Up from Poverty?

The 1832 Cherokee Land Lottery and the Long-run Distribution of Wealth[*]

Hoyt Bleakley[†] Joseph Ferrie[‡]

First version: December 4, 2012
This version: March 29, 2013

Abstract

The state of Georgia allocated most of its land to the public through a system of lotteries. These episodes provide unusual opportunities to assess the long-term impact of shocks to wealth, as winning was uncorrelated with individual characteristics and the eligible population was drawn from a broad cross section of adult white males. Using wealth measured in the 1850 Census manuscripts, we follow up on a sample of men eligible to win in the 1832 Cherokee Land Lottery. We assess the impact of lottery winning on the distribution of wealth 18 years after the fact. Winners are on average richer (by an amount close to the median of 1850 wealth), but mainly due to a (net) shifting of mass from the middle to the upper tail of the wealth distribution. The lower tail is largely unaffected. We discuss some possible mechanisms (fixed costs, interactions with ability, risk, and life-cycle consumption patterns) for this result.

---

[*] Draft version. Comments welcome. The authors thank Chris Roudiez for managing some of the transcription, and seminar participants at the University of Chicago and Yale University for helpful comments. We gratefully acknowledge funding support from the Stigler Center and the Center for Population Economics, both at the University of Chicago.

[†] Booth School of Business, University of Chicago. Postal address: 5807 South Woodlawn Avenue, Chicago, Illinois 60637. Electronic mail: bleakley at uchicago dot edu.

[‡] Department of Economics, Northwestern University. Postal address: 2001 Sheridan Road, Evanston, Illinois 60208. Electronic mail: ferrie at northwestern dot edu.

1 Introduction

One of the central concerns of policy is the inequality of economic outcomes. This in motivated, in part, by a concern for the plight of the poor and a preference for less inequality of outcome. In the background, nonetheless, is the concern that poverty alleviation and wealth transfers dull the incentive for productive activity. But attention to the distribution is also often motivated by preoccupation with so-called poverty traps, in which the poor might find it hard to make investments (in physical or human capital), even if the return is high. In such cases, an inequality of outcome (perhaps of the previous generation) implies an inequality of opportunity. In the present study, we turn our attention to the wealth distribution. Analysis of the distributional effect of transfers depends both on the measurement of the wealth distribution as well as credibly exogenous variation in transfers, which is usually an endogenous response to an individual's misfortune. One should also take care to distinguish between short-run effects of transfers, which change the wealth distribution purely as a matter of accounting, versus more persistent effects. Such longer-run effects could either amplify or attenuate the initial transfer, depending on the underlying causes of the initial distribution.

In the present study, we analyze a large-scale lottery to consider the effect of a random disbursement of wealth on the wealth distribution in the long run. Those eligible to be treated represent an approximately representative cross-section of household heads, unlike other studies of lotteries whose participants are a selective subset of the population. The prize in this lottery was a claim on a parcel of land. The average value of such parcels was large—comparable to the median level of wealth at the time. Winning in the lottery was close to a pure wealth shock: there were no strings attached to the land (such as a homesteading requirement) and the claim could be liquidated immediately. In addition, we consider a historical episode, which allows us to retrospectively examine the distributional effect in the long run, almost two decades after the lottery took place.

Specifically, we investigate the aftermath of the 1832 Cherokee Land Lottery in the US state of Georgia. In the early 19th century, Georgia opened almost three-quarters of its total land area to white[1] settlers in a series of lotteries. In the history of land opening,

---

[1] Slaves and free people of color were excluded from the lottery, as were Native Americans. Indeed, while the present study is focused on distributional changes for white men in Georgia eligible for the lottery, it bears mentioning that the land was expropriated (i.e., redistributed) from the Cherokees, who were subsequently expelled from northwest Georgia in a forced march known as the Trail of Tears.

this was an unusual allocation method, chosen in large measure for its sheer transparency in the wake of several tumultuous corruption scandals in Georgia in the 1790s. We conduct a follow-up on these random wealth shocks using a sample of over 14,000 men eligible to win land in that lottery. To ascertain the long-term effect on the wealth distribution, we transcribe information on wealth from the 1850 Census manuscripts, measured almost two decades after the lottery. The two measures of wealth available in the 1850 Census are real-estate and slave holdings. From this sample of eligibles, we identify winners using a list published by the state of Georgia (Smith, 1838). Those identified in the Smith list comprise the treatment group, and the lottery eligibles that were not linked to the Smith list serve as a control group. While, in theory, not all of the men in our sample of eligibles were technically eligible to win the lottery, our analysis in Section 4 suggests this was a minor subset in practice. Further, in our sample, lottery losers look similar to lottery winners in a series of placebo checks found in Section 4 and 5.5.

As a point of departure, consider first the mechanical effect on the wealth distribution of randomly assigned wealth. If everyone receives (and holds on to) the same dollar amount, this simply shifts the entire distribution of wealth, in levels, to the right by that same amount. It is nevertheless common to treat the wealth distribution in natural logarithms, which would be strongly compressed in such a circumstance. (A wealth shock of a given size represents a much larger fraction at the lower tail of the distribution.) If instead there are much higher returns to capital at the low-end, as argued by some,[2] then a constant-level disbursement would compress the distribution both in levels and, to a greater degree, in logs. A complication, however, comes from the heterogeneity in quality for the lotteried parcels. While this would increase the variance of the treatment wealth distribution relative to control, it would still have the effect of draining mass out of the lower tail by the random nature of the lottery, as long as the value of winning the lottery was positive. These cases provide a point of comparison for the empirical results, discussed next.

Almost two decades after the lottery, winners are, on average, \$700[3] richer than a compa-

---

[2] We review related literature in Section 2.

[3] Dollar figures reported in the study are in 1850 dollars unless otherwise specified. We suggest a few different ways to contextualize this number. First, as stated above, this is approximately equal to the median of wealth in our sample. If instead we convert this number to 2010 values using consumer prices, it is approximately \$20,200. In contrast, it would convert to \$142,000 in 2010 if adjusted by the relative value of the unskilled wage. (This latter figure translates to over ten years of earnings at the 2010 federal minimum wage of \$7.25 per hour for a full-time/full-year worker.) These conversion factors come from MeasuringWorth.com (Williamson, 2013).

rable population that did not win the lottery. The gains in wealth, however, are not evenly distributed among the lottery winners. Indeed, the poorest third of lottery winners are almost exactly as poor as the poorest third among those who did not win the lottery. Rather, the gains from lottery winning are almost entirely seen as a (net) shifting of mass from the middle of the wealth distribution to the upper tail. The lower tail is largely unaffected. Therefore this wealth shock tended to exacerbate inequality (at least when considering the poor versus the rest) rather than reduce it. These results are found in Section 5, where we compare the probability density functions (PDFs) and cumulative density functions (CDFs) of both control and treatment wealth distributions. Further, in Section 5, we use a quantile-regression estimator to show that winning the lottery affects wealth mostly in the upper half of the distribution.[4] We also show that these results are robust to controlling for various factors, including characteristics of the person's name. The latter strengthens the earlier conclusions in that, although we used the name to link to the list of winners, it did not appear to bias our estimate of the treatment effect.

Whether the wealth transfer actually increased the aggregate welfare of the treated depends on one's taste for equity, as we show in Section 5.4.[5] Various measures of inequality, such as the Gini coefficient or the standard deviation of log wealth, are higher in the treatment group than in the controls. We use a constant-elasticity-of-substitution aggregator, bootstrapped over both groups, to ask whether the treatment distribution shows a statistically distinguishable improvement over the control group under different preferences about the size of the pie versus how it is sliced. For very large elasticities of substitution (and correspondingly low weights on equity), the treatment group has a significantly higher aggregate outcome than control. But we cannot reject equality of outcomes for elasticities of substitution much below one, a far cry from a Rawlsian elasticity of zero in which social welfare depends exclusively on the well-being of the least-well-off individual.

---

[4] Absent the property of rank invariance across the distributions of potential outcomes, we cannot literally interpret these effects as the treatment effects at a given point in the control distribution, in that these results could have arisen through more complicated patterns of reshuffling from control to treatment. For example, all of the would-have-been-poor could have become rich and an equal number of the would-have-been-rich could have become poor as a consequence of treatment.

[5] For the purposes of the present study, we set aside issues of efficiency. The efficiency loss associated with the lottery could be twofold. First, by not selling the land at its market value, the state of Georgia was foregoing revenues which then would have to be raised from more distortionary taxes. Second, opening the land through a lottery was a peculiar form of "market design" that appeared to constrain land use well into the 20th century. This latter issue is discussed in detail in Weiman (1991) and Bleakley and Ferrie (2013).

In Section 6, we provide suggestive evidence on some of the mechanisms that might have prevented these wealth transfers from moving the wealth distribution, as alluded to in the title, "up from poverty." Our analysis suggests the main culprit is greater vulnerability to risk at the low end of the (counterfactual) distribution, and not a fixed cost, an interaction with observed proxies of ability, or life-cycle considerations. Apropos of fixed costs, we note that significant differences in the distribution are only seen for wealth in the thousands of dollars. This is too high to plausibly be related to the fixed cost of farming in antebellum Georgia. Indeed, we see farmers operating with less than $100 of land. Further, a slave (manifestly difficult to purchase in non-integral units) would have cost several hundred rather than several thousand dollars. An alternative possible mechanism is that lottery winnings alone are insufficient, but rather they must be paired with some complementary skill. Indeed, those at the bottom of the distribution (either control or treatment) may have been there precisely because they lack the ability to seize opportunities. Nevertheless, we do not find evidence of such interactions in our data. First, across the treatment distribution of 1850 wealth, there was a stable rate of claiming land by lottery winners; therefore those who ended up poor did not do so because they failed to collect their winnings at the outset. Second, we construct averages of wealth, fertility, and school attendance among (other) people in 1850 Georgia with the same surname, which serve as proxies of human capital along patrilineal lines. (Own literacy is another measure that we use for human capital.) While these measures are strongly predictive of own outcomes, they do not have strong interactions with winning the lottery. Thus, at the low end of the wealth distribution, a lack of observed ability does not seem to explain the heterogeneous effects of treatment. Additionally, these results are not apparently related to life-cycle considerations. Serendipitously, our sample is concentrated within ages where wealth is either rising or near its peak. Thus, the similarity in in the lower tails cannot be attributed to widespread decumulation of assets in old age. The remaining explanation, which admittedly is a residual category, is exposure to risk. By linking a subsample to the 1860 Census, we show that there was a large degree of risk and churn in the wealth distribution of that period. (Unfortunately, Census data on wealth is not available prior to 1850.)

Section 7 concludes the study.

2 Related Literature

The condition of the small entrepreneur is a topic that has received attention across a wide variety of contexts and disciplines. We cannot hope to give a proper survey here, so instead in this section we touch on a few relevant examples from various perspectives. For example, Thomas Jefferson and his later intellectual disciples argued for policies that would encourage yeomen farming (i.e., small-scale and owner-operated farms) over large estates and urban factories, both employing landless laborers.[6] This view gave rise to land policies in the 19th century US that distributed small landholdings on the frontier at low, often below-market, prices.

While the intent of the policy was to establish the dominance of small farming, the extent to which it did so may have been limited by other factors. Indeed, Paul W. Gates (1996) argues the so-called free-land policy was ineffective, because small-scale settlers were often capital constrained and probably were outbid, outmaneuvered, or bought out by those he called "frontier estate builders" (chapter 2 title, on page 23). Relatedly, Jeremy Atack (1988) shows that rates of landlessness among agricultural workers in the Midwest were at similar levels in 1860 (when ostensibly free land was still available on the frontier) and in 1880 (as the frontier was closing). To some extent this is a puzzle; giving free land in relatively small parcels to individuals will have the mechanical effect of compressing the land-wealth distribution in the short-run. The question remains, however, whether this compression will stick, or whether it will be unwound by some other feature of the economic environment that makes it difficult for farmers to operate at such a small scale.

More recently, Hernando de Soto (1987) and Tarrun Khanna (2007) highlight the ubiquity of small-scale entrepreneurs in developing economies, often hidden in the informal sector. De Soto (2002) also presents parallels between the antebellum US and developing economies today. While his central thesis is that economic development is (or was) held back by conflicting property rights, the undergirding theme is that capital markets fail(ed) to direct resources to a large class of small entrepreneurs (including small farmers), who could otherwise make productive use of such capital.

At some level, this line of thinking is related to the notion of a "poverty trap," which appears in a wide range of theoretical papers. There are many possible motivations for the

---

[6] See, for instance, Jefferson's *Notes on the State of Virginia*, in response to "Query 19", at http://etext.virginia.edu/toc/modeng/public/JefVirg.html.

existence of such traps; perhaps the easiest one to think about is a simple fixed cost of production. An example from this theoretical literature is by Abhijit Banerjee and Andrew Newman (1993) who, using a model with a poverty trap to analyze occupational choice (being a laborer versus an entrepreneur, e.g.), demonstrate the possibility of multiple steady-states for the wealth distribution.

But are such poverty traps empirically relevant for small entrepreneurs in developing economies? In principle, one could take detailed measurements of the production function, although demonstrating the poverty trap requires precise evidence on the third derivative of the production function, which likely renders the strategy impractical. An alternative approach would be to randomly disburse capital to entrepreneurs and attempt to measure how this changes the distribution of their outcomes. This is approach of the present study. Perhaps the most closely related work is by Suresh de Mel, David McKenzie, and Chris Woodruff (2008), who examine the impact on profitability of randomly assigned capital grants to a sample of self-employed in Sri Lanka. In the relatively short run, they find very large returns to additional capital. (Apart from the obvious difference in location and time period, an important distinction between their work and ours is that their follow-up was on the order of months rather than decades.)

Risk is another central feature of the environment that an entrepreneur faces, perhaps to an even greater degree at a small scale of operation. Banerjee and Esther Duflo (2011) argue that there is "so much risk in the everyday lives of the poor is that, somewhat paradoxically, events that are perceived to be cataclysmic in rich countries often seem to barely register with them (page 136)." They provide some illustrative anecdotes of such risk.[7] Note that high returns can exist in the short run perhaps as compensation for high risk that becomes more evident at longer horizons. Thus supernormal returns in the short run are not necessarily indication of a binding capital constraint. Instead, it might indicate a failure of diversification, a distortion that can itself hold back economic development, as in the model of Daron Acemoglu and Fabrizio Zilibotti (1997). Further, Dean Karlan, Robert Osei, Isaac Osei-Akoto, and Christopher Udry (2012) show how agricultural decisions change with the provision of insurance for small farmers in Ghana. Gavin Wright (1975) argued that small

---

[7] Richer detail, albeit from nonacademic sources, is presented by the journalist Katherine Boo (2012), who relates some of the difficult shocks endured by several families in an informal settlement in Mumbai, and by Laura Ingalls Wilder (1971), who details her own experience as a young mother on the 19th-century US frontier.

Antebellum Southern farmers practiced “safety first” farming because their risk exposure was so great. Roger Ransom (2005) labelled the Antebellum period as “the era of walk-away farming,” in which small farmers could cope with bad shocks by simply abandoning their land (and presumably their debts as well). In contrast, wealthier farmers were better able equipped to self-insure and thus not be obliged to abandon their wealth in response to a transitory negative shock.

Heterogeneity in returns might also arise for reasons that do not bring the specter of inefficiency. Consider Theodore W. Schultz’s (1975) argument that ability or human capital helps one take advantage of new opportunities. Indeed, a basic notion of economics is that factors of production should gravitate to their highest valued use. If the experimentalist somehow manages to perturb the distribution of factors away from the baseline, this logic suggests that we should expect a reduction in average returns. It is likely that skill and wealth are complementary, and furthermore that at least some at the bottom of the (treatment or counterfactual) wealth distribution were there precisely because they lacked the ability to seize opportunities such as winning the lottery.

