NBER WORKING PAPER SERIES

RUGGED ENTREPRENEURS:
THE GEOGRAPHIC AND CULTURAL CONTOURS OF NEW BUSINESS FORMATION

John M. Barrios
Yael Hochberg
Daniele Macciocchi

Working Paper 28606
http://www.nber.org/papers/w28606

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2021

We are grateful to the Startup Cartography Project and Jorge Guzman for the provision of data used in this project. All errors are our own. Corresponding Author: John M Barrios (john.barrios@wustl.edu), Washington University in St Louis, Olin School of Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Rugged Entrepreneurs: The Geographic and Cultural Contours of New Business Formation
John M. Barrios, Yael Hochberg, and Daniele Macciocchi
NBER Working Paper No. 28606
March 2021
JEL No. L26,N3,N9,O1,O43

ABSTRACT

How do geographic and historical-cultural factors shape new business formation? Using novel data on new business registrations, we document that 75% of the variation in new business formation is explained by time-invariant county-level factors and examine the extent to which such variation is driven by historical, cultural, and geographic factors. Current-day new business formation is positively related to historical attributes that presage individualist culture: frontier experience and historical birthplace diversity, as well as the county's topographical features. The relation holds when we exploit plausibly exogenous variation in frontier experience driven by shocks to the settlement process that arise from historical immigration flows. Our study points to the fundamental role of geographic and historical-cultural features, especially rugged individualism, in explaining contemporary new business formation in the U.S.

John M. Barrios
Washington University in St. Louis
John.barrios@wustl.edu

Yael Hochberg
Jones Graduate School of Business
Rice University
6100 Main Street, MS-531
McNair Hall Room 331
Houston, TX 77005
and Massachusetts Institute of Technology
and also NBER
hochberg@rice.edu

Daniele Macciocchi
University of Miami
Coral Gables, FL
dmacciocchi@miami.edu

1. INTRODUCTION

Entrepreneurial activity has long been recognized as a primary contributor to economic growth (Davis and Haltiwanger 1992; Jovanovic and MacDonald 1994; Haltiwanger, Jarmin, and Miranda 2013; Guzman and Stern 2019). Entrepreneurs are often considered drivers of urban growth in general and innovation-driven growth in particular (Glaeser and Kerr 2009; Glaeser et al. 2010; Agrawal et al. 2010; Glaeser et al. 2015). Despite this, our understanding of the factors determining the geographic distribution of new business formation remains limited. In particular, we know little regarding how geographic and historical-cultural factors shape new business formation, despite their importance to other economic activities (Alesina and Giuliano 2015). This paper seeks to fill this gap.

An extensive literature in sociology and cultural anthropology hypothesizes that cultural factors can significantly influence entrepreneurship. This literature dates back as far as Weber (1904), who argued for a role for the Protestant work ethic in encouraging entrepreneurial vocations. In the economics literature, cultural factors have been shown to affect economic activity more broadly. For example, Bazzi et al. (2020a) document that U.S. counties with longer historical exposure to frontier conditions exhibit greater individualism and opposition to redistribution and regulation. In the context of a pandemic, Bazzi et al. (2020b) show that such individualism undermines collective action in the face of crisis. Similarly, variation in the historical immigrant makeup of U.S. counties has long-run effects on their economic prosperity (Sequeira et al., 2019). These effects may arise due to culture, genetics, or networks (Ottaviano and Peri, 2006; Ager and Bruckner, 2013; Grosjean, 2014; Burchardi Chaney, and Hassan, 2019; Bandiera et al., 2017; Ager and Bruckner, 2017).

Similarly, through its impact on historical development, geography and topography have affected economic activity and outcomes (Nunn and Puga, 2012). On the one hand, geographic ruggedness can be an economic handicap, making it more expensive to transport goods, farmland, or generally do business. On the other hand, geographic ruggedness has historically served as a physical impediment to lawlessness, raiding, and other types of conflict activity that may impede economic development. Moreover, the local terrain's nature may contribute to the cultural attributes that develop in that area.

In this paper, we consider three cultural and geographic factors that may shape the contours of entrepreneurial entry and new business formation in the U.S. First, we consider the frontier experience, which leads to the development of a more individualist culture in the local area (Bazzi et al. 2020a). Entrepreneurship is often associated in the public mind with individualism, a desire to "be one's own boss," or one's wish to have flexibility in the workplace. Consistent with this notion, Hurst and Pugsley (2011) document that roughly 50% of small business owners in the U.S. report that such non-pecuniary benefits were one of the primary reasons for starting a business. Second, we consider the county's historical population's birthplace diversity, which may lead to a diversity of viewpoints and thought as immigrants transmit their values into the local area (Giuliano and Tabellini, 2020), spurring creativity that can translate into new business formation. Finally, we consider the ruggedness of the geographic terrain of the local area. Irregular terrain makes cultivation difficult,[1] and transportation over irregular terrain is slower and costlier. In addition, because of the very high costs of earthwork, building costs are significantly higher when

[1] On steep slopes, erosion becomes a potential hazard, and the control of water, such as irrigation, becomes much more difficult. According to the Food and Agriculture Organization (1993), when slopes are greater than 2 degrees, the benefits of cultivation often do not cover the necessary costs, and when slopes are greater than 6 degrees, cultivation becomes impossible.

the terrain is irregular (Rapaport and Snickars, 1999). These factors can affect the nature and extent of new business formation.

We begin by motivating our analysis with an assessment of how new business formation clusters geographically. To do this, we utilize a relatively novel dataset of new business registrations in a local region, provided by the Startup Cartography Project (SCP). Because a new company must not only incorporate in a state jurisdiction (which may not be the state they operate in), but also register with their local Secretary of State to do business (where the business actually operates), and because such registrations provide an actual operating address for the new company, utilizing business registration data allows us to observe the full universe of newly incorporated businesses (Guzman and Stern 2020). The SCP dataset provides us with counts, by zip-code and quarter, of all new for-profit businesses, allowing us to observe entrepreneurial entry at the micro-level. The SCP data also provides us with an estimate of the quality of the entrepreneurial cohort in each county in each quarter, where quality is measured as the estimated probability that a new business in that cohort will achieve a high growth exit (Guzman and Stern 2020; Fazio et al. 2020).

When we estimate a negative binomial regression at the county level, we find a highly significant excess variation parameter, suggesting that number and quality of new business formation is clustered more in certain counties than we would expect merely by chance. We observe a similar pattern when we expand the unit of analysis to county-quarter-years and estimate various fixed effects regressions. Specifically, we find that county fixed effects are highly significant and explain approximately 75% of the new business registration variation over the entire sample period (1988-2016). These estimates suggest that time-invariant factors drive the prevalence of entrepreneurial entry in a local area. With these basic patterns established, we next examine the extent to which such variation is driven by historical, cultural, and geographic factors that remain constant over time.

We begin by analyzing the relationship between individualism, as proxied for by Total Frontier Experience (TFE) (Bazzi et al. 2020a). During the process of westward expansion in the early years following the formation of the United States, living in a frontier state or county particularly rewarded those who demonstrated independence and self-reliance, as pioneers had little infrastructure to rely on. Historians suggest frontier locations historically demonstrated greater individualism (Turner 1893). Recent work in economics suggests that such individualism has persisted in these locations in the long run. Consistent with entry into entrepreneurship being characterized by individualistic nature, in linear panel models with a variety of demographic and geographic controls, we observe a positive relation between TFE and new business formation per capita in U.S. counties, with a one standard deviation increase in TFE being associated with a 3% to 5% increase in per capita new business starts. Similarly, we observe a positive association between TFE and the entrepreneurial quality index of Guzman and Stern (2020). The magnitude of this effect, however, is small: a one standard deviation increase in TFE associated with 1.25% increase in the predicted probability of high-growth exit in the current-day cohort. The small economic magnitude suggests that while the individualism brought about by TFE leads to increased entrepreneurial activity, it does not appear to meaningfully change the type of businesses started (small business versus innovation-driven high-growth startups).

