Prosperity, Inequality and Elites: The Determinants of Political Office-holding in Nineteenth Century Antioquia[1]

Camilo García Jimeno[2] and James A. Robinson[3]

Abstract

The poor economic performance of Latin America in the 19th century is often tied to the incentives of those who wielded political power. But what determined who held this power and what were their incentives? Though some have emphasized the persistence of colonial elites, there was also much entry into politics during this period. In this paper we use unique micro data on incomes, family structure and family background from 19th century Antioquia to examine some of the determinants of local political office holding. We show that the most important determinant of who became mayor was personal income. Richer people were significantly more likely to be mayor. We also find some evidence that those who enjoyed non-labor income were more likely to become mayor. We find no evidence however that elite background predicts office-holding. We argue that these findings suggest that the reason Antioquia was the most economically dynamic part of Colombia in the 19th century was that it had much more functional politics than the rest of the country. Our data also allows us to construct credible estimates of inequality and income per-capita. The data is in line with current estimates of Colombian income per-capita in the 1850s suggesting it was about 30% of the U.S. level, but it also suggests that inequality was already high at this point even before the integration of Antioquia into the World economy.

[1]We thank Daron Acemoglu for helpful insights and advice, Jorge Yepes and Alvaro Vidal for their help at the Archive in Medellin, Mauricio Ruiz and Ángela Fonseca for their help at Universidad de los Andes, and Harvard University and the Banco de la República de Colombia for financial support.
[2] Massachusetts Institute of Technology, Department of Economics, 50 Memorial Drive, Cambridge MA 02142 e-mail: cgarcia@mit.edu.
[3] Harvard University, Department of Government, IQSS, 1737 Cambridge Street N309, Cambridge, MA 01238; e-mail: jrobinson@gov.harvard.edu.

1. Introduction

Colombia was a poor country in the 19th century and like most Latin American countries it stagnated while North America began to experience rapid economic growth as a result of assimilating the new technologies generated by the industrial revolution. Although earlier generations of scholars emphasized geographical or even cultural impediments to economic progress in Latin America, the more recent literature has explained this in terms of the political economy of institutions (Engerman and Sokoloff, 1997, 2005, Acemoglu, Johnson and Robinson 2001, 2002). More specifically, Latin American societies emerged from colonialism early in the 19th century with highly oligarchic and hierarchical political institutions where power was wielded by narrow elites. These elites structured economic institutions to generate rents for themselves at great cost to the rest of society. For instance, Engerman and Sokoloff (1997) and Sokoloff and Khan (1990) emphasize the varied background of inventors and patentees in 19th century United States. Such upward social mobility was very difficult in the stratified societies of Latin America.

This view of Latin American and Colombian development argues that the power relations of the colonial period reproduced themselves in the 19th and 20th centuries and economic institutions continued to further the interests of elites. Though this perspective is certainly consistent with some basic patterns in the data, the mechanisms by which this power reproduced itself are much less understood. A simple hypothesis would be that the reproduction of power and lack of social mobility was manifested in the persistent dominance of particular elites or elite family dynasties whose power was based in the ownership of land and labor (see Stein and Stein, 1970, Schwartz, 1996). There is indeed evidence for this in some Latin American countries, such as Brazil, Guatemala or Peru.[4]

[4] This is best documented in the case of Central America where the social structure appears to have been extremely static, even in Costa Rica. Stone (1990) documented the striking number of presidents since independence in various Central American countries that are related to two conquistadors, Cristóbal de Alfaro and Juan Vázquez de Coronado, who both arrived in Central America in the 16th century. He found that no less that 48 presidents were direct descendents of Alfaro, 25 of them in Costa Rica. Alfaro's descendents also include the Somoza dynasty who ruled Nicaragua for most of the 20th century until the Sandinista Revolution in 1979. 29 presidents were descended from Vázquez, including again the Somoza clan and many of the same presidents of Costa Rica related to Alfaro. See also Paige (1997) on Central American elites.

In some cases traditional elites were able to control politics directly; in other cases even if economic elites did not become politicians themselves they have been able to indirectly control the political system. In either case, in the same way as Key (1949, p. 211) noted in his seminal analysis of the control of North Carolina's economic oligarchy over politics: "The effectiveness of the oligarchy's control has been achieved through the elevation to office of persons fundamentally in harmony with its viewpoint."

But elsewhere, particularly in Colombia, the picture is much more complex. It is not obvious that one can trace the roots of either the elites of the Liberal and Conservative parties in 19th century Colombia to either colonial or economic elites and the specialist literature emphasizes both that politics itself was a channel of upward social mobility in Colombian society and that politicians came from various backgrounds, often professions such as lawyers (Safford, 1972, Uribe-Uran, 2000). The fierce political competition between Liberals and Conservatives, sometimes at the ballot box and sometimes on the battlefield, seems qualitatively different from the stylized picture of political stasis even if some major 19th century politicians, such as José Maria Obando and Tomás Cipriano de Mosquera certainly fit the image of Latin American caudillos. Indeed, democratic politics, at least after the introduction of universal male suffrage by the Liberals, was vibrant. For instance, Bushnell (1971) calculated that as many as 40% of adult males may have voted in the 1856 presidential election.[5] It is also the case, however, that 19th century Colombian elections were often marred by pervasive fraud and violence.[6]

That the Colombian case deviates from the stylized picture of 19th century Latin American politics is illustrated by the one quantitative study of the relationship between economic and political elites by Acemoglu, Bautista, Querubín and Robinson (2008). Using data on land ownership from the 1879 and 1890 *Catastros de Cundinamarca* and data on the identities of mayors they show that while there was some overlap between the identities or large landowners and mayors, it was much less than one might have conjectured. Moreover,

[5] Though after the 1863 Constitution voting rights were determined at the state level, several of them maintained universal suffrage (see Delpar, 1981).

[6] Outside of Colombia there are also many problems with simplistic notions of persisting elites. For instance, Taylor (1972) showed how frequently the great haciendas of the Oaxaca valley were sold or turned over during the colonial period and this appears to have been quite general a pattern in Mexico (see Van Young, 1983).

the interpretation that the authors give to the correlations they find is that the political class was in many ways autonomous from the economic elites and because elections and political institutions worked very imperfectly to make politicians accountable, the presence of economic elites led to better development outcomes because they were able to use their power to discipline politicians. In Cundinamarca political elites often appear to have been those with a comparative advantage in solving the collective action problem, exercising violence, or political entrepreneurs such as the local *Gamonales* who mobilized support for the Liberal and Conservative parties (Deas, 1971, Christie, 1979).

Nevertheless, the direct continuity of colonial elites is only one way to explain persistence. Even this would not be sufficient if the interests of those elites changed over time, as surely happened in many cases. In addition, as Acemoglu and Robinson (2007), observe, it is not necessary either since even if the identity of elites change, there can be path dependence of political strategies in the sense that new elites can find it desirable to adopt the strategies of old elites. In this case the underlying economic incentive environment can persist even if the recipients of the rents change. In support of this, Acemoglu et al. (2008) find that politicians accumulated land much more rapidly than non-politicians in Cundinamarca between 1879 and 1890, suggesting that political entrepreneurs used their power to expropriate the property of others and become economic elites.

All of this suggests that we need a much deeper understanding of the politics of 19th century Latin America. We need to understand how power was articulated, who held power and why, what factors created political power, how that power was used, and how it reproduced itself over time. While the literature on Latin American development has tended to stress economic wealth and landownership as sources of power, the above discussion and recent social science research has tended to stress that power is much more multi-faceted. Even in well functioning democracies other elements than one person one vote are relevant. Wealth, prestige, social connections, the success of collective action or the influence of pressure groups all play a role (Becker, 1983, Stigler, 1988). In weakly institutionalized polities[7], even the de facto exercise of power, like direct violence or intimidation, can play a central role in how political power is allocated. Other literatures in economics have recently

[7] According to Acemoglu, Robinson and Verdier (2004), weakly institutionalized polities are those where political institutions place few constraints on what politicians or those holding political power can do.

started to study the roles of social capital and of social networks in several different settings, which suggests they may also play a role in politics, as might be implied by Putnam's (1994) argument that social capital was critical in determining how democratic governments function. Theoretical work on social networks has also highlighted how these are important in facilitating communication, the diffusion of information, promotion of pro-social behaviors, etc. Social networks may help enforce informal contracts, may alleviate moral hazard problems in human interaction, may facilitate risk-sharing, etc. (Möbius and Szeidl, 2007), and the family may represent one of the most critical social networks (Kandori, 1992). Not only the presence of social or family networks may be important, but how heterogeneous they are. For instance, Alesina and La Ferrara (2000) show for U.S. data that more homogeneous communities have systematically higher levels of social interactions, leading to more social capital.

In this research we use some unique data from 19th Century Antioquia to investigate the role of some of the potential determinants of political officeholding, and by implication, political power. We collected two sources with information on incomes from the Historical Archive in Medellín, the *Catastros de Ingreso* of 1853 (income censuses) and *Listas de Contribuyentes para Caminos* of 1856 (Censuses of Roads Tax Contributors), both of which contain information on individual incomes. We also collected data on the identity of local mayors of districts and on the identity of elite families. Our data also allows us to look at the size and heterogeneity of families.

As we discuss in the next section, the study of Antioquia is particularly interesting, especially as compared to Cundinamarca, because the historical literature suggests that Antioqueño society and institutions differed quite a lot from the rest of the economy. In the 19th Century Antioquia had an open frontier like the U.S. west, and at least from the 1880s onwards experienced rapid economic growth on the basis of coffee exports. It was also the area of the country which first experienced industrialization. Moreover, though there has been little research on the political economy of growth in 19th century Antioquia, the existing historical literature has stressed the importance of family networks and their values as central to the Antioqueño society (e.g, Twinam, 1982, Uribe de Hincapié, 1998, Londoño, 2002).

Our analysis generates several main findings. First, personal income was strongly correlated with the likelihood of a person becoming a mayor. In Antioquia, as opposed to Cundinamarca, the rich became politicians rather than the other way round. Second, belonging to an elite family does not seem to be a key determinant of the allocation of political posts. This is direct evidence that colonial elite families were not able, or perhaps did not desire, to monopolize politics. Third, we also find some evidence that larger and more heterogeneous families increase the probability of a mayoral appointment, while wealthier families reduce it. Finally, there is some evidence that being the recipient of non-labor income increases the probability that an individual will become a mayor.

Though these results are preliminary and one should be very cautious in interpreting the correlations we find as representing causal effects, we believe that these findings can help explain why Antioquia was much more dynamic economically than the rest of Colombia in the 19th century. They suggest that political power was held by relatively rich people who had probably experienced upward social mobility and who probably had a much greater interest in good economic institutions, such as secure property rights, than those who held political office in many of the municipalities of Cundinamarca. These results may be in line with the fact that historians have argued Antioqueño elite families were not very interested in politics directly, and rather were happy to keep politicians under control (Safford, 1967). Our results suggest that this may have been so because the interests of those who became politicians were aligned with the traditional Antioqueño elites. If this is correct, then an important reason why Antioquia did better economically than the rest of the country in the 19th century was that it had much less dysfunctional politics.

Our data on individual incomes is of independent interest because it allows us both to construct estimates of personal distribution of income and of income per-capita, which is unheard of for the 1850s in a Latin American country. The high levels of inequality in contemporary Latin America have been argued to play an important causal role in explaining the region's low growth. However, while Engerman and Sokoloff (1997) argue that this inequality emerged during the colonial period, others, such as Coatsworth (1998, 2008) claim that there is little evidence for this. Coatsworth sees the high levels of inequality in Latin America emerging in the second half of the 19th Century as a consequence of the integration of Latin American countries into the world economy.

We find that income inequality as measured by the Gini coefficient varied a lot across the municipalities in Antioquia, from 0.18 in Itagüí to 0.65 in Medellín. However, it was also surprisingly high on average particularly given that Antioquia was relatively neglected during the colonial period, with the exception of gold mining activity in Santa Fé de Antioquia, Zaragoza and in the Cauca valley. Indeed, the Gini for the entire state was 0.51 and this for a period well before the types of forces which Coatsworth has emphasized as creating Latin American inequality came into play. The data then appear to be more consistent with the claims of Engerman and Sokoloff that Colombia entered the 19th century with a very inegalitarian income distribution.

In terms of income per-capita our estimates imply that Antioqua's per capita income was around 30% of U.S. per capita income in 1853. This is a very plausible number since we know that by 1900 Colombian income per-capita was about 18% of GDP per-capita. This divergence between 1853 and 1900 is consistent with what we know from other sources about comparative development.

The paper proceeds as follows. Section 2 gives a brief account of the central aspects of the history of Antioquia during the 19th century and compares our findings to those in the historical and economic history literatures. Section 3 describes the data collected and used, section 4 presents the income distribution estimations and discusses the relevance of the findings, section 5 develops and explains the econometric model, section 6 presents the main results and section 7 concludes.

2. The Historical Setting

In the context of Colombia, the history of the department of Antioquia has been regarded as one of relative economic and social success. Especially during the 19th century, the region which comprises what is today Antioquia evolved from being a scarcely populated and backward area to become the most economically dynamic region, with the highest population growth rates in the country, with an increasing concentration of manufacturing activities in its capital city, Medellín (Brew, 1977, Palacios and Safford, 2002, p. 316), and with the highest rates of educational attainment (Helg, 1987, Ortiz, 1991). While the annual population growth rate of the Country as a whole was 1.89% in 1851, it reached 3.18% in Antioquia (Arrubla and Urrutia, 1970). Historians have suggested that in the late 19th century Antioquia had become the richest region in the country. As we will show below, this was probably not the case for the mid 19th century.

The most direct consequence of the economic and geographic expansion of Antioquia was the development of the coffee economy, which would be central for the economic development of Colombia during the 20th century. The literature exploring 19th century Antioquia and its relative success is ample. Although there has been debate about the central elements which may explain the region's economic history during this time period, there are some common elements which have been repeatedly highlighted.

The gold mining boom, which started after independence and allowed Antioquia not to fall into recession as the rest of the Country did, is no doubt very important, since it provided capital to finance other activities, and created the wealthiest families in Colombia at the time. As compared to other gold mining regions in Colombia (e.g., the Pacific region), Antioquia had at the beginning of the century a considerably lower slave population. Together with the institutions which developed around it, gold mining activities became more widespread among the population and may have allowed faster social mobility. Authors such as Poveda (1991) argue that mining was backed up by a very favorable legislation establishing institutions that facilitated the democratization of the activity and its security, like the *sociedad ordinaria de minas* (ordinary mining society). Under this contract one partner supplied capital and the other knowledge and work, and shared 50-50, creating "mining entrepreneurs" and fostering social mobility.

From a different perspective, authors such as Parsons (1961) and Hagen (1963) argued that the economic success of Antioquia was driven by the region's difficult geography. According to them, the region's isolation and harsh geographical conditions forced the people in Antioquia to work harder. The southern colonization would also be a product of the pressures imposed by the geographic adversity. Although these authors give no compelling evidence to support these claims, it is clear that geographic aspects have been pointed out as important in Antioquia's economic history. Actually, when compared to other departments of Colombia, the Antioqueño geography doesn't differ very much from the rest of the Andean region. The only significant difference during the mid 19^{th} century was that a larger portion of Antioquia was still frontier land.

A second strand intensively explored by the literature has tried to connect cultural traits of the Antioqueño society to its economic prosperity, but no systematic evidence has been offered either. For instance, supposed Jewish and Basque influences in Antioquia have been argued to have helped create an entrepreneurial society with reduced social tensions and a less hierarchical social structure (Hagen, 1963). At the same time, extended families and the characterization of family values in Antioquia are supposed to have given rise to strong social networks, solidarity and trust, translating into a more dynamic economy.

According to Parsons (1961), the rural society in Antioquia was in sharp contrast to other rural Latin American societies because of small landowners and homesteaders related with the large families and early marriage practices. At the same time, the particularities of the gold mining activity and its trade required high levels of social capital, which was backed up by the Antioqueño family values.

Surprisingly, little has been said in the literature about the relation between the working of politics in Antioquia and the region's economic development during the period. This is quite surprising given he thrust of most of the recent research on the comparative development of Latin America. This relationship, which we begin to explore in this paper, seems a very fruitful one to investigate because it is clear that Antioqueño politics was in several aspects different from the rest of the country. Antioquia became the most conservative of the States, and the Catholic church acquired a great deal of influence in the social and political realms. This contrasted with a predominantly liberal control of national politics starting in the early fifties, and implied a relative isolation of Antioquia from the rest of the nation.

