Essays on Inequality and Finance

Mintra Dwarkasing

Essays on Inequality and Finance

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op

dinsdag 17 januari 2017 om 16.00 uur door

Mintra Shanta Devi Dwarkasing

geboren op 12 april 1985 te Nijmegen

Promotiecommissie:

Promotor: prof. dr. Steven Ongena

Copromotor: dr. Fabio Braggion

Overige leden: dr. Marco Da Rin

dr. Rik Frehen

dr. Lyndon Moore

dr. María Fabiana Penas

To my parents

Acknowledgements

This thesis would not have been here without the help and support of many others. First of all, I would like to thank my advisors Fabio and Steven for their continuous support and guidance when writing this thesis. Fabio, thank you so much for your advice, encouragement, patience and believe in me. Steven, thank you for always being there, no matter from where in the world, to provide valuable feedback and for all your comments and advice along the way. I am very grateful to have learned so much from both of you.

I also would like to thank the members of my thesis committee, Marco Da Rin, Rik Frehen, Lyndon Moore and Fabiana Penas for their time to read this work and their invaluable comments and suggestions. This thesis has benefitted greatly from your input. Financial support from the Dutch Science Foundation (NWO) through its Mosaic-program (projectnumber 017.007.107) is greatly acknowledged.

Lastly, I would like to thank my family and friends for their continuous support and love.

My parents, Johan and Mientje, thank you for your unconditional love and understanding.

Mom, even though you are no longer here, you are forever present in our hearts.

I dedicate this thesis to you both.

Table of Contents

Introduction

In recent years there has been a growing interest in wealth inequality and its effects on financial outcomes, both from the academic world as well as in the public debate (see for example Piketty, 2014). Up to now however, little empirical evidence exists on how wealth inequality can affect and interacts with important economic outcomes. This thesis attempts to fill this gap by empirically examining how wealth inequality affects the financing choices of start-up firms, entrepreneurial dynamism and, additionally, how local inequality is associated with local judicial decision making. Shedding light upon these relations is important as a deeper understanding can eventually help in fostering economic growth.

The first part of this thesis studies how local household inequality in US counties and metropolitan areas affects entrepreneurial dynamism, in the form of (new) establishment entries and exits as well as the capital structure choices of start-up firms. Recent economic theory directly links the degree of wealth inequality to economic and financial outcomes. Engerman, 1997, Glaeser, 2003 and Sonin, 2003 for example describe how large levels of wealth inequality could impair the development of institutions that are conducive for economic growth. In unequal societies wealthy elites may prevent the sound development of basic institutions such as schools, the judiciary and capital markets, in order to maintain their grip on power. According to this view, an unequal society will be characterized by less effective schooling and law enforcement and by a poorly-developed financial system (see also Acemoglu, 2013, pp. 152-158, Perotti, 2006 and Rajan, 2011).

Overall, these theories predict that wealth inequality may negatively influence the decision of individuals to become entrepreneurs and affect the technology, ownership, and financing choices made by firms. In more unequal societies, either because of a lack of higher education or because of poorly-developed financial markets, we expect to observe less entry of new firms, and those that enter to more likely operate in traditional sectors, to come with a simpler corporate form. In more unequal societies, bank debt or family loans, as opposed to equity from institutional investors, will be the prevailing form of external finance. On the other hand, an alternative view sees some level of wealth inequality as a positive factor for entrepreneurship and growth. For example, wealthy individuals may also be engaged in charity and philanthropy: activities that could improve the provision of local public goods and have a positive impact on the economy. Hence, the effects of inequality remain unclear.

Using two measures of wealth inequality at the US MSA/county level: One based on the distribution of financial rents in 2004 and another one related to the distribution of land holdings in the late Nineteenth century, the empirical results suggest that in more unequal areas business creation, especially of high-tech ventures, is lower but more likely to be financed via bank and family financing and therefore less likely via equity from angels and venture capital. Wealth inequality seems to affect entrepreneurial choices and capital structure both via demand and supply channels: In more unequal counties the number of banks per capita is significantly lower compared to more equal counties, suggesting local credit rationing. Moreover, in more unequal counties the probability that an entrepreneur has a college degree or higher is significantly lower.

The second part of this thesis looks in more detail at how local household inequality is related to the judiciary and in specific to judges' decision making, using a unique dataset on individual second degree (appeal) sentences from US courts from 1984 onwards. According to established theory (Perotti and von Thadden, 2006), inequality may affect firms' financing. One of the implications of this model is that in more unequal societies bank financing will be more prevalent. In the model voters may vote for the supply of those forms of financing that are best aligned with their own preferences. Hence, they may also vote for judges that sentence in line with these preferences. Based upon these notions we investigate whether higher inequality in US counties is associated with banks having a higher probability to win a second degree trial. Additionally, we investigate whether, in the spirit of Engerman and Sokoloff (2006), lobbying reinforces the effect local inequality can have on judges' decision making in favour of banks. Using a historical measure of inequality dating as far back as 1890 at the US MSA/county level we find evidence in favour of the latter: Our results suggest that in more unequal counties where it is easier for elites to lobby and where judges are less independent, the effect of local inequality is more strongly related to whether or not a bank wins a second degree trial.

The third chapter of this thesis shows that a shock in trust and culture from the past can lead to an unequal probability of obtaining a mortgage loan for similar applicants today. More specifically, exploiting the random occurrence of battles during the American Civil War and using a spatial regression discontinuity approach it shows that being located in a battle county fosters the approval probability of mortgage applications. However, mortgage applicants that are of minority race have a more than 7 percent lower likelihood of obtaining a mortgage in a battle county compared to applicants of non-minority race. In addition, conditional upon approval, minority applicants also receive significant lower loan amounts in these counties (around 13 percent lower) compared to non-minority applicants. A channel through which this effect of battles still persists today is social capital: Battle counties have a considerable higher social capital index today compared to similar non battle counties. This result may suggest, in line with Putnam (1996) a 'dark side' to social capital: The common unity that was created within the white population in a county after a battle took place has been transmitted from generation to generation resulting in higher local social capital today. However, 'outsiders', i.e. non-white individuals benefit negatively from this higher level of social capital.

The last part of the thesis, chapter 4, studies the real effects of tax avoidance for US firms. This is an important question as little is known yet about the real effects of tax avoidance. On one hand tax avoidance may increase internal funds and thereby alleviate financing constraints, which can result in a positive effect on corporate innovation input and, to the extent that innovation input and output are positively related, innovation output. On the other hand, tax avoidance may hamper innovation due to decreasing corporate transparency and the accompanying costs. To push the empirical findings from correlation to causation, we exploit exogenous variation in tax avoidance caused by the introduction of Check-the-Box regulation in 1997 in the US. The results point in the direction of the latter: Tax avoidance has a negative effect on corporate innovation, both innovation input as well as output. This negative effect is less pronounced for financially constrained firms.

Chapter 1: Household Inequality, Entrepreneurial Dynamism and Corporate Financing[1]

Fabio Braggion, Mintra Dwarkasing, Steven Ongena[2]

[1] We thank Cédric Argenton, Fabio Castiglionesi, James Choi, Marco Da Rin, Joost Driessen, Thomas Hellmann, William Kerr, Alberto Manconi, Ramana Nanda, Fabiana Penas, Johann Reindl, David Robinson, Joacim Tåg, Florian Schuett, Janis Skrastins and Per Strömberg, participants at the 2015 ASSA-Meetings (Boston), 2015 FIRS Conference (Reykjavík), the Second European Workshop on Entrepreneurship Economics (Cagliari), the *Sveriges Riksbank* and EABCN Conference on Inequality and Macroeconomics (Stockholm), the 15th Workshop on Corporate Governance and Investment (Oslo), and the Bundesbank-CFS-ECB Joint Lunchtime Seminar, and seminar participants at Rotterdam School of Management, Tilburg University and the Tilburg Center for Law and Economics for very valuable comments. We also like to thank the *Kauffman Foundation* for providing access to its Firm Survey.

[2] Braggion is from EBC, CentER & Tilburg University. Dwarkasing is from EBC, CentER & Tilburg University. Ongena is from University of Zurich, SFI and CEPR.

Abstract

We empirically test if household wealth inequality determines entrepreneurial dynamism and corporate financing. We construct a measure of wealth inequality at the US MSA/county level based on the distribution of financial rents in 2004. We find that business creation, especially that of high-tech ventures, is lower in more unequal areas and more likely to be financed via bank and family financing. Inequality may undermine local institutions such as banks, schools, and courts and hamper entrepreneurial dynamism. Instrumentation with past rainfall and temperature, the use of 1890 landholdings to measure inequality, and the impact of changes in state "death taxes" corroborate these findings.

I. Introduction

Households' wealth inequality is a defining societal characteristic with important implications for economics and finance.[3] The growth of wealth inequality during recent decades has returned the issue to the top of the agendas of policymakers and social leaders in many Western economies. This paper sheds light on the economic consequences of households' wealth inequality and, in particular, studies its impact on firm creation, technology, and financing choices made by young entrepreneurs. This is an important issue because the creation of new business ventures, as well as the available means for financing them, are defining features of any economic system and likely have an important impact on economic development.

Beginning with Adam Smith's Wealth on Nations, various theories have linked wealth inequality with economic outcomes, producing different empirical predictions.[4] On one hand, large levels of wealth inequality could impair the development of institutions that are conducive for economic growth. In unequal societies, wealthy elites may prevent the sound development of basic institutions such as banks, schools, and courts in order to maintain their hold on power. According to this view, an unequal society will be characterized by a poorly developed financial system and by less effective schooling and law enforcement.[5] On the other hand, an alternative view sees some level of wealth inequality as a positive factor for entrepreneurship and growth. Thomas Malthus, for example, argued that the existence of few rich landlords may

---

[3] Recent academic work has consequently defined, measured and analyzed inequality (Chetty, Hendren, Kline and Saez (2014); Chetty, Hendren, Kline, Saez and Turner (2014); Piketty (2014); Saez and Zucman (2014)).

[4] In "The Wealth of Nations," Adam Smith expressed concern that an unequal distribution of land may have had a negative impact on the development of the New World colonies. In his words, "*The engrossing of land, in effect, destroys this plenty and cheapness*" (Smith (1776), p. 726).

[5] Engerman and Sokoloff (1997), Glaeser, Scheinkman and Shleifer (2003), Sonin (2003), Berkowitz and Clay (2011), Acemoglu and Robinson (2013), pp. 152-158, Perotti and von Thadden (2006) and Rajan and Ramcharan (2011).

help to generate high aggregate demand and stimulate economic activity. "*There must therefore be a considerable class of persons who have both the will and power to consume more material wealth than they produce, or the mercantile classes could not continue profitably to produce so much more than they consume. In this class the landlords no doubt stand pre-eminent"* (Malthus 1836), p. 466).[6] Wealthy individuals may also be engaged in charity and philanthropy: activities that could improve the provision of local public goods and have a positive impact on the economy.

Our study brings these alternative perspectives to the data and focuses on households' wealth inequality measured at either the US metropolitan statistical area (MSA) or county level. It then relates this inequality to individuals' decisions to start new businesses and their financing. Studying this topic at the MSA or county level substantially "shortens the distance" between local conditions and economic outcomes and therefore allows us to obtain more precise estimates of the effects of interest. Important for our purposes is the observation that US local administrations are often co-responsible (with state-level authorities) for many important elements of public life, such as the organization of schooling, the judiciary and the enforcement of the law and taxation.[7] We study start-up firms because their creation and financing is more likely to depend upon local institutional and credit market conditions and because we can observe their technology and production choices at the very beginning of their life cycle to identify entrepreneurial dynamics.[8]

---

[6] Matsuyama (2002) puts forward a related idea. While demanding luxury goods, wealthy individuals help firms to reduce their average cost of production. Such products become then affordable to a larger proportion of individuals, which, in turn, induce more entrepreneurs to produce them.

[7] In addition, Ramcharan (2010), Rajan and Ramcharan (2011), Galor, Moav and Vollrath (2009) and Vollrath (2013) relate county-level inequality to various economic outcomes, such as income redistribution, access to credit and schooling.

[8] Lerner (1995), Sorenson and Stuart (2001), Petersen and Rajan (2002), Guiso, Sapienza and Zingales (2004), Kerr and Nanda (2009), Chen, Gompers, Kovner and Lerner (2010) and Berger, Cerqueiro and Penas (2014).

Identifying the effect of inequality on entrepreneurial activities presents us with sizable empirical challenges. First, it is difficult to measure wealth inequality at a local level because direct and reliable data on household wealth are virtually impossible to find. Additionally, entrepreneurial outcomes (such as the ease of starting a *de novo* firm and the resultant distribution of cash flows) could easily determine local wealth inequality itself.

We address the first problem by constructing a proxy for local wealth inequality based on the amounts of dividends and interests earned by US households in 2004 (the first year for which these data are available) as reported by Internal Revenue Service (IRS) Statistics of Income (IRS-SOI) data. The IRS-SOI data report the total amount of dividends and interest income received by US households in each postal zip code. Under the assumption that a typical household holds the market index for stocks and bonds, the amount of financial rents it receives depends only on the quantity of stocks and bonds it holds – in other words, by the total amount of financial wealth it owns. We use this information to construct the distribution of financial rents at the local level and compute a Gini coefficient of financial wealth inequality.[9] In an additional test, we also construct a historical proxy of wealth inequality based on the distribution of land holdings in US counties in 1890 (*sic*), a measure that, given its historic nature, is strappingly pre-determined.[10]

The second empirical challenge consists of precisely identifying a causal relationship between wealth inequality and entrepreneurial outcomes, as wealth inequality itself may correlate with unobserved factors that are likely to affect our estimates. We tackle this problem in various ways. First, as our measures of wealth inequality are local (and because we know

---

[9] Mian, Rao and Sufi (2013) use a similar methodology to construct local measures of US Households' Net Worth. Saez and Zucman (2014) also use this methodology to construct US-wide long time series of wealth inequality.

[10] This measure has already been employed by Galor, Moav and Vollrath (2009), Ramcharan (2010), Rajan and Ramcharan (2011) and Vollrath (2013) to study US historical developments in education, banking and redistribution.

the precise location of the firms), we saturate our specifications with state, year, industry, state-year and/or industry-year fixed effects to account for any unobserved heterogeneity at those aforementioned levels. This procedure allows us to control for competing explanations of the deeply rooted determinants of institutions, such as individual states' type of colonization and legal traditions (see (Berkowitz and Clay 2011), pp. 16-59; Acemoglu, Johnson and Robinson (2001)), as well as changes in their legislation and regulation.

Second, we instrument the contemporary measure of wealth inequality with a set of variables related to the local historical averages of rainfall and temperature, which have been considered an exogenous predictor of contemporary wealth inequality. This strategy relies on the historical evidence provided by Engerman and Sokoloff (2002) that suggests that the quality of soil combined with the climate may have a persistent effect on the degree of inequality. In particular, regions whose soil and climate are best suited for large farms of crops such as cotton or tobacco should induce relatively high wealth inequality. The production of these crops entails high fixed costs. As a result, in equilibrium, the market can support only a few farms owned by a few wealthy individuals.[11]

The underlying assumption of the instrumental variable analysis is that local weather conditions matter because, via inequality, they determine local institutions and entrepreneurship. To address the concern of whether the exclusion restriction is satisfied, we perform a falsification test that links local weather conditions to local entrepreneurship in France: a developed country in which local authorities have very limited power to establish

[11] We measure average historical rainfall and temperature at the district level, where a district is defined by the National Climatic Data Center as a cluster of two or three counties sharing similar climatic conditions. The observed rainfall and temperature are good predictors of wealth inequality to the extent that inequality and institutions persist throughout time. Engerman and Sokoloff (1997), Engerman and Sokoloff (2002) and Rajan (2009) suggest that this is the case; ultimately, our first-stage regressions will address this empirical question. See also Easterly (2007), Galor, Moav and Vollrath (2009), Ramcharan (2010), Rajan and Ramcharan (2011) and Berkowitz and Clay (2011), pp. 102-104.

institutions such as schooling and the judiciary. In principle, we should not find any correlation between local weather conditions and entrepreneurship in France.

Third, we exploit state changes in Estate, Inheritance and Gift (EIG) taxes between 1976 and 2000 and assess their impact on (new) firm entry and exit using a difference-in-differences approach. EIG taxes may be related to wealth inequality because they define the amount of wealth transferred from one generation to another. As a result, lower EIG taxes should promote or maintain a high level of wealth inequality. Beginning in 1976, more than 30 states have eliminated their incremental EIG taxes imposed on top of the Federal tax, thus lowering the EIG tax burden on their citizens (Conway and Rork (2004)). If our conjectures on the relationship between wealth inequality and entrepreneurship are correct, we should find that states that lowered EIG taxes earlier experienced a significant change in entrepreneurship activity.

Our estimated coefficients robustly suggest that MSA-level inequality decreases firm entry into and exit from the MSA, indicating that wealth inequality has a negative effect on local business formation. Our estimates are not only statistically significant but also economically relevant. A one-standard-deviation increase in MSA-level wealth inequality leads to an approximately 10-percent increase in new establishments' entry and exit. Interestingly, more-equal areas experience higher closure of both young and old establishments, suggesting that, together with a genuine process of creative destruction (i.e., new establishments challenge the incumbents), there is a great deal of churning entry (i.e., many closures among newly formed establishments).

Our estimates also reveal that the negative effect of inequality is lowered (although not eliminated) when an area is racially more homogenous or when the wealthiest people in the community belong to "old" business families. The first result echoes Alesina, Michalopoulos and Papaioannou (2016), who find that a great deal of inequality between ethnic groups is associated with countries' worse economic performance. The second result suggests that "Old Money" may partially offset the negative effects of inequality, lending support to the notion that Old Money, by no longer being involved in businesses and especially active in philanthropy and charity, is not excessively detrimental to current entrepreneurial activities.[12]

Our analysis identifies a reduced-form relationship between wealth inequality and entrepreneurial activities and implicitly relates it to the quality of local institutions. In the second part of the paper, we assess whether we can find direct evidence of such mediating effects in the data. First, we study whether local institutions, such as schooling, the banking market and the judiciary, behave differently in areas with varying levels of wealth inequality. In particular, we find that wealth inequality is associated with an inefficient civil justice system: everything else equal, first-degree civil justice trials have a longer completion time in unequal counties. Higher wealth inequality is also associated with a lower percentage of the population with at least a college degree and a lower inflow of educated people from other geographical areas. Last but not least, in unequal counties, there are a lower number of bank branches per capita.[13] Second, we find the relationship between wealth inequality and establishments entries and exists disappears when we control for education, banking development and the quality of the judiciary in the regressions, suggesting that wealth inequality strongly correlates with the quality of local institutions.

In the last part of the analysis, using data from the Kauffman survey of entrepreneurial dynamics, we find that wealth inequality increases the likelihood that a firm is a sole-

---

[12] The view on Old Money and economic outcomes has been ambiguous. On one hand, Old Money may exacerbate problems related to institutional development. Old Money families could promote their interests at the expense of society via the control of large corporations and business pyramids (Morck, Wolfenzon and Yeung (2005); Morck, Yavuz and Yeung (2011). On the other hand, via philanthropy and charitable activities, Old Money may lessen the adverse effects of inequality and even contribute to sound institutional development.

[13] We follow Rajan and Ramcharan (2011) and use the number of bank establishments per capita as proxy for the supply of debt finance.

proprietorship and boosts its proportion of family and bank financing. In addition, these results are in line with theories that relate wealth inequality with bad institutions. If law enforcement is weak, entrepreneurs should more commonly use debt and family financing (Modigliani and Perotti (2000)). If the quality of education is poor, entrepreneurs may choose to work with simpler technologies relying on simple forms of financing.

In sum, our findings vividly demonstrate the importance of inequality in determining entrepreneurship and the type and amount of financing entrepreneurs receive.[14] Our analysis also adds to a growing literature on finance and inequality. While most of the work in this area studies how finance may affect the degree of income or wealth inequality (see Demirgüç-Kunt and Levine (2009) and, more recently, Beck, Levine and Levkov (2010) for a review), our paper studies how wealth inequality affects financial outcomes (and, in this sense, it is more similar to Rajan (2009), Degryse, Lambert and Schwienbacher (2013) and Bagchi and Svejnar (2015)).

The rest of the paper is organized as follows. Section II discusses the testable hypotheses and the empirical methods and introduces our measures of wealth inequality in greater detail. Section III discusses the results on local wealth inequality and firm creation. Section IV links local wealth inequality to local bank presence, education and the efficiency of the judicial system. Section V explores the relationship between wealth inequality and a firm's technology and financing choices. It also tackles further endogeneity issues using a falsification test and a difference-in-differences analysis that exploits the abandonment of EIG taxes by different states at different points in time. Section VI concludes the paper.

[14] See Black and Strahan (2002), Berkowitz and White (2004), Kerr and Nanda (2009) and Kerr and Nanda (2010), and reviews by Kerr, Nanda and Rhodes-Kropf (2014) and Carlino and Kerr (2015), among others.

II. Inequality and Entrepreneurial Outcomes

A. Background

In addition to appearing in Adam Smith's writings, the idea that wealth inequality may be detrimental to economic outcomes and business formation can also be found in North (1981), pp. 111, Engerman and Sokoloff (2002) and Acemoglu and Robinson (2013), e.g., pp. 152-165, among others. The common theme behind these studies is that the wealthy may use their influence to distort the development of local financial markets, schooling, and the judicial system to promote their own interests. While unambiguous anecdotes exist for many countries, for the US, there is published empirical evidence at the local level. Rajan and Ramcharan (2011), for example, present evidence that, in the 1930s, wealthy elites restricted the entry of banks in counties in which they held power. In addition, it is commonly accepted that wealthy neighborhoods have traditionally favored financing of public schools based on property taxes. This system implies that there is no redistribution to poorer neighborhoods and leads to large disparities in the amount of funding available to schools located in different areas. As states began to introduce some form of redistribution in school financing during the 1980s, wealthy neighborhoods lobbied for lower property taxes because they did not want to redirect resources to poorer districts (see Stark and Zasloff (2003) for a discussion). Card and Payne (2002), for example, show how the effects of the redistribution from wealthier to poorer school districts narrowed the gap in student outcomes between richer and poorer families. Other empirical analyses present evidence consistent with the idea that inequality could be detrimental to educational outcomes. Goldin and Katz (1998), for example, show that the degree of homogeneity in wealth across states positively correlates with the support for public education. Similarly, Galor, Moav and Vollrath (2009) show that expenditures on US public education at the beginning of the twentieth century were lower in states that displayed higher degrees of wealth inequality.

In addition, in terms of the judiciary, the election of state judges may give rise to the possibility that wealthy individuals distort judicial decisions in their favor by contributing to judges' electoral campaigns. Supporting this possibility, a *New York Times* article published in 2006, for instance, documents that Ohio Supreme court judges ruled in favor of their contributors 70 percent of the time (Liptak and Roberts (2006); also Berkowitz and Clay (2011), p. 133).

An alternative view considers wealth inequality as a positive factor of the economy. With their deep pockets, wealthy individuals may help stabilize aggregate demand and reduce the average cost of production of new, high-tech goods, making them more accessible to the middle class (Matsuyama (2002)). Philanthropy and charity may produce enough resources to extend schooling and economic opportunities to the poor, enhancing human capital formation. In a 2001 article, *Forbes* magazine, for example, espouses the view that, traditionally, charity in the United States "*concentrated in education and acculturation*" and "*stressed the skills and attitudes of self-reliance and personal responsibility.*"

B. Empirical Strategy

In our main analysis, we will estimate the impact of wealth inequality on the number of establishments' entries and exits using data at the MSA level, as well as on capital structure and technology choices of startups.[15] In particular, for entry and exit, we will estimate the following equation:

[15] In the analysis of entry and exit of new firms, we use data at the MSA level, as data at the county level are not freely available. In the other regressions, we will consider measures of inequality at the county level to consider firms and regions that do not belong to MSAs. The results do not change if we use MSA inequality measures in the capital structure and institutional regressions.

$$Y_{j,t} = \alpha + \alpha_s + \alpha_t + \beta Wealth\ Inequality_j + Controls_{j,t-1} + \varepsilon_{j,t}$$

Where, $Y_{j,t}$ indicates the natural logarithm for the number of establishments' entries and exits in the Metropolitan Statistical Area j at year t. In line with Kerr and Nanda (2010), we will focus on gross business entry and exit.[16] The variable *Wealth Inequality* indicates one of our measures of local wealth inequality. *Controls* stands for a set of MSA controls, such as population, income per capita and housing prices, all detailed in Appendix Table A.I.

C. Identification

Wealth inequality could be correlated either with omitted factors or with the degree of entrepreneurship itself. The possibility of reverse causality is based on the fact that entrepreneurs are a small fraction of the population but hold a large share of the total wealth (Cagetti and De Nardi (2008)). When entrepreneurial activities are successful, most of the rewards are accrued among a limited number of individuals, which, in turn, increases wealth inequality. From this perspective, we may expect to find a positive correlation between wealth inequality and entrepreneurship.

Wealth inequality could also be correlated with policies introduced by states, local income, income/wage inequality and the racial composition of the geographical area. While we introduce variables that expressively control for these factors, we also address this problem in several ways.

First, as we discussed, our measure of wealth inequality is constructed at either the MSA or the county level, which allows us to control for state fixed effects and state trends in the analysis. State fixed effects and state trends are a relevant feature of our identification strategy,

[16] Results are similar if we use per capita business entry and exit as alternative dependent variables.

as other main deeply rooted determinants of institutions, such as legal and colonial origins, are defined at the state level. Berkowitz and Clay (2011) give a precise overview of which US states have civil law (rather than common law) traditions and link their legal traditions to the countries of origin of early settlers.

Second, we make use of the available historical literature to generate an instrumental variable analysis. Engerman and Sokoloff (1997) and Engerman and Sokoloff (2002) describe the factors that can be underlying causes of persistent differences in inequality: different climates and geographical environments that may favor the production of one type of crop over another. These authors' argument suggests that climates that are best suited to large plantations, such as sugar or tobacco plantations, will induce relatively high economic inequality. The production of these crops comes at a high fixed cost; as a result, in equilibrium, the market can support only a few farms. The outcome is thus a society controlled by few wealthy landowners. Conversely, climates supporting crops such as wheat will result in a more equal society. The production of these crops does not require high fixed costs; hence, the market can "bear" more producers. These societies will be more equal and be composed mainly of small landowners. A feature of this theoretical framework is that inequality and "bad" institutions will be persistent over time and reinforce each other. Along these lines, Acemoglu and Robinson (2013) and Rajan (2009) also provide a theoretical framework and empirical evidence of how institutions may persist through time.

We will also construct a falsification test to corroborate the validity of our instrumental variables. Our identification strategy implicitly assumes that local inequality has an impact on entrepreneurship because of the quality of local institutions. We will relate local weather conditions to entrepreneurship in France, a country with a similar level of development as the US but where, unlike in the US, local authorities have a very limited or nonexistent role in shaping local institutions, such as schooling and the judiciary. Given that we do not have local inequality data for France, we will work with reduced-form equations linking local weather conditions to local business entry and exit for both France and the US (*à la* Nunn and Wantchekon (2011)). Finding that local weather conditions have an impact on business formation in the US – because, in the US, local institutions matter, but not in France − should justify the identification assumption.

Third, we exploit changes in EIG taxes that took place in various states between the 1970s and 2000. At different points in time beginning in 1976, 31 states repealed their "death" taxes. In particular, states switched from a system in which state EIG taxes were a percentage computed on top of the corresponding federal EIG tax to a "pick up" system in which the state only "picks up" a proportion of the Federal EIG tax applied to its citizens without increasing the total tax burden.

EIG taxes may matter for wealth inequality, as they define the amount of wealth transferred from one generation to the next. In principle, systems with very high EIG taxes should promote more equality, as wealthy parents will not be able to transfer all their wealth to their children. Conversely, low EIG taxes should make it easier to pass wealth from one generation to the next, promoting more inequality. By preserving high levels of wealth throughout generations, low EIG taxes, may contribute to maintain a system of local public policies that favour the wealthy.[17]

D. Measuring Wealth Inequality

Obtaining representative measures of wealth inequality at the local level proved to be difficult. As a result, we construct our own two proxies for local wealth inequality. The first

[17] Low EIG taxes may even reinforce such policies to the extent that they induce migration of wealthy elderly in the area. For instance, it appears that wealthy elderly favor a system of property taxes intended to finance only the local school district and they do not want to redistribute to other districts (see Brunner and Balsdon (2004)).

one is based on current levels of financial wealth and is broadly based on a methodology introduced by Mian, Rao and Sufi (2013) and Saez and Zucman (2014); it intends to construct local-level measures of household net worth. The second measure is based on historical records of land ownership.

The contemporary measure of wealth inequality looks at the amounts of dividends and interest earned by US households in 2004, the first year in our sample period, as reported by Internal Revenue Service (IRS) Statistics of Income (SOI) data. The IRS-SOI data report the total amount of dividends and interest income received by US households in a certain zip code. The information is reported as a total amount per zip code and is divided into five households' income groups, ranging from low income to high income. Under the assumption that a typical household owns the market index for stocks and bonds, the amount of financial rents it receives depends only on the quantity of stocks and bonds it holds. We use this information to construct a Gini index of wealth inequality based on financial rents. The procedure we adopted to construct the index is detailed in Appendix Table A.II. In an addition to the Gini index, we construct an alternative measure of wealth inequality: the proportion of households in the MSA/County that do not have financial wealth. The results we obtain with this alternative measure are the same (and, if anything, stronger) as those obtained with the Gini index.[18] The average Gini coefficient we obtain is 0.44 with a standard deviation of 0.14. These figures are in line with measures of household wealth inequality obtained at the aggregated level. For example, De Nardi (2004) shows that the Gini coefficient for the entire US is 0.78 based upon household wealth data from the Survey of Consumer Finances from 1989. Relying on the same survey, Wolff (2010) finds that the Gini coefficient is 0.83 for 2007.

---

[18] The granularity of the SOI data does not allow us to construct traditional measures of wealth inequality, such as the percentage of wealth detained by the top richest individuals in an area (because SOI distributes households into bins according to their total income and not according to the amount of dividend or interest income earned).

To construct our historical measure of wealth inequality, we obtain information on farmland sizes at the county level from the 1890 US Census. More precisely, for each county, we have information on the total number of farms that – based upon their total acres of farmland – fall within a certain size bin. Farms are assigned to one of seven bins: under 10 acres, from 10 to 19 acres, 20 to 49 acres, 50 to 99 acres, 100 to 499 acres, 500 to 999 acres, and 1,000 or more acres.[19]

We also find that 1890 land inequality displays a 36- and 46-percent positive correlation with our measures of dividend and interest inequality, respectively. The historical measure is also correlated with other contemporary socioeconomic measures that may reflect the degree of wealth inequality. It displays a positive correlation with local poverty rates (43 percent) and the number of crimes per capita (33 percent) and is negatively correlated with the number of white people (a rough proxy of the size of the middle class) living in a county (-53 percent).

III. Business Dynamics

A. Data Sources

We obtain data on establishments' entry and exit from the Business Dynamics Statistics (BDS), a database set up by the US census that provides annual measures of, amongst other things, establishment births and deaths, and firm startups and shutdowns. The BDS data are available only at the US MSA level and provide information from 1976 until 2012. It covers a wide range of industrial sectors including agriculture, manufacturing, wholesale trade, retail trade and services (amongst others). Following Kerr and Nanda (2009), we define entrepreneurship as the entry of new, stand-alone firms. From the Kauffman Firm Survey

---

[19] The average Land Gini coefficient in our dataset is 0.44, and its standard deviation is 0.14. These figures are slightly lower than other contemporary measures of household wealth inequality at the aggregated level.

(KFS) panel dataset, we extract the financial information for a five-year period from 2004 to (and including) 2008 on 4,928 individual US start-ups during their early years of operation (see Robb and Robinson (2014) for a comprehensive discussion of the capital structure choices of firms covered by this survey). This information is particularly useful to reconstruct the sources of financing of these young firms and allows us to distinguish between family, bank and venture capital financing. The Kauffman Firm Survey tracks start-up firms in their early years of operations and mainly consists of smaller, sole-entrepreneur firms. For example, in the first year of operations, 60% percent of the firms did not have any employee.

We collect the data on our main dependent variables for our later capital structure regressions from a restricted-access-only database, the so-called "Fourth Follow-Up Database," which is a longitudinal survey. We analyze the 3,419 firms of the baseline survey that either survived over the entire 2005-2008 period or were specifically identified as going out of business during the same period.[20] Hence, firms that dropped out in a specific year because their owners could not be located or refused to respond to the follow-up survey are not included in our analysis.[21] The dataset contains response-adjusted weights (which we use) to minimize the potential non-response bias in the estimates. From this database, we construct several crucial financial outcome variables, as well as control variables in the form of firm and main owner characteristics. We download the various state, MSA and county characteristics from the US Census Bureau.

Our proxy for Old Money is based on the *Forbes* list of the 500 wealthiest American individuals. We start from Kaplan and Rauh (2013), who augment the *Forbes Top 500* for 1982, 1992 and 2001. They provide additional information, such as the origin and the

[20] This time period allows us to abstract from confounding events related to the "Great Recession."

[21] This is common in the literature; see, e.g., Berger, Cerqueiro and Penas (2014) and Robb and Robinson (2014).

“generation” of the wealth involved.[22] We then integrate these data with information about the state of residence. This information is not available in the Kaplan and Rauh (2013) file, but we retrieve it from the original *Forbes* issues. For each state, we then compute the average generation of wealth for the top 500 resident individuals. Our premise is that when local inequality originates in Old Money, the generation of wealth will be greater.[23]

B. Results

1. Business Dynamics

We begin by testing how local wealth inequality affects business dynamics in US *Metropolitan Statistical Areas* (MSAs). Table II provides the first estimation results. We relate the local measure of inequality based on financial wealth to the yearly number of new establishments, as well as to the total number of establishments that become inactive in a given year.

[Table II around here]

In Column (1) in Table II, we begin with our baseline estimation using a standard Ordinary Least Squares (OLS) model. We include state and year fixed effects and control for the MSA population. In Column (2), we repeat the estimation, including an extra set of MSA characteristics. In Specification (3), we include two additional control variables that could be

[22] They compute the number of generations of wealth involved by identifying the founding date of the business that generated it and then by counting to which generation the current wealthy individual belongs. Hence, the resultant “generation” of the wealth involved is usually an integer, but if the individual inherited a relatively small business and built it into a much larger one, it could be coded as 1.5, for example.

[23] From Forbes, we also compute the proportion of the wealthiest individuals whose main source of wealth is inheritance. In an unreported test, we also use this measure as a proxy for Old Money. The results are fully in line with those based on the other two proxies.

correlated with (and could partially capture) wealth inequality: a measure of ethnic diversity in a given MSA and wage inequality. In Specification (4), we add state-year fixed effects. In all specifications, we find that the number of new establishment entries in an MSA decreases with inequality: a result that lends support to the notion that wealth inequality may be harmful for business formation. The effect we find is also economically significant and stable across specifications: a one-standard-deviation increase in MSA wealth inequality decreases the number of new establishment entries by approximately 8-12 percent.

