Land Openings on the Georgia Frontier and the Coase Theorem in the Short- and Long-Run*

Hoyt Bleakley† Joseph Ferrie‡

First version: September 25, 2012
This version: February 28, 2013

Abstract

The Coase Theorem, with low transaction costs, shows the independence of efficiency and initial allocations in a market, while the recent "market design" literature stresses the importance of getting initial allocations right. We study the dynamics of land-use in the two centuries following the opening of the frontier in the U.S. state of Georgia, which – in contrast with neighboring states – was opened up to settlers with a pre-surveyed grid in waves with differing parcel sizes. Using difference-in-difference and regression-discontinuity methods, we measure the effect of initial parcel sizes (as assigned by the surveyors' grid) on the evolution of farm sizes decades after the land was opened. Initial parcel size predicts farm size essentially one-for-one for 50-80 years after land opening. This effect of initial conditions attenuates gradually, and only disappears after 150 years. We estimate that the initial misallocation depressed the area's land value by 20% in the late 19th century. This episode suggests the relevance of the Coase Theorem in the (very) long run, but that bad market design can induce significant costs in the medium term (over a century in this case).

Keywords: market design, economic geography, history dependence, coordination problem

*Preliminary and incomplete. Comments welcome. The authors thank Brian Bettenhausen and Eric Hanss for help with spatial data, and seminar participants at the University of Illinois and the University of Chicago 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

The "market design" literature in economics has analyzed methods for allocating property rights in a variety of settings (Noll 1982; Hahn 1984). This work has shown that getting the initial allocation wrong can have long-run consequences for economic efficiency, a claim that stands in stark contrast with Coase (1960) who argued that the initial allocation of a property right should, in the absence of transaction costs, be irrelevant to the efficiency of the final outcome. After the initial allocation is made, according to Coase, agents should bargain and exchange until there are no longer gains from trade. But this outcome (as Coase recognized) can be long-delayed or even thwarted entirely due to transaction costs. These costs might be prohibitively high, both because there are real resource costs to bargaining and trading, or unequal market power, or, as argued by Myerson and Satterthwaite (1983), binding information constraints.

Theoretical results from the market-design literature have been applied to designing allocations for radio spectrum, logging on public lands, and internet domains. (We further review the related literature in Section 2.) Nevertheless, evidence is scarce for the long-run impact (and whatever improvement there might be in efficiency) when moving from a sub-optimal to an optimal mechanism. Compelling evidence on this question is hard to find for three reasons: (1) the usual econometric problem that the choice of the initial allocation was endogenous (for example, pollution permits might be assigned only to firms already present in the market and in possession of substantial market power); (2) the recent adoption of many of these designs, making it impossible to evaluate long-term effects on the efficiency of allocation mechanisms; and (3) the multiple dimensions along which particular market designs may differ from the optimum.[1]

We provide evidence on the long-run impact of differences in market design from a novel context: the opening of land for settlement on the frontier of the U.S. state of Georgia. In the initial third of the 19th century, Georgia made almost three quarters of its total land area available to (white) settlers in a series of eight lotteries. The historical idiosyncrasies

[1]For example, the comparison by Libecap and Lueck (2011) between areas of the U.S. settled under the rectangular survey (RS) system and the ancient metes-and-bounds (MB) system (in which parcels are identified by the locations of trees, rocks, and myriad natural and man-made markets) is complicated by differences across these areas in at least three dimensions: (1) the variation in parcel sizes (uniform sizes under RS, no restriction on sizes under MB); (2) the shape of parcels (rectangular under RS, irregular under MB); and (3) the presence of "interstitial" parcels (none under RS, potentially many under MB).

that led Georgia to do this are discussed in Section 3.1; suffice it to mention here that some particularly egregious corruption scandals in Georgia motivated this choice for its sheer transparency. The use of a lottery for allocation imposed two features on the market design in this case: (a) a grid survey to demarcate parcels available to win and (b) the allocation of property rights for *all* of the land to a diffuse set of owners at the moment of land opening. This latter feature in particular contrasts with how land was opened in the rest of the region, where land moved from public to private domain in a slow and piecemeal fashion. The waves of the lottery mainly differed along two margins: which swath of the state was opened and what initial farm size was imposed by the surveyors' grid. We investigate the persistence of these assigned farm sizes, which might reasonably be expected to be sticky because shifting the average farm size in a filled-out grid of farms represents a serious coordination problem. (In Section 3.3, we discuss further the comparative market-design issues in this context.)

Our results imply considerable persistence of the initial allocation, although not a strong path dependence (that is, limited rather than unlimited persistence). We estimate to what extent initial parcel size (as assigned by the surveyors' grid) affects the evolution of farm sizes decades after the land was opened. We find that the initial, assigned size of land lots in an area predicts average farm size essentially one-for-one more than a half century after the land was opened for settlement. Nevertheless, this footprint of initial conditions is more a case of delayed adjustment rather than path dependence. Indeed, the relationship between assigned and realized farm sizes attenuates gradually, and disappears (in both a statistical and economic sense) after 150 years. Thus, our results split the difference between the two opposing views discussed above. On the one hand, the Coase Theorem appears to be operative in the very long run. On the other hand, the poor market design of the lotteries induced significant costs in the medium run, albeit a medium run that covered more than a century. The supporting evidence for this conclusion is presented in Sections 5 and 6. (The data that we use are described in Section 4.)

We first demonstrate this persistence using a difference-in-difference design, shown in Section 5.1. We compare average farm sizes in the Georgia Lottery Zone (hereafter "Lottery Zone" or "LZ") to those in adjacent counties in a buffer around the LZ (hereafter, the "Buffer Zone"). In effect, we imagine the Buffer Zone as a pseudo-Lottery-Zone that extends past the LZ's boundaries, and extrapolate acreage assignments from the Lottery Zone to the Buffer Zone. (Details on the allocation methods in the Buffer Zone are found in Section 3.2.)

Adjoining areas in these two zones will, by and large, share land characteristics, which allows us to use the Buffer as a comparison group. By this assumption (which is a standard one for a difference-in-difference model), the Buffer filters out any relationship between assigned and realized farm size arising from the endogeneity of the assignment rule. We find that the farm sizes assigned by the surveyors' grid are very sticky in the Lottery Zone, over and above the (mostly attenuated) relationship in the Buffer Zone. Further, these results are not sensitive to a variety of spatial controls, county fixed effects, or to zooming in to sample only those areas close to the LZ's external boundary.

We find a similar pattern of persistence when looking within the Lottery Zone. The variation in assigned parcel sizes within the LZ was considerable and there were many boundaries at which parcel sizes varied substantially. We first use a regression analysis in Sections 5.2, and show the result is robust to a variety of spatial controls and to restricting the sample to only those areas that were very close to a boundary were the assigned acreage changed. Relatedly, in Section 6, we conduct regression-discontinuity (RD) tests testing for breaks at the boundary between two zones with different initial assigned parcel sizes. With either estimator, we find results that are broadly similar to those found with the difference-in-difference design.

Finally, we show in Section 7 that the medium-run cost of misallocation was substantial. In the post-bellum years, the relationship between imputed acreage assignment and realized farm size had gone to zero in the Buffer Zone. Taking the Buffer as a measure of the unconstrained optimum, we then use the difference-in-difference model to impute a gap between optimal and constrained average farm sizes in the Lottery Zone. In a auxiliary regression, we next find that the square of the gap measure (squared for the same reason that Harberger triangles are areas rather than lengths) predicts lower farm values. As a check, we also find that similar, albeit somewhat larger, effects of misallocation on value using the RD model. By the middle of the 20th century, when the excess relationship between assigned and realized farm size in the LZ has disappeared, estimates of both the gaps and the costs of misallocation go to zero. But the persistent misallocation of land resulting from the initial lottery parcel size had a substantial cost in the medium run: nearly 20% of land values in 1880, three quarters of a century after the first lottery was conducted.

2 Related Literature

In recent years, market design has emerged as the practical counterpart to mechanism design – they are related as engineering is to theoretical physics. Milgrom (2011) summarizes some of the practical complications in mapping mechanism design's findings into market design's real-world settings (e.g. "product definition," "message transmission," "incentives," and "linkage across markets"). An additional complication is the impact of the initial allocation of an asset on the efficiency with which the market subsequently allocates the asset. The concern with this issue conventionally begins with Coase (1960) and the "Coase theorem."[2] In a setting where transaction costs are present, the outcome will be shaped by the initial distribution of the property right. This insight has been extended in a variety of ways. Myerson and Satterthwaite (1983) propose that information constraints may be binding (e.g. each party may not know the other parties' true valuations for the asset), imposing frictions on *ex post* trade and preventing the achievement of the efficient allocation, while Hahn (1984) shows how disparate market power can prevent an efficient outcome.

The key lesson that emerges from attempts to design real-world markets in light of the concerns raised by Coase and others on the role of the initial allocation of property rights is that getting those rights "right" at the outset is crucial to even approximating the efficient outcome. In the presence of transaction costs, information asymmetries, or market power, a poor choice of initial allocation might not be remedied by the standard Pigouvian tax on the party that is initially advantaged. Though this challenge has been recognized in market design, the extent to which particular initial allocations lead to outcomes that depart from the efficient outcome has not been possible to gauge in most settings.

There are three challenges to such measurement: (1) the choice of the initial allocation of the asset to be distributed through a particular mechanism is often endogenous (for example, rights might be assigned to market incumbents rather than more broadly among the set of potential market participants); (2) many market-design exercises have occurred

[2]McCloskey (1998) traces its provenance back through Arrow and Debreu (1954), Edgeworth (1881), and Smith (1776), and states it simply as the proposition that "an item gravitates by exchange into the hands of the person who values it the most, if transactions costs (such as the cost of transportation) are not too high." (1998, p. 368). McCloskey also asserts that Coase's original intention was not to describe what would happen in a setting without transaction costs but to point out that transaction costs prevent a simple first-best solution (such as taxing the polluting party in the case of pollution) from achieving the efficient outcome. In short, in the real world, the initial assignment of the property right matters, according to Coase (McCloskey, 1998, *passim*).

only recently so long-term effects on the efficiency of the allocation mechanism in light of the initial allocation cannot yet be observed; and (3) the assets traded in many markets differ along multiple dimensions, complicating the identification of exactly how the initial allocation departs from that likely to produce the efficient outcome. The particular setting we exploit here – the distribution by lottery of land in parcels of different sizes in early nineteenth century Georgia – allows us to overcome these challenges.

This project is also related to a burgeoning literature, mainly in economic history, examining the persistent impact, even over spans of time as long as several centuries, of conditions at some point in the past. Much of this work has focused on land allocation and use rules. Libecap and Lueck (2011) compare outcomes in twentieth century Ohio where it is possible to locate plots distributed in the nineteenth century under the ancient system of metes and bounds immediately adjacent to plots distributed under the federal government's rectangular system. They find by a variety of measures that the areas with uniform, well-specified plot boundaries at the outset were better off more than a century and a half later. Dell (2010) examines areas on different sides of the border in Peru and Bolivia between (1) an area that was obliged to provide labor for the silver mines from 1573 to 1812 and (2) the adjacent region that was not, finding higher levels of twentieth century economic development outside the region from which labor was forcibly extracted.

Despite the insights this sub-genre has generated, we still lack a good metric by which to assess the cost of differences in early market arrangements, whether is the form of the system used to survey land, the political economy of labor extraction, or the very lay of the land. At the same time, the past differences generating these long-run outcome differences are multidimensional, so it is difficult to isolate the differences that matter. Our work allows us to be more precise in measuring the cost of the initial allocation, and in identifying the mechanisms that generate outcomes nearly two centuries after the initial conditions are observed.

Finally, our project also reflects an interest in patent law on whether the holder of an asset (a patentee or, in our case, a lottery winner) should be required to make use of the asset in order to secure the property right in that asset. Kitch (1977) claims that patent rights should be distributed before the value of the thing being patented is ascertained. This arrangement makes it easier for the patent holder to manage the development of the invention. In our case, individuals who won parcels in the Georgia land lotteries did not immediately know the quality of their randomly-assigned parcels and were not required to homestead them—they could immediately sell them without so much as visiting their properties. By contrast, areas of Alabama and Florida settled under federal land law were characterized by plots purchased for cash and then taken up by individuals who presumably knew the quality of the asset that they chose and, in the case of homesteaders, were obliged to remain on and improve the land for several years. Which of these arrangements yields the more efficient outcome (as measured by, say, average land price per acre) is a question we can investigate further with our data.

3 Background

3.1 Origins and Nature of the Georgia Land Lotteries

The state of Georgia distributed over two-thirds of its territory to settlers in a series of land lotteries held in the first third of the 19th century. Distributing parcels by lottery required first delineating parcels, which was conducted by state surveyors just prior to each lottery. For simplicity, the surveyors imposed a uniform grid when demarcating each wave of the land opening, and this resulted in equal-sized parcels within each wave of the lottery.

Using lotteries to allocate land on a large scale is quite uncommon and stands in marked contrast with much of the land opening in U.S. history. Contrast a lottery with homesteading, for example. In the case of homesteading, a large set of territory is opened for settlement, but no specific property rights were assigned ex-ante. Instead, individuals (or heads of households bringing their families) would go out to the open area and stake a claim. In the case of the territories opened under the rules of the Homestead Act of 1862, the claimant would have to spend several years working that parcel and demonstrate that the land had been improved in order for the claim to be considered valid. Much of the interior of the original Georgia colony (that is, the southwest side of the Savannah River watershed) was settled in like manner through the so-called Headright System. A common alternative mechanism elsewhere in the U.S. was the land grant, in which a large chunk of territory was given to a single individual (often in payment for military service), who could then subdivide and sell it. Importantly for our research design, Georgia's neighboring states did not use lotteries to allocate land, but rather a combination of claims and grants just discussed.

Why would the Georgia government have chosen to give out land in broad-based lotteries? The motivation starts with two notorious corruption scandals in Georgia (and U.S.) history: the Pine Barrens Fraud and the Yazoo Land Fraud, both of which took place in Georgia in the 1790s. In the former episode, a number of entrepreneurs bribed and 'packed' local land commissions to make and re-sell land claims (some of which were for non-existent land). In the latter, several land companies bribed a majority of the Georgia legislature to allow them to subdivide and sell vast stretches of land on the Georgia's then-western frontier.[3] These corruption scandals (particularly the Yazoo Land Fraud) caused a great deal of political tumult at the time and further cast a shadow on state politics for decades to come.

Nevertheless, by 1800, only one third of the Georgia's territory had been settled (by the non-indigenous). There was pressure on the state government to open up the rest of the state for white settlement. As a consequence, the Native Americans' lands were expropriated and they were evicted from their lands in various waves throughout the first four decades of the 19th century. (This particular aspect of settlement was of course not unique to Georgia.) The tribes that were evicted from these areas were principally the Creeks and the Cherokee, with the final set of evictions precipitating what is now known as the Trail of Tears.

When opening these newly available lands, there remained the question of how to allocate them in an orderly fashion. The Georgian politicians of the day were mindful of the acrimony generated by the earlier scandals. It was decided to allocate the rest of the state using a mechanism that was as transparent as possible. As Cadle (1991) puts it,

> Georgia adopted the rectangular system and the mechanism of the lottery, however, not so much as an escape from the shortcomings of the headright system [but rather] as an indignant reaction against the two great land scandals of the 1790s—the Pine Barrens Fraud and the Yazoo Fraud. After having suffered these experiences, the people of Georgia resolved in all future dispositions of public lands to guard against every opportunity for fraud and corruption. (169)

The lottery system with a pre-determined grid would have been difficult to corrupt. A typical lottery drawing was as follows. Following the surveying work, a drawing to set up by

[3]The Georgia government authorized these companies to sell land in the far West of its claimed territory, including land in the watershed of the Yazoo River, which is currently in the state of Mississippi. (Recall that immediately after the American Revolution, several of the original 13 colonies claimed land as far west as the Mississippi River.)

government commission to arrange the distribution of parcels. Names of eligible Georgians were collected from the local authorities, and each name was placed on a slip of paper. The slips of paper were placed in one barrel. Another barrel was filled up with slips of paper, one for each parcel available in the land opening plus enough blank slips to equalize the number in each barrel. Both barrels were rolled around to ensure adequate mixing (or proper randomization, in today's parlance). The blank slips were to further increase 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. Having the survey before the lottery also reduced the incentive to corrupt the parcel definitions. Even if you could influence the surveyors to produce a parcel that was bigger and better, there would be no way to ensure that you actually drew that parcel in the lottery. Furthermore, the parcels were very small relative to the grants in the Yazoo episode and actually existed unlike much of the 'land' in Pine-Barrens episode, so there was comparatively little for an rent-seeker to steal at any discrete step in the process. In sum, this system was hard to game, and the rewards for doing so were limited.

