		5.493	
Group 2: Immigrants	5.765***		5.684***		5.585***		5.443	
Difference	-0.272**		-0.190		-0.0918		0.0502	
Explained	-1.770**		-0.298		6.528		11.41	
Unexplained	1.498*		0.108		-6.620		-11.36	
	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	0.132*	-0.137	-0.0152	-0.560	0.0379	0.675*	-0.0435	5.180
ln(N)	-0.137	-6.026***	-0.0220	-34.52***	0.246	-11.20***	0.242	610.4
Mean month harv.	-0.0738	14.21*	-0.0166	14.16*	-0.0796	14.22*	-0.0200	14.16
1st month harv.	0.193*	-11.37**	-0.0455	-31.97***	-0.120	-18.39***	0.00945	498.4
Last month harv.	0.00326	3.050	0.0373	0.273	-0.00509	12.99**	-0.0699	-847.1
ln(HH size)	0.0230	-0.115	-0.00707	-0.0672	-0.00472	-0.500	0.0398	-1.674
ln(# members 7-60y)	-0.0859*	0.00304	-0.0573	-0.0933	-0.0250	0.490	-0.0667	1.141
Estaca ID Mean	-0.00335	0.424	0.000599	0.436	0.00609	0.101	-0.0102	180.6
Estaca ID MDEV	-0.00243	-0.332	0.0232	2.418	0.133	0.0450	-0.302	-39.93
Estaca ID S.D.	0.00264	0.225	-0.0182	-2.524	-0.129	-0.0871	0.210	41.97
Colonies (Ref. Col. Grande)								
Grama Larga	-0.0132	0.00482	0.00842	0.0680	0.00161	-0.0170	0.00469	0.0342
Lage	0.000850	0.000537	-0.00641	0.0112	-0.00185	0.00662	0.00220	0.00257
Morro Alto	-0.00589	0.0324	0.00159	0.0164	0.0178	0.0550	0.0204	28.13
Pão de Ló	0.00549	-0.0425	0.00975	0.110	-0.0241	-0.00819	0.0220	28.87
Saltinho & Pinheiro	0.00664	-0.161*	0.0530	-0.185*	0.00656	-0.165*	-0.00279	-70.20
Teteia	0.000431	-0.00951	0.0000739	-0.0126	-0.00152	0.00254	-0.00383	-0.00698
Col. Ibicaba	0.00232	0.00655	0.00304	0.00565	0.00482	0.00386	0.00464	0.00404
Linha	-0.00422	-0.0176	0	-0.0218	0.0798	-0.102	0.127	-0.149
Mariquita	-0.000883	-0.0000976	0	0	0	0	0	0
Portão Isabel	0	0	0	0	0.000402	-0.00707	0	0
Laborers	0.000379	0.00000310	0	0	0	0	0	0

Online Appendix IV - Table 11 – Twofold Oaxaca-Blinder decomposition: Earnings from the harvest season. Reference group: *Italians*

	Dep. Var.: $ln(Y_{ht}^{harv})$									
	White Brazilians		Non-Whites		Germanic		Portuguese		Spaniards	
group_1: Italians	5.765***		5.765***		5.765***		5.765***		5.765	
group_2: Ethnoling. group	5.557***		5.223***		5.684***		5.585***		5.443	
difference	0.208*		0.542***		0.0811		0.180*		0.322	
explained	1.422*		2.058***		-1.439*		0.131		0.134	
unexplained	-1.214		-1.516***		1.520*		0.0489		0.188	
explained	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.	explain.	unexpl.
ln(price)	-0.230*	0.110	0.167	-2.946	-0.133	-0.437	-0.171	0.889*	-0.199	5.341
ln(N)	0.174	9.767**	-0.00602	3.554**	-0.0285	-28.35**	0.397**	-5.192*	0.499	616.3
Mean month harv.	0.0767	-32.72**	0	0	0	0	0	0	0	0
1st month harv.	-0.235*	20.39**	-0.0350	3.690**	-0.0635	-20.78**	-0.181**	-7.154**	-0.0923	509.6
Last month harv.	-0.000607	2.134	0.0828	-4.707	0.0436	-2.787	-0.147	10.07	-0.165	-850.0
ln(HH size)	-0.0128	0.0480	-0.0681	0.675	-0.0102	0.0283	-0.00403	-0.408	-0.00143	-1.540
ln(# members 7-60y)	0.0379	0.0738	0.295***	-0.189	0.00414	-0.0718	0.0134	0.534**	0.0203	1.137
Estaca ID Mean	0.00369	-0.640	0.00210	1.085	0.00376	0.0120	0.00456	-0.319	0.00377	180.1
Estaca ID MDEV	0.00247	0.689	0.000561	1.346	-0.00668	2.782	-0.000226	0.513	0.0121	-39.91
Estaca ID S.D.	-0.00276	-0.913	-0.000630	-1.534	0.00907	-2.779	0.000117	-0.444	-0.0127	41.97
Colonies (Ref. Col. Grande)										
Grama Larga	0.00993	0.0195	0.0297	0.138	0.000655	0.0841	0.0120	-0.0190	0.0205	0.0267
Lage	-0.00224	-0.0130	-0.000758	0.00414	0.00320	0.000181	0.00156	0.00182	0.00414	-0.000760
Morro Alto	0.00518	-0.0498	0.0111	0.0101	0.00947	-0.0179	0.0118	0.0345	0.0141	28.11
Pão de Ló	-0.00730	0.0759	0.00275	-0.000163	-0.00807	0.165	0.00229	0.00243	-0.00933	28.94
Saltinho & Pinheiro	-0.0114	-0.511	0.00180	0.0204	0.0207	0.00149	-0.0230	0.0184	-0.206	-69.85
Teteia	-0.000324	-0.0913	-0.00173	-0.00000588	-0.000659	-0.00278	-0.000516	0.0106	-0.00117	-0.000562
Col. Ibicaba	0.00182	-0.00151	-0.0178	-0.0490	-0.000157	-0.0000304	-0.000156	-0.0000319	-0.000194	0.00000649
Linha	0	0	0.0519	-0.0519	0	0	0	0	0	0
Mariquita	0.000907	0.0000736	0.000948	0.0000321	0.000873	0.000107	0.00109	-0.000108	0.00105	-0.0000738
Portão Isabel	0	0	0	0	0	0	0.00128	-0.00795	0	0
Laborers	-0.000397	0.0000146	-0.000375	-0.00000753	-0.000431	0.0000489	-0.000529	0.000147	-0.000448	0.0000659

IV. Difference-in-means tests of households' characteristics: A complementary analysis to the OB Decompositions

A potential problem with OB decompositions is their sensitivity to the choice of reference group, especially for categorical variables (Jann 2008, p. 9), which may account for some implausibly large estimates for the timing of the harvest. To address this issue, this section reports simple difference-in-means tests for all independent variables used in the baseline specification, assuming unequal variances between groups (Online Appendix IV - Table 12). This approach implies that each ethnolinguistic group's mean is compared with the sample mean excluding that group, so that any difference between the two is attributed to endowments in the variables considered.

The results are reassuringly similar to the twofold OB decompositions for the explained components, particularly for prices (remunerations) and demographic characteristics.

Both *Non-White* and *Italian* households earned lower prices than the corresponding sample mean (significant at the 1% level), further indicating that variation in prices earned were not exclusively a result of 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 means in both seasons. This suggests that variation in prices earned did not result from xenophobia, either. That *Spaniards* earned higher-than-average prices in the harvest season further weighs against the possibility that these results reflect the longer integration of *Germanic* and *Portuguese* households. Finally, consistent with all previous results, *White Brazilian* and *Non-White* households were significantly smaller and had fewer economically active members, while *Italian* households showed exactly the reverse demographic pattern.

The measures of central tendency and dispersion for the *estacas* reveal statistically significant differences across groups. Although these numerical differences have no substantive interpretation, since the considered variables refer solely to identifiers, they nonetheless indicate that ethnolinguistic groups clustered within colonies of work. Also, the mean year when each group was most commonly registered in the sample reveals dynamic changes in the ethnolinguistic composition of *Ibicaba*'s labor force. As for the timing of the harvest – the most problematic estimates of the OB decompositions –. the difference-in-means tests show that *Non-White* households were engaged systematically Last month harv.r in the harvest season, starting and ending their harvests Last month harv.r than the corresponding sample means. *Italian* households, by contrast, started earlier in the first month and continued longer than the corresponding sample mean.[1] Finally, households headed by *White Brazilians* ended earlier in the last month of harvesting, yielding the earliest mean month of work among all groups.

[1] *Portuguese* and *Spaniards* showed the opposite pattern, starting Last month harv.r in their first month and ending earlier in the last month of harvesting.

Online Appendix IV - Table 12 – Difference-in-means tests of observable characteristics by ethnolinguistic groups (pooled sample)

	obs1	obs2	Mean1	Mean2	Diff.	S.D.	p-value
ln(price) in the lean season							
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
ln(price) in the harv. season							
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
ln(# prunings)							
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
ln(# months harvesting)							
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

Online Appendix IV - Table 12 (ctd.)

	obs1	obs2	Mean1	Mean2	Diff.	S.D.	p-value
			ln(HH size)				
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
			ln(# members 7-60y)				
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 ID 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 ID 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 ID S.D.				
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

Online Appendix IV - Table 12 (ctd.)

	obs1	obs2	Mean1	Mean2	Diff.	S.D.	p-value
			Mean month harv.				
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
			1st month harv.				
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
			Last month harv.				
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

Online Appendix V
Robustness Checks

Contents

I. Alternative categorizations of ethnolinguistic origins

Online Appendix V - Table 1 checks the sensitivity of the tests for taste-based discrimination to the different criteria for categorizing household ethnolinguistic origin presented in Online Appendix II. As a reminder, compared with *Origins* (baseline), the *Origins imposed*, *One drop*, and *One drop imposed* criteria increase the likelihood that household h is assigned to the category *Non-White*, or to the European ancestries culturally most distant from Brazilians (*i.e. Eastern European* $\succ$ *Germanic* $\succ$ *Italian* $\succ$ *Spaniard* $\succ$ *Portuguese*).

The results are robust to these alternative categorizations of ethnolinguistic origins. The rejection of the hypotheses of discrimination against *Non-Whites* and in favor of European immigrants and their descendants does not depend on any of these criteria in particular. In this regard, the marginally significant effect of *Non-Whites* and *Italians* under the *One drop* categorization is a statistical artifact, since it does not hold (or increase) with the *One drop imposed* categorization. The only group that raises concern are the *Spaniards*, which show statistically significant (at the 1% level) negative effects during the harvest season under the *One drop* and *One drop imposed* criteria. This result can be explained by the mixed origins of some *Spaniard* households and the comparatively limited number of their households in the sample (relative to other European). Please recall that the *Portuguese* and *Spaniards* were the only groups whose number did not increase monotonically in moving from *Origins* to *One drop* and *One drop imposed* (*cf.* Online Appendix II - Table 1).[1]

[1] *Spaniard* households in the sample intermarried with *Italians*; the *Portuguese* intermarried with individuals classified as *Non-Whites*, *Italians*, and, in one case, *Germanics.*

Online Appendix V - Table 1 – Testing for taste-based discrimination in prices earned by ethnolinguistic groups: Alternative classifications for ethnolinguistic origins

	Dep. var. Average price earned for							
	Cultivating one coffee grove (lean season)				Harvesting one *alqueire* (harvest season)			
	Criteria for classifying ethnolinguistic origins				Criteria for classifying ethnolinguistic origins			
	Origins	Origins imp	One drop	One drop imp	Origins	Origins imp	One drop	One drop imp
TFP agr. task	0.00144***	0.00143***	0.00144***	0.00143***	0.00731***	0.00790***	0.00687**	0.00769***
	(0.000279)	(0.000264)	(0.000276)	(0.000263)	(0.00268)	(0.00268)	(0.00268)	(0.00270)
Non-Whites	0.000321	0.000361	0.000587*	0.000541	0.00500	0.00341	0.00755*	0.00399
	(0.000207)	(0.000395)	(0.000303)	(0.000391)	(0.00372)	(0.00364)	(0.00407)	(0.00395)
Italians	-8.60e-05	8.58e-05	4.31e-05	0.000152	0.00443	0.00168	0.00711*	0.00243
	(0.000230)	(0.000265)	(0.000244)	(0.000287)	(0.00396)	(0.00394)	(0.00431)	(0.00429)
Germanics	0.000473	0.000632	0.000623	0.000683	0.000883	-0.000908	0.00314	-0.000266
	(0.000446)	(0.000437)	(0.000431)	(0.000432)	(0.00426)	(0.00411)	(0.00453)	(0.00431)
Portuguese	-0.000269	-0.000188	-0.000142	-0.000131	0.00601	0.00391	0.00858	0.00440
	(0.000254)	(0.000293)	(0.000277)	(0.000329)	(0.00470)	(0.00457)	(0.00546)	(0.00553)
Spaniards	-0.000184	-5.50e-05	-2.96e-05	6.78e-05	-0.0136	-0.0162*	-0.0219***	-0.0258***
	(0.000424)	(0.000446)	(0.000501)	(0.000524)	(0.00930)	(0.00896)	(0.00673)	(0.00637)
East. Europeans	-0.000556	-0.000104	-0.000424	-1.69e-05	0.00342	-0.0189	0.00617	-0.0176
	(0.000733)	(0.000580)	(0.000741)	(0.000590)	(0.00378)	(0.0227)	(0.00418)	(0.0223)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,360	2,360	2,360	2,360	1,470	1,470	1,470	1,470
Number of id	513	513	513	513	324	324	324	324

II. DEMOGRAPHIC CONTROLS

This set of robustness checks expands or replaces the variables in the baselines referring to households' labor composition. The additional variables drawn from the *Registries* include the number of males; the number of married people and widows, with singles as the reference category; and the number of nuclear families, defined as $(\#widows + (\#married/2))$ (*cf.* Section IV). Such variables may capture different (or additional) dimensions of intrahousehold division of labor, *e.g.* gender- and seniority-based behavior, or may be considered as substitute proxies for the number of economically active members. The number of males could capture the effect of having household members hired for seasonal labor within or outside the plantation; power dynamics governing the allocation of household members to different tasks during the lean and harvest seasons; and, most importantly, gender norms that may have differentiated freedmen from immigrants – particularly *Italians* – in the perceived appropriateness of women and children working as agricultural fieldhands (*cf.* Andrews 1988; Monsma 2010; Bassanezi 2019). The numbers of married couples and widows, in turn, indicate the coexistence of more than one nuclear family in the household, which could affect cooperation among household members, or serve as a substitute for household size.

Online Appendix V - Table 2 adds these demographic variables to the baselines one at a time. Online Appendix V - Table 3, in turn, replaces household size with the number of families and the numbers of married people and widows, and the number of economically active household members with the number of males, again one at a time. In both tables, Panel A reports the determinants of earnings in coffee production per task, and Panel B tests for taste-based discrimination. The TFP estimates used in Panel B are obtained from the residuals of the quantity-based productivity regressions, but now using the set of controls from each robustness check.

Crucially, the addition or replacement of demographic controls do not alter the baseline conclusions. In particular, the number of household members in the age range 7-60y old remains highly significant, with stable point estimates across specifications.

The number of nuclear families is not statistically significant as an additional control, indicating neither a benefit of having more families in a household – *e.g.* through the internal division of labor – nor problems of coordination. The same conclusion holds for the specification that controls for the number of married and widowed individuals; however, the number of observations decreases substantially with these controls, rendering even the TFP estimate for the harvest season statistically non-significant, which casts serious doubts on those specifications. Interestingly, the number of males is significant and robust as a control only for coffee cultivation in the lean season; in the harvest season, its significance is considerably less robust to the inclusion of other controls. This result is consistent with the prior that as many household members as possible were engaged in the harvest season, whereas a clearer gender-based division of labor prevailed during the lean season, with men cultivating the coffee groves and women likely working in domestic chores and, more intensively than the men, in subsistence agriculture.

Online Appendix V - Table 2 (Panel A) – Determinants of earnings. Additional demographic controls

	Dep. Var.: $ln(Y_{ht}^{cult})$					Dep. Var.: $ln(Y_{ht}^{harv})$				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
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(# widow)		-0.0926			-0.0942		-0.0649			-0.0671
		(0.0770)			(0.0767)		(0.179)			(0.180)
ln(# male)			0.150***	0.153***	-0.0405			0.000360	0.000468	-0.0421
			(0.0393)	(0.0407)	(0.0869)			(0.0596)	(0.0612)	(0.179)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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

Online Appendix V - Table 2 (Panel B) – Testing for taste-based discrimination in prices earned. Additional demographic controls

Physical output

	Dep. Varl.: $ln(Cult_{ht})$					Dep. Var.: $ln(Harv_{ht})$				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ln(HH size)	0.0408	0.257**	-0.106*	-0.112	0.285*	-0.101	-0.0289	-0.131	-0.109	0.0801
	(0.0547)	(0.121)	(0.0599)	(0.0697)	(0.171)	(0.0819)	(0.230)	(0.0883)	(0.101)	(0.360)
ln(# members 7-60y)	0.331***	0.276***	0.324***	0.327***	0.274***	0.391***	0.416**	0.391***	0.389***	0.416**
	(0.0415)	(0.0816)	(0.0377)	(0.0395)	(0.0815)	(0.0650)	(0.191)	(0.0623)	(0.0657)	(0.195)
ln(# families)	-0.0175			0.00217		-0.00156			0.000845	
	(0.0327)			(0.0337)		(0.0303)			(0.0310)	
ln(# married)		-0.00496			-0.00800		0.0152			-0.00263
		(0.0672)			(0.0694)		(0.147)			(0.144)
ln(# widow)		-0.0803			-0.0811		-0.0267			-0.0307
		(0.0771)			(0.0769)		(0.147)			(0.149)
ln(# male)			0.153***	0.157***	-0.0272			0.0211	0.00768	-0.0980
			(0.0393)	(0.0408)	(0.0850)			(0.0546)	(0.0568)	(0.186)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,627	214	1,635	1,621	213	1,627	214	1,635	1,621	213
R-squared	0.816	0.877	0.816	0.816	0.878	0.816	0.877	0.816	0.816	0.878
Number of id	332	59	335	332	58	332	59	335	332	58