The impact of the series of civil wars was much lower in Antioquia than in the rest of the country, according to authors such as Melo (1991), because the intensity of social tensions, although present, was less acute than in other regions due to the high social mobility, the absence of a consolidated aristocracy, and the opportunities offered by colonization. In the words of Payne (1968), Antioquia was "the deviant case" within the Colombian context. And although politics is believed to have been a matter of elites as in the rest of the country, even among the upper levels of the social structure historians have argued it did not raise much of an interest, at least as compared to other regions.

Authors such as Christie (1978, 1986), on the other hand, have quite compellingly shown that politics in the southern fronts of colonization was strongly controlled by small elites. Overall, economic historians argue that there was an implicit consensus among the elites and middle classes in Antioquia regarding the role of the public sector, which made politics not as harmful for economic development there. Contrary to what may have occurred in other regions, where political power rotated between liberals and conservatives constantly, and where it was intensively used to benefit very narrow interests at the expense of political losers, Antioquia's establishment seems to have favored political stability and a strong defense of property rights to protect the interest of miners and the nascent manufacturing sectors.[8]

In this view politics was much less dysfunctional in Antioquia than in the rest of the country and those who attained political office held similar interests to the society more generally. Traders, entrepreneurs, miners, and agriculturalists all tried to maintain control of politicians and were successful. These groups' main interests were economic development, avoiding taxing property, building roads, colonization and education. Above all, the Antioqueño elite was interested in preserving order and private property. In words of Safford (1965b), industry and capital were able to dominate the political frenzy. Our data suggests that they did this not by holding office themselves but rather by allocating political offices to economically successful individuals who had a vested interest in the same set of economic institutions. Our results clearly show that economic and political power went hand by hand also in Antioquia. The Descriptive Statistics in Table 1 clearly exemplify this point; while the average annual income of mayors was of 407 pesos, it was only 133 pesos for non-mayors, and in districts like Medellin the differences were even more striking. Average income for mayors in Medellin was 1,275 pesos, while for non mayors it was 228.

3. Data and Historical Sources

3.1 The *Catastros de Ingreso* of 1853

The historical database used for this paper has been constructed directly from the historical archives at the *Archivo Histórico de Antioquia*, located in Medellín. The archive keeps a large body of 19th century official documentation for all the department of Antioquia. In particular, we focused on four main sources of information, none of which have been analyzed before. First, the *Catastros* of 1853, which are detailed, handwritten lists of individuals containing information on annual income decomposed into its sources, be it labor, land or capital, for every individual earning an annual income of 100 pesos[9] or more. The collection of the *Catastro* was ordered by the Provincial legislature of the State of Antioquia through decree 57 of 1852. Its purpose was to collect a comprehensive list based

[8] Interestingly, Roldán's (2003) study of Antioquia during the period of *La Violencia* in the late 1940s and early 1950s shows that Antioqueño elites tried to avoid getting sucked into the political instability which was being created by elites of the Conservative party elsewhere in the country because they saw it as threatening the prosperity of the area.

[9] The smallest monetary unit of the time was the real, and 8 reales made up 1 peso.

on which an extraordinary (progressive) income tax could be raised to cover the budget deficit the State had run during the fiscal year 1851.

The decree stipulated that the *Catastro* should be done at the district[10] level by the local authorities (the district council and the Mayor's office), through the appointment of a *Catastro* commission. According to the documentation available, establishing the income of an individual was done through an "iterative" process. In the first place, the commission assigned an approximate income for every individual, and then published the list. Every individual in the list had then the right to ask for a revision of their assigned income (of course, individuals had the incentive to try to get their income revised downwards, given that the tax was progressive). The commission would then evaluate the request and make a reassessment of the income level. This process seems to have been done quite accurately given the possibilities of the time, because of the specific appointment of the commissions and because of the relatively small size of the communities which constituted the districts.

According to several sources of the time, 100 pesos were just around what could be considered a survival level of income for a household. As an illustration, below we reproduce a query made by the local council of the District of Zaragoza, asking the Provincial *Catastro* commission to exclude from the extraordinary tax all individuals earning less than 1,452 reales (181.5 pesos), since, according to the council, this income level was barely enough to cover their survival expenses:

> *The cabildo has considered unfair to include in this Catastro the "jornaleros", which are those in the list above for the following two reasons: 1st: because although most of the "jornaleros" earn annually in this district 6 reales a day, none claims to work continuously two thirds of the year, because of the insalubrities of the climate, and 2nd: because assuming that a "jornalero" works 242 days a year, which at a rate of 6 reales a day gives a total of 1452 reales, out of this product we would have to forcefully deduct 484 reales for their food during that same time, and 246 for the food during the 123 days in which he doesn't work, and from these deductions it will result that the income of the profession of a "jornalero" would only be of 722 reales, which being less than 800, must not be included in this Catastro* (Archivo Historico de Antioquia) [Our translation].

The decree ordering the *Catastro* also established that the income records should be disaggregated into several categories, depending on the source of income; labor income (from working activities), land income (from land rents), and capital income (from loans, or

[10] The name of the political-administrative unit of the time was the *Distrito Parroquial*. Today they are called municipalities.

censos[11]). As a result, the *Catastros* present a detailed picture of both the personal and functional distributions of income. Both men and women were to be included, as long as they had an income above 100 pesos. In practice, less than 1% of the individuals in the *Catastros* are women, and the few women who appear in the lists are usually wealthy. This is, of course, a result of the labor force composition during this period, where women were largely excluded from the formal labor market[12]. Although the *Catastro* was to be done in every district of the Province of Antioquia, presently we have been able to find and type only those for 28 districts[13] (at the time Antioquia had around 60 districts). The *Catastros* comprise individual data for 7,691 individuals.

3.2 The *Listas de Contribuyentes de Caminos* of 1856

Unfortunately the data in the *Catastros* only comprises individuals earning at least 100 pesos, and hence excludes those with an annual income below this level. The Provincial legislature considered these individuals did not have the capacity to contribute to the extraordinary tax, and as such, there was no need to record them in the lists. Given that one of our purposes is to estimate income distributions during 19th century Antioquia, not having data for the lower tail of the distribution appears as a serious issue. Fortunately, the Archive of Antioquia also has a series of handwritten documents called *Listas de Contribuyentes de Caminos* (Lists of Roads Tax contributors), which contain comprehensive lists, also at the district level, of all working age males (ages 15-55 approximately), and classify them according to their "income class". Although these lists do not contain income data, the classification by income classes provides enough information to fill in the gap left by the *Catastros*, in terms of the income distributions. These *Listas de Caminos* exist for several different years during the second half of the 19th century (1856-1857, 1865, 1876, 1884).

[11] Censos were loans which established the payment of interest in perpetuity while the initial capital did not have to be liquidated.

[12] In the 1870 National Census, for example, 92% of working age males were in the labor force, while less than 25% of working age women were. (1870 National Census, Biblioteca Luis Angel Arango, Bogota.)

[13] These are Amagá, Amalfi, Anapolis, Barbosa, Belén, Caldas, Cancán, Concepción, Concordia, Copacabana, Envigado, Fredonia, Girardota, Hatoviejo, Heliconia, Itagüí, La Estrella, Medellín, Nechí, Caramanta, Remedios, San Bartolomé, San Cristobal, Santo Domingo, Soledad, Titiribí, Yolombó, and Zaragoza.

Presently we have coded those for the year 1856, and for the districts for which we have *Catastros*.

These lists were collected after the Provincial legislature passed decree 31 in November, 1855, which created a tax for roads building. The tax should be paid by every working age male, and originally the idea was that it be paid in labor. The tax was also progressive, so that individuals classified in the lowest classes should contribute fewer working hours, but at the same time, individuals in higher income classes were allowed to pay an equivalent value in cash. This tax should be paid annually, and it remained in force during all the second half of the 19th century, when the agrarian frontier, especially towards the south, was being opened by waves of colonization. Roads were central to the process of economic expansion that Antioquia was experiencing. The classification of individuals by income classes was done for the purpose of assigning the appropriate tax rate to an individual, and according to the available documentation, the classes for the 1856 *Listas de Caminos* were largely determined using the information of the *Catastros*, elaborated three years before.

Given that the *Listas de Contribuyentes* were intended to consist of all working age males, we can use them to complete the income distribution pictures of each district in our sample. To do this we matched the individuals in the *Catastros* with those in the *Listas de Contribuyentes*. All of the individuals appearing in the *Listas de Contribuyentes* in the lowest income class but not in the *Catastro* can be confidently assumed to have received an income below 100 pesos[14]. Obviously the coverage of the *Listas de Caminos* is not perfect, and there appeared cases of individuals in higher income classes not appearing in the *Catastros*. We cannot be certain about why this could have happened, but overall the individuals in the *Catastros* are a subset of the individuals in the *Listas de Caminos*. Although the legislation established a total of 10 different income classes, each district classified its population in a different amount of classes according to the range of income levels in the district. This is due to the fact that income levels varied considerably across districts in our sample.

Currently we have the *Listas de Caminos* for 15 districts[15], comprising 7,901 individuals. Based on both the *Catastro*'s data and the *Listas de Caminos* data we were able to

[14] The comparison of both sources revealed that individuals appearing in the *Catastros* with an income of 100 pesos were usually classified in the second income class.
[15] Amagá, Amalfi, Belén, Caldas, Envigado, Fredonia, Girardota, Hatoviejo, Itagüí, La Estrella, Medellín, Caramanta, Santo Domingo, Titiribí and Zaragoza.

estimate the income distributions and income Gini coefficients of the districts of Antioquia for which we have data (See section 4). The fact that together *Catastros* and *Listas de Caminos* give a comprehensive picture of the whole income distribution in every district implies that the income Gini coefficients calculated from this data are a very good measure of overall income inequality.

A third source consists of the 1843 and 1851 National Censuses, which provide detailed information on population and on the labor force structure. Together with the income distribution data from the *Catastros* and *Listas de Caminos*, the population data also allowed us to construct approximate estimates for income per capita in Antioquia.

3.3 Mayoral appointments and Elite Data

Next we used data from the appointment decrees of the Governors of Antioquia, also found at the *Archivo Histórico de Antioquia*, to collect the names of the mayors appointed during a ten-year period around 1853 (the year in which the *Catastro* was done). Until 1851 mayors of every district were directly appointed by the Governor of the State of Antioquia for one-year terms. Up to this year the State was divided into three provinces (Antioquia, Medellín and Oriente). Between 1852 and 1856, a brief nationwide political reform, which also split the State of Antioquia into the three provinces, mandated the election of mayors, but after 1856, the State was rebuilt and gubernatorial appointment resumed. As a result, we were able to code the names of mayors appointed during the periods 1847-1851 and 1857-1859[16]. Currently we have data on 505 mayoral appointments for 32 districts[17].

Finally, to determine which families had elite status we used information in Ospina (1939) and Restrepo (1970). The former is a three volumes catalogue of short biographies of distinguished personalities in Colombia from colonial times up to the early 20th century, from which we made a list of all the last names of personalities from Antioquia living during the 19th century. The latter is also a catalogue of short biographies, this time of all the

16 Unfortunately, to date we have not been able to locate the names of the elected mayors covering the period 1852-1856 (since there are no appointments during this period), which would allow us to investigate other issues concerning changes in political institutions.

17 Amagá, Aná, Anápolis, Angostura, Anorí, Barbosa, Belén, Belmira, Cáceres, Caldas, Campamento, Carolina, Concordia, Copacabana, Don Matías, Envigado, Fredonia, Girardota, Guarne, Hatoviejo, Heliconia, Itagüí, La Estrella, Medellín, Caramanta, San Cristóbal, San Pedro, Santo Domingo, Soledad, Titiribí, and Yarumal. We have not yet coded mayor appointments data for Amalfi and Zaragoza, so these two districts are not part of the regression results.

governors of the State of Antioquia during the 19th century, so that by coding the last names of the governors we directly had information of the political elites in Antioquia.[18].

The lists of mayors along with the data on income, family structure and elite background allow us to explore the central question of this paper, what factors determined who held political power in 19th century Antioquia and how was this influenced by income, the strength of family networks, or elites background.

Merging together all of these historical data sources demanded a homogenization of the spelling rules of names and last names, since at the time spelling of proper names was relatively arbitrary, and oftentimes quite different from current spelling rules and customs. The basic rule of thumb was to use actual spelling, and in some specific cases to use the most common spelling at the time, of same names with differing spelling. For example, all the individuals with last names Meza and Mesa were coded as having the last name Meza, as is the rule today. This allowed us to match individuals from the three data sources in the most systematic way.

Table 1 presents descriptive statistics which show a few very interesting patterns. We split the sample according to the political status of individuals. For example, the left panel presents data for mayors while the right panel provides data on non-mayors. For instance in the district of Amagá we have observations on 973 individuals out of whom 20 were mayors and 953 non-mayors. Of the mayors 10 came from elite families, while 271 of the non-mayors came from elite families. Hence in Amagá mayors were more likely to be of elite background than non-mayors. Also of interest is that about 50% of mayors in Amagá had non-labor income, while this proportion was only 10% for the population of non-mayors. Note also that the average income of mayors was almost 3 times that of non-mayors, so mayors had substantially greater incomes. Nonetheless other characteristics show no difference between mayors and non-mayors. For example, both groups came from families of similar sizes. Looking at the last line we see that some of the important differences in Amagá are reflected on average over the whole of Antioquia. The incomes of mayors were in general about 3 times that of non-mayors, and mayors also tended to be more likely to

[18] Here we must note that necessarily this variable will be miss-measured since it will assign elite status to some families which randomly have an elite last name but are not really members of the political elites of Antioquia. As a result, it is likely that the coefficient estimates for this variable are somewhat biased downward.

have non-labor income and have elite surnames. These patterns are suggestive but of course whether or not they are robust remains to be seen.

4. Estimated Income Distributions

Before moving to the factors that determined political office holding we first look at some of the basic properties of the data in more detail. In particular we calculate estimates of income inequality and of the income distribution in Antioquia for the mid 19th century. The construction of the income distributions was done through a Maximum Likelihood Estimation procedure of fitting a log-normal curve[19] to the histogram of the sample data. Given a correctly specified likelihood function for the sample data, standard results on maximum likelihood estimation guarantee the parameter estimates will be consistent, under the correctly specified distribution function (See for example, Wooldridge, 2002).

4.1 Maximum Likelihood estimation of Income Distributions

A brief description of the steps performed may be valuable to offer a better overview of the data. Once the database coming from the *Catastros* was merged with the database from the *Listas de Caminos*, we obtained observations of individuals in three possible categories; either an individual appeared in both databases, in which case we have his income and his income class, or the individual appeared only in the *Catastro* in which case we have his income but not his income class, or the individual appeared only in the *Lista de Caminos* in which case we have his income class but not his income. The first step was then to use matched observations in both databases to assign a range of income levels to every income class. As a result, we created a set of income bins in which every individual in either database would be assigned to. The bounds of the bins were taken directly out of the resulting matches. This procedure obviously required some minor adjustments such as the reassignment of class for observations that were evidently anomalous. For example, an observation with income class 1 but with an income of 700 pesos was clearly inconsistent, so this individual would be reassigned to a higher income bin in which the income level 700 were close to average. Although the *Listas de Caminos* were supposed to be comprehensive in

[19] The log-normal curve is commonly accepted as the general shape that income distributions have. All of the procedure below assumes that the income distribution in each district and in Antioquia as a whole took a log-normal form.

terms of working age male population coverage, the crossing of these and the *Catastros* revealed that this was not necessarily the case. In some of the districts a significant number of individuals appear in the *Catastro* but not in the *Lista de Caminos*. We found that the set of individuals in the *Catastros* is not a strict subset of the individuals in the Roads lists, but overall a large proportion of the individuals in the *Catastros* are included in the Roads lists. For the observations of individuals in the *Catastros* but not in the Roads lists, the assignment of an imputed income class was done in a straightforward manner: Each individual was assigned to the income bin in which its income level fell.

Recall that the *Catastro* included only individuals with an annual income of 100 pesos or above. As a result, for individuals falling in income class 1 and not appearing in the *Catastros* we assume their income is strictly less than 100 pesos. This assumption is highly plausible, and completely in line with the way in which the *Catastro* was raised and with the nature of the data. This will be important for the appropriate construction of the likelihood function to estimate the income distribution, since our income variable is censored from below at 100, and this information is important in fitting the log-normal curve to the data. For individuals falling in all other income classes and not appearing in the *Catastro* we assigned them the average income of the respective income class bin[20].

Let y^* be true income and y observed income, with y^* distributed log-normally. Let $F_{y^*}(z;\mu,\sigma)$ be the log-normal cumulative distribution function, and

$$f_{y^*}(z;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left[\frac{\ln z - \mu}{\sigma}\right]^2 \right\} \tag{1}$$

be the log-normal density function, where μ and σ are the parameters characterizing the distribution[21]. Given the nature of our income data, censored from below at 100, we have that

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* \geq 100 \\ NA & \text{if } y_i^* < 100 \end{cases} \tag{2}$$

[20] We must note that this procedure is adopted for computational simplicity. A more appropriate procedure would be to specify the likelihood function taking into account that the income of an individual for which we only know the bin in which it falls is a random variable distributed log-normally in the respective bin range. This would be straightforward to implement but computationally much more demanding while of little practical importance, so we make the simplifying assumption of assigning to such an observation the mean value of the bin.