The grid size for each wave was chosen in accordance with what was known about the average land quality. For instance, the southeastern corner of the zone was dominated by the swampland, which was both unsuitable for agricultural use and was well known by Georgians of the era. To compensate for this, the surveyors grid for that wave of land opening was structured to set up relatively large parcels (490 acres). On the western edge of that same land opening wave, land quality was considerably higher, but nevertheless parcels were set to the same size as they were around the swampy eastern part of the zone. Indeed, the boundaries for these land-opening waves on the western side were often determined with very little knowledge of land quality on the frontier. Each wave was typically negotiated near the settled area between representatives of the incumbent native tribe (who may not have actually represented the tribe at all) and functionaries of the state government. While the surveyors laying out the grid could in theory have adjusted parcel sizes on the fly to accommodate differences in land quality within the waves, this was not done.

Another example was the northwestern corner of the state, which was opened up following the eviction of the Cherokees in the 1830s. This area was opened up in two simultaneous waves. One of the waves (comprising the northern two thirds of the land opened in that wave) saw the land divided into 160-acre parcels, because it was thought that land quality was relatively high. The other wave was subdivided into 40-acre parcels, largely because it was believed that the area was filled with gold deposits. (There was a modest gold rush into this area, but it largely was a bust and quickly eclipsed by other discoveries farther west, including the famous Gold Rush of 1849 in California.)

Other than differences in timing, location, and parcel size, the waves of land opening was quite similar. Lotteries were used throughout; men satisfying minimal Georgia residency requirement were eligible;[4] and, importantly, there was no homesteading requirement. This latter feature meant that winners of land could 'flip' their property just as soon as they had won it. The announcement of the winners served to coordinate buyers and sellers into what was a fairly liquid market in the early years of the land opening (Weiman, 1991). But, while the single parcel itself might have been initially a liquid asset, changing a farm's size after the lottery would require the coordination of multiple land owners. (See Section 3.3 below.)

3.2 Neighboring states

As discussed above, the areas surrounding the Lottery Zone did not use a lottery to assigned property rights on a grid at the moment the land was opened for settlement. Here we review the land-allocation mechanisms used in the Buffer Zone surrounding the part of Georgia allocated by lotteries. There were two dimensions to these allocation processes relevant for making comparisons to the Georgia lottery zones: the mechanism through which parcels were specified and their boundaries were defined and the mechanism through which property was conveyed to individual landholders.

The original colonies (Virginia, North Carolina, South Carolina, and the northeastern third of Georgia in our case) and the lands allocated prior to 1785 (Tennessee and Kentucky in our case) used metes and bounds, the system of defining parcel boundaries (which could be highly irregular) by reference to natural and man-made features of the landscape, such as waterways, large boulders, old stone fences, stone markers placed at property boundaries, and even large trees (Libecap and Lueck, 2011). The situation in Alabama and Florida is complicated by the passage of control of parts of these areas through the French, Spanish, and British before they were incorporated into the U.S. Grants from Spain and France did not use rectangular surveys. In the areas settled after the lands were ceded to the U.S. and surveyed by the federal government in the 1820s and 1830s, previous claims based on

[4] Widows and orphans were also eligible for draws. Veterans were often given extra draws.

Spanish and French grants were 'carved out' of otherwise strictly rectangular federal survey system (Gates, 1954; Price, 1995). By the 1820s in Alabama and by the 1830s in Florida, the remaining land was being distributed through the federal government's rectangular survey system, with lots of 160 acres, and also 80 acre lots from 1866-68 (White, 1983).

The means through which title to the land reached the ultimate landholder followed roughly the system through which parcels were defined: in the original colonies and the areas settled prior to 1785 (all of the neighboring areas that form our Buffer Zone, except Alabama and Florida), land was distributed through the headright system under which an individual acquired a quantity of land proportional to the number of settlers an individual brought to the colonies. In Alabama and Florida, however, except for the original grants made by the French and Spanish and some headright grants in northern Florida by the British, land was distributed through a system of cash sales, much as in the other federal land areas opened to settlement prior to the Homestead Act (1862). After the Civil War, some of the remaining federal land in Alabama and Florida was provided through homestead grants, with the requirement that the property be occupied and improved before final title was conveyed to the grantee.

The lottery zones in Georgia thus differed from the Buffer areas in two important respects. In the lottery zones, property rights to all parcels of land in a zone were assigned to specific individuals at the time the lottery occurred and the area was open to settlement. In the Buffer Zone, land was placed under the control of individuals only when those individuals were ready to use the land, with many interstitial parcels remaining unclaimed in the decades following settlement. At the same time, Georgia's lottery zones had rectangular surveys with parcels that differed in size across zones, while the neighboring states had either very irregular parcel sizes defined under metes and bounds or (as in Alabama and Florida) a mix of a small number of irregular parcels from grants made by Spain, France, and Britain and a large number of rectangular parcels of uniform size (160 acres, with a small number of 80 acre parcels granted 1866–68).

3.3 Market-Design Issues

In this subsection, we discuss the possible difficulty of adjusting farm sizes, particularly in the Lottery Zone of Georgia, where the land grid was fully allocated to a diffuse set of owners from the outset.

For our purposes, the main issue in designing the land opening is whether it could get close to setting an optimal farm size at the moment of opening, or at least whether it could facilitate the transition to optimally sized farms later on. In a frictionless world, the design and outcome of the lottery has purely distributional consequences. Such is the insight of the Coase Theorem. If cheap enough, transactions will occur until the optimal distribution of farm sizes and the efficient mapping of farmers to plots are obtained. How much and for how long costs of transaction might impede this march to efficiency are nonetheless empirical matters.

Relatedly, much of the market design literature is focused the issue of what could be called "strategy-proofness," meaning that agents would not wish to misrepresent their preferences during the process. At a very basic level, the land lottery easily passes this test because the space of signals is so limited: (i) register or not for the lottery; (ii) make a claim or not if you should happen to win. (As discussed above, it was designed to be hard to manipulate.) Nevertheless, strategy-proofness is a goal not for its own sake, but rather because *ex post* trade is costly.

These land openings differ from some other contexts to which market design is applied. First, one side of the market (the land lot) lacks agency or preferences with respect to its owner. This is a less important difference in that (a) the farming production function is analogous to preferences over the match and (b) the government itself might prefer an efficient use of the land, which would not necessarily occur given the absence of prices at the initial allocation. (This logic is similar in the case of allocating cadaveric organs for transplantation.) Second, while land lots are by definition discrete (like kidneys), there is a continuum of ways to partition land into lots (unlike kidneys). In this sense, the allocation of land is like a two-dimensional version of the allocation of broadcast spectra, a case that has received some attention in the market-design literature.

The land-reallocation problem is complicated by the fact that the assets fill out a two-dimensional space (the Earth's surface). Exploiting scale economies in agriculture typically requires the land of a given farm to be contiguous, which creates strong complementarities among neighboring pieces of land. In the case of *individual* land lots, trade is costly, although not prohibitively so, especially when such costs can be amortized over a long enough time horizon. (This is not to say that the costs are trivial. Information asymmetries could be particularly acute in that there is much that is hard to observe about land quality and land improvements.) But suppose instead that the grid itself needs adjusting.

Even if Georgia politicians somehow got the optimal farm size correct at the initial opening, over time changes in prices and technology would cause this optimum to drift. Furthermore, the fact that the grid restricted parcel sizes to be homogeneous within a wave would imply that some areas (such as the non-swampy areas of the wave mentioned above) would be wrong from the outset.

Consider the coordination problem involved in shifting an area with an already-allocated grid to a different average farm size. The simplest case is when the optimal farm size drops by some integral factor N. If so, the existing owners can just subdivide into N lots. A more complicated case would be if the optimal farm size goes up by some integral multiple M. To achieve these new gains from scale, M neighbors would need to get together and negotiate which one of them buys out the rest. This situation presents a classic "holdout" problem in that any one of those M neighbors could threaten to sink the deal unless he receives an outsized share of the surplus.

A more realistic and still more complicated case would arise if the optimal farm size changed more slowly over time, say by a few percent in a decade. Consider first the simple example with bilateral negotiation. In order for my farm to get α bigger, my neighbor's has to get α smaller. But a quick analysis of the optimization problem illustrates why this would be difficult to negotiate. Suppose that I and my neighbor are symmetric, and let $\pi^*(x)$ be the profit function (profits after optimizing the input mix), conditional on farm size x. We assume the usual conditions on the profit-maximization problem such that there exists an optimal farm size x^* and $\frac{d^2\pi^*}{dx^2} < 0$, at least in a neighborhood around x^*. Define $\tilde{x} \equiv x - x^*$, the deviation from the optimum, and $\tilde{\pi}(\tilde{x}) \equiv \pi^*(x) - \pi^*(x^*)$, the loss associated with not being at the optimum. The quadratic derived from a second-order Taylor approximation of $\tilde{\pi}$ at $\tilde{x} = 0$ is

$$\tilde{\pi}(\tilde{x}) \approx a\tilde{x}^2 + b\tilde{x} + c$$

By construction of π^*, $c = 0$ and $a > 0$. By virtue of x^* being optimal, $b = 0$. Assume for simplicity that both farms have an $\tilde{x} = \alpha$. My gain from moving from $\tilde{x}$ to 0 is approximately $a\alpha^2$, while my neighbor's loss if doubling his $\tilde{x}$ is $(a(2\alpha)^2 - a\alpha^2) = 3a\alpha^2$, which is greater than my potential gain by a factor of 3. Thus, there are no gains from trade, at least local to x^* (that is, for $|\alpha| \ll x^*$), because my moving closer to the optimum raises my value less than what it costs him to move the same amount away from the optimum.[5]

Allowing for negotiation among more (symmetric) neighbors does not necessarily help, however. The curvature of the optimization problem means that the gains from getting closer will be smaller than the losses from getting farther away. For α large enough, all of a farm's neighbors could 'carve up' the farm in between. But this might be hard to pull off in such a thin market, again with the possibility of holdout problems.

None of this would be a problem if there were no need for a farm to be a contiguous piece of property. This would break the complementarity amongst neighbors, and as a result the market for added farmland would be quite thick from the perspective of any one farmer trying to expand. There were indeed some discontiguous plantations in the antebellum South (Ransom and Sutch, 1977). This was likely a work-around if the planter was unable to acquire an adjacent parcel of land, and may be an exception that proves the rule in that the planter was taking on additional transportation costs and perhaps reducing the economies of scale associated with more land rather than being 'held up' by the owner of the desired adjacent parcel.

Politicians of the era were already paying attention to what would be regarded as market-design issues today. For example, even though the vast majority of the territory was allocated in a rectilinear grid, town and university sites were preselected because easier it was deemed easier to do so *ex ante* than to than assemble them from various owners later (Holder, 1982). This idea was also used by the Township and Range system that was used to allocate land in most of the central part of the U.S..

Already the choice to use a lottery indicated taking stands on a market-design question. An even more transparent alternative would be to chop up each wave of land into n pieces if n persons register for the lottery. In the case of the Cherokee Land Lottery of 1832, the lottery for 160-acre parcels was oversubscribed by a factor of nine. Chopping up the land would have resulted in parcels of approximately 18 acres, which would have been well below what would have been considered the minimum efficient scale of the time. For any of this land to have been useful, therefore, such an initial allocation would have required a massive amount of land swapping and would likely have been hampered by coordination problems in the attempt to assemble large, contiguous parcels.

[5]Note that this is related to Harberger's notion that the loss associated with being distorted farther from the optimum is second order, and thus grows like the square of the deviation.

The state of Georgia also published lists (for example, Smith, 1838) to facilitate land claims and reduce transaction costs. These lists identified the lottery winners by names and reported both their parcel won and their county/township of residence at the time of the lottery. The point of the list was to not just announce the winners, but also connect potential buyers and sellers. There are various anecotes of prospective buyers going to great lengths to find the winner of particularly valuable (and complementary) plots. Indeed, Weiman (1991) presents evidence that the lottery increased liquidity in land markets in the immediate aftermath of each land opening. Nonetheless, while the lottery may have increased the degree of 'flipping' of lots in the early years, it is less clear that it had much effect on the structure of land lots imposed by the surveyors' grid. A survey of one county in the antebellum years showed the vast majority of land transactions occurring for lots that were precisely the same size as those from the surveyors' grid. While this may have represented consolidation of lots into plantations, it did not reflect a change in the fundamental grid structure.

Decades after the land opening, it may have been more difficult to locate a lot's owner, which presents challenges to making marginal changes in farm size on a fully allocated grid. Using the state-disseminated lists to find a winner who wanted to sell might be easy in the lottery's immediate aftermath, but those lists would become dated rather quickly in a period of high westward migration. Nevertheless, the low cost to claim land winnings motivated most lottery winners to take formal possesion of their lot. But even owners who undertook to develop their lot may have become impossible to find if they subsequently abandoned their land.[6] Roger Ransom (2005) dubbed this period the era of "walk away farming," in which it was commonplace for farmers to simply leave their land in response to bad shocks. Obviously you can not negotiate with a neighbor that you could not find, and such 'walk away' neighbors would retain *de jure* ownership of their lot for some time. A farmer who wanted to expand on to this abandoned land could do so with some hope of eventually acquiring title, but this was not without risk.

[6] In relation to the question of what fraction of winners actually took up residence on the long-term basis, we estimate in other work (Bleakley and Ferrie, 2012) that, 18 years after winning in the 1832 Cherokee Land Lottery, only 2%pt. of winners resided in the Cherokee area, relative to less than 15% in the control group.

4 Data

In the present study, we trace out the relationship between initial, assigned parcel size and land outcomes over time. This requires, broadly defined, two types of data. The first type consists of spatial data. We construct information on the location and characteristics of each wave of land opening, and we link this information spatially to data on administrative units (counties and minor civil divisions). The second type of data is aggregate information (most prominently average farm size and farm value) which we relate to the assigned parcel size from the land opening. Below we discuss the sources, construction, and linkage of these data in turn.

We code the boundaries of each land-opening wave of the lotteries using historical maps and textual sources. Our principal reference for the location of these boundaries starts with a map of Georgia with details on land-opening districts (Hall, 1895) and a book detailing the history of land surveying in Georgia (Cadle, 1991). For comparison, county boundaries for each census year are drawn from the National Historical Geographic Information System (NHGIS, Minnesota Population Center 2004) and major rivers in the "Streams and Waterbodies" map layer, from NationalAtlas.gov.

We started with the geo-referenced version of Hall's map. From there, and with textual confirmation from Cadle's book, we identified the vast majority of district boundaries in either the spatial data for rivers or for historical county boundaries. (As it happened, the footprint of land-district boundaries persisted in the county boundaries for at least a decade and in many cases to the present day.) In a few additional cases, a district boundary was a simple straight-line extrapolation of a linear county boundary. The majority of the additional boundaries not identified with the above methods were in the northwestern corner of the state. Serendipitously, we were able to find scanned and geo-referenced versions of the original surveyors' maps for that area.[7] For the remaining boundary segments, we relied on textual descriptions in Cadle (1991) which were occasionally supplemented with information on features from USGS Quad maps (for example, for the location within Wilkinson and Laurens Counties of Turkey Creek, a stream that was too minor to appear on the 'rivers and streams' file). As with any spatial data set, there are no doubt errors in the exact placement of each feature. Note nevertheless that such errors are unlikely to be important

[7] These correspond to the maps found in Smith (1838), and were downloaded from data.georgiaspatial.org.

because (a) the waves of land opening were quite large and (b) we are mostly working with data on counties, which are large relative to the errors that might be reasonably expected in our coding of the boundaries. In what follows, we refer to the area within Georgia that was opened up via lotteries as the "Lottery Zone" or LZ.

Across the Lottery Zone, waves differed not only as to when and where the land was open, but also by the initial assignment of parcel size. As discussed above, the parcel sizes were imposed by the surveyors' grid, which was essentially the same within a wave.[8] Across waves of the lottery, grid squares varied from 40 to 490 acres. We show the assigned acreages in Figure 1. The boundary of the Lottery Zone is denoted by the solid, red line (the innermost thick line). (The region outside of the LZ is discussed below.)

Inside the Lottery Zone, the background shading denotes the initial parcel size of the land opening. Darker shades indicate smaller parcel sizes. For reference, we also display 1900 NHGIS county boundaries, drawn in black lines. The darkest shading, corresponding to 40 acres per parcel, is seen in the northern part of the LZ. This area was opened by the "Cherokee Gold Lottery of 1832," discussed above. The lightest shade represents parcels of 490 acres each, and is seen in the southeastern part of the Lottery Zone (near the Okefenokee Swamp, as mentioned above) and also in particularly mountainous areas of the northeastern LZ. The second darkest shading is found in the northwest corner of the state and denotes acreages of 160 per parcel. The second lightest shading denotes 250 acres per parcel, and is seen in the southwest corner and just south of the northeast corner of the LZ. The middle shade denotes parcels of 202.5 acres and comprises the bulk of the middle section of the state's Lottery Zone. Note that the Fall Line[9] roughly bisects the 202.5 zone, and is thus quite far from most of the boundaries that we will use for identification of the models below. Note finally that a maximal set of contiguous land-opening waves that share a common initial acreage will be referred to below as an "acreage zone."