The decline in Column (3) is driven mostly by the introduction of wage inequality, but the economic significance remains sizable: a one-standard-deviation increase in inequality reduces establishment formation by 8 percent.

Columns (5) to (8) confirm that local inequality matters for business formation dynamics. These columns relate local inequality to the yearly number of establishments that become inactive (natural logarithm). Similar to the new establishment entries regressions, we include in Column 5 state and year fixed effects and control for MSA population. In Column 6 we add an additional set of MSA characteristics, whereas in Columns 7 and 8, we include a measure of ethnic diversity in an MSA, a measure of wage inequality, and state-year fixed effects, respectively. These variables' economic relevancy is practically similar: a one-standard-deviation increase in MSA wealth inequality decreases the number of establishments that become inactive by approximately 10-12 percent, indicating that business formation (i.e., establishment entries and exits) is less dynamic in more unequal metropolitan areas. Additionally, this result confirms that larger values of wealth inequality are related to lower business dynamics. In the spirit of Kerr and Nanda (2009), we also divide establishments exits between Churning exits (establishment closures within 36 months of formation) and Schumpeterian exits (closures beyond 36 months). We present the results in the Appendix Table A.III. Remarkably, wealth inequality has a negative impact both on Churning and

Schumpeterian exits. The economic significance is also similar: a one-standard-deviation increase in wealth inequality reduces both the Churning and Schumpeterian exits by approximately 10 percent.

2. *Various Forms of Inequality*

In Table III, we assess whether the relationship between inequality and establishment "churn" is altered by ethnic diversity or Old Money. There is a large body of literature connecting ethnic diversity with economic performance (Alesina and Ferrara (2005)), but few analyses relate ethnic diversity through inequality. An exception is Alesina, Michalopoulos and Papaioannou (2016), who show that economic inequality has a strong negative impact on economic growth, especially when inequality is high between ethnic groups. Wealthy ethnic groups are in a better position to control institutional developments to their own advantage, and large economic disparities between ethnicities may increase the probability of conflicts and wars.

Following Easterly and Levine (1997), we measure ethnic diversity with a Herfindahl index:

$$Ethnic\ Diversity = 1 - \sum_{i} s_i^2$$

Where s_i is the share of group i over the total population. An ethnic diversity index equal to zero means a fully homogeneous population, while an index equal to one corresponds to complete heterogeneity.

Table III in Column (1) looks at the ethnic fractionalization of the local population measured with this ethnic diversity index. We find that inequality has a negative and statistically significant effect on establishment formation, while the interaction term with ethnic diversity is negative and statistically significant. This result suggests that the effects of wealth inequality are stronger when the community is ethnically more heterogeneous. In communities with a maximum degree of ethnic heterogeneity (i.e., ethnic diversity is equal to 1), a one-standard-deviation increase in wealth inequality leads to an 18-percent decline in establishment formation. This effect is reduced to minus 8 percent when ethnic diversity in the interaction term is equal to its sample mean. If, across MSAs, the wealthiest individuals are white and the poorest are African-American, for example, our results suggest that wealth inequality may have an even stronger negative impact on business formation. This result is in line with Alesina, Michalopoulos and Papaioannou (2016), who find that, across countries, ethnic inequality is linked to worse economic performance.

[Table III around here]

We then consider the interaction between inequality and the generation of wealth. Recall that, on one hand, Old Money may exacerbate the stunting of institutional development, as Old Money may promote their interests at the expense of society, for example, via their control of large corporations and business pyramids (Morck, Wolfenzon and Yeung (2005); Morck, Yavuz and Yeung (2011)). On the other hand, Old Money may mitigate the adverse effects of inequality and even contribute to sound institutional development via philanthropy and charitable activities.

Table III Column (2) interacts wealth inequality with the Forbes 500 average wealth generation in the state. We find the sign of the coefficient on wealth inequality to be negative and statistically significant, suggesting that wealth inequality alone is negatively associated with establishment entry. However, the sign of the interaction term is positive and statistically significant: the impact of wealth inequality on establishment entry is lowered when the wealthiest individuals in the region belong to Old Money. In terms of economic significance, when setting the Old Money proxy to its mean, the effect of wealth inequality is about halved.

An increase in wealth inequality by one standard deviation reduces establishment entry by 13 percent when all richest individuals are first generation, but it reduces entry by only 7 percent in MSAs where wealth generation equals 1.8 (i.e., the sample mean). This result is consistent with the idea that Old Money mitigates the effects of inequality on entrepreneurship. Either because of philanthropy or because subsequent generations of wealth lose their ability or interest in lobbying against business entry, Old Money inequality may be less harmful to new business formation than New Money.

In Columns (3) and (4), we consider the relationship with establishment exit. The coefficients on wealth inequality remain negative and statistically significant, while the coefficients on the interaction terms are again positive but not statistically significant.

3. Instrumental Variable Analysis

In this section, we more directly tackle the endogeneity issue by performing an IV analysis. In the spirit of Engerman and Sokoloff (2002) and Easterly (2007), the instrumental variables are based on past local weather conditions, i.e., they are based on the historical rainfall and temperature between 1895 and 2003 and their corresponding standard deviations. We obtain information from the National Climatic Data Center (NCDC) on local monthly precipitation and temperature (measured in inches and degrees Fahrenheit, respectively) and their corresponding standard deviations for the entire period between 1895 and 2003. We then construct simple averages of these series. The NCDC provides this weather information at the so-called "divisional" level, i.e., each state is subdivided into at most 10 divisions that comprised areas that are known to have similar climatic conditions. We assign each MSA to the state division to which it belongs. We proceed in a similar way when constructing the instrumental variables at the MSA level.

As already shown by Rajan and Ramcharan (2011) and Vollrath (2013), even within states, there is a significant amount of diversity in terms of temperature and rainfall. In Kansas and

Texas, for instance, some counties experience a yearly rainfall average of 20 inches, while others exceed 40 inches. A bit less extreme but still important are the differences in Illinois, where some counties have an average rainfall of 28 inches, while others have 30 percent more (approximately 36 inches). Similarly, in California, some counties had an average temperature of 50 F, while others have an average of 64 F.

[Table IV around here]

The *'First Stage'* column in Table IV provides the results of the first-stage regression from 2SLS regressions and indicates that, indeed, rain and temperature are significant determinants of current MSA inequality for the dependent variable *Total Establishment Entries*. All climate variable coefficients except for the standard deviation of temperature are statistically significant at the 1 percent level, and all enter with the expected sign: Higher rainfall levels and temperatures are associated with higher current MSA inequality, but at a decreasing rate, as indicated by the negative signs on the coefficients of their respective standard deviations. In all the specifications, the F-statistic of the first stage (not reported) is well above 20, confirming a powerful first stage.

The following columns in Table IV report the second-stage regressions. The results confirm our previous findings: in more-unequal MSAs, business formation is less dynamic, and the number of new establishment entries and exits is significantly lower. Again, the results are not only statistically significant but also economically relevant: a one-standard-deviation increase in MSA inequality decreases the total number of new establishments that are established in a given year by 12 percent.

IV. Wealth Inequality and Local Institutions

Our analysis so far has identified a reduced-form relationship between wealth inequality and business formation. This relationship could be mediated by many factors related to the local institutional environment. For instance, wealth inequality may result in inefficient financial markets and yield restrictions to the supply of external finance, which, in turn, may prevent local business formation. At the same time, a more-inefficient judicial system may simply discourage individuals from starting their own business.

We use data from the US Census and the Bureau of Justice Statistics to evaluate the relative importance of possible mediating factors that may underlie our results. We focus in particular on banking development, education, and the efficiency of the civil justice system.

In Table V, we first examine the effect of contemporary county inequality on banking development. As a measure of banking development we follow Rajan and Ramcharan (2011) and use the number of bank establishments per 1,000 capita. Rajan and Ramcharan (2011) shows that, in the 1930s, US counties displaying more wealth inequality had a significantly lower number of bank establishments per capita. In Column (1), we use a specification that includes state and year fixed effects and a comprehensive set of county controls. Column (2) presents the results from the second stage of a 2SLS IV regression, where we instrument local inequality with the average historical rainfall, temperature and their standard deviations. The results indicate that inequality indeed hampers banking development: the county inequality coefficient is negative and statistically significant throughout the specifications. The results are also economically meaningful: a one-standard-deviation increase in inequality decreases the number of banks per 1,000 capita, for example, by approximately 9 percent of its mean in Columns (1) and (2).

[Table V around here]

Turning to education as another institutional feature of the local environment, we present the results in Columns (3) to (6). In Columns (3) and (4), the analysis shows that, in more unequal counties, the percentage of adults with a college degree or higher is lower. In fact, a one-standard-deviation increase in county inequality decreases the percentage of adults with a college degree or higher between 16 to 45 percent of its mean, depending upon the specification. Moreover, the population inflow of educated individuals (i.e., those with at least a college degree) is also lower in more unequal counties, as seen in Columns (5) and (6). A one-standard-deviation increase in inequality results in a significantly lower inflow of educated individuals. The economic effect is also sizable: a decrease between 30 percent and 50 percent evaluated at the mean of county inflow.

In Columns (7) and (8), we assess the effect of local wealth inequality on another local institution: the judiciary. We investigate how local inequality affects the judicial efficiency as measured by the length of time to the verdict for a first-degree civil trial. We obtain data on individual civil cases from the Bureau of Justice Statistics (BJS) from 2005. The BJS reports data on civil litigations for the 75 most populous US counties. At the individual case level, we observe whether the number of days (its natural logarithm) it takes to come to a verdict in a case is affected by local inequality. We observe that local inequality matters for judicial decision making. When controlling for both state fixed effects and case controls (in the form of the number of plaintiffs involved as a proxy for case complexity and whether an appeal was granted, as well as the total number of cases in a certain court), the findings suggest that, in more-unequal counties, court rulings take more time and therefore are less efficient.

[Table VI around here]

Overall, we conclude that the results point in the direction that inequality also affects the quality of local institutions in the form of banking development, education and judicial efficiency. Indeed, when horseracing inequality with the measures of banking development and education (which, in contrast to judicial efficiency, are available for all MSAs) to explain establishment turnover (Table VI), inequality is robbed of all significance, suggesting that its impact flows mainly through institutional development.[24]

V. Wealth Inequality, Firm Type and Firm Financing

C. Background

In previous tables, we find evidence in support of less "creative destruction": fewer establishment entries and exits in more-unequal MSAs. However, does inequality also affect the type of firm formation and entrepreneurs' financing choices? To the extent that inequality is related to worse institutions, as suggested by our results, we may expect that, in unequal societies, entrepreneurs themselves prefer to undertake ventures in simple technologies or may find it difficult to finance riskier ventures because poorer schooling may lead individuals to opt for simple technology choices or because local capital markets might be wary of financing new technology. We therefore might expect to observe fewer high-technology start-ups when county inequality is larger. At the same time, if poorer institutions reduce the amount of external finance available to entrepreneurs, families will have a more important role in financing business ventures in more unequal areas. Modigliani and Perotti (2000) provide a theoretical framework that shows that debt contracts are easier to enforce (vis-a-vis equity contracts) in areas where law enforcement is weaker. Overall, these works suggest that more

[24] Including the average length of first-degree civil trials does not change our results; it simply lowers the number of observations, as we have data on civil litigations for only the 75 most populous counties.

inequality should be related to less technological change and a higher use of debt and family financing.[25]

We bring this idea to the data and test the following equation:

$$Y_{i,j,t} = \alpha + \alpha_s + \alpha_{Ind} + \alpha_t + \beta Wealth\ Inequality_j + Controls_{j,t-1} + \varepsilon_{i,t}$$

Where, $Y_{i,,j,t}$ indicates different capital structure variables: the technology choice and corporate form of startup *i* located in county *j* at year *t.*

Because, in this case, we have firm-level data, we additionally control for industry fixed effects together with the state and year fixed effects. *Controls* capture two sets of county and firm characteristics.

D. Results

We present results on this possibility in Table VII, where we examine the effect of wealth inequality on firm type, firm ownership, and firm financing. To examine whether local county inequality affects firm type choices, we construct a dummy variable that indicates whether a Firm is high tech and run a linear probability model in Columns (1) to (3). Because the definition of high tech is based on the NAICS industry classification, we do not include industry fixed effects and trends in these regressions. We do, however, include firm, owner and county controls and broad sets of fixed effects in the different specifications, i.e., state, year and state*year fixed effects, respectively. These regressions, as well as the following capital

---

[25] These hypotheses are also consistent with Perotti and von Thadden (2006), who build a model in which the median voter in an unequal society owns only her non-diversifiable human capital. As a result, she may prefer a financial system dominated by family and banks, “institutions” with whom she shares her aversion to risk. In more equal societies, the median voter may also own diversifiable financial wealth and may prefer a system that also relies on equity financing and is characterized by risk-taking dynamism.

structure regressions, use measures of inequality at the US county level. The results are similar when we use MSA measures of inequality.[26]

[Table VII around here]

Interestingly, the coefficient on county inequality is always negative and statistically significant at the 5 percent level, indicating that the likelihood that newly created firms are high tech decreases with local inequality. The effects are also economically relevant. Depending upon the specification, a one-standard-deviation increase in inequality decreases the likelihood that a firm is high tech by between 8 and 9 percent of its mean when introducing a contemporary inequality measure. In unequal societies, entrepreneurs themselves may prefer simpler technologies or may find it difficult to finance riskier ventures.

The results are suggestive of the possibility that banks are more willing to extend financing to conservative industries, making it difficult for new start-ups in more unconventional industries (such as high-tech industries) to obtain bank financing in order to establish a business. Another explanation may be demand driven: in more-unequal counties, entrepreneurs may prefer simpler technologies themselves.

We continue by testing how local wealth inequality affects firm ownership as measured by the corporate form in which the firm is established. Table VII, Columns (4) to (6) provide the first estimation results for the dependent variable *Firm Is Proprietorship*. We relate local inequality to the probability that a start-up is a sole proprietorship. In all specifications, we find that the probability of a start-up to be a sole proprietorship increases with inequality. The point

[26] When possible, we prefer using county measures of inequality for two reasons: (1) many of the institutional activities taking place in the local areas, such as schooling and the judiciary, are organized at the county level; (2) by estimating on a more local level and “closer to the firms,” county-level measures of inequality allow us to obtain a better control of omitted factors.

estimates increase slightly across Columns (5) and (6). The effect we find is also economically significant and stable across specifications: A one-standard-deviation increase in county wealth inequality increases the probability for a start-up to be a sole proprietorship by approximately 14 percent (evaluated at the mean of the proprietorship indicator variable).

Next, we consider firm financing. Columns (7) to (12) in Table VII present the results of our OLS analyses.[27] We focus on the proportion of bank and family financing in Columns (7) to (9), controlling for the usual firm, owner and county controls, as well as dense sets of fixed effects to capture unobserved state, year, industry, state-year and industry-year heterogeneity, respectively. The results in Column (7) show that the coefficient on county inequality is positive although not statistically significant. The result indicates that a one-standard-deviation increase in county inequality increases the proportion of bank and family financing by 7 percent (of its own mean) when using a contemporary inequality measure. When including a set of county controls, firm characteristics and industry fixed effects in Column (8), the effect doubles to 14 percent, and the coefficient is statistically significant at conventional levels. When we add state-year and industry-year effects in Column (9), the estimated coefficient on inequality is again statistically significant (at the 5 percent level) and implies that a one-standard-deviation increase in county inequality increases the proportion of bank and family financing again by approximately 14 percent.[28]

[27] We also re-estimate all specifications using a Tobit model. Maximum likelihood estimators of marginal effects in Tobit models are found to be overall much less biased due to the incidental parameter problem (than those in binary dependent variable models); however, when many fixed effects are introduced, expected biases in the slope estimators (in terms of marginal effects) emerge away from zero; at the same time, the estimated standard errors may be biased towards zero (Greene (2004)). We therefore report the results of using OLS.

[28] Our finding that, in more-unequal areas, there are a lower number of bank establishments per capita while entrepreneurs use more bank financing may seem contradictory. In reality, these results are consistent with two non-mutually exclusive explanations. The first possibility is that the decline in the supply of bank establishments in more unequal areas is more than offset by a larger demand for bank debt requested by local entrepreneurs. This explanation seems sensible in light of the fact that, in more-unequal areas, entrepreneurs use more-traditional technologies that do not require complex forms of financing. A second explanation looks more heavily at the monopoly power gained by banks in more-unequal areas. Banks become more monopolistic, reduce the supply of

In the remaining columns of Table VII, we again explore using OLS estimation how county inequality affects the proportion of angel and venture capital financing. The proportion of angel and venture capital financing decreases with county inequality, as we can see from the results in Columns (10) to (12). The coefficient on inequality is negative and statistically significant and robust across specifications; however, the economic significance is somewhat larger in the specifications in Columns (10) and (12) when including sets of firm characteristics, county controls, industry, state-year and industry-year effects, respectively. A one-standard-deviation increase in county inequality decreases the proportion of angel and venture capital financing by between 26 and 44 percent, depending upon the specification. External equity financing decreases with county inequality either because of a lower supply, in line with Perotti and von Thadden (2006), or a lower demand because entrepreneurs prefer more-traditional technologies. The results show that local inequality matters for firm financing. The findings are also consistent with the papers by Perotti and von Thadden (2006) and Modigliani and Perotti (2000).

Chen, Gompers, Kovner and Lerner (2010) show that the distribution of venture capitalists in the US is concentrated in three areas: San Francisco, Boston and New York.[29] We therefore verify whether our results are not simply driven by firms located in these areas by excluding all firms located in the states of California, Massachusetts and New York. We find that the results are unaffected (and therefore not tabulated). We also repeat the previously employed

---

their services, and simultaneously block competing forms of financing to supply services in the area. This second explanation is in line with the findings of Rajan and Ramcharan (2011).

[29] Chen, Gompers, Kovner and Lerner (2010) reports that more than 49 percent of US-based companies financed by venture capital firms are located in one of these three areas. The Kauffmann survey data we use combine venture capital with angel financing. The number of firms with non-zero angel and venture capital financing located in California, Massachusetts and New York corresponds to only 19 percent of the total number of firms, suggesting that angel financing may be less concentrated.

instrumentation strategy across all specifications in Table VII and tabulate the estimates in Appendix Table A.IV. The results are unaffected.

E. Further Endogeneity Issues

1. Inequality from 1890

To rule out any reverse causality problems and provide an alternative proxy for local wealth inequality, we obtain a historical measure of local wealth inequality based on historical farmland data from 1890, which we introduce as a second measure of local inequality in Table VIII. In addition, to account for omitted variables because of unobserved heterogeneity at the state, year, industry, state-year and industry-year level that could affect our estimates, we correspondingly introduce a broad set of fixed effects.

[Table VIII around here]

The results confirm our previous findings: Columns (1) and (2) show that business dynamics are hampered in more-unequal MSAs, i.e., in more-unequal MSAs there is not only less business entry but also less exit.[30] The estimated coefficient on inequality for establishment entry is also statistically significant at conventional levels. A one-standard-deviation increase in historical wealth inequality reduces establishments' entries and exits by approximately 3 percent. The probability that a start-up firm is of a high-tech nature decreases with county inequality, whereas the probability that a new venture is a proprietorship increases with county inequality, as seen in Columns (3) and (4). Both coefficients are statistically significant, and a

[30] Because we use a measure of wealth inequality based on 1890 data, in these regressions, we use the full time period (i.e., 1976-2012) available for the data on establishment entries and exits. The results are the same if we restrict the sample to the post-2004 period.

one-standard-deviation increase in county inequality alters both probabilities by approximately 3 percent (evaluated at their respective means). For *Firm Bank and Family Financing*, the coefficient on historical county inequality enters significantly and with a positive sign, in line with previous findings, in Column (5): a one-standard-deviation increase in inequality increases financing obtained from banks and family members by a bit less than 5 percent (evaluated at its mean). The coefficient on county inequality enters with a positive sign in Column (6) for the dependent variable *Angel and Venture Capital Financing*, contrary to previous findings, but it is not statistically significant.

2. *Falsification Test*

Our identification strategy implicitly assumes that local inequality has an impact on entrepreneurship via the quality of local institutions. To address the concern about whether the exclusion restriction is satisfied, we perform a falsification test that links local weather conditions to local entrepreneurship in France.

In France, local authorities have much more limited power to organize public life than authorities in the US. In our test, we understand departments as the French equivalent to US counties and French Administrative Regions as equivalents to US states. French departments constitute the second of three levels of government below the national government. They are smaller than the 27 administrative regions but larger than towns. The main areas of responsibility of a department include the organization of various welfare allowances, the maintenance of school buildings and local roads, and the contribution to municipal infrastructures. Importantly, and differently from the US, French departments do not have any role in organizing the judiciary, and their administrative activities are supervised by a prefect, the high representative of the national government.

Because we do not have local inequality data for France, we work with reduced-form equations linking local weather conditions to local business entry and exit for both France and the US. We obtain data on establishments' entry and exit between 2006 and 2014 from INSEE, the national statistics office, and data on historical temperature and rainfall from METEO-France, the national meteorology institute.

We run reduced-form regressions linking the local degree of firms' entry and exit to local weather patterns and control for regional and year fixed effects. Naturally, weather conditions are more extreme and likely to be more diverse within the same state in the US compared to the French departments. To control for this problem, we also perform the regression for the US, restricting the sample to US MSAs, whose average and standard deviation of rainfall and temperature fall within the respective French minimum and maximum averages and standard deviations. We cluster standard errors at the state/regional level.[31] The results are shown in Table IX.

[Table IX around here]

Our instrumental variables – the average historical rainfall and temperature and their respective standard deviations – significantly affect establishment entry and exit levels in the US. Especially when restricting the sample of observations to within the minimum and maximum of French rain and temperature means and standard deviations in Columns (3) and (4), this significant effect is very pronounced. In contrast, there is no significant relationship between our weather variables and business dynamics in France, as seen in Columns (5) and (6). Overall, these findings support the validity of the exclusion restriction when using rain and temperature as instrumental variables for local inequality.

[31] Because France has only 21 regions, the low number of clusters may lead to bias towards results that are not statistically significant. The results remain the same when we cluster at the department level: in this case, we have 83 clusters.

3. *The Removal of State EIG Taxes and Entrepreneurship Dynamics*

To further address endogeneity concerns, we exploit changes in Estate Inheritance and Gift taxes that took place in various states between the 1970s and 2000. These widespread changes in state EIG taxes show substantial cross-sectional and time series variations, which we exploit in our analysis. In this analysis, we focus only on the entry and exit of new establishments, as the Kauffman data are available from 2004 onwards, and we estimate the following equation:

$$Y_{j,t} = \alpha + \alpha_{msa} + \alpha_t + \beta Post_{jt} + Controls_{j,t-1} + \varepsilon_{j,t}$$

Where, $Y_{j,t}$ indicates the natural logarithm of the number of firms' entries and exists in the Metropolitan Statistical Area *j* in year *t*. The variable of interest is $Post_{jt}$. Post is a dummy variable that takes the value of 1 for the years following the introduction of the so-called 'pick-up' system in the state to which the MSA belongs. Similar to Kerr and Nanda (2010), we also study specifications in which we substitute $Post_{jt}$, with the (log) number of years since the reform was introduced. Because the data have a panel dimension, we control for MSA and year fixed effects. Because the EIG tax reforms are defined at the state level, we cluster the standard errors at the corresponding state level. We consider data between 1976 and 2000. In 2001, the federal government introduced legislation that phased out the pick-up system, generating an important confounding event.[32]

[Table X around here]

[32] Most states responded to the federal legislation by re-initiating a state EIG tax equal to the amount of the federal credit as determined by the IRS code as of January 2001.

We present the results in Table X. In Columns (1) and (2), the dependent variable is the log of the total number of firm entries. Column (1) reveals that the coefficient on the post variable is negative and statistically significant at the 5 percent level, indicating that switching to a pick-up system reduces the number of new firms that enter the MSA. Column (2) shows that the coefficient on the number of years since reform is also negative and statistically significant. To the extent that lower EIG taxes promote more wealth inequality, this result is in line with our baseline results that show a negative relationship between local wealth inequality and entrepreneurship dynamics. After the introduction of the new pick-up system, new business formation reduced by 4 percent. This effect is sizable, considering that the mechanism we attempt to identify relies on institutions and the provision of local public goods, which take a long time to change. Importantly, our effect is especially driven by States that deregulated earlier and, in principle, had more time to change the institutional settings.

The introduction of the pick-up system tends to decrease the number of firm exits when we consider the number of years since reform. The economic significance is similar to the economic significance of firms' entry. After the enactment of the pick-up system, the number of exits declined by 3.5 percent.

VI. Conclusions

We empirically test hypotheses emanating from recent theory showing how household wealth inequality may determine both entrepreneurial dynamism and corporate financing. Local wealth inequality may be associated with poorer institutions, leading entrepreneurs to choose simpler corporate forms for their businesses and to rely on bank and family finance.

In our empirical analysis, we employ two measures of wealth inequality: one based on the current distribution of dividends and another that relies on the distribution of landholdings within US counties in 1890. To overcome endogeneity problems, we saturate specifications with comprehensive sets of fixed effects and characteristics and estimate instrumental variable models. Additionally, we exploit the removal of EIG taxes in various states between the 1970s and 2000 in a difference-in-differences framework.

The estimated coefficients suggest that local-level wealth inequality robustly decreases firm creation, particularly of the high-tech type, and decreases firm exit. At the same time, inequality increases sole-ownership and the proportion of equity, family and bank financing, yet decreases angel and venture capital financing.

The effects of wealth inequality on entrepreneurs' financing and technology could be mediated by various factors. We find that, in more unequal counties, there are fewer bank establishments per capita, consistent with the existence of local credit rationing. However, we also find that, in more-unequal counties, entrepreneurs are less likely to have a college degree (or higher). In principle, these entrepreneurs could be more likely to work with traditional technologies that require simpler forms of financing. Moreover, our findings also suggest the presence of a less effective judiciary in more-unequal counties.

References

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

Acemoglu, D., and J. A. Robinson, *Why Nations Fail. The Origins of Power, Prosperity and Poverty* (New York NY: Random House, 2013).

Alesina, A., and E. L. Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43 (2005), 762-800.

Alesina, A., S. Michalopoulos, and E. Papaioannou, "Ethnic Inequality," *Journal of Political Economy*, (2016), Forthcoming.

Bagchi, S., and J. Svejnar, "Does Wealth Inequality Matter for Growth? The Effect of Billionaire Eealth, Income Distribution, and Poverty," *Journal of Comparative Economics*, 43 (2015), 505-530.

Beck, T., R. Levine, and A. Levkov, "Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States," *Journal of Finance*, 65 (2010), 1637-1667.

Berger, A. N., G. Cerqueiro, and M. F. Penas, "Market Size Structure and Small Business Lending: Are Crisis Times Different from Normal Times?," (Columbia SC: University of South Carolina, 2014).

Berkowitz, D., and K. B. Clay, *The Evolution of a Nation: How Geography and Law Shaped the American States* (Princeton NJ: Princeton University Press, 2011).

Berkowitz, J., and M. J. White, "Bankruptcy and Small Firms' Access to Credit," *RAND Journal of Economics*, 35 (2004), 69-84.

Black, S. E., and P. E. Strahan, "Entrepreneurship and Bank Credit Availability," *Journal of Finance*, 57 (2002), 2807-2834.

Brunner, E., and E. Balsdon, "Intergenerational Conflict and the Political Economy of School Spending," *Journal of Urban Economics*, 56 (2004), 369-388.

Cagetti, M., and M. De Nardi, "Wealth Inequality: Data and Models," *Macroeconomic Dynamics*, 12 (2008), 285-313.

Card, D., and A. A. Payne, "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores," *Journal of Public Economics*, 83 (2002), 49-82.

Carlino, G., and W. R. Kerr, "Agglomeration and Innovation," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics Volume 5A* (Amsterdam: Elsevier, 2015), 349-404.

Chen, H., P. A. Gompers, A. Kovner, and J. Lerner, "Buy Local? The Geography of Successful Venture Capital Expansion," *Journal of Urban Economics*, 67 (2010), 90-102.

Chetty, R., N. Hendren, P. Kline, and E. Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," (Cambridge MA: National Bureau for Economic Research, 2014).

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner, "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility?," (Cambridge MA: National Bureau for Economic Research, 2014).

Conway, K. S., and J. C. Rork, "Diagnosis Murder: The Death of State Death Taxes," *Economic Inquiry*, 42 (2004), 537-559.

De Nardi, M., "Wealth Inequality and Intergenerational Links," *Review of Economic Studies*, 71 (2004), 743-768.

Degryse, H., T. Lambert, and A. Schwienbacher, "The Political Economy of Financial Systems: Evidence from Suffrage Reforms in the Last Two Centuries," (London: CEPR, 2013).

Demirgüç-Kunt, A., and R. Levine, "Finance and Inequality: Theory and Evidence," *Annual Review of Financial Economics*, 1 (2009), 287-318.

Easterly, W., "Inequality Does Cause Underdevelopment: Insights from a New Instrument," *Journal of Development Economics*, 84 (2007), 755-776.

Easterly, W., and R. Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112 (1997), 1203-1250.

Engerman, S. L., and K. L. Sokoloff, "Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States," in Stephen Haber, ed., *How Latin America Fell Behind* (Stanford CA: Stanford University Press, 1997), pp. 260-304.

Engerman, S. L., and K. L. Sokoloff, "Factor Endowments, Inequality, and Paths of Development Among New World Economies," (Cambridge MA: National Bureau of Economic Research, 2002).

Engerman, S. L., and K. L. Sokoloff, "Factor Endowments, Inequality, and Paths of Development among New World Economies," *Journal of LACEA Economia*, 3 (2002), 41-109.

Galor, O., O. Moav, and D. Vollrath, "Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence," *Review of Economic Studies*, 76 (2009), 143-179.

Glaeser, E., J. Scheinkman, and A. Shleifer, "The Injustice of Inequality," *Journal of Monetary Economics*, 50 (2003), 199-222.

Goldin, C., and L. F. Katz, "The Origins of State-Level Differences in the Public Provision of Higher Education: 1890 to 1940," *American Economic Review Papers and Proceedings*, 88 (1998), 303-308.

Greene, W., "The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects," *Econometrics Journal*, 7 (2004), 98-119.

Guiso, L., P. Sapienza, and L. Zingales, "Does Local Financial Development Matter?," *Quarterly Journal of Economics*, 119 (2004), 929-970.

Kaplan, S. N., and J. D. Rauh, "Family, Education, and Sources of Wealth among the Richest Americans, 1982-2012," *American Economic Review*, 103 (2013), 158-162.

Kerr, W. R., and R. Nanda, "Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship," *Journal of Financial Economics*, 94 (2009), 124-149.

Kerr, W. R., and R. Nanda, "Banking Deregulations, Financing Constraints, and Firm Entry Size," *Journal of the European Economic Association*, 8 (2010), 582-593.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf, "Entrepreneurship as Experimentation," *Journal of Economic Perspectives*, 28 (2014), 25-48.

Lerner, J., "Venture Capitalists and the Oversight of Private Firms," *Journal of Finance*, 50 (1995), 301-318.

Liptak, A., and A. Roberts, "Campaign Cash Mirrors a High Court's Rulings," *New York Times* (New York NY: 2006).

Malthus, T. R., *Principles of Political Economy* (London: W. Pickering, 1836).

Matsuyama, K., "The Rise of Mass Consumption Societies," *Journal of Political Economy*, 110 (2002), 1035-1070.

Mian, A., K. Rao, and A. Sufi, "Household Balance Sheets, Consumption, and the Economic Slump," *The Quarterly Journal of Economics*, 128 (2013), 1687-1726.

Modigliani, F., and E. Perotti, "Security Markets versus Bank Finance: Legal Enforcement and Investors' Protection," *International Review of Finance*, 1 (2000), 81-96.

Morck, R., D. Wolfenzon, and B. Yeung, "Corporate Governance, Economic Entrenchment, and Growth," *Journal of Economic Literature*, 43 (2005), 655-720.

Morck, R., M. D. Yavuz, and B. Yeung, "Banking System Control, Capital Allocation, and Economy Performance," *Journal of Financial Economics*, 100 (2011), 264-283.

North, D. C., *Structure and Change in Economic History* (New York NY: Norton, 1981).

Nunn, N., and L. Wantchekon, "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101 (2011), 3221-3252.

Perotti, E., and E.-L. von Thadden, "The Political Economy of Corporate Control and Labor Rents," *Journal of Political Economy*, 114 (2006), 145-174.

Petersen, M. A., and R. G. Rajan, "Does Distance Still Matter? The Information Revolution in Small Business Lending," *Journal of Finance*, 57 (2002), 2533-2570.

Piketty, T., *Capital in the Twenty-First Century* (Cambridge MA: Harvard University Press, 2014).

Rajan, R., and R. Ramcharan, "Land and Credit: A Study of the Political Economy of Banking in the United States in the Early 20th Century," *Journal of Finance*, 66 (2011), 1895-1930.

Rajan, R. G., "Rent Preservation and the Persistence of Underdevelopment," *American Economic Journal: Macroeconomics*, 1 (2009), 178-218.

Ramcharan, R., "Inequality and Redistribution: Evidence from U.S. Counties and States, 1890–1930," *Review of Economics and Statistics*, 92 (2010), 729-744.

Robb, A. M., and D. T. Robinson, "The Capital Structure Decisions of New Firms," *Review of Financial Studies*, 27 (2014), 153-179.

Saez, E., and G. Zucman, "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," (Cambridge MA: National Bureau of Economic Research, 2014).

Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations* (London: W. Strahan and T. Cadell, 1776).

Sonin, K., "Why the Rich May Favor Poor Protection of Property Rights," *Journal of Comparative Economics*, 31 (2003), 715-731.

Sorenson, O., and T. E. Stuart, "Syndication Networks and the Spatial Distribution of Venture Capital Investments," *American Journal of Sociology*, 106 (2001), 1546-1588.

Stark , K. J., and J. Zasloff, "Tiebout and Tax Revolt: Did Serrano Really Cause Proposition 13?," *UCLA Law Review*, 50 (2003), 101-162.

Vollrath, D., "Inequality and School Funding in the Rural United States, 1890," *Explorations in Economic History*, 50 (2013), 267-284.

Wolff, E. N., "Recent Trends in Household Wealth in the United States: Rising Debt and the Middle-Class Squeeze - An Update to 2007," (Annandale-On-Hudson NY: Levy Economics Institute, 2010).