Of course, TFE may be correlated with economic development and other cultural and demographic factors in those areas, which likely affect entrepreneurial entry. To exclude unobservable location-specific confounders and isolate plausibly exogenous variation in TFE, we follow Bazzi et al. (2020a) and instrument TFE with historical shocks to the settlement process, which are driven by historical immigration flows to the U.S. in the late 1800s. Immigrants created population pressures that lead locals and immigrants themselves to move west, pushing out the frontier. We observe a similar relationship between instrumented TFE and new business formation in a county, with a somewhat higher economic magnitude (~8% increase in new business registrations for a one standard deviation increase in TFE). Overall, the analysis suggests that individualism, as proxied by the county's historical frontier experience, is a significant factor in explaining the current geographic distribution of new business formation.

Next, we examine a second key cultural factor: the county's historical diversity in immigrant backgrounds. To the extent that immigrant values spill over into the local population and these cultural values persist over time (e.g., Giuliano and Tabellini 2020), the broader diversity of initial immigrant backgrounds can lead to the diversity of viewpoints and thought, which can lead to creative recombinations that support innovation and entrepreneurial activity. Alternatively, such diversity may lead to societal fragmentation and a lack of social cohesion, impeding new business formation.

We observe a positive and significant relation between a county's historical diversity (HHI of population birthplaces by country) and per capita new business starts, suggesting that on average, the diversity of viewpoints and values brought by immigrants creates a persistent local culture that facilitates new business formation. The estimated effect is also economically significant: a one standard deviation increase in diversity of birth countries is associated with a 14% increase in new business registrations per capita. We observe no significant relationship, however, between diversity in birthplaces and entrepreneurial quality.

Next, we examine the topographical features of counties. Using the terrain ruggedness index developed by Riley, DeGloria, and Elliott (1999), we document a positive and significant relation between a county's terrain ruggedness and the number of new business registrations per capita. A one standard deviation increase in geographical ruggedness is associated with a 14% increase in new business registrations per capita, consistent with the notion that rugged terrain imposed historical constraints that may contribute to individualistic culture and self-reliance. As was the case for TFE, we also observe a small positive and statistically significant relationship between geographic terrain ruggedness and entrepreneurial quality.

Of course, the three factors we examine above may not be entirely independent of each other. In our final set of analyses, we estimate linear models in which we include all three factors (TFE, historical birthplace diversity, and terrain ruggedness) as right-hand side variables. When all three factors are horse raced in the same model, all retain their statistical and economic significance, with similar magnitudes to those obtained in our single-factor models. We then examine the interactive effects of these measures, adding interaction terms to our models one by one. We observe positive and significant interaction effects for historical birthplace diversity with both terrain ruggedness and TFE, but there is no interaction effect for TFE and terrain ruggedness. In other words, the models suggest that counties characterized by greater historical diversity see a magnification in the effects of terrain ruggedness and TFE on new business formation.

A natural follow-up question related to historical birthplace diversity's long-term effects is whether such diversity may benefit counties that today hold large concentrations of minority groups. We thus interact historical diversity with the county's current fraction of the population that is Hispanic or African-American. The estimates suggest that counties with large concentrations of minority groups, but low historical birthplace diversity have lower per capita new business starts. However, high minority areas experience more new business starts when the county is also characterized by high levels of historical diversity. Hence, our results suggest that while large concentrations of minority groups in counties with no historical diversity lead to societal fragmentation and harm new business formation, minority groups integrated into a historically diverse environment lead to greater entrepreneurial activity. This finding suggests that the positive effects of societal diversity may take time to materialize fully.

Our paper makes several contributions to the literature. First, and foremost, our paper contributes to a growing literature on the factors that drive entrepreneurial entry, a phenomenon that, despite much research, is still not well-understood. While studies have explored the effects of non-pecuniary benefits (Hurst and Pugsley 2011), wealth constraints (Hurst and Lusardi 2004), safety nets (Barrios et al 2020), and regulatory aspects (Djankov et al. 2010), these determinates fail to fully explain the significant variation in new business formation observed across both time and space. Our paper documents new factors that can explain such variation. Specifically, our findings suggest a fundamental role for America's frontier culture and individualism in shaping new business formation and entrepreneurial activity in the U.S. This, in turn, should spur further research on the cultural drivers of new business formation.

Second, and relatedly, by focusing on the historical, geographic, and cultural features that drive new business entry, our paper contributes to the existing literature on the importance of culture in business transactions. For example, Guiso, Sapienza, and Zingales (2006) highlight the role of cultural factors like trust or social capital for business transactions, while Alesina and Giuliano (2015) focus on the role of culture in institutions. By examining the role that historical geographic and cultural factors play in shaping new businesses formation, we extend the literature on culture into the sphere of entrepreneurship in the spirit of Iyer and Schoar (2010) and Landier and Thesmar (2008).

Finally, the empirical patterns we document contribute to the literature on agglomeration economies. The spatial variation in new business formation that we document, and its association with terrain ruggedness and other historical features, provides new insights that inform the discussion on the clustering of economic activities in certain areas of the country. Specifically, variation in clustering fostered by the terrain and historical-cultural features of U.S. regions sheds light on the origins of economic clustering in the U.S. (Duranton and Pagu, 2004; Glaeser 2010).

2. SAMPLE AND DATA

2.1. Sample

Our sample consists of all U.S. counties with a population greater or equal to 10,000 people in 2010, observed quarterly from 1988 to 2016, the last available business registration data year. The initial sample consists of 360,364 observations for 3,142 counties (sample size varies in the analysis depending on the data availability for any particular specification).

2.2. New Business Formation

Our outcomes of interest are the quantity and quality of new business launches, measured using business registrations by location and time period. For this purpose, we obtain data on new for-profit business registrations from the Startup Cartography Project (SCP, Guzman and Stern, 2020).[2] The SCP leverages business registration records, which are public records created when an individual register's a new business as a corporation, LLC or partnership. Importantly, as noted by Guzman and Stern (2020), while it is possible to found a new business without business registration (e.g., a sole proprietorship), the benefits of registration are substantial, and include limited liability, various tax benefits, the ability to issue and trade ownership shares, and credibility

[2] See http://www.startupcartography.com.

with potential customers. Furthermore, all corporations, partnerships, and limited liability companies must register with a Secretary of State or equivalent to take advantage of these benefits. The act of registering the firm triggers the company's legal creation. As such, these records reflect the population of incorporated businesses operating in a location (which may differ from their state of incorporation) that take a form that is a practical prerequisite for growth.

The SCP dataset provided to us covers 49 states plus the District of Columbia, and for each state the data includes records on the complete population of firms satisfying one of the following two conditions: (i) a for-profit firm physically located in the local jurisdiction, or (ii) a for-profit firm whose jurisdiction is in Delaware but whose principal office address is in the local state.

The SCP dataset provides a number of variables of interest to entrepreneurship researchers. We focus here on two specifically: (i) the quantity of new business registrations in a Census incorporated place in a given year-quarter, and (ii) an Entrepreneurial Quality Index, which is a measure of average quality within any given group of firms and represents a prediction for the probability of a growth outcome for a firm within a specified population of start-ups in a specific period (Guzman and Stern 2020).

To measure the geographic intensity of new business formations, we need to consider the number of new business registrations relative to the local population's size. We define our main variable of interest, *Log New Business Starts per Capita*, as the natural logarithm of one plus the number of new business registrations in the county-year-quarter scaled by the county's population in thousands. We graph the county variation of *New Business Starts per capita* in Figure 1-A. Our second variable of interest is *Entrepreneurial Quality Index* (EQI). Data on EQI is available for only a subset of the SCP data. Our sample is thus restricted to 211,394 observations in the analyses of the EQI.