[8]The surveying methods of the time left something to be desired, and as a consequence there was some variance in the actual size of parcels. Most notably, there were the cases of the so-called "fractional parcels." Each district was divided into a series of subdistricts, which were then subdivided into individual parcels by the surveyors. If a surveyor reached the end of his row or column in the assigned subdistrict without enough space to fit a whole parcel, a fractional parcel was created instead. These could be as little as a third of the size of a standard parcel. Based on our own estimates from the Cherokee Land Lottery of 1832, somewhat less than 2% of parcels were fractional. Aside from these exceptions, parcels deviated very little from the acreage assigned for the whole district.

[9]The Fall Line divides the Piedmont from the Coastal Plain and also denotes the last point where rapids occur on rivers in the region. Phillips (1906) and Bleakley and Lin (2012), *inter alia*, discuss the importance of the Fall Line for regional urban development.

We next define a "Buffer Zone" as a comparison group. This zone corresponds to the area outside of of the LZ, but within 200 km of the LZ boundary. The 200-kilometer buffer is shown in Figure 1 with the short-dashed, red line. A 100-kilometer buffer is drawn with the long-dashed, red line. As discussed in Section 3.2, states intersecting the Buffer Zone did not use lotteries, but rather a combination of grants and homesteading. Some used similar rectilinear systems of land demarcation, but none allocated property rights to a broad set of owners at the moment of land opening, as was done in the Georgia lotteries. This buffer includes Georgia counties that were not in the Lottery Zone. It also includes counties in Florida, Alabama, Tennessee, Kentucky, Virginia, and the Carolinas.

Our goal in constructing the Buffer Zone is to form a comparison group for the Lottery Zone. To fix ideas, consider the western edge of the 40-acre zone in Georgia's LZ. A reasonable comparison area might be found in the counties just to the west in Alabama. These would have similar land quality, elevation, distance to the ocean (or to the fall line), etc. But they would differ in the method by which they were opened for settlement. Constructing our data following this idea, we assign an acreage to a point in the Buffer Zone according to the initial acreage of the closest point within the Lottery Zone. This extrapolation from the LZ generates spurious acreage assignments in the Buffer.[10] These assignments are also seen in Figure 1 by the background shading. Again, lighter colors mean larger assigned acreages, with the same color key as within the Lottery Zone.

Next, we construct the average assigned acreage at the county level. We use the NHGIS data, which defines U.S. County boundaries by decade starting in 1790. We construct an average assigned acreage for the counties in either the Lottery or Buffer Zones by averaging the above mentioned raster file within each counties county's boundaries.[11] We also computed the average distance[12] within each U.S. county to the Lottery Zone, and included in

[10]Specifically, we convert the acreage assignments within the LZ to a raster file with the resolution of 300 meters. Then, we use the "Euclidean Allocation" feature of arcGIS 10 to extrapolate these values into the Buffer Zone. An alternative extrapolation using inverse-distance weighting is shown in Appendix Figure 1. Note that below when we use rasters their resolution is 300m. We have explored using 30 meters instead and obtained essentially identical results when aggregating to the county or MCD level.

[11]This is done with the "zonal statistics" package within Spatial Analyst of arcGIS. In most cases, a county falls entirely within a single acreage zone (real or extrapolated). For those counties that straddle more than one acreage zone, this procedure constructs the area-weighted average of assigned acreages.

[12]We computed the average distance by first constructing a raster that coded the distance to the Lottery Zone boundary for each pixel. We used "zonal statistics" in arcGIS to take the average over the entire area of each county. The seemed preferable to a "spatial join," which instead takes the distance to the closest county boundary. Other references below to distances should also be read as being computed by this raster-average

our sample all those counties that were within 200km of the LZ. Counties that have at least 50% of their area in the Lottery Zone are coded as being within the LZ.[13] We also record the land area of each county and the predominant (or nearest) land-opening district. These are used below as weights and clusters, respectively.

Information on county-level outcomes was drawn from Census tabulations, as reported in ICPSR study 2896 (Haines, 2010). We compute consistently defined variables for average farm size (measured in acres) and average farm value (nominal values per acre, converted to natural logs). Note that the census notion of a farm was in reference to operation rather than ownership, and therefore the practical subdivision of many farms in the late 19th century into sharecropper and tenant farms is reflected in these data.

As an alternative data sources, we digitized information on minor civil divisions (MCDs) in Georgia for the year 1930. This is the first year for which we could obtain the approximate locations of these sub-county units. There are approximately seven MCDs per county in Georgia in that year. The higher spatial precision in the MCD data allows us to estimate models with regression discontinuities, because fewer MCDs straddle acreage-zone boundaries. We manually constructed a shapefile with the approximate centroids for each 1930 MCD using a scanned and geo-referenced census map (Census, 1934). We also transcribe the information on farm size and farm value aggregated by MCD from tabulations in Census (1931). The centroids of the MCDs are shown in Appendix Figure D, superimposed on maps of the acreage zones and counties in 1930.

5 Assigned Acreage and Later Farm Size, 1850–1970

How much does the initial assignment of farm size predict average farm size in the (almost) two centuries following land opening? In this section, we compare the initial parcel size that comes from the surveyors' grid to the county-level averages separately by year from 1850 to 1970. First, we employ a difference-in-difference strategy using the Buffer Zone as a comparison group in Section 5.1. Then, in Section 5.2, we look in greater detail at this relationship within the Lottery Zone. We find results that are generally consistent across methodologies and specifications, the latter of which include spatial controls, county fixed

method.

[13]Counties were almost all in either the LZ or the Buffer. The only exceptions were a few counties straddling the LZ and the Headright section of Georgia

effects, and samples that 'zoom in' to sets of counties proximate to acreage-zone boundaries. Initial, assigned farm size is a strong predictor of realized farm size in the 19th and early 20th century. Indeed, the initial predicts the realized size at a rate of one for one as late as 1880, the better part of the century after land opening. However, this relationship attenuates to a statistical and economic zero by the middle of the 20th century.

5.1 Difference in difference, Buffer versus Lottery Zone

Here we use the Buffer Zone as a control group to measure the effect of initial, assigned parcel size on farm sizes in the Lottery Zone in later years. The Buffer consists of places that are nearby, but whose land was allocated using the headright, grant, and direct-purchase systems, and was not assigned ownership at the moment the land was opened, unlike the Lottery Zone. Areas in the Buffer are coded with counterfactual initial acreages using the extrapolations described above in Section 4. Such extrapolation is of course a spurious assignment in that the Georgia surveyors did not construct a grid outside of the Lottery Zone. But, to the extent that areas just across the boundary of the Lottery Zone have similar land quality, the spuriously assigned acreage outside of the LZ might predict farm size and land value.

Thus, the identification assumption that drives the difference in difference can be motivated as follows. Suppose there exists a relationship between farm size and grid spacing within the Lottery Zone. This may be because the surveyors' grid was 'sticky'. But this might also be because each wave's parcel size was set in response to the optimal local farm size, which is itself persistent. If the Buffer Zone is an adequate control for neighboring areas in the LZ, this relationship should be present in the Buffer as well. In the LZ, any relationship between initial and later farm size over and above what is found in the Buffer might reasonably be ascribed to the persistent effect of the spacing of the surveyors' grid. This is therefore a difference-in-difference model, albeit with a continuous treatment variable (initial acreage). Note that the regression model allows us to include controls that can improve the comparability of nearby parts of the Lottery and Buffer Zones.

The regression equation that implements this strategy is as follows:

$$Y_{it} = \beta_t \left(Y_{i0} - \mu_{Y_0}\right) \times LZ_i + \alpha_t Y_{i0} + \gamma_t LZ_i + \delta_t + X_{it}\Gamma_t + e_{it} \tag{1}$$

where LZ_i is a dummy for being in the Lottery Zone, Y_{it} is county average farm size for $t > 0$, i indexes county, and t indexes time. The variable Y_{i0} is the assigned acreage, described above, while μ_{Y_0} is its sample mean.[14] The coefficient (α_t) on the main effect of assigned acreage (Y_{i0}) arises from the endogeneity of acreage assignment, as estimates of that coefficient use data in the Buffer Zone. The coefficient of interest is β_t, which measures the effect of the interaction between assigned acreage and being in the Lottery Zone. All coefficients are subscripted by t to allow for this relationship to change over time. In the default specification, the error term e_{it} is allowed to cluster by land-opening wave. All estimates are weighted by county area to account for the changing cross section of counties in each year.

5.1.1 Graphical presentation of estimates

Estimates of β_t are shown in Panel A of Figure 2. We separately estimate 13 sets of coefficients, one for each decade of Census data from 1850–1970. We present only the decade-specific coefficient estimates of β_t. Displayed in the graph are the point estimates (the solid line) and the 95% confidence intervals (the short-dashed lines). For reference, the vertical long-dashed line denotes the date of Lee's surrender at Appomattox (the end of the U.S. Civil War).

These estimates suggest that reallocation is characterized by quite a lot of rigidity. Consider first the estimate for 1850: the $\hat{\beta}_{1850}$ is close to one. This indicates that, if the assigned parcel size had been one acre larger, the average farm size in 1850 would also have been one acre larger. Note that 1850 was already two to five decades after the land openings. Even as late as 1880, we cannot reject one as a point estimate for β. It is evident in the graph that there was not a large jump downwards with the Civil War, although the attenuation of this coefficient appears to have begun in the 1870s. Again, note that this is in comparison to what was going on in the Buffer Zone.

As can be seen in Panel A of Figure 2, this pattern of one-for-one persistence can be rejected by 1890 in that the β coefficient has attenuated to below 0.8 and the confidence interval now excludes unity. The attenuation of this coefficient continues for the rest of the sample, although we can rule out a β of zero as late as 1930. The last year in the sample is

[14] Removing the mean of Y_{i0} before interaction has the effect that the estimates of α_t are evaluated at the mean value of Y_{i0}.

1970, for which we estimate a β close to zero and with a fairly tight confidence interval.

5.1.2 Alternate strategies for managing county-boundary changes

Estimates of equation 1 above come from the cross-section of counties, as they are constituted at the time. Thus the sample is different each decade as county boundaries change. To account for the changing nature of the sample above, we use county land area as a weight in the regression. This strategy means that any given piece of land has a similar weight throughout the years in the sample, although it might be represented via the average of a different jurisdiction at different points in time. In contrast, estimates of β from unweighted regressions might be destabilized if, as we add smaller counties through subdivision of existing larger ones, the newly sampled counties changed the correlation between initial acreage and farm size.

Another way to account for county-boundary change is to take boundaries from some base year and project the data from the other years as best we can onto the base-year counties. Then we can rerun the estimation of the decade-specific beta on a sample that has a constant composition of counties over time. To implement this idea, we created raster files that encoded the average farm size by county for each decade using the data described above. Then we took the average value[15] of each raster for all the counties in the base year, for which we choose 1900 as the approximate middle of the sample. This gives us a panel of farm size defined over 1850–1970 for the year-1900 set of counties in the Lottery and Buffer Zones. Note that the caveat 'as best we can' above refers mainly to the implicit assumption here that farm size is uniform within a county. The lack of historical data at the sub-county level requires that we make this assumption. We drop a county/year observation in the case of a large discrepancy[16] between the area of the 1900 county and the area over which the average is computed. This applies mainly to observations in the antebellum years.

[15]We performed this procedure separately in arcGIS and also QuantumGIS and obtained identical results, excepting minor numerical error. Further, we also checked the accuracy of this method by performing of "round trip" of the base-year data to a raster and then back to its original boundaries. The original versus the round-trip data in 1900 had a correlation of 0.999. Finally, note that we did not use "spatial join" procedures to conduct this harmonization of county boundaries, which may have been the first instinct of some GIS-inclined readers, for two reasons. First, note that the spatial-join procedure generates an unweighted average, which is clearly inappropriate this context. Our initial implementations of standard workarounds for constructing weighted spatial joins in some instances resulted in significant numerical error (of as much as 10%). Second, there seem to be some incompatibility in the early years between NHGIS files and the spatial-join routine in arcGIS that resulted in "topology" errors that we were unable to diagnose.

[16]Specifically, an absolute log difference greater than 0.5.

Yet another strategy is to take this sample of consistently defined counties and rerun the estimation, but also include fixed effects for the base-year counties. For this strategy, we jointly estimate equation 1 across all t and include a δ_i as a dummy variable for each county i. With the county fixed effects, we can no longer estimate a full set of β_t, so we set the β to zero for the years 1950–70, in accordance with the above estimates.

Estimates from these two alternative research designs are seen in Panel B of Figure 2, plotted alongside the estimates from Panel A above. For estimates of equation 1 using the sample of harmonized counties (the short-dashed line), we see that essentially the same picture emerges: assigned acreage is strongly predictive in the early years, but this result attenuates over the middle years in the sample. The shape of the curve is quite similar if we estimate the β_t controlling for county fixed effects (the long-dashed line). The pattern of statistical significance across years is also quite similar to the results in Panel A. (See Appendix Figure B for each series separately plotted with confidence intervals. The results in Panel C, in which the diff-in-diff estimates are separated into their single-difference components, are discussed in Section 5.2 below.)

5.1.3 Sensitivity analysis

Estimates of equation 1 are not sensitive to a series of alternative controls, subsamples, or methods for computing standard errors. These results are shown in Table 1. Estimates are displayed from various years (denoted in the row heading) and from various specifications (denoted in the column heading). Each cell displays the results from a separate regression. The first Column contains the estimates from the baseline specification, which match the values shown in Panel A of Figure 2. The estimate for β_t is close to unity in 1880, but has been cut to half of that by 1900. It then attenuates to a precisely determined zero by the middle of the 20th century.

We show in Columns 2–6 of Table 1 that these results are robust to accounting for a variety of spatial factors. In Column 2, we control for state$\times$year fixed effects. Columns 3 and 4 report results controlling for year-specific linear and quadratic[17] functions of latitude and longitude. Another approach to filtering out spatial trends is to zoom closer into the Lottery-Zone boundary (the thick red line in Figure 1). The buffer of 200 km may be so large

[17]The quadratic polynomial includes the linear and squared terms for both variables as well as the interaction of latitude$\times$longitude.

that the counties at the extreme edge of the Buffer Zone are not comparable. Instead we can zoom into 100km or 50km from the lottery boundary. This should reduce issues associated with the control group being inappropriate because of distance. As can be seen in Columns 5–6, coefficients when using this smaller sample are similar (within less than a standard error) albeit a bit smaller. (We have experimented with even smaller buffers, but these result in drastic reductions in the number of counties, and further have the undesirable property that often there will be no county at all on one side of the boundary. This latter feature is particularly undesirable given that our goal at the outset was to improve the matchup between control treatment groups. Unfortunately, our research design is not suitable to a matching estimator in that the treatment is not binary.)

The main exception is that the antebellum (1860) estimates lose statistical significance when zooming in to within 100km of the LZ boundary. This happens because, in the Buffer, the relationship between initial (imputed) acreage and farmsize is a bit higher and much less precisely estimated in the counties closer to the LZ boundary in the antebellum years. If we use instead an inverse-distance-weighting (IDW) method to extrapolate to the Buffer Zone, we also find this pattern of statistical significance in the postbellum but not the antebellum years. These results are found in Appendix Figure C, a replication of Figure 2 with an IDW rather than nearest-neighbor extrapolation into the Buffer.[18] If the Buffer Zone were indeed a good control group for the LZ, the antebellum pattern of coefficients suggests some combination of two scenarios. First, the Georgia grid spacing may have been set so badly that the LZ's β_t was adjusting downwards through the first half of the 19th century. Second, perhaps the grid was well calibrated to the correct farm size at the moment of land opening, but the farms in the Buffer were able to keep pace with changes in optimal scale while the LZ counties were not. Lacking information on farm size by county before 1850, we cannot sort out which of these two stories best fits the pre-1850 facts. Note that, in Section 5.2 when we compare within the LZ only, antebellum estimates are robust, even when restricting to counties near an acreage-zone boundary. This would seem to favor the "bad grid" hypothesis, at least with respect to neighboring areas that got assigned different initial parcel sizes.

Next, we show that these results are not sensitive to using alternative methods of inference. Recall that the baseline results use standard errors that are clustered on the lottery

[18]The IDW extrapolation in effect puts higher weight on Buffer counties close to the LZ boundary because it (disproportionately) reduces the variance of imputed acreage assignments in Buffer counties farther away from the LZ by averaging over multiple acreage zones.

wave (or the closest lottery wave in the case of the Buffer Zone). In Column 7, we report the heteroskedasticity-robust ('Hal White') standard errors instead. These are actually quite similar to the baseline standard errors. In Column 8, we instead use the state as a cluster, in which we find considerably smaller standard errors. (With only eight states in the sample, one might worry that the Central Limit Theorem has not yet kicked in.) Finally, in Columns 9 and 10, we use Conley's (1999) estimator, which allows for serial correlation within a given radius around each geocoded observation. Here we present results for 25 km and 50 km cutoffs, although we find similar results using 15km and 75km cutoffs. By construction, the point estimates are the same as Column 1; only the estimated standard errors differ.