Tables

TABLE I
DESCRIPTIVE STATISTICS FOR THE EMPIRICAL ANALYSIS OF BUSINESS FORMATION, FIRM OWNERSHIP, FINANCING AND TYPE

Variable Name	Number of Observations	Units	Mean	Standard Deviation	10%	Median (50%)	90%
Dependent Variables							
MSA Total Establishment Entries	3.276	-	1.703	4.360	199	520	3.705
MSA Total Establishment Exits	3.276	-	1.005	2.532	133	318	2.220
Firm is High Tech	15.328	0/1	0,31	0,46	0	0	1
Firm Is Proprietorship	14.051	0/1	0,35	0,48	0	0	1
Firm Bank and Family Financing	10.540	-	0,11	0,25	0,00	0,00	0,50
Firm Angel and Venture Capital Financing	7.229	-	0,02	0,11	0,00	0,00	0,00
Main Independent Variable							
County Inequality in 2004	13.875	-	0,85	0,05	0,79	0,85	0,90
MSA Inequality in 2004	3.276	-	0,87	0,04	0,82	0,86	0,92
County Inequality in 1890	13.908	-	0,44	0,14	0,28	0,42	0,64
MSA Inequality in 1890	3.080	-	0,43	0,13	0,27	0,41	0,61
Instrumental Variables							
Rain	12.757	Inch	3,04	1,08	1,35	3,16	4,35
Temperature	11.787	°F	54,73	8,45	45,12	52,89	68,92
Control Variables							
Firm Characteristics							
Firm Total Assets	14.015	(USD)	534490,10	11700000,00	5,00	27759,00	405000,00
Firm ROA	12.016	-	0,26	2,26	-0,91	0,04	1,67
Firm Tangibility	12.602	-	0,56	0,37	0,00	0,64	1,00
Firm Number of Owners	14.039	-	1,92	4,71	1,00	1,00	3,00
Main Owner Characteristics							
Main Owner Is Female	14.006	0/1	0,27	0,44	0	0	1
Main Owner Is African-American	14.050	0/1	0,07	0,25	0	0	0
Main Owner Is Hispanic	14.050	0/1	0,04	0,20	0	0	0
Main Owner Is Asian	14.050	0/1	0,04	0,20	0	0	0
Main Owner Is Born in the US	13.997	0/1	0,91	0,29	1	1	1
Main Owner's Work Experience	14.002	-	13,49	10,96	1	11	30
State and County Characteristics							
State GDP	13.875	('000 USD)	42.594	7.038	36.647	41.694	49.070
County Population	13.875	-	905.644	1.557.066	42.269	405.142	2.015.355
County Catholic to Protestant Ratio	13.870	-	0,68	0,92	0,04	0,34	1,90
County Personal Income Per Capita	13.875	(USD)	37583,18	11735,81	26208,99	35147,01	51710,01
County Wage Inequality	13.875	-	0,55	0,04	0,50	0,54	0,60
County Ethnic Diversity	13.875	-	0,26	0,13	0,10	0,24	0,48
County Land Area	13.875	sq mi	2091954,00	938190,20	964100,90	1901434,00	3472570,00
County Bank Population per Capita	15.695		0,49	0,31	0,23	0,40	0,84
County Percentage of Adult Population with a College Degree	15.695		0,19	0,08	0,10	0,17	0,30
County Population Inflow with at least a College Degree	12.580	Thousands	109,53	701,54	0,07	1,02	96,46

NOTES. The table provides the number of observations, mean, standard deviation, 10th percentile, the median (50th percentile) and the 90th percentile of all variables used in the empirical analysis. The definition of the variables can be found in Appendix Table A.I. Due to confidentiality the minimum and maximum are not reported. For the sake of brevity we do not include MSA characteristics separately.

TABLE II
MAIN SPECIFICATIONS EXPLAINING BUSINESS FORMATION

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*MSA Total Establishment Entries*				*MSA Total Establishment Exits*			
MSA Inequality in 2004	-3.600***	-2.563***	-1.906**	-2.363***	-3.464***	-2.189***	-1.574**	-1.952**
	(0.783)	(0.785)	(0.768)	(0.806)	(0.878)	(0.721)	(0.723)	(0.742)
MSA Population$_{t-1}$	1.032***	1.009***	1.011***	1.011***	1.002***	0.995***	0.991***	0.995***
	(0.018)	(0.022)	(0.022)	(0.025)	(0.020)	(0.021)	(0.021)	(0.024)
MSA Catholic to Protestant Ratio		0.015	-0.028	0.011		0.001	-0.025	-0.002
		(0.037)	(0.032)	(0.040)		(0.032)	(0.037)	(0.035)
MSA Average Age		0.208	-0.366	0.177		0.870***	0.431	0.855***
		(0.203)	(0.260)	(0.210)		(0.267)	(0.293)	(0.272)
MSA Personal Income Per Capita		0.005**	0.005**	0.005*		0.005**	0.005**	0.005*
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
MSA Land Area		0.030	0.024	0.027		0.021	0.016	0.020
		(0.025)	(0.024)	(0.027)		(0.022)	(0.022)	(0.025)
MSA House Price Index		0.002***	0.002***	0.003***		0.000	0.000	0.002***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
MSA Ethnic Diversity			-0.260*				-0.096	
			(0.144)				(0.152)	
MSA Wage Inequality in 2004			2.536***				2.309***	
			(0.494)				(0.535)	
State Fixed Effects	Yes	Yes	Yes	--	Yes	Yes	Yes	--
Year Fixed Effects	Yes	Yes	Yes	--	Yes	Yes	Yes	--
State*Year Fixed Effects	No	No	No	Yes	No	No	No	Yes
Number of Observations	3.276	3.144	3.144	3.144	3.276	3.144	3.144	3.144

NOTES. All Models are estimated with a linear regression (OLS) model. The dependent variables are expressed in natural logarithms. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE III
TYPES OF INEQUALITY

Model	(1)	(2)	(3)	(4)
Dependent Variable	*MSA Total Establishment Entries*		*MSA Total Establishment Exits*	
MSA Inequality in 2004	-1.752*	-3.385***	-1.458	-2.286**
	(0.916)	(1.058)	(0.889)	(0.977)
MSA Inequality in 2004 * Ethnic Diversity	-3.513*		-3.098	
	(1.925)		(2.326)	
MSA Inequality in 2004 * Wealth Generation		2.914*		1.390
		(1.782)		(1.411)
MSA Ethnic Diversity	2.843*	-0.268*	2.640	-0.100
	(1.676)	(0.147)	(2.008)	(0.153)
County Controls	Yes	Yes	Yes	Yes
State Fixed Effects	--	--	--	--
Year Fixed Effects	--	--	--	--
State*Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	3.144	3.144	3.144	3.144

NOTES. All Models are estimated with a linear regression (OLS) model. The dependent variables are expressed in natural logarithms. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE IV

MAIN SPECIFICATIONS EXPLAINING BUSINESS FORMATION: MSA INEQUALITY 2004 - INSTRUMENTED

Model	First Stage	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	*MSA Inequality in 2004*	*MSA Total Establishment Entries*			*MSA Total Establishment Exits*		
MSA Inequality in 2004		-5.652***	-5.556***	-5.284**	-7.130***	-7.478***	-7.239***
		(0.827)	(1.980)	(2.189)	(1.127)	(2.296)	(2.438)
MSA Population$_{t-1}$	-0.014***	1.022***	0.965***	0.968***	0.984***	0.920***	0.920***
	(0.002)	(0.015)	(0.028)	(0.029)	(0.017)	(0.032)	(0.033)
MSA Catholic to Protestant Ratio	0.012*		0.058	0.053		0.076	0.073
	(0.006)		(0.061)	(0.063)		(0.064)	(0.066)
MSA Average Age of the Population	-0.184***		-0.417	-0.413		-0.223	-0.204
	(0.026)		(0.524)	(0.541)		(0.582)	(0.589)
MSA Personal Income Per Capita	-0.000		0.004**	0.004*		0.004*	0.004*
	(0.000)		(0.002)	(0.002)		(0.002)	(0.002)
MSA Land Area	0.010***		0.073***	0.069***		0.090***	0.088***
	(0.003)		(0.020)	(0.021)		(0.026)	(0.026)
MSA House Price Index	-0.000***		0.001***	0.002***		-0.000	0.000
	(0.000)		(0.000)	(0.001)		(0.000)	(0.001)
Rain	0.010*						
	(0.004)						
Rain Standard Deviation	0.002***						
	(0.001)						
Temperature	-0.017**						
	(0.008)						
Temperature Standard Deviation	0.003						
	(0.002)						
State Fixed Effects	Yes	Yes	Yes	--	Yes	Yes	--
Year Fixed Effects	Yes	Yes	Yes	--	Yes	Yes	--
State*Year Fixed Effects	No	No	No	Yes	No	No	Yes
Number of Observations	3.116	3.240	3.116	3.116	3.240	3.116	3.116

NOTES. All models are estimated with a 2SLS IV model. The dependent variables are expressed in natural logarithms. The first column contains the results of the first stage regression. MSA Inequality in 2004 is instrumented with average division rain fall and temperature between 1895 - 2003 and their corresponding standard deviations. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE V

MEDIATING FACTORS MAIN SPECIFICATIONS EXPLAINING COUNTY BANK FINANCING SUPPLY, COUNTY EDUCATION, EDUCATED COUNTY POPULATION INFLOW AND CIVIL JUSTICE EFFICIENCY

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*No. of Bank Establishments per 1,000 Capita*		*Percentage of Adults with College Degree or More*		*Population Inflow with at Least a College Degree*		*Judicial Efficiency*	
County Inequality in 2004	-0.695***	-4.513***	-0.873***	-1.957***	-5.041***	-10.465*	1.172**	-1,177
	(0.221)	(1.575)	(0.119)	(0.627)	(0.953)	(6.086)	(0.442)	(1.15)
County Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Case Controls	No	No	No	No	No	No	No	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Number of Observations	12.520	11.544	12.520	11.544	10.036	9.164	8.300	8.234

NOTES. Models (2), (4), (6) and (8) are estimated with a 2SLS IV model. In columns (5) and (6), the dependent variable is in log terms. All other Models are estimated using OLS. County Inequality in 2004 is instrumented with average division rain fall and temperature between 1895 - 2003 and their corresponding standard deviations. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE VI
MAIN SPECIFICATIONS EXPLAINING BUSINESS FORMATION

Model	(1)	(5)
Dependent Variable	*MSA Total Establishment Entries*	*MSA Total Establishment Exits*
MSA Inequality in 2004	-0.819	-0.919
	(0.783)	(0.711)
No. of Bank Establishments per 1,000 Capita	1,076.905***	1,028.422***
	(215.263)	(243.110)
Percentage of Adults with College Degree or More	0.053***	0.006
	(0.019)	(0.020)
Population Inflow with at Least a College Degree	0.003	0.006**
	(0.003)	(0.003)
MSA Population$_{t-1}$	0.953***	1.011***
	(0.034)	(0.040)
MSA Catholic to Protestant Ratio	0.031	0.022
	(0.036)	(0.028)
MSA Average Age	0.222	0.838**
	(0.321)	(0.398)
MSA Personal Income Per Capita	3.562*	3.902*
	(1.877)	(2.025)
MSA Land Area	0.022	0.016
	(0.027)	(0.027)
MSA House Price Index	0.001***	-0.000
	(0.000)	(0.000)
State Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
State*Year Fixed Effects	No	No
Number of Observations	2.760	2.760

NOTES. All Models are estimated with a linear regression (OLS) model. The dependent variables are expressed in natural logarithms. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE VII

MAIN SPECIFICATIONS COUNTY INEQUALITY 2004: EXPLAINING FIRM TECHNOLOGY, OWNERSHIP AND FINANCING

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent Variable	*Firm Is High Tech*			*Firm Is Proprietorship*			*Firm Bank and Family Financing*			*Firm Angel and Venture Capital Financing*		
County Inequality in 2004	-0.670***	-0.555***	-0.550***	1.159***	1.058***	1.060***	0.185	0.347*	0.359**	-0.113*	-0.189***	-0.193***
	(0.171)	(0.165)	(0.168)	(0.213)	(0.257)	(0.259)	(0.128)	(0.177)	(0.177)	(0.0635)	(0.0521)	(0.0557)
County Population	0.0208***	0.00836**	0.00828**	-0.0462***	-0.0451***	-0.0452***	-0.0137***	0.359**	-0.00961*	0.00326***	0.00311**	0.00349**
	(0.00322)	(0.00327)	(0.00335)	(0.00888)	(0.00749)	(0.00764)	(0.00326)	(0.177)	(0.00483)	(0.000972)	(0.00136)	(0.00167)
Firm Total Assets$_{t-1}$	--	-0.0137***	-0.0140***	--	-0.0523***	-0.0531***	--	0.0235***	0.0239***	--	0.00246	0.00266
	--	(0.00258)	(0.00257)	--	(0.00544)	(0.00550)	--	(0.00264)	(0.00265)	--	(0.00155)	(0.00160)
Firm ROA$_{t-1}$	--	0.00538**	0.00521**	--	0.00906***	0.00924***	--	-0.00295	-0.00292	--	-0.00324***	-0.00330***
	--	(0.00242)	(0.00249)	--	(0.00256)	(0.00254)	--	(0.00205)	(0.00217)	--	(0.00112)	(0.00118)
Firm Tangibility$_{t-1}$	--	-0.182***	-0.184***	--	0.195***	0.197***	--	0.0548***	0.0548***	--	-0.00477	-0.00497
	--	(0.0111)	(0.0111)	--	(0.0224)	(0.0223)	--	(0.0110)	(0.00973)	--	(0.00439)	(0.00445)
Firm Number of Owners$_{t-1}$	--	0.0248	0.0252	--	-0.367***	-0.366***	--	-0.00722	-0.00584	--	0.0412***	0.0421***
	--	(0.0192)	(0.0195)	--	(0.0422)	(0.0436)	--	(0.0133)	(0.0130)	--	(0.00948)	(0.00970)
Main Owner Is Female	--	-0.0487***	-0.0491***	--	0.0517**	0.0502**	--	0.00412	0.00386	--	-0.00987***	-0.00981***
	--	(0.0146)	(0.0147)	--	(0.0233)	(0.0238)	--	(0.0116)	(0.0119)	--	(0.00287)	(0.00308)
Main Owner Is African-American	--	0.0361	0.0364	--	0.00421	-0.000277	--	-0.0331*	-0.0326*	--	-0.00337	-0.00324
	--	(0.0270)	(0.0271)	--	(0.0420)	(0.0429)	--	(0.0184)	(0.0187)	--	(0.00284)	(0.00336)
Main Owner Is Hispanic	--	-0.0123	-0.0121	--	-0.00413	-0.00679	--	-0.00478	-0.00606	--	0.00520	0.00587
	--	(0.0263)	(0.0265)	--	(0.0769)	(0.0793)	--	(0.0134)	(0.0136)	--	(0.00723)	(0.00847)
Main Owner Is Asian	--	-0.00323	-0.00184	--	-0.0108	-0.0105	--	-0.00279	-0.00503	--	0.00489	0.00576
	--	(0.042)	(0.0422)	--	(0.0443)	(0.0456)	--	(0.0248)	(0.0242)	--	(0.0118)	(0.0119)
Main Owner Is Born in the US	--	-0.0951***	-0.0945***	--	0.0775**	0.0778**	--	-0.00165	-0.00177	--	-0.00112	-0.00124
	--	(0.0278)	(0.0280)	--	(0.0338)	(0.0343)	--	(0.0158)	(0.0163)	--	(0.00465)	(0.00458)
Main Owner's Work Experience	--	0.00388***	0.00388***	--	0.000401	0.000449	--	-0.000480	-0.000511	--	0.0000576	0.0000681
	--	(0.000404)	(0.000407)	--	(0.000718)	(0.000734)	--	(0.000433)	(0.000442)	--	(0.000150)	(0.000146)
State GDP$_{t-1}$	--	0.0820	--	--	0.0481	--	--	-0.0248	--	--	-0.0435	--
	--	(0.132)	--	--	(0.139)	--	--	(0.128)	--	--	(0.0485)	--
County Control Variables	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
State Fixed Effects	Yes	Yes	--	Yes	Yes	--	Yes	Yes	--	Yes	Yes	--
Year Fixed Effects	Yes	Yes	--	Yes	Yes	--	Yes	Yes	--	Yes	Yes	--
2-digit Industry Fixed Effects	No	No	No	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
State*Year Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Industry*Year Fixed Effects	No	No	No	No	No	Yes	No	No	Yes	No	No	Yes
Number of Observations Panel A	13.875	8.830	8,837	13.865	8.826	8,833	10.404	6.456	6,460	7.122	4.505	4,509

NOTES. All Models are estimated with a linear regression model (OLS) and take into account cross-sectional Kauffman Firm Survey weights. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE VIII

ROBUSTNESS ANALYSIS - INEQUALITY 1890: EXPLAINING BUSINESS FORMATION, FIRM TECHNOLOGY, OWNERSHIP AND FINANCING

Model	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	*MSA Total Establishment Entries*	*MSA Total Establishment Exits*	*Firm Is High Tech*	*Firm Is Proprietorship*	*Firm Bank and Family Financing*	*Firm Angel and Venture Capital Financing*
MSA Inequality in 1890	-0.214*	-0,162	--	--	--	--
	(0.122)	(0.114)	--	--	--	--
County Inequality in 1890	--	--	-0.182**	0.243***	0.120***	0.00415
	--	--	(0.0710)	(0.0773)	(0.0443)	(0.0200)
County Control Variables	No	No	Yes	Yes	Yes	Yes
MSA Control Variables	Yes	Yes	No	No	No	No
State Fixed Effects	Yes	Yes	--	--	--	--
Year Fixed Effects	Yes	Yes	--	--	--	--
Number of Observations	10.189	10.189	8,777	8,773	6,420	4,487

NOTES. All Models are estimated with a linear regression model (OLS). The dependent variables are expressed in natural logarithms. Models (3) - (7) take into account cross-sectional Kauffman Firm Survey weights. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE IX

FALSIFICATION TEST INSTRUMENTAL VARIABLES: US VERSUS FRANCE

Model	(1)	(2)	(3)	(4)	(5)	(6)
	US Full Sample		US Restricted Sample		France Full Sample	
Dependent Variable	*MSA Total Establishment Entries*	*MSA Total Establishment Exits*	*MSA Total Establishment Entries*	*MSA Total Establishment Exits*	*Dept. Total Establishment Entries*	*Dept. Total Establishment Exits*
Rain	-0.075**	-0.054	-1.477***	-0.364	-0.002	0.001
	(0.036)	(0.033)	(0.186)	(0.415)	(0.003)	(0.010)
Rain Standard Deviation	0.117*	0.096**	-0.102***	-0.050**	0.003	0.014
	(0.058)	(0.044)	(0.002)	(0.012)	(0.002)	(0.016)
Temperature	-0.017**	-0.019***	0.127**	-0.054	0.027	0.006
	(0.007)	(0.007)	(0.034)	(0.064)	(0.024)	(0.054)
Temperature Standard Deviation	-0.034**	-0.022	-0.039**	-0.081**	-0.009	0.355
	(0.015)	(0.016)	(0.014)	(0.020)	(0.079)	(0.280)
MSA/ Department Population$_{t-1}$	1.049***	1.051***	-0.087**	-0.016*	1.127***	0.890***
	(0.024)	(0.023)	(0.025)	(0.007)	(0.032)	(0.082)
MSA/ Department Land Area	0.015	-0.001	0.181***	0.133**	-0.212***	-0.199***
	(0.025)	(0.023)	(0.020)	(0.038)	(0.035)	(0.058)
State/Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	12.600	12.600	595	595	747	747

NOTES. All models are estimated with a linear regression model (OLS). The dependent variables are expressed in natural logarithms. Column (1) and (2) contain the results of the regressions including all observations from the US sample. Columns (3) and (4) provide the results when restricting the US sample to all those observations within the minimum and the maximum of the French averages and standard deviations. Column (5) and (6) present the results for the entire French data sample. The unit of analysis is the MSA for the US and the department for France. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. Standard errors are clustered at the state/ regional level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE X

DIFFERENCE-IN-DIFFERENCES ANALYSIS: STATE ELIMINATIONS OF ESTATE, INHERITANCE AND GIFT TAXES

Model	(1)	(2)	(4)	(5)
Dependent Variable	*MSA Total Establishment Entries*		*MSA Total Establishment Exits*	
Post	-0.026**		0.008	
	(0.013)		(0.017)	
Log Number of Years since Reform		-0.024*		-0.022**
		(0.014)		(0.009)
MSA Population$_{t-1}$	0.730***	0.737***	1.022***	1.024***
	(0.103)	(0.105)	(0.060)	(0.062)
MSA Catholic to Protestant Ratio	-0.033	-0.026	0.012	0.016
	(0.060)	(0.058)	(0.059)	(0.061)
MSA Whites to Total Population Ratio	2.005***	1.997***	-0.218	-0.198
	(0.685)	(0.654)	(0.719)	(0.703)
MSA House Price Index	1.338**	1.265**	0.663	0.593
	(0.602)	(0.550)	(0.441)	(0.418)
MSA Personal Income Per Capita	0.004***	0.004***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.000)
MSA Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Number of Observations	6.047	6.047	6.047	6.047

NOTES. All models are estimated with a linear regression model (OLS). The dependent variables are expressed in natural logarithms. Post is an indicator variable that takes on the value of 1in the years after the state in which an MSA is located abandons the EIG taxes and 0 otherwise. Number of Years since Reform indicates the total number of years that have passed after the EIG Reform was introduced in the state in which the MSA is located, with a maximum of 4 years. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

APPENDIX

A. Variables

We study the impact of county and MSA inequality on, in total, six corporate outcome variables. Table A.I collects all definitions of all these dependent variables, and also of all controls, and indicates the relevant data sources (Table I in the paper provides summary statistics).

[Table A.I around here]

For the purpose of summarizing we categorize the dependent variables into business formation, firm type, firm ownership and firm financing, though we recognize this categorization is not entirely descriptive. In terms of business formation we capture *MSA Total Establishment Entries*, the yearly total number of new establishments in an MSA and *MSA Total Establishments Exits*, which is the yearly total number of establishments that became inactive in an MSA (the estimates for these specifications will be reported in Tables IV and VI respectively). On average around 459 new establishments are set up each year, whereas a slightly higher number of establishments exits the market yearly, 539 on average.

As the sole firm type variable we feature: *Firm is High Tech* which equals one if the firm operates in a high technology industry, and equals zero otherwise. We classify firms as being high technology intensive (*High Tech*) based upon the High Technology Industries NAICS list from the Science and Engineering Indicators 2006 from the NSF. Based upon this classification 31 percent of the firms are considered as being high tech at inception.

Concerning firm ownership, we feature the following variable (the estimates for this specification will be reported in Table V and VII): *Firm Is Proprietorship* equals one if the firm is a proprietorship, and equals zero otherwise;[33] From the 3,419 firms in our dataset 1,294 firms (38 percent) are proprietorships at the start of the sample period.

As firm financing variables we feature (the estimates will be in Table V and VII as well): *Firm Family and Bank Financing* is the total amount of business and owners' personal bank financing plus family financing divided by total firm financing; *Firm Angel and Venture Capital Financing* is the amount of equity obtained from angels and venture capitalists divided by total firm financing.[34]

---

[33] We assign the value of one to the variable *Firm is Proprietorship* if a firm is either a sole proprietorship or a limited partnership. We assign all other firms the value of zero, i.e., these are firms that are classified as a limited liability firm, subchapter S-Corporations, C-Corporations, general partnerships or any other legal form. Alternatively defining a general partnership as a form of proprietorship does not change the results.

[34] Family Financing is the amount of equity invested by parents and/or spouse. In the main analysis, we add bank and family financing in the same dependent variable. In additional tests, we consider the two sources of financing as separate entities. It is evident from the descriptive statistics that a small percentage of firms rely on venture capital financing as well as bank financing. This generates many "zeros" in some of our dependent variable. To verify the robustness of our results in respect to this issue, we run two additional tests. First, we restrict our sample only to companies that rely on some form of external finance; this almost mechanically reduces the number of zeros by dropping from the sample firms that only rely on internal equity. Second, we keep the whole sample, but we transform the variables "bank and family financing" and "angels and venture capital financing" in (0/1) binary variables. We then run a rare event logit model (as in King, 2001) and see if the explanatory power of our inequality measures still holds. In both robustness checks, inequality maintains the statistically and economic significance we observe in our main results.

Firms in our sample tend to rely little on equity obtained from angels and venture capitalists; only 4 percent of the firms make use of this type of financing at start-up. On average only 2 percent of total funding is obtained through these sources. The mean amount of equity from angels and venture capitalists at inception equals around $37,500. Contrary to equity obtained from angels and venture capitalists, start-ups tend to rely more on debt financing in the form of bank and family financing (on average 11 percent of total financing comes from both business and owner's personal bank and family loans).[35]

We have three sets of control variables, i.e., firm characteristics, main owner characteristics, and state and county characteristics. We discuss each of these sets of control variables now in turn. For the analysis at the MSA level (for business formation dynamics) in tables IV, VI and VIII we include a set of MSA characteristics similar to the county characteristics included in the remaining analyses.[36] For the sake of brevity we do not discuss them.

Total Assets is the logarithm of one plus total assets, which is the sum of cash, accounts receivable, product inventory, equipment or machinery, land and buildings, vehicles, other business owned property and other assets; *ROA* is the Return on Assets, i.e., the amount of net profit divided by total assets, which we winsorize at the 1 percent level; *Tangibility* is the amount of property, plant and equipment divided by total assets; and, *Number of Owners* is the logarithm of one plus the total number of owners.

In their first year of operations the start-ups in our dataset have total assets worth of, on average, $172,709. Some firms have a negative return on assets. For example, in the first year of operations 54 percent of all start-ups have a *ROA* below zero percent. Tangible assets make up around 56 percent of total assets on average. However, 26 percent of the businesses report no tangible assets at all in the first year of business. The majority of firms (61 percent) have one owner, whereas the remaining 39 percent of businesses is owned by multiple owners.

The main owner characteristics comprise five dummies that equal one if the condition embedded in its label is fulfilled, and equals zero otherwise. These main owner dummies are: *Is Female*, *Is African-American*, *Is Hispanic*, *Is Asian* and *Is Born in the US*. *Work Experience* is the number of years of work experience of the main owner in the firm's industry.

We identify the main owner in the same way as Robb (2014) who consider the owner with the largest amount of equity invested to be the primary owner.[37] Overall, entrepreneurs of the nascent businesses in our sample are mostly white (only 15 percent is either African-American, Hispanic or Asian), male and born in the US. This is consistent with the owner characteristics of firm owners from the Survey of Consumer Finances (SCF) (see Puri (2007)). Less than a third (27 percent) of the main owners is female, and only 9 percent is born somewhere outside of the United States. In addition, primary owners tend to have quite some work experience in the same industry as their new business is operating in, the average years of experience is a little less than 14 years (median of 11 years).

As State characteristics we include its *GDP* which is the logarithm of one plus the gross domestic product of the State during the year.

As county characteristics we have: *Population* which is the total county population at year-end (one year lag); the *Catholic to Protestant Ratio* which is the (natural logarithm of the) ratio

---

[35] The largest part, i.e., 7 percent, comes from personal loans obtained by the owner(s). The average amount of bank financing that firms rely upon in our sample is $28,277 at inception.

[36] More specifically, the MSA characteristics included are: *Population, Wage Inequality, Catholic to Protestant ratio, Personal Income per Capita, Federal Government Expenditures per Capita, Land Area and the number of Whites to Total Population.*

[37] See their paper for the exact methodology on how to define the primary owner in case multiple owners invest an equal amount of equity into the firm.

of the total number of Catholics divided by the total number of Evangelicals in the county at year-end 2000; the *Personal Income Per Capita* which is the logarithm of one plus the per capita county personal income at year-end; the *Nonfarm Establishments Per Capita* which is the total number of nonfarm establishments divided by the total population in the county at year-end; *Wage Inequality* which is the Gini coefficient of wages earned in each US-county coming from the IRS-SOI data; the *Federal Government Expenditures Per Capita* which is the total Federal government expenditures in thousands of US Dollars during the year in the county divided by the total population in the county; and the *Land Area* which is the logarithm of one plus the total county area in square miles at year-end 2000.[38]

We include *GDP* to control for the state of the local economy and in the same line for per capita county personal income; their means are respectively 10.65 and 10.48. With respect to county demographics, between counties there is considerable heterogeneity when it comes to religion. County federal government expenditures per capita differ quite a lot between counties as well; the 10th percentile (of its logarithm) is 4 whereas the 90th percentile is almost three times higher. Additionally, counties' Wage Inequality, as measured by the Gini coefficient based on incomes, is quite high, with a mean value of 0.55, but differs less substantially between counties: its standard deviation is 0.04.

B. Measuring Wealth Inequality

It is very difficult to readily obtain representative measures of wealth inequality at the local level. As a result we construct our own two proxies for local wealth inequality. The first one is based on current levels of financial wealth and broadly based on a methodology introduced by Mian (2013) and Saez (2014), and it intends to construct local level measures of household net worth; the second measure is based on historical records of land ownerships.

The contemporary measure of wealth inequality looks at the amounts of dividends and interests earned by US households in 2004, the first year in our sample period, as reported by Internal Revenue Service (IRS) Statistics of Income (SOI) data. The IRS-SOI data report the total amount of dividends and interest income received by US households in a certain zip code. The information is reported as a total per zip code, but also divided in five households' income groups, ranging from low income to high income. Under the assumption that a typical household owns the market index for stocks and bonds, the amount of financial rents it receives depends only on the quantity of stocks and bonds it holds. IRS-SOI provides three pieces of information important to construct our proxy:

a) The total number of households belonging to each income group;

b) For each income group, the number of households who declared non-zero dividend and non-zero interest income (we will call these non-zero households); and,

c) For each income group, the total amount of dividends and interests earned by all households.

We now report the procedure we adopted to construct our inequality proxy. For simplicity, we just describe the case where we consider only dividends as a financial rent. The procedure is exactly the same when we also include interest income and comprises six steps.

[38] As we control both for population and land, we implicitly control for population density. Controlling directly for population density does not change our results.

(1) We aggregate the IRS-SOI figures at a county level.[39]

(2) For each county, we compute the number of households who declared zero dividend income and we place them into a separate category.

(3) For each county and each income group, we compute the average dividend earned by non-zero households. We do this by dividing the total amount of dividends for each income group by the respective number of non-zero households.

(4) We assume that each household in the same income group earned the average dividend computed in (2).

(5) We assume that each household owns the same type and composition of stock: the equity index. As a result, the amount of dividend received depends only on the quantity of stock owned.

(6) We use the number of non-zero households belonging to each income group, the number of households declaring zero dividends, and the average interests and dividend earned to compute a Gini coefficient that measures the distribution of dividend earnings within each county. Recall that the Gini coefficient is a measure of inequality that ranges between 0 and 1, where a coefficient close to 0 can be interpreted as full equality, whereas a coefficient of 1 indicates perfect inequality. We perform the same procedure with the interest income data.[40]

Table A.II provides an example of this computation. Column 1 lists the five income groups, Column 2 provides the number of households belonging to each income group, Column 3 the number of households declaring a non-zero dividend, and Column 4 the sum of all dividends received by all households in each group.

First, we compute the number of non-dividend earners by taking the difference between the total of Columns 1 and 2, which we report Column 4 in the row indicated as "Households with no Dividend Income". In the remainder of Column 4, we place the number of households that declared dividends, the same as in Column 2. We then compute the average dividend earned by non-zero households by dividing Column 3 by Column 4; we report this ratio in Column 5. We then compute the Gini coefficient, using the six dividend income groups. The first one consists of the 1,576,927 households that earned zero dividends, the second contains the 31,604 households that earn 1,181 dollars and up to the sixth group composed by 73,620 households that earn about 11,800 dollars in dividends. In this example, the Gini coefficient is equal to 0.91.

[Table A.II around here]

Naturally, this is a proxy, and it may be subject to measurement error. It performs well in identifying perfect equality and perfect inequality. In the former case, we would observe each household earning the same financial rents independently of the income group it belongs to, and our Gini coefficient would correctly have the value of zero. In the latter situation, our data would reveal all households but one receiving a financial rent and the Gini coefficient would correctly receive the score of one. The proxy does not work very well in every situation where in each income group, the distribution of dividends is very dispersed around the mean. In all these situations, we underestimate the degree of inequality. Measurement error may produce biased estimates of the coefficients when relating wealth inequality to financial outcomes. We

[39] For all analyses at the MSA level we execute this and other instructions below (that mention execution at the county level) at the MSA level.

[40] As we do not know the amount of dividends and interest income each individual household declares we cannot compute a unique Gini coefficient based on the sum of these amounts.

will be able to alleviate this problem by instrumenting this wealth inequality measure in various specifications.[41] We will present our main results using a Gini coefficient based on dividends. Results are basically the same if we use a Gini coefficient based on interest income.

To construct our historical measure of wealth inequality we obtain information on historical farm land sizes at the county level from the 1890 US Census. More precisely, for each county we have information on the total number of farms that – based upon their total acres of farm land – fall in a certain size bin. Farms are assigned to one of the following seven bins: under 10 acres, from 10 to 19 acres, 20 to 49 acres, 50 to 99 acres, 100 to 499 acres, 500 to 999 acres, and 1,000 or more acres.

First, we assume that the lower bound farm size of each bin is the average farm size of all the farms in this bin (for the first bin we set the lower bound equal to 0.001). Next, we use these lower bounds to calculate a county Gini coefficient in a similar way as in Rajan (2011).

Notice that we are unable to calculate a Gini coefficient for those counties that became incorporated after 1890, as the information on 1890 farm size distribution is unavailable. For these counties we manually look up the 1890 counties which these missing counties were then still part of before their own incorporation and take (simple) averages of the corresponding Gini coefficients. As the entire State of Oklahoma was incorporated well after 1890 (in 1907) we only have information on those eight counties that existed when it was still a territory. Based upon the information from these counties we construct a State Gini coefficient which we use for all counties in Oklahoma. To calculate this State Gini coefficient we sum the number of farms in each bin across the counties.[42]

In our dataset the average Land Gini coefficient is 0.44 and its standard deviation is 0.14. This is slightly lower compared to other contemporary measures of household wealth inequality at the aggregated level (contemporary measures of household wealth inequality at the county level do not exist). For example, De Nardi (2004) shows that the Gini coefficient for the entire US is 0.78 based upon household wealth data from the Survey of Consumer Finances from 1989. Relying on the same survey, Wolff (2010) finds that the Gini coefficient is 0.83 in 2007.

We also find that 1890 land inequality display a 36 and 46 percent positive correlation with our measures of dividend and interest inequality. The historical measure is also correlated with other contemporary socio economic measures that may reflect the degree of wealth inequality. It displays a positive correlation with local poverty rates (43 percent) and the number of crimes per capita (33 percent) and it is negatively correlated with the number of white people (a rough proxy of the size of the middle class) living in a county (-53 percent).