Finally, there is earlier work that also analyzes the wealth shock coming from lottery winnings. Guido W. Imbens, Donald B. Rubin and Bruce I. Sacerdote (2001) follow up on the consumption behavior of people who had won large jackpots in state-run lotteries in Massachusetts. Scott Hankins, Mark Hoekstra, and Paige Marta Skiba (2010) examine medium-sized jackpots in the Florida lottery and relate this to bankruptcy filings over the following several years. Both of these studies are strongly related to the present one by using lotteries to analyze wealth, although neither considers a developing-economy context and in neither case is the sample size large enough to permit the distributional analysis that we conduct below. Further, a perennial concern about examining the shock from gambling winnings is that one can only analyze the effect on gamblers, who are typically are highly selected population. As we discuss below, the people who participation in the 1832 lottery was so widespread (at least, among white adult men resident in Georgia circa 1830) that this selection issue is less important in our case.

3 The Cherokee Land Lottery of 1832

The state of Georgia is quite unusual in the U.S. in that much of the state's territory was distributed through a series of land lotteries. The initial Georgia colony was concentrated around the Savannah River, and this land was distributed through a more traditional grant-based system. However, a corruption scandal in the 1790s (the Yazoo Land Fraud) provoked such popular outrage that the Georgia Legislature opted to use lotteries as methods of distributing land from then forward. The first lottery took place in 1805 and the last ones were held in the early 1830s.

For this study we consider the 1832 lottery of Cherokee County in northwest Georgia. We choose to focus on the 1832 lottery because the list of winners was available and the later date increases the chance of tracking these people in census data. The land in this area was made available to white settlers by the eviction of the Cherokee from that area.

Essentially every adult male who had been resident for at least three years in Georgia by 1832 was eligible to one draw in this lottery. Widows, orphans, and certain veterans were eligible for two draws. (Because we would not know in the control group who was a widow, orphan, or veteran, we exclude them from the treated group in our analysis. Practically speaking, this is of little consequence because our sample excludes females and excludes years of birth that the veterans or orphans would disproportionately populate.) A group of highwaymen called the "Pony Club" that operated in old Cherokee County was also explicitly excluded from the lottery, but this group was trivially small compared to the population of the state. In theory, winners in previous lottery waves were excluded from participating, and there was also a 12.5¢ registration fee. It is not immediately evident the extent to which either of these was enforced, but the numbers suggest that neither was much of an impediment to participating. We do not know the exact population in late 1832 of white men meeting the requirements for age (18+) and for residency (3+ years in Georgia), but the 1830 Census reports the white male population of Georgia ages 15 and over in 1830 as approximately 80,000.[8] There were close to 15,000 winners (excluding widows and orphans) in the 1832 Land Lottery (Smith, 1838), which implies a winning rate of around 19%. These are close to, albeit a bit higher than, rates in Columbia County (16.0%) and Oglethorpe County (16.8%), two counties for which we have lists of actual (rather than inferred) registrants and winners.[9]

---

[8] Authors' calculation from the 1830 Census data in ICPSR study 2896 (Haines, 2010).

[9] Cadle (1991, page 278) reports that the total number of registrants was around 85,000, but does not give

Lists of the eligible population were constructed by each county government and forwarded to the state capital in Milledgeville.

Concurrent with this, the area known at the time as Cherokee County was divided into four sections, which were further subdivided into dozens of districts. The districts were generally square, except for those that were on the boundaries of the original Cherokee County, which were defined by the state border to the north and west, and by the Chattahoochee River to the southeast. Surveyors were sent to each district with the aim of further subdividing it into an 18×18 grid of square parcels of 160 acres each.

After the surveys were completed and the lists of eligibles were collected, the lottery began. The drawing proceeded as follows. One drum was filled with slips of paper containing the registration information on each eligible person. Another drum was filled with slips of paper specifying a parcel. Blank slips were added to parcel barrel to equalize the number of pieces of paper in each barrel. A slip of paper was drawn simultaneously from each barrel to determine who had won which parcel. (Thus, lottery losers were those matched to a blank piece of paper.) This implies that winning and losing was assigned randomly, and also that the specific parcel awarded to an individual, even conditional on winning, was random.[10] Over 18,000 parcels were assigned in this manner.

Very few requirements were imposed on the winners of the lottery. They were not required to homestead the parcel for any amount of time. They were not even required to set foot on their parcel. They simply had to register their claim with the state government and pay a nominal fee ($18). If they wished, they could immediately resell title to that parcel. Indeed, it is likely that many of the winners took this route. One factor that made this sort of "flipping" attractive is that it took six years before the state of Georgia could effectively exercise its jurisdiction over this land. The Cherokee nation fought the eviction through

---

the breakdown by single- versus double-draw categories. We use the distribution of single-draw and double-draw winners in the Smith (1838) book to infer this breakdown, and compute that approximately 75,822 registered for the single draws. The 1830 Census reports 77,968 white men aged 15 and older in Georgia in 1830. Comparison of these numbers indicates that around 97.2% of group in 1830 indeed registered for the lottery. The remaining 2.8% might easily be explained by a combination of mortality or emigration between 1830 and 1832, in-migration to Georgia between 1829 and summer 1830 (thus missing the full three years of the residency requirement), and that only some of the 15-year-olds in on June 1 of 1830 would have attained 18 years of age by the fall of 1832, when the drawing was held.

[10] Weiman, 1991, argues that the lottery's outcomes appeared approximately random. Both barrels were rolled around to ensure adequate mixing (or proper randomization, in today's parlance). The blank slips of paper further increased the transparency of the process; it was thus more difficult to increase your odds by excluding other names from ever making it into the barrel. There were a few instances of corruption after the fact that were easily discovered by virtue of the transparent nature of the lottery (Cadle, 1991).

the legal system, and the state of Georgia was not able to evict the Cherokees until 1838. Information on the parcels as well as a list of winners was circulated throughout the state and compiled into a single source by Smith (1838).

A rough measure of the value of a winning draw in the lottery can be obtained by calculating the average value of a farm in the 10 counties of Northwest Georgia in 1850, when the U.S. Census first provides the information[11] necessary to make this calculation. These counties contained 1.5 million acres of farmland (improved and unimproved); the 7,236 farms in these counties had a total cash value of \$8.9 million (1850 dollars), of which \$407,000 was implements and machinery. This yields a land value in 1850 of \$5.71/acre, or just over \$900 for a farm of 160 acres. If winners held their land until 1850, we would expect them to be \$900 richer. If they sold it before and bought land with a similar net present value (NPV), we would expect the same. These effects might be attenuated, however: wealth could be held in other forms (e.g., slaves, which we observe by linking to the 1850 Slave Schedule) or financial assets (very rare, except for the wealthiest); wealth could be consumed (in a variety of forms: direct consumption goods or larger family sizes). Additionally, those who flipped the land quickly may have received less than the land's NPV because of uncertainty about the exact timing of the expulsion of the Cherokee. There should have been little doubt about their eventual eviction, however. The Indian Removal Act was passed in 1830, and had been applied several times already in the region.

Roughly the bottom third of Cherokee County was distributed in 40 acre parcels instead as part of a Gold Lottery. It was thought that this area was particularly rich in gold deposits, an assumption which proved to be overly optimistic. (For this study, we examine only winners in the Land Lottery section of old Cherokee County.)

4 Data

4.1 Sources and Construction

The present study follows up on the outcomes of lottery winners and losers. There are two principal ingredients to this exercise. First, we need to identify who was eligible, and who won. Second, we need to find these individuals in later, publicly available data sources,

---

[11] These numbers are reported in ICPSR study 2896 (Haines, 2010).

so as to follow up on their outcomes. For the most part, we search for these individuals in the Census manuscripts of 1850 using a preliminary version of the full-count file for the 1850 Population Census from the IPUMS project, indexed and scanned images of the 1850 manuscript pages from Ancestry.com, and index of the 1850 Slave Schedule on Ancestry.com.

The original source for the names of lottery winners in the 1832 Georgia land lottery is Smith (1838). He lists, in numerical order, each parcel that was available and the associated lottery winner, along with the winner's county and minor civil division in 1832. Smith's list was partially transcribed and available on accessgenealogy.com, which we downloaded, cleaned, and compared with a copy of Smith (1838) that we scanned and transcribed with an OCR program.

In order to generate a control and treatment group for this lottery, we took advantage of the lottery's entry requirements: individuals had to be 18 years or older in 1832 and resident in Georgia for at least three years by 1832. We extracted all males from the complete count file of the 1850 U.S. Census who met two criteria: (1) they had at least one child born in Georgia in the three years prior to 1832; and (2) they had no children born outside of Georgia in those same years. This yielded a population of 14,306 individuals. Of these, 1,758 were then identified in the list of lottery winners based on their surname and given name. These individuals were then sought in the 1850 census manuscripts to obtain their 1850 real estate value; the complete count file directly provided the other outcomes we will explore below (county of residence in 1850, and marital status in 1850 and the number of children born between the 1832 lottery and the 1850 census). For example, Josiah Spivey, a winner found in the Smith list, was located in 1850 in Township 23 of Tallapoosa County, Alabama. He appears with his wife Mary and son James (born 1828 in Georgia) beginning at line 7 and reported $1,200 in real estate wealth in 1850.

4.2 Summary Statistics and Balancing Tests

We present summary statistics for the sample in Table 1. Each row presents a different variable, and variables are grouped thematically into panels. Means and standard deviations (in parentheses) are shown for each variable. These values for the whole sample are seen in Column 1, and then we provide decompositions based on each individual's likely lottery status in Columns 2 and 3, which report the summary statistics for lottery losers and winners, respectively. Additionally, in Column 4, we report the p-value of a test of the difference in means between these two subsamples. We implement this test with a bivariate regression on a dummy variable for being a lottery winner. In the cases below in which there is a grouped-data structure, such as that the household or surname level, we cluster the standard errors. The number in square brackets in each row reports the sample size used to compute this test statistic.

In the present study, we consider two measures of whether the person won land in the drawing for the Cherokee Land Lottery of 1832. Summary statistics for these variables are found in Panel A of Table 1. The first measure is coded to one if that person is a unique match to a name found on the list of winners published by Smith (1838). Anyone else is coded to zero, including individuals who were among several persons matched to the same winner's name. As is seen in the table, 12.4% of our sample is matched to the list of lottery winners. By construction, this variable takes on means of zero and one in Columns 2 and 3. In the second measure, we attempt to accommodate the relatively small fraction of individuals that tie for a match to the Smith list with others in our sample. In the case of a tie among n observations, we recode the match variable to $1/n$. This recoding of the variable is motivated by the belief that one member of the tying set did in fact win in the lottery, but we do not know which and thus distribute the probability of winning evenly across the group as if we had a uniform prior. More sophisticated (i.e., nonuniform) versions of a assigning partial treatment values within such groups are possible, but we shied away from this approach because of the lack of appropriate benchmark data with which to calibrate such an approach. The average value for this variable is 15.5% in our sample, which is approximately 3% higher than the binary match variable and just slightly below the rates discussed above. The vast majority of differences occur because numerous groups of small-n ties were recoded from zero up to $1/n$. These two lottery-status variables are extremely highly correlated: the regression coefficient of the second measure on the first has a t statistic of 329.

Next, we consider in Panel B of Table 1 a series of outcomes that were determined prior to the realization of the lottery, and therefore should be unaffected by whether the individual won land in the 1832 lottery. Analysis of these outcomes therefore serves as a balancing test when comparing the control and treatment samples. The lottery-eligible men in the sample are approximately 51 years old in 1850, and average age is similar between winners and losers. Almost 50% of the sample was born in Georgia, with the bulk of the remainder being born in the Carolinas. These fractions are statistically similar across groups. By the construction of the sample, these individuals have at least one child born in Georgia in the three years prior to the 1832 lottery. But there is no reason why lottery status should correlate to the number of children born in this earlier period, if our sampling design has drawn an appropriately matched treatment and control group. Indeed, we do find that the sample has approximately 1.33 children born in the pre-lottery window, and this number is quite similar between the two subsamples. The next variable that we consider is whether the individual could read and write. While this variable is measured in 1850 and could theoretically be affected by the lottery some 18 years prior, literacy was more likely realized in childhood. These men, if they had won the lottery or not, would be unlikely to undertake remedial education in literacy given that they were already adults in 1832 and had on their shoulders the demands of supporting a family in a largely agrarian society. By this measure, almost 15% of our sample was illiterate, with insignificant differences between the control and treatment groups. (Note that this was probably a fairly weak test of literacy in that many enumerators classified someone as literate if they could read and write their name. Rates of illiteracy were considerably higher if a more modern standard of literacy was applied.)

In the rest of Panel B, we examine characteristics based on the individual's surname, which was inherited from the father at birth and therefore predates the lottery. As there was probably very little phonetic change in the surname over the life course (or even across generations), the low rates of literacy and somewhat lax orthography of the time might have occasioned some drift in how the surname was spelled. For example, in the census manuscripts the surname "Blakely" has variants "Blakeley," "Bleakley," "Blakelee," and others, as does "Ferry" have the variant "Ferrie." To accommodate this heterogeneity in spelling, we use the Soundex version of the name, which reclassifies names that are phonetically similar into a single code. The first surname-based outcome that we consider is the number of letters in that name (and for this outcome alone we use the original surname rather than the Soundex version). On average, surnames have 6.2 characters, and this average is indeed slightly lower in the subsample of lottery winners. Next we find that the average person has a surname that appears 36 times in the sample, and this is not significantly different between subsamples. We also find that 10% of the sample has a surname that begins with the letter 'M' or 'O' (correlated with Celtic origin), and this rate is insignificantly different between the group of winners and losers. Indeed, for a cross tabulation of lottery status and the first letter of last name, a chi-squared test (d.f.=26) of the equality of distribution across groups has a value of 20 (p=0.8).

The final set of surname-related outcomes that we present in Panel B are constructed from the average characteristics of others in Georgia with the same surname. We restrict ourselves to Georgia in part to maintain similarity with our sample and also because we had access to a full transcription of the 1850 census for the counties in Georgia starting with the letters A-J that was provided to us by the IPUMS project. We took this transcription file and formed averages by surname (again using the Soundex recoding of surname) for various outcomes. To prevent any mechanical contamination from our lottery-eligible sample, we exclude anyone in our sample from the construction of the surname-level averages. The mean surname-average of real estate wealth for our sample (again, not their real estate wealth but the average wealth of those people with the same surname) is approximately \$1200. Because wealth is right-skewed, the mean presents a somewhat misleading picture, and accordingly we find the median wealth among individuals with the same surname is considerably lower: less than \$300. The surname-level illiteracy rate is almost 22%. None of these surname-level outcomes show a statistically significant difference when comparing the lottery winners versus losers. (Some readers might argue that this is a weak test because perhaps the surname-level averages are measured with considerable noise. Nevertheless, we show below in Section 6.2 that the surname averages are strong predictors of individual-level behavior, even when conditioning on demographic and locational covariates. We also test for interactions of winning the lottery with these surname averages below as a test of heterogeneity in the response to wealth shocks.)

In Panel C, we present summary statistics for measures of wealth in 1850. Note that this panel and the rest of the table can no longer be considered part of a balancing test in that we examine outcomes that might very well be affected by winning the lottery. For this panel, the numbers in curly brackets display the 25th, 50th, and 75th percentiles, respectively. The first measure that we consider is real estate wealth. The whole-sample mean is approximately \$2000 and the median is \$650. Unlike many of the outcomes above, here the mean differences by lottery status is significant for an $\alpha = 10\%$ level. Real-estate wealth also shows differences at the median, although not in the upper or lower tails. Next we consider statistics for slave wealth, which had a mean of approximately \$1340, and a statistically significant difference in means by lottery status. The final row of Panel C displays the sum of these two wealth components, which we label "total wealth" throughout the paper.[12] This variable, whose mean is over \$3000, shows a several-hundred-dollar difference between control and treatment groups, which is both economically and statistically significant. The mean difference in total wealth that we observe between lottery winners and losers is close in magnitude to our earlier back-of-the-envelope estimate of the value of the land won in the lottery. The median and 75th percentile is higher in the treatment versus control, but the 25th percentile is the same. Further, a Kolmogorov-Smirnov test rejects equality of the control and treatment wealth distributions at an $\alpha = 5\%$ level.