2.3. Cultural and Geographic factors

We consider three cultural and geographic factors that may shape the contours of entrepreneurial entry and new business formation in the U.S.. First, we consider the frontier experience (*Total Frontier Experience*). Our measure is derived from Bazzi et al. (2020a), and identifies the frontier line in each year from 1790 to 1890 as the contour lines longer than 500 kilometers, after removing all "inner island lines", that are east of the main frontier line. Counties within 100 kilometers of this line are defined as frontier counties. This historically grounded definition of the frontier captures two dimensions of frontier life: population sparsity and isolation from urban centers. Similar to Bazzi et al. (2020a), we construct a county-level measure of TFE, which captures the number of decades that a given county spent on the frontier from 1790 to 1890, the end of the frontier era according to Turner (1893) and the U.S. Census Bureau. Our baseline analysis focuses on the 3,109 counties whose entire frontier history can be observed from 1790 to 1890. The geographic variation in TFE can be observed in Figure 1-B. In the multivariate analysis, we standardize the TFE measure to be mean zero and standard deviation one to ease the interpretation of coefficients.

Second, we measure a county's historical population's birthplace diversity along the country of origin, we follow Manson et al. (2019). We measure a county's historical birthplace diversity, *Birthplace Diversity in 1890*, as one minus the Herfindahl-Hirschman concentration index for country of birth in the county's population in 1890. Birth country is defined as the U.S. or a given country or country grouping abroad. We plot the geographic variation in *Birthplace Diversity in 1890* in Figure 1-C. In the multivariate analysis, to ease our findings' interpretation, we standardize this measure to mean zero and standard deviation one.

In addition to historical birthplace diversity, we also measure the current concentration of minority groups in a county, *Minority*, defined as the average fraction of a county's population in the period 2000 to 2016 that is African-American or Hispanic. Data on minorities is available for the period 2000 to 2016, and for a smaller sample of counties (1,161). Hence, our analysis for *Minority* is limited to a sample with 76,500 observations. Consistent with the prior measures, we standardize this variable for ease of coefficient interpretation.

Finally, we measure terrain ruggedness using the county-level average *Terrain Ruggedness Index* (TRI) computed using 30-arc grid data on terrain variability based on Nunn and Puga (2012). The geographic distribution of TRI is depicted in Figure 1-D. We standardize this measure to be mean zero and standard deviation one.

2.4. Controls

Our regression models include a vector of covariates that control a county's demographic and geographic characteristics. We obtain annual income per capita from U.S. Census. We control for the county's average population density, measured as the ratio between the county's population and area using digitalized U.S. Census data on population for every decade in 1790-2010. The data on area is calculated using the 2010 county shapefiles from NHGIS using GIS software (Manson et al., 2019). The county-level population data and other pre-2010 data are harmonized to the 2010 county boundaries. The data for intercensal years is imputed using the procedure detailed in Bazzi et al. (2020a). We also control for the average Republican vote share in presidential elections. The data is obtained from ICPSR (8611) Presidential and Congressional Races, 1840–1972 for 1900–1972 elections, General Election Data for the United States 1950–1990 (ICPSR 13) for 1976–1996 elections, and Dave Leip's Atlas of U.S. Presidential Elections for 2000–2016 elections (2017). We also control for the average effective property tax rate per $100 of value, calculated as the county's ratio of the average real estate tax over the average house value, with data obtained from the National Association of Home Builders. They calculated the rates based on the 2010–2014 American Community Survey (ACS) data from the Census Bureau.

We also utilize data on the geographic characteristics from Manson et al. (2019) and Bazzi et al. (2020a). We include geographic controls for the log of a county's surface area in square miles, the absolute latitudinal distance from the equator in decimal degrees, the absolute longitudinal distance from the Greenwich Meridian in decimal degrees, and the average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. These variables are calculated from each county's centroid using GIS software and the 2010 county shapefiles from NHGIS. We provide descriptive statistics for all our variables in Table A1 of the appendix.

3. RESULTS

We begin our analysis by examining the geographic distribution of new business formation across U.S. counties. We start at the most basic level, taking a single observation for each county, and asking whether new business formation rates differ more than would be expected from sampling variation. We then expand the unit of analysis to the county-year-quarter pair, allowing us to estimate dynamic panel models with state and then county fixed effects.

We begin by analyzing whether new business formation tends to cluster regionally more than predicted by chance. In Exhibit 1 Figure 1-A, we observe preliminary visual evidence suggesting the clustering of new business formation in some regions of the U.S. However, the figure does not tell us whether these differences are larger than we would expect from random variation. Exhibit 2 Panel A, presents the estimates from a more rigorous analysis, using a single observation for each county. We examine three dependent variables: the number of year-quarters for which new business registrations occurs in a given county (Column 1), the number of year-quarters for which new businesses have above the median EQI (Column 2), and the number of year-quarters for which new businesses have above the 75th percentile EQI (Column 3), with the population at risk (the independent variable) being the total number of year-quarters for a given county in the sample.

Our models are based on the intuition that in a standard Poisson regression, the dependent variable is Poisson distributed with mean $e^{X_i\beta}$, where X_i are the covariates that determine the distribution. Using the Poisson distribution, we allow the distribution to vary by county, so that the mean becomes $e^{X_i\beta}u_i$, where u_i is a multiplicative gamma-distributed term with a mean equal to one and variance equal to α. A value of 0 for α corresponds to a standard Poisson model, whereas $\alpha > 0$ indicates unequal variance, or excess clustering, in new business registrations across counties, after accounting for each county's size population of year-quarters.

In Exhibit 2 Panel A we find that the coefficient on the number of county-year-quarters is significant and close to one across the three models, consistent with expectations. More importantly, however, the estimated variance parameter α is consistently positive and highly significant across the three models, allowing us to reject an equal distribution of new business starts across counties.[3]

Our preceding tests have low statistical power and are further limited because county-level heterogeneity is modeled based on a Poisson functional form (if the true distribution of u_i is not gamma, the regression is misspecified). Thus, we next estimate panel models at the county-year-quarter level and model regional heterogeneity using state fixed effects and county fixed effects, which are unconstrained by a specific a-priori distribution.

Panel B of Exhibit 2 reports the results. We estimate the fixed effects model for our two dependent variables: *Log of New Business Starts per capita* and the *Entrepreneurial Quality Index*.

[3] To interpret the magnitude, the result indicates that, for a county one standard deviation above the mean, new business formation is expected to occur about 38% more frequently than average.

We include year fixed effects as a benchmark model, and then add state fixed effects and county fixed effects alternatively, to examine their relative importance in explaining variation in our outcome of interest. The state fixed effects are highly significant and explain approximately 22% of the new business registration variation over the entire sample period. Similarly, state fixed effects explain 11% of the variation in the EQI. More interestingly, however, county fixed effects alone explain approximately 75% of the variation in new business registration (and 16% of the variation in EQI). These results indicate that most of the variation in new business starts is explained by time-invariant county level factors. This serves as the basis for examining the county's geographical and historical-cultural aspects, which we turn to next.

3.1. The Geographic and Cultural Contours of Variation in New Business Formation

From this point forward, we take the previously documented patterns as given, and attempt to better understand why they exist—examining how geographic and historical-cultural factors affect the quantity and quality of new business formation. We consider three important factors discussed in the introduction: the county's exposure to frontier conditions, which have been shown to lead to a persistent culture of rugged individualism; the historical diversity in country of birth; and the topographical features of counties, as proxied for by terrain ruggedness. Our analyses are motivated by theories of cultural persistence. While both the individualism induced by the historical frontier or rugged terrain, and the creativity or fragmentation induced by birthplace diversity may dissipate over time, they may well have shaped the subsequent evolution of culture in the county.