C. IV and Capital Structure

We turn to the IV results for the dependent variables capturing firm type, ownership and financing in Table A.IV. The *'First Stage'* Column in Table A.IV again shows the results of the first stage regression, indicating that also for US counties in addition to MSA's both rain and temperature are suitable instruments for local wealth inequality. The following columns in Table A.IV report the second stage regressions for the salient dependent variables. The results in Columns (1) and (2) indicate that county inequality indeed lowers the probability that a start-up firm is of High Tech nature. In Column (1) the effect is not only statistically significant but also very much economically meaningful: a one standard deviation increase in county

[41] Another possible source of measurement error may come from tax evasion. US financial institutions automatically report to the IRS dividends and interest income earned by their clients, making tax evasion through US banks virtually impossible. But taxpayers can avoid taxes by holding wealth at foreign banks.

[42] Results are unaffected if we exclude Oklahoma from the analysis.

inequality lowers the probability that a firm is of high tech nature by 21 percent. When including sets of firm and county characteristics and industry*year and state*year fixed effects in Column (2) however the coefficient is not statistically significant at conventional levels, although the sign remains negative as expected. In the same line the results mostly confirm our previous findings concerning firms' financing: in more unequal counties start-up firms rely more on bank and family financing.

[Tables A.IV around here]

Moreover, start-up firms are more likely to be of a simpler business form, i.e., a proprietorship, and, although the coefficients are not statistically significant at conventional levels when adding sets of county and firm characteristics, they also rely less on angel and venture capital financing and are less likely to be of a more complex high-tech nature. Again, the results are not only statistically significant but also very economically relevant: a one standard deviation increase in county inequality increases the probability that the start-up firm is of a simpler form (proprietorship) by more than half (53 percent increase of its mean) and increases the reliance on family and bank financing by 24 percent (again evaluated at its mean).

References

De Nardi, M., "Wealth Inequality and Intergenerational Links," *Review of Economic Studies*, 71 (2004), 743-768.

King, G., and L. Zeng, "Logistic Regression in Rare Events Data," *Political Analysis*, 9 (2001), 137-163.

Mian, A., K. Rao, and A. Sufi, "Household Balance Sheets, Consumption, and the Economic Slump," *The Quarterly Journal of Economics*, 128 (2013), 1687-1726.

Puri, M., and D. T. Robinson, "Optimism and Economic Choice," *Journal of Financial Economics*, 86 (2007), 71-99.

Rajan, R., and R. Ramcharan, "Land and Credit: A Study of the Political Economy of Banking in the United States in the Early 20th Century," *Journal of Finance*, 66 (2011), 1895-1930.

Robb, A. M., and D. T. Robinson, "The Capital Structure Decisions of New Firms," *Review of Financial Studies*, 27 (2014), 153-179.

Saez, E., and G. Zucman, "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," (Cambridge MA: National Bureau of Economic Research, 2014).

Wolff, E. N., "Recent Trends in Household Wealth in the United States: Rising Debt and the Middle-Class Squeeze - An Update to 2007," (Annandale-On-Hudson NY: Levy Economics Institute, 2010).

APPENDIX TABLE A.I

VARIABLE NAMES, DEFINITIONS, AND DATA SOURCES FOR THE EMPIRICAL ANALYSIS OF FIRM OWNERSHIP, FINANCING AND TYPE

Variable Name	Variable Definition	*Source*
Dependent Variables		
MSA Total Establishment Entries	The logarithm of the yearly total number of new establishments in the MSA between 2005-2012	*USC*
MSA Total Establishment Exits	The logarithm of the yearly total number of establishments that became inactive in the MSA between 2005-2012	*USC*
Firm is High Tech	= 1 if firm operates in a high technology industry, = 0 otherwise	*NSF*
Firm Is Proprietorship	= 1 if firm is a proprietorship, = 0 otherwise	*KFS*
Firm Bank and Family Financing	The amount of business and owners' personal bank financing and the amount of equity invested by parents and/or spouse divided by total firm financing	*KFS*
Firm Angel and Venture Capital Financing	The amount of equity obtained from angels and venture capitalists divided by total firm financing	*KFS*
Main Independent Variables		
County Inequality in 2004	The Gini coefficient of the distribution of wealth as measured by the distribution of the amount of declared dividends from household tax filings in the county	*IRS*
MSA Inequality in 2004	The Gini coefficient of the distribution of wealth as measured by the distribution of the amount of declared dividends from household tax filings in the metropolitan statistical area (MSA)	*IRS*
County Inequality in 1890	The Gini coefficient of the distribution of farm land in 1890 in the county (for counties in Oklahoma the state-level coefficient is used)	*USC*
MSA Inequality in 1890	The Gini coefficient of the distribution of farm land in 1890 in the metropolitan statistical area (MSA)	*USC*
Instrumental Variables		
Rain	The average district precipitation between 1895-2003, where a district is defined as a group of clustered counties with similar climatic conditions	*NCDC*
Temperature	The average district temperature in degrees between 1895-2003, where a district is defined as a group of clustered counties with similar climatic conditions	*NCDC*
Control Variables		
Firm Characteristics		
Firm Total Assets$_{t-1}$	The logarithm of one plus total assets, which is the sum of cash, accounts receivable, product inventory, equipment or machinery, land and buildings, vehicles, other business owned property and other assets	*KFS*
Firm ROA$_{t-1}$	Return on Assets, i.e., the amount of net profit divided by total assets winsorized at the 1% level	*KFS*
Firm Tangibility$_{t-1}$	The amount of property, plant and equipment divided by total assets	*KFS*
Firm Number of Owners$_{t-1}$	The logarithm of one plus the total number of owners	*KFS*
Main Owner Characteristics		
Main Owner Is Female	= 1 if main owner is a female, = 0 otherwise	*KFS*
Main Owner Is African-American	= 1 if main owner is African-American, = 0 otherwise	*KFS*
Main Owner Is Hispanic	= 1 if main owner is Hispanic, = 0 otherwise	*KFS*
Main Owner Is Asian	= 1 if main owner is Asian, = 0 otherwise	*KFS*
Main Owner Is Born in the US	= 1 if main owner was born in the US, = 0 otherwise	*KFS*
Main Owner's Work Experience	Number of years of work experience of the main owner in the firm's industry	*KFS*
State and County Characteristics		
State GDP$_{t-1}$	The logarithm of one plus the gross domestic product of the state during the year	*USC*
County Population	Total county population at year-end	*USC*
County Catholic to Protestant Ratio	Ratio of the total number of Catholics divided by the total number of Evangelicals in the county at year-end 2000	*ARDA*
County Personal Income Per Capita	County total personal Income at the end of the year divided by population (in 1,000 of habitants)	*BEA*
County Wage Inequality	The Gini coefficient of the distribution of wages as measured by the distribution of the amount of labor from household tax filings in the county	*IRS*
County Ethnic Diversity	One minus a Herfindhal index of ethic diversity based on the relative proportion of the ethnic groups living in the county.	*USC*
County Land Area	The logarithm of one plus the total county area in square miles at year-end 2000	*USC*
County Bank Population per Capita	Total number of banking establishment in a county divided by total population in the same county	*USC*
County Percentage of Adult Population with a College Degree	Total number of individuals with a college degree or more in a county divided by total population in the same county	*USC*
County Population Inflow with at least a College Degree	Total number of individuals with a college degree or more moving in a county	*USC*

NOTES. The table defines the variables used in the empirical analysis of firm ownership, financing and type, as well as the corresponding data sources used. Total firm financing is the sum of total debt and equity financing. *t-1* indicates a one year lag is used in the empirical analysis. For the sake of brevity we do not report the MSA characteristics separately. Data sources include: *ARDA* = Association of Religion Data Archives; *BEA* = Bureau of Economic Analysis; *IRS* = Internal Revenue Service; *KFS* = Kauffman Firm Survey; *NCDC* = National Climatic Data Center; *NSF* = National Science Foundation; *USC* = US Census.

APPENDIX TABLE A.II
EXAMPLE COUNTY INEQUALITY 2004 CONSTRUCTION

Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Income Group Category by Size of Adjusted Gross Income and Zip-Code	Total No. of Returns	Taxable Dividends: No. Of returns	Taxable Dividends: Total Amount Reported	No. of Returns	Average Dividend per Household	Taxable Interest: No. of Returns	Taxable Interest: Total Amount Reported
Total	1.882.964	306.037	1.287.291			651.013	1.189.469
Households with No Dividend Income			0	1.576.927	0,00		
Under $10,000	387.555	31.604	37.351	31.604	1,18	67.710	70.567
$10,000 under $25,000	553.957	42.503	64.756	42.503	1,52	114.076	150.519
$25,000 under $50,000	454.236	60.982	104.993	60.982	1,72	152.215	184.266
$50,000 under $75,000	231.139	55.051	113.500	55.051	2,06	123.892	156.056
$75,000 under 100,000	124.646	42.277	98.721	42.277	2,34	84.098	116.665
$100,000 or more	131.431	73.620	867.970	73.620	11,79	109.022	511.396

NOTES. The table provides an example of the data used to construct our *County Inequality* measure from 2004. We obtain data from the SOI (Statement of Income) database from the IRS on the total number of tax returns in thousands (one per household) filed in 2004 classified by zipcode and the adjusted gross income as shown in Column (1). In addition we obtain information on the number of returns that declared to have obtained a dividend and the accompanying total dividend amounts reported (reported in thousands and thousands of US $), again classified by zipcode and the adjusted gross income of the household (shown in Columns (2) and (3) respectively). Based upon this data we calculate the average dividend amount per household reported for each income group in Column (5). The average dividend amount is reported in thousands of US $. We create an extra category of the number of households that did not declare any dividend (which is the total reports filed minus all reports that declared a dividend) which we report in the row 'Total', column (4) and (5) respectively. We use these average dividends as well as the income group classification to construct a Gini index in line with Rajan (2011). We create a second Gini coefficient in the same way, only now based upon the amount of interests received by households in 2004, as reported in Columns (6) and (7). Again, we obtain this information from the SOI database. The correlations between the Gini's based upon dividends and interest income received by households is very large and we therefore only report the results from our anlysis in which we introduce the county inequality measure based upon dividends received. To construct our *MSA Inequality* measure from 2004 we follow the same procedure but aggregate the zipcode data to the corresponding metropolitan statistical area (MSA) as opposed to the county.

APPENDIX TABLE A.III
SPECIFICATIONS EXPLAINING ESTABLISHMENTS CHURNING AND SCHUMPETERIAN EXITS

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*MSA Total Establishment Exits within 36 months*				*MSA Total Establishment Exits after 36 months*			
MSA Inequality in 2004	-3.674***	-2.269***	-1.651**	-2.048**	-4.134***	-2.796**	-2.070**	-2.316**
	(0.785)	(0.813)	(0.772)	(0.830)	(1.100)	(1.071)	(0.923)	(1.149)
MSA Population$_{t-1}$	1.036***	1.029***	1.037***	1.029***	0.957***	0.968***	0.948***	0.966***
	(0.019)	(0.024)	(0.024)	(0.026)	(0.025)	(0.039)	(0.039)	(0.042)
MSA Catholic to Protestant Ratio		0.038	-0.017	0.035		-0.075	-0.069	-0.083
		(0.049)	(0.039)	(0.053)		(0.073)	(0.078)	(0.079)
MSA Average Age		0.499**	-0.143	0.472**		1.150	0.903	1.161
		(0.225)	(0.277)	(0.234)		(0.819)	(0.952)	(0.861)
MSA Personal Income Per Capita		0.004**	0.004**	0.004*		0.009**	0.007**	0.008**
		(0.002)	(0.002)	(0.002)		(0.004)	(0.003)	(0.004)
MSA Land Area		0.016	0.010	0.015		0.008	0.005	0.008
		(0.029)	(0.029)	(0.032)		(0.035)	(0.033)	(0.037)
MSA House Price Index		0.002***	0.002***	0.003***		-0.003**	-0.003**	-0.001
		(0.000)	(0.000)	(0.000)		(0.001)	(0.001)	(0.001)
MSA Ethnic Diversity			-0.398**				0.294	
			(0.160)				(0.209)	
MSA Wage Inequality in 2004			2.451***				2.545*	
			(0.615)				(1.393)	
State Fixed Effects	Yes	Yes	Yes	--	Yes	Yes	Yes	--
Year Fixed Effects	Yes	Yes	Yes	--	Yes	Yes	Yes	--
State*Year Fixed Effects	No	No	No	Yes	No	No	No	Yes
Number of Observations	3.276	3.144	3.144	3.144	3.276	3.144	3.144	3.144

NOTES. All Models are estimated with a linear regression (OLS) model. The dependent variables are expressed in natural logarithms. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

APPENDIX TABLE A.IV

MAIN SPECIFICATIONS COUNTY INEQUALITY 2004 - INSTRUMENTED: FIRM TECHNOLOGY, OWNERSHIP AND FINANCING

Model	First Stage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*County Inequality in 2004*	*Firm Is High Tech*		*Firm Is Proprietorship*		*Firm Bank and Family Financing*		*Firm Angel and Venture Capital Financing*	
County Inequality in 2004	--	-1.520***	-0.285	0.724	4.175**	0.770	1.374*	-0.366***	-0.284
	--	(0.442)	(1.290)	(1.141)	(2.017)	(0.642)	(0.812)	(0.109)	(0.395)
County Population	0.0105***	0.0201***	0.000575	-0.0398***	-0.0730***	-0.0164***	-0.0195*	0.00296**	0.00239
	(0.000542)	(0.00343)	(0.0189)	(0.00729)	(0.0241)	(0.00291)	(0.0105)	(0.00119)	(0.00485)
Firm Total Assets$_{t-1}$	0.000113	--	-0.0122***	--	-0.0540***	--	0.0237***	--	0.00175
	(0.000243)	--	(0.00299)	--	(0.00621)	--	(0.00276)	--	(0.00153)
Firm ROA$_{t-1}$	0.000265	--	0.00487*	--	0.0105***	--	-0.00149	--	-0.00336***
	(0.000220)	--	(0.00264)	--	(0.00254)	--	(0.00210)	--	(0.00129)
Firm Tangibility$_{t-1}$	0.000851	--	-0.190***	--	0.208***	--	0.0533***	--	-0.00616
	(0.00129)	--	(0.0120)	--	(0.0296)	--	(0.00918)	--	(0.00526)
Firm Number of Owners$_{t-1}$	-0.000745	--	0.0238	--	-0.355***	--	0.00170	--	0.0496***
	(0.00113)	--	(0.0219)	--	(0.0476)	--	(0.0126)	--	(0.00897)
Main Owner Is Female	0.000698	--	-0.0518***	--	0.0407	--	0.0146	--	-0.0119***
	(0.00101)	--	(0.0142)	--	(0.0253)	--	(0.0138)	--	(0.00358)
Main Owner Is African-American	0.0109***	--	0.0391	--	-0.0138	--	-0.0443**	--	-0.00495
	(0.00202)	--	(0.0346)	--	(0.0424)	--	(0.0224)	--	(0.00367)
Main Owner Is Hispanic	0.00533***	--	-0.0168	--	-0.0175	--	-0.0146	--	0.00660
	(0.00206)	--	(0.0259)	--	(0.0728)	--	(0.0137)	--	(0.0106)
Main Owner Is Asian	-0.000239	--	0.00102	--	-0.00654	--	-0.0393*	--	0.00921
	(0.00229)	--	(0.0459)	--	(0.0525)	--	(0.0201)	--	(0.0130)
Main Owner Is Born in the US	-0.00554***	--	-0.0882***	--	0.0729*	--	-0.0126	--	-0.00453
	(0.00163)	--	(0.0283)	--	(0.0411)	--	(0.0199)	--	(0.00509)
Main Owner's Work Experience	-0.0000441	--	0.00386***	--	0.00171**	--	-0.000397	--	0.0000155
	(0.0000428)	--	(0.000387)	--	(0.000767)	--	(0.000527)	--	(0.000152)
State GDP$_{t-1}$	0.0100	--	--	--	--	--	--	--	--
	(0.0214)	--	--	--	--	--	--	--	--
Rain	0.00752***	--	--	--	--	--	--	--	--
	(0.00262)	--	--	--	--	--	--	--	--
Rain Standard Deviation	-0.0156***	--	--	--	--	--	--	--	--
	(0.00310)	--	--	--	--	--	--	--	--
Temperature	0.000982***	--	--	--	--	--	--	--	--
	(0.000265)	--	--	--	--	--	--	--	--
Temperature Standard Deviation	-0.00444***	--	--	--	--	--	--	--	--
	(0.000611)	--	--	--	--	--	--	--	--
County Control Variables	Yes	No	Yes	No	Yes	No	Yes	No	Yes
State Fixed Effects	Yes	Yes	--	Yes	--	Yes	--	Yes	--
Year Fixed Effects	Yes	Yes	--	Yes	--	Yes	--	Yes	--
2-digit Industry Fixed Effects	Yes	No	No	No	--	No	--	No	--
State*Year Fixed Effects	No	No	Yes	No	Yes	No	Yes	No	Yes
Industry*Year Fixed Effects	No	No	No	No	Yes	No	Yes	No	Yes
Number of Observations	5,442	11,696	7,467	11,687	7,463	8,796	5,469	5,983	3,787
R-Squared	0.675	0.015	0.107	0.094	0.291	0.026	0.125	0.017	0.185

NOTES. All models are estimated with a 2SLS IV model. The first column contains the results of the first stage regression. County Inequality in 2004 is instrumented with average division rain fall and temperature between 1895 - 2003 and their corresponding standard deviations. All models take into account cross-sectional Kauffman Firm Survey weights. The definition of the variables can be found in Appendix Table A.I. t-1 indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. "--" indicates that the indicated set of characteristics or fixed effects are comprised in the wider included set of fixed effects. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

Chapter 2: Inequality and Judges' sentencing

Fabio Braggion, Mintra Dwarkasing, Steven Ongena[43]

[43] Braggion is from EBC, CentER & Tilburg University. Dwarkasing is from EBC, CentER & Tilburg University. Ongena is from University of Zurich, SFI and CEPR.

Abstract

We empirically test how wealth inequality in US counties affects local judges' decision making. According to established theory (Perotti and von Thadden, 2006), inequality may affect firms' financing. One of the implications of this model is that in more unequal societies bank financing will be more prevalent. In the model voters may vote for the supply of those forms of financing that are best aligned with their own preferences. Hence, they may also vote for judges that sentence in line with these preferences. Based upon these notions we investigate whether higher inequality in US counties is associated with banks having a higher probability to win a second degree trial. Additionally, we investigate whether, in the spirit of Engerman and Sokoloff (2006), lobbying reinforces the effect local inequality can have on judges' decision making in favour of banks. Using a historical measure of inequality dating as far back as 1890 at the US MSA/county level we find evidence in favour of the latter: Our results suggest that in more unequal counties where it is easier for elites to lobby and where judges are less independent, the effect of local inequality is more strongly related to whether or not a bank wins a second degree trial.

I. Introduction

That inequality can have a potential effect on financial outcomes such as firms' financing and may hamper the development of basic institutions such as schooling and the judiciary is recently acknowledged in the literature (see for instance Engerman and Sokoloff, 1997 and Perotti and Von Thadden, 2006 and more recently Braggion et al., 2015). In this paper we empirically test how wealth inequality in US counties is associated with local judges' decision making. More specifically, we see whether voting and the general preferences of the public may have a role in shaping the local financial system. In the spirit of Perotti and von Thadden (2006), voting is the mechanism that allows citizens to express their own preferences. In the model voters may vote for the supply of those forms of financing that are best aligned with their own preferences. Hence, they may also vote for judges that sentence in line with these preferences. One of the implications of the model is that in more unequal societies bank financing will be more prevalent. Based upon these notions we investigate whether higher inequality in US counties is associated with banks having a higher probability to win a second degree trial. However, thinking about the relationship between the quality of institutions and inequality, election outcomes and resultant policymaking will also be influenced by campaign contributions and lobbying by or on behalf of the wealthy elite, an idea consistent with Engerman and Sokoloff`s work. Additionally, Calomiris (2014) also notes that wealthy elites may capture the banking sector for themselves, a notion that is empirically supported in Rajan (2011) who find that in those counties where the agricultural elite had a large proportion of land holdings access to credit was more limited and credit itself more costly, suggesting that a wealthy elite may limit financial development in order to restrict access to credit. On a related note Pagano and Volpin (2005) also indicate that voting systems and the political economy in general can result in specific regulatory outcomes. They provide a model that shows that voting can produce specific regulatory outcomes regarding shareholder and employment protection. More specifically, they show that systems characterized by proportional voting, i.e. a majority of the votes is necessary to win, produce weak shareholder protection and strong employment protection, whereas a majoritarian system, i.e. a majority of districts implies victory, produces an outcome in favor of rentiers: strong shareholder protection but weak employment protection. Indirectly this may also explain financial outcomes such as the extent of stock market development and the reliance on equity financing through the extent of shareholder protection.

In this chapter we take to the data the key ingredients of these frameworks, that is, agents can vote. We will provide some indirect evidence to see whether the local level of wealth inequality has an impact on voting outcomes and whether this impact is determined by pure citizens' preferences or by some campaign contribution or lobbying activity.

Our geographical unit of analysis is the US County. We look at a county-level institution that (at least in many US States) citizens are called to elect, i.e., judges of first degree courts. In particular, we examine whether, in more unequal counties, the decisions of judges in civil cases are more likely to favour banks over businesses (or broadly speaking equity). We will also provide some suggestive evidence that will discern whether voting results are mainly driven by people's preferences or are more related with lobbying activities.

Given that the database we use on civil trials in the US goes back to the 1980s we make use of a historical measure of inequality dating as far back as 1890 at the US MSA/county level based upon the distribution of landholdings (as in Braggion et al., 2015) and relate it to contemporary civil trial outcomes. This approach relieves any endogeneity issues in the form of reverse causality and, given that we are able to control for state fixed effects, county controls as well as a broad set of case fixed effects at the individual case level, we are able to go from correlations to causation.

We find evidence in favour of the notions set forward by Engerman and Sokoloff (1997): Our results suggest that in more unequal counties where it is easier for elites to lobby and where judges are less independent, the effect of local inequality is more strongly related to whether or not a bank wins a second degree trial.

II. Data and Hypotheses

We obtain data on civil trials from Westlaw US, a database that contains opinions and descriptions of US trials since 1948. A limitation of the Westlaw database is that it only contains information about second degree trials.[44] Luckily, from the discussion of the second degree hearings we can obtain vital information about the first degree trial which we use in our analysis. We collect data for a 30 year period, i.e., between 1984 and 2014.

To select the data we employed the following procedure. We restricted our search to civil trials. We searched for civil trials that involved either a bank or a business as one of the parties. We made a keyword search for party names specifying the search to be restricted to the words "bank" and "partnership", "business" or "corporation". The output we obtained still contained many cases where a bank was facing an individual in a business unrelated case. We resolved this problem by manually selecting the cases in which a bank was actually facing a business.

[44] This will generate selection bias in our analysis as it is very likely that second degree cases are not a random sample of the whole population of trials in a county. It is very difficult to give an account of the importance and the direction of the bias. To the extent that second degree cases are the most controversial or somehow they deal with "new situations", the bias may actually be beneficial for our analysis. These are the situations where the discretion of the judge is more important in deciding the case, which is exactly what we want to capture. To the best of our knowledge the Westlaw database is the only database with such detailed information on the parties involved that enables us to exactly identify those cases between banks and corporations. We resort to another database, the Civil Justice Survey from the Bureau of Justice Statistics, to see whether the results are similar when using first degree trial information. However, in this database it is not possible to exactly identify those cases between banks and corporations, but only with huge bias.

The Westlaw trial report contains the name of the county where the first degree trial took place, the name of the parties involved, whether the first degree trial was a summary judgment, and brief description of the case which includes the result of the first degree trial. Westlaw also reports the West Headnotes of each trial. West Headnotes are a standard categorization which divides the law into major categories (for example *deposit and escrows*, *contracts*, *compromise and settlement*). We use these basic categories to create case (categories) dummies.

Each West Headnotes major category is also divided in various subcategories. In principle, each trial may belong to various West Headnotes subcategories depending on its nature and its complexity. We counted the number of subcategories each trial belongs and consider this a proxy for the complexity of the case. In principle, cases that belong to more subcategories should be more complex.

From the Westlaw report we also use some information items related to the second degree trial. In particular, whether the second degree sentence involved dissenting judges or not: We consider this another proxy of the complexity of the case. We also check whether the first degree sentence was affirmed or not in second degree.

We read the description of the first degree trials and determined whether the bank won it or not. We define as a victory a situation where the bank obtained in full what requested to the court, if the bank was a plaintiff, or a situation where the request of the opposing party was denied, when the bank was a defendant.

The process of judicial selection and the length of their appointments differs considerably between States. A total of 20 States elect its own judges. Some States (in total 7) have implemented a Partisan election as the mode of selection, where the candidates are listed on the ballot along with a label designating the political party's ballot on which they are running; in other States judges are not affiliated with a political party on the ballot (for example the so-called Non-partisan and Missouri Plan selection methods). In principle, in States that select their judges via elections it is possible for voters to select a judge that best coincides with their preferences.

Partisan elections and length of the term in office have been usually considered as measures of judicial independence (Hanssen (2004a)). Judges are more likely to be independent from their own constituency the longer they stay in office, as they do not need to face an imminent re-appointment. In partisan elections the importance of campaigning and campaign contributions tends to be an important factor to increase judges' chances of being selected. A summary of the variables used in our analysis as well as the corresponding descriptive statistics can be found in Tables I and II respectively.

[Table I and II about here]

Our hypotheses are broadly based upon the work of Engerman and Sokoloff (1997), Calomiris (2014) and Perotti and von Thadden (2006). In principle, our empirical analysis would provide some support to Perotti and von Thadden (2006) if we find that in counties that are more unequal, judges are more likely to rule in favour of banks. Hence we would expect to find a positive association between inequality and the probability that a bank wins a second degree trial. This effect should be stronger in counties where judges are elected. The institutional story would predict a stronger relationship between wealth inequality and judicial decisions in counties where judges are less independent, in other words were judges stay in office for a shorter period of time and face partisan elections. Hence we would expect to find that the coefficient on the interaction term between Inequality and the length a judge can stay in office will be *negative* (as a longer term is usually seen as greater independence since a judge does not face immediate re-election). Additionally, this negative effect of the interaction term combined with a *positive* coefficient on the standalone Inequality variable may indicate that a wealthy elite can capture the banking system and may try to accomplish favourable outcomes for banks during trials through lobbying, a finding in the spirit of Calomiris (2014).

To test these hypotheses we create a dummy called *Elected* which we code to be equal to 1 if a firm is located in a State/County where judges are elected. We also code a partisan dummy *Partisan* which to be equal to 1 if a firm is located in a State where judges are elected based upon a Partisan election and 0 otherwise. In both cases, we interact the dummy variable with our inequality measure. We also control for the length of the term in office of judges and also interact it with the inequality measure.[45] Since, our data starts in 1984, we consider here a predetermined historical measure of inequality. Table III presents the results of this analysis.

III. Results

We start by analysing whether local inequality affects judicial decision making in US counties in table III. Each regression controls for state fixed effects, years fixed effects and case fixed effects using the West Headnotes. We consider particularly important this last set of controls, as it absorb every specificity regarding any particular topic that may affect judges' decisions. The regressions also control for a dummy that takes the value of 1 if the Bank was a plaintiff in the case. As the plaintiff usually choses the trial court, the dummy controls whether banks systematically chose "friendly" courts.[46]

---

[45] In Rhode Island, Vermont and Massachusetts judges have life time appointments. We set their values equal to infinity and employ a monotonic transformation of all values of the term in office (tio): $\frac{1-e^{-tio}}{1+e^{-tio}}$.

[46] In many of the trials we obtain from Westlaw it is not possible to see who were the original plaintiffs in the case, because Westlaw only reports who are the appellants and appellees. By considering in our analysis only those cases where the plaintiff is clearly identified we lose almost 600 observations. However our results remain otherwise unaffected if we consider the larger sample for which we cannot control in the specifications whether the bank was a plaintiff or not.

[Table III about here]

Column 1 presents the effect of local wealth inequality on the probability that a judge will rule in favour of a bank. The coefficient on County Inequality 1890 is negative and not statistically significant. Column 2 of Table III considers the interaction between Wealth inequality and whether the State elects its judges. Also in this case neither wealth inequality alone nor the interaction term is statistically significant. In Column 3, we interact wealth inequality with the partisan dummy: The coefficient of the interaction term is positive and statistically significant. In more unequal counties where judges are elected, banks are more likely to win first degree civil trials. The economic effect is also sizable: A standard deviation increase in wealth inequality is associated with an increase of about 15 percent of the probability that the bank will win the case. Columns 4 and 5 consider the same interaction term, but also consider whether those banks that are plaintiffs or banks headquartered in the States where the trial takes place are more likely to win in partisan States. These additional controls are never statistically significant and more importantly, the coefficient on the interaction between County Inequality and Partisan remains unaltered.

Columns 6 to 9 look at the interaction between County Inequality and Judges' Length of the term in office. The coefficient on County Inequality alone is now positive and statistically significant while the coefficient on the interaction term is negative and also statistically significant. On average, in unequal counties, judges are more likely to decide in favour of banks, but this effect fades away the longer is the term a judge officially may remain in office. Taken altogether, these results suggest that the degree of judicial independence matters in determining judges' decisions in counties that display larger levels of wealth inequality. Hence, they provide some support on institutional explanations for the relationships between inequality and financial decisions in the spirit of Engerman and Sokoloff (2002).[47] Additionally, they may point in the direction as argued by Calomiris (2014) that elites can capture important institutions such as the banking sector and lobby for judges' decisions in favour of banks.

Given that the Westlaw cases in our dataset are from 1984 onwards we used a historical measure of inequality to test the effect of local inequality on judicial outcomes. In a robustness analysis we rerun our regressions using a contemporary measure of local wealth inequality. We construct a gini coefficient at the county level based upon dividend incomes at the zip code level from 1998 obtained from the IRS SOI as in Braggion (2015). The results are shown in table IV.

[Table IV about here]

We restrict our analysis using only those trials from 1999 onwards. This leads to a huge decrease in our sample size to 467 cases. In Column 1 we relate our contemporary inequality measure to the probability that a bank wins a second degree trial. The coefficient on our inequality is negative and not statistically significant. In Column 2 and 3 we interact the inequality measure with a dummy variables indicating whether judges are elected and whether the county is in which the trial takes place is located in a state with partisan election respectively. Neither the standalone coefficients on inequality nor the coefficients on the interactions are statistically significant. The coefficient on the interaction between inequality and our Partisan indicator is positive however, similar to our findings when using an historical measure of inequality. Also in Columns 4 and 5 the interaction term between inequality and the Partisan

[47] We also run specifications that include state and case trends. The estimated coefficients on the interaction terms between wealth inequality and length of the judicial terms remained virtually unaffected (both in terms of statistical significance and in terms of economic relevance). The interaction terms between wealth inequality and the partisan elections dummy maintain their original signs, but lose statistical significance.

indicator is positive but not statistically significant. This loss in significance may be due to the large decrease in sample size when restricting our analysis to only those cases from 1999 onwards. When interacting inequality with the length a judge can stay in office as a measure of judge independence in Columns 6 to 9 we see that the results are similar to the findings when using an historical measure of inequality, although only in Column 7 statistically significant: The coefficient on inequality is positive and the interaction with the length of the term a judge can stay in office is negative, indicating that the effect of inequality is reinforced when judges stay on for a shorter time i.e. are less independent. In unreported regressions we rerun our analysis instrumenting our contemporary inequality measure with district level rainfall and temperature (as in Braggion 2015 and following Easterly 2007). The results are similar to the OLS findings, only lose some statistical significance.

Additionally, to corroborate our findings, we resort to another database with information on first degree civil trials (as opposed to appeal cases) from 2005 from the Bureau of Justice Statistics, the Civil Justice Survey. Unfortunately, in this database we are only able to exactly identify that a bank is involved in 37 out of the 5000 cases. We are, however, able to filter out those cases in which a bank most likely is involved and in which the parties involved are not other types of organizations such as government organizations and insurance companies. Still, this approach most likely results in a tremendous bias in assessing whether a bank is involved or not. We rerun our analysis based upon those cases for which we suspect a bank is involved. The results are similar to the results obtained using the second degree trials from Westlaw.

IV. Conclusions

We empirically test how wealth inequality in US counties affects local judges' decision making. Using a detailed database on second degree (appeal) cases from the US from 1984 onwards and restricting our analysis to those cases that involve a bank and a business, we are able to relate local inequality to the probability that a bank wins second degree trial. Relying on a historical measure of inequality in the spirit of Rajan (2007) and following Braggion (2015) we find evidence suggesting that the degree of judicial independence matters in determining judges' decisions in counties that display larger levels of wealth inequality: On average, in unequal counties, judges are more likely to decide in favor of banks, but this effect fades away the longer the term is a judge officially may remain in office (i.e. the more independent a judge is). Additionally, the effect of inequality on judges' decision making is reinforced in those states where lobbying is more easy. Our findings provide some support on institutional explanations for the relationships between inequality and financial decisions in the spirit of Engerman and Sokoloff (2002) and show that inequality is associated with judicial decision making.

References

Braggion, F., Dwarkasing, M., and Ongena, S., 2015, “Household Inequality, Entrepreneurial Dynamism and Corporate Financing”, working paper.

Calomiris, C. W., & Haber, S. H., 2014, “Fragile by Design: The Political Origins of Banking Crises and Scarce Credit: The Political Origins of Banking Crises and Scarce Credit”, Princeton University Press, Chicago.

Easterly, W., 2007, "Inequality Does Cause Underdevelopment: Insights from a New Instrument," *Journal of Development Economics* 84, 755-776.

Engerman, S. L., and K. L. Sokoloff, 1997, "Factor Endowments, Institutions, and Differential Paths of Growth Among New World Economies: A View from Economic Historians of the United States," in S. Haber, ed., *How Latin America Fell Behind*, Stanford University Press, Stanford.

—, 2002, "Factor Endowments, Inequality, and Paths of Development among New World Economies," *Journal of LACEA Economia* 3, 41-109.

Hanssen, F. Andrew, 2004, "Learning about judicial independence: Institutional change in the State courts." *The Journal of Legal Studies* 33, 431-473.

Perotti, E., and E.-L. von Thadden, 2006, "The Political Economy of Corporate Control and Labor Rents," *Journal of Political Economy* 114, 145-174.

Rajan, R., and R. Ramcharan, 2011, "Land and Credit: A Study of the Political Economy of Banking in the United States in the Early 20th Century," *Journal of Finance* 66, 1895-1930.

Ramcharan, R., 2010, "Inequality and Redistribution: Evidence from U.S. Counties and States, 1890–1930," *Review of Economics and Statistics* 92, 729-744.