The results above are not sensitive to dropping whole wedges of the Buffer Zone. We do this by state (in a some cases grouping states with few observations in the sample). For example, in Column 11, we show estimates from a subsample that excludes Alabama. The pattern of coefficients is quite similar to the baseline. We also find similar results in Columns 12–15 when dropping Florida; Tennessee and Kentucky; The Carolinas and Virginia; and the Headright part of Georgia.

The similarity of results when dropping different sets of control states suggests that the sluggish adjustment in the LZ is related to the assignment of all property rights in a lottery wave at the time it was opened, rather than the grid *per se*. Recall the two main features of allocation in the Lottery Zone discussed above: (1) the rectangular grid, and (2) the *ex ante* assignment of property rights. Indeed, this latter aspect was particularly extreme: essentially all of the land in the LZ was parceled out to a large number of persons at the time of land opening. Areas to the north and east of the Lottery Zone used metes and bounds for demarcation and a homesteading-like system for claiming land. In other words, these areas did not employ rigid grid nor could initial ownership be decoupled with use. In contrast, Alabama and Florida both used rectilinear grids to demarcate property and would sell the lots from the public domain as interested buyers presented themselves.[19] Both states were surveyed according to the Public Land Survey System (PLSS) in the 1820s and 1830s. Even though this system was a grid, it was *de facto* more flexible in the early decades because

[19]The exception of the Spanish land grants within Florida was discussed above. In any event, the grid system would have applied to a substantial fraction of north Florida (the part in the Buffer Zone). Further, only the extreme southern portion (roughly, south of the line extending west from the Fla/Ga border) of Alabama would have been subject to Spanish jurisdiction.

any given plot was not completely surrounded with parcels already claimed by others at the moment of land opening. Thus we see that the feature that distinguishes the LZ from all of its neighbors in the Buffer is the earlier assignment of property rights.

5.2 Single difference within the Lottery Zone

We now estimate the relationship between initial acreage and subsequent farm sizes using data only from within the Lottery Zone. This has a few advantages over the diff-in-diff strategy. First, there is no need for assumptions about how to extrapolate what the surveyors would have done outside the LZ. Second, a higher fraction of LZ counties are close to boundaries at which the assigned acreage changes. And, third, this exercise provides a more natural comparison with estimates in Section 6 below.

The regression equation is as follows:

$$Y_{it} = \beta_t' Y_{i0} + \delta_t' + X_{it}\Gamma_t' + e_{it}' \tag{2}$$

in which the notation is the same as above for equation 1. Because the comparison now is among areas in the Lottery Zone, the coefficient of interest is β_t', the effect of initial acreage assignments on average farm size in later decades. The 'primes' on the coefficients indicate that these parameters could conceivably differ from their counterpart in equation 1. Note that there can also be a series of controls (X_{it}), although the most basic specification includes just the assigned parcel size and decade-specific constant term. As above, we use the predominant wave of land opening in the county as a clustering variable when computing the standard errors.

Estimates of β_t' are found in Table 2, which is similar in structure to Table 1. The baseline results are found in Column 1. The pattern of coefficients is quite similar to those estimated above in the double-difference specification, although perhaps the attenuation is a bit slower.

These estimates of β' for the full set of decades can also be seen in the thin, upper line in Panel C of Figure 2. With the exception of 1850, these estimates are quite similar to the estimated coefficients of the difference-in-difference model displayed in the other Panels of that Figure. As above, we cannot reject a coefficient of unity for any year before 1890. The difference in results for 1850 is understood (at a superficial level anyway) by inspecting the other line (the thick black one) in Panel C. This line presents the coefficients that come from estimating equation 2 using the Buffer Zone instead of the LZ. (Note that, by construction, the double-difference coefficients are just the gap between the two lines in Panel C.) Apart from the antebellum years, the relationship between the assigned acreage and eventual farm size is basically zero in the Buffer Zone.

The rest of Table 2 presents sensitivity analysis for the estimation of equation 2. For Columns 2–9, we code areas based on their proximity to boundaries between the acreage zones[20] in the LZ. We also classify these boundaries according to whether they are either natural (formed by a waterway or ridge, principally) or rectilinear and by the difference in acreage on either side. We also add to the specification dummy variables for the different categories of acreage change at the nearest acreage-zone boundary.[21] In Columns 2 and 3, We use subsamples of counties that are within 50 and 25 km, respectively, of an acreage-zone boundary. The pattern of coefficients over time is quite similar to the baseline. (Note that results in 1860 are not materially sensitive to this 'zoom in' check, unlike the results with the Buffer Zone as a comparison group in Table 1.) In Columns 4–9, we decompose the sample into whether the closest a acreage-zone boundary is rectilinear or natural and perform the same subsampling by boundary proximity. The pattern of high coefficients and then attenuation over time is seen here, just as in the baseline.

In Columns 11–12 of Table 2, we check for the sensitivity of the results to alternative formulations of the standard errors. As above, we find essentially similar results when we use heteroskedasticity-robust standard errors or Conley's spatial adjustments with cutoffs of 25 and 50km. (Unlike above, we cannot cluster at the state level, because the LZ is contained entirely within one state.)

A final feature of Table 2 worth mentioning is the parallel use of MCDs in the 1930 specification. We manually transcribed agricultural data for the MCDs and geocoded their centroids. (Such data were not available for earlier years.) Estimates of equation 2 given are similar to those obtained using the 1930 county sample instead. The coefficient of approximately 0.1 means that one additional assigned acre at the opening phase translates into 1/10 of an acre by 1930.

[20]There were many boundaries between waves of land opening that did not involve changes in acreage. For the construction of the acreage zones here, we ignore such boundaries.

[21]There are seven such categories of acreage changes at acreage-zone boundaries in the LZ.

6 Regression-Discontinuity Estimates, 1930 MCDs

In this section, we use the discontinuity in assigned acreage at boundaries among acreage zones to estimate the effect on farm sizes. We utilize the 1930 MCD data, which measure outcomes at a higher spatial resolution than the county data. (Appendix Figure D gives the precise locations of MCDs in relation to the Lottery Zone.) As discussed above, 1930 was the first year for which we could obtain geocoded data on farm sizes for MCDs. In theory, we could have used this regression-discontinuity (RD) design with county data, but the sample sizes were too small in practice to obtain precise results. Notice that the exercise with MCDs in 1930 is itself something of a weak test insofar as estimates above imply a considerable degree of attenuation by 1930. Nevertheless, we do find some areas for which the effect of initial assignment has not completely gone away.

We first compute an estimate from the whole sample with all of the acreage-zone boundaries within the LZ. To fix ideas, let x be the distance to the closest acreage-change boundary, and let d be the running variable for the RD design. If, at the closest acreage-change boundary, the change in acreage is negative, then we assign $d \equiv x$. If instead the acreage change is positive at the closest acreage-change boundary, then $d \equiv -x$. This has the effect that, when moving from left to right in the usual RD graph (i.e., d on the x axis), there is the jump *up* in assigned acreage at $d = 0$. We take the MCD data set discussed above and estimate the RD model using the running variable just defined. A locally smoothed polynomial average is used to estimate the relationship between farmsize and d. This average is computed using a triangular kernel and the optimal bandwidth. We use code described in Nichols (2007). As above, we weight by the MCD's land area.

Note that the RD design is an alternative method of estimating equation 2, albeit with a flexible control for d on either side of $d = 0$. At the RD, the ratio of the jump in farmsize to the jump in assigned acreage is an estimate of β'_{1930}. The resulting Wald/IV estimate of β'_{1930} from the RD is 0.08. The 95% confidence interval for this parameter excludes zero, but includes the estimated β in 1930 obtained from the difference-in-difference model. In economic terms, the estimated β'_{1930} here is essentially the same as what we found above for MCDs in 1930 (shown in Table 2).

As an alternative approach, we examine specific acreage-change boundaries. We focus here on the two boundaries where the change in acreage is largest: between 40 and 160 acres in the northwestern LZ and between 490 and 202.5/250 acres in the south/southeastern LZ. These two areas are discussed in detail in Section 3.1 and Section 4. Results from this exercise can be seen in Figure 4, which displays the data and a smoothed average for farmsize (the y axis) versus the running variable d (the x axis). The sample for each Panel is limited to the indicated acreage zones, and the x and d variables are constructed from distance to the specified boundary, regardless of whether another acreage-zone boundary is closer. (Acreage zones that are no adjacent to the specified boundary are excluded from the sample, even if their assigned acreage matches one of those of the particular boundary used.) The vertical line denotes a distance d to the boundary of zero. The thin black horizontal lines indicate the assigned acreages from the surveyors' grid. Again, when passing through zero in each graph, there is a discontinuous jump up in initial, assigned acreage. The MCDs from the smaller acreage zones are assigned negative d and thus occupy the area to the left of zero in each graph. Each dot denotes an MCD, unlike some RD graphs in which the dots are averages. (In our case, the number of observations was small enough that we could accommodate this information on the graph.)

The solid and dashed gray lines are the local-polynomial smoothed average of the MCD data, estimated separately for $d < 0$ and $d > 0$, respectively. This average is computed using a triangular kernel and the optimal bandwidth. Again, we use code described in Nichols (2007). Panel A considers the boundary on the southeastern quadrant of the lottery zone that divides the 220 and 250-acre zones from the 490-acre zone. Panel B considers the boundary on the northwestern quadrant of the lottery zone that divides the 40-acre zone from the 160-acre zone. At conventional levels of confidence, we cannot reject that the jump at the RD is zero for Panel A, but we can reject a zero jump for Panel B. Furthermore, the Wald/IV estimate for the 40/160 case is approximately double what we found above in the full-sample RD.

7 The Costs of Initial Misallocation

Now we turn to the question of misallocation. We attempt to measure the cost of the market design imposed by the Georgia land-lottery system. This in turn hinges on the curvature of the production function with respect to farm size. In the case of strong curvature, there is a well-defined optimum, such that pushing a farmer away from the optimal farm size is costly.

In contrast, if all that these historical accidents do is shift among perfect substitutes (such as the case of constant returns to scale in farm production), there would be little point in worrying about this kind of persistence. Here we attempt to measure the actual cost of the friction that arose from the opening the Lottery Zone with the wrong market design.

To accomplish this, we work with estimates of the squared deviations of LZ farm sizes from the optimum, in what we call the gap-squared model. This model is a system of equations. The first equation just repeats the difference-in-difference equation from Section 5.1. The next equation uses the previous one to impute a gap between current and optimal farm sizes for the counties in the Lottery Zone. Finally, in a third equation, we take the square of this gap variable and add it as a regressor in a version of equation 1 with the log of farm value on the left-hand side. Estimation of the gap-squared model suggests that farm values were depressed by the persistent misallocation induced by the lottery process.

To define a gap, first we need a comparison area that tells us about the optimum. As before, we use the Buffer Zone for this purpose. These areas might not themselves be at the optimum. Nevertheless, by the logic of the Envelope Theorem, farm value for the controls would be a decent indicator of optimized value so long as their farm size is not significantly different from the profit-maximizing scale. Also, we can show that our results are not sensitive to including different controls that help us estimate the gap. Further, we saw above that transitory dynamics associated with extrapolated acreage assignment had dissipated quite early in the Buffer Zone. This suggests that those areas were close to their equilibria, at least when measured along that specific dimension.

Therefore the gap measures are relative to the Buffer. We estimate this using equation 1 once again. Recall that the decade-specific coefficient β relates farm size to initial acreage. We modify the formula for fitted value to allow for an arbitrary $\hat{\beta}$ in equation 1. Let this pseudo-fitted value for farm size in county i at year t be $\hat{Y}_{it}(\beta)$. The estimated gap is therefore

$$gap_{it} = \hat{Y}_{it}(\hat{\beta}_t) - \hat{Y}_{it}(0) \quad (3)$$

$$= \hat{\beta}_t \left(Y_{i0} - \mu_0\right) LZ_i \quad (4)$$

such that the gap is zero in the Buffer Zone. Within the Lottery Zone, the gap depends on assigned acreage and the estimates of β.[22]

[22]Note that the generally positive estimates of β do not imply that the gap itself is positive. Instead it

We work with the squared version of gap_{it} because basic optimization theory says that the loss from a failure to optimize rises like the square of the misallocation. Consider Harberger's eponymous triangles. As he points out, deadweight loss rises with the square of the distortion because, when viewed graphically, it is the area of a triangle. A more mathematical justification comes from a Taylor expansion about the optimum, after an application of the Envelope Theorem. This was seen above in Section 3.3 in the discussion of coordinating shifts to newly optimal farm sizes in a grid of farms. (This is standard reasoning, so we do not repeat that analysis here.)

In the third equation of the gap^2 model, we regress county farm value on the imputed gap^2. Farm value is measured originally in nominal terms per acre, but transformed into natural logarithms. Coefficients in this regression can be interpreted in percentage terms, and the year dummies will absorb differences in the price level over time. The equation is as follows:

$$\log V_{it} = \omega\ gap_{it}^2 + \beta_t \left(Y_{i0} - \mu_{Y_0}\right) \times LZ_i + \alpha_t Y_{i0} + \gamma_t LZ_i + \delta_t + X_{it}\Gamma_t + e_{it} \tag{5}$$

This equation has all the same terms as the equation 1, except for having gap_{it}^2 as a control and the log of farm value (instead of farm size) as its dependent variable. Because gap^2 is a generated regressor, we block-bootstrap this entire system of equations using the county as a cluster. As an alternative, we also present standard errors by simply clustering the errors in equation 5 at the county level. The coefficient of interest is ω, the relationship between log farm value per acre and gap^2.

In Table 3, we present estimates of the gap^2 model. Here we display only the estimated ω in the farm-value equation. These coefficients are normalized to be interpreted as the effect at

is evaluated relative to the sample mean of assigned acreage (around 200 acres). An alternative method for constructing the gap that we tried involved setting not β but LZ to zero, which results in

$$gap'_{it} = \hat{\beta}_t \left(Y_{i0} - \mu_0\right) LZ_i + \hat{\gamma}_t LZ_i.$$

Because $\hat{\gamma}$ tended to be small, these two gap variables were highly correlated. One complication with both of these variables is that they impose a single cut-point where the gap is zero. To check the importance of this, we defined a third version of the gap variable, Z, which is the residuals of a regression of $\left(Y_{i0} - \mu_0\right) LZ_i$ on the other control variables in the model. Then the $gap'' \equiv \hat{\beta}_t Z_{it}$. This variable was highly correlated with the other two measures of the gap. Results below are similar using the alternative gap measures. We also attempted, but abandoned, a strategy of estimating the gap using the within-county dynamics of farm size (specifically with the implied steady state of an AR(1)) model, but the estimates of gap were quite sensitive to specification and complicated by the difficulties in estimating AR(1) models with near-unit roots.

the mean of gap^2 in 1880. We do not have a strong theoretical presumption has to whether farm size should enter in this system of equation as levels or logs, so we present results for both in this table, in Panels A and B, respectively.[23] Consider first the 1880 results, seen in Rows 1–3. Estimates indicate that a county with the mean level of gap-squared in 1880 had a land value that was depressed by 11-16%. These estimates are statistically significant at conventional confidence levels. The baseline estimates, shown in Column 1, use all the same controls as the usual difference-in-difference equation, as indicated across the first row describing the specifications. The results are broadly consistent as we add controls for latitude/longitude and state dummies. For the 1880 sample, we cannot use county fixed effects, because there is only one period. But the rest of the table uses a longer panel. All of the variables that appeared in the 1880 model will be included in turn for the panel data, but will appear in the specifications interacted with year for flexibility. These models also allow for county fixed effects (in Columns 7 and 11). As can be seen in the table, the estimates are broadly similar using these different controls and samples. The highest and lowest point estimates in the table are approximately 11% and 22%, respectively.

In Figure 3, we consider how these estimates change by decade. Panel A presents the estimates of mean of gap^2 over time. The average gap^2 declines throughout the sample, in keeping with the declining β. Panel B displays regression estimates of ω. (To facilitate comparison with Panel C, ω is not normalized, unlike above.) The horizontal lines display the point estimate and 95% confidence interval estimated in the 1880–1950 panel for comparison, while the dots indicate the point estimate (within its associated confidence interval) separately by decade. Up through the first decades of the 20th century, the estimates of the coefficients are fairly consistent with each other. Later in the 20th century, the point estimates are quite erratic and the confidence intervals are enormous (so large as to be clipped in this graph). This is because, by then, there are basically no gaps with which to estimate the model, as the β coefficient have gone to zero. One can only estimate a model with gap^2 if there are gaps, which have basically disappeared in the latter part of the sample.