Finally, the vast majority of the sample still lives in Georgia in 1850, and the bulk of the remainder resides in Alabama. Appendix Figure 1 displays the geographical distribution of our sample by county in 1850. The black lines display the 1850 county boundaries from the NHGIS project (Minnesota Population Center, 2004). A red dot in the center of a county indicates the presence of observations from our sample, and the size of the dot increases in proportion to the number of observations. (There are several counties in the densely packed region that have few or no observations in our sample, such as Carroll and Union Counties in Georgia and Benton/Calhoun County in Alabama. These are counties with large numbers of missing or illegible pages from the census manuscripts. Counties with no dots on the periphery of the map, however, likely reflect the absence of any households that satisfy our sampling criteria.) Again, we see that the bulk of the sample was contained within Georgia and Alabama, particularly inland of the fall line. Nevertheless, we do not see significant differences across subsamples in the propensity to be in either of these states. However, a chi-squared test overwhelmingly rejects the equality of the distribution of the subsamples across counties. Below, we show that one aspect of this difference is increased propensity of lottery winners to be in a county whose land was opened up by the 1832 Cherokee Land Lottery (shaded in blue in the map). Throughout the paper, we refer to such counties as old Cherokee County.

---

[12] Plainly, this is not a global total; there are other components of wealth that we cannot measure, such as non-slave personal property (which was only reported in the 1860 and 1870 censuses) and the individual's human wealth.

5 Estimated Change in the Wealth Distribution

In this section, we characterize the difference in the control versus treatment distributions of 1850 wealth using a variety of estimators. In Section 5.1 we define a simple regression equation that forms the basis of our empirical analysis. In Section 5.2, we show that the treatment group of lottery winners had, almost two decades after the lottery, higher mean wealth than the control group of lottery non-winners. This result is robust to a variety of controls derived from the characteristics of surnames and given names. However, results from quantile regressions show that the effect of the lottery on the treatment group is concentrated in the upper part of the wealth distribution. Then, in Section 5.3, we present estimates of the PDF for control and treatment groups, as well as estimates of the difference in the CDF between the two groups. Again, we show that the treatment associated with lottery winnings perturbs the distribution of wealth primarily in the upper half of the distribution. Using both the quantile and ΔCDF estimators, we find very little effect of treatment on the lower 40% of the wealth distribution. (To be clear, we are thinking of the distribution itself as an object that is being treated. None of our results in this section is meant to imply anything about the mapping from control to treatment, in the sense of characterizing the precise relationship between potential outcomes.) Next, in Section 5.4, we evaluate the gains from treatment (relative to control) under various preferences for distributional equity. Finally, in Section 5.5, we conduct a placebo exercise using a sample defined by having children born within the pre-lottery window, but within South Carolina instead of Georgia. Matches to the Smith list in this case are entirely spurious, and, accordingly, this placebo variable does not predict differences in wealth between the control and treatment groups.

5.1 Estimation strategy

The basic research design of the study is to compare the long-run outcomes of winners and losers among participants in the 1832 Cherokee Land Lottery. Above we discussed how we assign lottery status (winning vs. losing) in a sample of men who, by their characteristics, were eligible to participate in the lottery. With such a sample, estimating the treatment effect of winning the lottery is as simple as a comparison of means across the subsamples of winners and losers or, equivalently, a bivariate regression with the outcome on the left-hand side and a dummy variable for winning the lottery on the right-hand side. Throughout the present study, we opt for the regression approach, which is able to accommodate additional control variables on the right-hand side as well as the $1/n$ measure of lottery status, which is not dichotomous. At some level, the random nature of the lottery should obviate the need for control variables as fixes for omitted-variable problems. Nevertheless, controls might be useful to absorb some of the residual variation and perhaps improve the precision of the treatment estimates. Further, the methods that we use for tracking the lottery-eligible sample and imputing lottery status might introduce biases that control variables could clean up. (The fact that lottery status is not predictive of predetermined variables, as seen in Section 3 and Table 1, casts doubt on this supposition, but we can never rule it out entirely. We return to this issue in Section 5.5 with an alternative placebo test.)

The basic regression equation, which we generally estimate using OLS, is as follows:

$$Y_{ik} = \gamma T_i + BX_{ik} + \delta_a + \delta_k + \epsilon_{ik} \tag{1}$$

in which i, a, and k index individuals, ages, and 1850 counties of residence. The variable of interest, T_i, is a binary variable that denotes treatment—meaning winning the lottery—and the control variables are as follows: δ_a is a set of dummies for age; δ_k is a set of dummies for location (county×state k), which we include to account for differences in settlement patterns in the control and treatment groups; and X_{ik} is a vector of other control variables, as specified below. The random assignment of treatment by the lottery allows us to recover an unbiased estimate of γ.

A principal alternative specification used below also incorporates characteristics of the surname (last name). The main variant of the specification includes fixed effects at the surname level. The specification controls for a host of differences that might persist across lines of patrilineal descent. One way of thinking about the specification is measuring the impact of lottery winning within extended families (again, defined patrilineally). Recent work by Clark and Cummins (2012) and Guell *et al.* (2011) highlights the persistence in outcomes across patrilineal lines, and this effect would be absorbed by surname fixed effects. Furthermore, specification problems that are introduced by our use of surname in constructing the lottery variables would also be absorbed by these fixed effects. (As we discussed above, we use the Soundex version of names to account for minor spelling differences.) Note that this is a stronger test to pass in that we effectively ignore individuals whose surnames are unusual enough that the sample does not contain both a winner and loser with that surname.

5.2 Baseline regression results

We estimate a large effect on 1850 wealth from having won the lottery almost two decades earlier. Table 2 presents the estimates of equation (1) with total wealth (the sum of real estate and slave wealth) as the dependent variable, and results are shown for both levels and natural logs. The baseline estimates are found in Column 1. On average, lottery winners have approximately \$750 or 14% more wealth in 1850. This number is similar in magnitude to the unconditional difference seen in Table 1. It is also similar to, perhaps a bit smaller than, the back-of-the-envelope estimate of the value of land won in 1832. It is possible that the winnings were partially spent or saved in some other kind of wealth, although there was a relatively limited set of assets that could be used to store value in the rural Deep South at this time, and we are measuring two of the most important components (land and slaves).

In any event, the baseline estimates suggest substantial persistence. The remaining columns of Table 2 reports specifications that use different sets of fixed effects as controls. In Columns 2-4, we control for characteristics of the surname: the first letter, the number of letters, and the frequency of that surname in our sample. These estimates are within a third of a standard error of the baseline. In Column 5, we report specifications that include a full set of dummies for each surname (using the Soundex concept, as discussed above). Estimates drop by about half the standard error in this case, but we still estimate that lottery winners were almost \$600 richer 18 years after the lottery. In Column 6, we control for a full set of dummies for given (first) names rather than surnames, and we see that the estimates instead rise by about half a standard error relative to the baseline. Finally, in Column 7, we include fixed effects both for given name and for surname. (Note that these are two sets of fixed effects; fixed effects for each given-name-x-surname cell would absorb the lottery-status variable, which uses the full name for linkage to the Smith list.) These estimates are a bit below the baseline, but a bit above the estimates that we obtain when controlling for surname alone.

Table 3 continues the analysis of lottery status and wealth by presenting specifications with alternative ways of constructing the wealth variable. Panel A presents results for total wealth in levels or logs. Estimates from the baseline specification are repeated here for reference. Also in this panel, we attempt to adjust this variable for the truncation of the lower tail. Specifically, census enumerators were instructed to leave real estate wealth blank if the value was under \$100. It is common in studies of variables that are censored or truncated like this to impute a value of zero in levels and in logs (=ln(\$1)). In the previous analysis, we assume the blanks were zeros in levels and missing values in logs. It is difficult to check these assumptions, but they seem ad hoc. An inspection of the distribution of real estate wealth reveals that the truncation at \$100 is important: there is a nontrivial amount of density at and just above \$100. Furthermore, the distribution looks approximately log-normal above \$100. If we fit a truncated normal to the distribution above \$100, we estimate that the expected value of wealth below \$100 is approximately \$59.34. We use this number to impute wealth to those whose real estate wealth is below \$100 and rerun the regressions from above. As can be seen in Panel A, this adjustment for truncation of the lower tail results in trivial differences in the estimated coefficient on lottery winning. While this adjustment for truncation is also imperfect, the fact that the results change so minutely when moving around the lower-tail imputations by so much suggests that lottery status has very little impact on the lower tail of wealth. We test this directly in the next panel.

Panel B of Table 3 presents the results of quantile regressions[13] that allow us to explore the effect of winning the lottery on wealth at various points in the wealth distribution. We estimate very little effect of winning on wealth in the lower tail, seen in the first row of the panel where we estimate the treatment effect at the 25th percentile of wealth. (Note that the person at the 25th percentile of the wealth distribution in the sample has zero wealth.) In contrast, we estimate an effect of approximately \$200 at the median and over \$500 midway into the upper tail. We see even larger differences in wealth at the 95th percentile, although this result is only statistically significant for one of the two specifications. At such high levels of wealth, it is likely that any treatment effect of winning the lottery is overwhelmed by noise (be it statistical noise or variations in fortune/endowments/etc.), especially if the noise grows in magnitude as wealth increases even as the dollar value of winnings does not.

This pattern of results across the distribution is also shown in Figure 1, which presents quantile-regression estimates of lottery winning across the distribution of total wealth in 1850. The points are the quantile-specific estimates of the treatment effect, and the dashed

---

[13] It was not computationally feasible to estimate the quantile regressions with large sets of dummy variables, so the results reported here are from bivariate quantile regressions.

line is a local-polynomial-smoothed mean of these estimates. Here we use the 'unique match to Smith' definition of lottery winning. (Appendix Figure 2 displays analogous results using the $1/n$ match instead.) Again, we see that shifts in the distribution are quite small in the lower tail, become larger in the middle, and then grow quite large in the upper tail. (We omit the display of quantiles above 0.985, where estimates are larger still, so as to not obscure the shape of the curve for the vast majority of the distribution.) Note that, while the average coefficient is \$525, the gains are quite concentrated in the upper third of the distribution.

These results are, on their face, inconsistent with the simple hypothesis that the random disbursement of a fixed amount of wealth shifts the distribution equally at all points. For certain, there was variance in the value of lots won, but the random nature of the lottery insures us that both the variance and the expected value would have been independent of a winner's counterfactual point in the control distribution. Thus, if all of the winnings were at least positive, then the lottery should have to some degree drained mass from the lower tail of the distribution, relative to control. In any case, we cannot interpret these estimates as the treatment effects at a given point in the control distribution, unless the mapping from control to treatment (which is inherently unobservable) preserves the relative rank of each observation in the outcome distributions. Absent this rank-invariance property, the interpretation of quantile-regression estimates is somewhat awkward to render in words, so we return to this issue with graphical presentations of the differences in the distributional function in Section 5.3 below.

In Table 3, Panels C and D, we consider the subcomponents of measured wealth: real-estate wealth and wealth held in the form of slaves. First, consider the intensive margin. We estimate positive treatment effects of winning the lottery for both categories of wealth, with a somewhat higher coefficient on slave wealth. The estimate for real estate is about half the median real estate wealth, while the estimate for slaveholding is considerably larger than the median (of zero) in that category of wealth. Second, consider the extensive margin of wealth. We estimate essentially no effect of winning the lottery on holding real estate valued at least \$100 (the truncation point imposed by the census as discussed above). In contrast, lottery winners are four to five percent more likely to own at least one slave.

Finally, estimates are similar when using either the binary or the $1/n$ match variables to impute lottery winning. Throughout the rest of the study, we present only the binary variable to save space. Nevertheless, it does seem from the results in Table 2 that the specification including controls for surname fixed effects is a bit more conservative, and so we present that specification as an alternative to the baseline in the tables to follow.

5.3 Comparison of Density Functions

In this section, we compare the probability density functions and cumulative density functions (PDF and CDF, respectively) of the control and treatment groups. These results are shown graphically in Figure 2. To construct these graphs, we use the same sample of lottery-eligible household heads as above and we define "treated" to be the binary variable indicating a unique match to Smith (1838). The y axes in Figure 2 denote density (or probability) and the x axes measure the natural log of total wealth (displayed in thousands of dollars for legibility).

Relative to control, the empirical PDF of the treated group appears to be missing mass in the middle of the distribution and have extra mass in the upper tail. Panel A of Figure 2 shows the estimated PDFs of both the control and treatment groups (dashed and solid lines, respectively).[14] The vertical line at 0.1 (that is, \$100) denotes a level below which some enumerators censored real estate wealth in the 1850 Census. The control PDF is approximately log-normal and roughly similar to the treatment distribution below a few hundred dollars. Between roughly \$300 and \$2500, however, the treatment PDF is noticeably lower than that of the control. Above \$2500, this pattern is reversed, with control below treatment. (As mentioned in Section 4, the two distributions are significantly different from one another in a Kolmogorov-Smirnov test.)

The CDF for the treated is shifted up relative to control in a statistically significant manner, but only for wealth between approximately two and ten thousand dollars. This result is seen in Figure 2, Panel B, which displays the estimated differences in CDFs between the two groups at various points. We implement this estimator by constructing an indicator variable for wealth being below a given $\bar{w}$: $d_i^{\bar{w}} \equiv I(w_i \leq \bar{w})$. We then regress $d_i^{\bar{w}}$ on the treatment dummy, with controls as in equation 1. Sweeping the $\bar{w}$ threshold across the distribution of w_i, we recover a treatment effect and confidence interval that estimates the difference in CDFs across a series of wealth levels. (Note that shifting of the treatment PDF

---

[14] These densities are estimated using a kernel estimator in stata ("kdensity") with a Epanechnikov kernel and stata's calculation for the optimal bandwidth. We omit those observations with zero wealth rather than using the imputation. Otherwise the assumption of smoothness would be violated for the kernel density estimator. The question of the extensive margin of wealth was treated above in Section 5.2.

will generate negative coefficients in this procedure, by the definition of the CDF.) These estimates are shown in Panel B, with the solid gray line being the point estimate and the dotted lines describing the 95% confidence interval. A solid black horizontal line is drawn at zero for reference. As can be seen, the lower half (or more) of the CDFs are statistically indistinguishable, as are the extreme upper tails. Nevertheless, there is a range of wealth values, in the several thousands but not the tens of thousands, for which the treatment CDF has significantly more mass at higher wealth than has the control CDF. (These results are similar if we use alternative specifications of the linear-probability model. See Appendix Figure 3.)

5.4 Incorporating Tastes for Equity

The magnitude of the improvement for the treated group depends on one's taste for equity (i.e., distaste for inequality). There is greater inequality in the treatment than control distributions, as measured by the coefficient of variation, the standard deviation of logs, and the Gini coefficient, as well as with the Mehran, Piesch, Kakwani, Theil entropy, and Theil mean log deviation measures. Furthermore, a statistical test that the standard deviation of log total wealth (with the imputation adjustment) is greater for treatment than control (versus the null of equality) has a p value of 0.0321. But how should we weighed this increase in inequality (a presumed cost) with the higher average wealth (an obvious benefit)?