Tables

TABLE I

VARIABLE NAMES, DEFINITIONS, AND DATA SOURCES FOR THE EMPIRICAL ANALYSIS OF JUDGES' RULING

Variable Name	Variable Definition	*Source*
Dependent Variable		
Bank Wins	A bank wins a case when it obtained in full what requested to the court, if it was a plaintiff, or a when the request of the opposing party was denied, when it was a defendant.	*WL*
Main Independent Variable		
County Inequality in 1890	The Gini coefficient of the distribution of farm land in 1890 in the county (for counties in Oklahoma the state-level coefficient is used)	*USC*
Control Variables		
Case Characteristics		
Bank is Plaintiff	= 1 if the Bank was a Plaintiff in the case, = 0 otherwise	*WL*
Number of West Headnotes	Number of West Headnotes Subcategories attributed to the trial by the Westlaw Database	*WL*
First Degree Trial Was a Summary Judgement	= 1 whether the first degree trial was a summary judgement, = 0 otherwise	*WL*
Second Degree Affirms First Degree	= 1 if the second degree trial affirms the sentence of the first degree trial, = 0 otherwise	*WL*
Dissenting Judges in the Second Degree Sentence	= 1 if there were any judges dissenting from the second degree sentence, = 0 otherwise	*WL*
More than Four Parties in Trial	= 1 if there are more than four parties involved in the first degree trial, = 0 otherwise	*WL*
Bank Located in the Same State where the Trial takes place	= 1 if the bank is located in the same state where the trial takes place, = 0 otherwise	*WL*
State and County Characteristics		
Length of Judges Term in Office	Number of Years judges stay in office according to the State legislation before facing re-election or re-appointment	*JS*
State has Judicial Partisan Elections	= 1 if in the State judges are electd via Partisan elections, = 0 otherwise	*JS*
County Population	Total county population at year-end	*USC*
County Catholic to Protestant Ratio	Ratio of the total number of Catholics divided by the total number of Evangelicals in the county at year-end 1980, 1990 and 2000	*ARDA*

NOTES. The table defines the variables used in the empirical analysis of judges' ruling, as well as the corresponding data sources used. *t-1* indicates a one year lag is used in the empirical analysis. Data sources include: *ARDA* = Association of Religion Data Archives; *JS* = Judicial Selection.com *USC* = US Census; *WL* = Westlaw.

TABLE II

DESCRIPTIVE STATISTICS FOR THE EMPIRICAL ANALYSIS OF JUDGES' RULING

Variable Name	Number of Observations	Mean	Standard Deviation	10%	Median (50%)	90%
Dependent Variable						
Bank Wins	1,392	0.52	0.50	0	1	1
Main Independent Variables						
County Inequality in 1890	1,392	0.32	0.16	0.18	0.27	0.55
Control Variables						
Case Characteristics						
Bank is Plaintiff	1,392	0.38	0.49	0	0	1
Number of West Headnotes	1,392	7.93	6.57	2	6	16
First Degree Trial Was a Summary Judgement	1,392	0.36	0.48	0	0	1
Second Degree Affirms First Degree	1,392	0.53	0.50	0	1	1
Dissenting Judges in the Second Degree Sentence	1,392	0.08	0.27	0	0	0
More than Four Parties in Trial	1,392	0.21	0.41	0	0	1
Bank Located in the Same State where the Trial takes place	1,392	0.55	0.50	0	1	1
State and County Characteristics						
Length of Judges Term in Office	1,381	5.81	2.12	3	6	8
State has Judicial Partisan Elections	1,392	0.27	0.44	0	0	1
County Population	1,378	0.56	0.59	0.02	0.38	1.82
County Catholic to Protestant Ratio	1,392	0.95	0.81	0.11	0.67	2.19

NOTES. The table provides the number of observations, mean, standard deviation, 10th percentile, the median (50th percentile) and the 90th percentile of all variables used in the empirical analysis.

TABLE III
MAIN SPECIFICATIONS COUNTY INEQUALITY IN 1890 EXPLAINING JUDGES' RULING

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable					*Bank Wins*			
County Inequality in 1890	-0.106	0.065	-0.205	-0.206	-0.194	0.935*	0.959*	0.909
	(0.149)	(0.243)	(0.135)	(0.136)	(0.134)	(0.541)	(0.503)	(0.546)
State Has Judicial Elections * County Inequality in 1890	--	-0.205	--	--	--	--	--	--
	--	(0.302)	--	--	--	--	--	--
State Has Judicial Partisan Elections * County Inequality in 1890	--	--	0.716*	0.715*	0.662*	--	--	--
	--	--	(0.369)	(0.371)	(0.366)	--	--	--
Length of Judges Term in Office * County Inequality in 1890	--	--	--	--	--	-1.098*	-1.123*	-1.078*
	--	--	--	--	--	(0.622)	(0.586)	(0.622)
Bank is Plaintiff	0.044	0.044	0.043	0.040	0.044	0.044	0.105	0.043
	(0.033)	(0.033)	(0.033)	(0.043)	(0.033)	(0.033)	(0.184)	(0.032)
Bank Located in the Same State where the Trial takes place	-0.043	-0.043	-0.043	-0.043	-0.010	-0.044	-0.044	0.273
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.034)	(0.034)	(0.198)
Number of West Headnotes	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
First Degree Trial Was a Summary Judgement	0.045	0.045	0.044	0.044	0.044	0.046	0.046	0.046
	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)	(0.050)
Second Degree Affirms First Degree	0.020	0.020	0.020	0.020	0.018	0.020	0.020	0.019
	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)	(0.033)	(0.033)	(0.033)
Dissenting Judges in the Second Degree Sentence	-0.021	-0.020	-0.020	-0.020	-0.020	-0.019	-0.019	-0.021
	(0.057)	(0.056)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)	(0.057)
More than Four Parties in Trial	-0.127***	-0.127***	-0.127***	-0.127***	-0.125***	-0.128***	-0.128***	-0.126***
	(0.047)	(0.047)	(0.046)	(0.047)	(0.045)	(0.047)	(0.047)	(0.047)
State Has Judicial Partisan Election s*Bank is Plaintiff	--	--	--	0.011	--	--	--	--
	--	--	--	(0.059)	--	--	--	--
State Has Judicial Partisan Elections*Bank Located in the Same State where the Trial takes place	--	--	--	--	-0.124	--	--	--
	--	--	--	--	(0.400)	--	--	--
Length of Judges Term in Office*Bank Located in the Same State where the Trial takes place	--	--	--	--	--	--	-0.064	--
	--	--	--	--	--	--	(0.181)	--
Bank Located in the Same State where the Trial takes place * Length of Judges Term in Office	--	--	--	--	--	--	--	-0.331
	--	--	--	--	--	--	--	(0.212)
County Population$_{t-1}$	-0.009	-0.009	-0.027	-0.026	-0.024	-0.010	-0.009	-0.009
	(0.032)	(0.032)	(0.034)	(0.033)	(0.034)	(0.032)	(0.032)	(0.032)
County Catholic to Protestant Ratio	-0.009	-0.009	-0.003	-0.004	-0.000	-0.007	-0.007	-0.007
	(0.036)	(0.036)	(0.035)	(0.034)	(0.035)	(0.036)	(0.036)	(0.036)
Constant	0.283**	0.281**	0.253*	0.251*	0.211	0.284**	0.285**	0.286**
	(0.138)	(0.137)	(0.142)	(0.140)	(0.141)	(0.136)	(0.137)	(0.137)
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	1,337	1,337	1,337	1,337	1,337	1,337	1,337	1,337
R-squared	0.157	0.157	0.159	0.159	0.161	0.157	0.157	0.158

NOTES. All models are estimated with a Linear Probability Model. The definition of the variables can be found in Table IX. *t-1* indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE IV

ROBUSTNESS ANALYSIS: MAIN SPECIFICATIONS COUNTY INEQUALITY IN 1998 EXPLAINING JUDGES' RULING

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable	*Bank Wins*							
County Inequality in 1998	0.365	0.752	-0.007	0.018	-0.038	21.441	23.462*	24.692
	(0.917)	(1.132)	(0.828)	(0.828)	(0.852)	(13.669)	(12.194)	(14.901)
State Has Judicial Elections * County Inequality in 1998	--	-0.466	--	--	--	--	--	--
	--	(1.853)	--	--	--	--	--	--
State Has Judicial Partisan Elections * County Inequality in 1998	--	--	2.747	2.587	2.794	--	--	--
	--	--	(2.033)	(2.019)	(2.102)	--	--	--
Length of Judges Term in Office * County Inequality in 1998	--	--	--	--	--	-21.392	-23.480*	-24.694
	--	--	--	--	--	(14.052)	(12.554)	(15.347)
Bank is Plaintiff	0.102	0.101	0.106	0.147*	0.104	0.103	1.130**	0.104
	(0.065)	(0.065)	(0.065)	(0.078)	(0.066)	(0.065)	(0.487)	(0.065)
Bank Located in the Same State where the Trial takes place	-0.027	-0.028	-0.025	-0.020	-0.053	-0.027	-0.028	0.372
	(0.070)	(0.073)	(0.072)	(0.073)	(0.094)	(0.071)	(0.071)	(0.688)
Number of West Headnotes	-0.028	-0.027	-0.029	-0.028	-0.030	-0.028	-0.029	-0.028
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.036)
First Degree Trial Was a Summary Judgement	0.096**	0.096**	0.103**	0.106**	0.103**	0.093**	0.098**	0.094**
	(0.046)	(0.046)	(0.045)	(0.044)	(0.045)	(0.045)	(0.043)	(0.044)
Second Degree Affirms First Degree	-0.006	-0.006	-0.006	-0.008	-0.004	-0.013	-0.014	-0.015
	(0.058)	(0.059)	(0.057)	(0.057)	(0.056)	(0.058)	(0.057)	(0.058)
Dissenting Judges in the Second Degree Sentence	0.155*	0.153	0.149	0.169*	0.143	0.140	0.141	0.135
	(0.092)	(0.093)	(0.090)	(0.088)	(0.088)	(0.099)	(0.097)	(0.098)
More than Four Parties in Trial	-0.088	-0.088	-0.088	-0.085	-0.088	-0.085	-0.089	-0.083
	(0.063)	(0.063)	(0.063)	(0.061)	(0.063)	(0.064)	(0.063)	(0.064)
State Has Judicial Partisan Elections*Bank is Plaintiff	--	--	--	-0.174	--	--	--	--
	--	--	--	(0.131)	--	--	--	--
State Has Judicial Partisan Elections*Bank Located in the Same State where the Trial takes place	--	--	--	--	0.120	--	--	--
	--	--	--	--	(0.162)	--	--	--
Length of Judges Term in Office*Bank Located in the Same State where the Trial takes place	--	--	--	--	--	--	-1.055**	--
	--	--	--	--	--	--	(0.512)	--
Bank Located in the Same State where the Trial takes place * Length of Judges Term in Office	--	--	--	--	--	--	--	-0.410
	--	--	--	--	--	--	--	(0.709)
County Population$_{t-1}$	0.143	0.141	0.177	0.172	0.171	0.127	0.128	0.125
	(0.102)	(0.108)	(0.124)	(0.126)	(0.127)	(0.108)	(0.108)	(0.109)
County Catholic to Protestant Ratio	-0.035	-0.036	-0.028	-0.020	-0.035	-0.021	-0.028	-0.019
	(0.093)	(0.093)	(0.092)	(0.092)	(0.091)	(0.097)	(0.095)	(0.097)
Constant	0.378	0.381	0.185	0.146	0.197	0.083	0.078	0.050
	(0.904)	(0.900)	(0.850)	(0.839)	(0.870)	(0.801)	(0.787)	(0.780)
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	467	467	467	467	467	467	467	467
R-squared	0.264	0.264	0.268	0.271	0.270	0.267	0.272	0.268

NOTES. All models are estimated with a Linear Probability Model. The definition of the variables can be found in Table I. *t-1* indicates a one year lag. "Yes" indicates that the set of fixed effects is included. "No" indicates that the set of fixed effects is not included. Standard errors are clustered at the state level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

Chapter 3: The Dark Side of Social Capital? Battles and Mortgage Lending[48]

Mintra Dwarkasing[49]

[48] I would especially like to thank Fabio Braggion and Steven Ongena for helpful comments and suggestions. In addition I would like to thank Narly Dwarkasing, Fabiana Penas and seminar participants at Tilburg University, the EBC Banking Research Day and the 8th Finance and history workshop (Nijmegen).

[49] Dwarkasing is from EBC, CentER & Tilburg University.

Abstract

I study how a shock in culture and trust in the past still affects mortgage lending today. More specifically, I investigate the long run effects of the American Civil War, a defining event for culture in American history, on current mortgage lending approval between 2005 and 2011. Using a spatial regression discontinuity design, exploiting the random occurrences of battles during the Civil War, I find that location matters for credit extension: Being located in a county in which a battle took place fosters the probability of loan approval. However, applicants from minority groups, such as African Americans, have a significant lower probability to obtain a mortgage loan in counties where a battlefield during the Civil War was located compared to similar applicants in adjacent non-battle counties. Conditional upon approval they also receive lower loan amounts. I show that a channel through which this battle effect persists is culture: Counties in which soldiers actively fought during the Civil War show higher levels of social capital today. Additionally, I find that remembrances of Civil War battles are important for the persistency of local social capital: Those battle counties that actively remember through re-enactment groups have even higher social capital today compared to those battle counties that do not. Moreover, applicants from minority groups have a significant lower probability to see their application currently being approved in these counties compared to minority applicants from battle counties that do not actively remember through such groups. Furthermore, conditional upon approval, they are also granted a lower loan amount, suggesting a possible 'dark side' of social capital.

I. Introduction

That history can have a long run effect on economic outcomes and culture has gained attention in both the economics and finance literature in the past few years (Acemoglu et. al., 2001; Nunn, 2008). Historical events such as slave trade have been generally linked to negative contemporary economic outcomes such as underdevelopment (Nunn, 2008). Confirming these findings, Dell (2010) shows that Peru's mining mita, an extensive forced labour mining system in effect between the 15th and 18th century, lowers today's household consumption and increases the prevalence of stunted growth in children in those households located in districts that were subjected to the mita. More recently, Voth and Voigtlander (2012) show that the occurrence of plague era pogroms predict anti-Semitic behaviour in the same locations centuries later.

That location and current local factors also matter in lending, for instance for local credit availability and loan pricing, is generally recognized in the finance literature (see for example Ghent, 2014; Gilje, 2010). In mortgage lending it is for example known that, before the introduction of the Home Mortgage Disclosure Act in 1975, it was nearly impossible for minorities to secure mortgages for property located in so-called redlined zones. However, whether history can have a long run effect on lending, through its persistent influence on local factors, is something that has been overlooked in the current literature.

This study adds to the existing literature by investigating whether a defining event in US history still matters for local lending. More specifically, I investigate in this paper the long run effect of one of the major historical events in American history, the American Civil War, on individual mortgage lending approval in counties in the Southern US states between 2005 and 2011. Additionally, I investigate a channel of persistence through which the Civil War still has an effect on current credit extension. Looking at individual mortgage applications allows me to not only investigate the average effect of the Civil War on current mortgage extensions, but also to look at more detailed effects such as whether the effect can differ for different groups of applicants. I employ a spatial regression discontinuity design to address endogeneity problems arising from omitted variables bias and hence draw causal inferences. For this I exploit the random occurrence of battles in US counties in the southern States during the War. At that time the southern states heavily relied on slavery and were against its abolishment. During the Civil War 122 counties experienced a battle whereas other counties did not[50]. Additionally I explore channels of persistence. After the battle counties were usually left devastated, meaning that at least part of the county had to be rebuilt. According to anecdotal evidence the necessary reconstruction of the county lead to a common unity within the county population and a subsequent increase in social capital, suggesting that social capital may be a channel of persistence.

My identification strategy builds on three steps. Firstly, using a spatial regression discontinuity design, I am able to compare mortgage applications of individuals located near the border of a county that was struck by a battle to applications from individuals located at the other side of the border in similar adjacent counties in which no such battle took place. Figure 1 illustrates my regression discontinuity approach for one specific battle county. I find that for mortgage applications there indeed is a discontinuity at the border with respect to approval probability as well as (conditional upon approval) the loan amount granted between individuals from battle and non-battle counties. This discontinuity is far larger for minority applicants. In my empirical analysis I further asses the local effect at the border of a battle county and its similar adjacent counties. Specifically, I examine whether being located in a county in which a battle occurred during the Civil War matters for individual mortgage approval today. My results confirm the conventional agreement that history can have a persistent effect on economic outcomes (Dell, 2010; Nunn, 2008). Contrary to the mostly negative welfare effects Civil War

[50] Out of the current 3144 county and county equivalents.

usually is associated with (see for instance Collier et. al, 2013), battle occurrences have a positive contribution: Being located in a county in which a battle took place fosters the probability of loan approval. However, this effect is mitigated for minority applicants. In fact, minority applicants' mortgage applications are less likely to be approved when the applicant is located in a county that experienced a battle as opposed to applications from minority applicants located in non-battle adjacent counties on the other side of the border. Moreover, conditional upon approval, the loan amount granted is also significantly smaller for minority applicants that are located in a battle county.

Exploring social capital as a channel of cultural persistence, where culture can be defined as "transmission from one generation to the next, via teaching and imitation, of knowledge, values, and other factors that influence behaviour" (North, 1990) points out an explanation for these findings: I document, in line with anecdotal evidence, that counties that were hit by a battle indeed showed significantly higher levels of social capital shortly after the Civil War compared to similar adjacent counties in which no such battle took place[51]. This difference in the level of social capital cannot be explained by differing levels of social capital between the two groups of counties shortly before the beginning of the Civil War and suggests a natural discontinuity at the border between battle counties and adjacent non-battle counties.

Moreover, I find evidence consistent with the notion that one of the channels of persistence is indeed social capital: Battle counties have a considerable higher social capital index today compared to similar non-battle counties. These results may suggest that social capital can lead to positive outcomes in the form of a higher mortgage approval probability. The common unity that was created within the population in a county after a battle took place has been transmitted from generation to generation resulting in higher local social capital today

[51] Social capital is measured by total county church value. For an exact definition see section 3.2.

and, on average, a higher mortgage approval probability compared to non-battle counties. However, they also suggest in line with Putnam (1996) and Fukuyama (1999), that there is a 'dark side' to social capital: On the contrary, 'outsiders', i.e. minority (non-white) individuals benefit negatively from this higher level of social capital.

Additionally, I find that remembrances of Civil War battles are important for the persistency of local social capital: In battle counties in which at least one military unit is present that actively re-enacts battles, both social capital (around 90 percentage points) and the likelihood that a non-minority mortgage application is approved (around 7 percent) are higher compared to battle counties with no such re-enactment groups.

This paper relates to and builds on several strands of literature. Firstly, it extends the growing literature that acknowledges that there can be a long term effect of historical events on economic outcomes (Nunn, 2008; Dell, 2010). Secondly, it relates to the current research on the effect of local factors such as a possible existence of discriminatory lending practices on mortgage pricing in the US mortgage market (see Ghent, 2014) and the role of culture in lending (Fisman et. al., 2012). Additionally, this paper also is more broadly related to the literature on the persistence of cultural traits (Voth and Voigtlander, 2012) and the general welfare effects of civil war (Collier, 2013). Moreover, it broadly contributes to the literature on group association and social cohesion after civil wars (Gilligan et. al, 2013), that shows that social capital increases after a civil war in those regions that were most heavily hit by the war.

The rest of the paper is organized as follows. Section 2 provides the historical background. Section 3 discusses the theoretical model, whereas the data is touched upon in section 4. The empirical results on the effect of battles during the American Civil War on local mortgage lending as well as a channel of cultural persistence are discussed in section 6. Section 7 concludes.

II. Background: The American Civil War

The American Civil War, also known as the 'War between States" was the bloodiest war in American history. During this war, that started when seven initial southern states seceded themselves from the Union and became the Confederate States of America (during the war four more states followed the initial confederate states and seceded from the Union as well) over 600,000 soldiers died at the battlefield and much of the South's infrastructure was left devastated. The cause of the secessions was a sectional friction over slavery between the Confederacy and the Union (the 'North'). The southern states, which relied heavily upon slavery, supported the possible expansion of slavery into the West, whereas the remaining states in the Union did not. During four years armies from the Confederates fought against the Union. The battles of the American Civil War were fought between April 12, 1861 and May 1865 in 23 states, but the South was most heavily hit (see map 1 in the Appendix)[52]. In my initial analysis I exclude the battles that took place in the 7 Northern states, as at that time the North was already starting to industrialize and hence northern counties were not entirely comparable to their southern counterparts[53].

II.1. To what extent were battle locations random?

Where exactly it came to clashes during the American Civil war was to a certain extent random, especially when conditioning upon counties in close proximity of a battlefield. For example, a current local newspaper documents on the battle of Lynchburg, VA that: '*That the*

---

[52] I include all battles in the sixteen southern states as defined by the US Census. Of course, the South and the geographical defined Southern states do not fully coincide. I therefore rerun all my analyses twice, firstly including only the 11 Confederate states and secondly including the states from the Old South, i.e. southern states in which slaves were legally held before 1860. The results remain largely unaltered.

[53] Additionally I rerun the analysis including the few battles from non-Southern states. Again, the results remain largely unaltered.

climactic battle began in Lynchburg on July 1 was as much by accident as design. (..)Lynchburg happened to be where some units under Lee finally engaged Union troops under Gen. George Meade.'

However, several circumstances were taken into account by generals from both sides when determining their optimal war strategy. From historical resources it is known that the War largely followed the lines of communication and supply, i.e. water ways and rail roads were of great importance and therefore also important points of interest for the enemy (Clark, 2004).

Another factor of importance for generals and their soldiers was the elevation of a location. A more elevated location, such as a hill or even mountain, tended to be of great advantage for defenders during the Civil War especially in combination with a so called 'unobstructed field of fire'.

Given that lines of communication, in the form of railroads/train stations and river access as well as elevation were important factors that contributed to the exact location of battles I include three variables that capture and control for these factors in the empirical analysis: A dummy variable, *Rail*, indicating whether a county in which a battle occurred had access to a railroad in the county in 1860, a year before the start of the war. *Water*, a dummy variable indicating whether a main river was present in the county in 1860. And lastly I include a county's elevation in meters, as measured by the elevation of a county's centroid.

III. Method

To investigate the long term effect of the American Civil War on mortgage lending I exploit the random occurrences of battles during the American Civil War. I first test the long run effect of the American Civil War on contemporary local lending practices. In particular I investigate whether being located in a battle county matters for mortgage approval. Moreover, I examine whether an applicant's race conditional upon being located in a county that experienced at least one battle during the Civil War alters the probability that the mortgage application is approved. Battle treatment is a function of location: the latitude and longitude of an observation, hence suggesting the use of a regression discontinuity approach. Exploiting the random location of battles during the Civil War and the general notion that the reconstruction of the county after a battle created a common unity within the county population I use the following spatial regression discontinuity design to test this effect:

$$A_{ict} = \propto + \beta_1 * BattleCounty_c + \beta_2 * BattleCounty_c * B_i + \beta_3 * B_i + X'_c\, \beta + f(geographical\ location_{ic}) + \varphi_b + \gamma_s + \delta_y + \varepsilon_{ict}$$

Where A_{ict} is a dummy variable indicating whether a mortgage application from applicant *i* in county *c* in year *t* is approved yes or no. $BattleCounty_c$ indicates whether the application from individual *i* is from a county that endured a battle during the US Civil War or from one of its neighbouring counties, $BattleCounty_c*B_i$ indicates the interaction between the battle indicator and an indicator for the race of the applicant (B_i: Minority (non-white) yes or no) and X'_c includes specific county characteristics: The county land and water area in miles respectively as a battle could only be fought on land (I do not include naval battles, as they are not included in the battles list from the American Battlefield Program). As well as whether a county had access to a railroad or waterway in 1860, county elevation , the number of soldiers that died during battles in a county and a historical measure of inequality: A county's gini coefficient for 1860 as in Nunn (2011). Additionally I control for a county's current urbanization level and applicant characteristics in the form of yearly income and gender. $f(geographical\ location_{ic})$ contains the regression discontinuity polynomial, which is a control function that captures the functional form of an applicant's location and defines the relationship between an applicant's location and the outcome variable, the approval indicator and loan amount respectively. In line with Dell (2010) and Michalopoulos (2014) I explore several forms: 1. The location (latitude and longitude) of the property[54] 2. The distance to the nearest battle 3. Distance to the border of the county where the nearest battle took place. Φ_b includes battle fixed effects. γ_s captures state fixed effects and δ_y includes year fixed effects. I include these broad sets of fixed effects to control for unobservable constant factors at the state level as well as for possible specific year effects that can influence mortgage approval (following Michalopoulos et al., 2014). Additionally I restrict my sample to those applications that are approved and rerun above specification replacing the approval indicator on the left hand side with a variable capturing the loan amount granted.

III.1. Hypotheses

The general intuition on what long run effect the Civil War can have on local mortgage approval leads to two testable hypothesis, that both are related to the extensive general literature on group association and social cohesion after wars (Gilligan et. al., 2013), social capital and the cultural traits that are part of it (Voth and Voigtlander, 2012) as well as the dark side of social capital (Voth and Voigtlander, 2012). In specific, considering the historical anecdotal evidence that battles resulted in an increase in social capital in those counties affected and hence may be a channel of cultural persistence through which a battle effect still persists today, I expect that mortgage applications in areas that experienced a battle during the Civil War are:

1. *More likely to be approved on average.* The devastating effect of a battle contributed to the local community feeling, thereby positively contributing to the 'within group feeling' and social capital which was passed on to other generations, in line with the formation of a so called 'collective memory' of American Wars that was passed

[54] I use a cubic polynomial in latitude and longitude and also a cubic polynomial for the distance to the nearest battle.

through to younger generations (Zaromb et al., 2013) and social cohesion after civil wars (Gilligan et. al., 2013). Based upon this hypothesis I therefore expect a positive relationship between a battle and the probability that a mortgage is approved. This expectation is in line with the positive effects of higher social capital in lending due to higher levels of trust it is commonly associated with (see Guiso et. al., 2004).

In addition I expect that, if there indeed is a 'dark side' to social capital:

2. *Less likely to be approved when the applicant is minority, i.e. of non-white race.* The Civil War started because seven Southern slave states seceded themselves from the Union as they opposed the abolition of slavery and supported its expansion into the West. Conditioning upon the Southern States, I therefore expect that applications from minority applicants are less likely to be approved compared to non-minority race applications in counties where a battle took place as opposed to similar applications near the border from adjacent non-battle counties. I expect a possible channel through which cultural persistence exists to be social capital: Because the local community was forced to rebuild a great part of the county because of the devastating effect of the battle, this created a common unity within the local community which was maintained throughout the years, positively contributing to local social capital. However, 'outsiders', i.e. individuals of another race, will be less likely to benefit from this local social capital. Consistent with the social capital channel of cultural persistence, I therefore expect that battle counties also have *higher* levels of social capital compared to non-battle counties.
3. Conditional upon being granted a mortgage, the loan amount granted will be *smaller* for minority applicants compared to non-minority applicants in battle counties as opposed to non-battle counties.

IV. Data

Battles

I obtain a list including all battles that took place during the Civil War from the Civil War Sites Advisory Commission (CWSAC). As the county in which a battle took place is known I am able to construct a battle county indicator based upon this information. I code the location (latitude and longitude) of each battle to be the centroid of the county in which it took place.

Mortgage applications

From the Home Mortgage Disclosure Act database (HMDA) I obtain all individual mortgage applications received by financial institutions between 2005- 2011 (including). This rich dataset comprises, amongst others, information on whether the application was denied or approved, applicant and, if any, co-applicant's race as well as the location of the property (its census tract), which allows me to calculate the distance to the nearest battle, border of the battle county and to use these as well as the locations' latitude and longitude as RD polynomials. I focus on all mortgage applications from the 16 Southern states as defined by the US Census. I do this for two main reasons: 1. Southern states as well the southern counties were (and still are) homogenous, a necessary condition in order for my identification strategy to be valid. Additionally this will limit any problems concerning omitted variables 2. Most of the battles took place in the Southern states[55].

I drop all observations with missing data and code my approval indicator to be equal to one

[55] When I do include all battles (and hence all states) the results remain largely unaltered as well as when including only the 11 southern Confederate States or the southern states belonging to the 'Old South', i.e. those states in which slavery was common before 1860.

if the mortgage application was approved (irrelevant of whether it was accepted by the applicant or not) and 0 otherwise. I construct a dummy variable, B_i, based upon the race of the main applicant and, if present, the co-applicant. I code this variable to be equal to one if both the applicant and (if present) the co-applicant are of non-white race and to be equal to 0 otherwise. The HMDA (Home Mortgage Disclosure Act) data includes public loan data from lending institutions and was enacted by the US Congress in 1975. The approach of the HMDA data is largely local of nature as it was initially set up for the purposes of identifying whether financial institutions are serving the housing needs of their communities, to assist public officials in attracting private investments to areas where it is needed and to identify discriminatory lending practices. One of the requirements for financial institutions that are required to report under HMDA is that an institution should be actually physically based in a certain community with at least one branch office. This local feature of the data allows to validly assuming that the applications as reported in HMDA deal with a lender that is active with a branch or office in the same county as the location of the property.

County controls

I obtain county level data from various sources. Historical data from 1860 and 1870 on churches, rail and water access as well as prosperity and inequality from the US decennial censuses from these years. Elevation is obtained from the Geographic Names Information System database and the level of county urbanization is from the US Census. Data on the total number of soldiers that died during Civil War battles in a county is obtained from the CWSAC.

IV.1. Summary Statistics – Corroborating the Identification assumption

For my identification strategy to be valid two identifying assumptions are required (see Angrist and Pischke (2008)): Firstly, the location of a battle should not have been influenced by local factors that influence mortgage approval and/or economic prosperity, i.e. they should be random. As discussed in section II, anecdotal evidence suggests that this indeed was not the case, except for some important factors, the presence of rail and water ways as well as county elevation. I control for these factor in the analysis. Moreover, counties need to be similar on all dimensions except for treatment (whether there was a battle or not). I explore these issues in more detail in the following tables.

[Table I around here]

I start by exploring whether the counties in my sample were similar a year before the start of the Civil War, in 1860. For this I look at four important factors obtained from the US Census: The number of slaves and white males in a county as a proxy for the extent to which a county relied on slavery, the number of manufacturing establishments as a proxy for the extent of industrialization and the value of real estate as a measure of economic prosperity. The means for each of these factors are reported in Table I for battle and non-battle counties respectively[56]. As can be seen, battle and non-battle counties were similar; all means do not differ statistically at conventional significance levels between both groups of counties. In Table II I further explore whether the counties are still similar nowadays, to rule out that my findings are caused by current observational differences between battle and adjacent non-battle counties. For this I collect data on three relevant factors: county median household income and the percentage of population living in poverty as measures for economic prosperity as well as the number of bank branches in a county as a measure of the availability of credit supply.

[Table II around here]

[56] I report robust standard errors as well as Conley standard errors that account for spatial correlation.

The results again indicate that both battle and non-battle counties still have similar economic outlooks, household income per capita and the percentage of people living in poverty do not differ significantly. Additionally, my measure for credit supply, the number of bank branches per capita, does not differ significantly between the two groups of counties as well; it is 0.00034 for both county groups. To be able to use a spatial regression discontinuity design it is also important to know whether there indeed is a discontinuity at the border. I investigate whether there is significant difference between approval rates and loan amounts of mortgage applications for minorities in battle versus non-battle counties in tables IV and V respectively.

[Table III and IV around here]

The results indeed indicate a discontinuity at the border: Minority applicants have a lower approval rate and lower mortgage amount in battle counties compared to their neighbouring non-battle counties. Even though the differences are also statistically significant for non-minority applicants, the absolute differences are far larger for non-minority applicants. Additionally, Graph 1 shows that there indeed is a discontinuity exactly at the border for the approval rates and loan amounts granted to minorities for battle and non-battle counties respectively. Both the mean approval rates as well as the loan amounts are significantly lower for minority applicants in battle counties that are located within 5,10,15,20 or 25 kilometres from the border of their county compared to minority applicants from adjacent counties that live in a similar distance to the border of the battle county.

From the descriptive statistics in Table IV-b it can be derived that the total dataset comprises of more than 22 million mortgage applications in the period 2005-2011. Around 75 percent of all applications is approved and the average loan amount is $169 000. From all applications, 21 percent is from counties in which at least one battle took place. The majority of all applicants is male and is located in a county that had access to a water and or railway in 1860. Mean county elevation is 119 meters. On average counties have a size of 500 km^2.

[Table IV-b around here]

I turn to the testable hypotheses in the next section.

V. Results

V.1. Mortgage Approval Results

I start by estimating the effect of a battle during the American Civil war on the probability that a mortgage application is approved. Following Dell (2010) I explore several forms as regression discontinuity polynomial. Table V reports the results of the specification that uses a cubic polynomial in latitude and longitude to control for a smooth function of an observation's geographical location[57]. Column (1) limits the sample to counties within 10 kilometres of a battle, and columns (2), (3) and (4) restrict the sample to fall within 15, 20 and 25 kilometres to the border of a battle county respectively. To control for unobserved state and year effects I include state and year fixed effects in all specifications as well as battle fixed effects to capture possible differences between battles.

[Table V around here]

[57] Of course, there is no clear prediction on how the expected relationship between an applicant's location (its longitude and latitude) and the probability of approval should be specified in the RD polynomial. I therefore specify several polynomials, from a single polynomial to a fourth order polynomial. The results are similar and therefore I only report the results on the cubic (third order) polynomial.

Columns (1) – (4) all estimate that the occurrence of a battle in a county during the Civil War increases the probability that a mortgage application is approved by around 4 percentage points or around 6 percent compared to the average approval rate's standard deviation. This result suggests support for the finding that higher social capital increases lending (Guiso et. al., 2004). More interestingly however, the findings in columns (1) – (4) indicate that applicants of non-white race have a more than 3 percentage points lower probability to see their application being approved (5.2 percent) in a county where a battle occurred compared to applicants of another race. The estimated coefficients are always statistically significant at conventional levels. Moreover, the point estimates remain fairly stable when restricting the sample to be within a smaller distance of the nearest battle. These results indicate that the average positive long term effect of the Civil War on mortgage approval is mitigated for minority applicants.

I explore a single dimension RD polynomial (in line with Dell, 2010), distance to the nearest battle in kilometres in Table VI. Again I control for unobserved state and year effects by including state and year fixed effects in all specifications as well as battle fixed effects to capture possible differences between battles.

[Table VI around here]

The results confirm the findings from table V: Applicants of non-white race have a 3 percentage points (or 5.2 percent) higher probability to see their application being declined in counties in which a battle occurred. These results are reinforced when using as a single cubic RD polynomial in the distance to the border, the coefficients can be found in table VII.