3.1.1. Total Frontier Experience

We begin by estimating the linear relation between total frontier experience (TFE) and new business formation in a panel regression model with fixed effects:

$$Y_{c,t} = \beta_0 + \beta_1 TFE_c + \beta_2 Controls_{c,t} + \tau_t + \lambda_{s(c)} + \epsilon_{i,c}, \quad (1)$$

Where $Y_{c,t}$ indicates the *log of New Business Starts per capita* or the *Entrepreneurial Quality Index* (EQI) for county *c* at year-quarter *t*. TFE_c indicates a county's *Total Frontier Experience*, which is the number of decades that a given county spent on the frontier from 1790 to 1890. $Controls_{c,t}$ denotes a vector of covariates that control for important time variant and time invariant geographic and demographic characteristics, as described in Section 2.4. We include state fixed effects ($\lambda_{s(c)}$) to control for unobserved time-invariant local characteristics. We also include year fixed effects (τ_t) to capture time trends in new business registrations. The coefficient β_1 identifies a local effect of TFE after accounting for factors that may correlate with both TFE and modern culture. Standard errors are clustered by county. We conduct all our analysis first including only fixed effects, then adding demographic controls, and, finally, adding geographic controls.

Results are reported in Exhibit 3. Panel A graphs the relationship between TFE and new business starts controlling for demographic and geographic characteristics, as well as state and year fixed effects. For both outcome variables, we observe a positive relation to TFE. Panel B presents the estimates for Equation (1). We find that a one standard deviation increase in TFE is associated with an increase of between 3.7% and 5% in new business registrations (Columns (1) to (3)). Columns (4) to (6) estimate the relation between TFE and EQI, which allows us to examine the quality of the new business starts (the mix between new small business and high-growth oriented business in the county). We observe a positive relation between TFE and EQI, but of minimal economic magnitude: a one standard deviation increase in TFE is related to an increase in EQI of approximately 1.25% over the unconditional mean.

The main threat to causal identification lies in omitted variables correlated with both contemporary culture and TFE. We address this concern in two ways. First, we note that, across the different specifications with different controls, the coefficient of interest remains stable, and the increment in the R^2 is small. This suggests that omitted correlated variables are unlikely to affect our results (Oster 2019). Second, we use an IV strategy that isolates plausibly exogenous variation in TFE due to changes in national immigration flows over time, as in Bazzi et al. (2020a).

Our IV is based on historical shocks to the settlement process driven by immigration flows to the U.S. These immigrants contributed to westward expansion by exerting population pressure on the eastern seaboard, and by going west themselves. Thus, the flow of immigration determined the time it took for frontier locations in different periods to become established settlements. We isolate push factors by predicting migrant outflows from 16 European countries from 1820 to 1890 based on climate shocks (Sequeira et al. 2020). As in Bazzi et al. (2020a), we use country-specific regressions to predict the annual migrant outflows from each European country to the U.S. as a function of country-specific shocks to temperature and rainfall in the prior year. We then aggregate across countries to obtain the total predicted migrant inflows to the U.S. for each year. Finally, we construct the IV for each county by calculating the average annual predicted migrant inflow to the U.S. over the 30 years starting from the first year in which the given county is just west of the frontier. Hence, for each county in our sample, the instrument captures predicted climate-induced emigration flows to the U.S. starting just before the onset of local frontier settlement. These time-varying, national population shocks are unrelated to frontier counties' local conditions and help move us closer to a causal interpretation. We label our IV "*Log Average Predicted National Migration Inflows*".

Exhibit 3 of Panel C reports the IV estimates. The first stage regressions (Columns (1) and (3)), demonstrate a large and statistically significant effect of the IV on TFE, with first-stage F statistics above relevant thresholds (Stock and Yogo 2005), indicating the relevance of the instrument. Our instrumented TFE (Columns (2) and (4)) has a positive and slightly larger effect on *New Business Starts per Capita* and *Entrepreneurial Quality Index* than in the OLS estimates. A one standard deviation increase in TFE leads to an 8% increase in new business starts, and a 4% increase in EQI (relative to the unconditional mean of 0.048). Taken together, these estimates suggest that our findings are not driven by unobserved local conditions determining both TFE and outcomes of interest today.

3.1.2. Historical Birthplace Diversity

Next, in Exhibit 4 Panels A and B, we examine the role of historical diversity in immigrant backgrounds in the county. We estimate a linear panel model similar to that in Equation (1), where we replace TFE with *Historical Birthplace Diversity in 1890*. As before, we introduce demographic and geographic controls gradually across the three models, and all specifications include state and year fixed effects. Panel A shows the relationship between birthplace diversity and new business starts controlling for demographic and geographic characteristics, as well as state and year fixed effects. We observe a positive relationship between historical birthplace diversity and new business registrations, but no relation between historical birthplace diversity and EQI.

Similarly, the estimates of Equation (1) in Panel B exhibit a positive and significant association between the historical birthplace diversity and new business starts per capita, suggesting that on average, the diversity of viewpoints and values brought by immigrants creates a persistent local culture that facilitates new business formation. The estimated effect is economically significant: a one standard deviation increase in historical birthplace diversity is associated with a 14% increase in new business registrations per capita. In contrast, we observe no significant relationship between historical birthplace diversity and EQI. Taken together, evidence in Panels A and B of Exhibit 4 suggests that a wider diversity of initial immigrant backgrounds supports greater entrepreneurial activity, but not necessarily a greater proportion of high-growth-oriented business starts.

3.1.3. Terrain ruggedness and new business formations

Finally, in Exhibit 4 Panels C and D we examine the relation between the topographical features of the U.S. counties and new business formations. Once again, we estimate a linear panel model similar to that in Equation (1), replacing TFE with the *Terrain Ruggedness Index.* In the graphs reported in Panel C, and the estimates in Panel D, we observe a positive and significant association between a county's terrain ruggedness and the number of business starts per capita. A one standard deviation increase in geographical ruggedness is associated with a 14% increase in new business registrations per capita. Similar to the results for TFE, we also observe a small positive and statistically significant relationship between geographic terrain ruggedness and entrepreneurial quality, in the order of magnitude of a 2% increase in the unconditional mean.

3.1.4. Interactive Effects

The geographic and cultural factors we examine above are unlikely to be fully independent of each other. Thus, in our next set of analyses, we estimate linear model similar to Equation (1) in which we include all three factors as right-hand side variables (Exhibit 5, Panel A). Consistent to the prior analysis, we first examine the model without controls (Column 1), then introduce demographic controls (Column 2), and finally add geographic controls (Column 3). For robustness, we also examine the model with all controls included for a subsample of years post-2000 (Column 4) and post-2005 (Column 5). As can be seen from the exhibit, the coefficients on all three historical determinants retain their statistical and economic significance, with similar magnitudes to those obtained in our single factor models.

Next, in Panel B, we examine the interactive effects of the three factors. We augment our specification with interactions of TFE with terrain ruggedness, interactions of historical diversity with terrain ruggedness, or interactions of TFE with historical diversity. Panel B plots the relationship between new business starts per capita and TFE across the quintiles of terrain ruggedness and birthplace diversity, and similarly, the relationship between new business starts per capita and birthplace diversity across quintiles of terrain ruggedness. While we observe a positive relation between new business starts and historical diversity amongst quintiles of terrain ruggedness, and a positive relation between new business starts and TFE amongst the upper quintiles (quintile three and four) of terrain ruggedness, we find that the positive relation between new business starts and TFE is stronger in the top quintile of historical diversity.

In Panel C, we report formal estimates of our determinates' interactions in the spirit of Equation (1). Column (1) adds the interaction of terrain ruggedness with TFE. While the individual determinants maintain their positive and significant coefficients, the interaction term of TFE and terrain ruggedness does not load significantly. In contrast, in Column (2), when we add the interaction of terrain ruggedness with birthplace diversity, both the individual factors and the interaction term load positively and significantly, and in Column (3), we similarly observe positive and statistically significant loadings both for the individual factors and for the interaction of birthplace diversity and TFE. Overall, the figures and tables suggest that in counties defined as having higher historical birthplace diversity levels, the effects of terrain ruggedness and TFE on new business formation are magnified.