A simple and standard way to vary the equity weight when considering a group's distribution of outcomes is to use an aggregator with a constant elasticity of substitution (CES):

$$\bar{U}_j = \left[\sum_{i \in Z_j} \frac{1}{N_j} w_i^{\frac{\rho-1}{\rho}} \right]^{\frac{\rho}{\rho-1}} \tag{2}$$

in which i indexes individuals in our sample, j is an indicator for treatment status ($j = 0, 1$) Z_j is the set of indices i belonging to group j, w_i is i's 1850 total wealth, and ρ is the elasticity of substitution, which relates to the preference for equity in a manner described below. The question we ask is whether, for a given ρ, can we reject the equality of the CES aggregator between the two groups or, more formally, $H_0(\rho) : [\bar{U}_0 = \bar{U}_1]$?

Figure 3 shows the relationship between the taste for equity, ρ, and the ratio of the $\bar{U}$ for the treated divided by that of the controls. Importantly, the graph also displays the

95% confidence interval of this ratio, so we can see whether the two groups are statistically distinguishable under varying assumptions for ρ. These statistics are computed with 5000 bootstrapped samples for each ρ. The sample is the same as in Figure 2, Panel B, and the computations use the untruncated level[15] of wealth.

For a sufficiently high ρ we can reject the null hypothesis of equivalence between the two groups. As $\rho \longrightarrow \infty$, $\bar{U}_j$ becomes simply the arithmetic mean for each group. In this limit, a dollar held by any given person becomes a perfect substitute, in social-welfare terms, for a dollar held by someone else. Thus, the taste for equity disappears as ρ gets larger, and social welfare eventually depends only on the average outcome. Average wealth is indeed higher in the treatment group, as seen above. Therefore, as we see in Figure 3, we can reject the null hypothesis of identical $\bar{U}$ for very large ρ. As $\rho \longrightarrow 1$, the $\bar{U}_j$ aggregator becomes the Cobb-Douglas function, with each individual's wealth as its arguments. The H_0 is rejected for this case, which is equivalent to the test above for the natural log of wealth.

Nevertheless, we cannot reject the equality of the two (re-weighted) wealth distributions if we place a greater weight on equity. Specifically, the control and treatment distributions are only statistically distinguishable if we use a ρ much less than one. This is seen in Figure 3, where the confidence intervals overlap with zero for ρ below and including 0.95. This number is very far from, for example, John Rawls' (1971) maximin preference, an elasticity of zero in which social welfare depends exclusively on the well-being of the least-well-off individual.

5.5 Placebo test using South Carolina

We perform a falsification exercise using South Carolina rather than Georgia and do not find statistically significant results. One of the challenges in identifying the treatment effect associated with winning the 1832 Lottery is that our method of imputing lottery status via name matching may introduce biases through sample selection. To check for this possibility, we construct a placebo sample using households with only children born in South Carolina (rather than Georgia) during the same pre-lottery window (the three years prior to the Cherokee Land Lottery of 1832).[16] We use the names among this South Carolina sample to impute lottery status per the Smith (1838) list. As above, we use both a dummy for a unique

---

[15] Using natural logarithms would be an inappropriate starting point tantamount to imposing a maximum $\rho = 1$.

[16] We are grateful to Petra Moser for suggesting this procedure.

match to the Smith list and a variable that allows for probabilistic matches, deflated to $1/n$ in case of ties. By the eligibility rules of the Cherokee Land Lottery, any matches to this list from the placebo sample must be spurious. It is then reassuring that the fraction of unique matches in the placebo sample derived from South Carolina is only one quarter of the fraction in the Georgia sample. In Table 4, we estimate Equation 1 using this placebo sample, for the different variables indicating lottery status, and using both the basic specification and the one that includes surname/Soundex fixed effects. These results are found in Panels A and B, with analogous results from the Georgia sample provided for reference in Panel C. The first four columns of Table 4 show outcomes that were determined prior to the 1832 Lottery, and there are no statistically significant results. (Note that a series of falsifications checks using pre-lottery variables was also performed for the Georgia sample, as shown in Table 1, Panel B.) The remaining columns show post-lottery outcomes in 1850 such as residing in old Cherokee County and real-estate wealth. There is no statistically significant pseudo-treatment effect for the South Carolina sample, in contrast to what we find for Georgia.

6 Discussion of Mechanisms

What are some mechanisms that might have prevented these wealth transfers from impacting the lower tail of the 1850 wealth distribution? We consider several possibilities in this section. We consider, and find lacking, explanations based on fixed costs, interactions with ability, or life-cycle de-cumulation of wealth. Instead, greater vulnerability to risk, particularly at the low end of the (counterfactual) distribution, emerges as a leading candidate for such a mechanism.

6.1 Fixed Costs

It is implausible that a fixed cost of production, in and of itself, caused the long-run insensitivity to treatment of the lower tail of the wealth distribution. We say this not because we doubt the existence of fixed costs (we do not), but rather because it is inconsistent with what we know about farming in this period. Without doubt there were farms (plantations, really) that had large fixed costs of start-up and of operation. But a prospective farmer who could not produce such a large initial investment would have had the option to farm at a lower scale, perhaps with a different crop and/or in a different area.

Within our data itself, it is difficult to square the results with a large fixed cost of production. Differences in the PDFs do not emerge until over \$400 and the CDFs do not significantly diverge until over \$2000 in wealth. In contrast, we see farmers in our own data working with much lower levels of wealth. (See Figure 2, Panel A.) The fact that the control wealth distribution is approximately log-normal, even in this low-wealth range, suggests that we are observing something close to the steady-state level of operation rather than some transitory range that farms pass through on the way to either closing or dramatically expanding. In any event, the parcels won in the 1832 Cherokee Land Lottery were roughly 160 acres, which would have been more than enough for small-scale farming, except on the lowest-quality land. Even if 160 acres was too small for some purposes, lottery winners could have simply sold their land in Old Cherokee County (as many did) and purchased a larger farm with cheaper land farther west on the frontier.

While slaves were manifestly difficult to purchase in non-integral units, this nonconvexity need not have been an impediment to productivity using the lottery winnings. First, purchasing a prime-age, male fieldhand would have cost several hundred rather than several thousand dollars in 1850. Second, there were substitutes for purchasing slave labor, such as hired labor or less slave-intensive crops. Furthermore, while some areas of the South were dominated by slave-based agriculture, other were not. Indeed, Old Cherokee County itself was in the Upcountry where smaller yeomen farms were commonplace.

Land improvement was itself an up-front cost, but did not need to be either large or lumpy. Initial land clearance was often done in a low-cost, slash-and-burn fashion. In any case, improving land could be done incrementally, perhaps hiring oneself out as labor until at least enough land was cleared for subsistence.[17]

6.2 Interactions and Human Capital

One possible mechanism is that wealth shocks alone are insufficient, but rather they must be paired with some complementary skill. Indeed, those at the bottom of the distribution (either control or treatment) may have been there precisely because they lack the ability to seize opportunities. If so, the bottom parts of the two distributions might look similar because the low-skilled winners could not take advantage of their winnings.

---

[17] We are grateful to Tim Guinane for pointing this out.

6.2.1 Did the would-be poor not claim their winnings?

Failure to claim lottery winnings cannot account for the lower tails of the treatment group being similar to control. Evidence supporting this claim is seen in Figure 4, which displays estimates of the fraction of parcels claimed prior to 1838 (as reported in Smith, 1838) versus total 1850 wealth (adjusted for truncation at the lower tail) for the subsample of lottery winners only. (By definition, lottery losers were not assigned parcels, so parcel characteristics are unavailable for the full sample.) The solid line displays a local-polynomial smoothed estimate of the mean claim rate for each level of total wealth, and the short-dashed lines denote 95% confidence intervals. For reference, the long-dashed line presents the PDF for log total wealth.

Across the distribution of 1850 wealth in the treated group, there was a relatively stable rate of claiming land, as shown in Figure 4. Claiming rates are, if anything, a bit higher in the lower tail of the 1850 wealth distribution. In any event, note the range of the curve: within 5 percentage points of one half for almost all of the distribution. If less than 50% failed to claim their winnings by the end of 1837,[18] this will attenuate the effect of treatment, but it cannot explain the apparent 100% markdown of winnings in the lower tail of treatment. Therefore the lottery winners who ended up poor did not do so because they all failed to collect their winnings at the outset.

6.2.2 Surname-average Characteristics

In this subsection, we consider whether the response to the shock of winning the lottery is related to characteristics of other people who share the same surname and are therefore likely related along patrilineal lines of descent.

To assess this possibility, we construct surname-specific averages of wealth, fertility, literacy, and school attendance as possible proxies for differences across extended families in either preferences or prices. We used the 1850 100% census file to construct the average fertility, school attendance, and real-estate wealth among Georgia-resident households for each (Soundex) surname. Those individuals that appear in our lottery-eligible sample are excluded from the construction of the averages. We first check for the statistical power of

---

[18] Eventual claiming rates are probably even higher because the Georgia state government processed claims for several more years. Banks (1905) estimates that approximately a quarter of the land lots opened by lottery went unclaimed.

these proxies by regressing the individual-level outcome on the surname average:

$$Y_{ijks} = \alpha Y'_s + \delta_a + \delta_k + \epsilon_{ijk} \tag{3}$$

where s denotes the surname for each observation, Y'_s is the surname-average of the Y' variable, and each regression contains dummies for age and for state/county of residence. (The 'prime' on Y' allows for the possibility of a different Y variable's average on the right-hand side of the equation.) Furthermore, recognizing the group-level regressor, we adjust the standard errors for clustering at the surname level. The base sample for these regressions is the same as for analogous estimates of Equation 1 displayed in earlier tables, with the exception that some households are omitted if there were no other households in Georgia with the same surname and therefore no one with which to form the surname-level averages. We consider three surname-averaged wealth outcomes: the level of 1850 total wealth, its natural log (both adjusted for truncation in the lower tail), and a dummy variable if total wealth exceeds $5000. The baseline estimate for the treatment dummy is shown in Column 1 of Table 5.

The estimates indicate that the surname-averaged variable is indeed a strong and statistically significant predictor of individual wealth, although results are weaker for surname-average fertility. Estimates of equation 3 are found in even-numbered columns of Table 5. The coefficient of zero is rejected in most cases for conventional confidence intervals. A mechanistic model in which the patrilineal dynasty (this proxied by surname) predicts outcomes one-for-one is even more strongly rejected, however; the coefficients are closer to 1/4th or 1/8th. See, for example, Panel A, Column 2, or Panel B, Columns 4 and 6, for apples-to-apples comparisons. (Note that we do not argue that this is a causal effect of the behavior of their relatives on the individuals' choices, but rather a proxy for some shifter that is common within the group. Thus the interaction term should be interpreted as interacting with the shifter as well.) Surname-average fertility is a weaker predictor of wealth (Column 8). Nevertheless, men have statistically and economically higher wealth if their surnames are associated with higher rates of school enrollment among children aged 5-15 or of literacy among adult men (Columns 10 and 12, respectively). In sum, these surname-level measures are generally strong predictors of own wealth, and thus may be suitable predetermined proxies of own human capital. (Recall from Table 1, Panel B, that lottery status did not predict these characteristics.)

The odd-numbered Columns 3–13 of Table 5 report results in which we interact the surname-average characteristic with winning the lottery. The specific estimating equation is as follows:

$$Y_{ijks} = \gamma T_j + \beta T_j z(Y_s') + \alpha Y_s' + \delta_a + \delta_k + \epsilon_{ijk} \tag{4}$$

Note that this is a modified version of Equation 1, to which we add the interaction of the treatment variable (T_j) with the z-score of the surname-average variable ($z(Y_s')$). The main effect of the patrilineal dynasty loads on to $\alpha Y_s'$, and standard errors are clustered at the surname level. We report the coefficients on treatment, the surname average, and the interaction of treatment with the surname average. The estimated coefficients on being a lottery winner throughout the table are similar to those reported above, and the coefficients on the surname averages are similar to those in the even-numbered columns. The estimated interaction terms are generally of the expected sign but not statistically significant. For example, in Panel B, Column 7, we display estimates in which the outcome variable is the natural log of total wealth and the surname-level variable used to form the interaction with treatment is the natural log of the median total wealth among others adult males in Georgia with the same surname. The coefficient of .009 implies an additional .036 of log total wealth from treatment as we sweep across four standard deviations of the distribution of the surname-average variable. The difference is approximately one third of the main effect of treatment. However, this interaction coefficient is not significantly different from zero. Indeed, only one of the 18 interactions with surname averages yields a significant (at the 10% level) estimate, although 13 of 18 are of the expected sign.

6.2.3 Own Illiteracy

We do not find a significant interaction between treatment and the lottery winner's own illiteracy. While this could in theory have been affected by the lottery, it is unlikely that many men would have become literate during adulthood, whether they won in the lottery or not. Nor does lottery winning predict illiteracy, as we saw in Table 1. With these facts, we take license to use the lottery eligible's illiteracy (defined as unable to read or write) as a predetermined (prior to the lottery, that is) variable. These estimates should be taken *cum grano salis* in case this assumption is incorrect. In Table 5, Column 14, we show that illiterate men indeed had substantially lower wealth by all three measures. However, estimates for the interaction term (shown in Column 15) are not significantly different from zero.

All told, evidence in support of both a statistically and an economically important interaction between treatment and human capital is weak, at least with the proxies of human capital at our disposal.

6.3 Life-Cycle and Family Considerations

The results above are not likely because of a life-cycle-related wealth decumulation (be it for spending or bequests). First, note that *inter vivos* transfers to current members of the same household are already included in measure of wealth used above. Second, the age distribution of and life-profile of wealth do not favor this hypothesis of decumulation. The distribution of age and the average wealth by age, in our sample, are shown in Appendix Figure 4. Wealth accumulation peaks around 60 years, and over 80% of our sample is below this age. In any event, it is probable that some would have maintained their gross asset position (which is what we measure above) into old age to use as collateral against a mortgage or as a kind of social collateral to keep their children around waiting for bequests. We argue, therefore, that the similarities of the lower tail of the control and treatment wealth distributions are not because people had spent or bequeathed all of their wealth for standard life-cycle motives.

We argue that changes in fertility are not an important mechanism in explaining the pattern of results across the distribution. Using a similar research design, Bleakley and Ferrie (2012) found that lottery winners had a higher fertility rate (approximately 0.1 more children by 1850, significant at conventional levels of confidence) and that the effect on fertility[19] is strongest in the lower half of the wealth distribution. This result is shown in Appendix Figure 5, Panel A, where the control and treatment means show a gap of roughly 0.1 over approximately the same range for which the ΔCDF results in Figure 2, Panel B were insignificantly different from zero. Nevertheless, this coincidence does not imply that the pattern of excess fertility among lottery winners explains the pattern of the wealth results. First, if the average number of additional children is approximately 0.1, it implies that the vast majority of the treated did not have additional children, and therefore had

---

[19] Note that same study, focused on the quantity and quality effects of wealth, reports that lottery winners did not buy the increase in quantity with a decrease in quality. The children of lottery winners did not go to school more in 1850, nor did they hold more wealth in 1870, nor have higher-paid occupations in 1880.

no reason to lose their extra wealth because of extra fertility. Second, even if we ignore the integral nature of childbearing, comparing 0.1 extra children with the approximate value of winnings of \$500-\$1000 would imply a price of childbearing on the order of \$5000 of wealth per child. This is implausibly large.

Finally, we do not find locational choice as an important consideration in explaining differences in wealth between control and treatment. First, note that the lottery winners were slightly more likely to be in old Cherokee County in 1850, by approximately 2 percentage points. But apart from this, lottery winnings did not appear to bring them to a place that had peculiar characteristics, at least across a wide range of observables from aggregate county data, including fertility, schooling, farm values, farm sizes, land improvement, slave density, urbanization, or access to transport. (These results are found in Appendix Table 1.) In any event, we obtain fairly similar results for wealth whether or not we condition on state×county dummies; therefore, the change across the wealth distribution is not because some part of the treatment group happened to pick counties that saw faster appreciation of land values, for example. Note that, as seen in Appendix Figure 5, Panel B, 1850 residence in old Cherokee County by the treated seems to be strongest at the lower/middle part of the wealth distribution. This is consistent with claiming rates being slightly higher in the lower part of the distribution, as seen above. Generally, this pattern makes sense in that much of the work of actually going out to improve land would have been labor-intensive, and therefore a task disproportionately taken up by those with a low opportunity cost of time.