[Table VII around here]

V.2. Loan Amount results

Table VIII explores whether location matters for the loan amount granted. The results show that this is indeed the case: Conditional upon approval of an application, minority applicants are granted significantly lower loan amounts compared to non-minority applicants when located in a battle county versus minority applicants that are located across the border in a similar non-battle county. The result, using a cubic polynomial in distance to the nearest battle and the usual county controls, is both statistically as well as economically significant. Minority applicants receive lower loan amounts in general. However, minority applicants from a battle county receive a loan amount that is an extra 7.5 percent lower. Surprisingly, being located in a battle county does not significantly affect the loan amount on average. The coefficient is not statistically significant in any of the specifications.

5.3 Higher ex-post probability of default

To investigate whether the findings may be due to differences in default probabilities between applicants from battle and adjacent non-battle counties, ideally I would like to have information on the mortgage performances of those applications that are approved. Unfortunately, such information is not available. I therefore re-estimate the regressions including the county's yearly change in the percentage home-owners that are of non-white and white race as well as the percentage of mortgages that are in default in a respective year in a corresponding zip code area as proxies for individual mortgage performance. I calculate the yearly percentage change in the number of home-owners that are minority and non-minority (i.e. of non- white race and of white race respectively) using the US Census ACS data and rely on Fannie Mae Loan Performance dataset, which includes the performance of their single family mortgage loans, for yearly zip code level mortgage default percentages. The results remain largely unaltered when controlling for this information.

Why would applicants of non-white race have a lower probability to have their mortgage application approved as well as higher loan amounts in counties that experienced a battle 150 years before? I turn to a likely channel of cultural persistence in the next section.

VI. Channel of Persistence : Social Capital

In this section I provide a channel of cultural persistence as documented in the historical literature through which a battle effect still persists today: Social capital. Before I turn to the questions of why the occurrence of battles during the Civil War would increase social capital in those counties in which they took place and whether the occurrence of battles really led to an increase in social capital, I first provide a more formal definition of what social capital exactly entails in the following section.

VI.1. Social Capital : A definition

I follow Putnam (1995) and Woolcock (2001) by defining social capital as "features of social life—networks, norms, and trust—that enable participants to act together more effectively to pursue shared objectives". This definition suggests, in accordance with most other definitions of social capital that community engagement, in the form of group formation as well as forms of civic activity and collective actions are central elements (Rupasingha et.al. (2006)). I therefore measure social capital by constructing a county social capital index based upon the density of the following horizontal organizations, building on Rupasingha et.al. (2006): Bowling centres, golf clubs, fitness centres, sports organizations, religious organizations as well as county voter turnout. It is this measure of social capital that will be used in the regression discontinuity design in this paper to test whether social capital is one of the channels through which the influence of Civil War battles persist. This measure best reflects county community engagement as well as the extent to which a community is close knit. Moreover, previous studies have already indicated that community level social cohesion, as measured by voting and community organization increases after a civil war (see Gilligan et. al, 2013). One explanation for this is the 'collective coping mechanism', indicating that individuals band together after violence in order to cope with other threats.

VI.2. Battles and the creation of social capital

Why would the occurrence of battles during the Civil War increase social capital in those counties in which they took place? And did the occurrence of battles really lead to an increase in social capital? Already from a very early start, even as early as before the actual end of the war, when the real fighting had not even been terminated, battles were being re-played in so-called battle re-enactments (Hadden, 1999). Counties struck by a battle had to deal with severe devastation of both its infrastructure and buildings. Apparently, the destruction and the subsequent necessary rebuilding of counties (often county courthouses and other important facilities that played a vital role in the county community, were left destroyed as well (see Clark, 2004)) led to a common unity within the community that was being remembered through battle re-enactments. Lauderdale County, Mississippi, for example, was the scene of the Battle of Meridian (its county seat), an event which helped shape the community into what it is today (Putnam, 2011). The battle of Meridian had a destructive effect on Lauderdale County. Railroads were destroyed and much of the area was burnt down to the ground. Although soldiers had not attacked local citizens during the attack, most citizens were without food for some days after the attack. In addition, the destruction was of such a large magnitude that Maj. General Sherman, the commander of the Union forces during the battle reportedly said: "Meridian (Lauderdale County's county seat) with its depots, store-houses, arsenal, hospitals, offices, hotels, and cantonments no longer exists." However, the county community put everything in her power to rebuild everything that was so brutally destroyed during the battle, as Michelle

Putnam, a local Lauderdale County citizen remarks: *'The county flourished as a vital and vibrant hub of railroad commerce until the Civil War brought destruction and devastation. But its resilient citizens rose from the ashes and soon an area once ravaged by war became a home for industry and innovators.'* (Putnam, 2011).

Several studies have already indicated that social capital increased after civil war. For example, community level social cohesion, as measured by voting and community organization increased after a civil war (see Gilligan et. al, 2013). One explanation for this is the 'collective coping mechanism', which suggests that individuals band together after violence in order to cope with other threats. Additionally, a battle created a shared experience within the county community, as the community together rebuilt what was damaged and endured the battle together. This should positively contribute to social capital as measured by civic engagement (Costa and Kahn, 2003). Whether social capital indeed increased in those counties that experienced a battle remains an empirical question. To investigate this one would ideally like to obtain historical data, for both just before and just after the end of the war, on civic engagement to construct a similar social capital index as the contemporary index that is used in the spatial regression discontinuity design. Unavailability of such historical data on a county level however, does not allow for such an analysis. I therefore use as a proxy for social capital one specific factor of civic engagement: churches. In table IX I test whether the value of all churches significantly differs between counties that endured a battle and (neighbouring) counties that did not both at the eve of the war in 1860 as well as four years after the termination of the war in 1870[58]. Even though this proxy for social capital is limited to one factor of civic

[58] The total value of county churches is obtained from the Census of Religious Bodies and is defined as: "The estimated value of the church buildings owned and used for worship by the reporting organizations together with the value of the land on which these buildings stand and the furniture, organs, bells, and other equipment owned by the churches and actually used in connection with religious services."

engagement, it is widely considered by many scholars that social capital is embedded in groups such as churches (see for example Putnam, 1995).

[Table IX around here]

The results in table IX indicate that before the start of the war, in 1860, the total churches value did not differ statistically between counties that later on would be the scene of a battle. In 1870 however, four years after the war, battle counties show significantly higher amounts of social capital, as proxied for by the total value of all churches present in a county, compared to non-battle counties. The empirical evidence therefore indeed suggests that the occurrence of a battle during the war is associated with a higher level of social capital shortly after the end of the war[59].

VI.3. A persistent effect of battles on Social Capital

This section examines whether changes in local county social capital brought about by battles may have persisted throughout time. Given the notion that the local (mostly) white community was forced to rebuild a great part of the county because of the devastating effect of the battle, this created a common unity within the local community which was maintained throughout the years, positively contributing to local social capital. I examine this hypothesis in Table X. The dependent variable is a measure of contemporary social capital similar to Ruphasinga et. al (2006). This social capital index is based upon the eigenvalues from the first principal component of the following number of establishments in a county in 2009: (a) civic

---

[59] I rerun the analysis using the total church values instead of scaling it by total county population, to see whether the effect indeed is caused by a change in the numerator (values) as opposed to a change in county population after a battle (denominator effect). Still battle counties show significant higher church values shortly after the Civil War compared to non-battle counties. Before the Civil War there is no significant difference.

organizations; (b) bowling centres; (c) golf clubs; (d) fitness centres; (e) sports organizations; (f) religious organizations as well as voter turnout. I control for the usual state, year and battle fixed effects in the analysis.

[Table X around here]

The results indeed suggest that counties in which a battle occurred during the Civil War have higher local social capital today compared to non-battle counties. The effect is quite significant: A battle county has a social capital index that is around 1.20-1.70 points higher (143 percent) compared to a non-battle county and this effect is statistically significant at conventional levels. Moreover, the effect is similar when I either include as multidimensional RD polynomial a cubic polynomial using the counties centroid latitude and longitude or when using a single dimension cubic RD polynomial in the distance to nearest battle as measured from the county seat location. The effect is quite significant, as the average index is 1.71, meaning that being located in a battle county almost doubles the value of the index.

These findings suggest that the common unity that was created within the white population in a county after a battle took place has been transmitted from generation to generation resulting in higher local social capital today of which applicants of white race benefit in the form of a higher probability of mortgage approval. However, 'outsiders', i.e. non-white individuals benefit negatively from this higher level of social capital as the probability of approval of a mortgage application for them is lower in counties in which a battle occurred .

VI.4. Persistency of Civil War effect through collective memory: Battle re-enactments

How can the occurrences of battles during the American Civil War, some 140 years ago, still have a persistent effect on mortgage approval through local social capital? I look deeper into this question next. One of the main ways in which the memory of the civil war has been kept alive is through the yearly re-enactments of battles. If the active passing of memories of the Civil War through re-enactments indeed results in a collective memory which may be associated with higher social capital, I can hypothesize the following:

1. The probability of approval of minority mortgage applications is, conditioning upon counties that experienced at least one battle, *lower* in those counties in which memories of the Civil War are actively passed to other generations through battle re-enactments compared to battle counties in which this is not the case. In the same line, the loan amount granted is expected to be *lower* for these applicants as well.

2. Counties that remember their Civil War Battles actively should have *higher* current levels of social capital compared to other battle counties that do not.

To measure counties' active remembrance of the Civil War I construct a variable, *Reenactment Group*, which is a dummy variable that =1 if there is at least one military unit currently located in a county that participates in re-enactment battles of the Civil War and =0 otherwise. I obtain information on re-enactment groups, also called 'military units' from the Civil War Re-enactments headquarters website, on which units and their location are listed. The results can be found in tables XI, XII and XIII respectively.

[Table XI, XII and XIII around here]

Hypothesis 1 is indeed confirmed: The active remembrance of the Civil War matters for mortgage approval and the effect is both statistically and economically significant at conventional levels. Minority mortgage applicants are less likely to see their application being approved. When they live in a battle county in which the war is actively being remembered through re-enactment groups, this effect is even 3.5 percentage points larger (5 percent). Moreover, the loan amount granted, conditional upon approval, is around 11 percent lower. For the loan amount, however, an increase in local social capital through re-enactments is associated with higher loan amounts of on average, 17 percent. Turning to the question whether active remembrance counties also have higher social capital levels, table XIII indicates that this is indeed the case: The presence of a re-enactment group is indeed associated with a higher social capital index, indicating that re-enactments matter for the persistence of social capital.

Taken together these results suggest that re-enactments matters for the persistence of social capital and its accompanying dark side in the form of lower probability of loan approval and loan amounts for minority applicants in battle counties compared to similar applicants on the other side of the border in adjacent non-battle counties.

VI.5. County Migration during and after the Civil War

A possible concern is that shortly after or even during the Civil War many people migrated from counties that were affected by a battle to other places where they could rebuild their lives and have better prospects. Historical Census Records however, indicate that this was not the case. According to Ferrie (2006), there was a decrease in mobility and internal migration from the eve of the Civil War to 1900. He finds that, for white native male individuals of age 55 at the time of the census 40 percent lived outside the state of birth in 1900, whereas the percentage was considerably higher in 1850 (45%). Additionally, the rate is even lower in 1990, 39 percent, suggesting that lifetime interstate migration was less common at the end of the twentieth century compared to the second half of the nineteenth century. Moreover, current high migration rates should work against finding any effect of an applicants' race and location in a battle or non-battle county on mortgage approval.

VII. Channel of Persistence : Discrimination

Another channel through which a battle effect can persist via social capital is discrimination as well as hostile behaviour towards minorities (see Durlauf, 2002 and Field, 2008). The exclusion of outsiders can result in a culture of discriminatory practices against minorities. In this section I explore this explanation by looking at whether battles significantly influenced aggressive behaviour against blacks in the form of lynchings in 1882, some years after the end of the War. Additionally I investigate whether violence against individuals of black race is also more pronounced today. For this I investigate whether a battle effect significantly influenced the number of reported hate crimes against blacks. The results are reported in Table XIV. Column (1) suggests that discrimination indeed mattered: Counties in which a battle occurred during the Civil War also had a significant higher probability of 20 percent to observe at least one lynching of a black individual in 1882. Moreover, as indicated by the results in column (2), the number of hate crimes against black individuals is also higher today. This finding is statistically significant at conventional levels: Battle counties reported a 9.8 percent higher amount of hate crimes committed against black individuals in 2009 compared to non-battle adjacent counties.

VIII. Conclusions

In this paper I investigate the long run effects of one the most important events in American history, the American Civil War, on current mortgage lending practices. The random occurrence of battles during the American Civil War provides a natural opportunity to test this effect while taking into account endogeneity by means of the use of a spatial regression discontinuity approach. The results indicate that, contrary to the general accepted notion that Civil War can have negative welfare effects, being located in a battle county fosters the approval probability of mortgage applications. However, they also suggest that mortgage applicants that are of minority race have a more than 7 percent lower likelihood of obtaining a mortgage in a battle county compared to applicants of non-minority race. In addition, conditional upon approval, minority applicants also receive significant lower loan amounts in these counties (around 13 percent lower) compared to non-minority applicants. I show that a channel through which this effect of battles still persists today is social capital: Battle counties have a considerable higher social capital index today compared to similar non battle counties. This result may suggest, in line with Putnam (1996) a 'dark side' to social capital: The common unity that was created within the white population in a county after a battle took place has been transmitted from generation to generation resulting in higher local social capital today. However, 'outsiders', i.e. non-white individuals benefit negatively from this higher level of social capital. When looking deeper into social capital as the channel of persistence I find that conditioning upon battle counties, those counties that still actively remember Civil War battles through the presence of re-enactment groups, show even higher levels of social capital than those battle counties that do not. Additionally, in these counties non-minority applicants have a significant higher probability to obtain a mortgage compared to their minority counterparts. Also, conditional upon approval, minority applicants receive even lower loan amounts in these counties compared to similar minority applicants from nearby non-battle counties.

References

Acemoglu, D., Johnson, S., Robinson, J.A., *“Colonial origins of comparative development: an empirical investigation”*, American Economic Review, vol. 91, 2001, p. 1369–1401.

Angrist, J.D.,Pischke, J.S., *“Mostly harmless econometrics: An empiricist's companion”*, 2008, Princeton University Press.

Clark Jr.,J.E., *“Railroads in the Civil War : The Impact of Management on Victory and Defeat”,* 2001, Louisiana State University Press, Baton Rouge.

Collier, P., *“Breaking the conflict trap: Civil war and development policy”*, World Bank Publications, 2003.

Costa, D.L., Kahn, M.E., *“Civic engagement and community heterogeneity: An economist's perspective”*, Perspective on Politics, vol.1, no.1, 2001, p.103-111.

Dell, M., *“The persistent effects of Peru's mining mita”*, Econometrica, vol.78, no.6, 2010, p.1863-1903.

Duarte, J.,Siegel, S.,Young, L., *“Trust and credit: the role of appearance in peer-to-peer lending”*, Review of Financial Studies, vol. 25, no.8, 2012, p.2455-2484.

Ferrie, J.P., *“Internal Migration”*, book chapter in “Historical Statistics of the United States: Millennial Edition”, vol. 1, 2006, p. 489- 504.

Fukuyama, F., *“Social capital and civil society”*, 1999.

Ghent, A.C., Hernández-Murillo, R., Owyang, M.T., “*Differences in subprime loan pricing across races and neighborhoods*”, Regional Science and Urban Economics, vol. 48, 2014, p. 199-215.

Gilligan, M.J., Pasquale,B.J., Samii, C., *“Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal”*, American Journal of Political Science, 2013.

Guiso, L., Sapienza, P., Zingales,L., *“Does Culture Affect Economic Outcomes?”*, Journal of Economic Perspectives, vol. 20, no. 2, 2006, p. 23-48.

Guiso, L., Sapienza, P., Zingales,L., *“The Role of Social Capital in Financial Development”*, The American Economic Review, vol. 94, no. 3,2004, p. 526-556.

Hadden, R.L., *“Reliving the Civil War: A reenactor's handbook”*, 1999, Stackpole Books, Pennsylvania.

Nunn, N., *“The Long Term Effects of Africa's Slave Trades”*, Quarterly Journal of Economics, vol. 123, no.1, 2008, p.139-176.

Nunn, N, Wantchekon, L., *“The Slave Trade and the Origins of Mistrust in Africa”*, American Economic Review, vol. 101, no.7, 2011, p.3221-52.

Putnam, R., *“Lauderdale County, Mississippi: A Brief History”*,2011, The History Press.

Putnam, R.D., *“Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America”*, Political Science & Politics, vol. 28, 1995, p.664-683.

Putnam, R.D., *“Bowling alone: The collapse and revival of American community”,* Simon and Schuster, 2000.

Rupasingha, Anil, Stephan J. Goetz, and David Freshwater. "The production of social capital in US counties." The journal of socio-economics 35.1 (2006): 83-101.

Stelios Michalopoulos and Elias Papaioannou National Institutions and Subnational Development in Africa The Quarterly Journal of Economics 2014 129: 151-213.

Tabellini, Guido. "The scope of cooperation: Values and incentives." The Quarterly Journal of Economics 123.3 (2008): 905-950.

Voigtländer, Nico, and Hans-Joachim Voth. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany", The Quarterly Journal of Economics 127.3 (2012): 1339-1392.

Woolcock, Michael. "The place of social capital in understanding social and economic outcomes." Canadian Journal of Policy Research 2.1 (2001): 11-17.

Zaromb, Franklin, et al. "Collective memories of three wars in United States history in younger and older adults." Memory & cognition (2013): 1-17.

Appendix

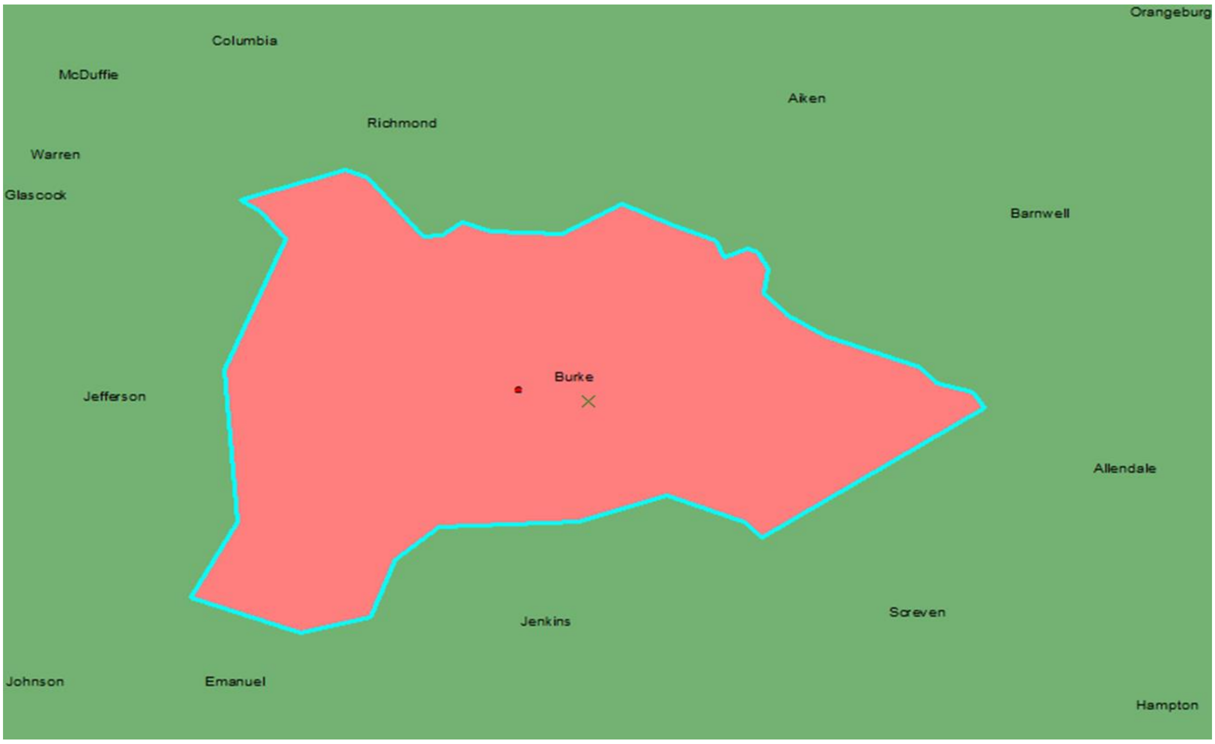
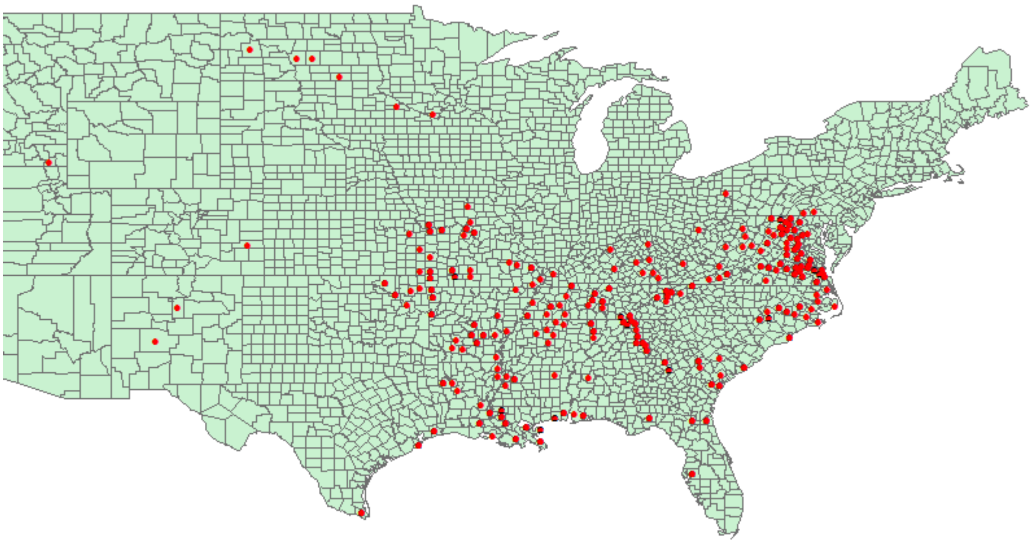


Figure 1. Example of regression discontinuity: Battle county and its corresponding border and non-battle similar adjacent counties. The blue line indicates a bandwidth of maximum 25 km on both sides of the border.



Map 1. All Battle locations during the American Civil War

Tables

Table I

Summary Statistics - 1860

Mean	*Battle County*	*Non Battle County*	*Standard error*
Slaves	0.31	0.31	(0.026)
			[0.026]
White Males	771.93	591.52	(193.137)
			[196.769]
Manufacturing establishments	58.24	44.48	(16.791)
			[16.934]
Real estate value	6,096,371	4,628,018	(1128972)
			[1196795.9]
Observations	89	139	

NOTES. The unit of observation is the county. The number of slaves is scaled by total population in a county and the real estate value is expressed in thousands of US dollars. Robust standard errors for the difference in means between battle and non battle counties are reported in parentheses. Conley standard errors, that take into account spatial correlation are reported in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. The data are taken from the 1860 US Census file through ICPSR.

Table II

Summary Statistics - 2005

Mean	*Battle County*	*Non Battle County*	*Standard error*
Household Income	44721.04	42780.63	(1309.243)
			[1559.213]
Poverty	14.01	14.48	(.603)
			[.709]
Bank branches	0.00034	0.00034	(.000018)
			[.000016]
Observations	92	140	

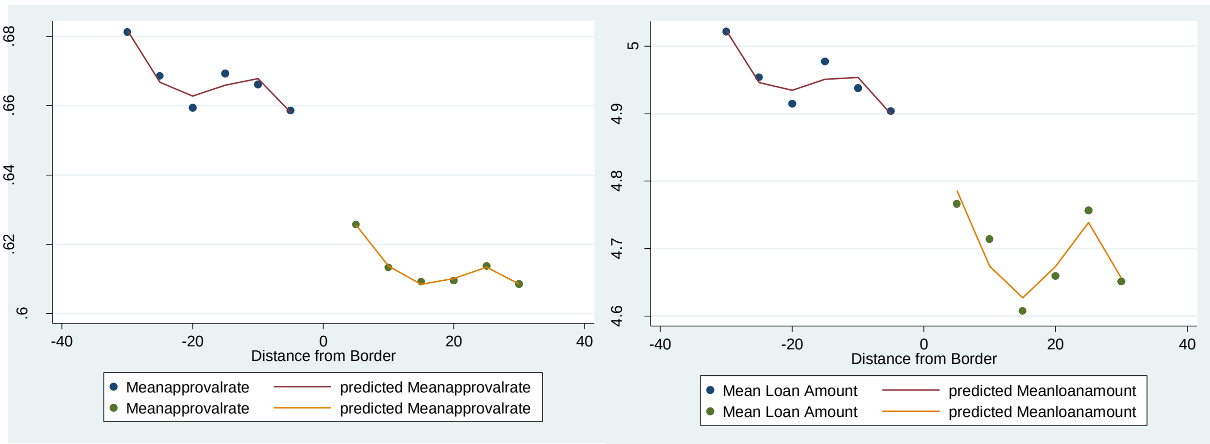
NOTES. The unit of observation is the county. The household income is the median household income expressed in thousands of US dollars in a county and poverty is expressed in percent, i.e. the percentage of all county population living in poverty. The number of bank branches are scaled by total county population. Robust standard errors for the difference in means between battle and non battle counties are reported in parentheses. Conley standard errors, that take into account spatial correlation are reported in brackets. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

Table III: Percentage of applications approved: Within 20 KM of county border

	Battle County	Non Battle County
Minority	60%	67%***
Non Minority	78%	78%***

Table IV: Mean Loan Amount: Within 20 KM of county border

	Battle County	Non Battle County
Minority	150.82	188.2108***
Non Minority	173.10	186.7477***



Graph 1. Discontinuity at the border: Approval rate and mean loan amount

Distance from the border is measured in kilometers. Each point represents the mean approval rate/loan amount (measured in natural logarithms) respectively in the corresponding 5km bin. The points with negative distances to the border represent 5km bins from adjacent non-battle counties.

Table IV-b: Summary statistics data sample

	Mean	St.Dev.	Min	Max	p10	p50	p90	Obs.
Approved	.7423277	.4373526	0	1	0	1	1	2.23e+07
Loan Amount	4.813918	.8336369	0	11.51292	3.688879	4.890349	5.799093	2.23e+07
Applicant income	100.8297	148.4097	1	9999	32	72	180	2.23e+07
Applicant gender	.6758279	.4680647	0	1	0	1	1	2.23e+07
County Battle	.217904	.4128218	0	1	0	0	1	2.23e+07
Minority	.2175939	.4126098	0	1	0	0	1	2.23e+07
County Battle * Minority	.0545478	.2270955	0	1	0	0	0	2.23e+07
Land area	21.17143	.7504563	15.4597	22.8873	20.37356	21.24717	22.12654	2.23e+07
Water area	18.08311	1.706449	1.94591	22.25562	16.0369	17.85415	20.5705	2.23e+07
County Minority percentage	76.22035	14.52625	19.4	98.9	56.7	77.6	93.1	2.23e+07
Died soldiers	1.472347	3.171831	0	11.57806	0	0	8.055158	2.15e+07
Urbanization	2.209926	.9584535	1	5	1	2	4	2.23e+07
Railroad access (County)	.5305573	.4990654	0	1	0	1	1	1.71e+07
River access (County)	.5477444	.4977153	0	1	0	1	1	1.71e+07
Elevation	3.489792	2.047118	0	7.110696	0	4.077538	5.686975	2.23e+07

NOTES. Loan amount is the natural logarithm of mortgage loan amount granted in thousands \$. Applicant income is measured in thousands of \$. Gender equals 1 if the applicant male and 0 otherwise. Minority equals 1 if the applicant's race is non-white and 0 otherwise. Land and water area are both measured in natural logarithms of the area in squared meters. Railroad and river acces equal 1 if an observation is located in a county in which a railroad or waterway is present and 0 otherwise. Elevation is measured as the natural logarithm of the elevation of a county's centroid in metres.

TABLE V
BATTLES AND MORTGAGE APPROVAL: CUBIC POLYNOMIAL IN LONGITUDE AND LATITUDE

Model	(1)	(2)	(3)	(4)
Distance to the border	< 10 KM	< 15 KM	< 20 KM	< 25 KM
Dependent Variable		*Mortgage Approved Dummy*		
Battle County	0.00578	0.0210	0.0296**	0.0341**
	(0.675)	(0.158)	(0.048)	(0.027)
Battle County * Minority	-0.0301*	-0.0308**	-0.0283**	-0.0273**
	(0.088)	(0.027)	(0.032)	(0.034)
Minority	-0.132***	-0.130***	-0.133***	-0.135***
	(0.000)	(0.000)	(0.000)	(0.000)
Landarea	-0.00931	-0.0109*	-0.0123**	-0.0139**
	(0.157)	(0.083)	(0.039)	(0.027)
Waterarea	-0.000563	0.00498	0.00552	0.00752**
	(0.818)	(0.149)	(0.116)	(0.035)
Percentage Non Minority Population	0.000872***	0.000813***	0.000716***	0.000511**
	(0.001)	(0.000)	(0.001)	(0.017)
Soldiers died	0.000458	-0.000399	-0.00169	-0.00238
	(0.795)	(0.838)	(0.367)	(0.205)
Urbanization	-0.00833	0.00289	-0.000577	0.00242
	(0.175)	(0.586)	(0.916)	(0.672)
Railroad access	0.0189***	0.0107*	0.0114**	0.0127**
	(0.005)	(0.067)	(0.042)	(0.028)
River access	0.0117	0.00988	0.00531	-0.00799
	(0.204)	(0.295)	(0.543)	(0.419)
Elevation	0.00245	0.00929**	0.00761*	0.00930**
	(0.517)	(0.025)	(0.054)	(0.023)
Inequality 1860	0.0551	0.0135	-0.000251	0.00635
	(0.323)	(0.773)	(0.995)	(0.882)
Battle Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Applicant characteristics	Yes	Yes	Yes	Yes
Number of Observations	4523803	5730115	6429856	6885294
Clusters	206	217	225	232
R-squared	0.040	0.039	0.039	0.039

NOTES. The unit of observation is an individual mortgage loan application. Standard errors are robust and adjusted for clustering at the county level. The dependent variable is a dummy variable taking on the value of 1 if a mortgage application is approved and 0 otherwise. All models include battle, state and year fixed effects respectively. All models include a cubic polynomial in latitude and longitude from the observation's census tract centroid to the nearest battle (as measured by the centroid of the county in which the battle took place). Model (1) includes only observations within 10 kilometers of a battle, model (2) within 15 kilometers, model (3) within 20 kilometers and model (4) within 25 kilometers. All regressions include county controls to control for land and water area, the percentage of contemporary non minority population,the amount of soldiers died during the Civil War battles, the level of contemporary urbanization, county railroad and river access in 1860, county elevation and county inequality in 1860 respectively. Railroad and river access are dummy variables taking a value =1 if a county had access to a railroad or was situated near a river and =0 otherwise. Elevation is measured in metres. Applicant characteristics included are gender and income. All regressions include controls for county land and water area respectively. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE VI

BATTLES AND MORTGAGE APPROVAL: CUBIC POLYNOMIAL IN DISTANCE TO THE NEAREST BATTLE

Model	(1)	(2)	(3)	(4)
Distance to the border	< 10 KM	< 15 KM	< 20 KM	< 25 KM
Dependent Variable	*Mortgage Approved Dummy*			
Battle County	0.0138	0.0316*	0.0328**	0.0345**
	(0.388)	(0.063)	(0.047)	(0.037)
Battle County * Minority	-0.0315*	-0.0338**	-0.0304**	-0.0290**
	(0.073)	(0.016)	(0.023)	(0.026)
Minority	-0.131***	-0.129***	-0.132***	-0.134***
	(0.000)	(0.000)	(0.000)	(0.000)
Landarea	-0.00747	-0.0106*	-0.0119**	-0.0147**
	(0.231)	(0.091)	(0.048)	(0.020)
Waterarea	-0.000603	0.00457	0.00423	0.00603**
	(0.804)	(0.147)	(0.165)	(0.048)
Percentage Non Minority Population	0.000779***	0.000669***	0.000595***	0.000390*
	(0.005)	(0.002)	(0.005)	(0.071)
Soldiers died	-0.000213	-0.00154	-0.00214	-0.00235
	(0.898)	(0.403)	(0.222)	(0.177)
Urbanization	-0.00683	0.00128	-0.00213	0.000962
	(0.281)	(0.823)	(0.664)	(0.841)
Railroad access	0.0122**	0.00399	0.00544	0.00907
	(0.041)	(0.452)	(0.298)	(0.103)
River access	0.00618	0.00440	0.00285	-0.00847
	(0.548)	(0.632)	(0.733)	(0.429)
Elevation	0.00343	0.0141***	0.0131***	0.0143***
	(0.379)	(0.002)	(0.001)	(0.000)
Inequality 1860	0.0133	-0.0472	-0.0683*	-0.0606
	(0.796)	(0.291)	(0.088)	(0.159)
Battle Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Applicant characteristics	Yes	Yes	Yes	Yes
Number of Observations	4523803	5730115	6429856	6885294
Clusters	206	217	225	232
R-squared	0.039	0.039	0.039	0.039

NOTES. The unit of observation is an individual mortgage loan application. Standard errors are robust and adjusted for clustering at the county level. The dependent variable is a dummy variable taking on the value of 1 if a mortgage application is approved and 0 otherwise. All models include battle, state and year fixed effects respectively. All models include a cubic polynomial in distance to the nearest battle from the observation's census tract centroid to the nearest battle (as measured by the centroid of the county in which the battle took place). Model (1) includes only observations within 10 kilometers of a battle, model (2) within 15 kilometers, model (3) within 20 kilometers and model (4) within 25 kilometers. All regressions include county controls to control for land and water area, the percentage of contemporary non minority population,the amount of soldiers died during the Civil War battles, the level of contemporary urbanization, county railroad and river access in 1860, county elevation and county inequality in 1860 respectively. Railroad and river access are dummy variables taking a value =1 if a county had access to a railroad or was situated near a river and =0 otherwise. Elevation is measured in metres. Applicant characteristics included are gender and income. All regressions include controls for county land and water area respectively. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE VII