The final step, seen in Panel C, is to compute the cost of misallocation. We do this by multiplying each county/year's gap^2_{it} by the cost of gap^2 (the $\hat{\omega}$ parameter) estimated from the panel data. As above, we use a block bootstrap using the county as the cluster. To be

[23] Above, farm size was always in levels. Also, it was always on the left-hand side, but here it is being used to construct a right-hand-side variable.

more specific, note that we are again bootstrapping the entire system of equations, including both the panel estimates of ω and the calculations of the mean cost of misallocation. The general pattern of attenuation is similar to that seen in Panel A (and in Figure 2, although in this case it starts from a negative number because we estimate a reduction in value. We can also see that the average county in the Lottery Zone experienced a reduction in its land value of approximately 20% in the latter decades of the 19th century. Note again that this cost was present 50 to 80 years after the lottery. Further down the road, the estimated cost drops to just a few percent by 1920.

Revisiting the RD framework with farm value as an outcome, we find confirmation (or at least not falsification) of the gap^2 model. Specifically, do we see differences in farm values at the acreage boundaries examined in Figure 4? We address this question in Figure 5. First, in Panel A we consider the southeastern boundary of 202.5/250 versus 490 acres. This comparison was a 'dog that did not bark' above in that the 1930 differences in farm size had already attenuated to zero. As a result, we should not expect there to be differences in farm value across the boundary, and we do not see them in the graph. (A test of a discontinuity at that point has a p-value of 0.7.) On the other hand, we did see a difference in farm sizes along the northwestern 40/160-acre boundary. Accordingly, in Panel B, we see that there is also a jump in farm value along that boundary.

We compare estimates in both RD figures to generate a Wald estimate of the ω coefficient from the gap^2 model. The ratio of the effect sizes between farm size and log farm value suggest a coefficient on gap^2 of approximately three—higher than what comes out of the county-level regressions. (Compare with Panel B of Figure 3, for example.) Whether we can compare these numbers is uncertain, however. Implicitly this calculation uses the 40-acre zone (which has the higher land value) as a benchmark for the optimum, and this might not be appropriate. Both zones were constrained by the surveyors' grid, and both zones might yet be out of equilibrium. If both are converging to some common acreage in between, then the comparison at the discontinuity would understate the loss from misallocation in that farm sizes on both sides are down from the top of the curve. If both zones are converging to some average farm size that is below the 1930 averages, then the difference in values overstates the effective gap-squared when considered relative to the optimal farm size. On the other hand, the estimates above from the county-level data might be biased by measurement error in the gap if the buffer counties do not adequately measure the counterfactual for nearby counties in the lottery zone. To the extent that these errors are classical noise, we would expect the estimates for the effect of gap squared above to be biased downwards relative to our observation here with the RD design.

In any event, while it might be easy to criticize the earlier diff-in-diff estimates of the gap^2, it appears from the RD evidence that the double-diff estimates are understating the cost of misallocation, if anything.

8 Conclusions

The Georgia land lotteries represent a unique environment in which to assess the effect of the initial allocation of an asset on the subsequent allocation of the asset through the marketplace. The concern raised in the market design literature that initial allocations can have lasting effects of the efficiency of the market's operation appear to be borne out in the short to medium term: by a variety of measures (difference-in-differences, regression discontinuity) it appears that average farm sizes in Georgia were strongly influenced by the parcel size at the time the land was initially distributed. This might not represent an inefficiency if the minimum efficient scale for farming did not vary across a broad range of farm sizes. But in comparison counties in surrounding states, parcel sizes were considerably less "sticky." Average land values suggest that farms in the Georgia lottery zones were less efficient than in comparison counties, with a differential of 20% as late as 1880, nearly eight decades after the first lottery. In the long run—after 150 years—the effect of the parcel size at allocation has dissipated, suggesting that getting the initial allocation "wrong" can impede the movement of the market toward an efficient allocation for a considerable length of time.

References

Arrow, K. J. and G. Debreu, G. (1954). "Existence of an Equilibrium for a Competitive Economy." *Econometrica*, 22: 265-90.

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

Coase, R. H. (1960). "The Problem of Social Cost." *Journal of Law and Economics*, 3:1-44.

Dell, M. (2010). “The Persistent Effects of Peru’s Mining Mita.” *Econometrica*, 78: 1863–1903.

Edgeworth, F. Y. (1961). *Mathematical physics; an essay on the application of mathematics to the moral sciences.* Kegan Paul, 1881. New York: AM Kelley.

Gates, Paul W. (1954). “Research in the History of American Land Tenure: A Review Article.” *Agricultural History*, 28:121-126.

Hahn, R. W. (1984). “Market power and transferable property rights.” *The Quarterly Journal of Economics*, 99(4), 753-765.

Kitch, E. (1977). “The Nature and Function of the Patent System.” *Journal of Law and Economics* 20:265

Libecap, G. D., and Lueck, D. (2011). “The Demarcation of Land and the Role of Coordinating Property Institutions.” *Journal of Political Economy*, 119(3), 426-467.

Libecap, G.D., and D. Lueck (2011b). “13 Land demarcation systems.” *Research Handbook on the Economics of Property Law*, 257.

McCloskey, D. (1998). “The So-Called Coase Theorem” *Eastern Economic Journal*, 24: 367-371.

Merges, R. and R. Nelson (1990). “On the Complex Economics of Patent Scope.” *Columbia Law Review*, 839.

Milgrom. P. (2011). “Critical Issues in the Practice of Market Design.” *Economic Inquiry* 49: 311–320.

Myerson, R. B., and Satterthwaite, M. A. (1983). “Efficient mechanisms for bilateral trading.” *Journal of Economic Theory*, 29(2), 265-281.

Noll, R. G. (1982). “Implementing marketable emissions permits.” *The American economic review*, 120-124.

Price, E. T. (1995). *Dividing the land: Early American beginnings of our private property mosaic* (Vol. 238). Chicago: University of Chicago Press.

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

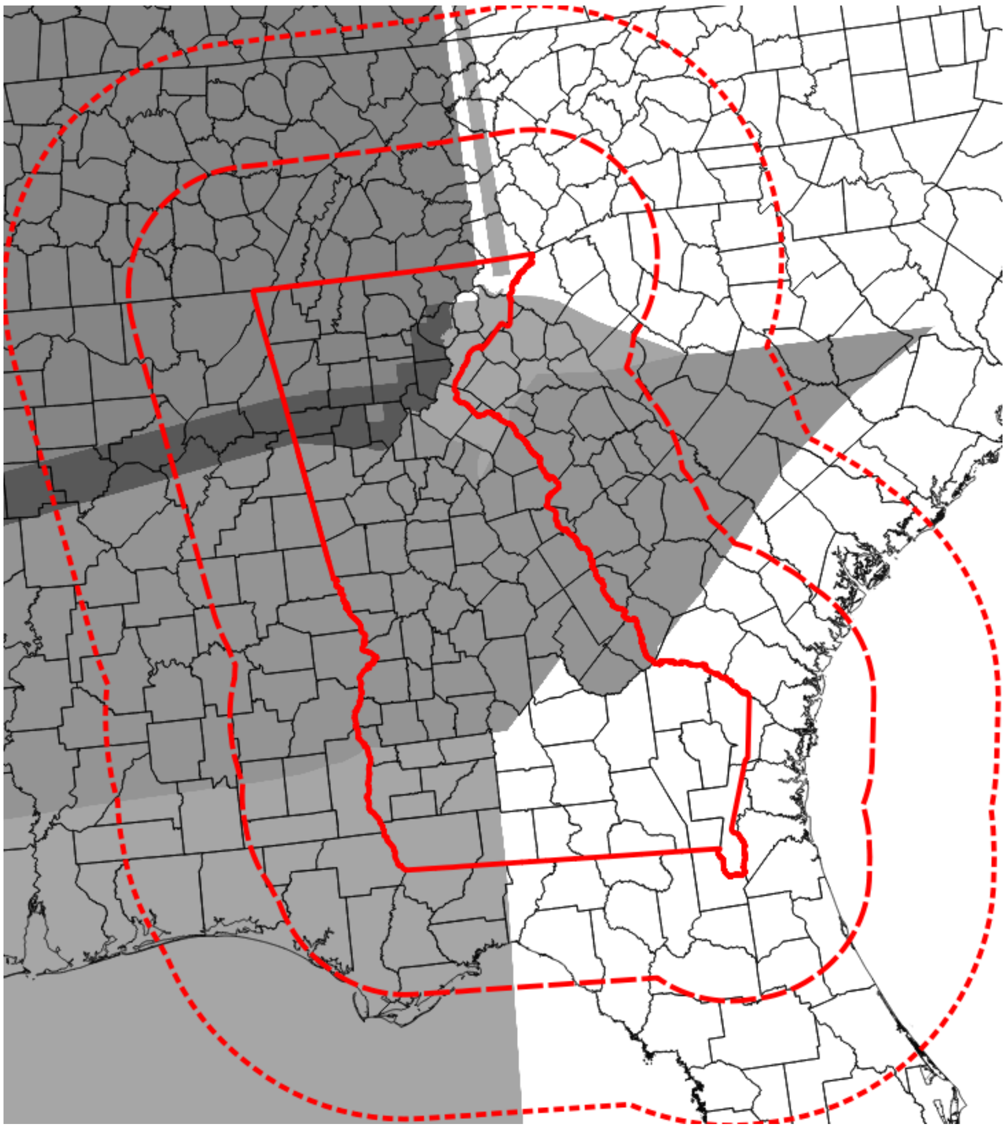
Smith, A. (1937). *The Wealth of Nations (1776).* New York: Modern Library.

Smith, J.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 & Brothers.

Weiman, D. 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.

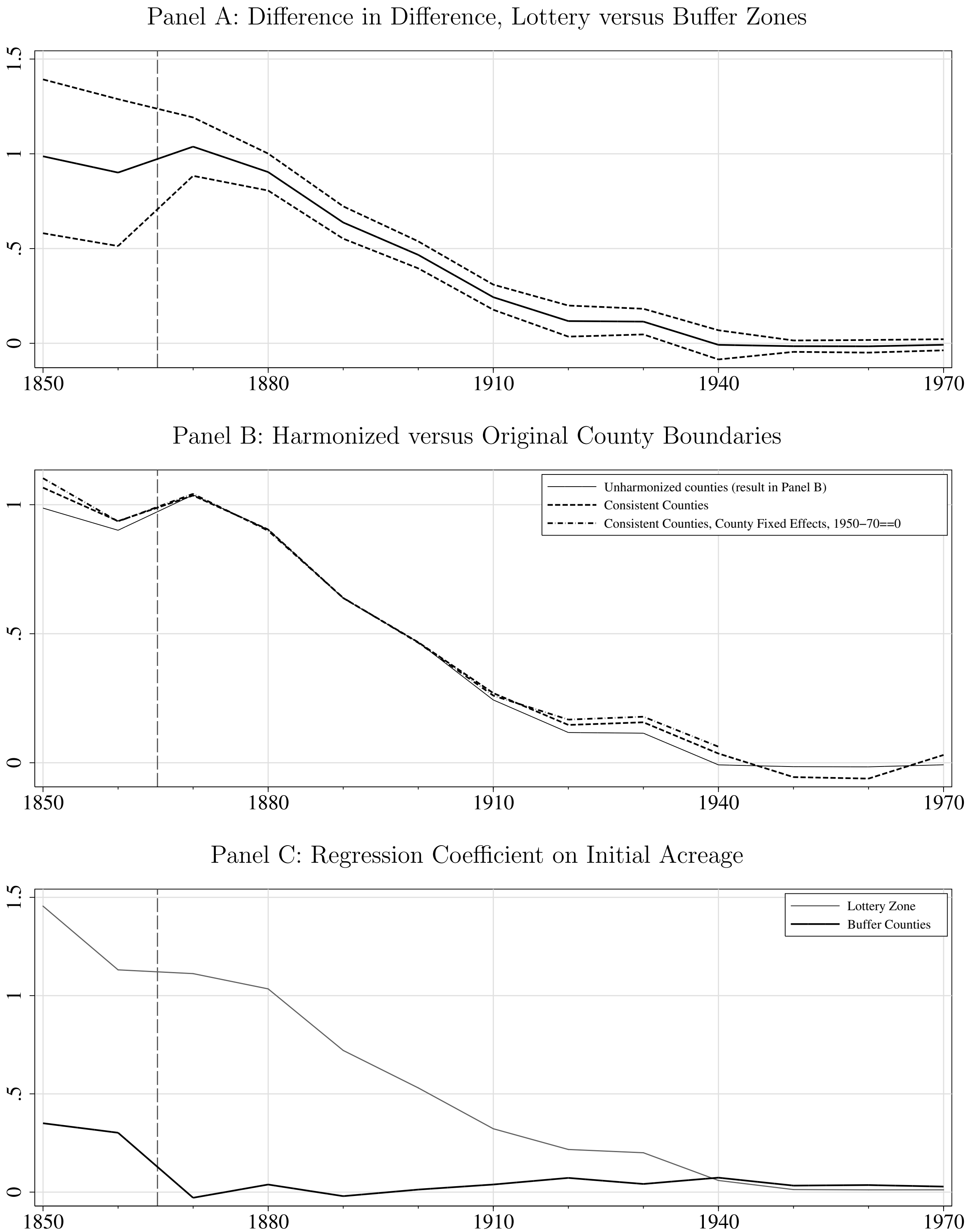
White, C. Albert. *A history of the rectangular survey system.* Vol. 2. U.S. Dept. of the Interior, Bureau of Land Management, 1983.

Figure 1: Lottery Zone, Initial Parcel Size, and 1900 County Boundaries



Notes: this map displays various features in and around the Georgia lottery zone. County boundaries (as defined in 1900) are denoted in black and are drawn from NHGIS data. The solid red line denotes the outer boundary of the Georgian lottery zone. The dashed and dotted red lines denote, respectively, a 100 km and 200 km buffer around the lottery zone. Inside the lottery zone, the background shading denotes the initial parcel size. Outside of the lottery zone, the background shading denotes the initial parcel size of the closest point within the lottery zone. Darker shades indicate smaller parcel sizes. See the text and data appendix for further details and definitions.

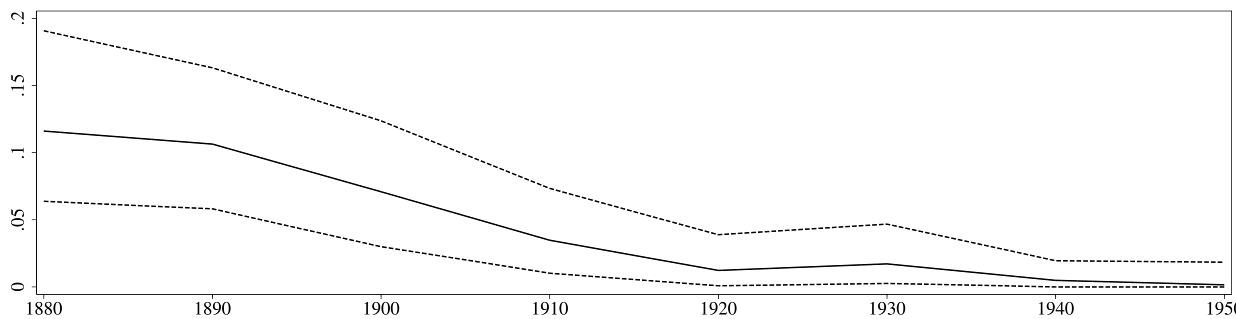
Figure 2: Farm Size versus Initial (Assigned) Acreage, by Decade



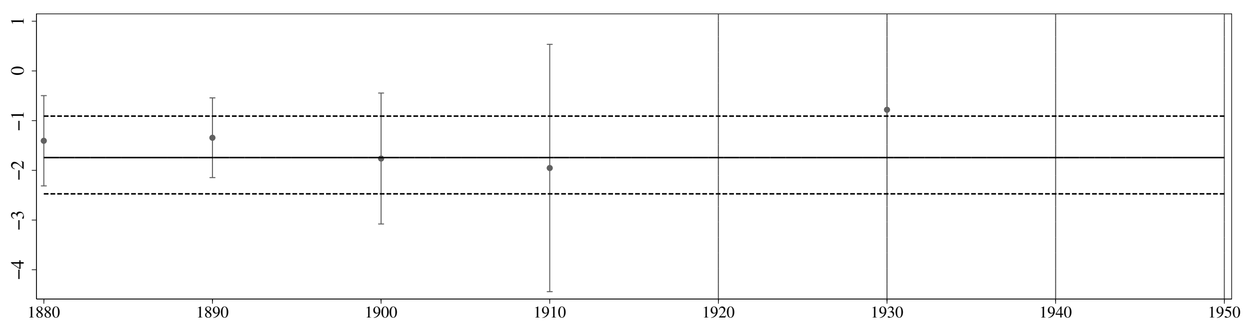
Notes: These graphs present decade-specific coefficient estimates of the relationship between farm size and initial parcel size assigned by the surveyors' grid in the land opening. The vertical dashed line in each Panel denotes the end of the Civil War. Panel A presents estimates from equation 1, with the 95% confidence intervals displayed by the dashed lines. Estimates are weighted by county area to account for the changing cross section of counties in each year. Standard errors are clustered by lottery wave. Panel B compares the point estimates from Panel A, estimated from the cross-section of counties in each decade, with similar results from a panel of consistently defined county data. The consistent counties are constructed by harmonizing the data onto the 1900 county boundaries. The dashed line displays estimates of equation 1 for this panel of consistent-boundary counties. The dash-dot line also displays estimates of the equation 1 for the consistent counties, but pools the decades and includes county fixed effects in the specification. The β_t in this case are normalized to zero for the years 1950–70. Panel C presents estimates of equation 2 for the lottery zone in Georgia and the nearby comparison ("buffer") areas. (Note that the difference in difference in Panel A is equal to the gap between the curves in Panel C.)