6.4 Risk and Churn

The remaining explanation, which admittedly is something of residual category, is exposure to risk. Figure 6, which displays estimates of the fraction holding no more than \$100 in 1860 as a function of 1850 total wealth. (The structure of this figure is therefore similar to that of Figure 5.) The dependent variable (on the y axis) is a dummy for whether the 1860 total wealth (personal plus real estate) is less than or equal to \$100 in 1860 dollars. The independent variable (on the x axis) in this figure is total measured wealth in 1850. The sample size for this figure is 5603, because only 40% of the sample is linked to 1860. The dashed line displays a local-polynomial smoothed estimate of the indicated fraction for each level of total wealth, and the grayed area denotes the 95% confidence interval. For reference, the solid gray line presents the PDF for log total wealth (excluding the imputation for those with zero wealth). Even for the very wealthy men, there is a roughly 5% chance that they drop to the lowest measured wealth status over the course of a decade. This probability reaches one-quarter for the very poor. These results highlight the large degree of risk and churn present in that period.

7 Conclusions

The 1832 Cherokee Land Lottery in Georgia represents an unusual environment in which to assess the long-term impact of shocks to wealth on the wealth distribution. Winning should have been uncorrelated with individual characteristics by the random nature of the lottery, and indeed our model passes numerous falsification tests using predetermined variables and in a placebo sample from South Carolina. Further, participation in the lottery was quite broad (albeit limited to adult white men largely) thus ameliorating some of the possible problems with external validity that can arise with self-selected samples.

Using wealth measured in the 1850 Census manuscripts, we follow up on a sample of men eligible to win in the 1832 Cherokee Land Lottery. With these data, we can assess the effect of winning that lottery on the distribution of wealth almost two decades after the fact.

We show that winners are, on average, richer, but mainly because the middle of the distribution is thinner and the upper tail is fatter. In contrast, the lower tail is largely unaffected. This stands in contrast with a 'mechanical' short-run effect of the lottery, which would tend to compress the distribution of (log) wealth. The results are also inconsistent with the view that the effect of winning, while always positive, would have been greatest at the lower tail where the returns to capital were been high because of credit constraints. We argue that these results are not plausibly due to fixed costs or life-cycle considerations. Moreover, there is weak (at best) evidence that difference in ability across the (counterfactual) distribution caused these results. Instead, we suggest that risk, particularly for people with lower wealth, played a role in preventing the pushing the lower tail of the treatment distribution "up from poverty".

8 References

accessgenealogy.com. 2009. "1832 Cherokee Country Georgia Land Lottery". [Accessed May 29, 2009.]

Acemoglu, Daron, and Fabrizio Zilibotti. 1997. "Was Prometheus Unbound by Chance? Risk, Diversification, and Growth," *Journal of Political Economy*, 105(4, August):709-751

Atack, Jeremy. 1988. "Tenants and Yeomen in the Nineteenth Century," *Agricultural History*, 62(3, Summer):6-32.

Banerjee, Abhijit V. and Esther Duflo. 2011. *Poor economics: a radical rethinking of the way to fight global poverty.* New York: Public Affairs.

Banerjee, Abhijit and Andrew Newman. 1993. "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101:(2, April):274-298.

Banks, Enoch Marvin. 1905. *The economics of land tenure in Georgia.* New York: The Columbia University Press, The Macmillan Company.

Bleakley, Hoyt and Joseph Ferrie. 2012. "Shocking Behavior: The Cherokee Land Lottery of 1832 in Georgia and Outcomes Across the Generations," Unpublished manuscript, September.

Bleakley, Hoyt and Joseph Ferrie. 2013. "Land Openings on the Georgia Frontier and the Coase Theorem in the Short- and Long-Run," Unpublished manuscript, March.

Boo, Katherine. 2012. *Behind the beautiful forevers.* New York: Random House.

Cadle, Farris W. (1991). *Georgia Land Surveying History and Law.* Athens, Georgia: Univ. Of Georgia Press.

Clark, Gregory and Neil Cummins. 2012. "What is the True Rate of Social Mobility? Surnames and Social Mobility, England 1800-2012." Unpublished manuscript.

de Mel, Suresh, David McKenzie, Christopher Woodruff, 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 123(4):1329-1372, November.

de Soto, Hernando. 1987. *El Otro Sendero: La Revolución Informal.* Editorial Sudamericana: Buenos Aires.

de Soto, Hernando. 2002. *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.* New York: Basic Books.

Hankins, Scott, Mark Hoekstra, and Paige Marta Skiba. 2010. "The Ticket to Easy Street? The Financial Consequences of Winning the Lottery." Unpublished manuscript.

Imbens, Guido W., Donald B. Rubin and Bruce I. Sacerdote. 2001. "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players," *The American Economic Review*, 91(4):778-794.

Khanna, Tarun. 2007. *Billions of Entrepreneurs: How China and India are Reshaping Their Futures–and Yours.* Boston, Mass.: Harvard Business School Press.

Minnesota Population Center, 2004. *National Historical Geographic Information System: Pre-release Version 0.1*, University of Minnesota, nhgis.org.

Ransom, Roger L., 2005, *The Confederate States of America.* New York : W.W. Norton & Co..

Rawls, John. 1971. *A Theory of Justice.* Cambridge, MA: Belknap Press.

Schultz, Theodore W. 1975. "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature*, 13(3):827-846.

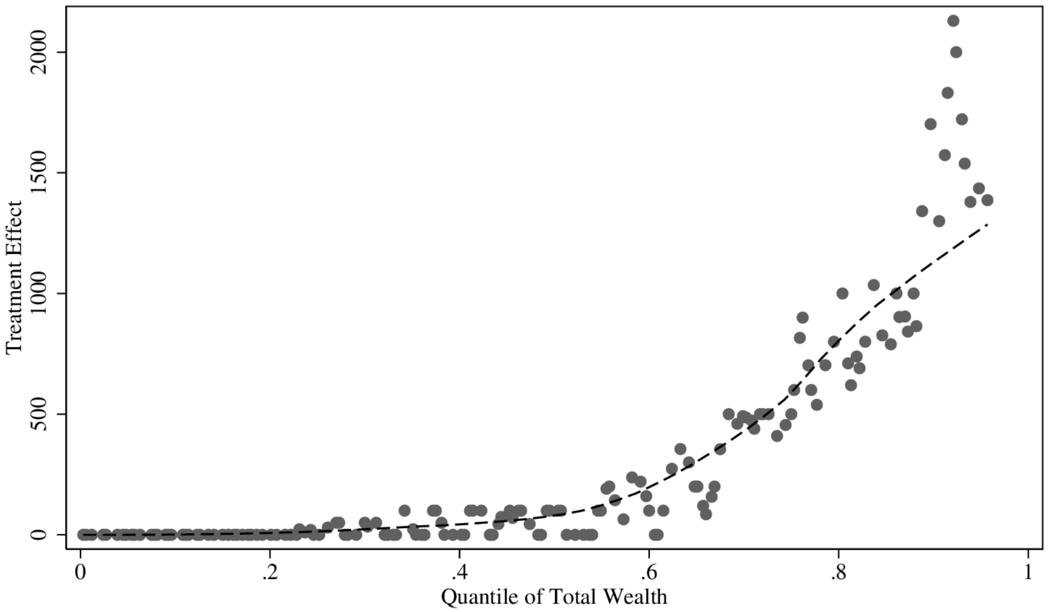
Smith, James F. 1838. *The Cherokee Land Lottery, Containing a Numerical List of the Names of the Fortunate Drawers in Said Lottery, with an Engraved Map of Each District.* New York: Harper and Brothers.

Weiman, David F. (1991). "Peopling the land by lottery? The market in public lands and the regional differentiation of territory on the Georgia frontier." *The Journal of Economic History*, 51(4):835-860.

Wilder, Laura Ingalls. 1971. *The first four years.* New York: Harper & Row.

Williamson, Samuel H. 2013. "Seven Ways to Compute the Relative Value of a U.S. Dollar Amount, 1774 to present," MeasuringWorth.com, accessed March 15, 2013.

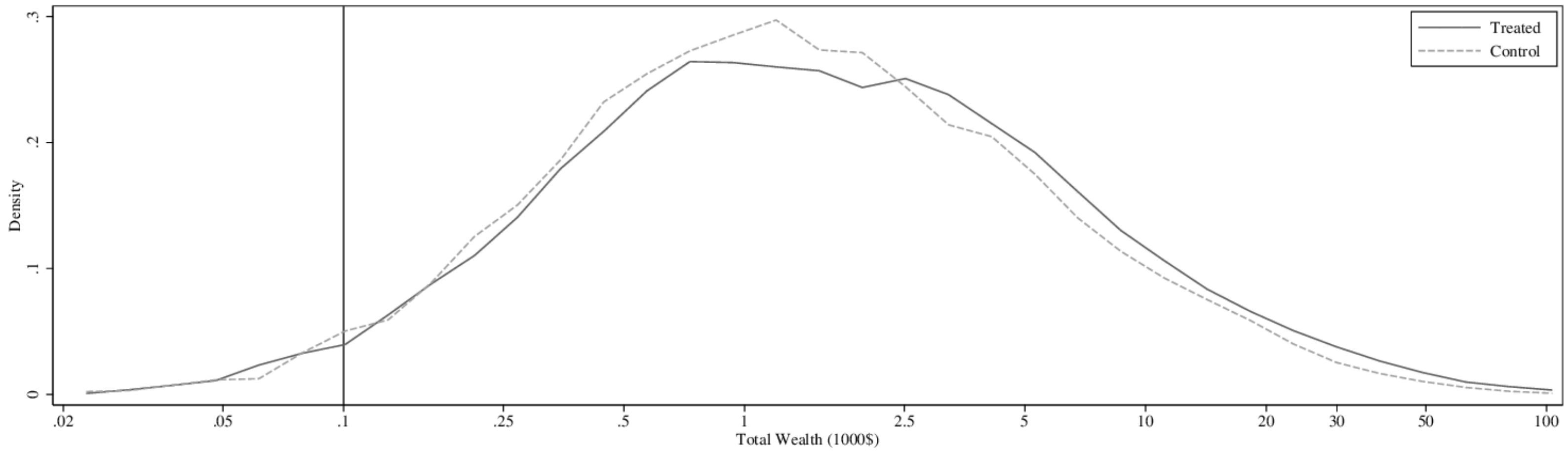
Figure 1: Quantile Regression Estimates on Total Wealth and Lottery Winning



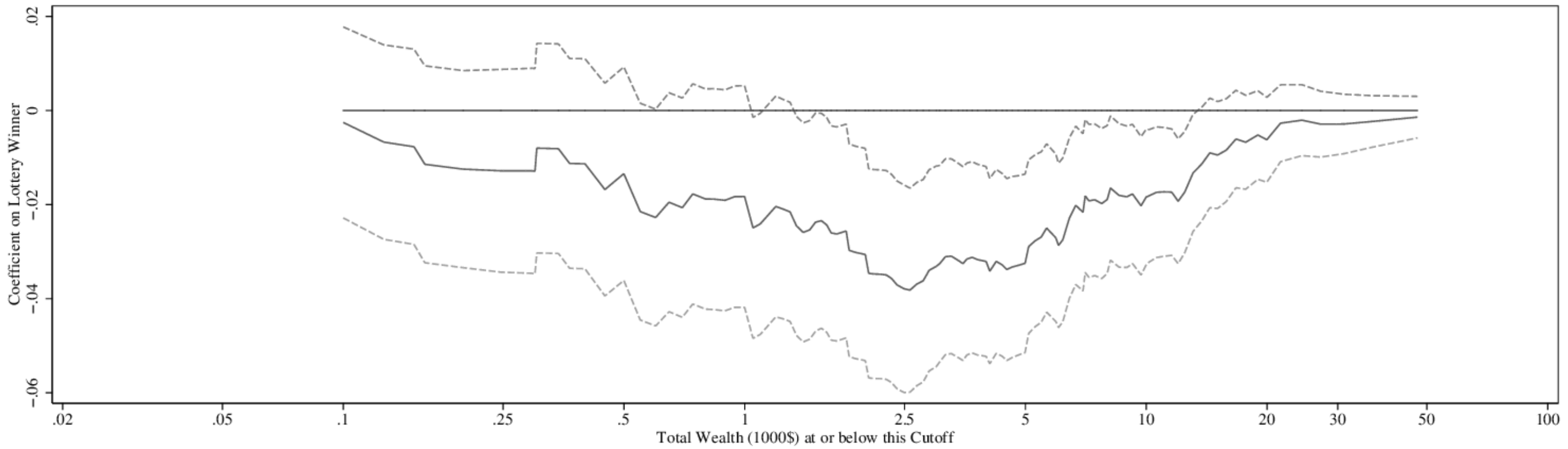
Notes: This figure displays estimates from a quantile regression of the effect of winning the lottery ("treatment") on total wealth in 1850. The points are the quantile-specific estimates of the treatment effect at various quantile points. The dashed line is a local-polynomial-smoothed (Epanechikov kernel, with a bandwidth of .11) estimate of the treatment effect. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. An observation is coded as a lottery winner ("treated") if he is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero ("control"). The dependent variable in this figure is total measured wealth in 1850, the sum of real-estate wealth and slave holdings. Real-estate wealth is as reported on and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. The sample size for this figure is 13094. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Figure 2: Differences in the Distribution of Total Wealth Between Lottery Winners and Losers