BATTLES AND MORTGAGE APPROVAL: CUBIC POLYNOMIAL IN DISTANCE TO THE BORDER

Model	(1)	(2)	(3)	(4)
Distance to the border	< 10 KM	< 15 KM	< 20 KM	< 25 KM
Dependent Variable		*Mortgage Approved Dummy*		
Battle County	0.0418**	0.0434***	0.0363**	0.0379**
	(0.013)	(0.008)	(0.024)	(0.017)
Battle County * Minority	-0.0347**	-0.0383***	-0.0369***	-0.0331**
	(0.011)	(0.004)	(0.004)	(0.013)
Minority	-0.125***	-0.122***	-0.123***	-0.128***
	(0.000)	(0.000)	(0.000)	(0.000)
Landarea	-0.0158***	-0.0168***	-0.0151**	-0.0112*
	(0.010)	(0.008)	(0.011)	(0.081)
Waterarea	0.00110	0.000648	0.00159	0.00212
	(0.729)	(0.832)	(0.592)	(0.463)
Percentage Non Minority Population	0.000896***	0.000814***	0.000625***	0.000574***
	(0.000)	(0.000)	(0.001)	(0.005)
Soldiers died	-0.00340	-0.00321	-0.00231	-0.00283
	(0.114)	(0.130)	(0.258)	(0.151)
Urbanization	0.00201	0.00488	0.00737	0.00532
	(0.758)	(0.396)	(0.122)	(0.240)
Railroad access	0.00514	0.00465	0.00135	0.00344
	(0.419)	(0.414)	(0.792)	(0.487)
River access	0.00231	0.00159	-0.00646	-0.00536
	(0.834)	(0.899)	(0.567)	(0.630)
Elevation	0.00810	0.0130***	0.0120***	0.00960***
	(0.103)	(0.002)	(0.001)	(0.008)
Inequality 1860	-0.0591*	-0.0693*	-0.0710*	-0.0634
	(0.093)	(0.065)	(0.063)	(0.131)
Border Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Applicant characteristics	Yes	Yes	Yes	Yes
Number of Observations	4523803	5730115	6429856	6885294
Clusters	206	217	225	232
R-squared	0.040	0.040	0.040	0.040

NOTES. The unit of observation is an individual mortgage loan application. Standard errors are robust and adjusted for clustering at the county level. The dependent variable is a dummy variable taking on the value of 1 if a mortgage application is approved and 0 otherwise. All models include border, state and year fixed effects respectively. All models include a cubic polynomial in distance to the nearest border of a battle county measured from the observation's census tract centroid. Model (1) includes only observations within 10 kilometers of a border, model (2) within 15 kilometers, model (3) within 20 kilometers and model (4) within 25 kilometers. All regressions include county controls to control for land and water area, the percentage of contemporary non minority population,the amount of soldiers died during the Civil War battles, the level of contemporary urbanization, county railroad and river access in 1860, county elevation and county inequality in 1860 respectively. Railroad and river access are dummy variables taking a value =1 if a county had access to a railroad or was situated near a river and =0 otherwise. Elevation is measured in metres. Applicant characteristics included are gender and income. All regressions include controls for county land and water area respectively. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE VIII

BATTLES AND LOAN AMOUNT: CUBIC POLYNOMIAL IN DISTANCE TO THE NEAREST BATTLE

Model	(1)	(2)	(3)	(4)
Distance to the border	< 10 KM	< 15 KM	< 20 KM	< 25 KM
Dependent Variable		*Loan Amount*		
Battle County	-0.0562	-0.0409	-0.0267	-0.0193
	(0.445)	(0.617)	(0.732)	(0.810)
Battle County * Minority	-0.0727*	-0.0793**	-0.0660*	-0.0626*
	(0.095)	(0.039)	(0.085)	(0.089)
Minority	-0.0748**	-0.0683***	-0.0803***	-0.0843***
	(0.018)	(0.008)	(0.001)	(0.000)
Landarea	-0.00559	-0.0175	-0.0150	-0.0127
	(0.850)	(0.636)	(0.656)	(0.695)
Waterarea	-0.000415	0.0236	0.0224	0.0254*
	(0.975)	(0.146)	(0.149)	(0.095)
Percentage Non Minority Population	0.00226*	0.00179*	0.00170*	0.000836
	(0.052)	(0.064)	(0.062)	(0.335)
Soldiers died	0.00970	0.0133	0.00906	0.00871
	(0.214)	(0.200)	(0.343)	(0.377)
Urbanization	-0.101***	-0.0273	-0.0494*	-0.0372
	(0.005)	(0.453)	(0.063)	(0.103)
Railroad access	0.0858**	0.0172	0.0218	0.0352
	(0.012)	(0.551)	(0.386)	(0.167)
River access	0.0445	0.0366	0.0275	-0.00122
	(0.401)	(0.444)	(0.488)	(0.976)
Elevation	0.0174	0.0748**	0.0774***	0.0805***
	(0.368)	(0.020)	(0.007)	(0.007)
Inequality 1860	0.0832	-0.309	-0.365	-0.308
	(0.711)	(0.227)	(0.179)	(0.253)
Battle Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Applicant characteristics	Yes	Yes	Yes	Yes
Number of Observations	3014745	3817795	4284577	4586326
Clusters	206	217	225	232
R-squared	0.157	0.167	0.174	0.175

NOTES. The unit of observation is an individual mortgage loan amount. Standard errors are robust and adjusted for clustering at the county level. The dependent variable is the loan amount in thousands of US $. All models include battle, state and year fixed effects respectively. All models include a cubic polynomial in distance to the nearest battle from the observation's census tract centroid to the nearest battle (as measured by the centroid of the county in which the battle took place). Model (1) includes only observations within 10 kilometers of a battle, model (2) within 15 kilometers, model (3) within 20 kilometers and model (4) within 25 kilometers. All regressions include county controls to control for land and water area, the percentage of contemporary non minority population,the amount of soldiers died during the Civil War battles, the level of contemporary urbanization, county railroad and river access in 1860, county elevation and county inequality in 1860 respectively. Railroad and river access are dummy variables taking a value =1 if a county had access to a railroad or was situated near a river and =0 otherwise. Elevation is measured in metres. Applicant characteristics included are gender and income. All regressions include controls for county land and water area respectively. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

Table IX

Social Capital before and after the Civil War: Church value			
Mean	**Battle County**	**Non Battle County**	**Standard error**
Church value 1860	3.42	2.73	(.4456213)
Church value 1870	4.27*	3.36	(.5542734)
Observations	89	139	

NOTES. The unit of observation is the county. The church value is the total value of all church buildings in a county measured in thousands of US dollars and scaled by total county population. Robust standard errors for the difference in means between battle and non battle counties are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE X

CHANNEL OF PERSISTENCE: SOCIAL CAPITAL

Model	(1)	(2)
Dependent Variable	*Social Capital Index*	
Battle County	1.724*	1.202*
	(0.086)	(0.079)
Cubic RD Polynomial in	Latitude and longitude	Distance to nearest battle
Controls	Yes	Yes
Battle Fixed Effects	Yes	Yes
Number of Observations	232	232
R-squared	0.663	0.600

NOTES. The unit of observation is the county. The dependent variable is a county's social capital index from 2009. Standard errors are robust. The dependent variable is a social capital index based upon the first principal component of several variables such as sports membership and the number of civic organizations (similar to Rupasingha et. al, 2006). All models include battle fixed effects and county controls. Model (1) includes a cubic regression discontinuity polynomial in the counties' centroid longitude and latitude. Model (2) includes a cubic polynomial in the distance to the nearest battle as measured from the county seat location. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. When adjusting standard errors for spatial correlation using Conley standard errors the results remain unaltered.

TABLE XI

BATTLES REENACTMENT GROUPS: CUBIC POLYNOMIAL IN DISTANCE TO THE BORDER

Model	(1)
Dependent Variable	*Mortgage Approved Dummy*
Reenactment Group	0.0141
	(0.183)
Reenactment Group * Minority	-0.0350*
	(0.082)
Minority	-0.146***
	(0.000)
Landarea	0.00548
	(0.602)
Waterarea	-0.0142**
	(0.032)
Percentage Non Minority Population	0.00102**
	(0.046)
Soldiers died	-0.00227
	(0.384)
Urbanization	0.0143*
	(0.065)
Railroad access	0.0337*
	(0.092)
River access	0.0489***
	(0.001)
Elevation	0.000264
	(0.876)
Inequality 1860	0.0618
	(0.345)
Applicant characteristics	YES
State Fixed Effects	YES
Year Fixed Effects	YES
Number of Observations	3860318
Clusters	92
R-squared	0.044

NOTES. The unit of observation is an individual mortgage loan application. Standard errors are robust and adjusted for clustering at the county level. The dependent variable is a dummy variable taking on the value of 1 if a mortgage application is approved and 0 otherwise. The model includes border, state and year fixed effects respectively. The model includes a cubic polynomial in distance to the nearest border of a battle county measured from the observation's census tract centroid. The regression includes county controls to control for land and water area, the percentage of contemporary non minority population,the amount of soldiers died during the Civil War battles, the level of contemporary urbanization, county railroad and river access in 1860, county elevation and county inequality in 1860 respectively. Railroad and river access are dummy variables taking a value =1 if a county had access to a railroad or was situated near a river and =0 otherwise. Elevation is measured in meters. Reenactment group is a dummy variable that =1 if there is at least one military unit group present in a county that currently participates in battle reenactments and =0 otherwise. All observations are from counties in which at least one battle took place during the Civil War. All regressions include controls for county land and water area respectively. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE XII

BATTLES REENACTMENT GROUPS: CUBIC POLYNOMIAL IN DISTANCE TO THE BORDER

Model	(1)
Dependent Variable	*Loan amount*
Reenactment Group	0.160***
	(0.006)
Reenactment Group * Minority	-0.102*
	(0.087)
Minority	-0.0913***
	(0.002)
Landarea	0.110*
	(0.079)
Waterarea	-0.176***
	(0.001)
Percentage Non Minority Population	0.00101
	(0.712)
Soldiers died	-0.0118
	(0.298)
Urbanization	0.00308
	(0.928)
Railroad access	0.250**
	(0.026)
River access	0.231**
	(0.020)
Elevation	0.0462***
	(0.001)
Inequality 1860	0.855*
	(0.057)
Applicant characteristics	YES
State Fixed Effects	YES
Year Fixed Effects	YES
Number of Observations	3860318
Clusters	92
R-squared	0.178

NOTES. The unit of observation is an individual mortgage loan application. Standard errors are robust and adjusted for clustering at the county level. The dependent variable is the (natural logarithm of) loan amount in thousands of $. The model includes border, state and year fixed effects respectively. The model includes a cubic polynomial in distance to the nearest border of a battle county measured from the observation's census tract centroid. The regression includes county controls to control for land and water area, the percentage of contemporary non minority population,the amount of soldiers died during the Civil War battles, the level of contemporary urbanization, county railroad and river access in 1860, county elevation and county inequality in 1860 respectively. Railroad and river access are dummy variables taking a value =1 if a county had access to a railroad or was situated near a river and =0 otherwise. Elevation is measured in meters. Reenactment group is a dummy variable that =1 if there is at least one military unit group present in a county that currently participates in battle reenactments and =0 otherwise. All observations are from counties in which at least one battle took place during the Civil War. All regressions include controls for county land and water area respectively. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.

TABLE XIII

SOCIAL CAPITAL: REENACTMENT GROUPS

Model	(1)
Dependent Variable	*Social Capital Index*
Reenactment Group	0.937**
	(0.027)
Cubic RD Polynomial in	Latitude and longitude
Controls	Yes
Number of Observations	92
R-squared	0.638

NOTES. The unit of observation is the county. The dependent variable is a county's social capital index from 2009. Standard errors are robust. The dependent variable is a social capital index based upon the first principal component of several variables such as sports membership and the number of civic organizations (see Rupasingha et. al, 2006). Only counties in which at least one battle occured during the Civil War are included in the analysis. The model includes a cubic regression discontinuity polynomial in the counties' centroid longitude and latitude. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. When adjusting standard errors for spatial correlation using Conley standard errors the results remain unaltered.

TABLE XIV

CHANNEL OF PERSISTENCE: DISCRIMINATION

Model	(1)	(2)
Dependent Variable	*Lynching Dummy 1882*	*Hate crimes 2009*
Battle County	1.328**	0.183**
	(0.048)	(0.028)
Cubic RD Polynomial in	Distance to nearest battle	Distance to nearest battle
Controls	Yes	Yes
Number of Observations	232	232
R-squared	0.491	0.571

NOTES. The unit of observation is the county. The dependent variable is a dummy variable indicating whether one or more lynchings occured in the county in 1882 or not in cloumn (1) and the number of hate crimes reported against blacks in a county 2009 per 10,000 of population. Standard errors are robust. All models include battle fixed effects and county controls. Both models include a cubic polynomial in the distance to the nearest battle as measured from the county seat location. P-values are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. When adjusting standard errors for spatial correlation using Conley standard errors the results remain unaltered.

Chapter 4: The real effects of Tax Avoidance: The effect of Tax Avoidance on Corporate Innovation

Mintra Dwarkasing, Narly Dwarkasing, Jasmin Gider[60]

[60] Narly Dwarkasing is from University of Bonn, Institute for Financial Economics and Statistics. Mintra Dwarkasing is from EBC, CentER & Tilburg University. Gider is from University of Bonn, Institute for Financial Economics and Statistics.

Abstract

We empirically investigate the impact of tax avoidance on corporate innovation. On one hand, tax avoidance may be conducive to long-term investment because it increases internal funds and thereby alleviates financing constraints. On the other hand, tax avoidance may hamper innovation due to decreasing corporate transparency and agency conflicts, which may result in difficulties in obtaining external financing. We find evidence in favor of the latter: Exploiting exogenous variation in tax avoidance introduced by the so-called Check-the-Box regulations in 1997 in the US, we document a negative effect of tax avoidance on both innovation input and output, which is statistically and economically significant.

I. Introduction

The engagement of firms in tax avoidance is an increasing worldwide phenomenon and has been the subject of research to a growing number of scholars and policy makers. Yet, very little is known about the real effects of tax avoidance and its effect on economic growth remains unclear. In the U.S., for instance, there has been a steady decrease in corporate income tax receipts in the last years, whereas corporate profits have remained relatively stable over time. Contrary to tax evasion, where firms engage in illegal practices to decrease their ultimate tax burden, tax avoidance is a legal way in which firms can mitigate their tax payments, but can be characterized by firms moving in so called 'grey areas' of the tax regime present. Typically, firms can reduce tax payments through intercompany transactions between entities in countries with different tax rates which maximize (minimize) income in low (high) tax regimes to reduce the overall tax burden of the firm.

This paper studies how tax avoidance intensity of U.S. firms affects their innovation practices. More specifically, we investigate whether innovation input, as measured by R & D expenditures changes when the corporate engagement in tax avoidance is altered. Additionally, we investigate the effect tax avoidance intensity has on innovation output *quality* in terms of the number of patents filed and future patent citations scaled by total patents. To the extent that innovation input and output are positively correlated, tax avoidance can be beneficial or detrimental for innovation output in a similar way of reasoning as innovation input. Using a panel of firms from the entire Compustat universe between 1990-2004, including information on firms' financial characteristics, tax avoidance activities (as measured by a firm's book tax gap) as well as other firm characteristics and supplementing this information with yearly data on their patenting practices and research and development expenditures, our results provide evidence on how tax avoidance affects innovation practices.

Relating tax avoidance to innovation activity in the form of innovation input and output is particularly interesting as innovation is central to firms' long term growth and productivity (Bloom, 2014). This paper contributes to the so far rather underexplored question.

On the one hand, tax avoidance may have positive real effects and hence may enhance innovation because firms have more internal funds available to fund complex investment projects. Atanassov and Li (2014) and Mukherjee, Singh and Zaldokas (2015) for example analyse the effect of staggered state tax rate changes and find that higher tax rates are associated with less innovation.[61] These findings support the notion that lower tax payments alleviate financial constraints and thereby increase investment. Similar to tax rate decreases, tax avoidance effectively increases after-tax cash flows. If tax avoidance is not different from lower tax rates, following Atanassov and Li (2014), we would expect that tax avoidance is conducive to corporate innovation input as well as output, to the extent that both innovation input and output are positively correlated. We label this hypothesis the *alleviated financial constraints hypothesis*.

The range of the existing evidence is wide and does not solely point out a positive impact of tax avoidance on innovation. Existing evidence also suggests that tax avoidance entails costs which makes it qualitatively very different from lower tax rates. Several scholars note that an increase in the intensity of tax avoidance behavior may increase the complexity of transactions and decrease corporate transparency (Balakrishnan, Blouin and Guay, 2012) which in turn makes it easier for managers 'to engage in opportunistic behavior such as earnings management and rent extraction' (Guo, 2014; Katz et al. 2015).[62] Additionally, banks may perceive tax

[61] Becker, Jacob and Jacob (2013) investigate the impact of payout taxes on corporate investment and document a large impact of payout taxes on the dynamics of corporate investment and growth which is driven by financial constraints.

[62] Some papers also acknowledge a reverse relationship, where the level of tax avoidance can be influenced by the internal information quality (Gallemore, 2015).

avoidance as giving rise to significant informational risks because of the accompanying decrease in corporate transparency, which may lead to firms having more difficulties in obtaining bank funding to finance their investments. Hasan et. al. (2014) for example find evidence pointing in this direction: Firms with greater tax avoidance practices receive a 'tax avoidance penalty' as they pay higher spreads on their bank loans. Following Hasan et. al. (2014) we therefore posit another channel through which tax avoidance may affect corporate innovation, which we label the *opacity hypothesis*. According to this hypothesis we expect that when tax avoidance becomes more attractive and firms subsequently increase their tax avoidance practices, the accompanying decrease in internal corporate transparency and increase in organizational complexity result in costs that may outweigh the benefits of tax avoidance. Hence, we would expect to see a *negative* relationship between tax avoidance and corporate innovation input and also output (to the extent that innovation input and output are positively correlated): An *increase* in the intensity of tax avoidance should lead to a *decrease* in R&D expenditures as well as innovation output as measured by total patents and their forward citations.

Assessing a causal relationship from tax avoidance to corporate innovation provides us however, with quite some challenges, as for example corporate innovation in the form of high patent activity itself may help firms in sheltering income through the use of, amongst others, subjective assets. Additionally, the extent to which corporations engage in tax avoidance may be correlated with unobserved factors that also affect innovation. Successful firms may for instance also be those that are most efficient in tax avoidance practices. We address these concerns in several ways. Firstly, we exploit an exogenous shock in the variation in tax avoidance caused by the introduction of the so called 'Check-the-Box' regulations in 1997 in the US, which caused a positive shock in tax avoidance engagement (or more specifically, a reduction in the costs of tax avoidance) and enables us to employ an instrumental variable analysis. Additionally, to corroborate our findings, we investigate whether tax avoidance intensity has a differential effect on corporate innovation for those firms that are more financially constrained (as classified based upon the Kaplan and Zingales index, 1997). Moreover, we employ an event study methodology, to see how investors react to an unexpected announcement of a review of FIN48, a federal measure set in place by the US Treasury to make tax avoidance practices more costly, on May 20th, 2011. The unanticipated review announcement created uncertainty on a possible repeal of FIN48, i.e. a retraction of the measure.

In this study we focus on US firms for two main reasons. First of all, the introduction of an exogenous source of variation in tax avoidance for US firms allows us to asses a causal effect on corporate innovation. Secondly, the gap between tax receipts and corporate profits has been increasing in the US for several years, making it particularly interesting to study these firms.

To assess causality, and push our findings away from merely correlations to causation we apply an instrumental variable approach as in Desai and Dharmapala (2009) where we instrument our main measure of tax avoidance (the book tax gap) using instrumental variables based upon the introduction of Check-the-Box regulations in the US in 1997. The introduction of these regulation came with the unintended consequence of simplifying the creation of hybrid entities for US firms and hence making tax avoidance less costly. Following Desai (2009) we therefore use as instruments the interactions between a post Check-the-Box time indicator and three measures that proxy the availability of tax shields to a firm: The amounts of current and long term debt and the total net operating loss carry forwards respectively. The central idea behind the use of these instruments is that firms will engage more in tax avoidance after the introduction of the Check-the-Box regulations but that this increase will be smaller for those firms that have more tax shields available. Additionally we include in all our analyses firm and year fixed effects, to capture any unobserved heterogeneity at the firm and year level and we include a broad set of time varying firm characteristics, such as the Herfindahl Index, firms' leverage and sales growth to name a few.

Summarizing, our main results affirm the *opacity hypothesis*: Corporate tax avoidance intensity decreases innovation activity for US firms. Both future innovation input, as measured by the yearly amount of research and development expenditures as well as future innovation output in terms of quality (the number of forward patent citations per patent) and the number of patents decrease significantly when tax avoidance engagement, as measured by an increase in the book tax gap, increases.

Additionally we also find some weak evidence that the negative effect of tax avoidance on corporate innovation practices is partially offset for those firms that are more financially constrained. This finding suggests that for financially constrained firms also the *alleviated financial constraints hypothesis* matters.

We also look deeper into how the market reacts to the announcement of a review of FIN48 on May 20^{th} 2011. This regulation requires firms to only recognize an income tax benefit if it is very likely that it will be sustained in following years, implicitly making tax avoidance more costly. With the announcement of the review unexpected uncertainty was created on whether or not FIN48 would be repealed after the review. Hence, if the market would understand that tax avoidance comes with costs, we would expect to see a *negative* market response during the announcement of the FIN48 review. Our results indeed affirm this expectation. In the cross-section we find that the standardized CAR for a window of five days around (and including) the review announcement date is significantly more negative for firms that are weakly governed and for firms that are active in innovative industries. For the group of firms that are both weakly governed and active in innovative industries the negative market response is largest in magnitude.

Our findings are not only statistically significant but also very economically relevant: A one standard deviation increase in tax avoidance reduces future innovation input as measured by R&D expenditures by almost 60 percent. Moreover, a one standard deviation increase in tax avoidance decreases the number of future patents scaled by total assets and the number of citations per patents with around a factor of 3. Additionally, the negative market reaction to a possible retraction of FIN48, a measure that makes tax avoidance more costly, is also very economically relevant. The standardized CAR (cumulative abnormal return) is around 3 percentage points lower for firms that are weakly governed.

Our study fits into the existing literature that looks into the effect of tax avoidance on firm outcomes. Recent studies find evidence on a significant relationship between tax avoidance and firm outcomes such as firm value (Desai, 2009) asset utilization (Katz, 2013) and the cost of public debt (Shevlin, 2013). Katz (2013) finds that the main components of future profitability (margins, asset utilization) are negatively correlated with previous periods' tax avoidance intensity. In the same line, Shevlin (2013) documents that higher levels of tax avoidance is associated with higher bond offering yields as well as lower and more volatile future cash flows.

This paper is also related to Atanassov and Liu (2014), who find that a decrease in state level *tax rates* boosts corporate innovation practices. On a similar note, Mukherjee (2014) shows that tax rate increases hinder innovation practices. In contrast, this study sheds more light upon how changes in *tax avoidance* itself, affect corporate innovation. Our results suggest that firms indeed face a trade-off: More aggressive engagement in tax-avoidance may lead to higher after tax profits, but, as affirmed by the empirical results, this does come with a cost. Decreasing corporate transparency as a result of increasing tax avoidance engagement increases costs to pursue innovative projects, which in turn leads to lower innovation activity input and output.

Additionally, our study is also more broadly related to and extends the growing literature that looks into the (managerial) determinants of corporate innovation. For example, Chang

(2015) finds that non-executive employee stock options contribute to innovation output, as well as hostile take-overs (Atanassov, 2013) and institutional ownership (Aghion, 2013). On a different, although related note, Hsu (2013) finds that financial development also matters for technological innovation. Industries relying more on external finance and high tech technologies show significantly higher levels of innovation in those counties where equity markets are better developed.

The rest of the paper is organized as follows. Section II describes in more detail our measures of corporate innovation input and output, tax avoidance as well as the sample selection and control variables. Section III presents the main results on the effect of tax avoidance on innovation. Section IV explores more in depth through which channel tax avoidance can affect innovation. Section V concludes.

II. Tax Avoidance and Corporate Innovation

A. Sample Selection

Our sample starts with the universe of U.S. Compustat firms between 1990 and 2010. Patent data is taken from Kogan (2014) who provide a database with patent grants and filings up to 2010. This data is based upon information from the Google Patent database which captures innovation output at the firm level, hence for the US firms that are part of our dataset part of the innovation practices take place abroad in branches located abroad. We match this information to the CRSP database that includes firms' stock prices. This allows us to employ an event study and investigate whether an unexpected announcement regarding a possible review of FIN48 (federal regulations set in place to make tax avoidance more costly) is negatively valued by the market.

Additionally, we obtain governance data from Cain (2014), who have constructed a hand-collected take-over threat index at the firm level.

B. Measuring Tax Avoidance

We use the book-tax gap as our main measure of corporate tax avoidance intensity (following Desai and Dharmapala, 2009). The book-tax gap measures the difference between a firm's financial accounting income and the taxable income.[63] It basically measures the difference or gap between the (accounting) income that is generated by the firm in a given year and the income amount that is used for tax purposes (the firm pays taxes based upon this reported income amount), scaled by total assets. The lower the income amount for tax purposes is, the lower the amount of taxes a firm has to pay. A larger book tax gap hence indicates that the income reported in the financial statements is larger compared to the income reported for tax purposes and therefore indicates that a firm engages more aggressively in tax avoidance purposes. We define the book-tax gap as follows:

$$Booktaxgap_t = \frac{pidom_t - \left(\frac{txfed}{taxrate}\right)}{at_{t-1}}$$

Where $pidom_t$ is the pretax domestic income, $txfed$ current federal tax expenses, $txfed$ the federal tax rate and at the book value of total assets. Given that the divergence between a firm's book income and taxable income not necessarily completely can be attributed to tax sheltering but also can be caused by earnings management, we include as a control variable *total accruals* in all our regressions, to control for this possibility. The book-tax gap measures both permanent and temporary differences between financial statement income and taxable income. Some tax

[63] Optimally, we would like to observe the actual income reported for tax purposes. However, this information is confidential to the U.S. tax authorities and cannot be obtained from public filings. We follow the literature and infer taxable income based on taxes paid.

avoidance strategies which are reflected in this measure reduce net income through total tax expenses, however other tax avoidance strategies do not.[64]

C. Measuring Innovation

There are several ways in which innovation can be measured. To capture innovation input we focus on research and development expenditures. We measure innovation output based on patents and patent citations thereby following the vast majority of papers in the innovation literature and measure innovation that have already used the number of patent citations scaled by total patents for a firm in a given year. Looking at patents is particularly suitable as they capture technological change embodied in new products. Moreover, it is acknowledged that citations may better reflect the economic and technical impact of innovation and its social value (Hall, 2005; Atanassov, 2013). It is for instance known that those patents with more citations have a larger economic value (Harhoff, 1999). To better capture the timing of innovation we count the number of patent citations at the time of application and look at patent filings as opposed to granted patents. Hence, we follow the literature and use the natural logarithm of the number of forward citations by patents as our main measure of innovation quality. We measure patent and patent citations at the time of the application of the patent. We then look at future innovation measures using leads of 1, 2 and 3 years to reflect that there is usually a substantial time lag until innovation effort results in applications for patents. Further, we also measure innovation output by looking at the number of patents scaled by total assets.

A possible drawback of using patent activity as measure of corporate innovation is that in some industries patenting is more pronounced compared to other industries and also the so-

[64] For robustness, we, in unreported regressions, employ other measures of tax avoidance as well. We use three additional measures of tax avoidance suggested by the literature to check for robustness: the cash effective tax rate, dtax (the permanent book tax gap) and a slightly amended version of the book tax gap as used in Manzon and Plesko (2002).

called innovation 'life-cycle' in terms of innovation propensities and durations may differ considerably between industries. To alleviate this concern we feature firm fixed effects in all empirical analyses.

D. Control Variables

To control for possible other factors that may affect a firm's corporate innovation input and output levels we include a set of control variables. Firstly, we include a broad set of firm specific time varying characteristics, comprising of: total assets (its natural logarithm), its long-term debt (scaled by total assets), current debt (scaled by total assets), capital expenditures (*capex*) scaled by total assets; sales growth; *Tobin's Q*, which is calculated as total assets plus the value of shares outstanding at the end of the fiscal year minus common equity (scaled by total assets); *net operating losses*, *Tangibility* and *deferred taxes*. To capture possible earnings management practices, we compute *total accruals* as in Collins (1999). Additionally, we compute a Herfindahl index based upon yearly sales at the 4 digit SIC code level to capture competition intensity within industries. We also include the yearly Return on Assets (*RoA*) defined as in our set of control variables. Furthermore, to capture any unobserved, time-invariant heterogeneity at the firm and year level, we correspondingly introduce firm and year fixed effects in our analyses. A summary and corresponding description of the variables included in the empirical analyses can be found in Table I.

To see whether tax avoidance has a differential impact on innovation for firms that are financially constrained (the results are presented in Section IV) we compute the *KZ index* (based on Kaplan and Zingales, 1997) for every firm and year.[65]

Subsequently we create a dummy variable '*Financially constraint*', that takes on the value of 1 if a firm falls in the top tercile of the KZ index in a given year and 0 otherwise.

Summary statistics for our measures of tax avoidance and innovation as well as for the set of control variables are shown in Table II. Looking at our tax avoidance measure, our main independent variable of interest, we see that firms tend to have a negative book – tax gap, which is slight different from 0, as its mean is -0.039. This finding is in line with Desai and Dharmapala (2009). Firms' sales growth is on average positive with a 29 percent mean, although those firms in the 10th percentile make losses. Additionally, most firms in the sample have high growth opportunities, the mean Tobin's Q is more than 4.

[Tables I and II around here]

III. Main Results

A. OLS Results

---

[65] We calculate the KZ index as follows:

$$1.0019\times\frac{ib_t+dp_t}{ppent_{t-1}}+0.2826\times\frac{at_t+prccf_t csho_t-ceq_t-txdb_t}{at_t}+3.1391\times\frac{dltt_t+dlc_t}{dltt_t+dlc_t+seq_t}$$
$$-39.3678\times\frac{dvc_t+dvp_t}{ppent_{t-1}}-1.3147\times\frac{che_t}{ppent_{t-1}}$$

Where *ib* is income before extraordinary items, *dp* is depreciation and amortization, *txdb* is deferred taxes, *dltt* is long-term debt, *dlc* is debt in current liabilities, *dvc* denotes common dividends, *dvp* is preferred dividends, *ppent* is net property, plant and equipment and *che* is cash and short-term investment.

In a first step we estimate the following OLS model to gauge whether there is a correlation between our tax avoidance measure and innovation input and output:

$$Innovation_{i,t+n} = \beta_0 + \beta_1 . TaxAvoidance_{i,t} + \beta_2 . X_{i,t} + .y_t + .\delta_i + \epsilon_{i,t}$$

Our main variable of interest, $TaxAvoidance_{i,t}$ is measured by the book-tax gap, whereas our innovation measures are the yearly amount spent on research and development, the total number of patents filed in a given year scaled by the book value of total assets and the log of the *Number of cites per patent* (scaled by patents) in year $t+n$, where n ranges from 1 to 3.[66] The estimation results are presented in Table III. According to the alleviated financial constraints hypothesis the coefficient β1 is expected to be positive, indicating that there is a positive correlation between our measures of tax avoidance and innovation input, as an increase in tax avoidance intensity may increase resources available for R&D expenditures for more complex investments like innovating activities. A negative coefficient β1 is in line with the opacity explanation. We also control for a vector of several time varying firm controls, $X_{i,t}$, containing amongst other the log of Total Assets, Leverage, Capital Expenditures and Tobin's Q as a measure of a firm's growth opportunities. To account for omitted variables because of unobserved heterogeneity at the firm and year level, we include both firm and year fixed effects. Time fixed effects are denoted as $y_t$and time fixed effects as δ_i.

[Table III around here]

[66] In addition we use three alternative measures for *tax avoidance*: The cash effective tax rate, the permanent book tax gap and a slightly amended version of the book tax gap according as used in Manzon and Plesko (2002).The results are for all three measures both quantitatively and qualitatively similar. For our innovation measure we also use *innovation efficiency*, defined as innovation per R&D expenditure. Also for *innovation efficiency* the results are both quantitatively and qualitatively similar as compared to the log cites per patent specifications. Due to space limitations we only report results on our *book-tax gap* measure for tax avoidance.

The results in Table III indicate that there is a positive correlation between tax avoidance and innovation, except for column 1. In column 1, it is documented that there is suggestive evidence of a negative correlation between tax avoidance and innovation, although not statistically significant. In order to make causal statements however the estimations presented in Table III are likely to be biased, as they do not account for possible endogeneity. The issue of reverse causality could bias the results as highly innovative firms, i.e. firms with high patent activity and more innovation input, are better able to shelter income through the creation of subjective assets, i.e. assets that are difficult to value and for which a great degree of subjectivity is needed in the valuation procedure. Another potential problem arises from omitted variables. Our measure(s) of tax avoidance may be correlated with some unobservable characteristics that also affect innovation at the same time. In order to address these possible endogeneity issues we rely on an instrumental variable approach, by exploiting the differential impact of an exogenous shock to levels of tax avoidance (see Desai and Dharmapala 2009): the Check the Box regulation, introduced in 1996 by the U.S. Treasury. The following section will explain in more detail the use of the Check the Box regulation as an instrumental variable for our measure(s) of tax avoidance in order to tackle endogeneity problems.

B. Instrumental Variable Analysis: Exploiting Check-in-the-box Regulation

Entity classification before 1996 was extremely complicated and cumbersome. In order to make life easier for smaller firms to choose their organizational form for tax purposes, thereby relieving them from the many administrative burdens, the Check-the-box regulation was introduced in late 1996. However, besides the main objective of this particular regulation, making entity classification easier for smaller firms, it had an unanticipated side effect: It lowered the cost of tax avoidance (Desai and Dharmapala, 2009). The introduction of the CTB made it possible to create hybrid entities (Altshuler & Grubert, 2005), treated one way in a foreign jurisdiction and another by the United States, thereby facilitating tax avoidance. These hybrid entities are generally treated as separately incorporated subsidiaries under the tax rules of one country and as unincorporated branches under the tax rules of another country (Desai and Dharmapala, 2009). Exploiting these hybrid entities can create possibilities to use disregarded loans to strip earnings out of high tax jurisdictions (like the United States) and relocate these profits to low or no tax countries (Scott, 2014).[67]

We exploit that the CTB provides an exogenous negative shock to the costs of tax avoidance. After the introduction of the CTB, due to this negative cost shock, more firms avoided taxes leading to larger book tax gaps. Further, following Desai and Dharmapala (2009) we construct several instrumental variables for our tax avoidance measure and exploit that the introduction of CTB has a differential impact on the firm depending on how attractive tax avoidance is for a given firm. More specific, our instruments, like Desai and Dharmapala (2009) consist of firm-year-level variables that capture the incentive to engage in tax avoidance interacted with a post-CTB indicator, a 0/1 indicator that equals “1” for all years from 1997 (incl.) onwards. The idea behind these IVs is that some firms have a greater propensity to engage in tax avoidance activities compared to others. However, for a given level of tax avoidance, it can be expected that after the introduction of the CTB, a firm will *increase* its level of tax avoidance post-CTB, *ceteris paribus,* if its availability of tax shields is low compared to a similar firm whom can rely on a larger availability of tax shields.