Figure 3: Estimates from the Gap^2 Model by Decade

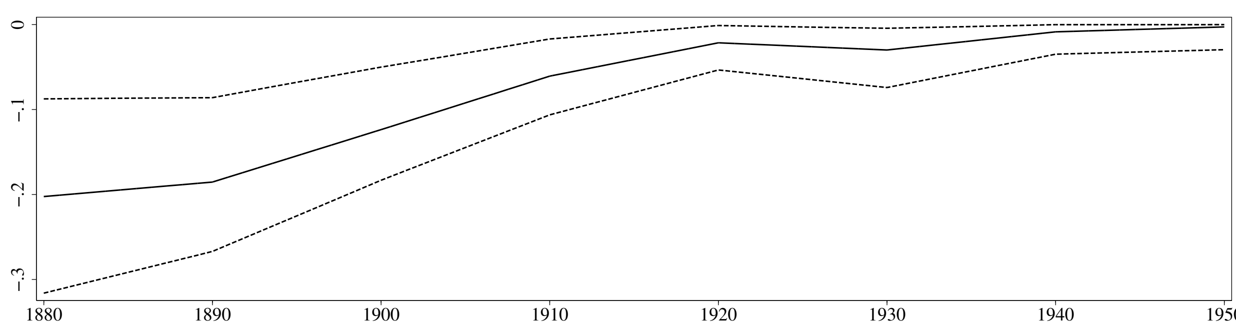
Panel A: Average Gap^2 in the Lottery Zone



Panel B: Regression Coefficient of Log Farm Value per Acre on Gap^2



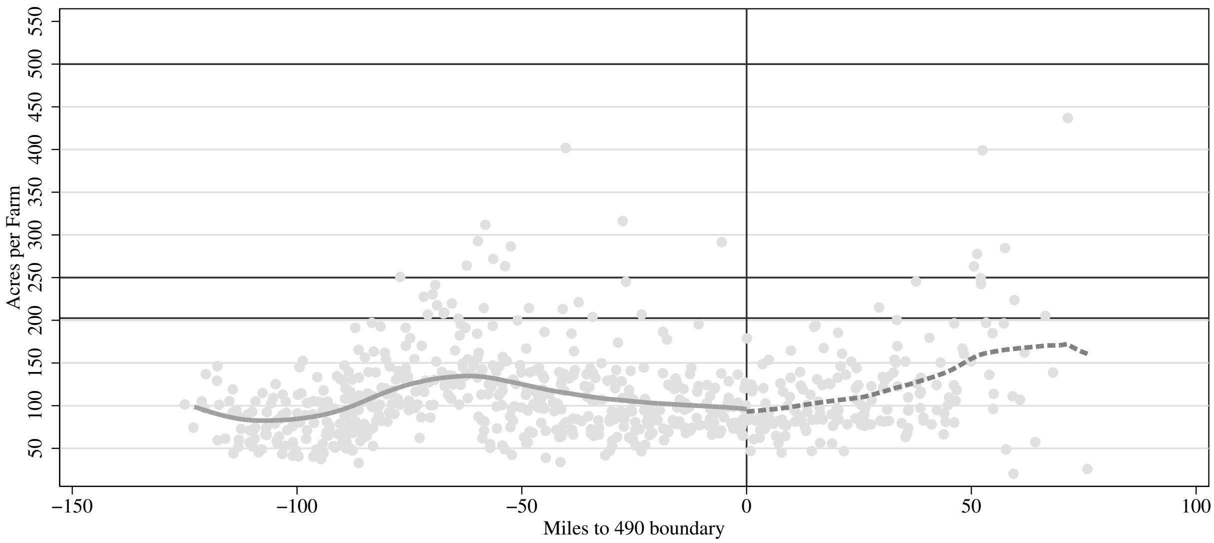
Panel C: Average Percentage Loss in the Lottery Zone from Misallocation



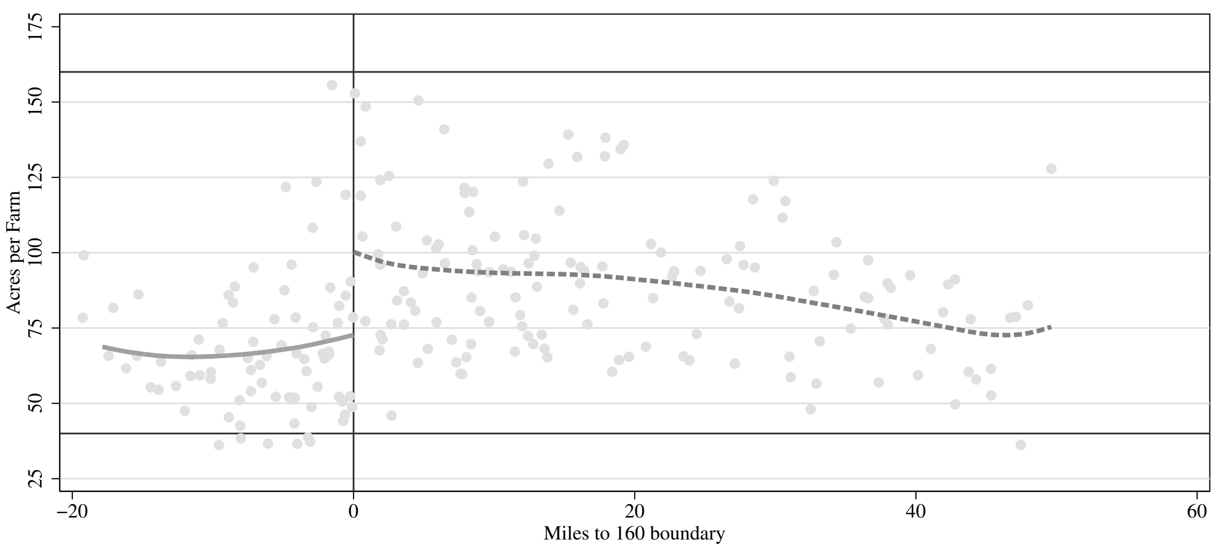
Notes: These graphs present various estimates of the gap^2 model, estimated from equations 3 and 5. The underlying data are the "consistent county" sample described in the text, with data harmonized to 1900 county boundaries. This particular sample is needed here because all of the estimates are block bootstrapped using the (consistent) county as a cluster. Throughout the three panels, the dashed lines graph the 95% confidence intervals of the mean (computed from the percentiles of the bootstrap results). The solid line in Panel A graphs the average gap^2 in the lottery-zone counties, by decade, estimated with equation 3. The solid line in Panel B denotes the estimated effect of gap^2 on the natural log of farm value per acre estimated in the OLS/FE model of equation 5. The dot (and whiskers) in Panel B are the decade-specific estimates (and 95% confidence intervals) of equation 5 using OLS on the cross section for each census year. (Note that, unlike those in Table 3 these estimates of ω are not normalized here to facilitate comparison across panels.) Panel C contains the decade-specific estimates (and 95% confidence intervals) of the average loss (in natural logs) from misallocation in the lottery zone. This is computed by multiplying each counties' estimated gap^2 by the estimates of ω from the OLS/FE estimate of equation 5. The confidence interval is computed by block bootstrapping both equations and the average of the loss simultaneously.

Figure 4: Farm Size versus Distance to Assigned Acreage Change, Select Zones, 1930

Panel A: Boundary in Southeastern Georgia between 220/250 and 490 acres



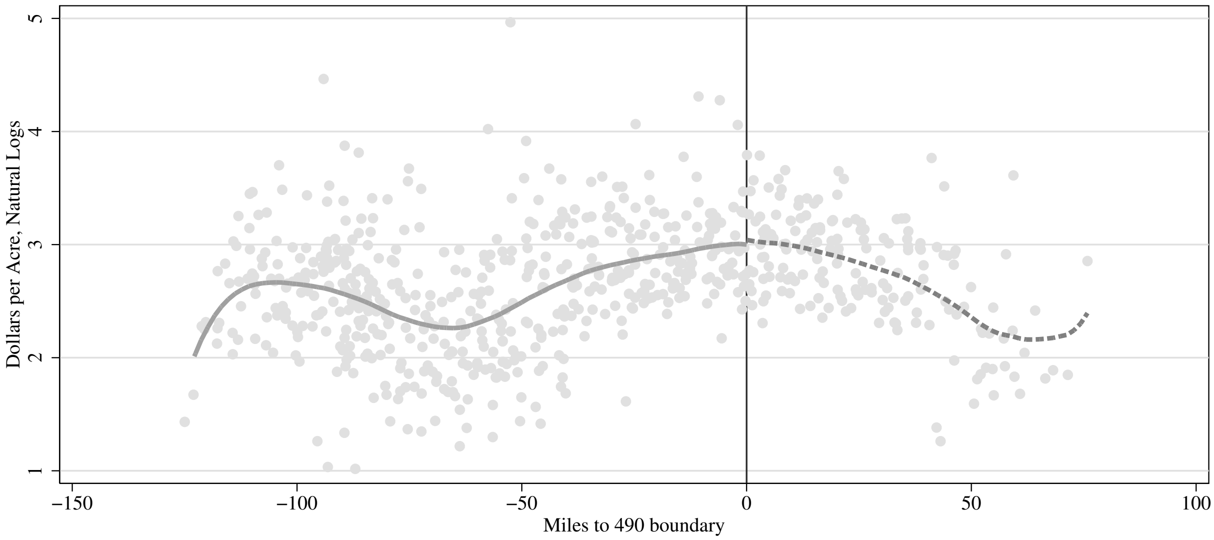
Panel B: Boundary in Northwestern Georgia between 40 and 160 acres



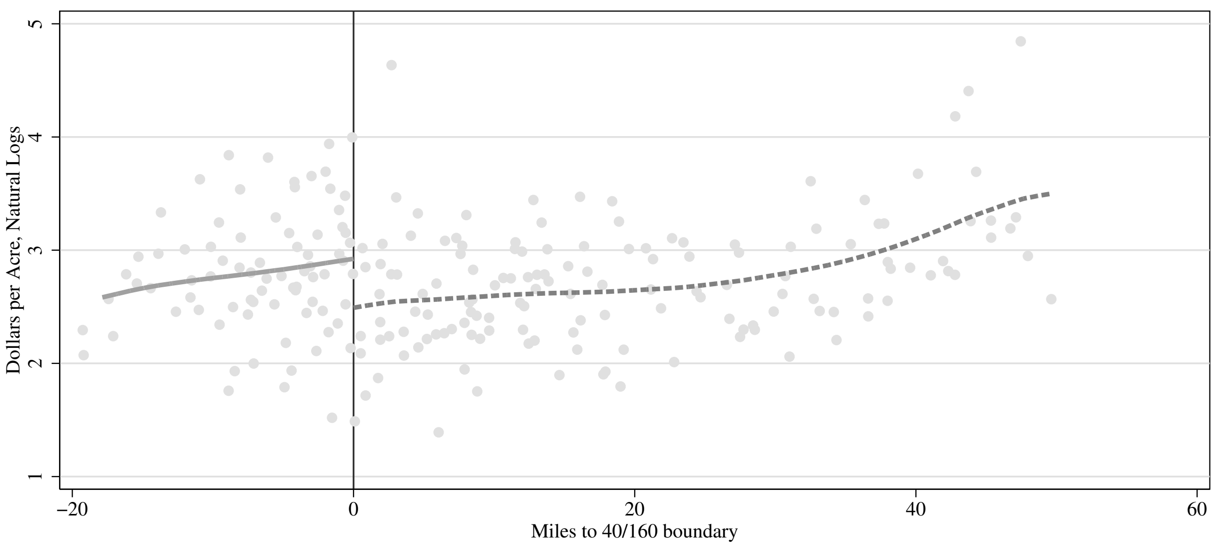
Notes: These graphs display the data and a smoothed average for farm size (the y axis) for the indicated acreage zones within the Georgia lottery area. Each dot denotes a Minor Civil Division (MCD) in Georgia, as defined in 1930. Centroids of MCDs were digitized from a 1930 Census map (Census Bureau, 1934), and the distance from each centroid to the indicated acreage-zone boundary is represented on the x axis. The solid and dashed gray lines are the local-polynomial smoothed average of the MCD data. This average is computed using a triangular kernel and the optimal bandwidth; see Nichols (2007) for more details on this estimator. The vertical line denotes a distance to the boundary of zero. The thin, black horizontal lines indicate the assigned acreages from the surveyors' grid. Panel A considers the boundary on the southeastern quadrant of the lottery zone that divides the 220 and 250-acre zones from the 490-acre zone. Panel B considers the boundary on the northwestern quadrant of the lottery zone that divides the 40-acre zone from the 160-acre zone. (See in the Figure D for precise locations.) The MCDs from the smaller acreage zones are assigned negative distances and thus occupy the area to the left of zero in each graph. Thus, when passing through zero in each graph, there is a discontinuous jump up in initial, assigned acreage. At conventional levels of confidence, a test of the equality of the two smoothed lines at zero is not rejected for Panel A but is rejected for Panel B. See the text and data appendix for further details and definitions.

Figure 5: Farm Value versus Distance to Assigned Acreage Change, Select Zones, 1930

Panel A: Boundary in Southeastern Georgia between 220/250 and 490 acres



Panel B: Boundary in Northwestern Georgia between 40 and 160 acres



Notes: These graphs display the data and a smoothed average for farm value (the y axis) for the indicated acreage zones within the Georgia lottery area. Each dot denotes a Minor Civil Division (MCD) in Georgia, as defined in 1930. Centroids of MCDs were digitized from a 1930 Census map (Census Bureau, 1934), and the distance from each centroid to the indicated acreage-zone boundary is represented on the x axis. The solid and dashed gray lines are the local-polynomial smoothed average of the MCD data. This average is computed using a triangular kernel and the optimal bandwidth; see Nichols (2007) for more details on this estimator. The vertical line denotes a distance to the boundary of zero. The thin, black horizontal lines indicate the assigned acreages from the surveyors' grid. Panel A considers the boundary on the southeastern quadrant of the lottery zone that divides the 220 and 250-acre zones from the 490-acre zone. Panel B considers the boundary on the northwestern quadrant of the lottery zone that divides the 40-acre zone from the 160-acre zone. (See in the Figure D for precise locations.) The MCDs from the smaller acreage zones are assigned negative distances and thus occupy the area to the left of zero in each graph. Thus, when passing through zero in each graph, there is a discontinuous jump up in initial, assigned acreage. At conventional levels of confidence, a test of the equality of the two smoothed lines at zero is not rejected for Panel A but is rejected for Panel B. See the text and data appendix for further details and definitions.