[21] This is, the log of true income is normally distributed with mean μ and standard deviation σ.

Now, when $y_i = NA$, we have:

$$\Pr(y_i = NA) = \Pr(y_i^* < 100) = F_{y^*}(100) \tag{3}$$

and when $y \geq 100$, we have that $y_i = y_i^*$, so that

$$F_y(k) = \Pr(y_i < k) = \Pr(y_i^* < k) = F_{y^*}(k) \tag{4}$$

where $F_y(k)$ is the cumulative distribution function of the observed data. As a result,

$$f_y(k \mid y \geq 100) = F_{y^*}'(k) = f_{y^*}(k) \tag{5}$$

Hence, the likelihood function for an observation y_i takes the form:

$$f(y_i) = \left[F_{y^*}(100)\right]^{1-I\{y_i \geq 100\}} \left[f_{y^*}(y_i)\right]^{I\{y_i \geq 100\}} \tag{6}$$

where $I\{\cdot\}$ is the indicator function for the event inside the brackets.

The estimation of the income distribution is done through Maximum Likelihood for the parameters μ and σ from the log of the likelihood function above (equation (6)). This likelihood function satisfies the standard conditions for consistency of MLE estimators. We estimate the log-normal income distribution for each of the districts independently and for all the districts of Antioquia for which we have data together[22].

With the estimated log-normal income distribution $f_y(k)$ we can estimate conditional moments of interest for y. For example the conditional mean for unobserved incomes is:

$$\hat{E}[y^* \mid y^* < 100] = \frac{\int_0^{100} y f_y(y)\,dy}{\int_0^{100} f_y(y)\,dy} \tag{7}$$

Table 2 presents the estimated parameters μ and σ for every district and for the sample as a whole, as well as the mean and conditional mean (equation (7)) from the estimated log-normal distributions. The table also presents the sample mean. The standard deviation of mean income across districts is of 31.25 pesos, this is, more than a fourth of the overall mean income of 111.5 pesos. Not surprisingly we find that Medellin, the State's capital, presents the highest mean income, but at the same time, the conditional mean for individuals with incomes below 100 pesos is quite low for this district. Figures 1 through 7 present the estimated log-normal curves for some districts and the whole sample, and the same curve

[22] The MATLAB ® code for the MLE procedure is available upon request from the authors.

superimposed over the histogram of the data[23]. These districts were chosen to reflect different parts of the state. For example, Belén is in the northwest towards the Urabá region while Santo Domingo is in the northeast in the *Magdalena Medio* region. Fredonia on the other hand is in the south in the prime coffee growing region towards *Viejo Caldas*, while Itagui is now a suburb of Medellín. The histograms all show a high density for low income values and a rapid fall as income increases, consistent with the standard shape of log-normal distributions. We must note that, although asymptotically consistent, the MLE estimates obtained here (in finite sample) may overstate the true σ given that it is precisely the region with highest density the one for which we have the least precise information (we only know that income is below 100 pesos). This means that inequality may have been somewhat lower than what our estimates suggest.

4.2 Gini coefficients and Income Per capita Estimates

Although the MLE estimates of the income distributions by themselves provide reliable information about income inequality in Antioquia, more direct and comparable income inequality statistics commonly used in the literature are the Lorenz curve and the income Gini coefficient[24]. The Gini coefficient is widely used because it satisfies scale-invariance and translation-invariance, which are desirable properties for income inequality measures (Cowell, 2000). The Gini coefficient for district *i* is given by:

$$g_i = \frac{1}{N^2}\frac{1}{\bar{y}_i}\sum_{j=1}^{N}\sum_{k=1}^{N}\left|y_i - y_j\right| \tag{8}$$

where N is the number of individuals in the community, and $\bar{y}_i$ is average income. Table 2 also presents the estimates for the income Gini coefficients, both from the sample data and from the ML estimated income distributions. The income Gini estimates in table 2 are some of the first direct estimates of income inequality during the mid 19th century for Latin America. The estimates allow us to conclude that the income Gini in Antioquia was around

[23] Although the MLE estimation procedure was done in MATLAB, the figures were plotted in EasyFit ® software, which allows the depiction of clear and neat graphs of density curves.

[24] The Lorenz curve plots the proportion of total income held by each cumulative proportion of individuals in a sample. The Gini coefficient is the area between the 45 degree line and the Lorenz curve, as a fraction of the area below the 45degree line, and ranges between 0 (perfect equality) and 1 (perfect inequality).

0.55, which is quite high if we take into account that for 1853 Antioquia's economy had not yet fully integrated into the world markets.[25]

The final column in table 2 provides an estimate of income per capita, based on the estimated income distributions, together with the demographic data from the 1843 and 1851 National Censuses. While the 1843 census has data on the labor structure of the population, the 1851 census has the data chronologically closest to the *Catastros* (1853) and *Listas de Caminos* (1856) information. Let N_i be the total population of district *i* taken from the 1851 census, W_i be economically active population in district *i*[26], and C_i the total number of individuals in the *Catastros* or *Listas de Caminos*. Then the estimate of income per capita in district *i* is:

$$y_i^{pc} = \frac{1}{N_i}\left[\sum_{j=1}^{C_i} y_{i,j} + (W_i - C_i)E[y_i]\right] \tag{9}$$

where $y_{i,j}$ is income of individual j in the *Catastro* or *Lista de Caminos* of district *i*.

Estimates of income per capita for 19th century Latin America are also very scarce; these are very important for the study of comparative development across the Americas, and to understand the process of divergence occurred since then. Engerman and Sokoloff (1997) present estimates for Mexico, Brazil, Argentina and Chile around 1850, but to date, the only income per capita estimate for Colombia during the 19th century is the recent contribution by Kalmanovitz (2006). Our estimates here can allow us to asses the relative performance of the Antioqueño, and Colombian economy compared to the U.S. and to other Latin American countries, during the 19th century. Robinson and Urrutia (2007) estimate that Colombia's GDP per capita remained around 18% of U.S.'s GDP per capita throughout the 20th century. Table 2 shows that per capita income varied considerably across districts, ranging from 53.6 pesos in the mining district of Zaragoza, to only 17.3 in Girardota, for an overall income per capita of around 30.8 pesos in the state. The difference between the average income per capita and the average income from the estimated distributions is due, obviously, to the fact

---

25 Nonetheless, we must take into account that in the Antioqueño case, it is not clear if the subsequent involvement with the world markets that took place starting in the mid 19th century increased or decreased income inequality, given the relatively egalitarian expansion of the agrarian frontier for coffee cultivation.

26 Economically active population is estimated based on the 1843 census figures for labor force participation, applied to the 1851 population.

that the income distribution only accounts for the working age male population, and hence excludes children, the elderly and women.

To make these estimates comparable, we will rely on Camacho Roldán (1895). He was during the second part of the 19th century a renowned liberal politician and occupied the post of minister of finance. Among his writings, he makes a comparison of per capita GDP between Colombia and the U.S., stated in pesos, for the year 1861. Although Maddison (2001) does not provide an estimate of income per capita for Colombia in 1861, his estimates for the U.S. allow us to express Camacho Roldán's estimate and ours in dollars (see table 4). Camacho Roldán's estimate implied that Colombia's per capita income was about 34% of that of the U.S. in 1861. Assuming the exchange rate was the same in 1850[27], our income per capita estimate of 30.8 pesos for Antioquia in 1853 would be equivalent to 533 1990 dollars[28]. On the other hand, Maddison's (2001) estimate of U.S. income per capita in 1850 is 1,806 1990 dollars, which means that Antioqua's per capita income was around 30% of U.S. per capita income in 1853.

5. The Model

5.1 The Appointment Decision

The main theoretical interest in the paper lies in trying to understand what factors determined who held political power. As we noted, all the data we have for mayors relates to those who were appointed by the governor, not elected. The appointment decision can then be thought of as a latent variable, of which we observe the event of being appointed as mayor or not. To make the appointment decision there are many factors that the governor could have taken into account, which could include the power base of individuals, their wealth as well as their family characteristics. Let *a* be the latent variable for the "appointment score". We model the governor's appointment decision as follows:

$$a_i = \alpha \ln y_i + X_i \beta + u_i \tag{10}$$

and

$$m_i = \begin{cases} 1 & \text{if } a_i \geq 0 \\ 0 & \text{if } a_i < 0 \end{cases} \tag{11}$$

[27] The gold value of the peso was extremely stable during the 19th century until the mid 1880's.
[28] 533 = 30.8 * (2178/126).

where m_i is a dichotomous variable which takes the value of 1 if individual i was appointed mayor, $ln\ y_i$ is the log of income, and X_i is a vector of family network characteristics and other controls. These include district dummies, family dummies, an elite family dummy, a non-labor income dummy and the family network characteristics we will describe in section 6. Among the family network characteristics that may be relevant for the allocation of political power we include family size (and a quadratic on family size), within-family heterogeneity, and family income variables[29]. The inclusion of a non-labor income dummy, which is coded from the information on land and capital income in the *Catastros*, is intended to capture the fact that the possession of assets like land or capital may have also influenced whether or not to have appointed someone as a mayor.

Now, as we know from our previous discussion about the nature of the income data, y^* is censored at 100, and our central assumption is that it is log-normally distributed: $y_i^* \sim F_{y^*}(z;\mu,\sigma)$. Hence, $ln\ y^*$ is normally distributed with mean μ and variance σ^2. In the previous section we consistently estimated the income distributions for every district in our sample, allowing us to compute the average incomes of those individuals for which we do not observe their income levels. As a result, we can impute average incomes for these individuals, and model income as a potentially endogenous variable due to measurement error. The measurement error in this case comes from the fact that for a particular sub-sample, that of individuals with incomes below 100 pesos, the observed income measure is the estimated conditional mean income. However, given that we know which individuals have an income below 100 pesos, we have a natural instrument to deal with the errors in variables issue; a dummy variable for individuals with income below 100 can play this role.

On the other hand, it is also likely that individuals in the *Catastros* with a reported income of exactly 100 pesos may have been assigned this income level just for a rounding-up purpose. As such, incomes of 100 pesos are likely to also have measurement error, which may be correlated with u_i in equation (10). Compelling evidence that this is the case is given by a

[29] Arguably we could imagine there may be a simultaneity issue, since the evidence of Acemoglu, et al. (2008) shows that being a mayor could influence income. Throughout the paper we will maintain an exclusion restriction which is that the mayor variable does not appear in the income equation. Our justification for this is that the historical evidence on Antioquia suggests that the situation was very different from Cundinamarca. Citizens seem to have been reluctant to accept mayoral appointments because these seemed to imply a high opportunity cost given they had to leave their own businesses aside for a year. Evidence on this comes from the high degree of quits and the abundant letters in the Archive where appointed citizens asked to be excused from the mayoral position. The evidence suggests accepting a mayoral appointment was not economically beneficial.

histogram of log-incomes for the whole sample, which presents a spike at *ln(100)*. The figure suggests the likelihood that a fraction of observations with an income of 100 really correspond to slightly lower or slightly higher income levels, since the histogram presents no density at incomes slightly below 100, and a very low density for incomes slightly above 100 (see figure 8). As a result, a dummy for individuals with income of exactly 100 pesos can also be used as an instrument for $ln\, y_i^*$.

5.1 The Reduced Form and Validity of Instruments

The reduced form for log income takes the form

$$\ln y_i^* = X_i\beta + D_{1i}\delta_1 + D_{2i}\delta_2 + v_i \qquad (12)$$

where D_{1i} takes the value of 1 if individual *i*'s log income measure is the conditional mean of the estimated income distribution of its district, given that it is for this part of the sample for which log income is miss measured, and D_{2i} takes the value of 1 if an individual's income is exactly 100 pesos.

This is,

$$D_{1i} = \begin{cases} 1 & \text{if } y_i^* < 100 \\ 0 & \text{if } y_i^* \geq 100 \end{cases} \qquad (13)$$

and

$$D_{2i} = \begin{cases} 1 & \text{if } \ln y_i = 100 \\ 0 & \text{otherwise} \end{cases} \qquad (14)$$

In fact, the error will be the difference between true income and the conditional mean, or between true income and 100:

$$Error_i = \begin{cases} \hat{E}[\ln y_i^* \mid y_i^* < 100] - \ln y_i^* & \text{if } y^* < 100 \\ \ln 100 - \ln y_i^* & \text{if } y^* = 100 \\ 0 & \text{if } y^* > 100 \end{cases} \qquad (15)$$

From (15) note that $E[Error]=0$[30]. These two instruments satisfy the conditions for valid instruments in an errors-in-variables context: they are uncorrelated with the measurement

[30] That this expectation is zero follows from the assumption (observation) that for individuals with a reported income of 100, the probability of having a true income slightly above or slightly below is the same.

error and the structural error, but correlated with the true income levels (Angrist and Krueger, 2001).[31]

The conditional expectation of log income is computed as follows:

$$\hat{E}[\ln y_i^* \mid y_i^* < 100] = \frac{\int_0^{100} \ln y f_y(y)dy}{\int_0^{100} f_y(y)dy} \tag{16}$$

where $f_y(y)$ is the estimated pdf of the log-normal income distribution for every district. Table 3 presents the values of the estimated parameters for every district and the respective conditional means of log income[32].

5.3 The Empirical Methodology

Notice that our assumption that (conditional on X) income is log-normally distributed implies that $ln\, y$ is normally distributed, which in turn implies v_i in equation (12) is normally distributed. This will be a central requirement for the estimation procedure implemented here to be valid, and it follows directly from the original log-normality of income distributions. The measurement error problem in this setting lies in the fact that v_i and u_i in equation (10) are correlated, and the central assumption we will maintain throughout is that this correlation is due only to the measurement error caused by the estimation of log income for individuals with annual incomes below or equal to 100 pesos. We will assume that (u_i, v_i) is independent of X and D, a bivariate normally distributed random vector, and will normalize the variance of u_i to be 1:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \eta \\ \eta & \tau^2 \end{bmatrix}\right) \tag{17}$$

The normalization of $var(u_i)=1$ gives the parameters of the structural latent variable equation (10) an average partial effect interpretation (See Wooldridge, 2002, p. 473), since the estimation procedure can only estimate the structural parameters up to scalar. We follow Rivers and Vuong (1988) in what follows[33], who develop a two-step approach. First note that the bivariate joint normality of (u_i, v_i) implies that u_i can be expressed as

---

[31] See the Appendix for a proof.

[32] Note that, by Jensen's inequality, $E[ln\, y^* \lfloor y^*<100] < lnE[y^* \lfloor y^*<100]$.

[33] An alternative way to estimate this model is to specify the joint density of m and y, and to do MLE.

$$u_i = \frac{\eta}{\tau^2} v_i + e_i \tag{18}$$

where e_i is normally distributed and independent of X_i, D_i , and v_i, with $E[e_i]=0$ and $var(e_i)=1-\varrho^2$ [34] , where $\varrho=corr(u_i,v_i)=\eta/\tau$.

Replacing equation (18) in the structural latent variable equation (10),

$$a_i = \alpha \ln y_i + X_i\beta + \frac{\eta}{\tau^2} v_i + e_i, \qquad e_i \mid X_i, \ln y_i, v_i \sim N(0,1-\rho^2) \tag{19}$$

Since $m_i=1$ if $a_i \geq 0$, we have that[35]

$$\Pr\left(m_i = 1 \mid X_i, \ln y_i, v_i\right) = \Phi\left(\frac{\alpha \ln y_i + X_i\beta + \frac{\eta}{\tau^2} v_i}{\sqrt{1-\rho^2}}\right) \tag{20}$$

Equation (20) above shows that if we knew v_i, we could consistently estimate the scaled coefficients $a/\sqrt{(1-\varrho^2)}$, $\beta/\sqrt{(1-\varrho^2)}$ and η/τ^2 in a probit regression. This motivates the two-step procedure of Rivers and Vuong (1988), in which estimates of v_i are obtained in a first step as the residuals from an OLS regression of the reduced form equation (12). Then a probit of m_i on $\ln y_i$, X_i and the residuals v_i is implemented, yielding consistent estimates of the scaled parameters of interest. Given that we have a certain measurement error problem on the income variable, $\eta \neq 0$ and we can only estimate the parameters up to scale. In a more general framework, the *t*-statistic on the coefficient estimate for η/τ^2 is a valid test for the endogeneity of the potentially endogenous variable (See Wooldrige, 2002). Since we are assuming $\eta \neq 0$, the usual probit standard errors are not valid. As a result, bootstrapped standard errors will be calculated.