Discrimination in pay

	Dep. Var.: Average price earned for									
	Cultivating one coffee grove (lean season)					Harvesting one *alqueire* (harv. season)				
TFP agr. task	0.00151***	0.00119**	0.00148***	0.00150***	0.00118**	0.00729***	-0.00272	0.00737***	0.00736***	-0.00335
	(0.00232)	(0.000517)	(0.000259)	(0.000238)	(0.000508)	(0.00271)	(0.00353)	(0.00268)	(0.00271)	(0.00376)
Non-Whites	0.000345*	0.00150*	0.00311	0.000339	0.00150*	0.00499	-0.00862	0.00507	0.00507	-0.00867
	(0.000207)	(0.000838)	(0.000209)	(0.000210)	(0.000840)	(0.00375)	(0.0132)	(0.00376)	(0.00378)	(0.0132)
Italians	-8.16e-05	0.00108	-9.45e-05	-7.39e-05	0.00108	0.00444	-0.00809	0.00448	0.00448	-0.00794
	(0.000225)	(0.000792)	(0.000229)	(0.000227)	(0.000799)	(0.00399)	(0.0129)	(0.00400)	(0.00402)	(0.0128)
Germanics	0.000490	0.00153	0.000445	0.000468	0.00154	0.00841	0.00260	0.000961	0.000938	0.000283
	(0.000447)	(0.00115)	(0.000450)	(0.000451)	(0.00116)	(0.0116)	(0.00427)	(0.00429)	(0.00116)	(0.00427)
Portuguese	-0.000187	0.000904	-0.000271	-0.000187	0.000916	0.00605	0.00763	0.00606	0.00609	0.00769
	(0.000254)	(0.000927)	(0.000256)	(0.000257)	(0.000929)	(0.00473)	(0.0115)	(0.00472)	(0.00475)	(0.0115)
Spaniards	-0.000139	0.000481	-0.000204	-0.000147	0.000493	-0.0136	-0.00798	-0.0138	-0.0137	-0.00800
	(0.000414)	(0.000734)	(0.000430)	(0.000423)	(0.000743)	(0.00935)	(0.0120)	(0.00927)	(0.00931)	(0.0120)
East. Europeans.	-0.000597		-0.000496	-0.000533		0.00341		0.00352	0.00344	
	(0.000765)		(0.000748)	(0.000786)		(0.00381)		(0.00381)	(0.00385)	
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,338	293	2,351	2,329	292	1,464	196	1,466	1,460	195
Number of id	508	79	512	507	78	322	57	324	322	56

Online Appendix V - Table 3 (Panel A) – Determinants of earnings. Exchanging demographic controls

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

Online Appendix V - Table 3 (Panel B) – Testing for taste-based discrimination in prices earned. Exchanging demographic controls

Physical output

	Dep. Var.: $ln(Cult_{ht})$					Dep. Var.: $ln(Harv_{ht})$				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ln(HH size)			0.178***					0.231***		
			(0.0506)					(0.0682)		
ln(# members 7-60y)	0.355***	0.375***				0.334***	0.402***			
	(0.0309)	(0.0537)				(0.0381)	(0.136)			
ln(# families)	-0.00787		-0.00301			-0.0250				-0.000739
	(0.0295)		(0.0307)			(0.0288)				(0.0316)
ln(# married)		0.0596		0.131		0.00984				0.123
		(0.0720)		(0.0885)		(0.145)				(0.153)
ln(# widow)		-0.0742		0.0236		-0.0262				0.0898
		(0.0802)		(0.0869)		(0.147)				(0.144)
ln(# male)			0.164***	0.284***	0.220***			0.0350	0.190***	0.129
			(0.0447)	(0.0301)	(0.0764)			(0.0601)	(0.0421)	(0.127)
Full set of controls	Yes	Yes	Yes	Yes	Yey	Yes	Yes	Yes	Yes	Yes
Observations	2,336	292	2,397	2,348	291	1,464	196	1,492	1,476	195
Number of id	507	78	537	507	77	322	57	334	324	56
R-squared	0.862	0.903	0.846	0.844	0.873	0.844	0.890	0.833	0.831	0.878

Discrimination in pay

	Dep. Var.: Average price earned for									
	Cultivating one coffee grove (lean season)					Harvesting one *alqueire* (harv. season)				
TFP agr. task	0.00152***	0.00125***	0.00128***	0.00125***	0.00137**	0.00725***	-0.00287	0.00648***	0.00626***	-0.00307
	(0.000229)	(0.000484)	(0.000228)	(0.000207)	(0.000640)	(0.000271)	(0.000362)	(0.000239)	(0.000229)	(0.000367)
Non-Whites	0.000344**	0.00155**	0.000308	0.000343	0.00158*	0.00498	-0.00862	0.00546	0.00552	-0.00900
	(0.000207)	(0.000834)	(0.000197)	(0.000214)	(0.000640)	(0.00375)	(0.0132)	(0.00359)	(0.00364)	(0.0134)
Italians	-8.81e-05	0.00108	-0.000106	-9.37e-05	0.00110	0.00448	-0.00808	0.00465	0.00460	-0.00819
	(0.000202)	(0.000794)	(0.000217)	(0.000228)	(0.000783)	(0.00398)	(0.0129)	(0.00388)	(0.00401)	(0.0130)
Germanics	0.000493	0.00152	0.000381	0.000401	0.00153	0.00779	0.000258	0.00966	0.000571	0.000236
	(0.000447)	(0.00115)	(0.000445)	(0.000458)	(0.00115)	(0.00429)	(0.0116)	(0.00416)	(0.00401)	(0.00117)
Portuguese	-0.000189	0.000905	-0.000306	-0.000225	0.000910	0.00608	0.00766	0.00619	0.00633	0.00735
	(0.000253)	(0.000925)	(0.000225)	(0.000259)	(0.000918)	(0.00473)	(0.0115)	(0.00463)	(0.00475)	(0.0117)
Spaniards	-0.000146	0.000446	-0.000228	-0.000134	0.000602	-0.0136	-0.00800	-0.0141	-0.0132	-0.00812
	(0.000413)	(0.000744)	(0.000436)	(0.000259)	(0.000719)	(0.00934)	(0.0120)	(0.00901)	(0.00932)	(0.0122)
East. Europeans.	-0.000596		-0.000518	-0.000584		0.00341		0.00290	0.00334	
	(0.000765)		(0.000657)	(0.000746)		(0.00381)		(0.00377)	(0.00386)	
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,338	293	2,399	2,350	292	1,464	196	1,492	1,476	195
Number of id	508	79	538	508	78	322	57	334	324	56

The partial effect of the number of members in the age range 7-60y has proved highly robust to changes in the demographic composition of households. The next robustness checks vary the age windows used to define economically active household members, substituting the variable *ln(# members 7-60y)* by the natural log of the number of members aged 17-45y, 17y+, 10-45y, 10y+,[2] 7-45y, and 7y+. The pairwise correlations between these age windows range from 0.7325 (17-45y & 7-60y) to 0.9796 (7y+ & 7-60), all significant at the 1% level (Online Appendix V - Table 4).

All age ranges are significant at the 1% level, indicating that each captures a substantial part of household labor composition (Online Appendix V - Table 5 - Panel A). In addition, none of them affect the conclusions on the absence of taste-based discrimination in remuneration (Panel B). The coefficients for each age range, however, vary considerably. For the lean and harvest seasons, the two smallest point estimates are those for the 17-45y and 17y+ age ranges, respectively. Consonant with the literature and descriptive evidence from the *Ibicaba Collection*, these ranges understate the number of economically active individuals: the first is too restrictive relative to the actual age span of the labor force, and the latter implies too late an age of entry. Interestingly, the largest point estimates for both seasons correspond to the 10y+ definition. Between these cases, the 10-45y and 7-45y age ranges are very close to the baseline estimates. The first underscores the complementary labor supplied by children in the household (in contrast to the 17-45y range), while the second reiterates that the resident labor force at *Ibicaba* was indeed quite young, with a pooled mean of maximum ages across households of about 48 years (*cf.* Section IV).

The 7-60y age range appears nonetheless to be preferable for three reasons. First, it avoids overstating the labor force participation of older individuals, whose age heaping may have been substantial. Second, it takes into consideration the labor force participation of children around 7, who certainly took part at least in domestic chores. Third, it preserves data consistency, given that the 10y+ range draws additional observations from summary tables not used in the compilation of the other variables.[3]

[2] The 10y+ variable draws on additional demographic data contained in a summary table for 1910, covering men and women recorded simply as younger or older than 10. In these cases, the data on household size and age composition were also derived from these summary tables. In principle, the 10y+ variable could therefore show discrepancies with the other data on labor composition. However, if the additional information from 1910 is dropped, the resulting variable (*enxada04_test* in the data compilation files) remains almost identical to the 10y+ variable itself (*enxada04*).

[3] Moreover, that these alternative age ranges are either too narrow or too broad is shown by the fact that they render *household size* statistically significant, with coefficients varying so widely as to change sign across specifications.

Online Appendix V - Table 4 – Pairwise correlations between number of household members in the corresponding age windows

	17-45y	17+y	10-45y	10+y	7-45y	7+y	17-45y
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	
17-45y	0.7325*	0.7483*	0.9398*	0.9350*	0.9654*	0.9796*	1.0000

Note: * indicates significance at 1%-level.

Online Appendix V Table 5 (Panel A) – Determinants of earnings. Different age ranges defining economically active household members

	$\ln(Y_{ht}^{cult})$						$\ln(Y_{ht}^{harv})$					
	Age ranges for (# active members)						Age ranges for (# active members)					
	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(# active members)	0.131***	0.172***	0.356***	0.440***	0.316***	0.436***	0.807***	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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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
Number of id	501	512	507	512	507	512	313	326	319	326	319	326
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

Online Appendix V - Table 5 (Panel B) – Testing for taste-based discrimination in prices earned. Different age ranges defining economically active household members

	Physical output											
	Dep. Var.: $ln(Cult_{ht})$						Dep. Var.: $ln(Harv_{ht})$					
	Age ranges for (# active members)						Age ranges for (# active members)					
	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)	0.307***	0.245***	0.0729*	0.0176	0.0633	-0.0267	0.275***	0.183***	-0.0418	-0.129**	-0.0318	-0.181**
	(0.0396)	(0.0422)	(0.0409)	(0.0394)	(0.0473)	(0.0520)	(0.0567)	(0.0564)	(0.0668)	(0.0574)	(0.0811)	(0.0793)
ln(# active members)	0.133***	0.170***	0.358***	0.440***	0.315***	0.431***	0.0796**	0.193***	0.397***	0.560***	0.325***	0.540***
	(0.0262)	(0.0369)	(0.0312)	(0.0399)	(0.0386)	(0.0536)	(0.0386)	(0.0491)	(0.0503)	(0.0585)	(0.0615)	(0.0838)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,270	2,378	2,309	2,380	2,311	2,380	1,580	1,658	1,609	1,659	1,611	1,659
Number of id	502	513	508	513	508	513	324	337	330	337	330	337
R-squared	0.466	0.441	0.536	0.518	0.496	0.487	0.806	0.807	0.821	0.823	0.813	0.816

	Discrimination in pay											
	Dep. Var: Average price earned for											
	Cultivating one coffee grove (lean season)						Harvesting one *alqueire* (harv. season)					
TFP agr. task	0.00128***	0.00139***	0.00144***	0.00151***	0.00143***	0.00148***	0.00666**	0.00702***	0.00763**	0.00810***	0.00704**	0.00704**
	(0.000234)	(0.000234)	(0.000259)	(0.000271)	(0.000248)	(0.000257)	(0.00285)	(0.00268)	(0.00301)	(0.00293)	(0.00278)	(0.00278)
Non-Whites	0.000332	0.000362*	0.000303	0.000351*	0.000294	0.000336	0.00548	0.00541	0.00480	0.00518	0.00489	0.00508
	(0.000211)	(0.000211)	(0.000209)	(0.000209)	(0.000208)	(0.000208)	(0.00386)	(0.00370)	(0.00378)	(0.00375)	(0.00374)	(0.00372)
Italians	-5.47e-05	-7.77e-05	-6.36e-05	-5.43e-05	-7.91e-05	-7.51e-05	0.00504	0.00466	0.00444	0.00468	0.00444	0.00462
	(0.000222)	(0.000225)	(0.000226)	(0.000225)	(0.000228)	(0.000226)	(0.00412)	(0.00395)	(0.00401)	(0.00392)	(0.00402)	(0.00394)
Germanics	0.000397	0.000493	0.000469	0.000523	0.000452	0.000497	0.000822	0.000961	0.000751	0.00134	0.000605	0.00104
	(0.000403)	(0.000443)	(0.000440)	(0.000441)	(0.000441)	(0.000442)	(0.00444)	(0.00427)	(0.00430)	(0.00431)	(0.00430)	(0.00429)
Portuguese	-0.000156	-0.000202	-0.000240	-0.000245	-0.000246	-0.000255	0.00782*	0.00631	0.00617	0.00623	0.00621	0.00617
	(0.000253)	(0.000252)	(0.000253)	(0.000248)	(0.000255)	(0.000250)	(0.00421)	(0.00469)	(0.00477)	(0.00469)	(0.00476)	(0.00468)
Spaniards	-2.00e-05	-0.000117	-0.000106	-0.000203	-0.000105	-0.000193	-0.0135	-0.0133	-0.0143	-0.0142	-0.0140	-0.0138
	(0.000470)	(0.000422)	(0.000462)	(0.000422)	(0.000458)	(0.000419)	(0.00950)	(0.00925)	(0.00939)	(0.00931)	(0.00942)	(0.00938)
East. Europeans.	-0.000641	-0.000664	-0.000516	-0.000539	-0.000562	-0.000592	0.00301	0.00266	0.00292	0.00344	0.00304	0.00359
	(0.000700)	(0.000731)	(0.000696)	(0.000762)	(0.000699)	(0.000768)	(0.00408)	(0.00388)	(0.00391)	(0.00374)	(0.00390)	(0.00376)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,270	2,378	2,309	2,380	2,311	2,380	1,416	1,485	1,441	1,486	1,443	1,486
Number of id	502	513	508	513	508	513	313	326	319	326	319	326

Much of the work behind this paper was devoted to converting historical accounting data into a harmonized dataset that matches information from sources across the *Ibicaba Collection*. As discussed in Section III and Online Appendix I, an obvious difficulty is that these sources were produced for practical managerial and accounting purposes, which entailed numerous glosses and corrections – at times indecipherable to the modern reader. Among these are the crossing-out of some household members in the *Registries*. What happened to those individuals cannot be determined with certainty from the ledgers alone: a crossed-out person could have moved out, married into a different household, or passed away –, or the new entry could simply reflect the correction of a previously erroneous family record in the *Registries*.

The next robustness checks exclude from the sample all crossed-out entries in the sources that were counted as household members in the previous estimates.

The specification corresponding to the baseline, but now excluding crossed-out individuals, is reported in Online Appendix V - Table 6. The specifications using the alternative age ranges, but likewise excluding crossed-out individuals, is reported in Online Appendix V - Table 7. Reassuringly, the results remain robust to the exclusion of household members crossed out in the sources, which further increases confidence in the methodology developed in this paper for cleaning and harmonizing *Ibicaba*'s managerial and accounting records.

Online Appendix V - Table 6 (Panel A) – Determinants of earnings. Excluding households crossed-out in the sources

	Dep. Var.: $ln(Y_{ht}^{cult})$					Dep. Var.: $ln(Y_{ht}^{harv})$				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ln(HH size) - check			0.175***					0.259***		
			(0.0503)					(0.0737)		
ln(# members 7-60y) - check	0.353***	0.369***				0.349***	0.446***			
	(0.0306)	(0.0551)				(0.0409)	(0.149)			
ln(# families) - check	-0.00800			-0.00227		-0.0150			0.00508	
	(0.0288)			(0.0299)		(0.0299)			(0.0321)	
ln(# married) - check		0.0546			0.128		-0.0286			0.0987
		(0.0725)			(0.0891)		(0.152)			(0.160)
ln(# widow) - check		-0.0918			0.00520		-0.0671			0.0535
		(0.0799)			(0.0861)		(0.182)			(0.187)
ln(# male) - check			0.158***	0.277***	0.207**			0.00986	0.184***	0.142
			(0.0450)	(0.0305)	(0.0786)			(0.0658)	(0.0457)	(0.123)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,334	292	2,395	2,346	291	1,462	196	1,490	1,474	195
Number of id	507	78	537	507	77	322	57	334	324	56
R-squared	0.862	0.903	0.845	0.844	0.873	0.844	0.890	0.833	0.831	0.877

Online Appendix V - Table 6 (Panel B) – Testing for taste-based discrimination in prices. Excluding households crossed-out in the sources

Physical output

	Dep. Var.: $ln(Cult_{ht})$					Dep. Var.: $ln(Harv_{ht})$				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
ln(HH size) - check			0.180***					0.231***		
			(0.0506)					(0.0684)		
ln(# members 7-60y) - check	0.354***	0.378***				0.332***	0.403***			
	(0.0308)	(0.0529)				(0.0381)	(0.138)			
ln(# families) - check	-0.00758			-0.00243		-0.0237			0.00182	
	(0.0295)			(0.0307)		(0.0287)			(0.0315)	
ln(# married) - check		0.0579			0.131		0.00949			0.124
		(0.0716)			(0.0882)		(0.145)			(0.153)
ln(# widow) - check		-0.0777			0.0215		-0.0290			0.0893
		(0.0801)			(0.0872)		(0.148)			(0.144)
ln(# male) - check			0.160***	0.280***	0.220***			0.0325	0.186***	0.124
			(0.0448)	(0.0302)	(0.0764)			(0.0600)	(0.0425)	(0.126)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,336	293	2,398	2,348	292	1,625	214	1,662	1,637	213
Number of id	508	79	539	508	78	332	59	346	334	58
R-squared	0.496	0.640	0.437	0.427	0.529	0.815	0.877	0.802	0.800	0.864

Discrimination in pay

	Dep. Var.: Average price earned for									
	Cultivating one coffee grove (lean season)					Harvesting one *alqueire* (harv. season)				
TFP agr. task	0.00152***	0.00125***	0.00128***	0.00124***	0.00138**	0.00724***	-0.00284	0.00647***	0.00626***	-0.00308
	(0.000229)	(0.000483)	(0.000227)	(0.000206)	(0.000640)	(0.00270)	(0.00359)	(0.00239)	(0.00229)	(0.00366)
Non-Whites	0.000344*	0.00155*	0.000310	0.000345	0.00159*	0.00500	-0.00863	0.00547	0.00553	-0.00902
	(0.000207)	(0.000834)	(0.000197)	(0.000214)	(0.000853)	(0.00375)	(0.0132)	(0.00359)	(0.00375)	(0.0134)
Italians	-8.84e-05	0.00103	-0.000106	-9.33e-05	0.00110	0.00448	-0.00809	0.00466	0.00461	-0.00818
	(0.000224)	(0.000794)	(0.000217)	(0.000228)	(0.000783)	(0.00398)	(0.0129)	(0.00388)	(0.00401)	(0.0130)
Germanics	0.000491	0.00152	0.000381	0.000402	0.00153	0.000788	0.000258	0.000982	0.000592	0.000238
	(0.000447)	(0.00115)	(0.000445)	(0.000458)	(0.00115)	(0.00429)	(0.0116)	(0.00415)	(0.00431)	(0.0117)
Portuguese	-0.000189	0.000907	-0.000306	-0.000224	0.000911	0.00608	0.00766	0.00619	0.00633	0.00735
	(0.000253)	(0.000925)	(0.000248)	(0.000259)	(0.000918)	(0.00473)	(0.0115)	(0.00463)	(0.00475)	(0.0117)
Spaniards	-0.000146	0.000446	-0.000229	-0.000133	0.000600	-0.0136	-0.00801	-0.0141	-0.0132	-0.00812
	(0.000413)	(0.000744)	(0.000436)	(0.000436)	(0.000719)	(0.00933)	(0.0120)	(0.00900)	(0.00932)	(0.0122)
East. Europeans	-0.000594		-0.000521	-0.000582		0.00334		0.00281	0.00326	
	(0.000772)		(0.000665)	(0.000751)		(0.00382)		(0.00379)	(0.00388)	
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,336	293	2,397	2,348	292	1,462	196	1,490	1,474	195
Number of id	508	79	538	508	78	322	57	334	324	56

Online Appendix V - Table 7 (Panel A) – Determinants of earnings. Different age ranges defining economically active household members, excluding households crossed-out in the sources

