NBER WORKING PAPER SERIES

LEGALIZING ENTREPRENEURSHIP

Dany Bahar
Bo Cowgill
Jorge Guzman

Working Paper 30624
http://www.nber.org/papers/w30624

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2022, Revised March 2023

The authors thank Pierre Azoulay, David Blei, Laura Boudreau, Ina Ganguli, Peter Hull, Namrata Kala, Ameet Morjaria, David Robinson, Szymon Sacher, Daniela Scur, Scott Stern, Kevin Shih, and Inara Tareque for valuable feedback. We also thank participants at NBER Summer Institute (Entrepreneurship), the Wharton Migration and Organizations Conference, and REER, for valuable feedback. Bahar thanks the Conrad N. Hilton Foundation for their support. Cowgill thanks the Ewing Marion Kauffman Foundation for their support. This work was partially funded by The Jerome A. Chazen Institute for Global Business at Columbia University. Marion Restrepo provided excellent research assistance. A previous version of this paper was titled, "The Economic Effects of Immigration Pardons: Evidence from Venezuelan Entrepreneurs." All errors and omissions are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Legalizing Entrepreneurship
Dany Bahar, Bo Cowgill, and Jorge Guzman
NBER Working Paper No. 30624
November 2022, Revised March 2023
JEL No. J26,K37,L26

ABSTRACT

We use administrative data linked to the complete formal business registry to study a 2018 policy shift in Colombia that made nearly half a million Venezuelan undocumented migrants eligible for a resident visa. Immigrants who receive the visa increase their economic activity in the form of higher entrepreneurship by a factor as high as 12, bringing it to parity with native Colombians four years later. To establish causal estimates, we develop a novel extension of a regression discontinuity design. Our design uses variation in the running-variable (coming from rain) to instrument for migrants' choices to apply for visas.

Dany Bahar
Brown University
111 Thayer St
Providence, RI 02912
dany_bahar@brown.edu

Bo Cowgill
Columbia Business School
New York, New York
bo.cowgill.work@gmail.com

Jorge Guzman
Columbia Business School
Kravis Hall, 975
655 W 130th St
New York, NY 10027
and NBER
jag2367@gsb.columbia.edu

1 Introduction

Research about immigration often focuses on the overall levels of immigration flows and their effects. A less explored but equally important set of questions is about how immigrants' legal rights affect their integration and economic contributions to their destination society.

A key barrier to understanding the effects of immigrants' rights is the absence of counterfactuals. Although immigrants differ in their legal status, these differences are mostly endogenous. The choice to obtain documentation may relate to hidden aspirations or abilities that could also influence integration through other channels. This self-selection makes it difficult to disentangle the effects of immigration, institutional context (such as legal status), and pre-existing characteristics. This paper aims to deal with this fundamental question.

We study a natural experiment in Colombia that made nearly 500 thousand undocumented Venezuelan immigrants eligible for a resident visa in 2019. The program, known as *Permiso Especial de Permanencia* (PEP), gave access to a collection of legal rights akin to those of citizens and residents. We use this policy experiment to study the entrepreneurship behavior of immigrants upon receiving this visa. We choose to focus on entrepreneurship because it is an economic activity that i) heavily relies on access to several formal markets and inputs, ii) requires risk-taking and long-term horizons (Puri and Robinson, 2013), and iii) has the potential to significantly contribute to the receiving communities through boosting job creation and economic dynamism (Haltiwanger et al., 2013a). Of course, residency rights can also increase opportunity costs of entrepreneurship (e.g., by increasing the returns to normal employment).[1] The effect of legal status on entrepreneurship is thus theoretically ambiguous.

We find that undocumented immigrants who receive a legal migratory status increase their likelihood of registering a new company by 1.2 to 1.8 percentage points. This corresponds, roughly, to up to 12 times relative to the mean level of firm creation among the studied population, putting them at par with the firm creation rate of locals.