3.1.5. Differential Minority Effects and Geographic and Cultural Determinates

Our analysis on the relationship between new business starts and historical birthplace diversity raises the natural question of whether such historical diversity may also affect underrepresented minorities today. To examine this relation, we look at counties that today hold large concentrations of minority groups, interacting historical diversity with the current fraction of the minority population in the county (see Column 4 of Panel C of Exhibit 5). The estimates suggest that counties with large concentrations of minority groups with low historical diversity have lower per capita new business starts. Yet, minority counties characterized by high levels of historical diversity are associated with higher levels of new business starts per capita.

Taken together, our results suggest that while large concentrations of minority groups in counties with no historical diversity lead to societal fragmentation and hinder new business formation, minority groups integrated in a historically diverse environment lead to greater entrepreneurial activity. This finding suggests that the positive effects of societal diversity may take time to fully materialize.

4. CONCLUSIONS

This paper examines the role of geographic and key historical-cultural factors on shaping new business formation. Using novel business registration data, we document that county fixed effects explain approximately 75% of the new business registration variation over the entire sample period (1988-2016). We provide evidence that time-invariant historical-cultural and geographic factors explain a significant portion of the prevalence of entrepreneurial entry in a local area. Our study document a fundamental role for America’s frontier culture in shaping new business formation and entrepreneurial activity in the U.S. These findings suggest a need for further research exploring the role of cultural drivers of new business formation.

REFERENCES

Agrawal, A., Cockburn, I., & Rosell, C. (2010). Not invented here? Innovation in company towns. Journal of Urban Economics, 67(1), 78-89.

Ager, P., & Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the US during the age of mass migration. European Economic Review, 64, 76-97.

Ager, Philipp, and Markus Brueckner (2018). Immigrants' genes: Genetic diversity and economic development in the United States. Economic Inquiry 56, no. 2, 1149-1164.

Alesina, A., & Giuliano, P. (2015). Culture and institutions. Journal of Economic Literature, 53(4), 898-944.

Bandiera, O., Burgess, R., Deserranno, E., Morel, R., Rasul, I., & Sulaiman, M. (2017). Social connections and the delivery of development programs. Work. Pap., London School Econ.

Barrios, J. M., Hochberg, Y. V., & Yi, H. (2020). Launching with a Parachute: The Gig Economy and Entrepreneurial Entry. University of Chicago, Becker Friedman Institute for Economics Working Paper, (2020-21).

Bazzi, S., Fiszbein, M., & Gebresilasse, M. (2020a). Frontier culture: The roots and persistence of "rugged individualism" in the United States. Econometrica, 88(6), 2329-2368.

Bazzi, S., Fiszbein, M., & Gebresilasse, M. (2020b). Rugged Individualism and Collective (In) action During the COVID-19 Pandemic (No. w27776). National Bureau of Economic Research.

Burchardi, K. B., Chaney, T., & Hassan, T. A. (2019). Migrants, ancestors, and foreign investments. The Review of Economic Studies, 86(4), 1448-1486.

Catalini, C., Guzman, J., & Stern, S. (2019). Passive Versus Active Growth: Evidence from Founder Choices and Venture Capital Investment (No. w26073). National Bureau of Economic Research.

Davis, S. J., & Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. The Quarterly Journal of Economics, 107(3), 819-863.

Djankov, S., Ganser, T., McLiesh, C., Ramalho, R., & Shleifer, A. (2010). The effect of corporate taxes on investment and entrepreneurship. American Economic Journal: Macroeconomics, 2(3), 31-64.

Duranton, Gilles, and Diego Pagu (2004). Micro-foundations of urban agglomeration economies, in J. Vernon Henderson, and Jacques-Francois Thisse, eds.: Handbook of Regional and Urban Economics (North-Holland, Amsterdam).

Ellison, G., & Swanson, A. (2016). Do schools matter for high math achievement? Evidence from the American mathematics competitions. American Economic Review, 106(6), 1244-77.

Fazio, C., Guzman, J., & Stern, S. (2020). The impact of state-level research and development tax credits on the quantity and quality of entrepreneurship. Economic Development Quarterly, 34(2), 188-208.

Giuliano, P., & Tabellini, M. (2020). The Seeds of Ideology: Historical Immigration and Political Preferences in the United States (No. w27238). National Bureau of Economic Research.

Glaeser, E. L. (Ed.). (2010). Agglomeration economics. University of Chicago Press.

Glaeser, E. L., & Kerr, W. R. (2009). Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain?. Journal of Economics & Management Strategy, 18(3), 623-663.

Glaeser, E. L., Kerr, S. P., & Kerr, W. R. (2015). Entrepreneurship and urban growth: An empirical assessment with historical mines. Review of Economics and Statistics, 97(2), 498-520.

Glaeser, E. L., Kerr, W. R., & Ponzetto, G. A. (2010). Clusters of entrepreneurship. Journal of urban economics, 67(1), 150-168.

Grosjean, P. (2014). A history of violence: The culture of honor and homicide in the US South. Journal of the European Economic Association, 12(5), 1285-1316.

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes?. Journal of Economic perspectives, 20(2), 23-48.

Guzman, J., & Stern, S. (2020). The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014. American Economic Journal: Economic Policy, 12(4), 212-43.

Haltiwanger, J., Jarmin, R. S., & Miranda, J. (2013). Who creates jobs? Small versus large versus young. Review of Economics and Statistics, 95(2), 347-361.

Hurst, E., & Lusardi, A. (2004). Liquidity constraints, household wealth, and entrepreneurship. Journal of political Economy, *112*(2), 319-347.

Hurst, E., & Pugsley, B. W. (2011). What do small businesses do? (No. w17041). National Bureau of Economic Research.

Iyer, R., & Schoar, A. (2010). Are there cultural determinants of entrepreneurship?. In International differences in entrepreneurship (pp. 209-240). University of Chicago Press.

Jovanovic, B., & MacDonald, G. M. (1994). The life cycle of a competitive industry. Journal of Political Economy, 102(2), 322-347.

Landier, A., & Thesmar, D. (2008). Financial contracting with optimistic entrepreneurs. The Review of Financial Studies, 22(1), 117-150.

Manson, S., Schroeder, J., Van Riper, D., & Ruggles, S. (2017). Ipums national historical geographic information system: Version 12.0 [database]. Minneapolis: University of Minnesota, 39.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. Review of Economics and Statistics, 94(1), 20-36.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187-204.

Ottaviano, G. I., & Peri, G. (2006). The economic value of cultural diversity: evidence from US cities. Journal of Economic geography, 6(1), 9-44.

Parsons, C. A., Sulaeman, J., & Titman, S. (2018). The geography of financial misconduct. The Journal of Finance, 73(5), 2087-2137.

Rapaport, E., & Snickars, F. (1999). gis-based road location in Sweden: A case study to minimize environmental damage, building costs and travel time. In Geographical information and planning (pp. 135-153). Springer, Berlin, Heidelberg.

Riley, S. J., DeGloria, S. D., & Elliot, R. (1999). Index that quantifies topographic heterogeneity. intermountain Journal of sciences, 5(1-4), 23-27.

Sequeira, S., Nunn, N., & Qian, N. (2020). Immigrants and the Making of America. The Review of Economic Studies, 87(1), 382-419.