Table 1: Initial Assigned Plot Size and Subsequent Farm Size, Georgia Lottery Zone versus Comparison Counties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
			Latitude/Longitude		Lottery Zone Boundary		Alternative clustering		Conley's spatial GMM		Drop Subsets of Buffer Counties				
Sample year:	Basic	State dummies	Linear	Quadratic	Within 150km	Within 100km	No clustering	Cluster on state	25km cutoff	50km cutoff	Alabama	Florida	Tennessee, Kentucky	Virginia, Carolinas	Headright Georgia
1860	0.901	0.919	1.195	1.090	0.743	0.586	0.901	0.901	0.901	0.901	0.912	0.729	1.076	0.729	1.083
	(0.296)***	(0.448)*	(0.397)***	(0.394)**	(0.398)*	(0.596)	(0.197)***	(0.279)**	(0.258)***	(0.258)***	(0.289)***	(0.368)*	(0.375)**	(0.536)	(0.208)***
	[326]	[326]	[326]	[326]	[247]	[204]	[326]	[326]	[326]	[326]	[283]	[294]	[253]	[280]	[296]
	{0.10}	{0.31}	{0.24}	{0.31}	{0.13}	{0.15}	{0.10}	{0.10}	{0.10}	{0.10}	{0.11}	{0.14}	{0.09}	{0.12}	{0.16}
1880	0.904	0.900	0.961	0.946	0.852	0.820	0.904	0.904	0.904	0.904	0.922	0.910	0.911	0.809	0.940
	(0.120)***	(0.136)***	(0.121)***	(0.122)***	(0.131)***	(0.140)***	(0.119)***	(0.055)***	(0.118)***	(0.128)***	(0.120)***	(0.128)***	(0.128)***	(0.137)***	(0.112)***
	[360]	[360]	[360]	[360]	[275]	[225]	[360]	[360]	[360]	[360]	[305]	[327]	[279]	[306]	[328]
	{0.40}	{0.43}	{0.44}	{0.44}	{0.41}	{0.40}	{0.40}	{0.40}	{0.40}	{0.40}	{0.39}	{0.45}	{0.40}	{0.41}	{0.46}
1900	0.467	0.488	0.469	0.458	0.426	0.382	0.467	0.467	0.467	0.467	0.481	0.467	0.455	0.409	0.502
	(0.076)***	(0.107)***	(0.080)***	(0.089)***	(0.091)***	(0.105)***	(0.077)***	(0.045)***	(0.077)***	(0.082)***	(0.077)***	(0.084)***	(0.084)***	(0.097)***	(0.066)***
	[369]	[369]	[369]	[369]	[279]	[225]	[369]	[369]	[369]	[369]	[314]	[334]	[287]	[309]	[337]
	{0.36}	{0.42}	{0.40}	{0.41}	{0.38}	{0.40}	{0.36}	{0.36}	{0.36}	{0.36}	{0.36}	{0.42}	{0.37}	{0.38}	{0.44}
1930	0.115	0.116	0.109	0.086	0.079	0.041	0.115	0.115	0.115	0.115	0.128	0.133	0.089	0.061	0.141
	(0.074)	(0.086)	(0.072)	(0.074)	(0.082)	(0.084)	(0.055)**	(0.042)**	(0.055)**	(0.060)*	(0.074)	(0.077)	(0.074)	(0.081)	(0.070)*
	[405]	[405]	[405]	[405]	[311]	[252]	[405]	[405]	[405]	[405]	[349]	[362]	[324]	[341]	[364]
	{0.11}	{0.18}	{0.17}	{0.20}	{0.11}	{0.13}	{0.11}	{0.11}	{0.11}	{0.11}	{0.09}	{0.15}	{0.12}	{0.13}	{0.13}
1950	-0.015	-0.009	-0.020	-0.026	-0.026	-0.020	-0.015	-0.015	-0.015	-0.015	-0.015	0.009	-0.017	-0.044	-0.014
	(0.022)	(0.018)	(0.011)*	(0.005)***	(0.024)	(0.013)	(0.009)*	(0.022)	(0.094)	(0.010)	(0.022)	(0.009)	(0.024)	(0.024)*	(0.024)
	[403]	[403]	[403]	[403]	[309]	[250]	[403]	[403]	[403]	[403]	[347]	[360]	[322]	[339]	[362]
	{0.06}	{0.28}	{0.30}	{0.39}	{0.09}	{0.11}	{0.06}	{0.06}	{0.06}	{0.06}	{0.05}	{0.12}	{0.05}	{0.18}	{0.06}
1970	-0.008	-0.004	-0.014	-0.020	-0.023	-0.017	-0.008	-0.008	-0.008	-0.008	-0.011	0.013	-0.003	-0.035	-0.007
	(0.026)	(0.024)	(0.015)	(0.016)	(0.029)	(0.023)	(0.011)	(0.020)	(0.011)	(0.013)	(0.026)	(0.020)	(0.028)	(0.026)	(0.028)
	[404]	[404]	[404]	[404]	[309]	[250]	[404]	[404]	[404]	[404]	[348]	[361]	[323]	[339]	[363]
	{0.06}	{0.28}	{0.39}	{0.40}	{0.07}	{0.07}	{0.06}	{0.06}	{0.06}	{0.06}	{0.06}	{0.11}	{0.02}	{0.13}	{0.08}

Notes: This table reports estimates of equation 1 for various years (denoted in the row headings) and specifications (indicated in the column headings). The coefficient displayed in each cell is the β_t, the effect of initial parcel size on current farm size, with both variables measured in acres. Reporting of additional terms is suppressed. The sample employed is the cross section of counties from each year within the lottery and buffer zones. Standard errors are reported in parentheses. Unless otherwise specified, standard errors are heteroskedasticity-robust and clustered by lottery wave. In Column 12, we report White's heteroskedasticity-robust standard errors with no clustering. In Column 13, we report standard errors clustered on state. In Columns 14-16, we report standard errors using Conley's adjustment for spatial autocorrelation. Single asterisk denotes statistical significance at the 90% level of confidence; double, 95%; and triple, 99%. The sample size and R^2 for each regression are reported, respectively, in square and curly brackets. See the text and data appendix for further details and definitions.

Table 2: Initial Assigned Plot Size and Subsequent Farm Size, Within Georgia Lottery Zone

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Acreage-Zone Boundary		Closest Zone Boundary is Rectilinear			Closest Zone Boundary is Natural				Conley's spatial GMM	
Sample year:	Baseline	Within 50km	Within 25km	All	Within 50km	Within 25km	All	Within 50km	Within 25km	No clustering	25km cutoff	50km cutoff
1860	1.171	1.263	1.055	1.119	1.217	0.937	1.399	1.392	1.483	1.171	1.171	1.171
	(0.138) ***	(0.153) ***	(0.167) ***	(0.157) ***	(0.149) ***	(0.149) ***	(0.259) ***	(0.264) ***	(0.268) ***	(0.147) ***	(0.164) ***	(0.164) ***
	[99]	[88]	[69]	[51]	[42]	[30]	[48]	[46]	[39]	[99]	[99]	[99]
	{0.35}	{0.34}	{0.26}	{0.43}	{0.44}	{0.38}	{0.25}	{0.24}	{0.22}	{0.35}	{0.35}	{0.35}
1880	0.945	0.981	0.925	0.796	0.781	0.846	1.520	1.530	1.317	0.945	0.945	0.945
	(0.110) ***	(0.155) ***	(0.170) ***	(0.058) ***	(0.081) ***	(0.121) ***	(0.330) ***	(0.335) ***	(0.507) **	(0.117) ***	(0.115) ***	(0.125) ***
	[103]	[91]	[72]	[53]	[43]	[32]	[50]	[48]	[40]	[103]	[103]	[103]
	{0.51}	{0.49}	{0.42}	{0.71}	{0.70}	{0.73}	{0.47}	{0.47}	{0.25}	{0.51}	{0.51}	{0.51}
1900	0.485	0.429	0.415	0.468	0.389	0.433	0.521	0.526	0.288	0.485	0.485	0.485
	(0.062) ***	(0.068) ***	(0.083) ***	(0.052) ***	(0.068) ***	(0.083) ***	(0.132) ***	(0.134) ***	(0.122) *	(0.077) ***	(0.076) ***	(0.082) ***
	[103]	[91]	[72]	[53]	[43]	[32]	[50]	[48]	[40]	[103]	[103]	[103]
	{0.54}	{0.54}	{0.53}	{0.57}	{0.59}	{0.63}	{0.46}	{0.50}	{0.20}	{0.54}	{0.54}	{0.54}
1930	0.145	0.139	0.154	0.175	0.175	0.187	0.091	0.071	0.032	0.145	0.145	0.145
	(0.063) **	(0.086)	(0.135)	(0.094) *	(0.124)	(0.164)	(0.024) ***	(0.022) **	(0.036)	(0.054) ***	(0.054) **	(0.058) **
	[118]	[102]	[80]	[59]	[48]	[36]	[59]	[54]	[44]	[118]	[118]	[118]
	{0.13}	{0.11}	{0.10}	{0.15}	{0.13}	{0.13}	{0.10}	{0.06}	{0.01}	{0.13}	{0.13}	{0.13}
1930 (MCDs)	0.109	0.096	0.113	0.118	0.105	0.136	0.080	0.069	0.056	0.109	0.109	0.109
	(0.024) ***	(0.028) ***	(0.041) ***	(0.031) ***	(0.038) ***	(0.056) **	(0.016) ***	(0.016) ***	(0.018) ***	(0.024) ***	(0.029) ***	(0.038) ***
	[1213]	[1073]	[840]	[646]	[541]	[401]	[567]	[532]	[439]	[1213]	[1213]	[1213]
	{0.04}	{0.03}	{0.04}	{0.05}	{0.04}	{0.05}	{0.03}	{0.03}	{0.01}	{0.04}	{0.04}	{0.04}
1950	0.105	0.058	0.081	0.165	0.092	0.097	0.008	0.003	0.056	0.105	0.105	0.105
	(0.055) *	(0.066)	(0.083)	(0.052) **	(0.070)	(0.080)	(0.066)	(0.068)	(0.116)	(0.046) **	(0.046) **	(0.056)
	[118]	[102]	[80]	[59]	[48]	[36]	[59]	[54]	[44]	[118]	[118]	[118]
	{0.04}	{0.02}	{0.02}	{0.11}	{0.05}	{0.04}	{0.00}	{0.00}	{0.01}	{0.04}	{0.04}	{0.04}
1970	0.012	0.010	0.019	0.022	0.017	0.023	-0.005	-0.004	0.008	0.012	0.012	0.012
	(0.017)	(0.020)	(0.022)	(0.020)	(0.023)	(0.021)	(0.014)	(0.016)	(0.030)	(0.008)	(0.008)	(0.010)
	[118]	[102]	[80]	[59]	[48]	[36]	[59]	[54]	[44]	[118]	[118]	[118]
	{0.01}	{0.01}	{0.02}	{0.04}	{0.03}	{0.04}	{0.00}	{0.00}	{0.00}	{0.01}	{0.01}	{0.01}

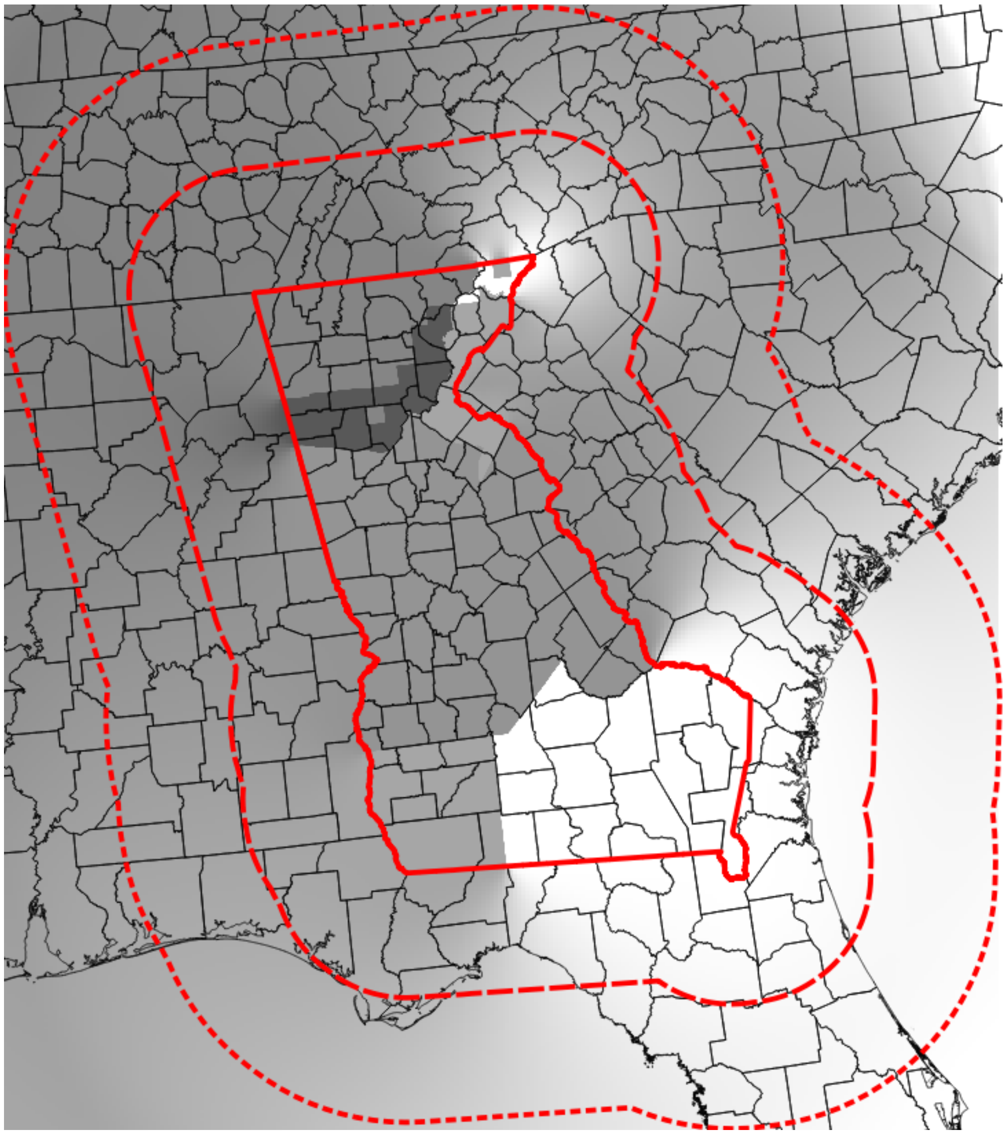
Notes: This table reports estimates of equation 2 for various years (denoted in the row headings) and specifications (indicated in the column headings). The coefficient displayed in each cell is the β_t', the effect of initial parcel size on current farm size, with both variables measured in acres. Reporting of additional terms is suppressed. The sample employed is the cross section of counties from each year within the lottery zone, except for the row labeled "MCDs" which uses minor civil divisions in 1930. Standard errors are reported in parentheses. Unless otherwise specified, standard errors are heteroskedasticity-robust and clustered by lottery wave. In Column 10, we report White's heteroskedasticity-robust standard errors with no clustering. In Columns 11-12, we report standard errors using Conley's adjustment for spatial autocorrelation. Single asterisk denotes statistical significance at the 90% level of confidence; double, 95%; and triple, 99%. The sample size and R^2 for each regression are reported, respectively, in square and curly brackets. See the text and data appendix for further details and definitions.

Table 3: Relationship between Log Farm Value and the Imputed Gap-Squared

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
		1880			Panel of 1850-1970				Panel of 1880-1950		
Dependent Variable:					*Panel A: Specification with Farm Size in Levels*						
Gap Squared	-0.111	-0.124	-0.123	-0.130	-0.135	-0.137	-0.115	-0.162	-0.176	-0.179	-0.088
	(0.055) **	(0.053) **	(0.050) **	(0.031) ***	(0.032) ***	(0.033) ***	(0.041) ***	(0.049) ***	(0.050) ***	(0.052) ***	(0.042) ***
	[0.043] **	[0.045] ***	[0.045] ***	[0.037] ***	[0.039] ***	[0.040] ***	[0.041] ***	[0.064] **	[0.064] ***	[0.065] ***	[0.043] **
					Panel B: Specification with Farm Size in Natural Logs						
Gap Squared	-0.163	-0.150	-0.162	-0.202	-0.182	-0.193	-0.216	-0.170	-0.151	-0.162	-0.167
	(0.034) ***	(0.028) ***	(0.029) ***	(0.030) ***	(0.031) ***	(0.033) ***	(0.036) ***	(0.029) ***	(0.030) ***	(0.031) ***	(0.027) ***
	[0.047] ***	[0.042] ***	[0.047] ***	[0.051] ***	[0.046] ***	[0.049] ***	[0.063] ***	[0.045] ***	[0.041] ***	[0.043] ***	[0.045] ***
Specification:											
Diff-in-diff variables	Yes	Yes	Yes	x Year	x Year	x Year	x Year	x Year	x Year	x Year	x Year
Latitude and longitude		Yes	Yes		x Year	x Year	x Year		x Year	x Year	x Year
State dummies			Yes			x Year	x Year			x Year	x Year
County fixed effects							Yes				Yes

Notes: This table reported estimates of the gap-squared model, estimated from the system of equations 3 (which constructs the gap^2 variable) and 5. Reported here are estimates of the ω parameter (the effect of gap^2 on log farm value per acre) using various specifications of the system of equations. Panel A uses county-average farm size in levels (acres), while Panel B uses this same variable transformed into natural logs. The estimates of ω are normalized to be interpretable as the reduction in log farm value at the mean value of gap^2 in 1880. The underlying dataset is the "consistent county" sample described in the text, with data harmonized to 1900 county boundaries. This particular sample is needed here because all of the estimates are block bootstrapped using the consistently defined county as a cluster. The phrase "x Year" indicates that the variable in the row heading enters into the regression interacted with year dummy variables. Various sample years are employed, as denoted in the column headings. Robust standard errors (clustered by county) are reported in parentheses. Alternative standard errors computed from the block bootstrap by county are reported in square brackets. Single asterisk denotes statistical significance (computed using both types of standard errors) at the 90% level of confidence; double, 95%; and triple, 99%. See the text and data appendix for further details and definitions.