	Dep. Var.: $ln(Y_{ht}^{cult})$							Dep. Var.: $ln(Y_{ht}^{harv})$						
	Age ranges for (# active members)							Age ranges for (# active members)						
	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) - check	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(# active members) - check	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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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
Number of id	501	512	507	512	507	512	512	313	326	319	326	319	326	324
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

Online Appendix V - Table 7 (Panel B) – Testing for taste-based discrimination in prices earned. Different age ranges defining economically active household members, excluding households crossed-out in the sources

Physical output

	Dep. Var.: $ln(Cult_{ht})$							Dep. Var.: $ln(Harv_{ht})$						
	Age ranges for(# active members)							Age ranges for (# active members)						
	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) - check	0.307***	0.244***	0.0716*	0.0175	0.0620	-0.0272	0.0269	0.273***	0.182***	-0.0425	-0.127**	-0.0333	-0.181**	-0.115
	(0.0395)	(0.0421)	(0.0409)	(0.0394)	(0.0475)	(0.0522)	(0.0484)	(0.0574)	(0.0569)	(0.0669)	(0.0577)	(0.0813)	(0.0794)	(0.0736)
ln(#active members) - check	0.132***	0.169***	0.359***	0.440***	0.316***	0.430***	0.333***	0.0757*	0.189***	0.394***	0.554***	0.323***	0.536***	0.393***
	(0.0263)	(0.0370)	(0.0313)	(0.0399)	(0.0387)	(0.0537)	(0.0402)	(0.0387)	(0.0493)	(0.0502)	(0.0584)	(0.0611)	(0.0836)	(0.0613)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,268	2,376	2,307	2,378	2,309	2,378	2,358	1,578	1,656	1,607	1,657	1,609	1,657	1,639
Number of id	502	513	508	513	508	513	513	324	337	330	337	330	337	335
R-squared	0.466	0.441	0.536	0.518	0.496	0.487	0.494	0.806	0.807	0.820	0.823	0.813	0.815	0.816

Discrimination in pay

	Dep. Var.: Average price earned for													
	Cultivating one coffee grove (lean season)							Harvesting one *alqueire* (harv. season)						
TFP agr. task	0.00127***	0.00138***	0.00144***	0.00151***	0.00143***	0.00148***	0.00151***	0.00663**	0.00699***	0.00759**	0.00806***	0.00701**	0.00717***	0.00728***
	(0.000233)	(0.000234)	(0.000258)	(0.000270)	(0.000247)	(0.000256)	(0.000252)	(0.00283)	(0.00266)	(0.00299)	(0.00291)	(0.00276)	(0.00273)	(0.00267)
Non-Whites	0.000333	0.000362*	0.000305	0.000352*	0.000295	0.000337	0.000321	0.00550	0.00543	0.00484	0.00522	0.00492	0.00511	0.00501
	(0.000211)	(0.000211)	(0.000209)	(0.000208)	(0.000208)	(0.000208)	(0.000206)	(0.00386)	(0.00370)	(0.00378)	(0.00375)	(0.00374)	(0.00372)	(0.00372)
Italians	-5.54e-05	-7.86e-05	-6.44e-05	-5.54e-05	-7.99e-05	-7.60e-05	-0.000103	0.00505	0.00466	0.00445	0.00468	0.00445	0.00462	0.00443
	(0.000222)	(0.000225)	(0.000226)	(0.000225)	(0.000228)	(0.000227)	(0.000226)	(0.00412)	(0.00395)	(0.00401)	(0.00392)	(0.00402)	(0.00394)	(0.00397)
Germanics	0.000395	0.000492	0.000465	0.000520	0.000449	0.000494	0.000468	0.000833	0.000967	0.000752	0.00133	0.000609	0.00104	0.000886
	(0.000403)	(0.000443)	(0.000440)	(0.000441)	(0.000442)	(0.000442)	(0.000445)	(0.00444)	(0.00427)	(0.00430)	(0.00431)	(0.00429)	(0.00429)	(0.00426)
Portuguese	-0.000156	-0.000202	-0.000240	-0.000245	-0.000246	-0.000255	-0.000273	0.00783*	0.00632	0.00619	0.00623	0.00622	0.00617	0.00601
	(0.000253)	(0.000252)	(0.000253)	(0.000248)	(0.000255)	(0.000250)	(0.000253)	(0.00421)	(0.00469)	(0.00477)	(0.00469)	(0.00476)	(0.00468)	(0.00470)
Spaniards	-2.12e-05	-0.000118	-0.000107	-0.000204	-0.000106	-0.000194	-0.000198	-0.0135	-0.0133	-0.0143	-0.0142	-0.0140	-0.0138	-0.0136
	(0.000470)	(0.000422)	(0.000462)	(0.000422)	(0.000458)	(0.000419)	(0.000420)	(0.00949)	(0.00925)	(0.00939)	(0.00930)	(0.00942)	(0.00937)	(0.00930)
East. Europeans	-0.000643	-0.000663	-0.000518	-0.000538	-0.000564	-0.000591	-0.000563	0.00292	0.00258	0.00284	0.00339	0.00297	0.00355	0.00338
	(0.000706)	(0.000737)	(0.000704)	(0.000771)	(0.000706)	(0.000776)	(0.000736)	(0.00410)	(0.00390)	(0.00393)	(0.00376)	(0.00392)	(0.00378)	(0.00379)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,268	2,376	2,307	2,378	2,309	2,378	2,358	1,414	1,483	1,439	1,484	1,441	1,484	1,468
Number of id	502	513	508	513	508	513	513	313	326	319	326	319	326	324

III. ALTERNATIVE CONTROLS FOR *ESTACAS*

As with crossed-out individuals, one of the main difficulties in compiling the final dataset was to decide whether crossed-out *estacas* and coffee groves in year *t* should be counted as agricultural capital of the household in that same year, since it cannot be determined whether the crossing-out was binding for year *t* or later (*cf.* Section III and Online Appendix 01).

The baselines use variables constructed from *estacas* and coffee groves that could be reliably assigned to year *t* on the basis of the total number of coffee groves reported in the sources (variable $Cult_{ht}$). The robustness checks will now consider three alternative variables. Dep. Var.: $Cult\ check_{ht}$ is a lower-bound estimate that assumes all crossing-outs of *estacas* or coffee groves to be binding in year *t*, *i.e.* it deletes all *estacas* and coffee groves crossed out in year *t*, while still benchmarking them against the sum of checked totals reported in the sources. The other two variables ignore the totals recorded in the sources and consider only individual entries, regardless of whether these sum up to the totals reported in the sources. *tree_tot_estaca_check* sums all coffee groves recorded for household *h* in year *t* and deletes all entries from *estacas* crossed out in that year. *tree_tot_estaca*, in turn, is an upper-bound estimate of the number of coffee groves assigned to households: it sums all coffee groves from each *estaca* for which data are available, including those crossed out in the sources and those in accounts nullified by *Ibicaba*'s managers.

Online Appendix V - Table 8 reports descriptive statistics and pairwise correlations for these variables. Differences notwithstanding, variations between them are minimal. The smallest correlation coefficients is 0.975, and all pairwise correlations are significant at the 1% level. Consequently, results shown in Online Appendix V - Table 9 are fully in line with the baseline results, with only minor variations in the significance of individual colonies and measures of *estacas* dispersion.[4]

[4] These specifications substitute each alternative metric for the baseline *tree_tot_check* as the dependent variable to estimate households' per task productivity in the lean season, as well as taste-based discrimination in prices earned (remuneration).

Online Appendix V - Table 8 – Descriptive statistics & Pairwise correlations: Alternative metrics for number of cultivated coffee groves & *Estacas*

	Descriptive statistics						
	N	n	Mean	S.D.	Median	Min	Max
$Cult_{ht}$	3,147	755	6,377,434	3,004,643	5,704	537	28,096
$Cult\ check_{ht}$	3,092	750	6,290,625	3,041,135	5,621	397	28,096
Tree_tot_estaca_check	3,077	749	6,352,585	3,034,916	5,663	517	28,096
Tree_tot_estaca	3,130	754	6,448,379	3,020,614	5,774	537	28,096

	Pairwise correlations			
	$Cult_{ht}$	$Cult_{ht}$	tree_tot_estaca_check	tree_tot_estaca
$Cult_{ht}$	1.0000			
$Cult\ check_{ht}$	0.9876*	1.0000		
tree_tot_estaca_check	0.9863*	0.9941*	1.0000	
tree_tot_estaca	0.9931*	0.9750*	0.9838*	1.0000

Note: * indicates significance at the 1% level.

Online Appendix V - Table 9 – Testing for taste-based discrimination in prices earned. Alternative metrics for the number of cultivated coffee groves & *Estacas*

	Physical output			
	Dep. Var.:			
	$ln(Cult_{ht})$	$ln(Cult\ check_{ht})$	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)
Estaca ID Mean	2.45e-06	3.67e-06	-3.48e-06	-8.37e-06
	(2.64e-05)	(2.72e-05)	(2.68e-05)	(2.74e-05)
Estaca ID MDEV	-4.84e-05	4.34e-05	-0.000154*	-0.000132
	(8.69e-05)	(0.000115)	(9.15e-05)	(9.69e-05)
Estaca ID S.D.	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. & Pinheiro	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)
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	-0.105	-0.0982	-0.107	-0.108
	(0.0651)	(0.0672)	(0.0659)	(0.0694)
Full set of controls	Yes	Yes	Yes	Yes
Observations	2,360	2,316	2,359	2,317
Number of id	513	506	512	505
R-squared	0.494	0.473	0.471	0.451

	Discrimination in pay			
	Dep. Var.: Average price earned for cultivating one coffee grove (lean season)			
	TFPs computed for:			
	$ln(Cult_{ht})$	$ln(Cult\ check_{ht})$	ltree_tot_estaca	ltree_tot_estaca_check
TFP agr. task	0.00151***	0.00153***	0.00138***	0.00143***
	(0.000252)	(0.000225)	(0.000278)	(0.000259)
Non-Whites	0.000321	0.000277	0.000335	0.000313
	(0.000206)	(0.000207)	(0.000208)	(0.000209)
Italians	-0.000102	-0.000113	-7.39e-05	-7.88e-05
	(0.000226)	(0.000225)	(0.000231)	(0.000230)
Germanics	0.000470	0.000464	0.000474	0.000480
	(0.000445)	(0.000445)	(0.000447)	(0.000446)
Portuguese	-0.000273	-0.000256	-0.000270	-0.000249
	(0.000253)	(0.000254)	(0.000255)	(0.000255)
Spaniards	-0.000198	-0.000149	-0.000181	-0.000132
	(0.000420)	(0.000419)	(0.000428)	(0.000426)
East. Europeans	-0.000564	-0.000605	-0.000543	-0.000572
	(0.000728)	(0.000730)	(0.000737)	(0.000735)
Full set of controls	Yes	Yes	Yes	Yes
Observations	2,360	2,316	2,359	2,317
Number of id	513	506	512	505

Beyond the number of coffee groves cultivated, the *Registries* include other variables that can proxy for households' physical yields during the lean season. The most obvious is the number of *estacas* itself. The variable *Estaca Count* is defined as the number of *estacas* cultivated by household *h* in year *t*, rather than the number of coffee groves in them.[5] Similarly, *Estaca Count - Check* corresponds to *Estaca Count* net of all crossed-out groves.

The average household worked about six *estacas* (S.D. = 2.8), with only minimal variations between the two metrics referred to above. By definition, these two proxies are also closely related to the baseline's dependent variable, as further confirmed by their pairwise correlations (Online Appendix - Table 10). It is nonetheless worth checking the sensitivity of the results to using them as dependent variables, particularly in light of the argument that the number of coffee groves cultivated per household was not simply a linear function of the number of economically active household members (*cf.* Section V of the main text).

The conclusions are unaffected if the number of *estacas* replaces that of cultivated coffee groves (Online Appendix V - Table 11). The significance of the measures relating to the location of *estacas* change, however. The mean of *estaca* IDs becomes statistically significant, while indicators for various colonies become non-significant. In other words, the colonies affected the number of coffee groves cultivated, but not the number of *estacas*. *Ceteris paribus*, this implies that the number of coffee groves per *estaca* must have varied substantially across colonies, which is again consistent with the evidence discussed in Section IV of the main text.

The next set of robustness checks return to the number of cultivated coffee groves as the dependent variable and vary instead the independent variables relating to *estacas*. First, more parsimonious specifications than the baseline's are adopted, controlling for only one measure of *estaca* ID dispersion at a time to limit potential problems with multicollinearity. Second, MAD replaces the MDEV indicators of *estaca* IDs used in the baselines. Finally, each of these steps is repeated with the alternative variable that excludes crossed-out *estacas* ("-check" variables). The results reported Online Appendix - Table 12 are robust to these alternative specifications and variables. Interestingly, the effects of *estaca* IDs and of the colonies remain almost unchanged in the more parsimonious models. Replacing the MDEV with the *estaca* MAD identifier shifts the significance levels of a few colonies, but minorly. Finally, the "-check" variables, which exclude crossed-out *estacas*, likewise leave the results unchanged, further strengthening confidence in the protocols adopted to generate the dataset.

[5] *Estaca Count* was compiled by counting the number of unique IDs assigned to each *estaca*, as a few were unlabeled in the original sources. Summary tables from the sources were excluded, as they do not identify the number of *estacas*.

Online Appendix V - Table 10 – Descriptive statistics & Pairwise correlations: Number of coffee groves & *Estacas* cultivated per household

Descriptive statistics

	N	n	Mean	S.D.	Median	Min	Max
tree_tot_check	3,147	755	6,377.434	3,004.643	5,704	537	28,096
estaca_count	2,940	620	6.179	2.824	6	1	20
estaca_count_check	2,836	612	6.120	2.845	6	1	20

Pairwise correlations

	tree_tot_check	estaca_count	estaca_count_check
tree_tot_check	1.0000		
estaca_count	0.9233*	1.0000	
estaca_count_check	0.9234*	0.9927*	1.0000

Note: * indicates significance at 1%-level.

Online Appendix V - Table 11 – Physical output and prices earned by different ethnolinguistic groups. Alternative metrics for number of cultivated coffee groves & *Estacas*

	Physical output		
	Dep. Var.:		
	$ln(Cult_{ht})$	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)
Estaca ID Mean	2.45e-06	-0.000183***	-0.000182***
	(2.64e-05)	(2.27e-05)	(2.34e-05)
Estaca ID MDEV	-4.84e-05	2.63e-05	0.000142*
	(8.69e-05)	(7.93e-05)	(8.02e-05)
Estaca ID S.D.	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. & Pinheiro	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)
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	-0.105	0.0904	0.0903
	(0.0651)	(0.111)	(0.111)
Full set of controls	Yes	Yes	Yes
Observations	2,360	2,359	2,294
Number of id	513	512	504
R-squared	0.494	0.448	0.471

	Discrimination in pay		
	Dep. var.: Avg. price earned for cultivating one coffee grove (lean season)		
	TFPs computed for:		
	$ln(Cult_{ht})$	lestaca_count	lestaca_count_check
TFP agr. task	0.00151***	0.00145***	0.00155***
	(0.000252)	(0.000216)	(0.000216)
Non-Whites	0.000321	0.000314	0.000304
	(0.000206)	(0.000209)	(0.000210)
Italians	-0.000102	-7.30e-05	-9.91e-05
	(0.000226)	(0.000225)	(0.000226)
Germanics	0.000470	0.000433	0.000427
	(0.000445)	(0.000448)	(0.000450)
Portuguese	-0.000273	-0.000264	-0.000261
	(0.000253)	(0.000256)	(0.000259)
Spaniards	-0.000198	-0.000105	-0.000103
	(0.000420)	(0.000420)	(0.000422)
East. Europeans	-0.000564	-0.000783	-0.000850
	(0.000728)	(0.000696)	(0.000681)
Full set of controls	Yes	Yes	Yes
Observations	2,360	2,359	2,294
Number of id	513	512	504

Online Appendix V - Table 12 (Panel A) – Determinants of earnings (lean season). Alternative measures for *Estacas* IDs

	Dep. Var.: $ln(Y_{ht}^{cult})$									
	(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 ID 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 ID MDEV	-4.53e-05	3.39e-06								
	(8.50e-05)	(2.51e-05)								
Estaca ID S.D.	3.34e-05		2.55e-06						0.000127***	
	(6.03e-05)		(1.77e-05)						(3.70e-05)	
Estaca ID 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 ID MDEV - check				-4.09e-05	-1.22e-06					
				(0.000115)	(2.62e-05)					
Estaca ID S.D.- check				2.72e-05		-1.52e-06				0.000137***
				(8.47e-05)		(1.91e-05)				(4.16e-05)
Estaca ID MAD							9.63e-05		0.000588***	
							(7.66e-05)		(0.000175)	
Estaca ID 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. & Pinheiro	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)
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	-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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,358	2,364	2,358	2,293	2,299	2,293	2,364	2,299	2,358	2,293
Number of id	512	512	512	504	504	504	512	504	512	504
R-squared	0.861	0.854	0.861	0.863	0.855	0.863	0.854	0.855	0.863	0.864

Online Appendix V - Table 12 (Panel A ctd.) – Determinants of earnings (harv. season). Alternative measures for *Estacas* IDs

	Dep. Var.: $ln(Y_{ht}^{harv})$									
	(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 ID 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 ID MDEV	5.14e-05	-2.04e-05								
	(0.000120)	(4.11e-05)								
Estaca ID S.D.	-5.06e-05		-1.55e-05						7.33e-05	
	(0.000101)		(3.01e-05)						(8.41e-05)	
Estaca ID 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 ID MDEV - check				-5.65e-05	-3.88e-05					
				(0.000207)	(4.46e-05)					
Estaca ID S.D.- check				1.29e-05		-2.71e-05				0.000113
				(0.000159)		(3.35e-05)				(8.55e-05)
Estaca ID MAD							0.000125		0.000419	
							(0.000111)		(0.000327)	
Estaca ID 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. & Pinheiro	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)
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)
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	-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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,470	1,470	1,470	1,447	1,447	1,447	1,470	1,447	1,470	1,447
Number of id	324	324	324	322	322	322	324	322	324	322
R-squared	0.844	0.844	0.844	0.846	0.846	0.846	0.844	0.846	0.845	0.846

Online Appendix V - Table 12 (Panel B) – Testing for taste-based discrimination in prices earned (lean season). Alternative measures for *Estacas* IDs