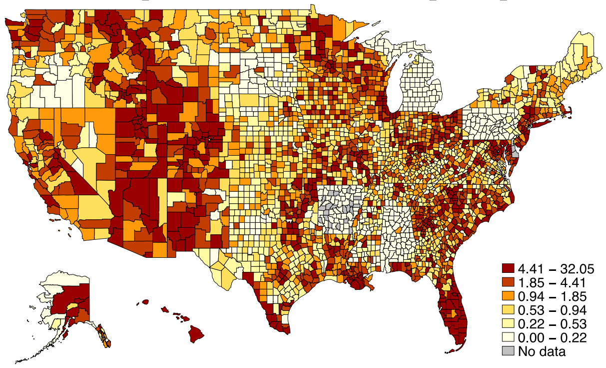
Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. Identification and inference for econometric models: Essays in honor of Thomas Rothenberg, 80(4.2), 1.

Turner, F. J. (1893). The Significance of the Frontier in American History. Proceedings of the State Historical Society of Wisconsin.

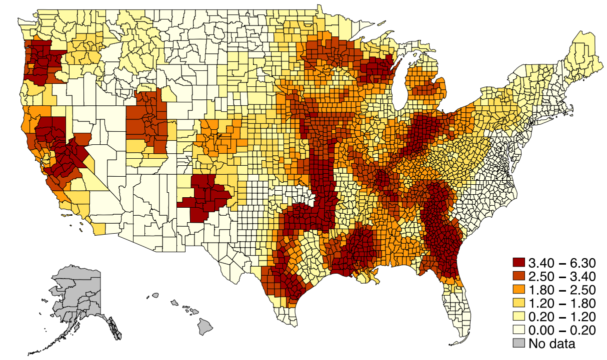
Weber, M. (1904). Die" Objektivität" sozialwissenschaftlicher und sozialpolitischer Erkenntnis. Archiv für sozialwissenschaft und sozialpolitik, 19(1), 22-87.

EXHIBIT 1

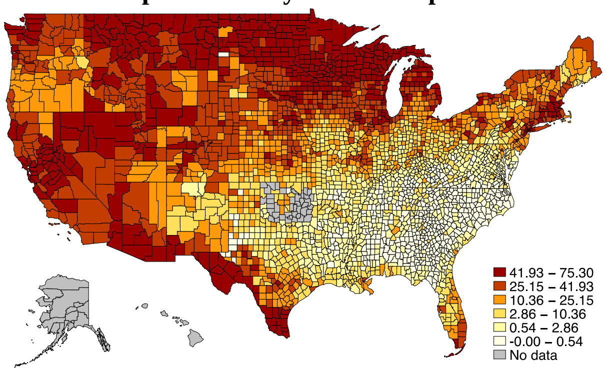
Panel A: Map of New Business Starts per capita



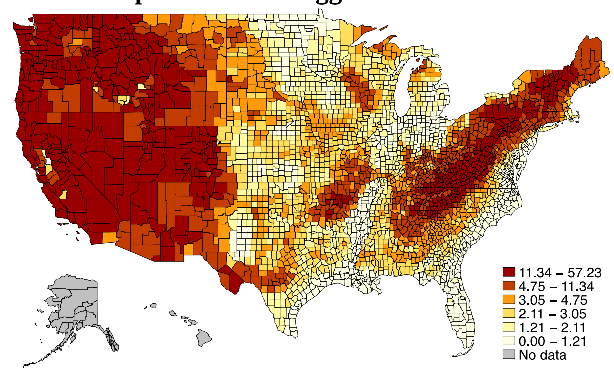
Panel B: Map of Total Frontier Experience



Panel C: Map of Diversity of Birth Population in 1890.



Panel D: Map of Terrain Ruggedness Index



Notes: This exhibit shows the geographic variation in our measure of new business formations as well as the measures of culture and geographic ruggedness. Panel A shows the county's median variation in the number of new firm registrations per 1,000 people. Panel B shows the variation in *Total Frontier Experience*. Total frontier experience is the total number of decades a county was within 100 km of the frontier line between 1790–1890 based on county-level data from NHGIS. The yellow areas to the west are beyond the 1890 frontier line. Panel C shows the variation in *Diversity of Birth Population in 1890*, which is defined as one minus the Herfindahl-Hirschman concentration index for country of birth in the county's population in 1890. Finally, Panel D shows the variation in the *Terrain Ruggedness Index*, which is measured using 30-arc grid data on terrain variability (Nunn and Puga, 2012).

EXHIBIT 2

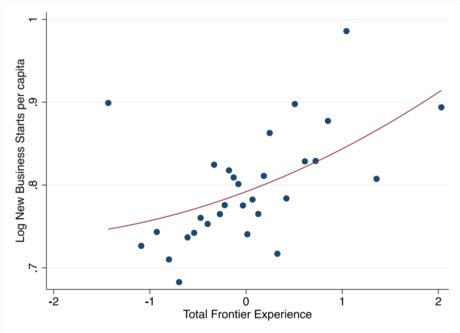
	Number of county-year-quarters with:		
	new business registrations	above the median EQI	top quartile EQI
	(1)	(2)	(3)
Log(Number of County-Year-quarters)	0.9659***	0.9174***	0.8414***
	(0.0409)	(0.0380)	(0.0437)
Costant	0.0076	0.0462	0.0252
	(0.1941)	(0.1805)	(0.2075)
N	3,142	3,142	3,141
Pseudo R2	0.0268	0.0611	0.0133
α	0.1436	0.0372	0.3155
p- value	<0.001	<0.001	<0.001

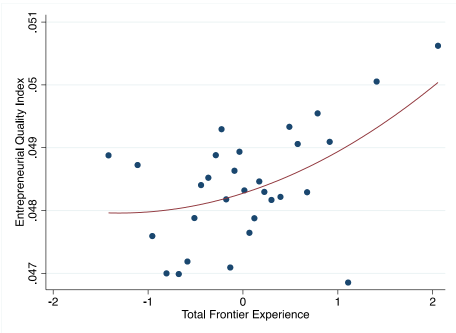
Panel A: This table presents estimates from negative binomial regressions predicting the number of new business registrations (NBR) and Entrepreneurial Quality Index (EQI) in each county. The dependent variable is the number of year-quarters with NBR in county c in Column (1), the number of year-quarters with firms with EQI above the median in Column (2), and the number of year-quarters with firms with EQI in the top quartile in Column (3). The independent variable is the log of the number of year-quarters in the county. The sample uses all counties in our sample. The parameter $\hat{\alpha}$ is the estimated standard deviation of random effects at the county level. The p-value corresponds to a $\chi 2$ test of the null that $\hat{\alpha}$ is 0. The significance levels are abbreviated with asterisks: *** indicates statistical significance at the 1% level. The observation unit is at the county level. The robust standard errors are reported in parentheses.

Dependent Variable	Year Fixed Effects	State Fixed Effects	County Fixed Effects	Observations	Adj R-squared
(1)	(2)	(3)	(4)	(5)	(6)
Log New Business Starts per capita	1115.99 (< .0001)			360,364	0.08
Log New Business Starts per capita	1444.73 (< .0001)	2017.09 (< .0001)		356,820	0.30
Log New Business Starts per capita	6169.63 (< .0001)		527.73 (< .0001)	360,364	0.83
Entrepreneurial Quality Index	524.62 (< .0001)			211,394	0.06
Entrepreneurial Quality Index	589.61 (< .0001)	577.24 (< .0001)		209,781	0.17
Entrepreneurial Quality Index	593.86 (< .0001)		15.11 (< .0001)	211,378	0.22

Panel B: This table presents estimates from fixed effects panel regressions, where standard errors are clustered at the county level. For each dependent variable (as reported in Column 1), the fixed effects included are row 1: year fixed effects; row 2: year and state fixed effects; row 3: year and county fixed effects. We also report F-tests for the joint significance of the various fixed effects. For each F-test we report the value of the F-statistic and the p-value. Column 5 reports the number of observations, and column 6 the adjusted R2s for each regression.