Panel A: Estimated Probability Distribution Functions

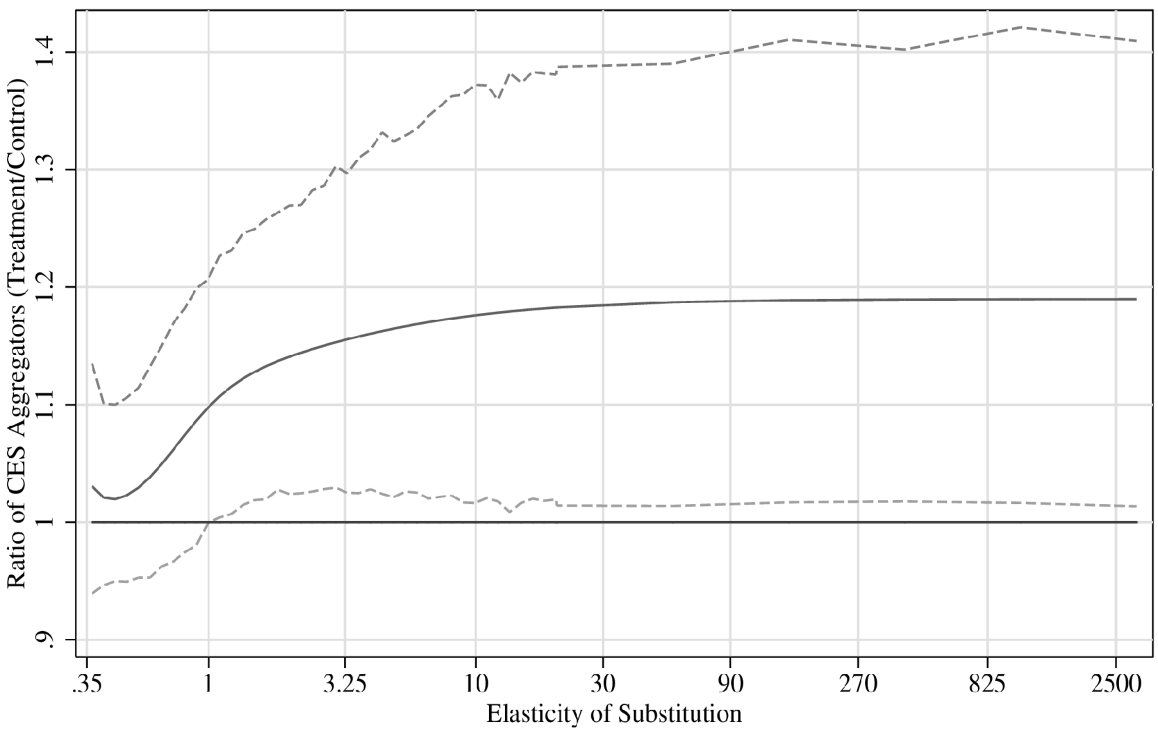


Panel B: Estimated Differences in the Cumulative Distribution Functions



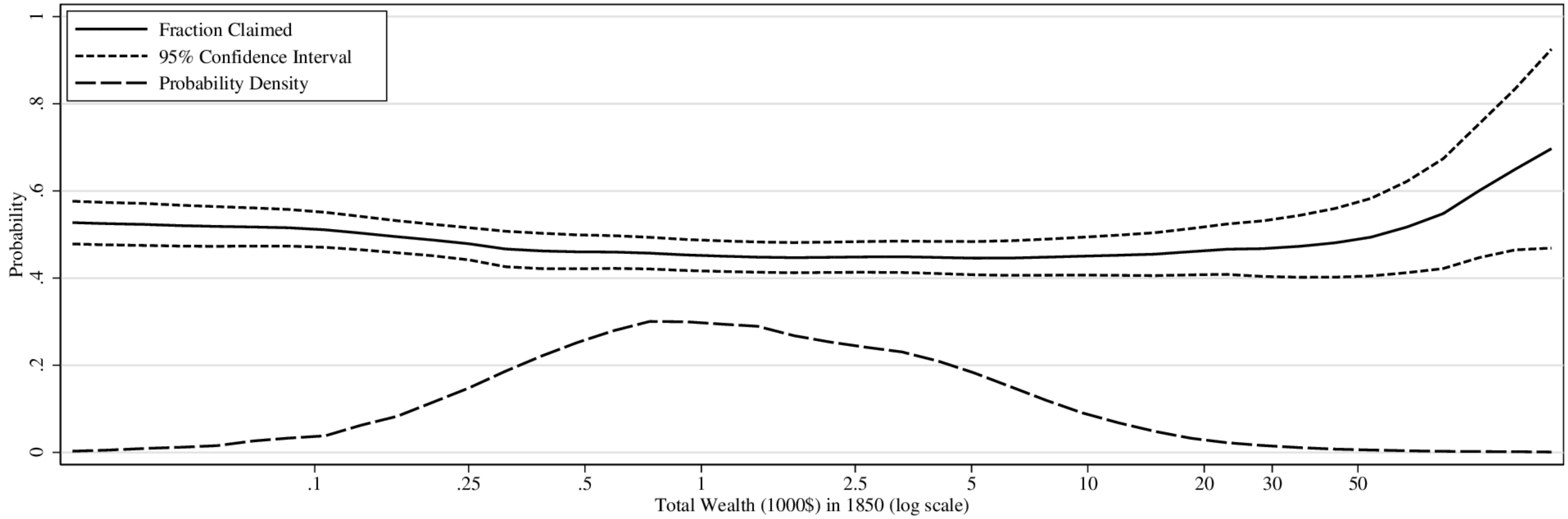
Notes: This figure displays estimates of the distribution, decomposed by lottery status, of total 1850 wealth for sample of lottery-eligible men. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. An observation is coded as a lottery winner ("treated") if he is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero ("control"). The dependent variable in this figure is total measured wealth in 1850, the sum of real-estate wealth and slave holdings. Real-estate wealth is as reported on and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. The sample size for this figure is 13094. Panel A presents the probability distribution functions, estimated using the "kdensity" command in stata. The vertical line denotes $100, the level below which some enumerators censored real-estate wealth. The solid line in Panel B presents the differences in the cumulative distribution function between groups (treatment minus control), estimated using a linear probability model (equation 1) of an indicator for being below 200 quantile cut-points for wealth. Heteroskedasticity-robust 95% confidence intervals are plotted with the dashed lines. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Figure 3: Total-Wealth Differences, Lottery Winners versus Losers, Under Various Tastes for Equity



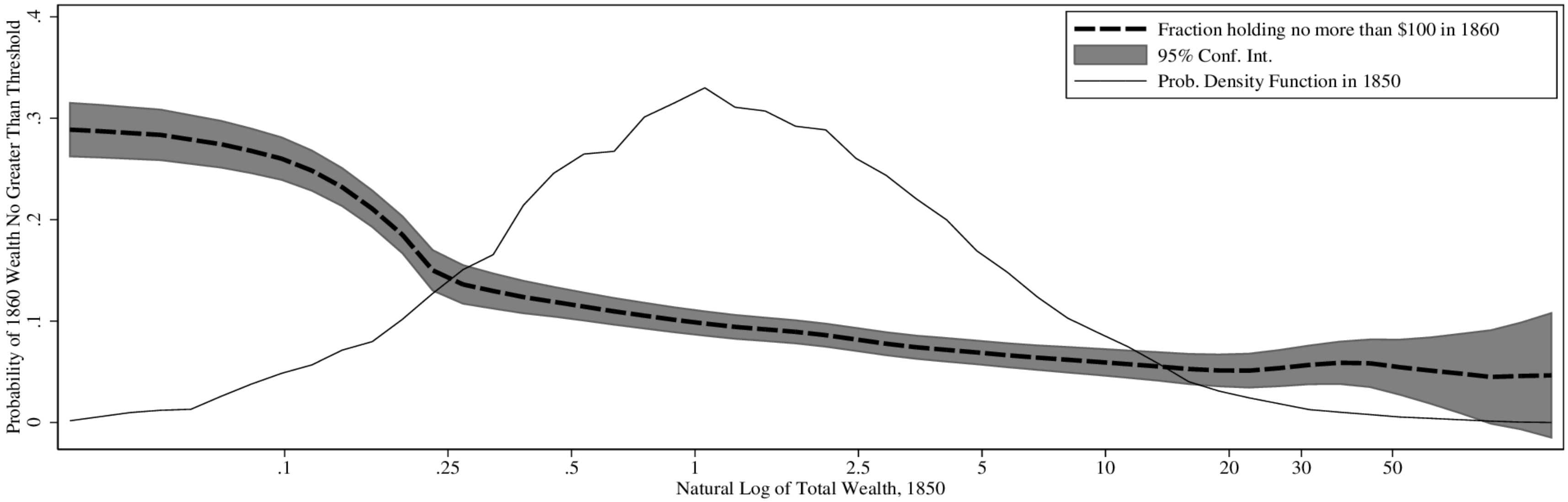
Notes: This figure compares treatment/control differences using a range of preferences for equity over total wealth. The summary statistic for each group is computed using a constant-elasticity-of-substitution (CES) aggregator. The ratio (treatment divided by control) of this statistic is indicated on the y axis. This ratio is computed for various values of the elasticity of substitution (rho), denoted on the x axis. The graph displays the ratio and 95% confidence interval, computed with 5000 bootstrapped samples for each point in the rho grid. The sample is the same as in Figure 1, Panel B. See the notes to Figure 1 for variable and sample definitions.

Figure 4: Probability that Parcel is Claimed by 1838 versus Realized 1850 Wealth among Lottery Winners



Notes: This figure displays estimates of fraction of parcels claimed as of 1838 (as reported in Smith, 1838) versus total 1850 wealth for the subsample of lottery winners only. (By definition, lottery losers were not assigned parcels, so parcel characteristics are unavailable for the full sample.) The base sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. An observation is coded as a lottery winner ("treated") if he is a unique match to a name found on the list of winners published by Smith (1838). The dependent variable (on the y axis) is a dummy for whether the lotter winner's parcel had been claimed by the time of publication of the Smith (1838) list. The independent variable (on the x axis) in this figure is total measured wealth in 1850, the sum of real-estate wealth and slave holdings. Real-estate wealth is as reported in and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. The sample size for this figure is 1607. The solid line displays a local-polynomial smoothed estimate of the mean claim rate for each level of total wealth, and the short-dashed lines denote 95% confidence intervals. For reference, the long-dashed line presents the probability distribution function for log total wealth (excluding the imputation for those with zero wealth). These curves are estimated using, respectively, the "lpoly" and "kdensity" commands in Stata version 12. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Figure 5: Fraction with 1860 Wealth Less Than or Equal to $100 versus 1850 Wealth



Notes: This figure displays estimates of the fraction holding no more than $100 in 1860 as a function of 1850 wealth. The base sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. An observation is coded as a lottery winner ("treated") if he is a unique match to a name found on the list of winners published by Smith (1838). The dependent variable (on the y axis) is a dummy for whether the 1860 total wealth (personal plus real estate) is less than or equal to $100. The independent variable (on the x axis) in this figure is total measured wealth in 1850, the sum of real-estate wealth and slave holdings. Real-estate wealth is as reported in and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. The sample size for this figure is 5603. The dashed displays a local-polynomial smoothed estimate of the indicated fraction for each level of total wealth, and the grayed area denotes the 95% confidence interval. For reference, the solid gray line presents the probability distribution function for log total wealth (excluding the imputation for those with zero wealth). These curves are estimated using, respectively, the "lpoly" and "kdensity" commands in Stata version 12. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Table 1: Summary Statistics

	(1) Whole Sample	(2) Lottery "Losers"	(3) Lottery "Winners"	(4) p-value, mean difference [N]
Panel A: Lottery Winner or Loser				
Dummy for unique match to Smith (1838) list	0.124 (0.329)	0	1	---
Dummy for match to Smith (1838), deflated to 1/n in case of ties	0.155 (0.335)	0.037 (0.121)	0.995 (0.053)	0.000 [14375]
Panel B: Predetermined Outcomes				
Age, in years	51.2 (8.5)	51.3 (8.5)	50.9 (8.6)	0.122 [14375]
Born in Georgia	0.497 (0.500)	0.497 (0.500)	0.498 (0.500)	0.889 [14375]
Born in South Carolina	0.212 (0.408)	0.210 (0.407)	0.222 (0.416)	0.263 [14375]
Born in North Carolina	0.180 (0.384)	0.180 (0.384)	0.178 (0.383)	0.804 [14375]
Number of Georgia-born children in the three years prior to the lottery	1.333 (0.542)	1.333 (0.541)	1.332 (0.542)	0.910 [14375]
Cannot read and write	0.147 (0.354)	0.147 (0.354)	0.142 (0.350)	0.593 [14340]
Number of letters in surname	6.19 (1.61)	6.20 (1.62)	6.13 (1.51)	0.072 [14375]
Frequency with which surname appears in sample	36.2 (46.3)	36.3 (46.9)	35.3 (41.9)	0.380 [14375]
Surname begins with "M" or "O"	0.101 (0.302)	0.101 (0.301)	0.104 (0.305)	0.740 [14375]
Mean wealth of families in Georgia with same surname	1186.3 (1257.8)	1185.4 (1288.4)	1192.3 (1021.8)	0.811 [13848]
Median wealth of families in Georgia with same surname	289.1 (716.6)	290.0 (717.6)	282.7 (709.9)	0.686 [13848]
Mean illiteracy of adults in Georgia with same surname	0.219 (0.107)	0.219 (0.108)	0.218 (0.098)	0.648 [13848]

Notes: Table continues on next page.

Table 1 (continued): Summary Statistics

	(1)	(2)	(3)	(4)
	Whole Sample	Lottery “Losers”	Lottery “Winners”	p-value, mean difference [N]
		Panel C: Measures of Wealth in 1850		
Real-estate wealth	1999.0	1970.8	2198.2	0.068
	(4694.2)	(4422.0)	(6290.1)	[13094]
	{0,650,2000}	{0,640,2000}	{0,700,2000}	
Slave weath	1339.1	1297.3	1635.3	0.021
	(5761.0)	(5329.7)	(8189.0)	[14375]
	{0,0,0}	{0,0,0}	{0,0,326}	
Total wealth (sum of wealth in real estate and slaves)	3323.7	3245.5	3876.5	0.006
	(8691.0)	(7952.9)	(12734.4)	[13094]
	{100,800,3000}	{100,800,3000}	{100,1000,3550}	

Notes: This table displays summary statistics for the main data used in the present study. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. Column (1) presents means and standard deviations (in parentheses) of variables for this entire sample. We use two measures of whether the person won land in the drawing for the Cherokee Land Lottery of 1832. The first measure is coded to 1 if that person is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. The second measure takes individuals that “tie” for a match to the Smith list with (n-1) other observations and recodes them to 1/n. These variables are summarized in Panel A. Columns (2) and (3) present means and standard deviations of variables for the subsamples of, respectively, lottery losers and winners (decomposed using the first measure). Column (4) presents the p-value on the test of zero difference in means between the subsamples of losers and winners. In square brackets, we report the sample size used for this test, although the test involving children or surnames adjust for the clustering of errors. With the exception of measure the surname length, we use the Soundex version of each name to account for minor spelling differences. For the variables that are means by surname, we use the 1850 100% census file to construct average fertility, school attendance, and real-estate wealth among Georgia-resident households for each (soundex) surname. (Those individuals that appear in our lottery-eligible sample are excluded from the construction of these indices.) Real-estate wealth is as reported on and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. Numbers in curly brackets in Panel C are the 25th, 50th, and 75th percentiles of the respective wealth measures. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Table 2: Lottery Status versus Total Wealth in 1850

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Binary Match to Smith (1838)							
Levels	723.4 (325.3) **	714.4 (319.5) **	710.1 (325.4) **	632.4 (311.2) **	593.6 (352.3) *	855.1 (348.9) **	677.8 (385.6) *
Natural Logs	0.127 (0.043) ***	0.128 (0.043) ***	0.126 (0.043) ***	0.121 (0.043) ***	0.098 (0.049) **	0.142 (0.045) ***	0.098 (0.053) *
Panel B: Allow for 1/n Matching to Smith (1838)							
Levels	777.7 (310.7) **	749.8 (303.0) **	762.5 (310.5) **	660.2 (300.2) **	572.0 (335.6) *	922.7 (331.3) ***	645.6 (332.6) **
Natural Logs	0.146 (0.042) ***	0.147 (0.042) ***	0.146 (0.042) ***	0.135 (0.042) ***	0.112 (0.049) **	0.158 (0.045) ***	0.110 (0.053) **
Additional Fixed-Effect Controls:	None	First letter of surname	Number of letters in surname	Freq. of surname in sample	Surname	Given name	Surname; Given name

Notes: This table displays OLS estimates of equation (1) in the text. Each cell presents results from a separate regression, and only the coefficient on winning the lottery is reported. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same lapse of time. The dependent variable in this table is total measured wealth. This variable is the sum of real-estate wealth, which was reported to enumerators on the population schedule, and slave wealth, which was computed from the slave schedule. Whether this variable enters the specification in levels or natural logs is indicated by the row headings. The sample size the levels regressions is 13094, and is 10013 for the logs regressions. The baseline specification also includes dummies for age and for (state x county) of residence. Additional sets of fixed effects are included in columns 2-7, as reported in the bottom row. In columns 4-7, we use the Soundex version of each name to account for minor spelling differences. Two variables are constructed to measure whether the person was a lottery winner. The first measure, used in Panel A, is coded to 1 if that person is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. The second measure, which is used in Panel B, takes individuals that "tie" for a match to the Smith list with (n-1) other observations and recodes them to 1/n. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Table 3: Lottery Status vs. Various Wealth Measures in 1850