	Physical output									
	Dep. Var.: $ln(Cult_{ht})$									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(HH size)	0.0278	0.0286	0.0279	0.0216	0.0232	0.0215	0.0250	0.0189	0.0321	0.0269
	(0.0483)	(0.0476)	(0.0484)	(0.0489)	(0.0482)	(0.0490)	(0.0473)	(0.0478)	(0.0468)	(0.0474)
ln(# members 7-60y)	0.332***	0.335***	0.332***	0.336***	0.338***	0.336***	0.336***	0.339***	0.326***	0.328***
	(0.0403)	(0.0403)	(0.0403)	(0.0409)	(0.0409)	(0.0408)	(0.0401)	(0.0407)	(0.0395)	(0.0404)
Estaca ID Mean	2.45e-06	6.23e-07	5.40e-07				-2.14e-05		-0.000299***	
	(2.64e-05)	(2.67e-05)	(2.60e-05)				(2.27e-05)		(8.84e-05)	
Estaca ID MDEV	-4.84e-05	3.04e-06								
	(8.69e-05)	(2.55e-05)								
Estaca ID S.D.	3.53e-05		2.38e-06						0.000134***	
	(6.17e-05)		(1.79e-05)						(3.80e-05)	
Estaca ID Mean - check				4.55e-06	2.95e-06	3.89e-06		-2.72e-05		-0.000306***
				(2.68e-05)	(2.69e-05)	(2.67e-05)		(2.38e-05)		(9.50e-05)
Estaca ID MDEV - check				-2.91e-05	-2.00e-06					
				(0.000121)	(2.66e-05)					
Estaca ID S.D.- check				1.80e-05		-2.38e-06				0.000140***
				(8.91e-05)		(1.94e-05)				(4.30e-05)
Estaca ID MAD							0.000104		0.000627***	
							(7.82e-05)		(0.000182)	
Estaca ID MAD - check								0.000119		0.000636***
								(8.05e-05)		(0.000191)
Grama Larga	0.133**	0.130**	0.132**	0.139**	0.136**	0.138**	0.124**	0.128**	0.0602	0.0626
	(0.0565)	(0.0564)	(0.0565)	(0.0571)	(0.0569)	(0.0571)	(0.0564)	(0.0568)	(0.0583)	(0.0592)
Lage	0.215***	0.215***	0.214***	0.214***	0.214***	0.214***	0.212***	0.210***	0.153**	0.148**
	(0.0721)	(0.0715)	(0.0721)	(0.0732)	(0.0725)	(0.0731)	(0.0711)	(0.0721)	(0.0743)	(0.0755)
Morro Alto	0.121*	0.121*	0.121*	0.129**	0.128**	0.128**	0.113*	0.118*	0.0375	0.0414
	(0.0620)	(0.0618)	(0.0620)	(0.0626)	(0.0625)	(0.0627)	(0.0623)	(0.0630)	(0.0654)	(0.0665)
Pão de Ló	0.185**	0.185**	0.186**	0.190**	0.190**	0.191**	0.176**	0.184**	0.248***	0.255***
	(0.0771)	(0.0769)	(0.0770)	(0.0771)	(0.0770)	(0.0770)	(0.0750)	(0.0750)	(0.0758)	(0.0764)
Salt. & Pinheiro	0.145*	0.144*	0.144*	0.160*	0.159*	0.160*	0.147*	0.166*	0.199**	0.208**
	(0.0797)	(0.0799)	(0.0797)	(0.0858)	(0.0857)	(0.0855)	(0.0801)	(0.0855)	(0.0783)	(0.0846)
Teteia	0.202***	0.207***	0.202***	0.217***	0.221***	0.216***	0.193***	0.206***	0.118*	0.133*
	(0.0653)	(0.0642)	(0.0652)	(0.0668)	(0.0657)	(0.0667)	(0.0644)	(0.0662)	(0.0688)	(0.0705)
Col. Ibicaba	-0.294**	-0.234*	-0.294**	-0.266*	-0.209	-0.266*	-0.237*	-0.212	-0.343**	-0.318*
	(0.140)	(0.130)	(0.140)	(0.153)	(0.138)	(0.153)	(0.132)	(0.141)	(0.162)	(0.174)
Mariquita	0.295***	0.295***	0.295***	0.293***	0.293***	0.292***	0.300***	0.298***	0.315***	0.312***
	(0.0292)	(0.0290)	(0.0292)	(0.0297)	(0.0295)	(0.0297)	(0.0290)	(0.0295)	(0.0302)	(0.0309)
Paineira	0.128***	0.128***	0.125***	0.271***	0.277***	0.276***	0.126***	0.282***	0.135***	0.231***
	(0.0253)	(0.0250)	(0.0255)	(0.0355)	(0.0268)	(0.0287)	(0.0249)	(0.0258)	(0.0253)	(0.0312)
Portão Isabel	0.465***	0.452***	0.465***	0.417***	0.402***	0.416***	0.451***	0.399***	0.466***	0.418***
	(0.0265)	(0.0294)	(0.0265)	(0.0343)	(0.0377)	(0.0342)	(0.0295)	(0.0377)	(0.0263)	(0.0341)
Iracema	-0.105	-0.106	-0.106	-0.101	-0.101	-0.101	-0.113*	-0.110*	-0.186***	-0.184***
	(0.0651)	(0.0647)	(0.0652)	(0.0658)	(0.0654)	(0.0658)	(0.0645)	(0.0652)	(0.0675)	(0.0677)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,360	2,366	2,360	2,295	2,301	2,295	2,366	2,301	2,360	2,295
R-squared	0.494	0.482	0.494	0.498	0.485	0.498	0.483	0.486	0.500	0.504
Number of id	513	513	513	505	505	505	513	505	513	505

	Discrimination in pay									
	Dep. Var.: Avg. price earned for cultivating one coffee grove (lean season)									
TFP agr. task	0.00151***	0.00153***	0.00151***	0.00161***	0.00163***	0.00161***	0.00151***	0.00161***	0.00146***	0.00154***
	(0.000252)	(0.000254)	(0.000253)	(0.000244)	(0.000247)	(0.000244)	(0.000252)	(0.000245)	(0.000249)	(0.000244)
Non-Whites	0.000321	0.000375*	0.000321	0.000321	0.000374*	0.000321	0.000376*	0.000374*	0.000313	0.000314
	(0.000206)	(0.000215)	(0.000206)	(0.000206)	(0.000215)	(0.000206)	(0.000215)	(0.000214)	(0.000207)	(0.000207)
Italians	-0.000102	-4.80e-05	-0.000101	-0.000135	-8.09e-05	-0.000135	-4.70e-05	-8.06e-05	-9.06e-05	-0.000119
	(0.000226)	(0.000233)	(0.000226)	(0.000226)	(0.000233)	(0.000226)	(0.000233)	(0.000233)	(0.000226)	(0.000227)
Germanics	0.000470	0.000526	0.000471	0.000458	0.000514	0.000458	0.000525	0.000512	0.000472	0.000462
	(0.000445)	(0.000448)	(0.000445)	(0.000447)	(0.000450)	(0.000447)	(0.000449)	(0.000450)	(0.000445)	(0.000447)
Portuguese	-0.000273	-0.000222	-0.000273	-0.000296	-0.000245	-0.000296	-0.000223	-0.000246	-0.000267	-0.000289
	(0.000253)	(0.000260)	(0.000253)	(0.000255)	(0.000262)	(0.000255)	(0.000260)	(0.000262)	(0.000253)	(0.000256)
Spaniards	-0.000198	-0.000136	-0.000197	-0.000229	-0.000165	-0.000229	-0.000138	-0.000168	-0.000185	-0.000215
	(0.000420)	(0.000420)	(0.000420)	(0.000421)	(0.000422)	(0.000421)	(0.000420)	(0.000421)	(0.000421)	(0.000424)
East. Europeans	-0.000564	-0.000547	-0.000563	-0.000580	-0.000566	-0.000580	-0.000545	-0.000563	-0.000568	-0.000584
	(0.000728)	(0.000749)	(0.000728)	(0.000721)	(0.000742)	(0.000721)	(0.000749)	(0.000742)	(0.000724)	(0.000718)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,360	2,366	2,360	2,295	2,301	2,295	2,366	2,301	2,360	2,295
Number of id	513	513	513	505	505	505	513	505	513	505

Online Appendix V - Table 12 (Panel B ctd.) – Testing for taste-based discrimination in prices earned (harv. season). Alternative measures for *Estacas* IDs

	Physical output									
	Dep. Var.: $ln(Harv_{ht})$									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ln(HH size)	-0.113	-0.114	-0.114	-0.130*	-0.133*	-0.132*	-0.116	-0.135*	-0.103	-0.120
	(0.0735)	(0.0725)	(0.0731)	(0.0747)	(0.0737)	(0.0743)	(0.0719)	(0.0725)	(0.0742)	(0.0742)
ln(# members 7-60y)	0.393***	0.394***	0.394***	0.398***	0.399***	0.399***	0.394***	0.399***	0.385***	0.387***
	(0.0616)	(0.0609)	(0.0614)	(0.0631)	(0.0622)	(0.0626)	(0.0603)	(0.0614)	(0.0624)	(0.0637)
Estaca ID Mean	-3.10e-05*	-3.14e-05*	-3.37e-05*				-6.27e-05**		-0.000277**	
	(1.74e-05)	(1.76e-05)	(1.86e-05)				(3.15e-05)		(0.000136)	
Estaca ID MDEV	-8.01e-05	-1.05e-05								
	(0.000107)	(3.07e-05)								
Estaca ID S.D.	4.95e-05		-5.37e-06						0.000102	
	(8.32e-05)		(2.29e-05)						(6.60e-05)	
Estaca ID Mean - check				-2.87e-05	-2.82e-05	-3.17e-05		-8.40e-05**		-0.000362**
				(1.86e-05)	(1.86e-05)	(2.06e-05)		(3.54e-05)		(0.000142)
Estaca ID MDEV - check				-0.000155	-2.42e-05					
				(0.000170)	(3.41e-05)					
Estaca ID S.D.- check				9.43e-05		-1.43e-05				0.000137*
				(0.000129)		(2.61e-05)				(7.15e-05)
Estaca ID MAD							9.29e-05		0.000503*	
							(8.44e-05)		(0.000269)	
Estaca ID MAD - check								0.000147		0.000673**
								(9.49e-05)		(0.000279)
Grama Larga	0.0706	0.0700	0.0693	0.0653	0.0652	0.0641	0.0620	0.0515	0.00968	-0.0184
	(0.0716)	(0.0716)	(0.0718)	(0.0721)	(0.0720)	(0.0723)	(0.0708)	(0.0710)	(0.0776)	(0.0791)
Lage	0.0310	0.0304	0.0297	0.0361	0.0359	0.0350	0.0252	0.0263	-0.0176	-0.0315
	(0.0724)	(0.0725)	(0.0726)	(0.0774)	(0.0773)	(0.0774)	(0.0725)	(0.0773)	(0.0773)	(0.0842)
Morro Alto	0.0523	0.0518	0.0511	0.0395	0.0393	0.0382	0.0424	0.0230	-0.0167	-0.0556
	(0.0843)	(0.0844)	(0.0846)	(0.0866)	(0.0866)	(0.0869)	(0.0845)	(0.0867)	(0.0922)	(0.0955)
Pão de Ló	0.0316	0.0316	0.0337	0.0303	0.0299	0.0329	0.0276	0.0295	0.0792	0.0976
	(0.0759)	(0.0756)	(0.0748)	(0.0770)	(0.0767)	(0.0757)	(0.0777)	(0.0791)	(0.0709)	(0.0686)
Salt. & Pinheiro	0.223	0.225	0.226	0.230	0.235	0.236	0.232	0.245	0.266*	0.287*
	(0.159)	(0.157)	(0.157)	(0.172)	(0.168)	(0.168)	(0.158)	(0.169)	(0.149)	(0.158)
Teteia	-0.0459	-0.0463	-0.0464	-0.0519	-0.0525	-0.0529	-0.0575	-0.0704	-0.110	-0.137
	(0.160)	(0.159)	(0.159)	(0.155)	(0.155)	(0.155)	(0.158)	(0.152)	(0.160)	(0.157)
Col. Ibicaba	0.151	0.151	0.151	0.172	0.173	0.172	0.148	0.165	0.112	0.116
	(0.274)	(0.274)	(0.273)	(0.315)	(0.315)	(0.314)	(0.277)	(0.320)	(0.282)	(0.326)
Mariquita	0.287***	0.287***	0.287***	0.288***	0.288***	0.288***	0.291***	0.295***	0.301***	0.307***
	(0.0465)	(0.0461)	(0.0463)	(0.0462)	(0.0458)	(0.0459)	(0.0461)	(0.0459)	(0.0492)	(0.0491)
Portão Isabel	-0.136***	-0.136***	-0.136***	-0.272***	-0.273***	-0.273***	-0.137***	-0.274***	-0.135***	-0.271***
	(0.0377)	(0.0378)	(0.0378)	(0.0494)	(0.0496)	(0.0496)	(0.0378)	(0.0496)	(0.0373)	(0.0489)
Iracema	-0.267***	-0.266***	-0.267***	-0.273***	-0.272***	-0.272***	-0.275***	-0.287***	-0.331***	-0.360***
	(0.0967)	(0.0967)	(0.0967)	(0.0993)	(0.0991)	(0.0991)	(0.0973)	(0.0998)	(0.100)	(0.104)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,641	1,642	1,641	1,595	1,596	1,595	1,642	1,596	1,641	1,595
Number of id	335	335	335	333	333	333	335	333	335	333
R-squared	0.816	0.816	0.816	0.818	0.818	0.818	0.816	0.818	0.816	0.819
	Discrimination in pay									
	Dep. Var.: Avg. price earned for harvesting one *alqueire* (harv. season)									
TFP agr. task	0.00730***	0.00731***	0.00733***	0.00789***	0.00793***	0.00795***	0.00724***	0.00784***	0.00725***	0.00778***
	(0.00268)	(0.00268)	(0.00268)	(0.00277)	(0.00277)	(0.00277)	(0.00268)	(0.00277)	(0.00267)	(0.00276)
Non-Whites	0.00499	0.00500	0.00500	0.00520	0.00521	0.00521	0.00500	0.00521	0.00496	0.00516
	(0.00372)	(0.00372)	(0.00372)	(0.00376)	(0.00376)	(0.00376)	(0.00372)	(0.00376)	(0.00372)	(0.00376)
Italians	0.00443	0.00442	0.00442	0.00437	0.00436	0.00436	0.00444	0.00438	0.00445	0.00440
	(0.00396)	(0.00397)	(0.00396)	(0.00400)	(0.00400)	(0.00399)	(0.00397)	(0.00400)	(0.00396)	(0.00399)
Germanics	0.000879	0.000878	0.000880	0.00110	0.00110	0.00110	0.000869	0.00109	0.000880	0.00109
	(0.00427)	(0.00426)	(0.00426)	(0.00422)	(0.00422)	(0.00422)	(0.00427)	(0.00422)	(0.00427)	(0.00423)
Portuguese	0.00600	0.00600	0.00600	0.00604	0.00604	0.00604	0.00601	0.00604	0.00601	0.00605
	(0.00470)	(0.00470)	(0.00470)	(0.00471)	(0.00471)	(0.00471)	(0.00470)	(0.00472)	(0.00470)	(0.00472)
Spaniards	-0.0136	-0.0136	-0.0136	-0.0144	-0.0144	-0.0144	-0.0136	-0.0143	-0.0136	-0.0143
	(0.00931)	(0.00930)	(0.00930)	(0.00908)	(0.00907)	(0.00907)	(0.00932)	(0.00909)	(0.00932)	(0.00911)
East. Europeans	0.00343	0.00342	0.00342	0.00262	0.00262	0.00260	0.00344	0.00263	0.00332	0.00247
	(0.00378)	(0.00378)	(0.00378)	(0.00383)	(0.00383)	(0.00383)	(0.00378)	(0.00383)	(0.00379)	(0.00384)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,470	1,470	1,470	1,447	1,447	1,447	1,470	1,447	1,470	1,447
Number of id	324	324	324	322	322	322	324	322	324	322

IV. Colonies of residence, colonies of work, and housing characteristics

As discussed in Section 04 of the main text and Online Appendix III, most entries in the *Registries* did not distinguish between colonies of residence and of work. This is not surprising, as most *colonos* indeed lived in the same colonies where they worked – given the costs of walking long distances and the value they placed on contracts allowing them to cultivate subsistence crops between coffee rows.[6] The pairwise correlation of 0.964 (significant at the 1%-level) between these variables shows their strong association. Consequently, swapping the two variables yields practically the same results, the only substantive differences being the reversal of significance for *Colônia Ibicaba* and *Morro Alto* (Online Appendix V – Table 13, Panels A and B).[7]

In addition, the descriptive statistics showed that while most households remained in the same colony upon arrival in *Ibicaba*, a few did move between colonies. To control for potential selectivity, especially if more productive households demanded more productive agricultural capital, variables *Dummy Δ colony (work)* and *Dummy Δ colony (resid)* identify those households that worked or resided in two different colonies for contiguous observations.[8] Results are robust to controlling separately for each. Also, their coefficients are negative and significant (at the 5%-level) during the harvest season, but not during the lean season, suggesting that households moved between seasons, after completing (and being paid for) the cultivation of their coffee groves (Online Appendix V – Table 14, Panels A and B).

Finally, Online Appendix V - Table 15 (separated between the lean and harvest seasons) assesses the effect of controlling for housing characteristics, which can be taken as a proxy for basic material wellbeing in the home. Columns 1-7 control separately for the number of houses, rooms, windows, doors, locks, and lightbulb sockets available to each household, while column 8 controls jointly for the number of houses, windows, and lightbulb sockets. Specifications controlling for the number of rooms and the number of locks render household TFP estimates marginally significant, at best – a result attributable to ensuing small samples.[9] The results also indicate that housing characteristics had no significant effect in themselves. If anything, occupying more than one house was significant for the number of coffee groves cultivated and for the earnings thereupon derived.[10] Since the number of amenities per house is not significant, this effect cannot stem from households' ability to pay for more amenities or from managers' favoritism in assigning them to better houses. Rather, occupying more than one house seems to have permitted a better intrahousehold allocation of labor.

---

[6] The household of Mathias Perez (ID = 0151) shows the largest number of such variations. In 1925, they were curiously living in the "house of a neighboring property" (ledger #0062, DI 128). Similar entries for workers living outside *Ibicaba* appears twice more in that same ledger (DIs 76, 101).

[7] 25 observations with complete data show differing colonies of residence and work, a figure that would rise to at most 200 if every missing observation were assumed to differ in that same way.

[8] Alternative indicators based on different measures of the number of known times a household moved within *Ibicaba* were all dropped due to collinearity.

[9] *Non-Whites* and *Eastern Europeans* are omitted in the estimates that control for the number of # locks.

[10] Although only 8 households are recorded as residing in three or four houses, a further 107 lived in two.