EXHIBIT 3





Panel A: This panel plots our two measures of new business formation, *log of New Business Starts per capita* (left graph), and *Entrepreneurial Quality Index* (right graph) against the county's *Total Frontier Experience*. Each plot includes controls for demographic characteristics of the county (population density, income per capita, republican vote share and average property tax) and geographic characteristics (log of a county's surface area in square miles, the absolute latitudinal distance from the equator in decimal degrees, the absolute longitudinal distance from the Greenwich Meridian in decimal degrees, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters), and year-state indicators.

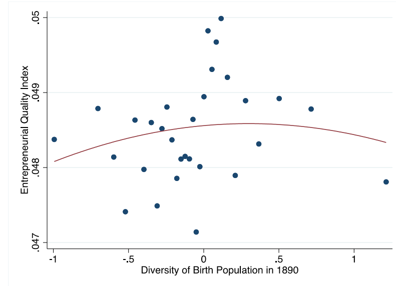
	Log New Business Starts per capita			Entrepreneurial Quality Index		
	(1)	(2)	(3)	(4)	(5)	(6)
Total Frontier Experience	0.0373**	0.0501***	0.0470***	0.0000	0.0006***	0.0006***
	(0.0155)	(0.0151)	(0.0154)	(0.0002)	(0.0002)	(0.0002)
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Geographic Controls	No	No	Yes	No	No	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	356,820	355,128	352,516	209,781	209,427	206,940
Adj. R2	0.299	0.344	0.357	0.172	0.186	0.189

Panel B: This table presents OLS estimates of equation (1). The dependent variables are *log of New Business Starts per capita* (Columns 1 -3) or the *Entrepreneurial Quality Index* (Columns 4-6) measured at the county *c* and year-quarter *t*. *Total Frontier Experience,* is the number of decades that a given county spent on the frontier from 1790 to 1890. The variable is standardized to have a mean of zero and standard deviation of one. The specifications include controls for demographic characteristics of the county: population density, per capita income, republican vote share and average property tax. We also control for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator in decimal degrees, the absolute longitudinal distance from the Greenwich Meridian in decimal degrees, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. All specifications include state and year fixed effects. Standard errors are clustered by county. *p<0.10. **p<0.05, ***p<0.01.

	Log New Business Starts per capita		Entrepreneurial Quality Index	
	(1)	(2)	(3)	(4)
	First Stage	Second Stage	First Stage	Second Stage
X: Pred Total Frontier Experience	–	0.0794**	–	0.0020***
		(0.0341)		(0.0005)
IV: Log Average Predicted National Migration Inflows	-0.6907***	–	-0.6476***	–
	(0.0287)		(0.0303)	
DemographicControls	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
N	352,516	352,516	206,940	206,940
Adj. R2	0.590	0.083	0.617	0.017
First Stage F Statistic	580.20		456.70	

Panel C: This panel reports instrumental variables estimates of equation (1) based on the instruments described in Section 3.1.1. The IV is based on annual migration inflows to the US predicted by weather shocks in Europe. The first stage is reported in Columns 1 and 3 for each of the outcome measures (*log of New Business Starts per capita* and *Entrepreneurial Quality Index*). The second stage results are reported in Columns 2 and 4. Each specification includes controls for demographic characteristics of the county: population density, per capita income, republican vote share and average property tax. We also control for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator in decimal degrees, the absolute longitudinal distance from the Greenwich Meridian in decimal degrees, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. All specifications include state and year fixed effects. The first-stage F statistics Standard errors are clustered by county. *p<0.10. **p<0.05, ***p<0.0.

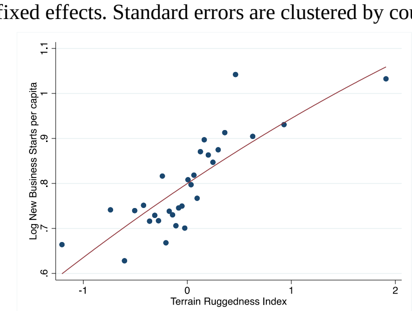
EXHIBIT 4

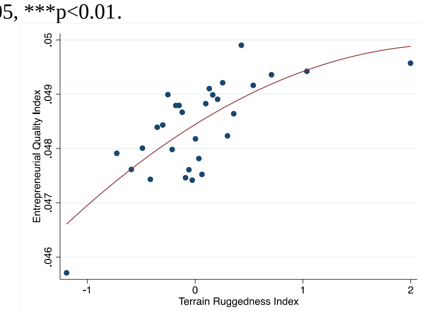


Panel A: This panel plots our two measures of new business formation, *log of New Business Starts per capita* (left graph) and *Entrepreneurial Quality Index* (right graph) against the county's *Diversity in Birth Population in 1890*. Each plot includes controls for demographic characteristics of the county (population density, income per capita, republican vote share and average property tax) and geographic characteristics (log of a county's surface area in square miles, the absolute latitudinal distance from the equator in decimal degrees, the absolute longitudinal distance from the Greenwich Meridian in decimal degrees, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters), and year-state indicators.

	Log New Business Starts per capita			Entrepreneurial Quality Index		
	(1)	(2)	(3)	(4)	(5)	(6)
Diversity of Birth Population in 1890	0.1922***	0.1450***	0.1421***	0.0031***	0.0001	0.0001
	(0.0270)	(0.0257)	(0.0273)	(0.0006)	(0.0004)	(0.0004)
Demographic_Controls	No	Yes	Yes	No	Yes	Yes
Geographic_Controls	No	No	Yes	No	No	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	352,064	350,372	347,760	207,670	207,316	204,829
Adj. R2	0.313	0.351	0.365	0.175	0.189	0.191

Panel B: This table presents OLS estimates of equation (1). The dependent variables are the *log of New Business Starts per capita* (Columns 1-3) or the *Entrepreneurial Quality Index* (Columns 4-6) measured at the county *c* and year-quarter *t*. *Diversity of Birth Population in 1890* is defined as one minus the Herfindahl-Hirschman concentration index for country of birth in the county's population in 1890. We standardize this measure to be mean zero and standard deviation one. The specifications include controls for demographic characteristics of the county: population density, per capita income, republican vote share, and average property tax. We also control for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator, the absolute longitudinal distance from the Greenwich Meridian, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. All specifications include state and year fixed effects. Standard errors are clustered by county. *p<0.10. **p<0.05, ***p<0.01.





Panel C: This panel plots our two measures of new business formation log of *New Business Starts per capita* (left graph) and *Entrepreneurial Quality Index* (right graph) against the county's *Terrain Ruggedness Index*. Each plot includes controls for demographic characteristics of the county (population density, income per capita, republican vote share and average property tax) and geographic characteristics (log of a county's surface area in square miles, the absolute latitudinal distance from the equator in decimal degrees, the absolute longitudinal distance from the Greenwich Meridian in decimal degrees, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters), and year-state indicators.