	(1) Binary match to Smith	(2) Allow 1/n match
Panel A: Total Wealth (N=13094)		
Levels	723.4	777.7
	(325.3) **	(310.7) **
Levels, Adjusted for Truncation of Lower Tail	723.6	777.6
	(325.2) **	(310.6) **
Natural Logs (N=10013)	0.127	0.146
	(0.043) ***	(0.042) ***
Natural Logs, Adjusted for Truncation of Lower Tail	0.121	0.142
	(0.049) ***	(0.049) ***
Panel B: Quantiles of Total Wealth (N=13094)		
Levels, 25th percentile	0.0	0.0
	(32.7)	(25.7)
Levels, 50th percentile (median)	200.0	200.0
	(39.8) ***	(26.7) ***
Levels, 75th percentile	550.0	511.8
	(109.7) ***	(116.1) ***
Levels, 95th percentile	1503.7	2022.1 **
	(1114.3)	(1076.6)
Panel C: Real-Estate Wealth (N=13094)		
Levels	286.3	295.2
	(159.7) *	(154.4) **
Indicator for Wealth At Least $100	0.002	-0.003
	(0.011)	(0.010)
Panel D: Slave Wealth (N=14375)		
Levels	391.8	431.8
	(201.8) *	(192.7) **
Indicator for Wealth > 0	0.039	0.052
	(0.011) ***	(0.011) ***

Notes: This table displays OLS estimates of equation (1) in the text, except for Panel B where a quantile regression is used. Each cell presents results from a separate regression, and only the coefficient on winning the lottery is reported. The specification also includes dummies for age and for (state x county) of residence. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. The dependent variables are various measures of wealth, as indicated in the Panel descriptions and row headings. The wealth variable in Panels A and B is the sum of real-estate wealth, which was reported to enumerators on the population schedule, and slave wealth, which was computed from the slave schedule. Panel C reports results for real-estate wealth. Enumerators in 1850 were instructed to record such wealth only if were at least $100, which is the cutoff we used for analyzing the extensive margin in Panel C as well as for estimating the truncated normal used to impute values below $100 in the truncation adjustment in Panel A. Two variables are constructed to measure whether the person was a lottery winner. The first measure, used in Column (1), is coded to 1 if that person is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. The second measure, used in Column (2), takes individuals that "tie" for a match to the Smith list with (n-1) other observations and recodes them with to 1/n. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Table 4: Falsification test using South Carolina instead of Georgia to construct sample

Dependent variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lottery-status variables:	Born in Georgia	Born in South Carolina	Number Ga.-born children, pre-lottery	Number SC-born children, pre-lottery	Resides in Old Cherokee County	Real-estate Wealth ($)	Real-estate Wealth >$100	Real-estate Wealth >$3000
Panel A: South Carolina, basic specification								
Dummy for unique match to Smith (1838) list	0.001	-0.017		-0.019	0.005	-41.1	0.001	0.007
	(0.004)	(0.013)		(0.019)	(0.007)	(236.2)	(0.016)	(0.013)
Dummy for match to Smith (1838), deflated to 1/n in case of ties	0.004	-0.016		-0.002	0.010	-15.6	0.001	0.004
	(0.004)	(0.012)		(0.018)	(0.007)	(232.0)	(0.015)	(0.012)
Panel B: South Carolina, including surname fixed effects								
Dummy for unique match to Smith (1838) list	0.000	-0.003		-0.016	0.006	-93.3	0.005	0.012
	(0.004)	(0.014)		(0.021)	(0.008)	(229.4)	(0.016)	(0.012)
Dummy for match to Smith (1838), deflated to 1/n in case of ties	0.003	-0.004		-0.004	0.009	-72.6	0.015	0.012
	(0.004)	(0.014)		(0.020)	(0.008)	-(72.6)	(0.015)	(0.014)
Panel C: Analogous results for Georgia, dummy for unique match to Smith list								
Basic specification	-0.004	0.014	0.002		0.022	295.2	0.002	0.020
	(0.012)	(0.011)	(0.014)		(0.008) ***	(154.4) *	(0.011)	(0.009) **
Control for surname fixed effects	0.001	0.012	0.009		0.023	315.8	0.002	0.021
	(0.014)	(0.012)	(0.016)		(0.008) ***	(146.8) ***	(0.011)	(0.010) **

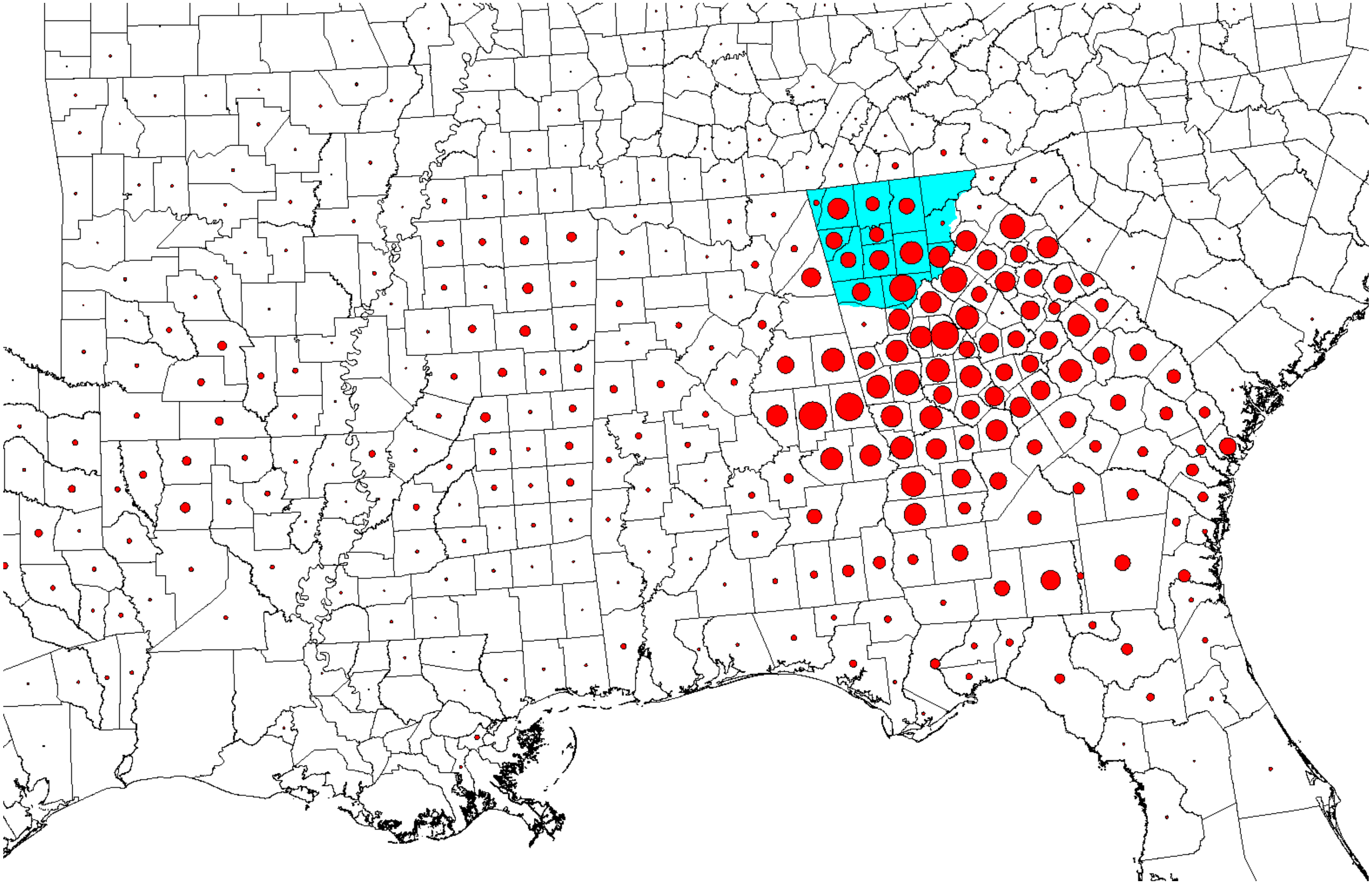
Notes: This table displays estimates of equation (1) in the text. Each cell presents results from a separate regression, and only the coefficient on "winning the lottery" is reported. The sample for Panels A and B consists of all households in the 1850 census with children born in South Carolina during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. The sample for Panel C, which repeats some results from earlier tables, uses households with Georgia-born children in this same window. We use two measures of whether the person won land in the drawing for the Cherokee Land Lottery of 1832. The first measure is coded to 1 if that person is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. The second measure takes individuals that "tie" for a match to the Smith list with (n-1) other observations and recodes them to 1/n. Note that these are spurious measures for the South-Carolina samples because the birthplace of their children implies that they lived outside of Georgia at some point during the three years prior to the lottery, and were therefore ineligible. The basic specification also includes dummies for age. The other specification used includes fixed effects for surname (soundex). The dependent variables are indicated in the column headings. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors (shown in parentheses) are heteroskedasticity robust and clustered on the lottery-eligible man in there are multiple observations per household. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Table 5: Interaction with Own Illiteracy and Surname Averages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
		Additional variables constructed from surname averages in 1850 Census data from Georgia (excluding own realizations)													
	Baseline	Mean wealth		Mean log wealth		Log median wealth		Fertility		In school, ages 5-15		Literacy rate		Own illiteracy	
	Panel A: Total wealth, Levels, Adjusted for Truncation of Lower Tail														
Dummy for unique match to Smith (1838) list	716		717		707		710		682		709		713		826
	(232) ***		(298) **		(298) **		(298) **		(283) **		(297) **		(298) **		(251) ***
Additional variable		0.158	0.144	821	684	443	361	126	182	1438	1260	1802	1631	-2186	-2084
		(0.067) **	(0.062) **	(145) ***	(150) ***	(86) ***	(87) ***	(223)	(224)	(433) ***	(469) ***	(932) *	(766) **	(227) ***	(241) ***
Interaction term (using z score for surname variables)			44		-12		25		346		42		-397		-811
			(216)		(241)		(224)		(304)		(198)		(390)		(650)
Number of observations	13036	12553	12553	12553	12553	12553	12553	12896	12896	12417	12417	12595	12595	13005	13005
	Panel B: Total wealth, Natural Logs, Adjusted for Truncation of Lower Tail														
Dummy for unique match to Smith (1838) list	0.115		0.120		0.115		0.117		0.117		0.121		0.118		0.115
	(0.048) **		(0.051) **		(0.051) **		(0.051) **		(0.050) **		(0.051) **		(0.051) **		(0.051) **
Additional variable		4.04E-5	3.83E-5	0.282	0.236	0.164	0.140	-0.063	-0.004	0.444	0.411	0.911	0.679	-0.974	-0.971
		(1.75E-5) **	(1.63E-5) **	(0.035) ***	(0.036) ***	(0.022) ***	(0.022) ***	(0.053)	(0.052)	(0.117) ***	(0.119) ***	(0.168) ***	(0.172) ***	(0.046) ***	(0.049) ***
Interaction term (using z score for surname variables)			0.083		0.043		0.009		-0.028		-0.056		0.038		-0.024
			(0.050) *		(0.050)		(0.049)		(0.063)		(0.051)		(0.049)		(0.132)
Number of observations	13036	12553	12553	12553	12553	12553	12553	12896	12896	12417	12417	12595	12595	13005	13005
	Panel C: Total wealth Greater than $5000														
Dummy for unique match to Smith (1838) list	0.029		0.029		0.028		0.028		0.027		0.028		0.028		0.034
	(0.009) ***		(0.010) ***		(0.009) ***		(0.009) ***		(0.009) ***		(0.010) ***		(0.009) ***		(0.010) ***
Additional variable		6.96E-6	7.09E-6	0.042	0.038	0.022	0.019	0.017	0.019	0.071	0.047	0.123	0.100	-0.108	-0.104
		(3.23E-6) **	(3.26E-6) **	(0.007) ***	(0.007) ***	(0.004) ***	(0.004) ***	(0.009) *	(0.009) **	(0.022) ***	(0.023) **	(0.030) ***	(0.032) ***	(0.009) ***	(0.009) ***
Interaction term (using z score for surname variables)			0.000		0.004		0.004		0.012		0.001		-0.005		-0.037
			(0.010)		(0.009)		(0.010)		(0.013)		(0.010)		(0.009)		(0.025)
Number of observations	14306	13780	13780	13780	13780	13780	13780	14144	14144	13615	13615	13824	13824	14271	14271

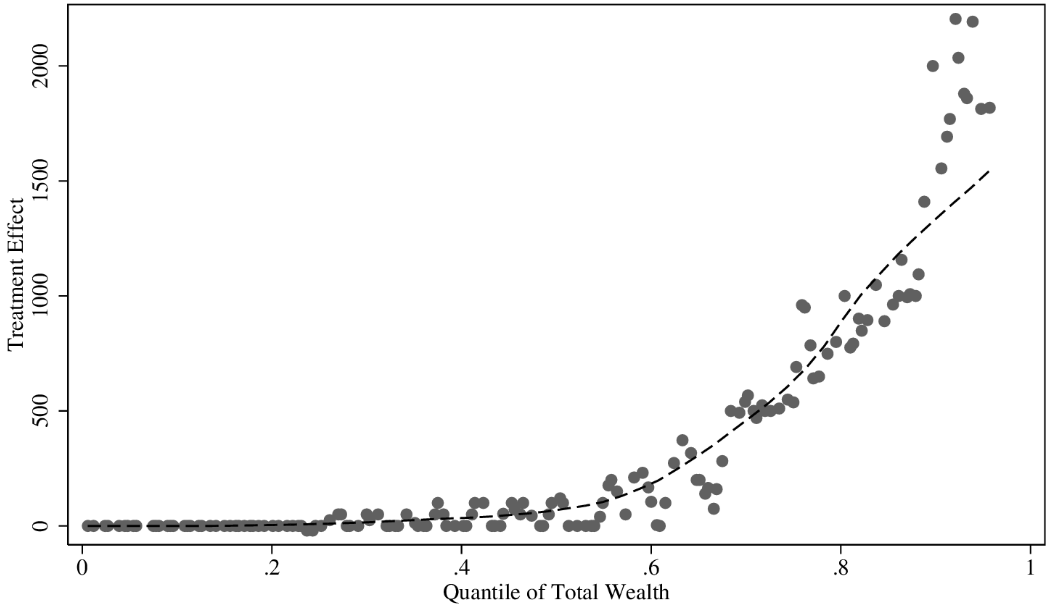
Notes: This table displays OLS estimates of equations (2) and (3) in the text. This table departs from previous ones in the use of surname-specific characteristics to proxy for differences across extended families ("dynasties") in child and wealth outcomes. We use the 1850 100% census file to construct average fertility, school attendance, and real-estate wealth among Georgia-resident households for each (soundex) surname. (Those individuals that appear in our lottery-eligible sample are excluded from the construction of these indices.) Each panel/column presents results from a separate regression. In addition to the displayed coefficients, regressions include dummies for age and for (state x county) of residence. The base sample for these regressions consists of all households in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. Households without a corresponding surname in the database of surname averages are excluded from the regressions. The dependent variables are indicated in the column headings. A household is coded as a lottery winner if the head is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors (shown in parentheses) are heteroskedasticity robust and clustered on the surname level to account for correlation induced by the surname-averages. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Appendix Figure 1: Old Cherokee County and the 1850 Locations of the Sample



Notes: This figure displays a map of the southeastern United States with information on the location (by county) in 1850 of the lottery-eligible households in our main sample. Black lines indicate the 1850 county boundaries, drawn from the NHGIS database. The area shaded in blue in northwest Georgia denotes old Cherokee County, which was allocated by the Cherokee Lottery of 1832. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. If households in our sample are resident in a county in 1850, we place a red dot at the county centroid. The area of a dot is proportional to the number of sample households resident in that county. A minor fraction of sampled households resides in counties outside the frame of this map. Such households are included in the econometric analysis, but we zoom in on this region to make the features legible in the map. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Appendix Figure 2: Replicate Figure 1 (Quantile Regressions) Using 1/n Match to Smith Instead



Notes: This figure displays estimates from a quantile regression of the effect of winning the lottery ("treatment") on total wealth in 1850. The points are the quantile-specific estimates of the treatment effect at various quantile points. The dashed line is a local-polynomial-smoothed (Epanechikov kernel, with a bandwidth of .11) estimate of the treatment effect. The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. An observation is coded to a lottery status using the 1/n match procedure as described in the text. The dependent variable in this figure is total measured wealth in 1850, the sum of real-estate wealth and slave holdings. Real-estate wealth is as reported on and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. The sample size for this figure is 13094. Data sources and additional variable and sample definitions are found in the text and in the appendices.