Finally, since our interest here relies on the determinants of mayoral appointments, average partial effects (APE) can be consistently estimated after this two-stage procedure. Our main interest lies on average partial effects such as

---

[34] $var(e_i)=var(u_i)-(\eta/\tau^2)^2 var(v_i)=1-(\eta/\tau^2)^2\tau^2=1-(\eta^2/\tau^2)$.
Since $\varrho=corr(u_i,v_i)=cov(u_i,v_i)/(\sqrt{(var(u_i))}\sqrt{(var(v_i))})=\eta/\tau$, we get: $var(e_i)=1-\varrho^2$.

[35] $\Pr(m_i = 1 | X_i, \ln y_i^*, v_i)=\Pr(a_i \geq 0 | X_i, \ln y_i^*, v_i)$
$=\Pr(\alpha \ln y_i^* + X_i\beta + (\eta/\tau^2)v_i + e_i \geq 0 | X_i, \ln y_i^*, v_i)$
$=\Pr(e_i \geq -\alpha \ln y_i^* - X_i\beta - (\eta/\tau^2)v_i \mid X_i, \ln y_i^*, v_i)$
$=\Pr(((e_i)/(\sqrt{(1-\varrho^2)})) \geq -((\alpha \ln y_i^* + X_i\beta + (\eta/\tau^2)v_i)/(\sqrt{(1-\varrho^2)})) \mid X_i, \ln y_i^*, v_i)$
$=\Phi(((\alpha \ln y_i^* + X_i\beta + (\eta/\tau^2)v_i)/(\sqrt{(1-\varrho^2)})))$

$$APE(x_j) = E\left[\frac{\partial \Pr(m_i = 1 \mid X_i, \ln y_i, v_i)}{\partial x_j}\right] = E\left[\frac{\partial \Phi\left(\frac{\alpha \ln y_i + X_i\beta + \frac{\eta}{\tau^2} v_i}{\sqrt{1-\rho^2}}\right)}{\partial x_j}\right] \quad (21)$$

where the expectation is taken over the distribution of v_i. Given the mean-zero normality of v_i, and that in this framework we can interchange differentiation and integration, the expression above is equivalent to:

$$APE(x_j) = \frac{\partial}{\partial x_j} E\left[\Phi\left(\frac{\alpha \ln y_i + X_i\beta + \frac{\eta}{\tau^2} v_i}{\sqrt{1-\rho^2}}\right)\right] \quad (22)$$

There are several ways to estimate consistently the expectation in equation (22). The more straightforward way to do it exploits the uniform weak law of large numbers, and the fact that $\hat{a}/\sqrt{(1-\hat{\rho}^2)}$, $\hat{\beta}/\sqrt{(1-\hat{\rho}^2)}$, $(\hat{\eta}/\hat{\tau}^2)/\sqrt{(1-\hat{\rho}^2)}$, and $\hat{v}_i$ are all consistent estimates of their respective population counterparts. Hence,

$$\frac{1}{N}\sum_{i=1}^{N} \Phi\left(\frac{\hat{\alpha}}{\sqrt{1-\hat{\rho}^2}} \ln y_i + X_i \frac{\hat{\beta}}{\sqrt{1-\hat{\rho}^2}} + \frac{\frac{\hat{\eta}}{\hat{\tau}^2}}{\sqrt{1-\hat{\rho}^2}} \hat{v}_i\right) \xrightarrow{p} E\left[\Phi\left(\frac{\alpha \ln y_i + X_i\beta + \frac{\eta}{\tau^2} v_i}{\sqrt{1-\rho^2}}\right)\right] \quad (23)$$

Consistent estimates of the Average Partial Effect with respect to a given x_j will then take the form:

$$APE(x_j) = \frac{1}{N} \frac{\hat{\beta}_j}{\sqrt{1-\hat{\rho}^2}} \sum_{i=1}^{N} \Phi\left(\frac{\hat{\alpha}}{\sqrt{1-\hat{\rho}^2}} \ln y_i + X_i \frac{\hat{\beta}}{\sqrt{1-\hat{\rho}^2}} + \frac{\frac{\hat{\eta}}{\hat{\tau}^2}}{\sqrt{1-\hat{\rho}^2}} \hat{v}_i\right) \quad (24)$$

We will be especially interested in the partial effects with respect to log income and the family network characteristics. As a robustness check, and given its computational simplicity, a linear probability model by 2SLS will also be estimated to deal with the Errors in Variables problem. Although a LPM is necessarily miss specified, it might provide fairly good estimates of the average effects which will serve for comparison (see Wooldrige, 2002, p. 472).

6. Estimation and Results

6.1 Explanatory Variables

Our model is intended to investigate what factors influenced the probability of being appointed as a mayor in 19th-century Antioquia. To this aim we exploit the cross-section of individuals in the State. Given the wide variability of income per-capita across the districts in our sample, obviously the effect of income on the appointment probability will differ with the overall income level of each district. For example, we should expect that, if richer individuals are more likely to be appointed, then the income level required to have a given probability of appointment should be higher in Medellín, the State capital, than in a small and backward district such as Santo Domingo. This suggests that it will be important to include district fixed effects among the control variables in all of the models to be estimated. On the other hand, the *Catastros* also provide us with information about the factor composition of income. Individuals who earn non-labor income are actually rather scarce in the data, and landholders or owners of capital may have been more likely to be appointed as mayors. Perhaps these agents were the real economic elite or perhaps their opportunity cost of time was lower since they were more like rentiers. As a result, we also include a dummy for whether or not an individual had non-labor income among the exogenous covariates as noted above.

Recall the four elements we hypothesize may be important determinants of whether or not an individual is appointed mayor in addition to individual income, are family size, family income, elite status, and within-family heterogeneity. Let P_i be the set of individuals in district i. Then P_i can be partitioned in J disjoint sets F_{i1}, F_{i2}, ... F_{iJ}, where F_{ij}, is the set of individuals in family j at district i. Family size for individual $k \in F_{ij}$, is measured simply as the number of individuals in district i in family j, as a proportion of all individuals in that district:

$$familysize_k = \frac{|F_{ij}|}{|P_i|} \tag{25}$$

where $|\cdot|$ denotes the cardinality of the set. What could the effect of family size on appointment decisions be, holding constant other characteristics as elite status and income? It is likely that larger families create more widespread social networks which should be taken into account by the governor at the time of appointment. Nonetheless, larger families are also probably more prone to face a collective action problem which may make it more difficult for them to influence political appointments. Hence the effect of this variable is really an empirical question which our estimates will address.

On the other hand, family income for individual $k \in F_{ij}$ is measured as the income of all individuals in family j at district i, as a proportion of total income in district i:

$$familywealth_k = \frac{\sum_{s \in F_{ij}} y_s}{\sum_{F_{ij} \subset P_i} \sum_{s \in F_{ij}} y_s} \tag{26}$$

Although own income is likely to be an important determinant of a mayoral appointment, we can also expect that the income of an individual's family influences the likelihood of an appointment. If richer families are more politically influential, we should see that for two individuals with the same income, the probability of being appointed mayor is higher for the one coming from a richer family. On the other side, richer families may face a higher opportunity cost of having their members be chosen as mayors, which may reduce their likelihood of being appointed, in case the governor foresaw who were more likely to accept a mayoral post.

To construct the elite status variable we coded a dummy variable based on Ospina (1939) and Restrepo (1970). The elite dummy takes the value of 1 if an individual has a last name in any of these lists.

Within-family heterogeneity has two dimensions. On the one hand, we try to capture how different individual k is from the rest of individuals in his family; this we call *personal heterogeneity*, and measure it as the absolute difference between k's income and the average income of his family, as a proportion of the total family income:

$$personalheterogeneity_k = \frac{\left| y_k - \frac{1}{\left|F_{ij}\right|} \sum_{s \in F_{ij}} y_s \right|}{\sum_{s \in F_{ij}} y_s} \tag{27}$$

The inclusion of this variable is motivated by the fact that how different or similar is an individual to his family or to his social networks is likely to affect how influential he is in his social environment. It is likely, for example, that a rich individual member of a relatively poor family is particularly influential within his family. On the contrary, it is also possible that individuals with income levels very different from those of their families end up being excluded and segregated, affecting the likelihood of them being appointed mayors.

We also try to capture how heterogeneous a family is directly with the variance of income in the family. The *family heterogeneity* measure for individual *k* is hence the variance of income within his family:

$$familyheterogeneity_k = \frac{1}{|F_{ij}|}\sum_{s\in F_{ij}}\left(y_k - \frac{1}{|F_{ij}|}\sum_{s\in F_{ij}} y_s \right)^2 \qquad (28)$$

6.2 Regression Results

Table 5 presents the basic regression results, which pools together the observations for all districts[36]. The first four columns are simple linear probability models, for comparative purposes, of the form

$$m_i = \alpha \ln y_i + \beta X_i + u_i \qquad (29)$$

which means the coefficient estimates can be interpreted as partial effects. Columns (1) and (2) present the OLS results, while columns (3) and (4) present the 2SLS results, where the errors-in-variables issue is addressed. All of these models have robust standard errors, which are also clustered at the family level. Equation (12) is the first stage in Panel B of columns (3) and (4), which uses (13) and (14) as instruments. The specification in column (1) excludes log income as well as district and family dummies, to take a first look at the relationship between the family network variables and the appointment probability. We can see that the polynomial terms on family size have an overall negative effect, while the personal heterogeneity variable is positively and significantly associated with the appointment probability. Within family heterogeneity and, surprisingly, the elite family dummy, are not significant. Overall, this regression suggests that the strongest correlation between mayoral appointments and family network variables occurs through personal heterogeneity. Individuals which, within their family, are considerably distinct in terms of their income, have higher probability of being mayors in their districts.

Once we include log income, the non-labor income dummy and district dummies (column (2)), results change dramatically. The family network characteristics lose their significance, while the coefficient on income appears positive and strongly significant (0.045 with standard error of 0.005). Columns (3) and (4) present 2SLS estimates including district

[36] These are Amagá, Belén, Caldas, Envigado, Fredonia, Girardota, Hatoviejo (currently named Bello), Itagüí, La Estrella, Medellín, Nueva Caramanta, Santo Domingo and Titiribí.

dummies, but without and with family dummies respectively. The first-stage R-squared is considerably high, above 60%, and both instruments D_1 and D_2 significantly reduce log income as should be expected (Table 5, Panel B). Together with the sample size, these first stages leave little room for weak instruments issues. Panel A presents the second stages of both models. The family network variables, as well as the elite and non-labor income dummies remain insignificant, while the coefficient on log income is highly significant and increases to 0.06 (standard error 0.009). The increase in the size of the coefficient, as compared to the OLS (columns (1) and (2)) specifications is probably due to the attenuation bias in the OLS models caused by the errors-in-variables problem, which 2SLS is solving. Income appears as the main and overriding determinant of mayoral appointments, above elite status and family connections considerations.

Columns (5) through (7) in Table 5 present simple maximum likelihood Probit regressions, to compare their results with columns (8) and (9) which present the correct specification derived in section 5, which deals with the errors in variables issue. The specification in column (5) only includes family network variables; Family size, family income and personal heterogeneity all appear as significant, and in particular, the latter has a positive coefficient. Columns (6) and (7) include district and district and family dummies respectively, plus log income, which modify the benchmark results. The square of family size is not significant, but family size does appear as significantly increasing the probability of a mayoral appointment at the 10% level (coeff=21.9, std. err=12.5). Interestingly, family income reduces it (coeff=-11.2, std. err=5.1). Moreover, the sign of the coefficient on the personal heterogeneity measure flips around becoming negative and significant. Once the unobserved fixed characteristics of districts and families are controlled for, individuals who are less similar to their families are less likely to be appointed. The significance of log income, on the other hand, remains very robust (coeff=0.8, std. err=0.07).

Finally columns (8) and (9) present the results of the model developed in section 5.3, where the errors-in-variables issue on the log income variable is addresses. Panel B presents the first step of the procedure (equation (12)), from which the predicted residuals v_i are obtained. In this first step the instruments D_1 and D_2 are highly significant and have the expected negative signs. The first observation worth noting in Panel A, which presents the second step, is that the coefficient on the estimated residual is highly significant, which is at the same time a test for endogeneity as noted previously. Given our certain errors-in-variables issue, this should indeed be the expected result, where we cannot reject that the residuals from the miss measured log income are correlated with the structural error in equation (10). Log income is again highly significant (coeff=1.57, std. err=0.14), but interestingly, the coefficient on family income shows a significant negative effect on the likelihood of appointment (coeff=-10.5, std. err=3.7). While richer individuals are more likely to be appointed as mayors, individuals from richer families are less likely. The economic magnitude of these countervailing effects can be assessed with the average partial effects presented below implied by the coefficient estimates.

On the other hand, and consistent with the results from the simple Probit models, the personal heterogeneity measure in column (8) is negatively related to the mayoral appointments (coeff=-5.3, std. err=0.9). Family heterogeneity, on the other hand, is positive, but significant only when family effects are included[37]. A final puzzling result also shows up concerning the elite dummy; although this variable is not significant in any of the previous models, column (9) shows a negative and significant effect; this is, among families which had a mayor appointed at some point in any of the districts in the sample, being member of an elite family reduces the likelihood of a mayoral appointment. Once more we find some evidence that wealth and social elite status may have been operating in very different directions concerning politics in Antioquia.

An economic interpretation of the results from the Probit models requires the computation of average partial effects, as developed in equation (26) in section 5.3. Table 6 presents the APE's for the statistically significant coefficients computed from the estimates in column (9), directly applying the formula in equation (26). As a result, these partial effects can be interpreted as the average over the sample, of the change in the probability of a mayoral appointment given a marginal change in any of the continuous variables, or given a change from 0 to 1 in the dummy variables.

Take, for example, the APE for log income (0.165). This implies that, holding everything else constant, changing an individual's income from 100 pesos to 200 pesos would, on average, increase the probability of him being appointed as mayor by 11%[38], which is quite large. On the other hand, take family size (ignoring the insignificant non-linear

[37] Note that the inclusion of family fixed effects effectively drops out of the regression all families without within-family variation in the mayor appointment variable, so column (9) can only be interpreted in terms of families in with at least a mayoral appointment.
[38] (ln(200)-ln(100))*0.165=0.11

term), which has an APE of 2.08. The average (relative) family size in our sample is 0.02. This is, the average family in a given district represents 2% of the total district's population. Hence, holding other things equal, an individual from a family which is 3% of its district will be 2% more likely to be appointed as mayor than an individual from a family which is 2% of the district's population[39]. Similar calculations can be performed based on the APE's in Table 6.

Table 7, on the other hand, presents analogous results to those from Table 5, but separately for Medellín, the State capital. This not only allows a robustness check on the overall results for Antioquia, but also may indicate if particularly deviant or specific events were occurring in the main economic and political center of Antioquia. For example, although the overall results in Table 5 suggest that elites were not particularly more likely to be appointed, the individual district regressions in Table 7 show that in Medellín, elite status was a significant burden for mayoral appointments (coeff=-10.6, std.err=1.06). This was probably due to the fact that the population of Medellín was considerably larger, and as a result many more families and last names were present and actively participating in politics.

Regression results for each district independently show that[40], on the one hand, income is especially important as a determinant of mayoral appointments when comparing only families with appointed mayors; when all families are considered, the significance of log income is reduced. On the other hand, income was not a key determinant of mayoral appointments in Girardota, Hatoviejo (Bello), and La Estrella. Another set of interesting results concern the personal heterogeneity variable. In most districts it usually significantly reduces the likelihood of a mayoral appointment, except for Girardota. This one is, without doubt, the family network characteristic that influences appointment probabilities the most, and suggests that the degree of similarity of an individual to his family group was central in determining his political prospects. Overall, the most dissimilar ones were less likely to be appointed as mayors. Also quite interesting is to note that power is very low for the significance tests on family size coefficients, which are relatively large across municipalities, but with very large standard errors. Only for Medellín is family size important, and in this case it reduces the likelihood of a mayoral appointment. This last result for Medellín

[39] (0.03-0.02)*2.08=0.021

[40] The regression results for each of the other districts separately can be provided by the authors upon request.

obviously contradicts the finding for the whole of Antioquia (Table 5), as described above, but can be reconciled by the fact that, given that Medellín was particularly large, the average family size there was smaller (0.01 as can be seen in Table 1). Only particularly small and cohesive family groups may have been able to exert political influence. Family heterogeneity does not have a remarkable influence in any of the districts. Regarding the elite status dummy, it is only significant in Hatoviejo (Bello) and Medellín, where it reduces the probability of an appointment.