[67] The typical structure would involve setting up a disregarded entity in a low tax country making a loan to a subsidiary in a foreign jurisdiction. For United States tax purposes both the loan and interest payments are not recognized. In the foreign country, as the tax haven will be seen as a corporation, the interest expenses can be deducted for tax purposes. This will effectively reduce taxes paid in the foreign jurisdiction and the U.S. subpart F income (Scott, 2014).

While capturing this propensity to engage in tax avoidance with certain variables and interacting them with a post-CTB indicator, should allow picking up any effect of a possible increase in tax avoidance caused by the introduction of the CTB given certain pre-CTB levels of tax avoidance. As variables that determine the incentive to engage in tax avoidance we rely on those selected by Desai and Dharmapala (2009), whom indicate that crucial determinant of the incentives to engage in tax avoidance is the availability of tax shields. Tax avoidance and the availability of tax shields can be seen as substitutes (see Graham and Tucker, 2006). Should a firm have many tax shields available (such as tax deductions from other sources, like interest deductions or net operating loss (NOL) carry forwards resulting from losses in previous years) a firm is *less likely* to engage in tax avoidance. The selected tax shield variables are:

- *NOL carry forwards*
- *Long term debt*
- *Debt in current liabilities*

The instruments for tax avoidance involve interacting a dummy variable for the post-CTB time period (all years after 1996) with each of these tax shield variables.

Using a two stages least square Instrumental Variable approach, in the first stage regression we regress our measure of tax avoidance on our instruments while conditioning upon several time varying firm level controls as well as year and firm fixed effects. From Desai and Dharmapala (2009) it can be expected that there is a *negative* relationship between our instruments and our measure(s) of tax avoidance (relevance condition). As firms whom can exploit many tax shields will have lower levels of tax avoidance compared to firms with few tax shields, since both can be seen as substitutes. In addition as the CTB was introduced with the objective to make entity formation less cumbersome and not to affect levels of innovation it is not expected that there is a direct influence of our instruments on innovation other than through tax avoidance.

In the first stage regression we regress our measure of tax avoidance on our instruments and several control variables:

$$TaxAvoidance_{i,t} = \beta_0 + \beta_1 PostCTB_t \,.\, CurrentDebt_{i,t} + \beta_2 PostCTB_t \,.\, LTDebt_{i,t}$$

$$+ \beta_3 PostCTB_t \,.\, NOLs_{i,t} + \beta_4 \,.\, X_{i,t} + y_t + \delta_i + \epsilon_{i,t}$$

Then in the second stage we take our instrumented values for the Tax Avoidance and explain whether there is a causal relationship that goes from tax avoidance to both innovation input as well as output.

$$Innovation_{i,t+n} = \beta_0 + \beta_1 \,.\, \widehat{TA}_{i,t} + \beta_2 \,.\, X_{i,t} + y_t + \delta_i + \epsilon_{i,t}$$

Table IV presents the results on the first stage regression as well as the second stage results. The negative relationship between our instrumental variables and tax avoidance is confirmed, indicating that the availability of tax shields and tax avoidance are negatively related.

The joint F-statistic for our three instrumental variables equals 18.2, higher than the general cut off level of 10, for appropriate instruments. Then in the second stage we take our instrumented values for tax avoidance and regress our measures of corporate innovation on tax avoidance.

[Table IV around here]

Table IV presents the second stage results where innovation is regressed on our instrumented tax avoidance measure. In Table IV we show the results using innovation input as measured by R&D expenditures, and innovation output as measured by the number of patents scaled by total assets and the number of cites per patent using different horizons for each dependent variable of interest. Columns (1), (4) and (7) show the results using the dependent variable at t+1, Columns (2), (5) and (8) at t+2 and Columns (3), (6) and (9) at t+3. In all specifications results are in line with the opacity hypothesis, where an *increase* in tax avoidance should *reduce* levels of innovation, as due to tax avoidance firms become more complex which comes with costs at the expense of long term investments and innovation. These results are both statistically as well as economically significant: a one standard deviation increase in tax avoidance leads to a 60 percent decrease in the future amount spent on research and development and decrease the amount of future patents and citations by a factor of 3.

The results remain robust when including other measures for tax avoidance and innovation and are in line with the opacity theory explanation as opposed to the alleviated financial constraints hypothesis.[68]

IV. Exploring Channels

A. Financial Constraints

In addition to the above presented results we explore in a further stage whether tax avoidance intensity has a differential effect on corporate innovation for those firms that are more financially constrained (as classified based upon the Kaplan and Zingales- 1997 index).

We look at the cross-sectional differences between firms by including a financial constraint measure. As indicated by the alleviated financial constraints hypothesis, it should be particularly beneficial for a firm that is financially constraint to engage in tax avoiding activities. It will then be able to increase internal cash levels via tax avoidance, thereby creating more room to finance innovative investment projects. It can therefore be expected when including a measure for being financial constraint and an interaction term *Financially constraint* tax avoidance* can capture the effect of any cross sectional difference of the effect

[68] In unreported robustness checks, we include indicators of state-level tax changes as time-varying control variables as independent variables. Our main results are qualitatively and quantitatively unchanged. In addition, our results remain robust to restricting our sample to firms which have at least one patent over the entire sample.

of tax avoidance on innovation for financially constrained firms versus not financially constrained firms. Table V presents the results. In addition to including our *Financially constraint* measure and the interaction term *Book Tax Gap * Financially Constraint* we include our benchmark set of time varying firm level control variables, as well as year and firm fixed effects. As before, we instrument our measure of tax avoidance, the book-tax gap and the interaction term *Book Tax Gap * Financially Constraint* with the instruments presented before.

[Table V around here]

Column (1) to (3) show the results for different time horizons *n* of our innovation input measure, where *n* ranges from 1 to 3 years. The results are robust for various *n* and indicate that, *ceteris paribus*, financially constrained firms on average spend more on research and development in the following years than those who are not constrained. In addition the *negative* effect of tax avoidance on innovation input is partly mitigated, i.e. *less severe*, for financially constrained firms as the interaction term *Book Tax Gap * Financially Constraint* is negative throughout different horizons and also statistically significant for a horizon of t+2. This result is in line with the alleviated financial constraints hypothesis: Financially constrained firms that want to innovate more benefit (partly) from increasing levels of tax avoidance as this will increase their internal cash flows, making it possible to finance innovative ventures. However, as the combined effect of *Book Tax Gap* and the interaction term *Book Tax Gap * Financially Constraint* is *negative*, the results indicate that firms that are financially constrained still suffer from opacity issues, but to a lesser extent compared to non-financially constraint firms.[69] The results for innovation output, as presented in Columns (4) – (9) confirm these findings: An

[69] We confirm these results in unreported robustness checks if we use firm size instead of the KZ index to capture financial constraints.

increase in tax avoidance practices decreases innovation output in subsequent years significantly, but this negative effect is again partially offset for those firms that are financially constrained.

B. Market Valuation: Reassessment of the Introduction of FIN48

In July 2006, the Financial Accounting Standards Board (FASB) released the interpretation No. 48, “Accounting for Uncertainty in Income Taxes” (Fin 48). It requires businesses to disclose reserves for uncertain tax benefits and standardizes their accounting (Erman, 2015).

In essence, the introduction of this regulation enforced firms to report profits more transparent for tax purposes, implicitly making tax avoidance *more* costly. On May 20th 2011, a special committee announced that FIN48 was going to be reviewed and possibly was going to be repealed. As this announcement came sudden and unexpected, it offers a perfect setting to study the market reaction towards a possible *decrease* in the cost of tax avoidance. As tax avoidance might become cheaper firms may have an incentive to engage in it more aggressively.

If, in line with the opacity theory, market participants then believe that benefits of tax avoidance activities do not outweigh the informational costs because of increased opacity it comes with, we would expect a *negative* reaction to the announcement of a possible repeal of FIN48. This negative effect should be particular pronounced for firms with weak governance, as for these firms’ agency problems as a result of increased opacity are more likely to be severe and managers are more easily able to divert resources into their own pockets. Our results indeed confirm this hypothesis.

[Table VI around here]

In column (1) of Table 12 we investigate the determinants of the market reaction towards the announcement of the review of FIN48, by looking at which factors explain abnormal share price reactions in an event window of (0,5) around the event announcement. Results indicate that the share prices of weakly governed firms react, as expected, negatively to the announcement of the FIN48 review, indicating that shareholders of weakly governed firms negatively value a decrease in the cost of tax avoidance. This is in line with the opacity explanation: Shareholders of weakly governed firms expect managers to engage more aggressively in tax avoiding activities, making the structure of the firm more complex, allowing managers to easily divert resources at the expense of shareholders.

Our findings are not only statistically significant but also very economically relevant: The negative market reaction to a possible retraction of FIN48, a measure that makes tax avoidance more costly, is also very economically relevant. In the cross-section, we find that the standardized CAR (cumulative abnormal return) is almost 3 percentage points lower for weakly governed firms.

V. Conclusions

We empirically examine the effect of tax avoidance engagement of US multinationals on their innovation practices, as measured both by innovation input (future R&D expenditures) and output (future patent filings and citations). On one hand this relationship can be positive as tax avoidance increases internal cash flows and therefore resolves possible financing constraints a firm may have. According to an opposing view, which we label the 'opacity hypothesis', greater tax avoidance engagement should lead to a negative effect on future firms innovation in- and output, as it decreases corporate transparency which comes with costs, for example in the form of worse lending conditions. To asses causality we employ an instrumental variable analysis.

For this, we exploit exogenous variation in tax avoidance introduced by the Check-the-Box regulations in 1996 in the US. Our empirical findings support the latter hypothesis. We find that a greater reliance of firms on tax avoidance decreases their future innovation in- and output, as measured by R&D expenditures and patent quantity (number of patents) and quality (number of patent citations). Additionally, we find that the market responds negatively to an announcement that may decrease the costs of tax avoidance. This negative market reaction is more pronounced for firms that are weakly governed and are present in innovative industries, suggesting that the market incorporates the negative effects of tax avoidance in their valuations. Our results show that tax avoidance is not just about the redistribution of cash flows but carries real costs, which eventually contribute to hampering economic growth.

References

Atanassov, J., 2013, "Does the Threat of Hostile Takeovers Stifle Innovation", *Journal of Finance* 68, 1097-1133.

Atanassov, J., Li, 2014, "Corporate Income Taxes, Financial Constraints and Innovation", Working Paper, University of Oregon.

Balakrishnan, K., Blouin, J., Guay, W., 2012, "Does Tax Aggressiveness Reduce Corporate Transparency?", Working Paper.

Becker, B., Jacob, M., Jacob, M., 2013, "Payout Taxes and the Allocation of Investment", *Journal of Financial Economics* 107, 1-24.

Cain, D., S. McKeon, and S. Solomon, 2014, "Do Takeover Laws Matter? Evidence from 5 decades of Hostile Takeovers", Working paper.

Cornaggia, J., Mao, Y., Wolfe, B. and Tian, X, 2015, "Does Banking Competition Affect Innovation?", *Journal of Financial Economics* 115, 189-209.

Desai, M., Dharmapala, D., 2009, "Corporate Tax Avoidance and Firm Value", *Review of Economics and Statistics* 91, 537–546.

Frank, L., Lynch, L., Rego, S., 2009, "Tax Reporting Aggressiveness and its Relation to Aggressive Financial Reporting", *The Accounting Review* 84, 467–496.

Gallemore, J., Labro, E., 2015, "The Importance of the Internal Information Environment for Tax Avoidance", *Journal of Accounting and Economics* 60, 149–167.

Graham, J., Tucker, A., 2006, "Tax Shelters and Corporate Debt Policy", *Journal of Financial Economics* 81, 563–594.

Hall, B., Jaffe, A., Trajtenberg, M., 2005, "Market Value and Patent Citations", *RAND Journal of Economics* 32, 101-128.

Hasan, I., Hoi, C. K. S., Wu, Q., and Zhang, H., 2014, "Beauty is in the eye of the beholder: The effect of corporate tax avoidance on the cost of bank loans", *Journal of Financial Economics* 113, 109-130.

Kaplan, S. N., Zingales, L., 1997, "Do Investment-Cash Flow Sensitivities Provide Useful Measures of Financing Constraints", *Quarterly Journal of Economics* 112, 169–215.

Katz, S., Khan, U., Schmidt, A.P., 2015, "Tax Avoidance and DuPont Measures of Future Performance", Working Paper, Columbia Business School and Poole College of Management, North Carolina State University.

Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2014, "Technological Innovation, Resource Allocation, and Growth", Working Paper.

Manzon, G., Plesko, G., 2002, "The relation between financial and tax reporting measures of income", *Tax Law Review* 55, 175–214.

Mukherjee, A., Singh, M., Zaldokas, A, 2015, "Do Corporate Taxes Hinder Innovation?", Working Paper.

Popp, D., Juhl, T., Johnson, D.K.N., 2004, "Time In Purgatory: Determinants of the Grant Lag for U.S. Patent Applications", *Topics in Economic Analysis & Policy* 4.

Rego, S., Wilson, R., 2012, "Equity Risk Incentives and Corporate Tax Aggressiveness", *Journal of Accounting Research* 50, 775–810.

Shevlin, T., Urcan, O., Vasvari, F., 2013, “Corporate Tax Avoidance and Public Debt Costs”, Working paper, University of California, Irvine.

Tables

TABLE I

VARIABLE NAMES, DEFINITIONS, AND DATA SOURCES FOR THE EMPIRICAL ANALYSIS OF THE REAL EFFECTS OF TAX AVOIDANCE

Variable Name	Variable Definition	*Source*
Dependent Variables		
R&D Expenditures	The yearly amount of Research and Development expenditures in thousands of US $ winsorized at the 1% level	*CS*
Number of Patents to Total Assets ratio	The ratio of the yearly number of patents filed divided by Total Assets	*CS*
Cites per Patent	The natural logarithm of the yearly number of Citations per Patent	*GP*
Main Independent Variables		
Book Tax Gap	The yearly difference between a firm's financial accounting income and taxable income as defined by Desai and Dharmapala (2009)	*CS*
Instrumental Variables		
Net Operating Loss Carryforwards	The yearly amount of net operating loss carryforwards in thousands of US $ winsorized at the 1% level	*CS*
Long Term Debt	The yearly total amount of long term debt in thousands of US $ scaled by total assets winsorized at the 1% level	*CS*
Current Debt	The total amount of debt in current liabilities in thousands of US $ scaled by total assets in a given year winsorized at the 1% level	*CS*
Control Variables		
Firm Characteristics		
Total Assets	The logarithm of one plus total assets, which is the sum of current assets plus net property, plant, and equipment plus other noncurrent assets (including intangible assets, deferred items, and investments and advances)	*CS*
Capital Expenditures (Capex)	The yearly amount of funds used for additions to property, plant, and equipment, excluding amounts arising from acquisitions in thousands of US $ scaled by total assets winsorized at the 1% level	*CS*
Sales Growth	The amount of sales in thousands of US $ (item 12) at t corrected for the amount of sales at t-1 divided by the amount of sales at t-1	
Tobin's Q	The market capitalization of equity adjusted for the amount of common equity including total assets divided by the book value of total assets	*CS*
Tangibility	1 minus the amount of intangible assets divided by the amount of total assets in a given year	*CS*
Deferred taxes	The amount of deferred taxes scaled by total assets in a given year	*CS*
Total Accruals	The difference between the amount of total current assets at time t and the amount of total current assets at t-1 minus the differences between current liabilities at t and t-1 and cash and cash equivalents at t and t-1 plus the difference between debt in current liabilities at t and t-1 minus the total amount of depreciation and amortization at time t	*CS*
Herfindahl Index	The yearly sum of the squared market shares (based upon annual sales) for all firms in a certain industry (4-digit SIC code)	*CS*
Return on Assets (ROA)	Return on Assets, i.e., the amount of net profit divided by total assets winsorized at the 1% level	*CS*

NOTES. The table defines the variables used in the empirical analysis of the real effects of tax avoidance, as well as the corresponding data sources used. *t-1* indicates a one year lag is used in the calculations of the corresponding variable. Data sources include: CS = Compustat; *GP* = Google Patents Database as obtained from Kogan (2014).

TABLE II

DESCRIPTIVE STATISTICS FOR THE EMPIRICAL ANALYSIS OF TAX AVOIDANCE ON INNOVATION OUTCOMES

Variable Name	Number of Observations	Mean	Standard Deviation	10%	Median (50%)	90%
Dependent Variables						
R&D Expenditures	161430	0.063	0.14	0.000	0.000	0.197
Number of Patents to Total Assets ratio	150307	0.017	2.58	0.000	0.000	0.004
Cites per Patent	161430	0.250	0.80	0.000	0.000	0.693
Main Independent Variable						
Book Tax Gap	134304	-0.039	0.11	-0.150	0.000	0.029
Instrumental Variables						
Net Operating Loss Carryforwards	150307	1.071	4.81	0.000	0.000	1.420
Long Term Debt	150307	0.204	0.27	0.000	0.113	0.519
Current Debt	150307	0.128	0.45	0.000	0.017	0.231
Control Variables						
Firm Characteristics						
Total Assets	150926	1424.960	4373.96	3.353	102.298	3004.936
Capital Expenditures (Capex)	148043	0.061	0.07	0.005	0.037	0.139
Sales Growth	128415	0.293	1.13	-0.236	0.078	0.701
Tobin's Q	126776	4.028	11.39	0.896	1.555	5.514
Tangibility	150307	0.284	0.25	0.027	0.199	0.710
Deferred taxes	150307	-0.001	0.02	-0.012	0.000	0.012
Total Accruals	134956	-0.064	0.31	-0.204	-0.043	0.097
Herfindahl Index	161430	0.247	0.19	0.075	0.188	0.499
Return on Assets (ROA)	150307	-0.173	1.09	-0.511	0.089	0.222

NOTES. The table provides the number of observations, mean, standard deviation, 10th percentile, the median (50th percentile) and the 90th percentile of all variables used in the empirical analysis.

TABLE III

OLS SPECIFICATIONS: THE EFFECT OF TAX AVOIDANCE ON INNOVATION OUTCOMES

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	*R & D Expenditures*			*Number of Patents to Total Assets*			*Patent Citations*		
Book Tax Gap	-0.006	0.002	0.001	0.007**	0.007**	0.007*	0.038	0.057	0.078*
	(-1.17)	(0.34)	(0.11)	(2.45)	(2.34)	(1.67)	(0.95)	(1.36)	(1.85)
Net Operating Loss Carryforwards	-0.000	-0.000	0.000	0.000	0.001	-0.002	0.000	0.001	0.002*
	(-0.56)	(-0.24)	(0.13)	(0.28)	(0.60)	(-1.03)	(0.25)	(1.13)	(1.81)
Long Term Debt	-0.001	0.002	0.001	0.002	-0.007	-0.001	-0.094***	-0.075***	-0.082***
	(-0.13)	(0.51)	(0.20)	(0.26)	(-1.07)	(-0.20)	(-4.67)	(-3.37)	(-3.52)
Current Debt	-0.011***	-0.012***	-0.012**	-0.001	0.004	0.009	-0.004	0.010	-0.002
	(-2.74)	(-2.67)	(-2.57)	(-0.29)	(0.70)	(0.60)	(-0.37)	(1.00)	(-0.18)
Total Assets	-0.008***	-0.003***	-0.002*	-0.002***	0.000	-0.002	0.028***	0.010	-0.006
	(-6.78)	(-2.93)	(-1.81)	(-3.01)	(0.08)	(-0.66)	(4.21)	(1.29)	(-0.82)
Capital Expenditures (Capex)	0.025***	0.010	-0.009	0.002	-0.002	0.009	0.042	0.114*	0.122*
	(3.55)	(1.34)	(-1.24)	(0.24)	(-0.19)	(0.55)	(0.74)	(1.89)	(1.89)
Sales Growth	-0.001	0.001	0.000	-0.001	-0.000	-0.001	0.001	0.008**	0.004
	(-1.06)	(0.98)	(0.66)	(-1.53)	(-0.23)	(-0.86)	(0.25)	(2.17)	(1.35)
Tobin's Q	-0.001***	-0.000	0.000	-0.000	-0.000	0.000	0.004***	0.003***	0.002***
	(-3.08)	(-0.20)	(0.65)	(-1.24)	(-0.36)	(0.11)	(4.63)	(3.90)	(2.65)
Tangibility	0.015***	-0.005	-0.003	-0.015	0.008	0.009	0.072**	0.059*	0.085**
	(2.73)	(-0.89)	(-0.54)	(-0.85)	(0.52)	(0.50)	(2.14)	(1.65)	(2.21)
Deferred taxes	0.064***	0.038**	0.009	0.021**	0.014*	0.006	-0.312	-0.410**	-0.279
	(4.14)	(2.19)	(0.49)	(2.04)	(1.65)	(0.55)	(-1.61)	(-2.22)	(-1.53)
Total Accruals	-0.004	-0.003	-0.004	-0.009	-0.009	0.015	-0.016	0.004	0.006
	(-1.48)	(-1.36)	(-1.34)	(-1.07)	(-1.29)	(0.95)	(-1.26)	(0.29)	(0.47)
Herfindahl Index	0.006	0.006	0.006	0.003	0.005	0.005	0.121**	0.107*	0.056
	(0.88)	(0.83)	(0.82)	(0.47)	(0.71)	(0.66)	(2.19)	(1.81)	(0.86)
Return on Assets (ROA)	-0.011***	-0.000	0.003	-0.001	0.004	-0.002	0.006	0.016**	0.010
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	69684	64260	59198	69438	63896	58793	69698	64270	59213

NOTES. All Models are estimated with a linear regression model (OLS). The definition of the variables can be found in Table I. "Yes" indicates that the set of fixed effects is included. Standard errors are clustered at the firm level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. In Model (1), (4) and (7) the dependent variable is measured at t+1, in Model (2), (5) and (8) at t+2 and Model (3), (6) and (9) at t+3 (a lead of 1,2 and 3 years respectively).

TABLE IV
MAIN SPECIFICATIONS OF THE EFFECT OF TAX AVOIDANCE ON INNOVATION OUTCOMES - INSTRUMENTED

Model	First Stage	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	*Book Tax Gap*	*R & D Expenditures*			*Number of Patents to Total Assets*			*Patent Citations*		
Book Tax Gap	--	-0.0169	-0.257**	-0.324**	-0.468**	-0.475**	-0.522**	-2.260**	-4.762***	-6.608***
	--	(0.881)	(0.038)	(0.015)	(0.021)	(0.040)	(0.046)	(-2.09)	(-4.16)	(-5.34)
Net Operating Loss Carryforwards	-0.001	-0.000302	-0.000231	-0.0000412	0.000154	0.000628	-0.00220	0.001	0.001	0.002
	(-0.81)	(0.575)	(0.698)	(0.944)	(0.885)	(0.617)	(0.277)	(0.31)	(0.52)	(1.04)
Long Term Debt	0.043***	-0.000412	0.00620	0.00575	0.00842	-0.00000653	0.00640	-0.097***	-0.045	-0.052
	(7.39)	(0.934)	(0.273)	(0.342)	(0.351)	(0.999)	(0.527)	(-3.06)	(-1.26)	(-1.31)
Current Debt	0.016***	-0.0108***	-0.0125***	-0.0125***	-0.00223	0.00323	0.00786	-0.018	-0.008	-0.022
	(3.13)	(0.006)	(0.006)	(0.007)	(0.648)	(0.546)	(0.578)	(-1.15)	(-0.55)	(-1.53)
Total Assets	0.002*	-0.00776***	-0.00273**	-0.00123	-0.000970	0.00154	-0.000194	0.087***	0.064***	0.043***
	(1.83)	(0.000)	(0.032)	(0.335)	(0.399)	(0.331)	(0.923)	(8.52)	(5.89)	(3.84)
Capital Expenditures (Capex)	-0.115***	0.0236	-0.0197	-0.0461***	-0.0515***	-0.0568	-0.0508	-0.166	-0.350**	-0.603***
	(-10.89)	(0.109)	(0.213)	(0.007)	(0.006)	(0.100)	(0.219)	(-1.17)	(-2.29)	(-3.64)
Sales Growth	-0.005***	-0.000595	-0.000872	-0.00137	-0.00337***	-0.00280*	-0.00345*	-0.013*	-0.018**	-0.034***
	(-9.43)	(0.464)	(0.297)	(0.125)	(0.008)	(0.052)	(0.070)	(-1.77)	(-2.21)	(-4.15)
Tobin's Q	-0.001***	-0.000734***	-0.000407	-0.000310	-0.000846***	-0.000737**	-0.000718	0.003	-0.000	-0.005**
	(-8.39)	(0.010)	(0.158)	(0.252)	(0.005)	(0.024)	(0.102)	(1.52)	(-0.21)	(-2.16)
Tangibility	0.053***	0.0153**	0.00865	0.0140	0.0103	0.0341	0.0373	0.185**	0.326***	0.466***
	(9.14)	(0.046)	(0.314)	(0.147)	(0.356)	(0.208)	(0.222)	(2.48)	(4.28)	(5.49)
Deferred taxes	0.779***	0.0721	0.238**	0.260**	0.389**	0.388**	0.415**	1.657*	3.430***	4.767***
	(19.19)	(0.419)	(0.015)	(0.014)	(0.015)	(0.029)	(0.040)	(1.86)	(3.68)	(4.73)
Total Accruals	-0.007**	-0.00394	-0.00514*	-0.00581**	-0.0123	-0.0122	0.0118	-0.045**	-0.031	-0.036*
	(-2.07)	(0.152)	(0.058)	(0.048)	(0.203)	(0.143)	(0.415)	(-2.33)	(-1.55)	(-1.85)
Herfindahl Index	0.012	0.00596	0.00872	0.00963	0.00812	0.00946*	0.0102*	0.204**	0.231***	0.188**
	(1.41)	(0.393)	(0.255)	(0.236)	(0.194)	(0.092)	(0.100)	(2.57)	(2.78)	(2.14)
Return on Assets (ROA)	-0.014***	-0.0116***	-0.00369	-0.00129	-0.00725***	-0.00169	-0.00830**	-0.031*	-0.045***	-0.074***
	(-8.71)	(0.000)	(0.245)	(0.670)	(0.001)	(0.488)	(0.043)	(-1.92)	(-2.59)	(-4.23)
Post CTB x NOLs	0.000	--	--	--	--	--	--	--	--	--
	(0.15)	--	--	--	--	--	--	--	--	--
Post CTB x long-term debt	-0.040***	--	--	--	--	--	--	--	--	--
	(-6.75)	--	--	--	--	--	--	--	--	--
Post CTB x current debt	-0.018***	--	--	--	--	--	--	--	--	--
	(-3.47)	--	--	--	--	--	--	--	--	--
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	69698	69684	64260	59198	69438	63896	58793	69698	64270	59213

NOTES. All models are estimated with a 2SLS IV model. The first column contains the results of the first stage regression. The F-statistic for the joint significance of the instruments is 18.2 and statistically significant at the 1% level. The Book Tax Gap is instrumented with the yearly amounts of Net Operating Loss Carryforwards, Long Term Debt and Current Debt interacted with a post check the box regulation introduction dummy which equals 1 from 1997 onwards. The definition of the variables can be found in Table I. "Yes" indicates that the set of fixed effects is included. Standard errors are clustered at the firm level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. In Model (1), (4) and (7) the dependent variable is measured at t+1, in Model (2), (5) and (8) at t+2 and Model (3), (6) and (9) at t+3 (a lead of 1,2 and 3 years respectively).

TABLE V

MAIN SPECIFICATIONS OF THE EFFECT OF TAX AVOIDANCE ON INNOVATION OUTCOMES: FINANCIAL CONSTRAINTS - INSTRUMENTED

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	*R & D Expenditures*			*Number of Patents to Total Assets*			*Patent Citations*		
Book Tax Gap	-0.0359	-0.280**	-0.364***	-0.524**	-0.526*	-0.596*	-3.282***	-3.656***	-4.393***
	(0.754)	(0.024)	(0.009)	(0.033)	(0.062)	(0.068)	(-3.62)	(-3.72)	(-3.78)
Financially Constraint	-0.00104	0.00205	0.00335*	0.00384***	0.00596***	0.00751***	1.795***	2.514***	2.794***
	(0.466)	(0.180)	(0.058)	(0.001)	(0.000)	(0.000)	(5.61)	(7.33)	(7.46)
Book Tax Gap x Financially Constraint	0.0182	0.0680**	0.0506	-0.0196	0.0693**	0.166**	0.052***	0.083***	0.095***
	(0.580)	(0.038)	(0.172)	(0.722)	(0.046)	(0.037)	(3.62)	(5.33)	(5.56)
Net Operating Loss Carryforwards	-0.000294	-0.000198	-0.000110	-0.0000162	0.000565	-0.00277	0.042***	0.031***	0.023***
	(0.626)	(0.765)	(0.870)	(0.987)	(0.626)	(0.293)	(6.32)	(4.32)	(2.75)
Long Term Debt	0.000760	0.00615	0.00475	0.00749	0.000295	0.00639	-0.009	-0.018**	-0.024***
	(0.874)	(0.271)	(0.447)	(0.432)	(0.972)	(0.625)	(-1.33)	(-2.21)	(-2.63)
Current Debt	-0.0101**	-0.0135***	-0.0140***	-0.00230	0.00472	0.00991	-0.084	-0.011	-0.069
	(0.014)	(0.004)	(0.006)	(0.698)	(0.494)	(0.570)	(-0.97)	(-0.11)	(-0.60)
Total Assets	-0.00808***	-0.00283**	-0.00121	-0.000948	0.00108	-0.000450	-0.011**	-0.023***	-0.022***
	(0.000)	(0.028)	(0.345)	(0.440)	(0.438)	(0.842)	(-2.40)	(-4.25)	(-3.65)
Capital Expenditures (Capex)	0.0217	-0.0192	-0.0486***	-0.0594**	-0.0573	-0.0499	0.002**	0.003**	0.004***
	(0.146)	(0.222)	(0.005)	(0.018)	(0.171)	(0.271)	(2.22)	(2.46)	(2.64)
Sales Growth	-0.000608	-0.00113	-0.00173*	-0.00376**	-0.00304*	-0.00387*	-0.001	-0.000	0.000
	(0.462)	(0.187)	(0.058)	(0.017)	(0.083)	(0.094)	(-1.01)	(-0.22)	(0.26)
Tobin's Q	-0.000713**	-0.000358	-0.000321	-0.00102**	-0.000848**	-0.000814	0.167***	0.190***	0.209***
	(0.023)	(0.228)	(0.278)	(0.010)	(0.048)	(0.116)	(4.32)	(4.47)	(4.22)
Tangibility	0.0165**	0.00926	0.0158	0.0113	0.0335	0.0393	1.727***	1.728***	2.006***
	(0.032)	(0.286)	(0.105)	(0.268)	(0.273)	(0.273)	(2.92)	(2.66)	(2.66)
Deferred taxes	0.0811	0.241**	0.284**	0.441**	0.413*	0.435*	0.028**	0.036***	0.044***
	(0.374)	(0.014)	(0.010)	(0.033)	(0.065)	(0.067)	(2.14)	(2.60)	(2.88)
Total Accruals	-0.00360	-0.00423	-0.00506*	-0.0148	-0.0138	0.0122	-0.000	0.002	0.002
	(0.199)	(0.125)	(0.099)	(0.221)	(0.186)	(0.410)	(-0.19)	(0.55)	(0.50)
Herfindahl Index	0.00616	0.00916	0.00915	0.00777	0.00902	0.00834	0.089*	0.114**	0.179***
	(0.374)	(0.229)	(0.258)	(0.221)	(0.113)	(0.191)	(1.90)	(2.12)	(2.87)
Return on Assets (ROA)	-0.0128***	-0.00470	-0.00245	-0.00929***	-0.00116	-0.00981**	-0.021*	-0.015	-0.014
	(0.000)	(0.146)	(0.458)	(0.001)	(0.709)	(0.048)	(-1.88)	(-1.22)	(-0.97)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	69096	63714	58693	68867	63371	58305	66101	61152	56631

NOTES. All models are estimated with a 2SLS IV model. The Book Tax Gap is instrumented with the yearly amounts of Net Operating Loss Carryforwards, Long Term Debt and Current Debt interacted with a post check the box regulation introduction dummy which equals 1 from 1997 onwards. The definition of the variables can be found in Table I. Financially Constraint is a dummy variable that equals one if a firm belongs to the top tercile of firms that are most financially constrained based upon the Kaplan and Zingales financial constraints index. "Yes" indicates that the set of fixed effects is included. Standard errors are clustered at the firm level. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively. In Model (1), (4) and (7) the dependent variable is measured at t+1, in Model (2), (5) and (8) at t+2 and Model (3), (6) and (9) at t+3 (a lead of 1,2 and 3 years respectively).

TABLE VI
MARKET RESPONSE TO FIN48 WITHDRAWAL

Model	(1)	(2)	(3)
Dependent Variable		*Standardized CAR (0,5)*	
Weak Governance	-0.110**	-0.107**	-0.089
	(-2.30)	(-2.23)	(-1.48)
Mean cites per patent (SIC 2-level)	--	-2.215***	-1.834
	--	(-2.62)	(-1.57)
Weak governance * Mean cites per patent (SIC 2-level)	--	--	-2.653**
	--	--	(-2.28)
Net Operating Loss Carryforwards	0.029**	0.029**	0.029**
	(2.02)	(2.01)	(2.01)
Long Term Debt	-0.083	-0.107	-0.108
	(-0.72)	(-0.93)	(-0.94)
Current Debt	-0.593**	-0.593**	-0.591**
	(-2.20)	(-2.23)	(-2.21)
Total Assets	-0.003	-0.003	-0.003
	(-0.23)	(-0.27)	(-0.25)
Capital Expenditures (Capex)	1.164***	1.183***	1.185***
	(2.78)	(2.83)	(2.84)
Sales Growth	0.002	0.004	0.004
	(0.08)	(0.14)	(0.14)
Tobin's Q	-0.015	-0.016	-0.016
	(-1.42)	(-1.55)	(-1.53)
Tangibility	0.260**	0.206*	0.204*
	(2.33)	(1.81)	(1.80)
Deferred taxes	-1.160	-1.270	-1.260
	(-1.03)	(-1.12)	(-1.11)
Total Accruals	0.229	0.239	0.237
	(1.19)	(1.25)	(1.24)
Herfindahl Index	-0.144	-0.152	-0.150
	(-1.49)	(-1.57)	(-1.56)
Return on Assets (ROA)	-0.387***	-0.402***	-0.398***
	(-3.74)	(-3.87)	(-3.81)
Number of Observations	2636	2636	2636

NOTES. All models are estimated with alinear regression (OLS) model. The dependent variable is the standardized cumulative abnormal return during the time window of 0 to 5 days around the announcement of the FIN 48 withdrawal on weak governance, the mean number of cites per patents in the SIC 2-digit industry, the interaction between these two variables and a further set of control variables. Standard errors are heteroscedasticity-robust. The definition of the variables can be found in Table I. Standard errors are given in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level respectively.