Online Appendix V - Table 13 (Panel A) – Determinants of earnings (lean season). Colonies of work and of residence

	Dep. Var.: $ln(Y_{ht}^{cult})$	
	Controls for colonies:	
	Work	Residence
ln(HH size)	0.0186	0.0157
	(0.0484)	(0.0485)
ln(# members 7-60y)	0.335***	0.337***
	(0.0397)	(0.0398)
Estaca ID Mean	1.92e-06	1.82e-06
	(2.60e-05)	(2.61e-05)
Estaca ID MDEV	-4.53e-05	-5.18e-05
	(8.50e-05)	(8.55e-05)
Estaca ID S.D.	3.34e-05	3.68e-05
	(6.03e-05)	(6.07e-05)
	Colonies	
Grama Larga	0.143**	0.142**
	(0.0568)	(0.0559)
Lage	0.213***	0.182***
	(0.0712)	(0.0659)
Morro Alto	0.129**	0.128**
	(0.0620)	(0.0593)
Pão de Ló	0.189**	0.189**
	(0.0751)	(0.0748)
Salt. & Pinheiro	0.157**	0.158**
	(0.0790)	(0.0785)
Teteia	0.200***	0.198***
	(0.0644)	(0.0632)
Col. Ibicaba	-0.269**	-0.177
	(0.134)	(0.139)
Mariquita	0.277***	0.197***
	(0.0303)	(0.0277)
Paineira	0.0876***	-0.240***
	(0.0282)	(0.0414)
Portão Isabel	0.433***	0.443***
	(0.0261)	(0.0282)
Laborers	-	0.0512
		(0.0805)
Iracema	-0.115*	-0.120*
	(0.0644)	(0.0639)
Full set of controls	Yes	Yes
Observations	2,358	2,342
R-squared	0.861	0.860
Number of id	512	508

Online Appendix V - Table 13 (Panel B) – Testing for taste-based discrimination in prices earned. Colonies of work and of residence

	Physical output			
	Dep. Var.: $ln(Cult_{ht})$		Dep. Var.: $ln(Harv_{ht})$	
	Controls for colonies:			
	Work	Residence	Work	Residence
ln(HH size)	0.0278	0.0248	-0.113	-0.110
	(0.0483)	(0.0484)	(0.0735)	(0.0741)
ln(# members 7-60y)	0.332***	0.334***	0.393***	0.392***
	(0.0403)	(0.0405)	(0.0616)	(0.0621)
Estaca ID Mean	2.45e-06	2.35e-06	-3.10e-05*	-3.13e-05*
	(2.64e-05)	(2.66e-05)	(1.74e-05)	(1.78e-05)
Estaca ID MDEV	-4.84e-05	-5.39e-05	-8.01e-05	-7.35e-05
	(8.69e-05)	(8.77e-05)	(0.000107)	(0.000112)
Estaca ID S.D.	3.53e-05	3.80e-05	4.95e-05	4.50e-05
	(6.17e-05)	(6.23e-05)	(8.32e-05)	(8.73e-05)
		Colonies		
Grama Larga	0.133**	0.132**	0.0706	0.0687
	(0.0565)	(0.0555)	(0.0716)	(0.0714)
Lage	0.215***	0.182***	0.0310	0.0151
	(0.0721)	(0.0676)	(0.0724)	(0.0793)
Morro Alto	0.121*	0.119**	0.0523	0.0471
	(0.0620)	(0.0593)	(0.0843)	(0.0819)
Pão de Ló	0.185**	0.184**	0.0316	0.0313
	(0.0771)	(0.0768)	(0.0759)	(0.0763)
Salt. & Pinheiro	0.145*	0.144*	0.223	0.223
	(0.0797)	(0.0792)	(0.159)	(0.159)
Teteia	0.202***	0.200***	-0.0459	-0.0500
	(0.0653)	(0.0642)	(0.160)	(0.160)
Col. Ibicaba	-0.294**	-0.206	0.151	-0.0609
	(0.140)	(0.141)	(0.274)	(0.378)
Mariquita	0.295***	0.204***	0.287***	0.230***
	(0.0292)	(0.0293)	(0.0465)	(0.0441)
Paineira	0.128***	-0.304***	-	-
	(0.0253)	(0.0379)		
Portão Isabel	0.465***	0.479***	-0.136***	-0.134***
	(0.0265)	(0.0321)	(0.0377)	(0.0382)
Laborers		0.0751		0.0191
		(0.115)		(0.0469)
Iracema		-0.111*	-0.267***	-0.272***
		(0.0646)	(0.0967)	(0.0945)
Full set of controls	Yes	Yes	Yes	Yes
Observations	2,360	2,344	1,641	1,632
Number of id	513	509	335	333
R-squared	0.494	0.491	0.816	0.815

	Discrimination in pay			
	Dep. Var.: Average prices earned for			
	Cultivating one coffee grove (lean season)		Harvesting one *alqueire* (harvest season)	
TFP agr. task	0.00151***	0.00145***	0.00730***	0.00725***
	(0.000252)	(0.000277)	(0.00268)	(0.00266)
Non-Whites	0.000321	0.000327	0.00499	0.00495
	(0.000206)	(0.000207)	(0.00372)	(0.00371)
Italians	-0.000102	-9.04e-05	0.00443	0.00440
	(0.000226)	(0.000228)	(0.00396)	(0.00396)
Germanics	0.000470	0.000474	0.000879	0.000881
	(0.000445)	(0.000446)	(0.00427)	(0.00428)
Portuguese	-0.000273	-0.000270	0.00600	0.00601
	(0.000253)	(0.000254)	(0.00470)	(0.00470)
Spaniards	-0.000198	-0.000194	-0.0136	-0.0136
	(0.000420)	(0.000420)	(0.00931)	(0.00932)
East. Europeans.	-0.000564	-0.000556	0.00343	0.00361
	(0.000728)	(0.000732)	(0.00378)	(0.00377)
Full set of controls	Yes	Yes	Yes	Yes
Observations	2,360	2,344	1,470	1,462
Number of id	513	509	324	322

Online Appendix V - Table 14 (Panel A) – Determinants of earnings. Mobility between colonies in *Ibicaba*

	Dep. Var.: $ln(Y_{ht}^{cult})$		Dep. Var.: $ln(Y_{ht}^{harv})$	
Dummy Δ colony (work)	-0.00197		-0.0846**	
	(0.0205)		(0.0397)	
Dummy Δ colony (resid)		-0.00211		-0.0900**
		(0.0232)		(0.0390)
ln(HH size)	0.00235	0.00317	-0.132	-0.105
	(0.0534)	(0.0528)	(0.0827)	(0.0855)
ln(# members 7-60y)	0.357***	0.356***	0.424***	0.414***
	(0.0434)	(0.0432)	(0.0701)	(0.0712)
Estaca ID Mean	7.16e-06	7.49e-06	-3.61e-05*	-2.94e-05
	(2.55e-05)	(2.55e-05)	(1.88e-05)	(2.04e-05)
Estaca ID MDEV	-3.79e-05	-5.76e-05	9.55e-05	9.20e-05
	(8.40e-05)	(7.95e-05)	(0.000123)	(0.000136)
Estaca ID S.D.	2.35e-05	3.72e-05	-7.19e-05	-8.09e-05
	(6.02e-05)	(5.74e-05)	(0.000104)	(0.000113)
	Colonies			
Grama Larga	0.129**	0.129**	0.0870	0.0841
	(0.0595)	(0.0598)	(0.0781)	(0.0776)
Lage	0.215**	0.214**	0.0452	0.0500
	(0.0831)	(0.0854)	(0.0730)	(0.0725)
Morro Alto	0.124*	0.125*	0.0648	0.0639
	(0.0681)	(0.0683)	(0.0883)	(0.0878)
Pão de Ló	0.157*	0.156*	0.0844	0.0782
	(0.0800)	(0.0802)	(0.0795)	(0.0788)
Salt. & Pinheiro	0.107	0.108	0.189	0.190
	(0.0846)	(0.0842)	(0.196)	(0.191)
Teteia	0.153***	0.154***	-0.0660	-0.0777
	(0.0587)	(0.0585)	(0.162)	(0.161)
Col. Ibicaba	-0.312**	-0.312**	0.225	0.242
	(0.150)	(0.151)	(0.251)	(0.241)
Mariquita	0.287***	0.288***	0.226***	0.198***
	(0.0280)	(0.0277)	(0.0415)	(0.0441)
Paineira	0.205***	-	-	-
	(0.0383)			
Portão Isabel	0.430***	0.429***	-0.0900	-0.109**
	(0.0257)	(0.0264)	(0.0560)	(0.0550)
Iracema	-0.171**	-0.171**	-0.140	-0.139
	(0.0749)	(0.0780)	(0.107)	(0.107)
Full set of controls	Yes	Yes	Yes	Yes
Observations	1,971	1,960	1,450	1,442
Number of id	380	377	317	313
R-squared	0.876	0.876	0.844	0.845

Online Appendix V - Table 14 (Panel B) – Testing for taste-based discrimination in prices earned. Mobility between colonies in *Ibicaba*

	Physical output			
	Dep. Var.: $ln(Cult_{ht})$		Dep. Var.: $ln(Harv_{ht})$	
Dummy Δ colony (work)	-0.00282		-0.0800**	
	(0.0206)		(0.0372)	
Dummy Δ colony (resid)		5.50e-05		-0.103***
		(0.0234)		(0.0379)
ln(HH size)	0.0127	0.0122	-0.148**	-0.116
	(0.0536)	(0.0529)	(0.0719)	(0.0755)
ln(# members 7-60y)	0.355***	0.354***	0.410***	0.397***
	(0.0439)	(0.0437)	(0.0615)	(0.0629)
Estaca ID Mean	8.18e-06	8.24e-06	-3.92e-05**	-3.45e-05*
	(2.62e-05)	(2.61e-05)	(1.74e-05)	(1.83e-05)
Estaca ID MDEV	-4.12e-05	-5.93e-05	-2.25e-05	-5.51e-05
	(8.56e-05)	(8.19e-05)	(9.81e-05)	(0.000109)
Estaca ID S.D.	2.50e-05	3.79e-05	1.82e-05	3.32e-05
	(6.13e-05)	(5.91e-05)	(8.03e-05)	(8.60e-05)
Grama Larga	0.117**	0.116*	0.0783	0.0795
	(0.0592)	(0.0595)	(0.0728)	(0.0735)
Lage	0.216**	0.214**	0.0528	0.0623
	(0.0835)	(0.0858)	(0.0705)	(0.0700)
Morro Alto	0.114*	0.114*	0.0527	0.0537
	(0.0678)	(0.0680)	(0.0845)	(0.0841)
Pão de Ló	0.155*	0.154*	0.0726	0.0688
	(0.0829)	(0.0828)	(0.0823)	(0.0857)
Salt. & Pinheiro	0.0947	0.0968	0.243	0.227
	(0.0868)	(0.0863)	(0.173)	(0.168)
Teteia	0.155***	0.157***	-0.0448	-0.0582
	(0.0599)	(0.0596)	(0.163)	(0.164)
Col. Ibicaba	-0.333**	-0.335**	0.217	0.238
	(0.162)	(0.164)	(0.265)	(0.254)
Mariquita	0.303***	0.303***	0.298***	0.270***
	(0.0281)	(0.0278)	(0.0469)	(0.0482)
Paineira	0.284***	-	-	-
	(0.0330)			
Portão Isabel	0.459***	0.459***	-0.131***	-0.163***
	(0.0264)	(0.0268)	(0.0375)	(0.0398)
Iracema	-0.164**	-0.168**	-0.167	-0.154
	(0.0766)	(0.0796)	(0.104)	(0.106)
Full set of controls	Yes	Yes	Yes	Yes
Observations	1,972	1,961	1,617	1,608
R-squared	0.533	0.536	0.818	0.818
Number of id	380	377	328	325

	Discrimination in pay			
	Average price earned for			
	Cultivating one coffee grove (lean season)		Harvesting one *alqueire* (harvest season)	
TFP agr. task	0.00160***	0.00162***	0.00672***	0.00674***
	(0.000278)	(0.000264)	(0.00226)	(0.00228)
Non-Whites	0.000180	0.000165	0.00605	0.00591
	(0.000242)	(0.000243)	(0.00386)	(0.00389)
Italians	-0.000291	-0.000311	0.00547	0.00538
	(0.000256)	(0.000257)	(0.00407)	(0.00409)
Germanics	0.000138	0.000145	0.00249	0.00251
	(0.000538)	(0.000552)	(0.00414)	(0.00416)
Portuguese	-0.000341	-0.000346	0.00604	0.00683
	(0.000276)	(0.000276)	(0.00457)	(0.00464)
Spaniards	-0.000724	-0.000758	-0.0165*	-0.0165*
	(0.000509)	(0.000511)	(0.00869)	(0.00867)
East. Europeans	-0.000947	-0.000972	0.00475	0.00469
	(0.000954)	(0.000954)	(0.00392)	(0.00395)
Full set of controls	Yes	Yes	Yes	Yes
Observations	1,972	1,961	1,450	1,442
Number of id	380	377	317	313

Online Appendix V - Table 15 (Panel A) – Determinants of earnings (lean season). Housing characteristics

	Dep. Var.: $ln(Y_{ht}^{cult})$							
	(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)
Estaca ID Mean	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)
Estaca ID MDEV	-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)
Estaca ID S.D.	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. & Pinheiro	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	-0.115*	-0.217***		-0.219**	-0.216**		-0.101	-0.220**
	(0.0644)	(0.0790)		(0.104)	(0.103)		(0.0656)	(0.103)
# houses		0.0646***						0.0912**
		(0.0199)						(0.0375)
# rooms			0.0258*					
			(0.0133)					
# windows				0.00914				0.00284
				(0.00735)				(0.00895)
# doors					-0.000673			
					(0.00732)			
# locks						0.110***		
						(0.0374)		
# lightbulb sockets							-0.0473	-0.0198
							(0.0392)	(0.0337)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,358	2,265	766	1,536	1,523	948	2,342	1,536
Number of id	512	494	152	336	332	227	508	336
R-squared	0.861	0.865	0.904	0.842	0.842	0.803	0.861	0.844

Online Appendix V - Table 15 (Panel A ctd.) – Determinants of earnings (harv. season). Housing characteristics

	Dep Var.: $ln(Y_{ht}^{harv})$							
	(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)
Estaca ID Mean	-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)
Estaca ID MDEV	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)
Estaca ID S.D.	-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. & Pinheiro	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)
Portão Isabel	-0.107**	-0.0621					-0.109**	
	(0.0543)	(0.0681)					(0.0546)	
Iracema	-0.235**	-0.310***		-0.113	-0.104		-0.278***	-0.214
	(0.0979)	(0.100)		(0.133)	(0.128)		(0.0970)	(0.141)
# houses		0.00524						-0.0674
		(0.0294)						(0.0564)
# rooms			-0.00623					
			(0.0230)					
# windows				0.00985				0.0113
				(0.00849)				(0.0109)
# doors					-0.000284			
					(0.00629)			
# locks						0.0193		
						(0.0319)		
# lightbulb sockets							0.0357	0.104*
							(0.0435)	(0.0600)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,470	1,424	443	996	987	673	1,462	996
Number of id	324	316	96	212	207	149	322	212
R-squared	0.844	0.867	0.891	0.876	0.876	0.866	0.844	0.877

Online Appendix V - Table 15 (Panel B) – Testing for taste-based discrimination in prices earned (lean season). Housing characteristics

	Physical output							
	Dep. Var.: $ln(Cult_{ht})$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(HH size)	0.0278	0.0521	0.209**	0.144***	0.146***	0.209***	0.0263	0.140***
	(0.0483)	(0.0475)	(0.102)	(0.0527)	(0.0548)	(0.0708)	(0.0484)	(0.0517)
ln(# members 7-60y)	0.332***	0.313***	0.258***	0.334***	0.338***	0.304***	0.332***	0.326***
	(0.0403)	(0.0406)	(0.0587)	(0.0522)	(0.0512)	(0.0838)	(0.0406)	(0.0519)
Estaca ID Mean	2.45e-06	-0.000435**	-0.000884**	-0.000646*	-0.000643*	-0.000273	2.37e-06	-0.000664*
	(2.64e-05)	(0.000197)	(0.000446)	(0.000349)	(0.000342)	(0.000558)	(2.77e-05)	(0.000346)
Estaca ID MDEV	-4.84e-05	0.000387*	0.00277***	0.00215**	0.00212**	0.00429***	-6.08e-05	0.00220**
	(8.69e-05)	(0.000230)	(0.000559)	(0.000964)	(0.000940)	(0.00158)	(8.70e-05)	(0.000986)
Estaca ID S.D.	3.53e-05	-7.66e-05	-0.00118***	-0.000889*	-0.000872*	-0.00222**	4.36e-05	-0.000914*
	(6.17e-05)	(9.18e-05)	(0.000305)	(0.000526)	(0.000513)	(0.000942)	(6.16e-05)	(0.000537)
Grama Larga	0.133**	0.0285	0.0988	-0.0385	-0.0324	0.00472	0.161***	-0.0251
	(0.0565)	(0.0724)	(0.125)	(0.103)	(0.101)	(0.135)	(0.0577)	(0.101)
Lage	0.215***	0.110		0.0601	0.0667	0.130	0.218***	0.0609
	(0.0721)	(0.0921)		(0.110)	(0.110)	(0.145)	(0.0724)	(0.107)
Morro Alto	0.121*	0.00787		0.0198	0.0242	0.0775	0.129**	0.00845
	(0.0620)	(0.0827)		(0.115)	(0.115)	(0.146)	(0.0612)	(0.115)
Pão de Ló	0.185**	0.259***	0.236**	0.305***	0.323***	0.145	0.181**	0.298***
	(0.0771)	(0.0789)	(0.0967)	(0.0992)	(0.0997)	(0.123)	(0.0769)	(0.101)
Salt. & Pinheiro	0.145*	0.153*		0.146	0.160	0.0441	0.148*	0.165
	(0.0797)	(0.0889)		(0.138)	(0.132)	(0.0641)	(0.0795)	(0.134)
Teteia	0.202***	0.0548		0.119	0.137	0.131	0.202***	0.0897
	(0.0653)	(0.0951)		(0.160)	(0.156)	(0.225)	(0.0647)	(0.162)
Col. Ibicaba	-0.294**	-0.492**					-0.285**	
	(0.140)	(0.227)					(0.135)	
Mariquita	0.295***	0.310***	0.284***	0.222***	0.210***	0.300***	0.302***	0.218***
	(0.0292)	(0.0302)	(0.0178)	(0.0340)	(0.0335)	(0.0442)	(0.0301)	(0.0367)
Paineira	0.128***							
	(0.0253)							
Portão Isabel	0.465***	0.398***					0.465***	
	(0.0265)	(0.0323)					(0.0268)	
Iracema	-0.105	-0.220***		-0.231**	-0.228**		-0.0915	-0.233**
	(0.0651)	(0.0788)		(0.103)	(0.101)		(0.0657)	(0.101)
# houses		0.0622***						0.0912**
		(0.0203)						(0.0372)
# rooms			0.0267*					
			(0.0135)					
# windows				0.00945				0.00310
				(0.00724)				(0.00883)
# doors					-0.000917			
					(0.00734)			
# locks						0.112***		
						(0.0376)		
# lightbulb sockets							-0.0466	-0.0184
							(0.0393)	(0.0334)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,360	2,267	766	1,536	1,523	948	2,344	1,536
R-squared	0.494	0.511	0.601	0.525	0.525	0.546	0.495	0.529
Number of id	513	495	152	336	332	227	509	336
	Discrimination in pay							
	Dep. Var.: Avg. price earned for cultivating one coffee grove (lean season)							
TFP agr. task	0.00151***	0.00151***	0.00120***	0.00211***	0.00207***	0.00192***	0.00153***	0.00212***
	(0.000252)	(0.000230)	(0.000416)	(0.000334)	(0.000326)	(0.000415)	(0.000247)	(0.000335)
Non-Whites	0.000321	0.000319	0.000429	0.000263	0.000258	0.000345	0.000310	0.000252
	(0.000206)	(0.000216)	(0.000377)	(0.000268)	(0.000271)	(0.000301)	(0.000207)	(0.000267)
Italians	-0.000102	-4.90e-05	0.000332	-0.000209	-0.000216	-8.08e-05	-0.000111	-0.000206
	(0.000226)	(0.000231)	(0.000371)	(0.000294)	(0.000296)	(0.000342)	(0.000226)	(0.000294)
Germanics	0.000470	0.000470	0.000493	0.000479	0.000478	0.000807	0.000457	0.000469
	(0.000445)	(0.000456)	(0.000922)	(0.000524)	(0.000524)	(0.000692)	(0.000445)	(0.000525)
Portuguese	-0.000273	-0.000228	-0.000199	-0.000482	-0.000538	-0.000436	-0.000290	-0.000492
	(0.000253)	(0.000262)	(0.000384)	(0.000339)	(0.000354)	(0.000484)	(0.000254)	(0.000339)
Spaniards	-0.000198	-0.000328	-0.000780	-0.000548	-0.000559	-0.000914*	-0.000194	-0.000544
	(0.000420)	(0.000399)	(0.000798)	(0.000464)	(0.000457)	(0.000528)	(0.000418)	(0.000465)
East. Europeans.	-0.000564	-0.000614	-0.000529	-0.000902	-0.000918	-0.00170	-0.000550	-0.000905
	(0.000728)	(0.000741)	(0.00102)	(0.000942)	(0.000938)	(0.00140)	(0.000732)	(0.000925)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,360	2,267	766	1,536	1,523	948	2,344	1,536
Number of id	513	495	152	336	332	227	509	336