7. Conclusions

This paper tried to provide some evidence on the determinants of political power in 19th century Colombia using some unique data from Antioquia. The recent literature on Latin American development has emphasized how political economy factors stopped Latin American countries taking advantage of the growth opportunities presented in the 19th century by the industrial revolution. Yet there are many puzzles about the nature of this dysfunctional political equilibrium and there is obviously a lot of variation both between and within Latin American countries. Though many scholars see this as a path dependent outcome of the nature of colonial society, as yet we lack a more detailed understanding of the origins and incentives of 19th century political elites. While in some places they were obviously related to colonial elites, in other places they were not and new forces and groups emerged. How did political institutions and the distribution of power influence who held office and how they used this power? What was the legacy of colonial institutions?

In this paper we investigated directly some of the potential determinants of political office holding in the districts of Antioquia to offer preliminary answers to two sets of questions. On the one hand, the unique historical income data directly taken out of the historical archive of Antioquia, which contains information for the province during the mid 19th century, allowed us to construct income distributions and inequality estimates, central for the comparative development debates that have recently been taking place. Novel and reliable data was used to asses the extent of income inequality in a region of Colombia which was beginning to develop an economy strongly based on mining and agricultural exports and manufacturing. The overall results show that income inequality was indeed high, but not so different to current levels of inequality in Colombia, nor significantly higher than historical levels of inequality in countries such as the U.S., where income and wealth inequality were especially high in the South during the same time period. On this ground, the contribution of this paper relies on the novel data it uses, and on its straightforward methodology. Income distribution and inequality estimates for 19th century Latin America tend to be highly speculative and not based on micro-income data, in opposition to what this paper does.

On the other hand, the determinants of the allocation of political power constitute a central issue for students of political economy. The main results based on an econometric specification for the probability of being appointed as a mayor suggest that local political power was indeed concentrated among the wealthiest individuals in their communities, but not necessarily among what could be considered traditional elites. Obviously this was only possible because in Antioquia the social elites did not necessarily coincide with the richest families. Future research should try to investigate more precisely how and to what extent economic and political elites in Antioquia coincided.

At the same time, a remarkable characteristic suggests itself as important in the appointment decision of governors; how different an individual was from its relatives would make him more likely to be appointed as a mayor, while family size does not seem to have played a very important role, despite the importance it has been given in the literature highlighting the Antioqueño culture. That money matters for the political process is not in general a novel result, but it is in the context of the non-democratic appointment institutions of the State of Antioquia, and in the context of the historical accounts which have argued that politics in Colombia have been relatively open to middle sectors which find in it a career. Not only mayors in Antioquia were not career politicians, but political posts seem to have been economically costly.

Evidently all of the analysis in the paper is based on the assumption that, apart from errors in variables considerations, income can be considered exogenous for the political allocation process. This is a quite strong assumption which, although plausible in the Antioqueño setting, will hopefully be relaxed in future research. In the meanwhile, the econometric results can be seen as historical correlations which put into question much of what has been, up to date, considered common knowledge about the political and economic history of 19th century Antioquia.

References

Acemoglu, Daron, María Angélica Bautista, James A. Robinson and Pablo Querubín (2008) "Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia" in Elhanan Helpman ed. *The Political Economy of Institutions*, Cambridge: Harvard University Press.

Acemoglu, Daron, Simon Johnson and James A. Robinson (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 5, 1369-1401.

Acemoglu, Daron, Simon Johnson and James A. Robinson (2002) "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *Quarterly Journal of Economics*, 118, 1231-1294.

Acemoglu, Daron, Simon Johnson and James A. Robinson (2005) "Institutions as the Fundamental Cause of Long-run Growth", in Aghion, Phillipe, and Stephen Durlauf (eds.), *Handbook of Economic Growth*, Amsterdam: North-Holland.

Acemoglu, Daron and James A. Robinson (2007) "A Model of the Iron Law of Oligarchy," Unpublished.

Acemoglu, Daron and James A. Robinson (2008) "Persistence of Power, Elites and Institutions," *American Economic Review*, 98, 267-293.

Acemoglu, Daron, James A. Robinson and Thierry Verdier (2004) "Kleptocracy and Divide and Rule: A Theory of Personal Rule" *Journal of the European Economic Association* 2, 162-192.

Alesina, Alberto and Eliana La Ferrara (2000) "Participation in Heterogeneous Communities", *American Economic Review*, 115, 847-904.

Angrist, Joshua and Alan B. Krueger (2001) "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives*, 15, 69-85.

Arrubla, Mario and Miguel Urrutia (1970) *Compendio de Estadísticas Históricas de Colombia.* Bogotá: Dirección de Divulgación Cultural, Universidad Nacional de Colombia.

Becker, Gary S. (1983) "A Theory of Competition among Pressure Groups for Political Influence", *Quarterly Journal of Economics*, 98, 371-400.

Berndt, Ernst, Bronwyn Hall, Robert E. Hall and Jerry Hausman (1974) "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*, 3, 653-666.

Brew, Roger (1977) *El Desarrollo Económico de Antioquia desde la Independencia hasta 1920*, Bogotá: Banco de la República.

Bushnell, David (1971) "Voter Participation in the Colombian Election of 1856," *Hispanic American Historical Review*, 51, 237-249.

Camacho Roldán, Salvador (1895) *Escritos Varios*. Bogotá: Librería Colombiana.

Coatsworth, John H. (1998) "Economic and Institutional Trajectories in 19th Century Latin America", in John H. Coatsworth and Alan M. Taylor (eds.) *Latin America and The World Economy since 1900*. Cambridge: Harvard University Press.

Coatsworth, John H. (2008) "Inequality, Institutions, and Economic Growth in Latin America," *Journal of Latin American Studies*, 40, 545-569.

Cowell, Frank (2000) "Measurement of Inequality", in Atkinson Anthony and François Bourguignon (eds.), *The Handbook of Income Distribution*, Vol. 1., Amsterdam: North-Holland.

Christie, Keith H. (1978) "Antioqueño Colonization in Western Colombia, A Reappraisal", *Hispanic American Historical Review*, 58, 260-283.

Christie, Keith H. (1979) "Gamonalismo in Colombia: An Historical Overview," *North/South: Canadian Journal of Latin American Studies*, 4, 42-59.

Christie, Keith H. (1986) *Oligarcas, campesinos y política en Colombia: aspectos de la historia socio-política de la frontera antioqueña*, Bogotá: Universidad Nacional de Colombia.

Deas, Malcolm (1971) "Algunas Notas sobre la Historia del Caciquismo en Colombia," *Revista del Occidente*, 127, 118-140.

Delpar, Helen (1981) *Red against blue: the Liberal Party in Colombian politics, 1863-1899,* Baton Rouge: University of Alabama Press.

Engerman, Stanley L. and Kenneth L. Sokoloff (1997) "Factor Endowments, Institutions, and Differential Growth Paths among New World Economies", in Stephen H. Haber (ed.), *How Latin America Fell Behind*. Stanford University Press, Palo Alto.

Engerman, Stanley L. and Kenneth L. Sokoloff (2005) "The Evolution of Suffrage Institutions in the New World," *Journal of Economic History*, 65, 891-921.

García, Antonio (1937) *Caldas: Geografía Económica de Colombia*. Bogotá, Imprenta Nacional.

Hagen, Everett E. (1963) *El cambio social en Colombia: el factor humano en el desarollo económico*, Bogotá: Ediciones Tercer Mundo.

Helg, Aline (1987) *La educación en Colombia, 1918-1957: una historia social, económica y política*, Bogotá: Fondo Editorial CEREC.

Kalmanovitz, Salomón (2006) "El PIB de la Nueva Granada en 1800: Auge Colonial, Estancamiento Republicano", in *Revista de Economía Institucional*, 8, 161-183.

Kandori, Michihiro (1992) "Social Norms and Community Enforcement", *Review of Economic Studies*, 59, 63-80.

Key, V.O. Jr. (1949) *Southern Politics: In State and Nation*, New York: Vintage Books.

Londoño, Patricia (2002) *Religión, Cultura y Sociedad en Colombia. Medellín y Antioquia, 1850-1930*. Bogotá: Fondo de Cultura Económica.

Maddison, Angus (2001) *The World Economy: A Millennial Perspective*, Paris: OECD Development Center.

Melo, Jorge Orlando (1991) *Historia de Antioquia*. Bogotá: Editorial Presencia.

Möbius, Markus and Adam Szeidl (2007) "Trust and Social Collateral," Working Paper, Harvard University.

Ortiz, (1991) "Antioquia bajo el Federalismo", in Melo, Jorge O. (ed.), *Historia de Antioquia*. Presencia, Bogotá.

Ospina, Joaquin (1939) *Diccionario Biografico y Bibliografico de Colombia*, Vols. 1,2 and 3. Bogotá: Aguila.

Paige, Jeffery M. (1997) *Coffee and power: revolution and the rise of democracy in Central America*, Cambridge: Harvard University Press.

Palacios, Marco and Frank Safford (2002) *Colombia: Fragmented Land, Divided Society*, New York: Oxford University Press.

Parsons, James J. (1961) *La Colonización Antioqueña en el Occidente Colombiano*, Bogotá: Imprenta del Banco de la República.

Payne, James L. (1968) *Patterns of Conflict in Colombia*, New Haven: Yale University Press.

Poveda, Gabriel (1991) "Breve Historia de la Minería", in Melo, Jorge O. (ed.), *Historia de Antioquia*. Presencia, Bogotá.

Putnam, Robert D. (1994) *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton: Princeton University Press.

Ramírez, Ernesto (1984) *Poder Económico y Democracia Política. El Caso de la Familia Ospina*. Bogotá.

Restrepo, Jose M. (1970) *Gobernadores de Antioquia*, Vol. 2. , Bogotá: Lumen Christi.

Rivers, Douglas and Quong Vuong (1988) "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models", *Journal of Econometrics*, 39, 347-366.

Robinson, James A. and Miguel Urrutia eds. (2007) *Economía Colombiana del Siglo XX: Un Análisis Cuantitativo*, Bogotá: Fondo de Cultura Económica.

Roldán, Mary (2002) *Blood and Fire: La Violencia in Antioquia, Colombia 1946-1953*, Durham: Duke University Press.

Safford, Frank P. (1965a) *Commerce and Enterprise in Central Colombia, 1821-1870*, Unpublished PhD Dissertation, Department of History, Columbia University.

Safford, Frank P. (1965b) "Significación de los Antioqueños en el Desarrollo Económico Colombiano. Un Examen Crítico de las Tesis de Eeverett Hagen", in *Anuario Colombiano de Historia Social y de la Cultura*., 3, 49-69.

Safford, Frank P. (1972) "Social Aspects of Politics in Nineteenth-Century Spanish America: New Granada, 1825-1850," *Journal of Social History*, 5, 344-370.

Schwartz, Stuart (1996) "The Landend Elite", in Louisa Schell Hoberman and Susan Midgen Socolow (eds.), *The Countryside in Colonial Latin America*, Albuquerque: University of New Mexico Press.

Sokoloff, Kenneth L. and B. Zorina Khan (1990) "The Democratization of Invention during early Industrialization," *Journal of Economic History*, 50, 363-378.

Stein, Stanley and Barbara Stein (1970) *The Colonial Heritage of Latin America*, New York: Oxford University Press.

Stigler, George (1988) *Chicago Studies in Political Economy*. Chicago: University of Chicago Press.

Stone, Samuel Z. (1990) *The heritage of the conquistadors: ruling classes in Central America from the Conquest to the Sandinistas*, Lincoln: University of Nebraska Press.

Taylor, William B. (1972) *Landlord and Peasant in Colonial Oaxaca*, Stanford: Stanford University Press.

Twinam, Ann (1982) *Miners, Merchants and Farmers in Colonial Colombia*. Austin: University of Texas Press.

Uribe de Hincapié, María Teresa (1998) *Las raíces del poder regional: el case antioqueño*, Medellín: Editorial Universidad de Antioquia.

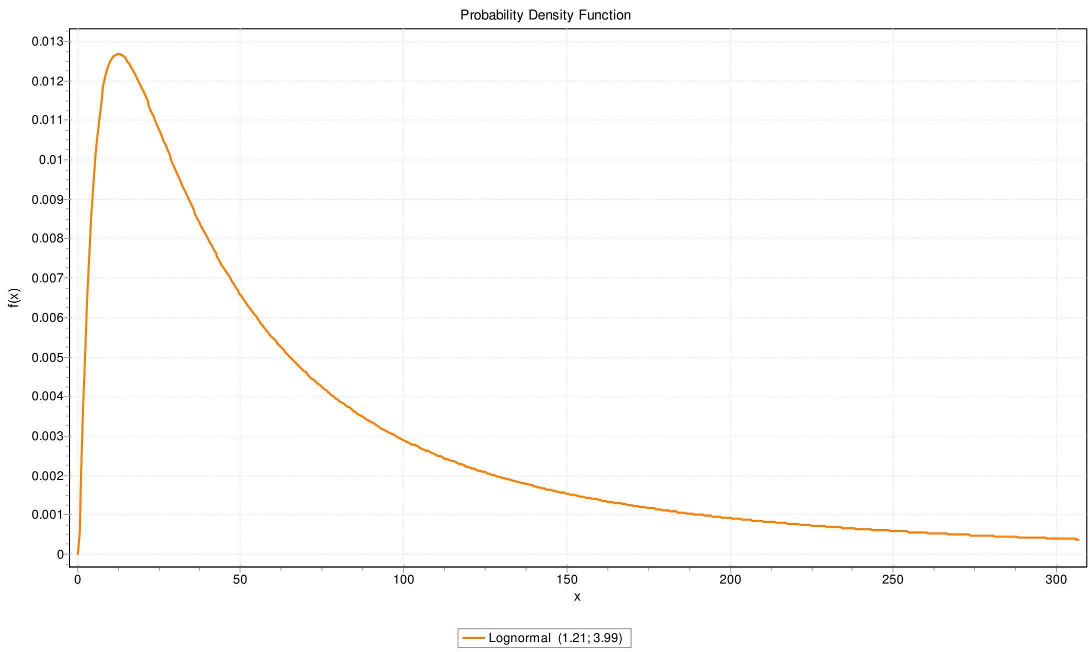
Uribe-Uran, Victor (2000) *Honorable Lives: lawyers, families and politics in Colombia, 1780-1850*, Pittsburgh: University of Pittsburgh Press.

Wooldridge, Jeffrey (2002) *Econometric Analysis of Cross Section and Panel Data*, Cambridge: MIT Press.

Van Young, Eric (1983) “Mexican Rural history since Chevalier: The Historiography of the Colonial Hacienda,” *Latin American Research Review*, 18, 5-61.