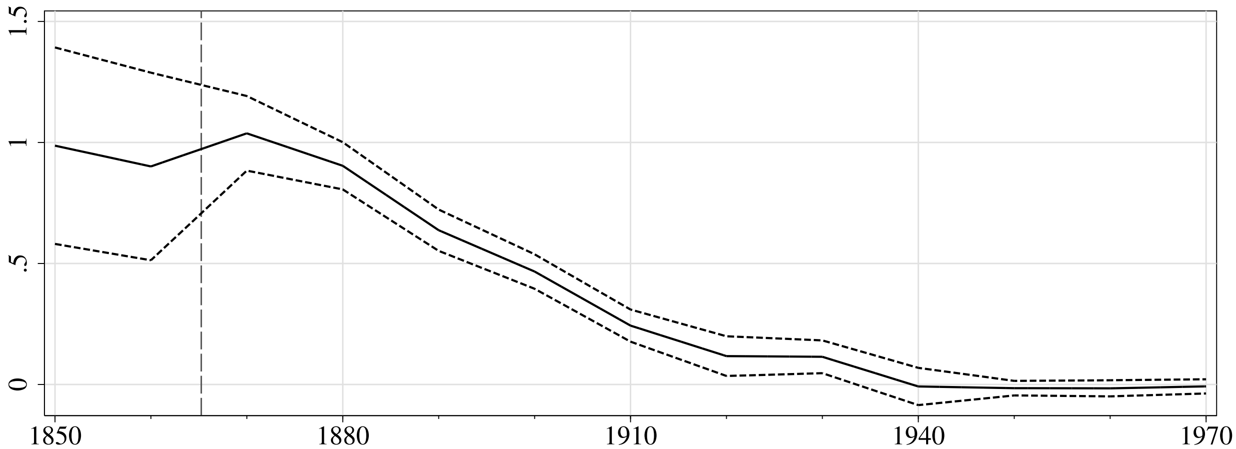
Appendix Figure A: Alternate Extrapolation of Initial Parcel Size to Buffer Zone



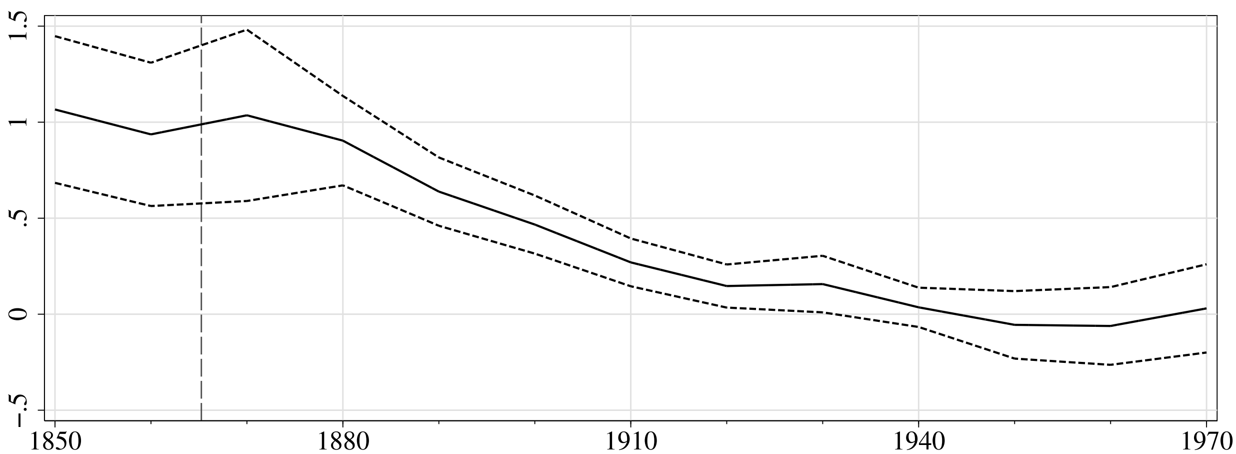
Notes: This is an alternative version of Figure 1. The map displays various features in and around the Georgia lottery zone. County boundaries (as defined in 1900) are denoted in black and are drawn from NHGIS data. The solid red line denotes the outer boundary of the Georgian lottery zone. The dashed and dotted red lines denote, respectively, a 100 km and 200 km buffer around the lottery zone. Inside the lottery zone, the background shading denotes the initial parcel size. Outside of the lottery zone, the background shading denotes the initial parcel size of the constructed as an inverse-distance-weighted average of points in the Lottery Zone. Darker shades indicate smaller parcel sizes. See the text and data appendix for further details and definitions.

Appendix Figure B: Replication of Figure 2, Panel B, with Standard Errors on Each Series

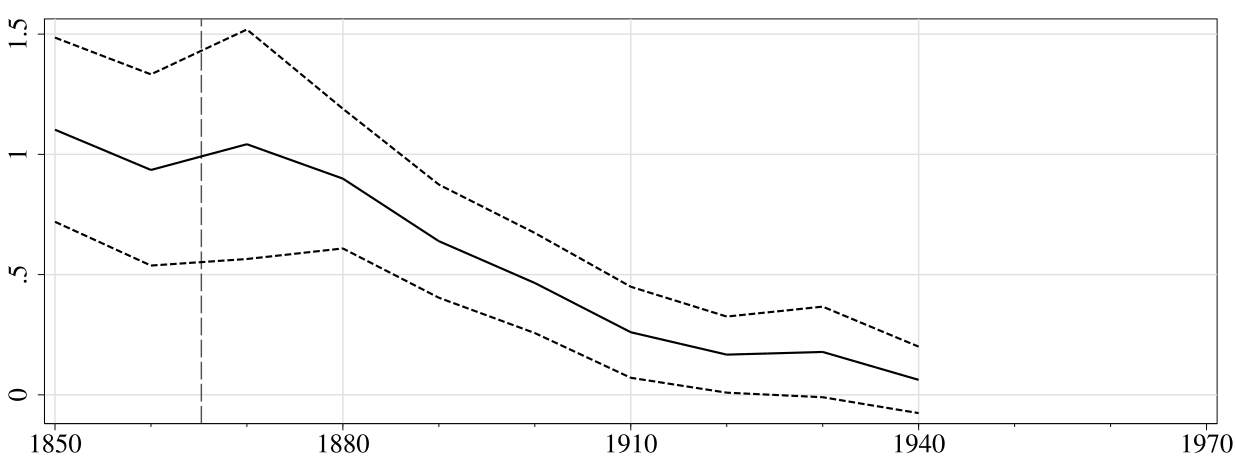
Panel A: Unharmonized counties



Panel B: Consistent counties



Panel C: Consistent counties with county fixed effects, $\beta_{1950}, \beta_{1960}, \beta_{1970} = 0$

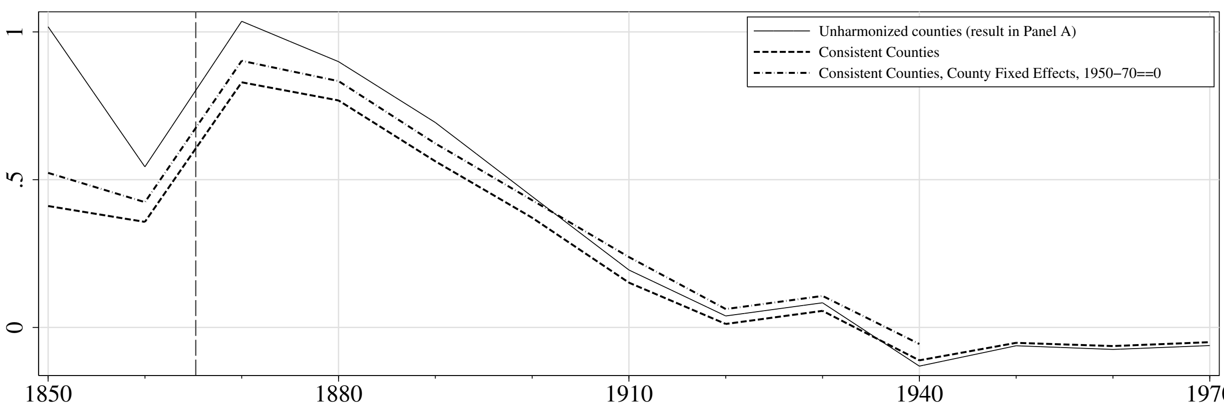


Notes: This figure expands upon the estimates shown in Figure 2, Panel B, by presenting the point estimates in separate graphs and including their confidence intervals. The graphs present decade-specific coefficient estimates of the relationship between farm size and initial parcel size assigned by the surveyors' grid in the land opening. The vertical dashed line in each Panel denotes the end of the Civil War. All Panels present estimates of equation 1 using the lottery zone in Georgia and the nearby comparison ("buffer") areas. The solid line graphs the point estimates of β_t and the 95% confidence intervals are displayed by the dashed lines. Panel A replicates the results for the cross section of unharmonized counties shown in Figure 2, Panel B. Panel B presents estimates using the sample of lottery and buffer counties using data harmonized onto 1900 county boundaries. Panel C replicates the results in Panel B, but also includes county-level fixed effects and the β_t in this case are normalized to zero for the years 1950–70.

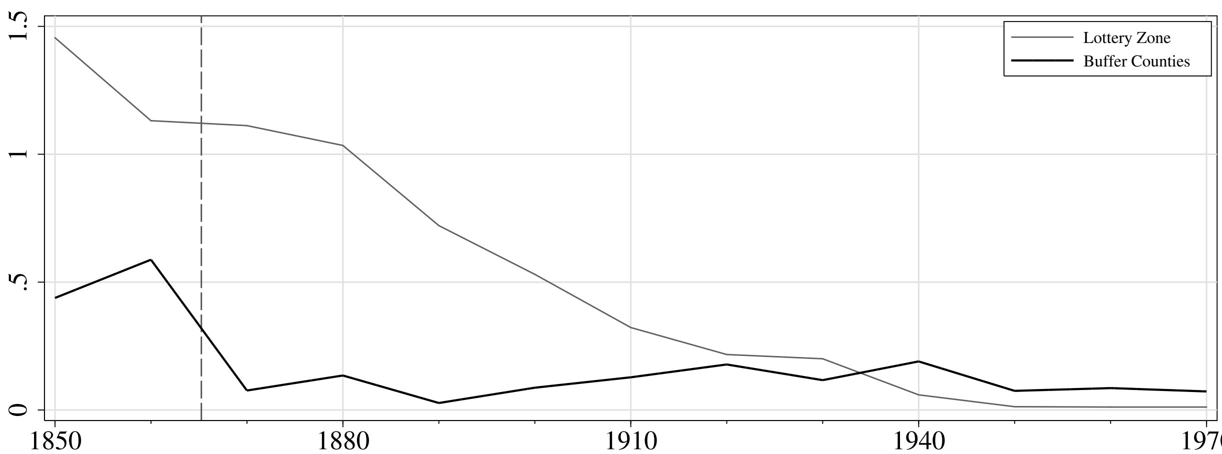
Appendix Figure C: Replication Figure 2 with Inverse-Distance-Weighted Extrapolation of Assigned Acreage to Buffer Zone

Panel A: Difference in Difference, Lottery versus Buffer Zones

Panel B: Harmonized versus Original County Boundaries

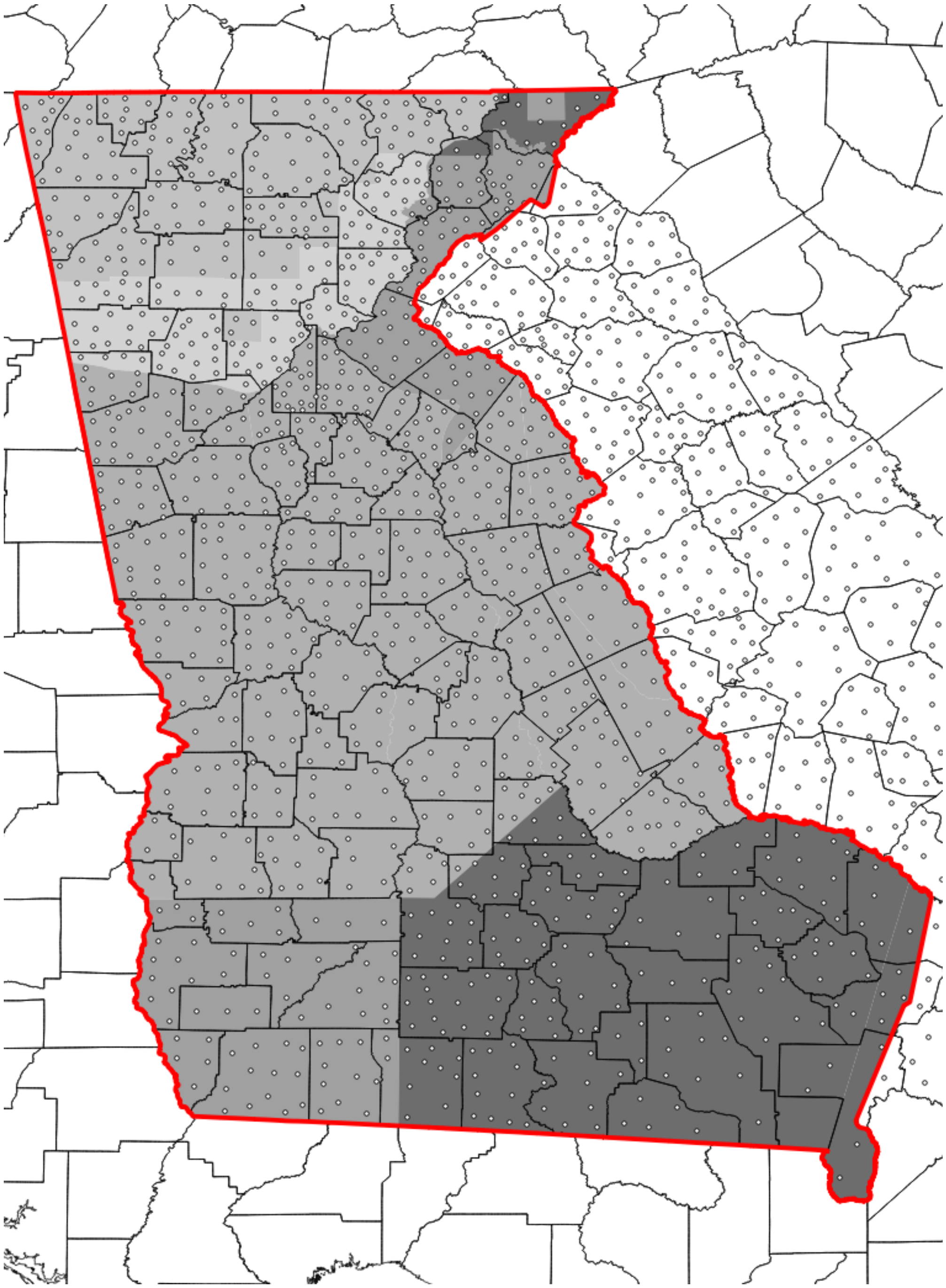


Panel C: Regression Coefficient on Initial Acreage



Notes: These graphs present decade-specific coefficient estimates of the relationship between farm size and initial parcel size assigned by the surveyors' grid in the land opening. The vertical dashed line in each Panel denotes the end of the Civil War. Panel A presents estimates from equation 1, with the 95% confidence intervals displayed by the dashed lines. Estimates are weighted by county area to account for the changing cross section of counties in each year. Standard errors are clustered by lottery wave. Panel B compares the point estimates from Panel A, estimated from the cross-section of counties in each decade, with similar results from a panel of consistently defined county data. The consistent counties are constructed by harmonizing the data onto the 1900 county boundaries. The dashed line displays estimates of equation 1 for this panel of consistent-boundary counties. The dash-dot line also displays estimates of the equation 1 for the consistent counties, but pools the decades and includes county fixed effects in the specification. The β_t in this case are normalized to zero for the years 1950–70. Panel C presents estimates of equation 2 for the lottery zone in Georgia and the nearby comparison ("buffer") areas. (Note that the difference in difference in Panel A is equal to the gap between the curves in Panel C.)

Appendix Figure D: Lottery Zone and Centroids of 1930 Minor Civil Divisions in Georgia



Notes: this map displays various features in and around the Georgia lottery zone. The dots represent the centroids of 1930 Minor Civil Divisions in Georgia. These were digitized from a 1930 Census map (Census Bureau, 1934). County boundaries (also defined in 1930) are denoted in black and are drawn from NHGIS data. The solid red line denotes the outer boundary of the Georgian lottery zone. Inside the lottery zone, the background shading denotes the initial parcel size. Darker shades indicate larger parcel sizes. See the text and data appendix for further details and definitions.