Online Appendix V - Table 15 (Panel B ctd.) – Testing for taste-based discrimination in prices earned (harv. season). Housing characteristics

	Physical output							
	Dep. Var.: $ln(Harv_{ht})$							
:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(HH size)	-0.113	-0.0936	0.0684	-0.123	-0.121	-0.0629	-0.114	-0.128*
	(0.0735)	(0.0611)	(0.101)	(0.0764)	(0.0769)	(0.0867)	(0.0738)	(0.0766)
ln(# members 7-60y)	0.393***	0.398***	0.340***	0.524***	0.529***	0.540***	0.395***	0.536***
	(0.0616)	(0.0548)	(0.0781)	(0.0771)	(0.0768)	(0.103)	(0.0621)	(0.0783)
Estaca ID Mean	-3.10e-05*	-0.000464*	-0.000229	-8.83e-05	-9.38e-05	0.000387	-3.07e-05*	-6.07e-05
	(1.74e-05)	(0.000272)	(0.000704)	(0.000460)	(0.000448)	(0.000558)	(1.74e-05)	(0.000465)
Estaca ID MDEV	-8.01e-05	0.000393	0.00821***	0.00382	0.00367	0.00331	-7.28e-05	0.00366
	(0.000107)	(0.000285)	(0.00266)	(0.00278)	(0.00279)	(0.00234)	(0.000112)	(0.00275)
Estaca ID S.D.	4.95e-05	-9.65e-05	-0.00610***	-0.00238	-0.00224	-0.00241	4.35e-05	-0.00229
	(8.32e-05)	(9.26e-05)	(0.00182)	(0.00216)	(0.00218)	(0.00183)	(8.72e-05)	(0.00215)
Grama Larga	0.0706	0.00160	0.288*	0.0263	0.0282	0.116	0.0522	-0.0517
	(0.0716)	(0.0841)	(0.157)	(0.121)	(0.117)	(0.138)	(0.0686)	(0.124)
Lage	0.0310	-0.0161		-0.0131	-0.00726	0.175	0.0341	-0.0102
	(0.0724)	(0.0884)		(0.0930)	(0.0917)	(0.106)	(0.0720)	(0.0949)
Morro Alto	0.0523	-0.0387		0.129	0.136	0.259*	0.0464	0.136
	(0.0843)	(0.0991)		(0.138)	(0.133)	(0.149)	(0.0832)	(0.140)
Pão de Ló	0.0316	0.0875	0.140	-0.0336	-0.0118	-0.191	0.0364	-0.0147
	(0.0759)	(0.0900)	(0.132)	(0.104)	(0.107)	(0.133)	(0.0775)	(0.103)
Salt. & Pinheiro	0.223	0.392**		0.193	0.215	-0.212**	0.224	0.192
	(0.159)	(0.153)		(0.205)	(0.210)	(0.0856)	(0.161)	(0.208)
Teteia	-0.0459	-0.201		-0.367	-0.351	-0.227	-0.0440	-0.302
	(0.160)	(0.171)		(0.227)	(0.223)	(0.248)	(0.155)	(0.232)
Col. Ibicaba	0.151	-0.464***					0.148	
	(0.274)	(0.121)					(0.277)	
Mariquita	0.287***	0.228***	0.0975***	0.138**	0.123**	0.299***	0.284***	0.126**
	(0.0465)	(0.0351)	(0.0248)	(0.0587)	(0.0607)	(0.0765)	(0.0479)	(0.0597)
Portão Isabel	-0.136***	-0.114**					-0.137***	
	(0.0377)	(0.0442)					(0.0379)	
Iracema	-0.267***	-0.349***		-0.140	-0.132		-0.299***	-0.233*
	(0.0967)	(0.0998)		(0.132)	(0.127)		(0.0940)	(0.138)
# houses		0.0204						-0.0527
		(0.0285)						(0.0583)
# rooms			-0.00513					
			(0.0201)					
# windows				0.00990				0.0106
				(0.00853)				(0.0110)
# doors					-0.000781			
					(0.00652)			
# locks						0.0127		
						(0.0323)		
# lightbulb sockets							0.0268	0.0930
							(0.0417)	(0.0595)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,641	1,591	505	1,018	1,009	674	1,632	1,018
R-squared	0.816	0.842	0.875	0.855	0.856	0.853	0.815	0.856
Number of id	335	328	104	212	207	149	333	212
	Discrimination in pay							
	Avg. price earned for harvesting one *alqueire* (harv. season)							
TFP agr. task	0.00730***	0.00735***	0.000535	0.00919**	0.00916**	0.00892*	0.00728***	0.00919**
	(0.00268)	(0.00278)	(0.00304)	(0.00390)	(0.00392)	(0.00519)	(0.00267)	(0.00390)
Non-Whites	0.00499	0.00460	0.0111	0.00672*	0.00653		0.00505	0.00678*
	(0.00372)	(0.00381)	(0.00698)	(0.00400)	(0.00398)		(0.00371)	(0.00401)
Italians	0.00443	0.00482	0.0129*	0.00339	0.00311	0.00541	0.00437	0.00334
	(0.00396)	(0.00409)	(0.00680)	(0.00466)	(0.00466)	(0.00610)	(0.00396)	(0.00465)
Germanics	0.000879	0.00121	0.00967	-0.000933	-0.00119	-0.00152	0.000890	-0.000810
	(0.00427)	(0.00436)	(0.00649)	(0.00484)	(0.00485)	(0.00646)	(0.00428)	(0.00483)
Portuguese	0.00600	0.00649	0.0119	0.00644	0.00634	0.00604	0.00599	0.00648
	(0.00470)	(0.00485)	(0.0119)	(0.00557)	(0.00581)	(0.00728)	(0.00470)	(0.00557)
Spaniards	-0.0136	-0.0140	-0.00786	-0.0155	-0.0151	-0.0202	-0.0135	-0.0156
	(0.00931)	(0.00961)	(0.0121)	(0.0105)	(0.0106)	(0.0136)	(0.00933)	(0.0104)
East.Europeans	0.00343	0.00405	0.0105	0.00145	0.00122		0.00346	0.000852
	(0.00378)	(0.00391)	(0.00692)	(0.00447)	(0.00448)		(0.00378)	(0.00452)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,470	1,424	443	996	987	673	1,462	996
Number of id	324	316	96	212	207	149	322	212

V. Alternative measures of prices, earnings, and timing of the harvest

This set of robustness checks varies, one at a time, the measures of nominal earnings (lean and harvest seasons), of prices (remunerations) in the lean season, and of the proxies for the timing of the harvest.

Firstly, variable for nominal earnings from the lean season includes figures recorded both directly in the sources, as well as estimates derived wherever the following data were available: (i) number of coffee groves cultivated; (ii) nominal prices earned for the cultivation of individual groves; and (iii) number of prunings executed in year *t*. Moreover, conditional on (i), nominal earnings could be derived also by interpolating (ii) and (iii) from other households in the sample. The variable $Y_{ht}^{cult}\,check$ excludes the latter, using only nominal earnings recorded in the sources or directly derivable from them without interpolation. Online Appendix V - Table 16 shows that results are robust to this measure. Point estimates for prices, frequency of executed tasks, and household size and composition are identical to those of the baseline up to the second decimal place.

Secondly, the variable for prices in the lean season were compiled as the mean price earned by household h for cultivating coffee in year t, weighted by the quantity of coffee groves cultivated under different prices (remunerations). Online Appendix V - Table 17 shows that results are essentially the same if the median is used instead, or the unweighted mean.

Thirdly, nominal earnings in the harvest season can also be derived with interpolated prices. $Y_{ht}^{harv\ interpol\ year}$ averages the price per *alqueire* harvested in year t across all households, whereas $Y_{ht}^{harv\ interpol\ HH}$ averages the price series for each household h, interpolating missing values with information from that household alone.[11] If households were discriminated against (or in favor of), then they should have been systematically assigned to lower (higher) paying coffee groves, which favors $Y_{ht}^{harv\ interpol\ HH}$. By contrast, if households were simply allocated to the groves that had to be harvested most urgently, then averaging over all prices per year is a better variable, which favors $Y_{ht}^{harv\ interpol\ year}$. The descriptive statistics in Online Appendix V - Table 18 show, however, that differences are minimal: the overall means of prices and of nominal earnings vary by no more than 0.008 and 1.099 *milréis*, respectively.[12]

Finally, the last exercises alter the combination of variables for the timing of the harvest. Online Appendix V - Table 19 replaces the mean month of harvesting with its median and S.D., or combines these variables with the first and last month of harvesting. The results for frequency, timing, and household composition are all robust to these changes, as well as the rejection of taste-based discrimination in remunerations.

[11] In both cases, earnings are computed as the interpolated price multiplied by the (known) volume harvested.
[12] These alternative dependent variables yield exactly the same results as the baseline, since the sample is restricted to observations with complete data. Results are not reported here, but are available on request.

Online Appendix V - Table 16 – Determinants of earnings. Excluding data interpolations in earnings (lean season)

	Dep. Var.:	
	$ln(Y_{ht}^{cult})$	$ln(Y_{ht}^{cult}\ check)$
ln(price)	1.286***	1.286***
	(0.0838)	(0.0838)
ln(N)	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)
Estaca ID Mean	1.92e-06	2.18e-06
	(2.60e-05)	(2.60e-05)
Estaca ID MDEV	-4.53e-05	-4.30e-05
	(8.50e-05)	(8.74e-05)
Estaca ID S.D.	3.34e-05	3.09e-05
	(6.03e-05)	(6.23e-05)
Full set of controls	Yes	Yes
Observations	2,358	2,354
Number of id	512	512
R-squared	0.861	0.861

Online Appendix V - Table 17 (Panel A) – Determinants of earnings. Different weights for prices by number of cultivated coffee groves (lean season)

	Dep. Var.: $ln(Y_{ht}^{cult})$		
	Controls for *price*		
	Mean (weighted)	Mean (unweighted)	Median
ln(price)	1.286***	0.987***	0.975***
	(0.0838)	(0.0824)	(0.0816)
ln(N)	1.138***	1.168***	1.176***
	(0.252)	(0.222)	(0.222)
ln(HH size)	0.0186	0.0269	0.0272
	(0.0484)	(0.0494)	(0.0493)
ln(# members 7-60y)	0.335***	0.339***	0.339***
	(0.0397)	(0.0416)	(0.0415)
Estaca ID Mean	1.92e-06	4.43e-06	4.73e-06
	(2.60e-05)	(2.46e-05)	(2.44e-05)
Estaca ID MDEV	-4.53e-05	-4.78e-05	-4.81e-05
	(8.50e-05)	(8.82e-05)	(8.82e-05)
Estaca ID S.D.	3.34e-05	3.27e-05	3.28e-05
	(6.03e-05)	(6.30e-05)	(6.30e-05)
Full set of controls	Yes	Yes	Yes
Observations	2,358	2,358	2,358
Number of id	512	512	512
R-squared	0.861	0.849	0.849

Online Appendix V - Table 17 (Panel B) – Testing for taste-based discrimination in prices earned. Different weights for prices by number of cultivated coffee groves (lean season)

	Dep. Var.: Average price earned for cultivating one coffee grove (lean season)	
	Price with which residuals were estimated	
	Mean (unweigh.)	Median
TFP agr. task	0.00112***	0.00114***
	(0.000215)	(0.000215)
Non-Whites	0.000199	0.000215
	(0.000211)	(0.000212)
Italians	-0.000145	-0.000125
	(0.000215)	(0.000215)
Germanics	0.000197	0.000218
	(0.000413)	(0.000422)
Portuguese	-0.000231	-0.000229
	(0.000252)	(0.000252)
Spaniards	-0.000267	-0.000269
	(0.000433)	(0.000434)
East. Europeans	-0.000769	-0.000753
	(0.000589)	(0.000588)
Full set of controls	Yes	Yes
Observations	2,360	2,360
Number of id	513	513

Online Appendix V - Table 18 – Descriptive statistics of prices and earnings (lean and harvest seasons) with and without interpolated data

	N	n	Mean	S.D.	Median	Min	Max
Price (harvest)	4,032	915	0.8902	1.1285	0.700	0.3074	15
Price (harvest) interpol year	4,661	990	0.8907	1.1554	0.700	0.3074	15
Price (harvest) interpol HH	4,561	915	0.8829	1.1659	0.700	0.3074	15
Y_{ht}^{harv}	4,029	915	364.9285	419.8206	276.000	0.500	9,712.500
$Y_{ht}^{harv\ interpol\ year}$	4,605	974	364.4085	401.7976	283.000	0.500	9,712.500
$Y_{ht}^{harv\ interpol\ HH}$	4,523	915	363.8296	403.8676	279.900	0.500	9,712.500

Note: The interpolation per year increases the number of uniquely identified HHs (and of N), while the interpolation per HH increases only N (per definition not changing the number of uniquely identified HHS).

Online Appendix V - Table 19 – Determinants of earnings (harvest season). Alternative measures for the timing of the harvest.

	Dep. Var.: $ln(Y_{ht}^{harv})$					
	(1)	(2)	(3)	(4)	(5)	(6)
ln(N)	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)
S.D. 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)
Estaca ID Mean	-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)
Estaca ID MDEV	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)
Estaca ID S.D.	-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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,470	1,470	1,455	1,455	1,455	1,455
Number of id	324	324	320	320	320	320
R-squared	0.844	0.844	0.840	0.840	0.843	0.843

VI. Limited samples, indicators for sources of data, and data quality

Besides glossing over, crossing out, and emending individual information on household members and *estacas*, *Ibicaba*'s accountants occasionally crossed out entire households in the *Registries*, in what appears to have been a nullification of such accounts (75 observations); in other cases, only earnings were crossed out (32 observations). These cases have been coded as the indicators *Nullified* and *Crossed-out earnings*. A further 22 observations record households that left *Ibicaba* during the harvest season, coded as the indicator *Quit*. Online Appendix - Table 20 shows that results are unaffected by dropping any of these cases.

Some ledgers also include yearly summaries with data on demographics (142 observations), number of coffee groves cultivated (58 observations), and housing (132 observations) aggregated at the household level. This implies that the more ingrained data on *estaca* IDs or on disaggregated demographic characteristics cannot be recovered for these cases. Similarly, some accounts and summary tables record demographic data for household heads only. Of these, 132 observations correspond to the heads of households otherwise known in the sample, whose composition can be derived from other years or other ledgers; another 60 are, however, from households otherwise not listed in the transcribed sources. Online Appendix V - Table 21 reports the results of dropping each of these cases. Dropping the summaries of cultivated coffee groves reproduces the baseline by definition, as the latter requires the location of *estacas*. The remaining results are similarly robust, especially the conclusions on household labor composition and the absence of taste-based discrimination in remuneration. The only difference worth noting is that household size becomes significant (at the 5% level) as a determinant of earnings in the lean season when entries for household heads of known households are dropped.

Online Appendix V - Table 20 (Panel A) – Determinants of earnings. Samples excluding nullified accounts, accounts with crossed-out earnings, and households that quit

	Dep. Var.: $ln(Y_{ht}^{cult})$			Dep. Var.: $ln(Y_{ht}^{harv})$		
	Sample excluding:			Sample excluding:		
	Nullified	Crossed-out earnings	Quit	Nullified	Crossed-out earnings	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)
Estaca ID Mean	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)
Estaca ID MDEV	-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)
Estaca ID S.D.	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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,330	2,345	2,358	1,458	1,465	1,470
Number of id	507	511	512	320	324	324
R-squared	0.861	0.863	0.862	0.848	0.845	0.844

Online Appendix V - Table 20 (Panel B) – Testing for taste-based discrimination in prices earned. Samples excluding nullified accounts, accounts with crossed-out earnings, and households that quit

	Dep. Var.: Avg. price earned for cultivating one coffee grove (lean season)						Dep. Var.: Avg. price earned for harvesting one *alqueire* (harvest season)					
	Residuals obtained from regressions conditional on:						Residuals obtained from regressions conditional on:					
	Nullified		Crossedout_income		Quit		Nullified		Crossedout_income		Quit	
	Estimations based on sample						Estimations based on sample					
	Unrestrict.	Nullified	Unrestrict.	Crossed-out earnings	Unrestrict.	Quit	Unrestrict.	Nullified	Unrestrict.	Crossed-out earnings	Unrestrict.	Quit
TFP agr. task	0.00139***	0.00145***	0.00143***	0.00143***	0.00144***	0.00144***	0.00745***	0.00821***	0.00732***	0.00738***	0.00731***	0.00731***
	(0.000295)	(0.000286)	(0.000279)	(0.000281)	(0.000279)	(0.000279)	(0.00277)	(0.00296)	(0.00268)	(0.00269)	(0.00268)	(0.00268)
Non-Whites	0.000317	0.000312	0.000320	0.000321	0.000321	0.000321	0.00463	0.00573	0.00498	0.00499	0.00500	0.00500
	(0.000207)	(0.000213)	(0.000207)	(0.000207)	(0.000207)	(0.000207)	(0.00376)	(0.00378)	(0.00372)	(0.00373)	(0.00372)	(0.00372)
Italians	-7.32e-05	-0.000105	-8.39e-05	-8.42e-05	-8.60e-05	-8.60e-05	0.00445	0.00440	0.00442	0.00444	0.00443	0.00443
	(0.000231)	(0.000235)	(0.000230)	(0.000230)	(0.000230)	(0.000230)	(0.00398)	(0.00404)	(0.00396)	(0.00397)	(0.00396)	(0.00396)
Germanics	0.000478	0.000455	0.000473	0.000472	0.000473	0.000473	0.000956	0.00101	0.000886	0.000915	0.000883	0.000883
	(0.000446)	(0.000455)	(0.000446)	(0.000446)	(0.000446)	(0.000446)	(0.00425)	(0.00429)	(0.00426)	(0.00426)	(0.00426)	(0.00426)
Portuguese	-0.000262	-0.000291	-0.000265	-0.000271	-0.000269	-0.000269	0.00598	0.00604	0.00605	0.00603	0.00601	0.00601
	(0.000254)	(0.000260)	(0.000254)	(0.000256)	(0.000254)	(0.000254)	(0.00470)	(0.00473)	(0.00470)	(0.00472)	(0.00470)	(0.00470)
Spaniards	-0.000174	-0.000192	-0.000183	-0.000187	-0.000184	-0.000184	-0.0138	-0.0136	-0.0136	-0.0137	-0.0136	-0.0136
	(0.000427)	(0.000425)	(0.000425)	(0.000425)	(0.000424)	(0.000424)	(0.00925)	(0.00927)	(0.00930)	(0.00928)	(0.00930)	(0.00930)
East. Europeans	-0.000550	-0.000581	-0.000557	-0.000553	-0.000556	-0.000556	0.00348	0.00340	0.00344	0.00340	0.00342	0.00342
	(0.000736)	(0.000740)	(0.000734)	(0.000730)	(0.000733)	(0.000733)	(0.00380)	(0.00386)	(0.00377)	(0.00378)	(0.00378)	(0.00378)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,360	2,331	2,360	2,347	2,360	2,360	1,470	1,458	1,470	1,465	1,470	1,470
Number of id	513	507	513	512	513	513	324	320	324	324	324	324