Using a panel data approach, we decompose the immigrant entrepreneurship effect into a *physical relocation effect* and a *visa effect*, and find that the effect of receiving a visa is over twice as large as the relocation effect. This suggests that the benefits of the bundle of

[1] Kossoudji and Cobb-Clark (2002) and Bahar et al. (2021) find some evidence for residency status improving labor market access. Fairlie and Lofstrom (2015) also note theoretical ambiguity.

legal rights is more important for migrants' entrepreneurial investment than their physical relocation into their destination country.

In our setting, the immigrants eligible for the visa were those who participated in the census of undocumented immigrants *Registro Unico de Migrantes Venezolanos* (RAMV) from April to June of 2018. In August of 2018, the outgoing President Juan Manuel Santos announced that the nearly 500 thousand participants in the RAMV census would be eligible for a resident visa (PEP), but only about 60% of those registered in RAMV eventually applied for and got the visa.[2]

This census dataset gives us access to a large sample of immigrants who were eligible for legal status, including those who did not apply for it. We later track outcomes for *all* migrants in administrative data. For identification, we use discontinuities in the RAMV census numbers. When immigrants participated in the census, they received a sequentially allocated census number. Although migrants took the census throughout the country (over 1100 locations in 413 Colombian cities), the census numbers were allocated sequentially at the national level (irrespective of where the migrant completed the census).

These census numbers were used in the policy implementation to batch access to PEP applications. To avoid an avalanche of applications, the government created 22 batches based on census number thresholds at approximately even spaces (4.5% per batch). Lower numbers were eligible to apply for the visa earlier, giving them more time to complete the application compared to higher numbers. We use these threshold and timing differences to handle the endogeneity of applying for a visa.

To do so, we introduce a novel extension of the regression discontinuity design (Imbens and Lemieux, 2008). Our approach can also be seen as an application of composite treatments (Borusyak and Hull, 2021). Our approach relies on knowledge of how the running variable is constructed to identify a source of random variation in running variable scores. We then map this back to a probability that each observation lies above or below each threshold.

In our setting, the running variable is the *census number*, which is a sequential ordering of immigrants based on their census timing. As we show, migrants' census timing was affected by local weather shocks. We use these weather shocks to generate a distribution of counterfactual running variables (census numbers) for each migrant under slightly different weather shocks. Our counterfactuals are based on permutations of weather at the

[2]Santos left office two weeks after the announcement, and the policy was implemented by his successor, Ivan Duque.

same location during the calendar week of the migrant's registration.

Using the distribution of potential census numbers, we can then calculate each migrant's probability of falling above (or below) each batch threshold — and thus the expected wait time to apply. We use each migrant's actual census number and wait time as if it were a random draw from this set of potential assignments. This approach allows us to use a principled, design-based approach for using observations farther from the cutoffs (depending on the distribution of potential running variables) and potentially include a larger sample in our analysis.

We use this method and a panel data approach for our main results as well as several other complementary findings. The effect of PEP on firm creation increases over time, with most firm creation happening 3 to 4 years after visas were issued. The rate of entrepreneurship for the treated group in 2018 is 0.3%, and it raises to nearly 0.8% percentage point by 2022, a value above the national entrepreneurship rate (which we estimate at 0.7%). We interpret this as evidence that the effect we are capturing actual firm creation instead of registration of pre-existing informal firms since, in that case, we would instead expect a large number of firms to register immediately after the PEP was rolled out.[3] Furthermore in a LATE framework — as in our IV/RD main design — any migrants who were (hypothetically) seeking only to formalize pre-existing firms are likely to be always-takers.

We also find that these new firms are meaningful new sources of economic activity. We find similar impacts on the creation of both employer and non-employer firms: While these employer firms create 1 to 6 new jobs and are not "high growth" by the typical standards of developed countries, they still represent meaningful economic spillovers. The effect also appears (in relative terms) to be about the same magnitude for both sole proprietorships (*personas naturales*) and limited liability companies (*sociedades*). Finally, we find that the PEP entrepreneurs come from different backgrounds in terms of employment category at baseline, and the effect is driven by employed, self-employed and unemployed individuals alike. This hints that our results are not solely a story about subsistence entrepreneurship.