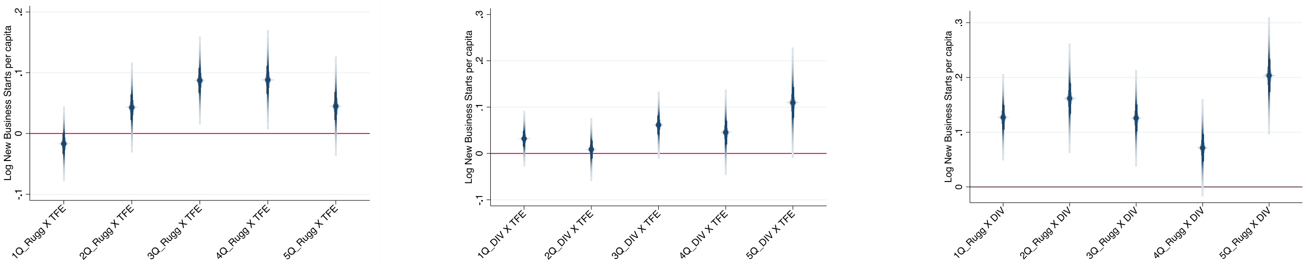
	Log New Business Starts per capita			Entrepreneurial Quality Index		
	(1)	(2)	(3)	(4)	(5)	(6)
Terrain Ruggedness Index	0.1120***	0.1348***	0.1428***	-0.0002	0.0006***	0.0009***
	(0.0210)	(0.0201)	(0.0211)	(0.0002)	(0.0002)	(0.0003)
Demographic_Controls	No	Yes	Yes	No	Yes	Yes
Geographic_Controls	No	No	Yes	No	No	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	356,820	355,128	352,516	209,781	209,427	206,940
Adj. R2	0.307	0.355	0.366	0.172	0.186	0.189

Panel D: This table presents OLS estimates in the spirit of equation (1). The dependent variables are the *log of New Business Starts per capita* (Columns 1-3) or the *Entrepreneurial Quality Index* (Columns 4-6) measured at the county *c* and year-quarter *t*. *Terrain Ruggedness Index* is using 30-arc grid data on terrain variability based on Nunn and Puga (2012). We standardize this measure to be mean zero and standard deviation one. The specifications include controls for demographic characteristics of the county: population density, per capita income, republican vote share, and average property tax. We also include controls for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator, the absolute longitudinal distance from the Greenwich Meridian, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. All specifications include state and year fixed effects. Standard errors are clustered by county. *p<0.10. **p<0.05, ***p<0.01.

EXHIBIT 5

	Log New Business Starts per capita				
	(1)	(2)	(3)	(4)	(5)
Total Frontier Experience	0.0318**	0.0398***	0.0336**	0.0432**	0.0435**
	(0.0156)	(0.0152)	(0.0158)	(0.0179)	(0.0182)
Terrain Ruggedness Index	0.1016***	0.1219***	0.1291***	0.1427***	0.1434***
	(0.0209)	(0.0202)	(0.0213)	(0.0240)	(0.0244)
Diversity of Birth Population in 1890	0.1887***	0.1317***	0.1329***	0.1532***	0.1520***
	(0.0268)	(0.0256)	(0.0270)	(0.0311)	(0.0320)
Demographic_Controls	No	Yes	Yes	Yes	Yes
Geographic_Controls	No	No	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Sample Years	ALL	ALL	ALL	Post-2000	Post-2005
Observations	352,064	350,372	347,760	191,872	131,912
Adj. R2	0.323	0.364	0.375	0.365	0.360

Panel A: This panel presents OLS estimates in the spirit of equation (1) where we include *Total Frontier Experience*, *Diversity of Birth Population in 1890*, and *Terrain Ruggedness Index* as contemporaneous right-hand side variables. In each specification we delineate which controls are used. We include controls for demographic characteristics of the county: population density, per capita income, republican vote share, and average property tax. We also include controls for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator, the absolute longitudinal distance from the Greenwich Meridian, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. The last two columns examine two different subsamples of years: post-2000 (Column 4) and post-2005 (Column 5). All specifications include state and year fixed effects. Standard errors are clustered by county. *p<0.10. **p<0.05, ***p<0.01.



Panel B: This panel plots the relationship between new business starts per capita and TFE across the quintiles of terrain ruggedness and birthplace diversity, and similarly, the relationship between new business starts per capita and birthplace diversity across quintiles of terrain ruggedness. The plotted coefficients are obtained by estimating an augmented equation (1). The specifications include interactions of TFE with terrain ruggedness, interactions of birthplace historical diversity with terrain ruggedness, or interactions of TFE with diversity of birth country. Each specification includes controls for demographic characteristics of the county: population density, per capita income, republican vote share, and average property tax. We also include controls for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator, the absolute longitudinal distance from the Greenwich Meridian, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. All specifications include state and year fixed effects. The plotted coefficients include a two-tailed 95% confidence intervals based on standard errors clustered at the county level. We provide a red line through the y-axis at zero for ease of interpretation.

	Log New Business Starts per capita			
	(1)	(2)	(3)	(4)
Total Frontier Experience	0.0342**		0.0565***	
	(0.0158)		(0.0168)	
Terrain Ruggedness Index	0.1397***	0.1236***		
	(0.0214)	(0.0202)		
Diversity of Birth Population in 1890		0.1324***	0.1544***	0.2154***
		(0.0273)	(0.0271)	(0.0548)
Minority				-0.0901**
				(0.0432)
Total Frontier Experience_X_Terrain Ruggedness Index	0.0111			
	(0.0159)			
Diversity of Birth Pop_X_Terrain Ruggedness		0.0373**		
		(0.0166)		
Diversity of Birth Pop_X_Total Frontier Experience			0.0358**	
			(0.0178)	
Diversity of Birth Pop_X_Minority				0.1364***
				(0.0418)
Demographic_Controls	Yes	Yes	Yes	Yes
Geographic_Controls	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
N	352,516	347,760	347,760	76,500
Adj. R2	0.367	0.375	0.367	0.407

Panel C: This panel presents OLS estimates in the spirit of equation (1) where we examine interactions between our various cultural/geographic measures. The dependent variable is the *Log New Business Starts per capita*. We include interactions between TFE and *Terrain Ruggedness Index* (Column 1), *Diversity of Birth Population in 1890* and Terrain Ruggedness (Column 2), TFE and *Diversity of Birth Population in 1890* (Column 3), and *Diversity of Birth Population in 1890* and *Minority*, which is the average fraction of a county's Hispanic or African-American population from 2000 to 2010. Each specification includes controls for demographic characteristics of the county: population density, per capita income, republican vote share, and average property tax. We also include controls for geographic characteristics of the county: log of a county's surface area in square miles, the absolute latitudinal distance from the equator, the absolute longitudinal distance from the Greenwich Meridian, the county-level average terrain elevation, the minimum distance to the coastline, rivers, and lakes in meters. All specifications include state and year fixed effects. Standard errors are clustered by county. *p<0.10. **p<0.05, ***p<0.01

APPENDIX

Table A1

Summary statistics	mean (1)	sd (2)	p10 (3)	p50 (4)	p90 (5)
Dependent variables:					
Log New Business Starts per capita	0.7969	0.8036	0.0000	0.5640	1.9211
Entrepreneurial Quality Index	0.0484	0.0397	0.0222	0.0411	0.0626
Independent variables:					
Total Frontier Experience	1.9478	1.4504	0.0000	1.8000	4.0000
Terrain Ruggedness Index	0.0617	0.0779	0.0082	0.0305	0.1670
Diversity of Birth Population in 1890	0.1854	0.1931	0.0028	0.1061	0.4953
Minority	0.2135	0.1860	0.0329	0.1520	0.4916
Control variables:					
Population density	196.66	910.30	4.4042	44.459	351.52
Income per capita	13,675	3,108	10,334	13,312	17,381
Average Republican Presidential vote share	59.577	13.038	42.187	60.562	75.582
Property tax rate	0.9776	0.4654	0.4734	0.8846	1.6147
Log county area	6.5232	0.7714	5.7275	6.4413	7.4947
Latitude	38.296	4.8554	31.724	38.363	44.829
Longitude	-91.775	11.496	-107.75	-90.186	-78.363
Mean elevation	435.13	534.50	46.000	277.90	1087.8
Distance to the coast	593,116	427,034	51,565	558,069	1,212,278
Distance to rivers	43,904	39,195	7,674	33,960	90,282
Distance to lakes	164,809	111,423	32,929	147,863	314,570

Notes: This table presents the summary statistics for the variables used in the analysis. We report the mean, the standard deviation, the bottom decile, the median and the top decile of the distribution for each variable. The variables are described in the Section 2.2, 2.3, and 2.4 in the paper.