Online Appendix V - Table 21 (Panel A) – Determinants of earnings. Samples excluding data from summary tables or data from household heads only

	Dep. Var.: $ln(Y_{ht}^{cult})$						Dep. Var.: $ln(Y_{ht}^{harv})$					
	Sample excluding						Sample excluding					
		Summary tables			Info on HH head, only			Summary tables			Info on HH head, only	
	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)
Estaca ID Mean	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)
Estaca ID MDEV	-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)
Estaca ID S.D.	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)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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
Number of id	508	512	511	507	510	507	320	324	324	320	321	323
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

Online Appendix V - Table 21 (Panel B) – Testing for taste-based discrimination in prices earned. Samples excluding data from summary tables or data from household heads only

Physical output												
	Dep. Var.: $ln(Cult_{ht})$						Dep. Var.: $ln(Harv_{ht})$					
	Sample excluding						Sample excluding					
	Summary tables				Info on HH head, only		Summary tables				Info on HH head, only	
	Demogr.	Coffee groves	Housing	All	Known HH	Unknown HH	Demogr.	Coffee groves	Housing	All	Known HH	Unknown HH
ln(HH size)	0.0474	0.0264	0.0297	0.0508	0.106**	0.0266	-0.104	-0.112	-0.110	-0.101	-0.0358	-0.108
	(0.0515)	(0.0484)	(0.0496)	(0.0530)	(0.0496)	(0.0485)	(0.0761)	(0.0733)	(0.0754)	(0.0783)	(0.0762)	(0.0741)
ln(# members 7-60y)	0.346***	0.333***	0.328***	0.342***	0.334***	0.333***	0.395***	0.393***	0.391***	0.393***	0.379***	0.393***
	(0.0413)	(0.0405)	(0.0415)	(0.0424)	(0.0421)	(0.0405)	(0.0617)	(0.0615)	(0.0634)	(0.0636)	(0.0633)	(0.0616)
Estaca ID Mean	3.55e-06	3.67e-06	5.80e-06	5.77e-06	4.66e-06	3.52e-06	-2.96e-05*	-3.14e-05*	-3.02e-05*	-2.82e-05	-3.07e-05*	-3.21e-05*
	(2.57e-05)	(2.62e-05)	(2.63e-05)	(2.58e-05)	(2.55e-05)	(2.62e-05)	(1.69e-05)	(1.74e-05)	(1.77e-05)	(1.72e-05)	(1.65e-05)	(1.74e-05)
Estaca ID MDEV	-3.20e-05	-4.56e-05	-5.72e-05	-4.37e-05	8.32e-05	-4.86e-05	-8.11e-05	-7.94e-05	-8.48e-05	-8.70e-05	7.97e-06	-0.000106
	(8.70e-05)	(8.74e-05)	(8.81e-05)	(8.77e-05)	(8.34e-05)	(8.89e-05)	(0.000103)	(0.000107)	(0.000107)	(0.000103)	(0.000103)	(0.000107)
Estaca ID S.D.	2.23e-05	3.25e-05	3.71e-05	2.69e-05	-6.24e-05	3.48e-05	4.69e-05	4.92e-05	5.12e-05	4.90e-05	-1.49e-05	6.87e-05
	(6.22e-05)	(6.21e-05)	(6.30e-05)	(6.30e-05)	(5.98e-05)	(6.33e-05)	(7.98e-05)	(8.32e-05)	(8.34e-05)	(8.00e-05)	(8.19e-05)	(8.32e-05)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,292	2,360	2,316	2,248	2,288	2,354	1,598	1,641	1,599	1,556	1,596	1,638
Number of id	509	513	512	508	511	508	331	335	333	329	332	333
R-squared	0.501	0.496	0.496	0.501	0.513	0.496	0.812	0.816	0.814	0.810	0.813	0.816

Discrimination in pay												
	Dep. Var.: Avg. price earned for cultivating one coffee grove (lean season)						Dep. Var.: Avg. price earned for harvesting one *alqueire* (harv. season)					
TFP agr. task	0.00158***	0.00144***	0.00143***	0.00157***	0.00148***	0.00144***	0.00754***	0.00731***	0.00731***	0.00754***	0.00789***	0.00741***
	(0.000278)	(0.000279)	(0.000280)	(0.000278)	(0.000286)	(0.000279)	(0.00274)	(0.00268)	(0.00268)	(0.00275)	(0.00290)	(0.00271)
Non-Whites	0.000265	0.000321	0.000309	0.000254	0.000291	0.000325	0.00482	0.00500	0.00498	0.00480	0.00476	0.00493
	(0.000206)	(0.000207)	(0.000208)	(0.000207)	(0.000207)	(0.000208)	(0.00377)	(0.00372)	(0.00372)	(0.00377)	(0.00379)	(0.00372)
Italians	-0.000129	-8.60e-05	-9.13e-05	-0.000134	-7.63e-05	-8.04e-05	0.00454	0.00443	0.00443	0.00454	0.00459	0.00440
	(0.000230)	(0.000230)	(0.000231)	(0.000231)	(0.000230)	(0.000230)	(0.00398)	(0.00396)	(0.00396)	(0.00398)	(0.00399)	(0.00396)
Germanics	0.000441	0.000473	0.000476	0.000444	0.000470	0.000481	0.00103	0.000883	0.000883	0.00103	0.00110	0.000846
	(0.000449)	(0.000446)	(0.000451)	(0.000455)	(0.000445)	(0.000447)	(0.00424)	(0.00426)	(0.00426)	(0.00424)	(0.00422)	(0.00427)
Portuguese	-0.000302	-0.000269	-0.000282	-0.000315	-0.000202	-0.000265	0.00606	0.00601	0.00599	0.00604	0.00614	0.00599
	(0.000253)	(0.000254)	(0.000255)	(0.000254)	(0.000254)	(0.000255)	(0.00472)	(0.00470)	(0.00470)	(0.00472)	(0.00473)	(0.00470)
Spaniards	-0.000244	-0.000184	-0.000207	-0.000271	-6.78e-05	-0.000171	-0.0143	-0.0136	-0.0136	-0.0143	-0.0147	-0.0136
	(0.000421)	(0.000424)	(0.000427)	(0.000424)	(0.000463)	(0.000424)	(0.00917)	(0.00930)	(0.00930)	(0.00917)	(0.00924)	(0.00931)
East. Europeans	-0.000521	-0.000556	-0.000581	-0.000546	-0.000569	-0.000566	0.00369	0.00342	0.00342	0.00370	0.00374	0.00335
	(0.000726)	(0.000733)	(0.000718)	(0.000711)	(0.000786)	(0.000743)	(0.00379)	(0.00378)	(0.00378)	(0.00379)	(0.00380)	(0.00378)
Full set of controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,292	2,360	2,316	2,248	2,288	2,354	1,427	1,470	1,470	1,427	1,426	1,469
Number of id	509	513	512	508	511	508	320	324	324	320	321	323

As discussed throughout the paper and in Online Appendix I, data quality varied substantially across sources. This variation should already be captured by the year-FEs in the baseline specifications, but the following robustness checks control for the sources of observations explicitly. The *Source** variable identify the ledger in the *Ibicaba Collection* containing the corresponding information on demographics (*Source01*), coffee cultivation (*02*), housing (*03*), and harvest (*04*); Online Appendix - Table 22 tabulates them. Therefore, the *Source** variables capture the overall quality of the sources and serve as an indicator of the abilities and competencies of *Ibicaba*'s managerial and accounting personnel.

In principle, it is also possible to control for the number of the PDF image from which each observation derives. The variables *PDF01-03* do that for the controls on demographics, coffee cultivation, and housing, respectively. An important caveat, however, is that a single point in the dataset could have been derived from multiple pages. Data on harvesting, in particular, were frequently spread over numerous pages recording volumes harvested daily during the season. I therefore dropped these controls, since no single page can be assigned to such observations, and replaced them by the indicator *Multiple pages (harv)* instead, which marks cases in which harvest data were drawn from more than one page of a ledger.

The paper's conclusions are unaltered after controlling for these additional variables (Online Appendix V - Table 23), which cements confidence in the data generating process – both as *Ibicaba*'s accountants and managers produced the records and as those are now interpreted as sources of quantitative data in economic history. The controls for sources and pages have no effect on the estimates for the lean season. For the harvest season, however, *Source03* is statistically significant at the 1%-level in the specification that does not control for *PDF**. Where those controls are included, *PDF03* is significant and positive: holding the ledger constant, higher page numbers are associated with higher earnings, already net of the time trends captured by the year-FEs.

Online Appendix V - Table 22 – Tabulation of the number of observations by sources from the *Ibicaba Collection*

Ledger #	Var. *Source01*		Var. *Source02*		Var. *Source03*		Var. *Source04*		
	Freq.	Percent	Freq.	Percent	Freq.	Percent	Ledger	Freq.	Percent
0029	332	11.17	648	18.93	597	17.48	0005	651	13.97
0043	787	26.48	923	26.96	966	28.28	0013	717	15.38
0062	835	28.10	835	24.39	835	24.44	0021	12	0.26
0083	530	17.83	556	16.24	538	15.75	0025	842	18.06
0093	33	1.11	33	0.96	33	0.97	0040	1,117	23.96
0098	20	0.67	20	0.58	20	0.59	0068	340	7.29
0099	45	1.51	38	1.11	50	1.46	0077	982	21.07
0100	70	2.36	51	1.49	58	1.70			
BPML0004	320	10.77	320	9.35	319	9.34			
Total	2,972	100.00	3,424	100.00	3,416	100.00		4,661	100.00

Online Appendix V - Table 23 (Panel A) – Determinants of earnings. Controls for sources and PDF images from which observations were compiled

	Dep. Var.: $ln(Y_{ht}^{cult})$		Dep. Var.: $ln(Y_{ht}^{harv})$	
Source 01	0.00280	0.0121	0.0122*	-0.0380
	(0.00471)	(0.0226)	(0.00666)	(0.0375)
Source 02	0.00671	-0.00898	-0.00196	-0.116*
	(0.00533)	(0.0234)	(0.00764)	(0.0644)
Source 03	-0.00542	-0.00458	-0.0186***	0.129*
	(0.00418)	(0.00583)	(0.00618)	(0.0698)
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)
Multiple pages (harv)				0.0859
				(0.0919)
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)
Estaca ID Mean	4.41e-06	5.16e-06	-2.91e-05	-2.71e-05
	(2.57e-05)	(2.56e-05)	(2.01e-05)	(1.96e-05)
Estaca ID MDEV	-5.45e-05	-5.95e-05	7.44e-05	6.17e-05
	(8.51e-05)	(8.54e-05)	(0.000135)	(0.000137)
Estaca ID S.D.	3.73e-05	4.01e-05	-6.71e-05	-5.79e-05
	(6.05e-05)	(6.08e-05)	(0.000112)	(0.000113)
Full set of controls	Yes	Yes	Yes	Yes
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) Specifications for the lean season (physical output and earnings) do not control for *Source04*, as that variable refers solely to the harvest season. (2) *Source04* is omitted due to collinearity in the specifications referring to the harvest season (physical output and income).

Online Appendix V - Table 23 (Panel B) – Testing for taste-based discrimination in prices earned. Controls for sources and PDF images from which observations were compiled

	Physical output			
	Dep. Var.: $ln(Cult_{ht})$		Dep. Var.: $ln(Harv_{ht})$	
Source 01	0.00276	0.0146	0.00988	-0.00569
	(0.00474)	(0.0227)	(0.00659)	(0.0323)
Source 02	0.00676	-0.0118	0.00121	-0.00498
	(0.00537)	(0.0235)	(0.00774)	(0.0333)
Source 03	-0.00556	-0.00496	-0.0214***	-0.0170
	(0.00417)	(0.00587)	(0.00615)	(0.0110)
PDF01		0.00178		-0.00228
		(0.00316)		(0.00459)
PDF02		-0.00266		-0.000752
		(0.00330)		(0.00457)
PDF03		8.19e-05		0.000612
		(0.000554)		(0.00129)
Multiple pages (harv)	-	-	-	-0.0279
				(0.131)
ln(HH size)	0.0241	0.0251	-0.107	-0.105
	(0.0485)	(0.0487)	(0.0727)	(0.0740)
ln(# members 7-60y)	0.334***	0.335***	0.391***	0.389***
	(0.0401)	(0.0400)	(0.0601)	(0.0609)
Estaca ID Mean	4.95e-06	5.75e-06	-3.18e-05*	-2.94e-05*
	(2.61e-05)	(2.60e-05)	(1.78e-05)	(1.76e-05)
Estaca ID MDEV	-5.65e-05	-6.20e-05	-7.22e-05	-7.89e-05
	(8.72e-05)	(8.75e-05)	(0.000113)	(0.000112)
Estaca ID S.D.	3.84e-05	4.14e-05	4.42e-05	4.74e-05
	(6.21e-05)	(6.24e-05)	(8.76e-05)	(8.68e-05)
Full set of controls	Yes	Yes	Yes	Yes
Observations	2,344	2,344	1,632	1,632
Number of id	509	509	333	333
R-squared	0.498	0.498	0.817	0.818

	Discrimination in pay			
	Dep. Var.: Avg. price earned for cultivating one coffee grove (lean season)		Dep. Var.: Avg. price earned for harvesting one *alqueire* (harv. season)	
TFP agr. task	0.00144***	0.00145***	0.00726***	0.00720***
	(0.000278)	(0.000276)	(0.00269)	(0.00269)
Non-Whites	0.000314	0.000314	0.00500	0.00504
	(0.000207)	(0.000207)	(0.00371)	(0.00372)
Italians	-9.75e-05	-9.89e-05	0.00433	0.00435
	(0.000229)	(0.000229)	(0.00397)	(0.00397)
Germanics	0.000464	0.000463	0.000833	0.000828
	(0.000446)	(0.000446)	(0.00428)	(0.00428)
Portuguese	-0.000282	-0.000282	0.00592	0.00594
	(0.000255)	(0.000255)	(0.00470)	(0.00470)
Spaniards	-0.000194	-0.000197	-0.0135	-0.0136
	(0.000423)	(0.000423)	(0.00934)	(0.00934)
East. Europeans.	-0.000571	-0.000573	0.00344	0.00334
	(0.000736)	(0.000734)	(0.00379)	(0.00381)
Full set of controls	Yes	Yes	Yes	Yes
Observations	2,344	2,344	1,462	1,462
Number of id	509	509	322	322

Note: *Cf.* notes on Online Appendix V - Table 23 (Panel A)

A final set of robustness checks assesses the effect that accounting errors might have had on *colonos*' earnings during the lean season.[13] The variable *Accounting error* measures the errors committed by the accountants in computing the earnings of *colono* households for pruning coffee, as detailed in Online Appendix I.

The first two columns of Table 24 add the variable *Accounting error* to the specification, with column (2) excluding the outlier of 120.694 *milréis* by restricting the sample to errors below 120 *milréis*. The other columns vary the sample by including only observations without accounting errors (column 3); only observations with accounting errors (column 4); and all observations except the outlier of 120.694 *milréis* (column 5).

The additional control and the subsample analyses do not affect previous conclusions. In columns 1 and 2, *Accounting error* is marginally significant (at 10%-level), but its economic effect is minimal: evaluated at the sample mean, its coefficient implies a reduction of only 0.05% in mean household earnings.[14] Moreover, in the sample restricted to households with accounting errors only (column 4), household size becomes significant and positive; the coefficient on the number of economically active members is correspondingly smaller than in the other samples, so that the two variables combined capture much the same effect.

Finally, Table 25 reports the conditional correlations between *Accounting error* and households' ethnolinguistic origins. The specifications vary in whether year-FEs are included and whether the error outlier is dropped. The ethnolinguistic origin of households is not a significant explanator of accounting errors in any model, and the corresponding R2s never exceed 1.5%.

[13] Recall that all accounting errors discussed in this paper refer solely to earnings in coffee cultivation during the lean season. These conclusions cannot be extrapolated to other potential accounting errors harvesting, or in *colonos'* current accounts. Please refer to Online Appendix I for a thorough discussion. In short, these robustness checks are not a historical auditing; rather, they intend to explore only the impact that potential accounting errors might have as measurement errors.

[14] (exp(-0.00763)-1)*100*0.065767.

Online Appendix V - Table 24 – Determinants of earnings. The impact of accounting errors (lean season)

	Dep. Var.: $ln(Y_{ht}^{cult})$				
	Specification		Restricted samples		
Accounting errors	Yes	< 120 *milréis*	No	Yes	< 120 *milréis*
Accounting error	-0.00763*	-0.00763*			
	(0.00401)	(0.00401)			
ln(price)	1.287***	1.287***	1.162***	1.239***	1.287***
	(0.0885)	(0.0885)	(0.0883)	(0.112)	(0.0884)
ln(N)	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)
Estaca ID Mean	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)
Estaca ID MDEV	-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)
Estaca ID S.D.	2.97e-05	2.97e-05	6.51e-06	-0.000600	2.73e-05
	(6.28e-05)	(6.28e-05)	(5.17e-05)	**	(6.31e-05)
				(0.000273)	
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)
Saltinho & Pinheiro	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)
Full set of controls	Yes	Yes	Yes	Yes	Yes
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) Column (1) adds "account_error" variable to the baseline specification, without any changes in the sample; (3) "<120 *milréis*" refers to the sample dropping the accounting error outlier of 120.649 réis.

Online Appendix V - Table 25 – Conditional correlations between ethnolinguistic origins and accounting errors (controls included separately in Panel A, and jointly in Panel B

	Dep. var.: Accounting errors	
Panel A - Ethnol. origins included jointly		
Non-Whites	-0.0251	0.0170
	(0.337)	(0.301)
Italians	-0.204	-0.173
	(0.281)	(0.250)
Germanics	-0.240	-0.251
	(0.290)	(0.304)
Portuguese	-0.243	-0.231
	(0.287)	(0.270)
Spaniards	-0.231	-0.193
	(0.280)	(0.246)
Eastern Europeans.	-0.218	-0.163
	(0.281)	(0.234)
Year-FE	No	Yes
Dropping outlier	No	No
Full set of controls otherwise	Yes	Yes
Observations	2,803	2,803
R-squared	0.001	0.015

Panel B - Ethnol. origins included separately				
White Brazilians	0.207	0.184	-0.0670	-0.0625
	(0.280)	(0.253)	(0.0561)	(0.0661)
Non-Whites	0.158	0.185	0.201	0.185
	(0.193)	(0.189)	(0.188)	(0.189)
Italians	-0.0632	-0.0518	0.0313	0.0233
	(0.103)	(0.0871)	(0.0414)	(0.0440)
Germanics	-0.0672	-0.107	-0.0237	-0.0160
	(0.0918)	(0.126)	(0.0809)	(0.0815)
Portuguese	-0.0766	-0.0933	-0.0291	-0.0327
	(0.0839)	(0.0946)	(0.0692)	(0.0710)
Spaniards	-0.0567	-0.0470	-0.0138	-0.00440
	(0.0482)	(0.0550)	(0.0219)	(0.0194)
Eastern Europeans	-	-	-	-
Year-FE	No	Yes	No	Yes
Dropping outlier	No	No	Yes	Yes
Full set of controls otherwise	Yes	Yes	Yes	Yes
Observations	2,905	2,905	2,904	2,904
R-squared	0.001	0.015	0.000	0.011

Notes: (1) Robust standard errors in parentheses whenever homoskedasticity has been rejected at the 10 percent level.