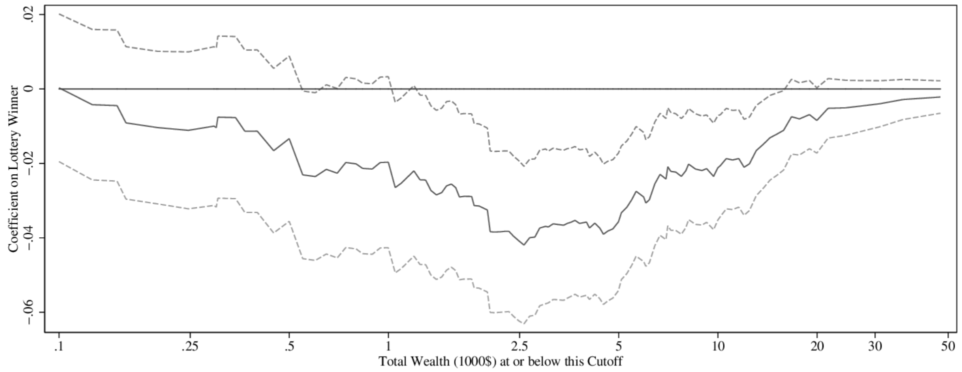
Appendix Figure 3: CDF Differences under Various Specifications

Panel A: Binary Match to Smith List, Baseline Specification (Shown in Figure 2, Panel B)

Panel B: Binary Match to Smith List, Bivariate Specification

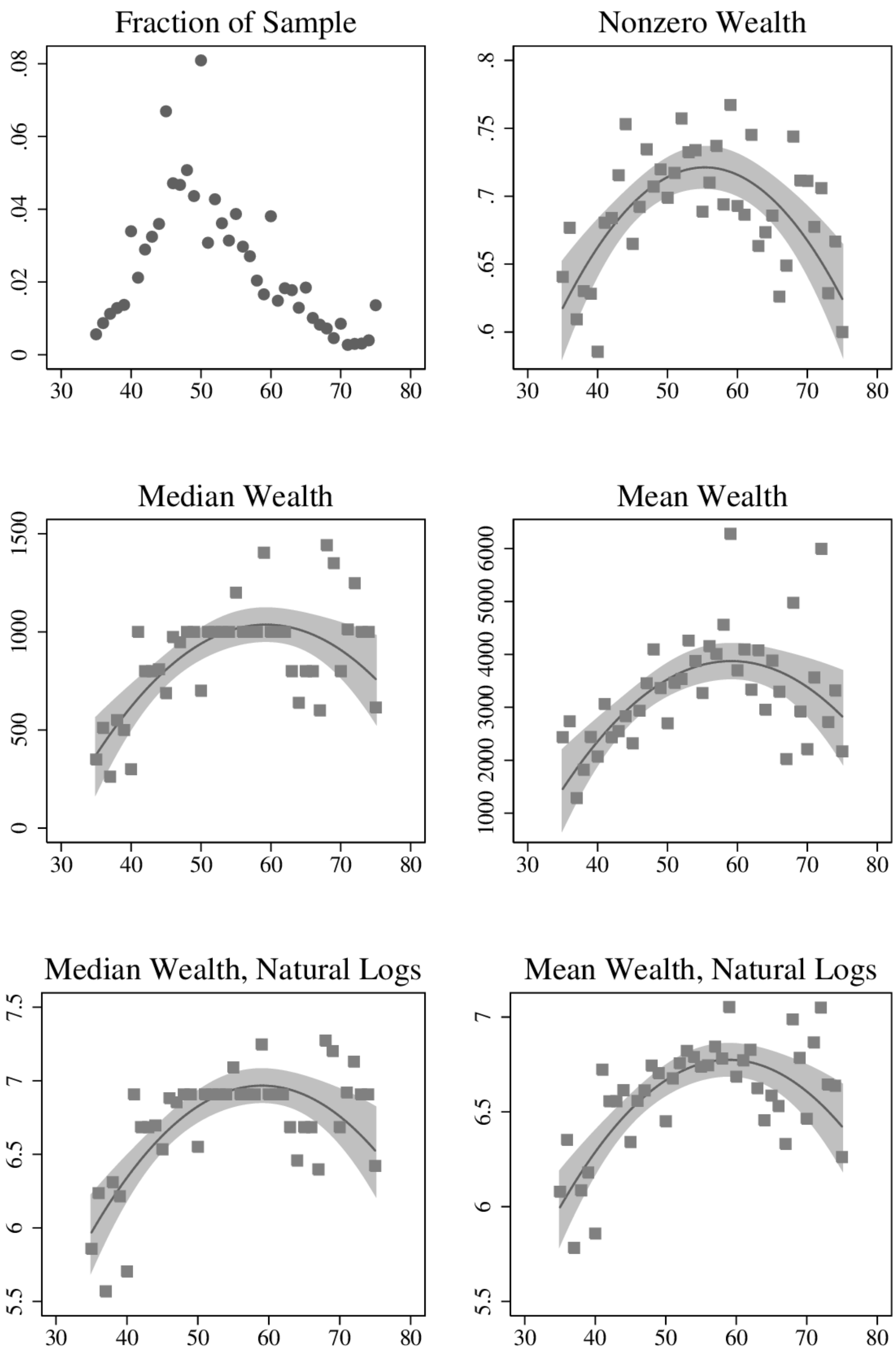
Panel C: Binary Match to Smith List, Baseline Specification with Soundex Fixed Effects

Panel D: 1/n Match to Smith List, Baseline Specification



Notes: This figure displays alternate estimates of the change in cumulative distribution functions in Figure 2, Panel B. The solid lines in each Panel presents the differences in the cumulative distribution function between groups (treatment minus control), estimated using a linear probability model (equation 1) of an indicator for being below 200 quantile cut-points for wealth. Heteroskedasticity-robust 95% confidence intervals are plotted with the dashed lines. See the notes for Figure 2 for definitions of the data and specifications.

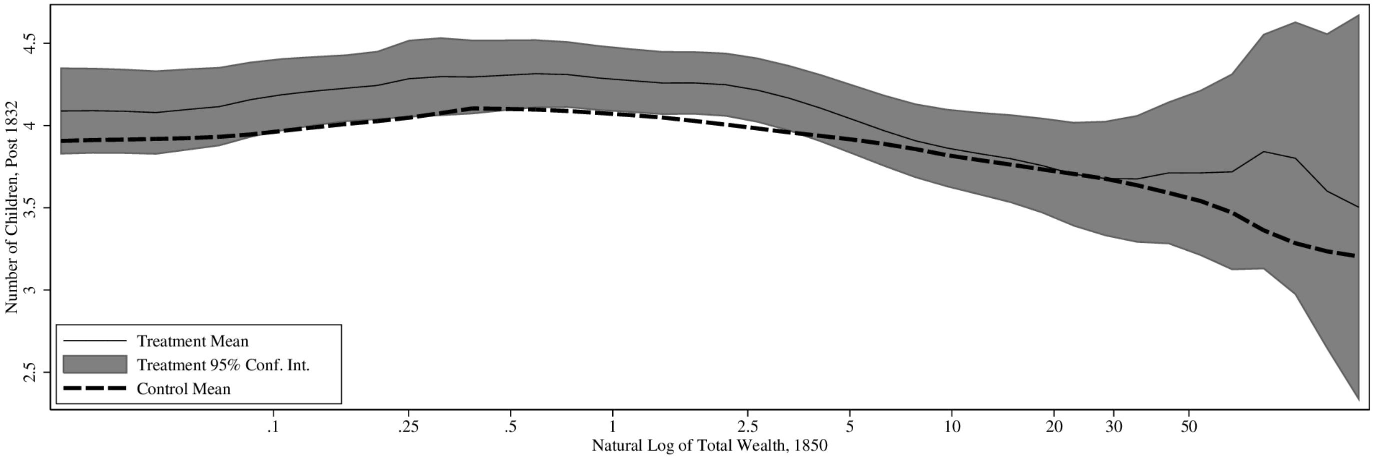
Appendix Figure 4: Wealth versus Age



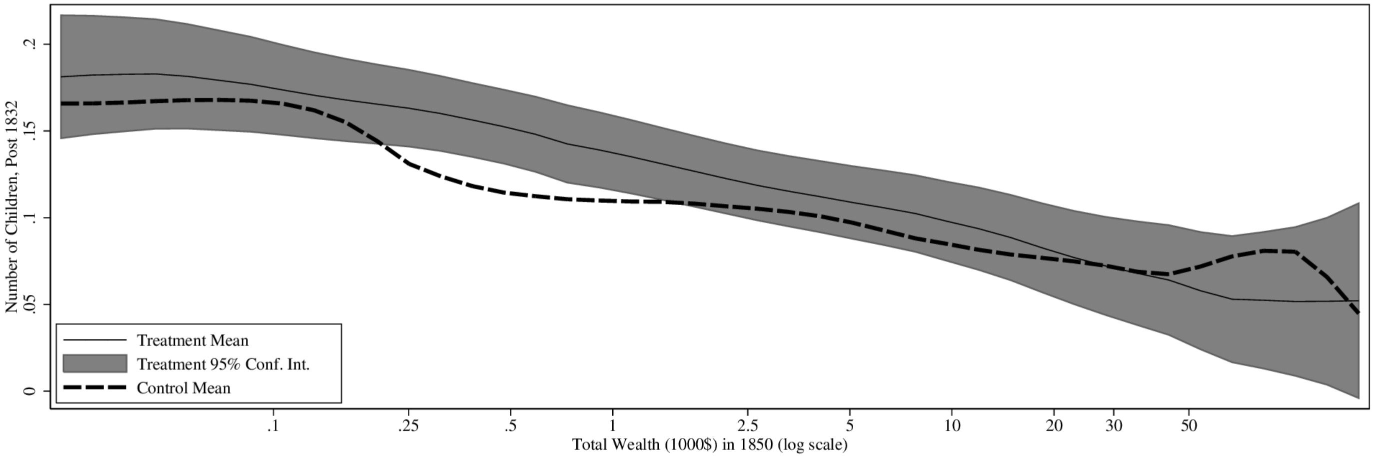
Notes: This figure plots the average total 1850 wealth by age in the sample of lottery eligibles, except for the upper left-hand panel, which displays the fraction of the sample in each age cell. The age-specific averages are denoted with the square symbols. A quadratic fit (plus confidence interval) is displayed with the solid line (and associated shading). The sample and data are defined as in Figure 1-3, except that we do not display results for the handful of observations with ages above 75.

Appendix Figure 5: Fertility, Residence, and Realized 1850 Wealth by Lottery Status

Panel A: Number of Children Born Post Lottery in 1850 Household



Panel B: Fraction Residing in Old Cherokee County in 1850



Notes: This figure displays estimates for 1850 of the number of children and residence in Old Cherokee County versus total wealth, by lottery status. The base sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. An observation is coded as a lottery winner ("treated") if he is a unique match to a name found on the list of winners published by Smith (1838). The dependent variable (on the y axis) is an indicated above. The independent variable (on the x axis) in this figure is total measured wealth in 1850, the sum of real-estate wealth and slave holdings. Real-estate wealth is as reported in and transcribed from the manuscript pages of the 1850 Census of Population. Slave wealth was estimated by linking the household to the 1850 Slave Schedule and imputing a market value of slave holdings adjusting for the reported ages and gender of slaves on the Schedule. The solid and dashed display local-polynomial smoothed estimates of the means for each level of total wealth for the treated and controls, respectively. The gray shaded areas denote 95% confidence intervals for the treatment-group conditional mean. These curves are estimated using the "lpoly" command in Stata version 12. Data sources and additional variable and sample definitions are found in the text and in the appendices.

Appendix Table 1: Differences in County-of-Residence Characteristics by Lottery Status

(1) Resides in Old Cherokee County	(2) Resides in Georgia	(3) Miles East	(4) Miles North	(5) School Enroll. Rate	(6) Total Fertility Rate (TFR5)	(7) Total Fertility Rate (TFR19)	(8) Log of Farm Value per Acre	(9) Log of Average Farm Size	(10) Log of Improved Land Ratio	(11) Log Slaves per Area	(12) Log Fraction Urban	(13) Access to Water Transport	(14) Access to Railroads
Panel A: Basic Specification													
0.022	0.005	4.320	-4.026	-0.003	0.006	0.011	-0.007	-0.014	-0.017	-0.001	0.072	-0.007	0.018
(0.008) ***	(0.011)	(3.643)	(2.211) *	(0.003)	(0.004)	(0.012)	(0.021)	(0.017)	(0.017)	(0.026)	(0.072)	(0.011)	(0.016)
Panel B: Control for Surname Fixed Effects													
0.022	0.004	4.265	-4.661	-0.004	0.005	0.009	-0.011	-0.005	-0.024	0.000	0.066	-0.001	0.015
(0.008) ***	(0.013)	(3.997)	(2.306) **	(0.003)	(0.004)	(0.012)	(0.022)	(0.017)	(0.018)	(0.026)	(0.062)	(0.011)	(0.016)
Panel C: Control for State Fixed Effects and Dummy for Residence in Old Cherokee County													
---	---	4.654	-5.924	-0.004	0.005	0.007	-0.012	-0.004	-0.015	0.008	0.072	-0.003	0.018
		(4.343)	(2.781) **	(0.003)	(0.004)	(0.012)	(0.022)	(0.016)	(0.017)	(0.025)	(0.072)	(0.011)	(0.015)
Panel D: Control for State Fixed Effects,Surname Fixed Effects, and Dummy for Residence in Old Cherokee County													
---	---	4.560	-6.569	-0.005	0.004	0.005	-0.016	0.004	-0.022	0.009	0.066	0.002	0.014
		(4.727)	(2.761) **	(0.003) *	(0.004)	(0.012)	(0.022)	(0.016)	(0.019)	(0.026)	(0.062)	(0.011)	(0.015)

Notes: This table displays OLS estimates of equation (1) in the text. Each cell presents results from a separate regression, and only the coefficient on winning the lottery is reported. The basic specification (shown in Panel A) also includes dummies for age. The specification used in Panel B includes fixed effects for surname (soundex). The sample consists of all household heads in the 1850 census with children born in Georgia during the three years prior to the Cherokee Land Lottery of 1832 and no children born outside of Georgia during the same period. The dependent variables are the locational county-specific characteristics denoted in the column headings. Location data used in Columns 3 and 4 are county centroids computed from NHGIS data, and are converted into miles east or north of the NAD83 reference point in central Oklahoma. County data used in Columns 5-14 are drawn from ICPSR study #2896. The number of observations for Columns 1-4 is 14375 and for Columns 5-14 is 14237 because of missing data for some (mostly unorganized) counties. A household is coded as a lottery winner if the head is a unique match to a name found on the list of winners published by Smith (1838); anyone else in the sample is coded to zero. A single asterisk denotes statistical significance at the 90% confidence level; double 95% and triple 99%. All standard errors are heteroskedasticity robust and, in Columns 3-14, clustered at the (state x county) level to account for multiple observations per county. Data sources and additional variable and sample definitions are found in the text and in the appendices.