Maximum Likelihood Fitted Log-normal Income Distributions (Estimated Distribution and Superimposed on the data):

Figure 1: All Districts (μ=3.99,σ=1.21)



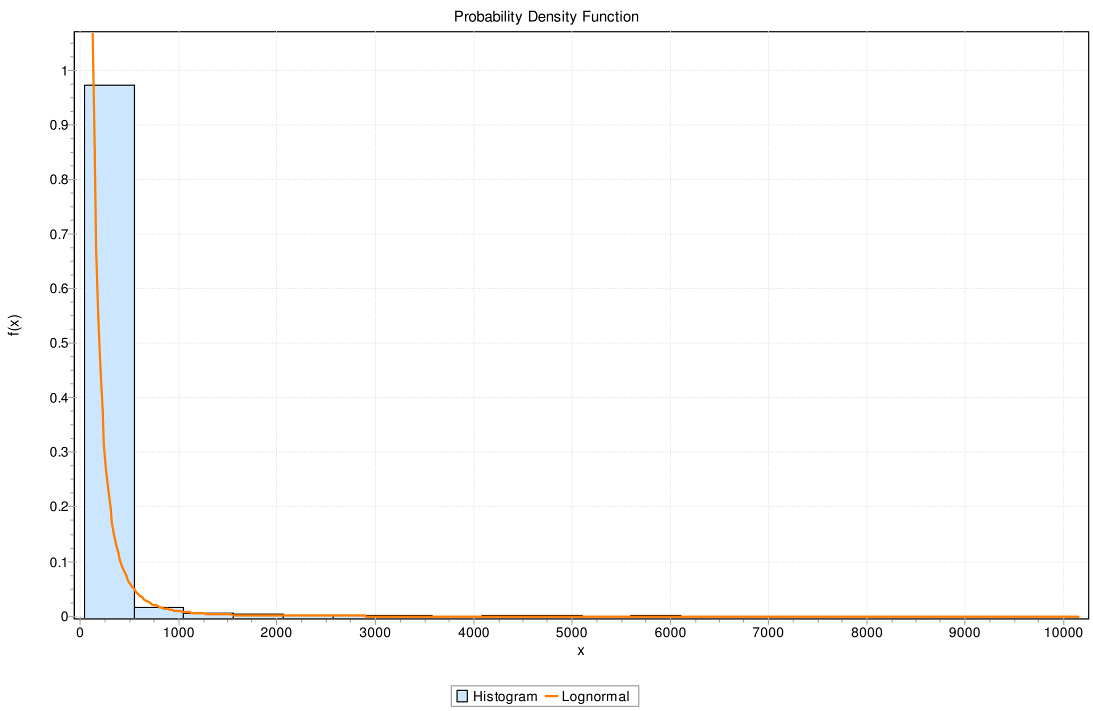
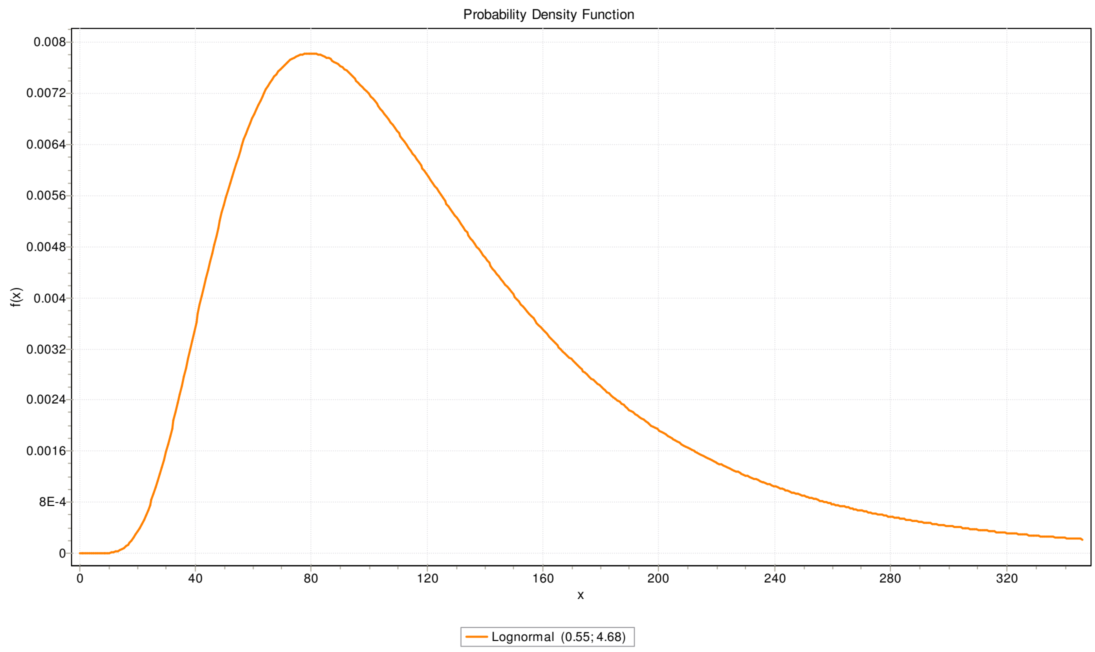


Figure 2: Belen (μ=4.68,σ=0.55)



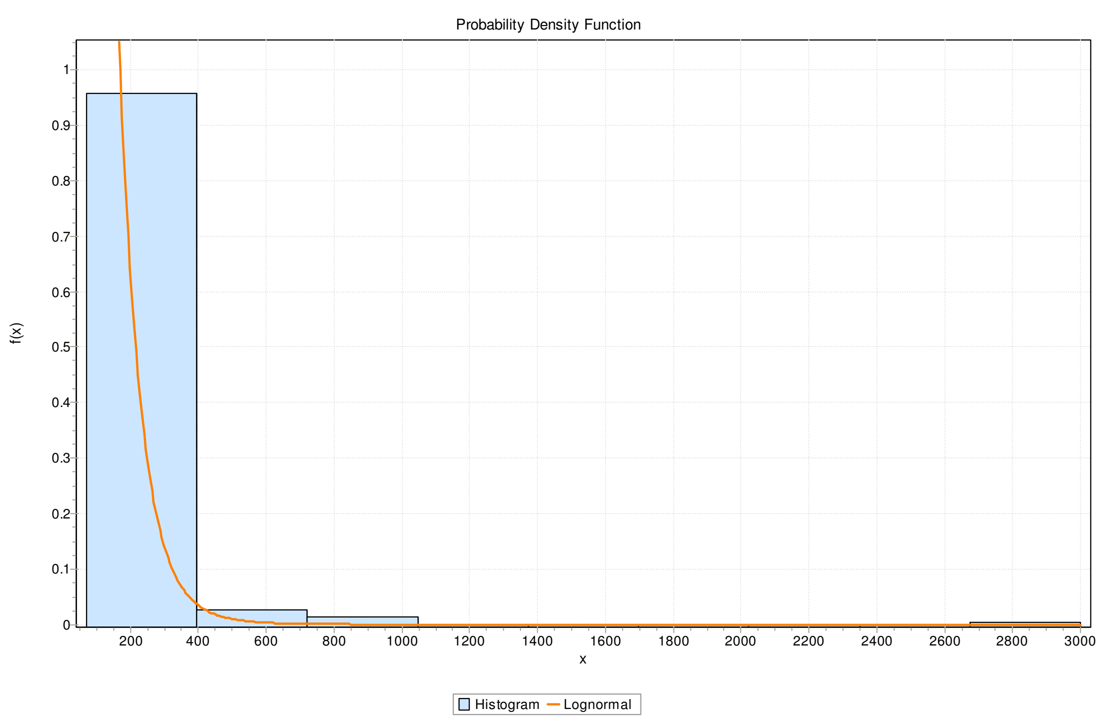
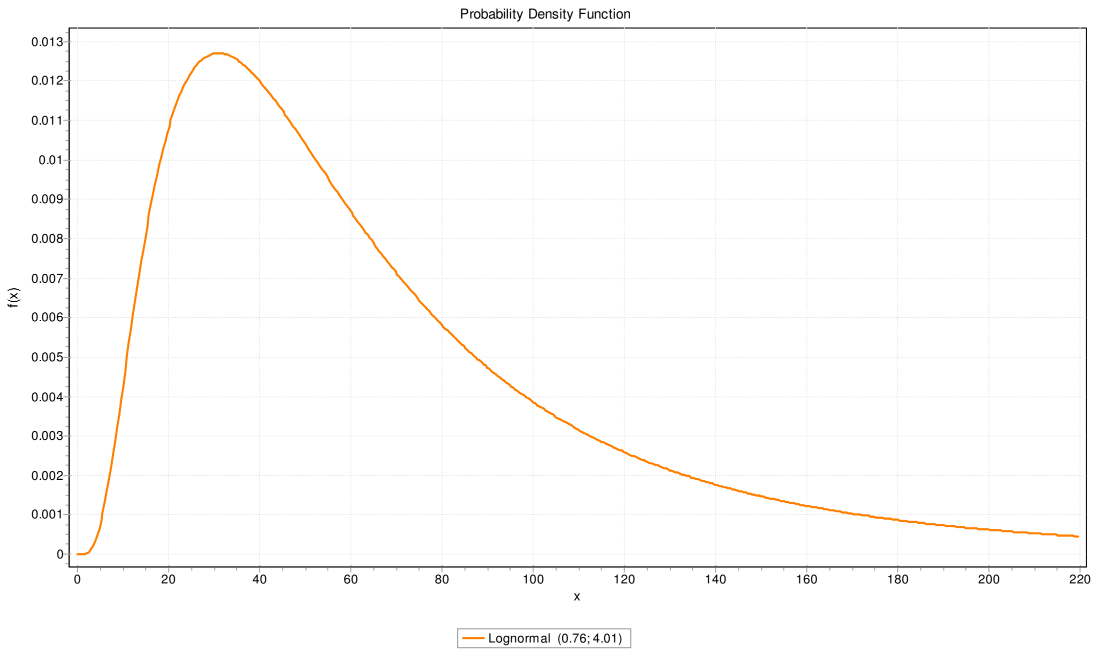


Figure 3: Fredonia (μ=4.01,σ=0.76)



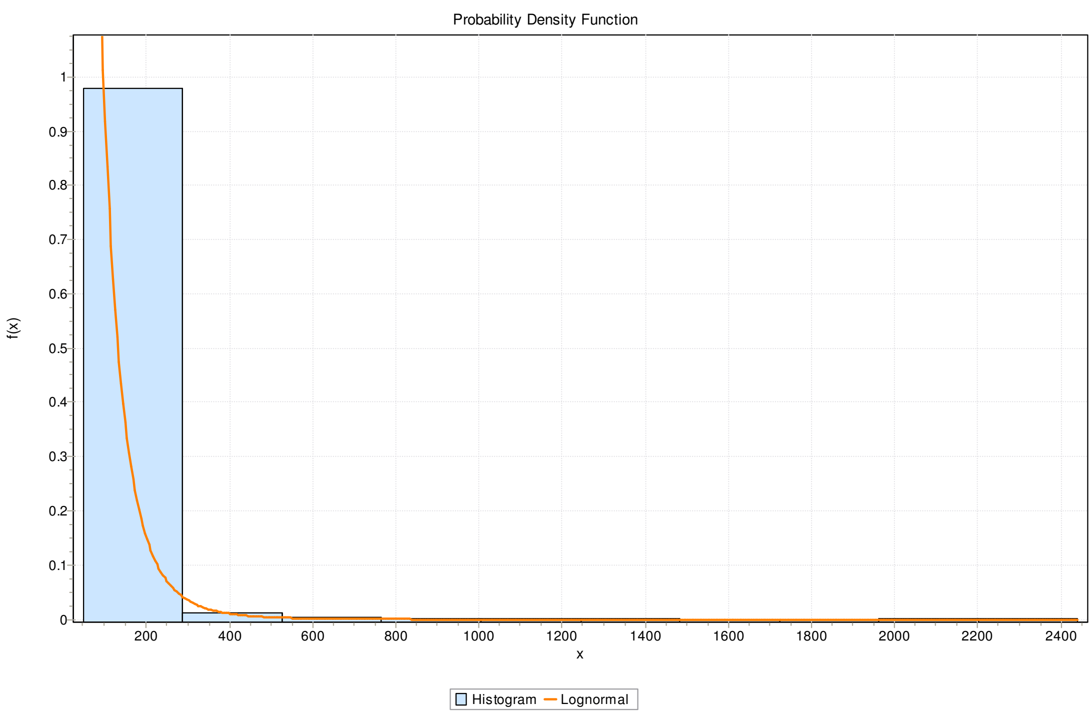
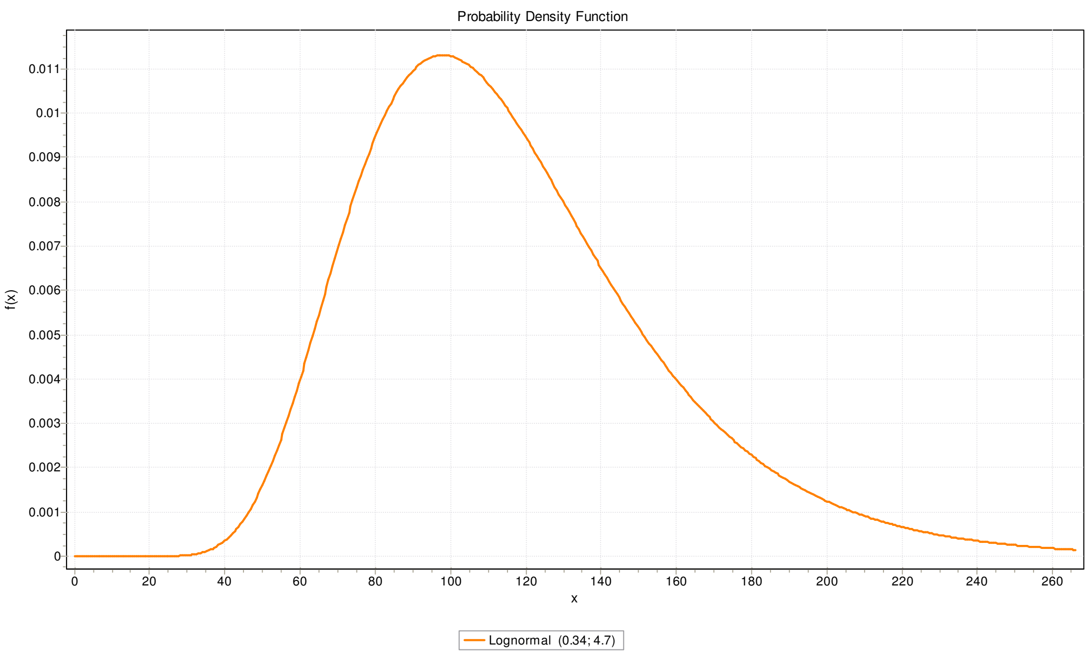


Figure 4: Itagui (μ=4.7,σ=0.34)



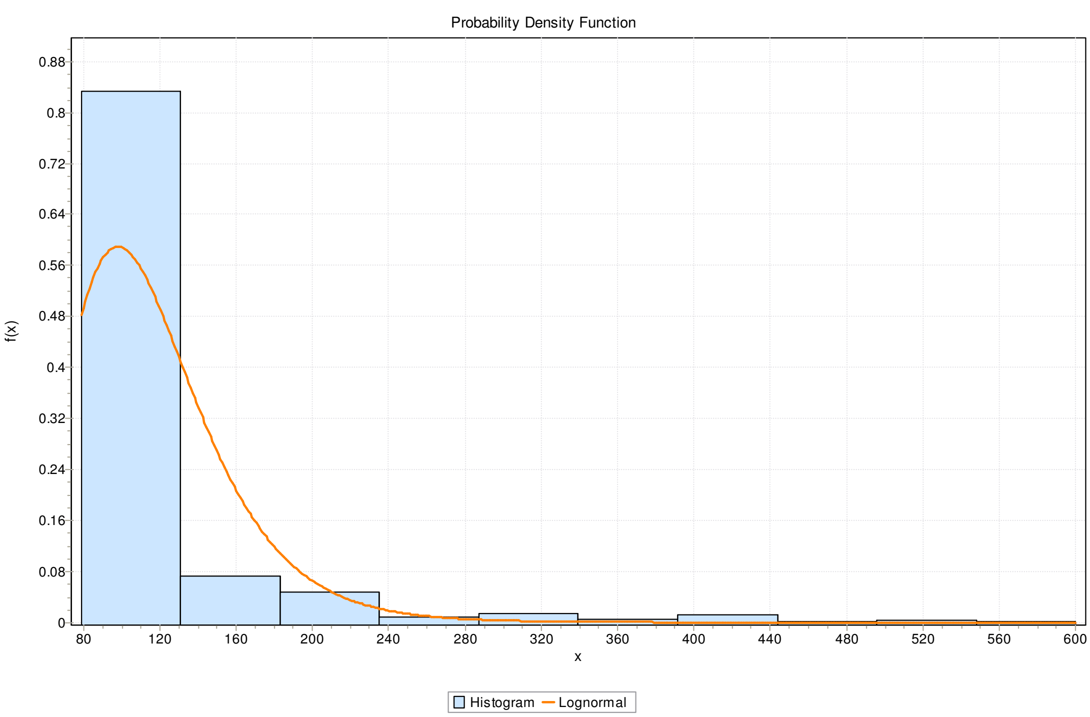
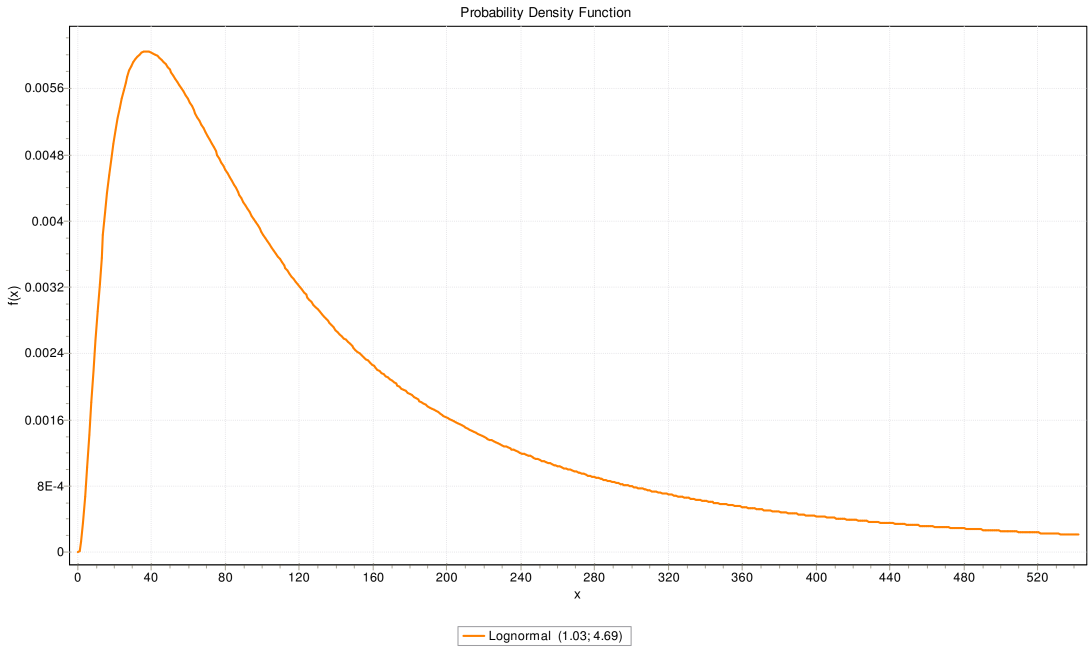