Online Appendix VI

The immediate post-abolition (1888-1890) in *Ibicaba*: Additional insights from Ledger Biblioteca Paulo Masuti Levy (BPML) #0004

Contents

I. Immediate Managers' assessments of laborers' behavior (*Ibicaba*, 1890)

Covering the years 1888 to 1890, ledger BPML #0004 is the last of the *Registries* produced by *Vergueiro & Co*. This uniquely rich source preserves detailed household characteristics, including gradations of skin color as perceived by the plantation managers, literacy, religion, and occupations (including non-agricultural), as discussed in the main text. In addition, the source provides a classification of household behavior in 1890, according to the perceptions of the plantation's managers. Household behaviors were classified as *máo* [SIC] (bad/poor), *bom* (good), and *exemplar* (exemplary). Online Appendix VI - Table 1 tabulates the number of households in each category, distinguished by ethnolinguistic origins.

Online Appendix VI - Table 1 – Household's behavior according to *Ibicaba's* manager(s), by ethnolinguistic origin (1890)

	Bad/Poor	Good	Exemplary	Total
White 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. Europeans	1	11	0	12
Total	1	104	8	113

Noticeably, among foreigners, the exemplary status was assigned only to the Portuguese. The other four households considered exemplary were all Brazilians, three of them Non-Whites. Contrary to racist outbursts in the press and public discourses – much highlighted by the historiography –, which disparaged the willingness of freedmen to become free fieldhands, practices on the ground, at *Ibicaba* in 1890, show managers satisfied with the behavior of *Non-White* households and, by implication, with their labor in the immediate post-Abolition.

II. The Likelihood of signing *colonato* contracts: The role of ethnolinguistic origins and other household characteristics

This section uses the additional data from ledger BPML #0004 to test whether the likelihood of households signing *colonato* contracts correlate with their ethnolinguistic origins, conditional on other controls (*cf.* Table 1 in the main text for the distribution of occupations, agricultural and otherwise, in 1889-1890). LPM models augmented stepwise with controls for household characteristics are reported in Online Appendix VI - Tables 2, 4, and 6. The corresponding odd-ratios from logit models are reported in Tables 3, 5, and 7. These sets of tables differ in their controls for ethnolinguistic origins: a binary category for *Whites* and *Non-Whites* (Table 2), gradations of skin color according to the source (Table 4), and categorizations of ethnolinguistic origins as in the baseline (Table 6). Predicted probabilities of signing a *colonato* contract were computed from the most complete model (column 5 of each table), evaluated at the means of (1) the sample; (2) *Non-Whites*; (3) *All Whites* (Brazilians and foreigners); and (4) *White Brazilians* only.

Consonant with the literature on the Age of Mass Migration to Brazil, the LPM and odd-ratios from the logit estimates reveal the relevance of the number of economically active members for a household's likelihood of signing a *colonato* contract. In general, the negative effect of being *Non-White* on the likelihood of being a *colono* household becomes either smaller in absolute terms or statistically weaker once other household characteristics are controlled for, in particular the number of economically active members (compare columns 1 and 4, for which the number of observations is the same across Tables 2, 4, and 6).[1]

At the same time, however, the negative effect of being *Black* on the likelihood of signing a *colonato* contract remains statistically significant in column (4) across Tables 2-5. Also, the predicted probabilities reported in Tables 2 and 3 show that *Non-White* households had a smaller chance of being *colonos* than *Whites*, irrespective of the covariate profile used. However, Tables 6 and 7 show that this advantage was driven by *Italians*, rather than by all white households.[2] *Non-White* households in fact had a higher predicted probability of being *colono* than *White Brazilians*, and closely tracked the *Portuguese* – with the minimal differences in their predicted probabilities reversing between the LPM and logit estimates.

Consistent with the broader conclusions of this paper, these results show the centrality of the number of economically active members for becoming a *colono*, also in the immediate post-Abolition era. More importantly, results show that *Non-White* households had a lower likelihood of signing a *colonato* contract, both because of their demographics characteristics and a *ceteris paribus* overall disadvantages relative to *Italian* households. Nonetheless, in immediate post-Abolition *Ibicaba*, *Non-White* household*s* were certainly not barred from *colonato*, and their likelihood of becoming *colonos* was practically as high as that of the *Portuguese* and higher than that of *White Brazilians*.

---

[1] Households with literate members were less likely to be *colonos*, perhaps because they were more frequently recruited to managerial or non-agricultural tasks. This may explain why *Germanics* were less likely than *Non-White* households to be *colonos* (*cf.* the predicted probabilities in Panel B of Tables 6 and 7).

[2] *Eastern European* households show a near-100% probability of being *colonos* by construction, as that was their only recorded occupation in the sample.

Online Appendix VI - Table 2 – LPM estimates of *colono* status (1 = *Colono*, 0 = Otherwise), with binary identifiers for *Non-Whites* (1888-1890)

	Dep. Var.: *Colono* (= 1)				
	(1)	(2)	(3)	(4)	(5)
Non-Whites	-0.243***	-0.206**	-0.229***	-0.167**	-0.116
	(0.0853)	(0.0916)	(0.0861)	(0.0770)	(0.0811)
Literate		-0.238**			-0.250**
		(0.109)			(0.100)
Religion			-0.0595		-0.00132
			(0.106)		(0.0726)
HH size				-0.0367**	-0.0268
				(0.0154)	(0.0164)
# members 7-60y				0.0915***	0.0812***
				(0.0209)	(0.0224)
Year-FE	Yes	Yes	Yes	Yes	Yes
Observations	314	209	307	314	208
R-squared	0.118	0.177	0.107	0.208	0.261
r2_a	0.109	0.161	0.0949	0.195	0.235

Notes: (1) Standard errors clustered at the household level. (2) All models include a constant.

Predicted probabilities of *colono* = 1
from (5), at mean values of

	Sample	Non-Whites	All Whites	White Brazilians
Non-Whites	0.770***	0.695***	0.783***	0.700***
	(0.0752)	(0.0782)	(0.0750)	(0.0850)
Whites	0.885***	0.810***	0.899***	0.816***
	(0.0227)	(0.0329)	(0.0224)	(0.0437)

Online Appendix VI - Table 3 – Odds-Ratios from logit estimates of *colono* status (1 = *Colono*, 0 = Otherwise), with binary identifiers for *Non-Whites* (1888-1890)

	Dep. Var.: *Colono* (= 1)				
	(1)	(2)	(3)	(4)	(5)
Non-Whites	0.277***	0.215***	0.277***	0.482*	0.478
	(0.113)	(0.114)	(0.120)	(0.187)	(0.255)
Literate		0.123***			0.104***
		(0.0879)			(0.0886)
Religion			0.636		0.885
			(0.452)		(0.697)
HH size				0.698**	0.658*
				(0.106)	(0.145)
# members 7-60y				2.615***	2.952***
				(0.601)	(1.023)
Year-FE	Yes	Omitted	Yes	Yes	Omitted
Observations	231	138	226	231	137
r2_p	0.0518	0.133	0.0479	0.202	0.283

Notes: (1) Standard errors clustered at the household level. (2) All models include a constant. (3) "-": omitted due to collinearity. Year-FEs reported as "Omitted" were omitted by the estimation procedure.

Predicted probabilities of *colono* = 1
from (5), at mean values of

	Sample	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)

Online Appendix VI - Table 4 – LPM estimates of *colono* status (1 = *Colono*, 0 = Otherwise), with skin colors as in the source (1888-1890)

	Dep. Var.: *Colono* (= 1)				
	(1)	(2)	(3)	(4)	(5)
Trigueiros	0.162***	0.0549***	0.151***	0.208***	0.0870**
	(0.0276)	(0.0163)	(0.0277)	(0.0387)	(0.0362)
Pardos	-0.00315	-	-0.00372	0.0876***	-
	(0.00311)		(0.00350)	(0.0220)	
Blacks	-0.259***	-0.217**	-0.243***	-0.182**	-0.123
	(0.0875)	(0.0973)	(0.0888)	(0.0786)	(0.0865)
Literate		-0.262**			-0.264***
		(0.111)			(0.0993)
Religion			-0.0546		0.00374
			(0.106)		(0.0712)
HH size				-0.0410***	-0.0302*
				(0.0155)	(0.0170)
# members 7-60y				0.0970***	0.0862***
				(0.0211)	(0.0229)
Year-FEs	Yes	Yes	Yes	Yes	Yes
Observations	315	210	308	315	209
R-squared	0.125	0.192	0.113	0.220	0.280
r2_a	0.110	0.172	0.0953	0.202	0.252

Notes: (1) Standard errors clustered at the household level. (2) All models include a constant.

Predicted probabilities of *colono* = 1
from (5), at mean values of

	Sample	Non-Whites	All Whites	White Brazilians
Blacks	0.758***	0.682***	0.774***	0.684***
	(0.0807)	(0.0829)	(0.0805)	(0.0908)
Trigueiros	0.967***	0.892***	0.983***	0.894***
	(0.0292)	(0.0363)	(0.0299)	(0.0480)
Whites	0.880***	0.805***	0.896***	0.807***
	(0.0227)	(0.0323)	(0.0223)	(0.0428)

Online Appendix VI - Table 5 – Odds-Ratios from logit estimates of *colono* status (1 = *Colono*, 0 = Otherwise), with skin colors as in the source (1888-1890)

	Dep. Var.: *Colono* (= 1)				
	(1)	(2)	(3)	(4)	(5)
Blacks	0.262***	0.207***	0.265***	0.450**	0.462
	(0.108)	(0.112)	(0.116)	(0.179)	(0.253)
Literate		0.109***			0.101***
		(0.0761)			(0.0840)
Religion			0.663		0.894
			(0.471)		(0.700)
HH size				0.677**	0.648*
				(0.104)	(0.145)
# members 7-60y				2.699***	2.998***
				(0.614)	(1.031)
Year-FE	Yes	Omitted	Yes	Yes	Omitted
Observations	230	138	225	230	137
r2_p	0.0547	0.145	0.0501	0.211	0.299

Notes: Notes: (1) Standard errors clustered at the household level a. (2) All models include a constant. (3) "-": omitted due to collinearity. Year-FEs reported as "Omitted" were omitted by the estimation procedure. (4) "Pardo" and "Trigueiro" omitted; as such, the comparisons reported refer to "Blacks only" and "Whites" (reference group).

Predicted probabilities of *colono* = 1
from (5), at mean values of

	Sample	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)

Online Appendix VI - Table 6 – LPM estimates of *colono* status (1 = *Colono*, 0 = Otherwise), with ethnolinguistic categories (1888-1890)

	Dep. Var.: *Colono* (= 1)				
	(1)	(2)	(3)	(4)	(5)
Non-Whites	-0.116	-0.0299	-0.0491	-0.0661	0.0504
	(0.124)	(0.144)	(0.131)	(0.112)	(0.131)
Italians	0.183*	0.210*	0.203*	0.154*	0.203*
	(0.0942)	(0.124)	(0.104)	(0.0858)	(0.120)
Germanics	-0.222	-0.193	-0.378**	-0.239	-0.310
	(0.180)	(0.206)	(0.186)	(0.168)	(0.208)
Portuguese	0.151	0.146	0.166	0.0905	0.0724
	(0.110)	(0.142)	(0.118)	(0.110)	(0.143)
Spaniards	0.151	0.154	0.166	0.124	0.0870
	(0.109)	(0.155)	(0.119)	(0.102)	(0.164)
Eastern Eurps.	0.297***	0.312**	0.230*	0.303***	0.270*
	(0.0973)	(0.120)	(0.119)	(0.0899)	(0.146)
Literate		-0.104			-0.103
		(0.124)			(0.116)
Religion			0.123		0.0724
			(0.0806)		(0.119)
HH size				-0.0240	-0.0179
				(0.0162)	(0.0167)
# members 7-60y				0.0780***	0.0776***
				(0.0213)	(0.0225)
Year-FEs	Yes	Yes	Yes	Yes	Yes
Observations	320	211	309	320	210
R-squared	0.183	0.232	0.165	0.265	0.335
r2_a	0.162	0.198	0.140	0.241	0.294

Notes: (1) Standard errors clustered at the household level. (2) All models include a constant.

Predicted probabilities of *colono* = 1
from (5), at mean values of

	Sample	Non-Whites	All Whites	White Brazilians
White Brazilians	0.728***	0.641***	0.744***	0.706***
	(0.111)	(0.111)	(0.111)	(0.103)
Non-Whites	0.780***	0.693***	0.795***	0.758***
	(0.0697)	(0.0723)	(0.0692)	(0.0785)
Italians	0.943***	0.856***	0.958***	0.920***
	(0.0275)	(0.0334)	(0.0260)	(0.0543)
Germanics	0.414**	0.327*	0.429**	0.392**
	(0.192)	(0.197)	(0.194)	(0.183)
Portuguese	0.798***	0.711***	0.813***	0.776***
	(0.0950)	(0.102)	(0.0942)	(0.0938)
Spaniards	0.822***	0.735***	0.837***	0.800***
	(0.116)	(0.121)	(0.115)	(0.124)
Eastern Europs.	0.998***	0.911***	1.013***	0.976***
	(0.0829)	(0.0887)	(0.0851)	(0.102)

Online Appendix VI - Table 7 – Odds-Ratios from logit estimates of *colono* status (1 = *Colono*, 0 = Otherwise), with ethnolinguistic categories (1888-1890)

	Dep. Var.: *Colono* (= 1)				
	(1)	(2)	(3)	(4)	(5)
Non-Whites	0.791	0.763	0.996	1.308	1.634
	(0.500)	(0.542)	(0.641)	(0.847)	(1.397)
Italians	5.292**	6.705**	5.694***	5.357**	5.497
	(3.436)	(5.938)	(3.737)	(3.803)	(6.226)
Germanics	0.367	0.510	3.51e-07***	0.205	4.28e-08***
	(0.341)	(0.513)	(4.01e-07)	(0.278)	(7.11e-08)
Portuguese	3.194	2.179	3.194	2.449	1.453
	(2.564)	(1.892)	(2.564)	(2.398)	(1.846)
Spaniards	3.627	2.111	3.628	4.438	1.510
	(3.237)	(2.667)	(3.250)	(4.209)	(2.233)
Literate		0.339			0.366
		(0.275)			(0.364)
Religion			1.116e+06***		3.070e+06***
			(1.515e+06)		(4.500e+06)
HH size				0.776	0.708
				(0.127)	(0.178)
# members 7-60y				2.283***	3.043***
				(0.520)	(1.234)
Year-FE	Yes	Omitted	Yes	Yes	Omitted
Observations	222	127	215	222	126
r2_p	0.114	0.171	0.107	0.253	0.360

Notes: (1) Standard errors clustered at the household level. (2) All models include a constant. (3) Ethnolinguistic "Eastern Europs." omitted. (4) "-": omitted due to collinearity. . Year-FEs reported as "Omitted" were omitted by the estimation procedure.

Predicted probabilities of *colono* = 1
from (5), at mean values of

	Sample	Non-Whites	All Whites	White Brazilians
White 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)
Spaniards	0.903***	0.692***	0.948***	0.841***
	(0.103)	(0.250)	(0.0593)	(0.169)

III. Colony of work and quality of coffee groves by ethnolinguistic origins

Complementing Table 3 (main text) and Online Appendix III, Online Appendix VI - Table 8 reports the distribution of households by ethnolinguistic origins across colonies in *Ibicaba*, pooled for 1888-1890. The great majority of the labor force lived in *Colônia Ibicaba,* where *Italians* formed the largest group: 104 out of 147 households in that colony, out of 250 recorded in total. *Non-Whites*, Europeans, and *White Brazilians* all lived in *Colônia Ibicaba*, *Grama Larga*, *Abundância*, and *Saltinho & Pinheiro* colonies. The majority of *Non-White colonos* (15 out of the 39 in the pooled sample), however, lived in *Boa Esperança* colony, where they were the only ethnolinguistic group.

Ledger BPML #0004 also provides a unique classification of coffee bushes as *novos* (newly planted), *velhos* (old), and *dando fructos* (bearing fruits). Among these, some bushes further carry the classifications *com gramma* (with weed) and *falhos* (failed), the latter likely indicating gaps and dead trees among the coffee rows). In the aggregate, these classifications need to be treated with caution, since they were recorded by *Vergueiro & Co.* in its last season as *Ibicaba*'s proprietary firm, when the owners had every incentive to downplay the plantation's valuation. There is no evident reason, however, why managers would have biased entries on the quality of groves assigned to households of particular ethnolinguistic origins.

Online Appendix VI - Table 9 reports conditional correlations for the share of coffee bushes *bearing fruit*, *newly planted*, *with weed*, and *failed* among all bushes a household cultivated. The data are pooled for 1888-1890.[3] Controls include literacy, religion, colony of residence, household size and age composition, and year-FEs. The specifications vary in their measures of ethnolinguistic origins: binary *Non-White* indicator, gradations of skin color, and *Origins*.

The *Non-White* indicator and the gradations of skin color do not correlate with the shares of any type of coffee bush cultivated, except for those with weed; in that case, however, all ethnolinguistic groups have a statistically significant higher share of poorer bushes than *White Brazilians* if the *Origins*' categories are used instead. Moreover, while *Germanic* and *Eastern European* households correlate positively with the share of bushes *bearing fruit* and negatively with the share of *newly planted* ones, they also correlate positively with the share of bushes *with weed* – showing no general advantage over other groups in the quality of the bushes they cultivated.[4] Even more interestingly, the indicator for *Boa Esperança* colony, where only *Non-White* households resided, is associated with a statistically significant higher share of coffee bushes of better quality and a lower share of those of poorer quality.

These results take *Grama Larga* as the reference category for being an ordinary colony in 1888-1890, with the usual high share of *Italians* and a minority of *Non-Whites*, *Portuguese*, and *Eastern Europeans*. Online Appendix VI - Table 10 uses *Boa Esperança* as the reference instead. Results are not sensitive to this change. The only difference is that all groups other than *Non-Whites* have non-significant coefficients on the share of coffee bushes *with weed*.

[3] The category "old" was not analyzed due to sample size.
[4] For the *Germanics*, being *Protestant* almost completely nets out their origin's effect.

Online Appendix VI - Table 8 – Distribution of households by ethnolinguistic groups across colonies, *Ibicaba* (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

Online Appendix VI - Table 9 – Conditional correlates: Quality of coffee bushes cultivated (1888-1890) – Ref. category for colonies: *Grama Larga*

	Dep. Var.: Share of coffee trees											
	Bearing fruit			Newly planted			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 include the full set of independent variables, including HH size and age composition (not reported to save space and non-significant throughout) and a time trend; (3) All models include a constant.

Online Appendix VI - Table 10 – Conditional correlates: Quality of coffee bushes cultivated (1888-1890) – Ref. category for colonies: *Boa Esperaça*

	Dep. Var.: Share of coffee trees											
	Bearing fruit			Newly planted			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: *Cf.* notes on Table Online Appendix VI - Table 9.