Figure 5: Medellín (μ=4.69,σ=1.03)



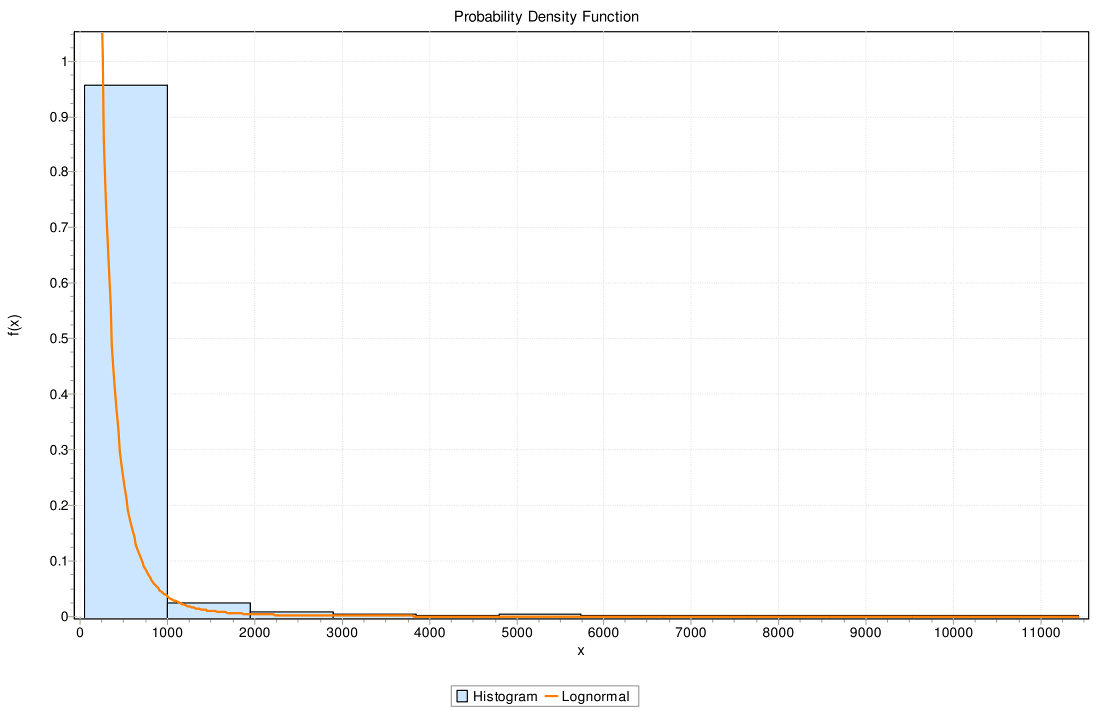
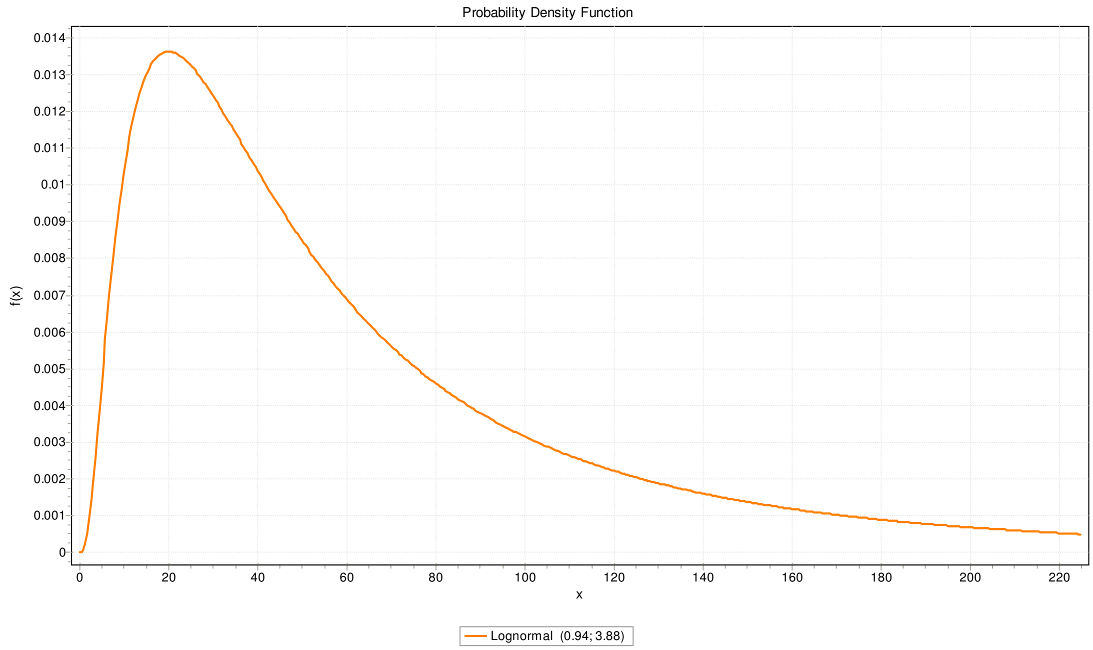


Figure 6: Santo Domingo (μ=3.88,σ=0.94)



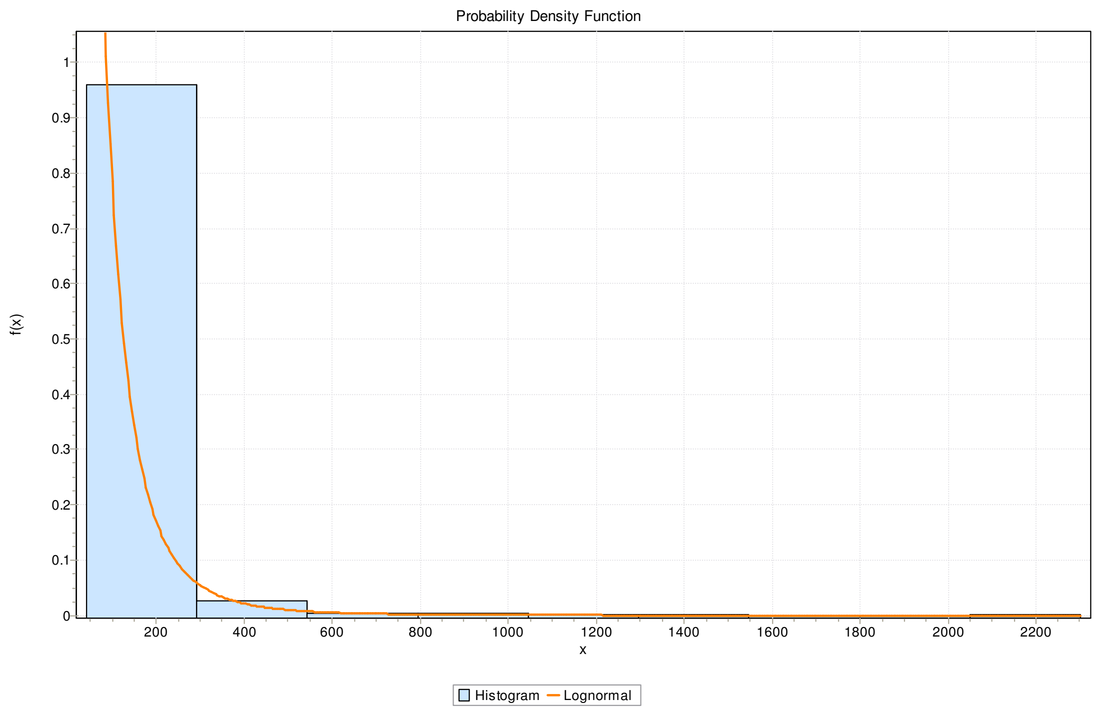
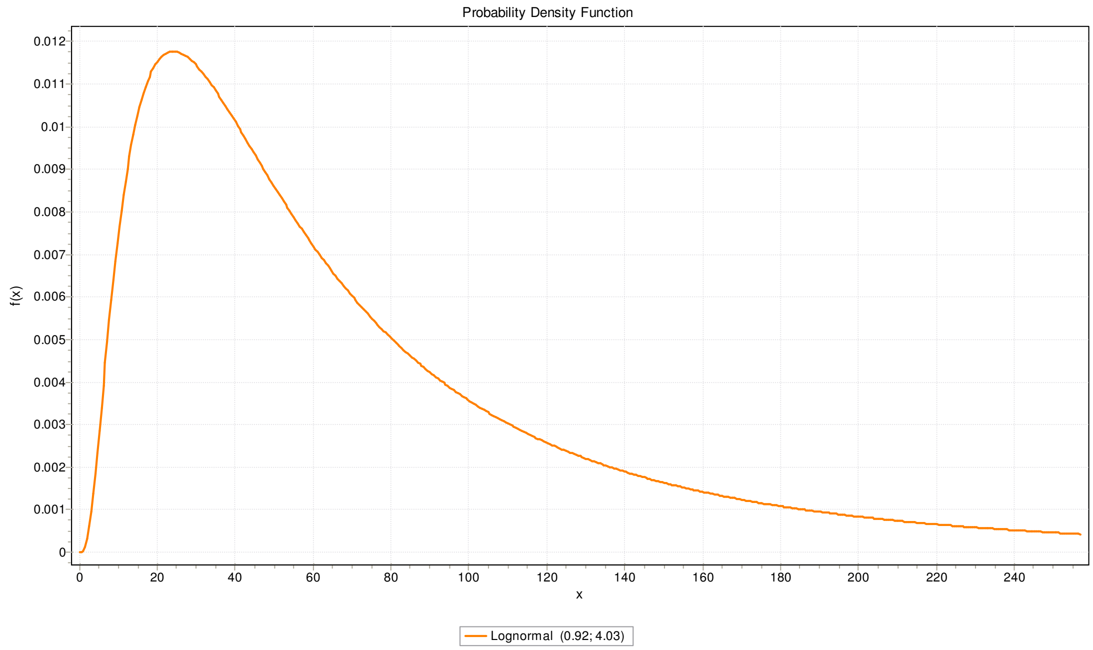


Figure 7: Titiribi (μ=4.03,σ=0.92)



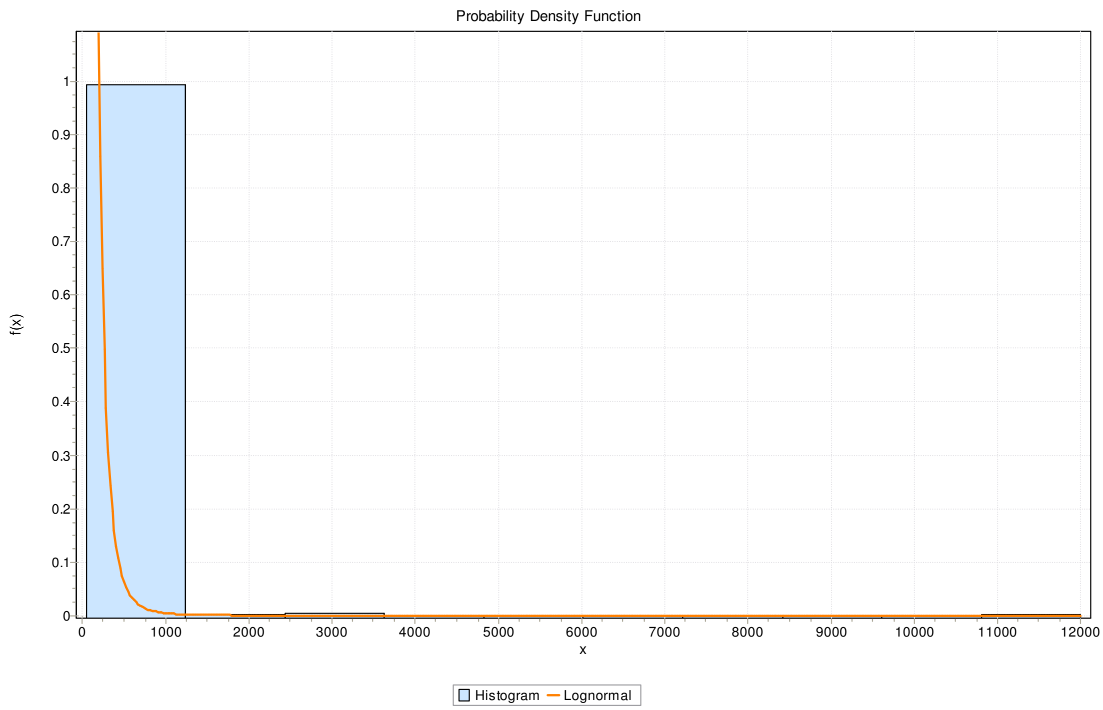
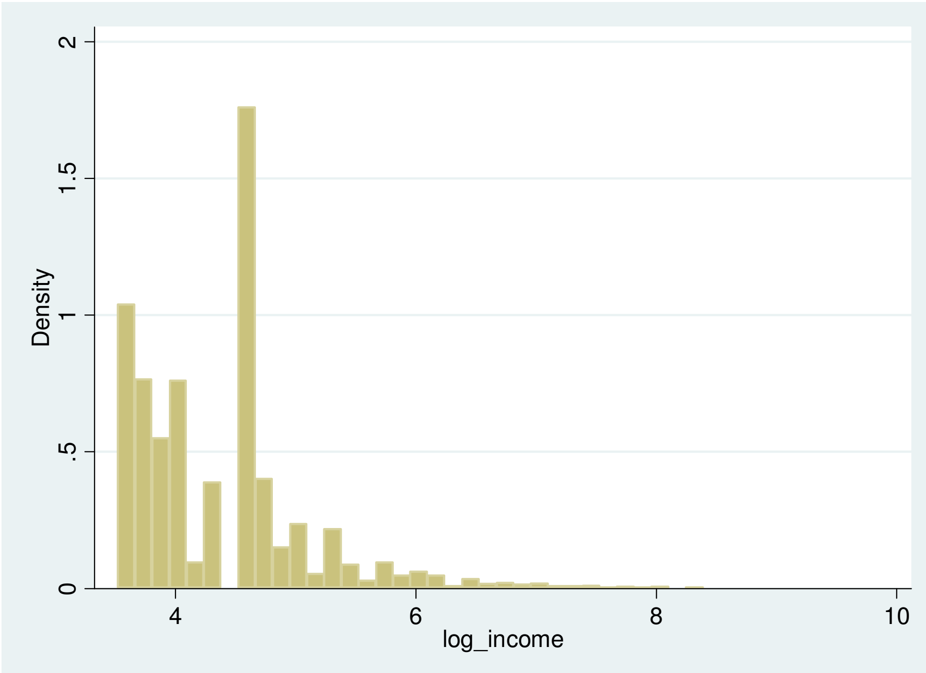


Figure 8: Histogram for log-incomes



Whole sample, log-incomes below 100 pesos are the estimated values from the log-normal distributions.

Table 1: Descriptive Statistics

Descriptive Statistics	Mayors								Not Mayors							
	Obs.	Income	Non-labor Income	Family Size	Family Income	Personal Heterogeneity	Family Heterogeneity	Elite	Obs.	Income	Non-labor Income	Family Size	Family Income	Personal Heterogeneity	Family Heterogeneity	Elite
Amaga		214.5	9	0.02	0.02	0.09	1.28	10		76.9937	102	0.02	0.01	0.02	0.52	271
No. Obs: 973	20	(194.01)		(0.01)	(0.01)	(0.11)	(0.91)		953	(56.01)		(0.01)	(0.01)	(0.01)	(0.71)	
Amalfi*										139.401	226	0.01	0.01	0.07	0.44	397
No. Obs: 1047									1047	(409.41)		(0.01)	(0.01)	(0.11)	(0.91)	
Belen		182.5	20	0.03	0.04	0.03	0.62	10		132.782	237	0.03	0.03	0.04	0.46	202
No. Obs: 382	20	(150.71)		(0.01)	(0.01)	(0.01)	(1.11)		362	(186.91)		(0.01)	(0.01)	(0.11)	(0.81)	
Caldas		326.7333	0	0.03	0.04	0.12	1.74	8		98.1383	0	0.03	0.03	0.04	0.66	173
No. Obs: 579	15	(217.11)		(0.01)	(0.01)	(0.11)	(0.61)		564	(81.71)		(0.01)	(0.01)	(0.11)	(0.61)	
Envigado		341.8571	18	0.03	0.03	0.05	1.02	9		145.017	347	0.02	0.02	0.04	0.70	337
No. Obs: 817	21	(154.31)		(0.01)	(0.01)	(0.11)	(0.51)		796	(173.81)		(0.01)	(0.01)	(0.11)	(0.71)	
Fredonia		196.7938	14	0.02	0.02	0.08	0.98	9		77.9754	162	0.01	0.01	0.03	0.38	344
No. Obs: 1001	16	(114.01)		(0.01)	(0.01)	(0.11)	(0.71)		985	(147.51)		(0.01)	(0.01)	(0.11)	(0.91)	
Girardota		266.1905	17	0.04	0.06	0.09	1.19	14		64.8228	109	0.03	0.03	0.03	0.58	197
No. Obs: 595	21	(307.11)		(0.01)	(0.01)	(0.11)	(0.81)		574	(59.91)		(0.01)	(0.01)	(0.11)	(0.81)	
Hatoviejo		140.8824	14	0.02	0.02	0.05	0.52	7		136.459	50	0.03	0.02	0.04	0.46	110
No. Obs: 245	17	(56.91)		(0.01)	(0.01)	(0.11)	(0.51)		228	(131.41)		(0.01)	(0.01)	(0.11)	(0.91)	
Itagui		264.5556	16	0.02	0.03	0.07	1.50	12		113.766	227	0.02	0.03	0.02	0.71	331
No. Obs: 643	18	(123.41)		(0.01)	(0.01)	(0.11)	(0.71)		625	(53.41)		(0.01)	(0.01)	(0.01)	(0.61)	
La Estrella		181.6429	0	0.02	0.02	0.08	1.17	5		93.8593	0	0.02	0.02	0.02	0.63	236
No. Obs: 626	14	(86.41)		(0.01)	(0.01)	(0.11)	(1.21)		612	(33.31)		(0.01)	(0.01)	(0.01)	(0.71)	
Medellin		1275.719	21	0.01	0.02	0.11	1.52	24		228.09	1176	0.01	0.01	0.04	0.53	1109
No. Obs: 5734	34	(1376.21)		(0.01)	(0.01)	(0.11)	(1.01)		2677	(641.41)		(0.01)	(0.01)	(0.11)	(0.81)	
Nueva Caramanta		312.5	4	0.02	0.04	0.08	1.15	0		119.696	87	0.02	0.02	0.06	0.69	31
No. Obs: 160	4	(131.51)		(0.01)	(0.01)	(0.11)	(1.01)		156	(80.01)		(0.01)	(0.01)	(0.11)	(0.51)	
Santo Domingo		412.5	6	0.02	0.05	0.14	1.33	1		75.5921	50	0.02	0.02	0.04	0.40	206
No. Obs: 503	8	(247.51)		(0.01)	(0.01)	(0.21)	(0.61)		495	(146.21)		(0.01)	(0.01)	(0.11)	(0.81)	
Titiribi		324.577	8	0.03	0.05	0.07	0.82	11		100.399	77	0.02	0.02	0.04	0.31	347
No. Obs: 920	20	(627.61)		(0.01)	(0.01)	(0.11)	(1.51)		900	(431.61)		(0.01)	(0.01)	(0.11)	(0.91)	
Zaragoza*										116.881	71	0.00	0.00	0.04	0.34	128
No. Obs: 937									937	(117.61)		(0.01)	(0.01)	(0.11)	(0.51)	
All**		407.3451	147	0.02	0.03	0.08	1.14	120		133.267	2624	0.02	0.02	0.03	0.53	3894
No. Obs:10155	228	(684.61)		(0.01)	(0.01)	(0.11)	(1.01)		9927	(374.31)		(0.01)	(0.01)	(0.11)	(0.81)	

* No data on majors

**Excluding Amalfi and Zaragoza for which we have no data on majors

Table 2: MLE Estimates of Income Distributions, Conditional Moments, Gini Coefficients and Income per-capita Estimates

District	Lognormal MLE estimates		E[y*]	E[y*|y*<100]	Gini for distribution	Empirical Gini*	Mean Income*	Income Per Capita* **
	mu	sigma						
Amaga	4.21	0.53	77.88	59	0.292	0.221	79.82	20
Amalfi	4.23	0.98	112.4	47.3	0.502	0.573	139.4	52.2
Belen	4.68	0.55	125.6	69.5	0.291	0.337	135.4	-
Caldas	4.4	0.65	101.1	59.75	0.363	0.331	104.1	26.15
Envigado	4.68	0.75	144.5	61.6	0.417	0.419	150.1	30.24
Fredonia	4.01	0.76	73.4	47.7	0.441	0.356	79.9	18
Girardota	3.9	0.84	70.3	44	0.443	0.341	71.9	17.3
Hatoviejo	4.76	0.44	129.17	77.1	0.243	0.257	71.9	25
Itagui	4.7	0.34	116.18	78.9	0.197	0.184	117.9	27.6
La Estrella	4.51	0.28	94.93	78.9	0.152	0.122	95.8	25.5
Medellin	4.69	1.03	185	52.44	0.552	0.645	241.2	41.76
Nueva Caramanta	4.64	0.58	122.87	67.6	0.32	0.313	124.5	25.9
Santo Domingo	3.88	0.94	74.68	41.7	0.485	0.433	80.9	17.74
Titiribi	4.03	0.92	85.94	44.97	0.463	0.513	105.3	22
Zaragoza	4.62	0.43	111.44	73.95	0.244	0.24	116.9	53.6
All Districts	3.99	1.21	111.5	39	0.571	0.51	137.7	30.82

Assuming all individuals with income below 100 have income of E[y|y*<100].

**Based on the 1851 population census

Table 3: Moments Estimates of the MLE Income distributions

District	Lognormal MLE estimates		E[lny*|y*<100]
	μ	σ	
Amaga	4.21	0.53	4.0026
Amalfi	4.23	0.98	3.6702
Belen	4.68	0.55	4.1924
Caldas	4.4	0.65	4.0045
Envigado	4.68	0.75	4.0331
Fredonia	4.01	0.76	3.7251
Girardota	3.9	0.84	3.6053
Hatoviejo	4.76	0.44	4.3048
Itagui	4.7	0.34	4.3656
La Estrella	4.51	0.28	4.3434
Medellin	4.69	1.03	3.8134
Nueva Caramanta	4.64	0.58	4.1548
Santo Domingo	3.88	0.94	3.5229
Titiribi	4.03	0.92	3.6188
Zaragoza	4.62	0.43	4.2674
All Districts	3.99	1.21	3.3791

Table 4: Comparative Per Capita Income between C olombia and the U.S. in 1861

Country	Income Per Capita 1861 Camacho Roldán (pesos)	Maddison (1990 US$)
U.S.	126	2178
Colombia	43	743
Colombia/U.S.	0.34	0.34

Camacho Roldan (1895) and Maddison (2001)

Table5: Regression Results, All Districts

	All districts								
	Linear Probability Model				Simple Probit**			Two-Step Probit**	
	OLS*		2SLS*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Second Stage: Dependent Variable: Major Dummy									
Family Size	0.347	-0.184	-0.271	0.074	19.001	4.031	21.910	1.527	21.190
	(0.500)	(0.454)	(0.441)	(0.486)	(5.498)	(6.866)	(12.527)	(7.502)	(12.902)
Family Size squared	-10.974	-1.683	1.332	-2.137	-240.915	-27.194	-164.894	69.924	-98.568
	(4.483)	(5.436)	(5.794)	(6.047)	(68.548)	(73.452)	(132.603)	(88.452)	(139.101)
Family Income	0.986	0.396	0.156	-0.070	7.457	-2.529	-11.262	-10.474	-19.868
	(0.481)	(0.403)	(0.422)	(0.434)	(2.494)	(3.588)	(5.154)	(3.776)	(5.424)
Personal Heterogeneity	0.301	0.008	-0.107	-0.122	3.356	-1.677	-2.388	-5.383	-6.598
	(0.049)	(0.044)	(0.061)	(0.071)	(0.349)	(0.649)	(0.682)	(0.922)	(0.989)
Family Heterogeneity	-0.0001	0.0027	0.0035	0.0034	0.054	0.060	0.141	0.088	0.208
	(0.006)	(0.006)	(0.006)	(0.008)	(0.046)	(0.060)	(0.087)	(0.064)	(0.089)
Elite family dummy	-0.0021	-0.0054	-0.0065	0.0039	0.0054	-0.037	-1.337	-0.077	-2.017
	(0.005)	(0.005)	(0.005)	(0.012)	(0.062)	(0.076)	(1.006)	(0.069)	(1.023)
Non-Labor-Income Dummy		0.017	0.0036	0.004		0.418	0.377	0.006	-0.056
		(0.006)	(0.006)	(0.007)		(0.096)	(0.100)	(0.101)	(0.121)
District Dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Family Dummies	No	No	No	Yes	No	No	Yes	No	Yes
log Income		0.045	0.067	0.062		0.741	0.804	1.573	1.675
		(0.005)	(0.009)	(0.009)		(0.071)	(0.072)	(0.145)	(0.165)
v_hat								-0.952	-1.004
								(0.157)	(0.168)
Panel B: First Stage: Dependent Variable: log Income									
Family Size			2.824	4.152				2.500	4.049
			(1.123)	(1.400)				(1.030)	(1.283)
Family Size squared			-107.028	-118.684				-108.142	-119.694
			(13.407)	(15.639)				(12.515)	(14.524)
Family Income			9.120	10.134				9.274	9.950
			(0.600)	(0.726)				(0.556)	(0.666)
Personal Heterogeneity			4.494	4.999				4.163	4.661
			(0.094)	(0.101)				(0.078)	(0.084)
Family Heterogeneity			-0.039	-0.092				-0.041	-0.087
			(0.011)	(0.012)				(0.009)	(0.011)
Elite family dummy			0.025	1.492				0.027	1.158
			(0.011)	(0.470)				(0.010)	(0.453)
Non-Labor-Income Dummy			0.012	0.010				0.010	0.006
			(0.017)	(0.018)				(0.016)	(0.016)
D1			-0.865	-0.853				-0.913	-0.901
			(0.019)	(0.020)				(0.017)	(0.017)
D2			-0.306	-0.292				-0.339	-0.324
			(0.017)	(0.017)				(0.015)	(0.015)
District Dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Family Dummies	No	No	No	Yes	No	No	Yes	No	Yes
R-squared	0.0279	0.0696	0.6092	0.6409				0.6294	0.6601
No. of Observations	10088	10088	10139	10088	10139	10139	5242	11792	11727

* Huber-White Robust and Clustered standard errors at the family level in parenthesis

**Bootstrapped standard errors

Constant not reported

Table 6: Average Partial Effects, All Districts for model (9) in table 5

Average Partial Effects	
Log Income	0.16458
Family Size	2.0814
Family Income	-1.9516
Personal Heterogeneity	-0.6481
Family Heterogeneity	0.020403
Elite Family Dummy	-0.19807

Whole Sample (Coefficient estimates from model (9))

Table 7: Regression Results, Medellín

	Medellin								
	Linear Probability Model				Simple Probit**			Two-Step Probit**	
	OLS*		2SLS*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Second Stage: Dependent Variable: Major Dummy									
Family Size	-0.048	-0.904	-0.970		-0.449	-14.840		-221.914	
	0.717	0.758	0.675		32.063	28.928		82.583	
Family Size squared	-2.013	23.621	25.636		-514.517	62.948		6306.193	
	22.965	24.103	22.180		1357.688	892.622		2429.642	
Family Income	-0.520	-0.821	-0.845		1.499	-13.236		-88.094	
	0.597	0.581	0.567		13.569	19.334		31.197	
Personal Heterogeneity	0.119	-0.045	-0.058	-0.052	1.922	-1.910	-3.295	-44.708	-78.997
	0.055	0.054	0.080	0.136	0.665	1.207	1.520	15.199	28.549
Family Heterogeneity	0.024	0.022	0.022	0.140	0.282	0.219	0.416	-0.408	2.966
	0.013	0.012	0.012	0.006	0.120	0.154	0.375	0.278	0.363
Elite family dummy	0.010	0.006	0.005	0.380	0.352	0.214	1.329	-0.384	-10.590
	0.007	0.007	0.007	0.030	0.186	0.198	0.834	0.292	1.064
Non-Labor-Income Dummy		-0.007	-0.008	-0.007		-0.137	-0.183	2.773	4.002
		0.005	0.008	0.009		0.203	0.254	1.130	0.260
District Dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Family Dummies	No	No	No	Yes	No	No	Yes	No	Yes
log Income		0.026	0.029	0.025		0.597	0.745	8.368	10.540
		0.006	0.010	0.013		0.127	0.146	2.770	0.159
v_hat								-7.894	-9.904
								2.794	0.217
Panel B: First Stage: Dependent Variable: log Income									
Family Size			26.535					26.535	
			5.326					5.326	
Family Size squared			-803.214					-803.214	
			149.771					149.771	
Family Income			9.737					9.737	
			2.328					2.328	
Personal Heterogeneity			5.492	7.755				5.492	7.755
			0.215	0.254				0.215	0.254
Family Heterogeneity			0.081	0.031				0.081	0.031
			0.033	0.116				0.033	0.116
Elite family dummy			0.073	0.125				0.073	0.125
			0.036	0.659				0.036	0.659
Non-Labor-Income Dummy			-0.405	-0.448				-0.405	-0.448
			0.050	0.051				0.050	0.051
D1			-1.219	-1.298				-1.219	-1.298
			0.053	0.058				0.053	0.058
D2			-0.615	-0.604				-0.615	-0.604
			0.056	0.057				0.056	0.057
District Dummies	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Family Dummies	No	No	No	Yes	No	No	Yes	No	Yes
R-squared	0.0319	0.0653	0.4716	0.5728				0.4716	0.5728
No. of Observations	2597	2597	2600	2597	2600	2600	512	2600	2597

* Huber-White Robust and Clustered standard errors at the family level
**Bootstrapped standard errors

Appendix

To see this, note that

$$\text{cov}(D_1, \ln y^*) = E[D_1 \ln y^*] - E[D_1]E[\ln y^*] \tag{A1}$$

Now we can note that $E[D_1 lny^*]=E[lny^* \lfloor y^*>100]=\mu_{100}{}^+$ given that D_1 is zero whenever y^* is less than 100, and 1 whenever y^* is at least 100. Also note that $E[D_1]=1-F_{y^*}(100)$, for the same reason. As a result,

$$\text{cov}(D_1, \ln y^*) = \mu_{100}{}^+ - (1 - F_{y^*}(100))\mu > 0 \tag{A2}$$

given that $\mu_{100}{}^+ > \mu$ and $1-F_{y^*}(100)<1$.

On the other hand, we can also see that $cov(D_1,Error)=0$:

$$\begin{aligned} &\text{cov}(D_1, Error) = E[D_1 Error] - E[D_1]E[Error] \\ &= E[D_1 Error] - (1 - F_{y^*}(100)) \cdot 0 \\ &= E[D_1 Error] \end{aligned} \tag{A3}$$

given that the expectation of the error is zero. Now, $E[D_1 Error]=E[Error \lfloor y^*>100]=0$, since the observations with a reported income above 100 pesos do not have measurement error, and D_1 is zero for $y^*<100$. As a result,

$$\text{cov}(D_1, Error) = 0 \tag{A4}$$

(A2) together with (A4) are the necessary conditions for the instrument D_1 to be valid.

Regarding D_2, we have that

$$\text{cov}(D_2, \ln y^*) = E[D_2 \ln y^*] - E[D_2]E[\ln y^*] \tag{A5}$$

From equation (A5) we must note the following: $E[D_2 lny^*]=E[lny^* \lfloor y^*=100]=ln(100)$, given that D_2 is zero whenever y^* is not 100, and 1 when $y^*=100$. On the other hand, $E[D_2]=f_{y^*}(100)$ for the same reason. As a result,

$$\text{cov}(D_2, \ln y^*) = \ln(100) - f_{y^*}(100)\mu \neq 0 \tag{A6}$$

Finally, the covariance between D_2 and the measurement error is:

$$\begin{aligned} &\text{cov}(D_2, Error) = E[D_2 Error] - E[D_2]E[Error] \\ &= E[D_2 Error] - f_{y^*}(100) \cdot 0 \\ &= E[D_2 Error] \\ &= 0 \end{aligned} \tag{A7}$$

Where the last line comes from equation (15) and the assumption from footnote 21, so that $E[D_2 Error]=E[Error|y^*=100]=E[ln(100)-lny^*|ln^*=100]=0$. Equations (A6) and (A7) are the necessary conditions for D_2 to be a valid instrument.