[3] La Porta and Shleifer's 2014 literature review about informal firms states, "informal firms almost never become formal." Reviews by Marx et al. (2013) and La Porta and Shleifer (2008, 2014) suggest that informal firms "start out and live out their lives informal" in part to for subsistence and tax avoidance, and find "it is difficult to lure them [into becoming formal], even with subsidies." Several field experiments around the world have attempted to induce informal firms to formalize, with low results (De Mel et al., 2013). These include experiments in Peru (Jaramillo, 2009), Brazil (De Andrade et al., 2013), and in Colombia itself (Galiani et al., 2017). Randomized microcredit interventions have not led to greater formalization (Karlan and Zinman, 2011).

Related Literature

Our results contribute to several streams of the economic literature at large.

Institutions and Entrepreneurship. First, we contribute to the understanding of the role of institutions, broadly defined, on entrepreneurship and its role in economic dynamism (Haltiwanger et al., 2013a; Decker et al., 2014). In particular, our paper contributes to the long-held discussion about the importance of institutions as motivators of individual investment and firm formation. An important portion of this prior work has focused on the design of institutions to promote growth-oriented entrepreneurship across regional ecosystems (e.g., Lerner, 2009; Murray and Stern, 2015; Chatterji et al., 2014). One of the longer-standing hypotheses in this area is that legal rules and regulations can change the allocation of talent by pushing people to invest time and effort into entrepreneurship (Baumol, 1990; Murphy et al., 1991; Acemoglu and Robinson, 2012). Yet, empirical work on the role of law on entrepreneurship has been so far limited to corporate law (Djankov et al., 2002; Guzman, 2020), without studying the legal design of rights for individuals and how these rights promote entrepreneurship. Our paper aims to make an initial contribution to this question using a large-scale policy experiment and administrative outcome data. A paper similar in spirit is Fairlie and Woodruff (2010), which uses a self-reported survey and synthetic control design to study whether U.S. immigration reforms in 1986 influenced self-employment.

Gains from Migration. Second, we also contribute to the literature on immigration and its externalities. While there is a large literature studying immigrants and their entrepreneurial choices (Azoulay et al., 2020; Kerr, 2013; Saxenian, 2002; Hunt, 2011), this work tends to focus on high skilled immigrants and the way their location in a region compares to the counterfactual of never arriving at this location. Our paper looks at undocumented immigrants all residing in the same country, but some of them get a legal migratory status due to the policy shock. While prior work has looked at the effects of legalization of migratory status in areas such as labor market outcomes (Kaushal, 2006; Kossoudji and Cobb-Clark, 2002; Amuedo-Dorantes et al., 2007; Amuedo-Dorantes and Bansak, 2011; Monras et al., 2018; Bahar et al., 2021), crime (Baker, 2015; Pinotti, 2017; Ibáñez et al., 2021), consumption (Dustmann et al., 2017), health (Giuntella and Lonsky, 2020), and education (Kuka et al., 2020), our paper is the first to study, to the best of our knowledge, investment choices of immigrants in the form of entrepreneurship.

In this line, our paper also contributes to a growing literature on the socioeconomic integration of immigrants in developing countries (Arendt et al., 2021; Foged et al., 2022b,a).

We add to this line of work by focusing on undocumented immigrants, showing that the *effective* legal rights faced by individuals, are key to their inclusion.

Treatment Effects from Thresholds. Finally, third, our novel regression discontinuity (RD) application contributes to the methodological literature by showing the application of composite treatments (Borusyak and Hull, 2021) for estimating treatment effects using threshold-based designs. This is related to "local randomization" interpretation of RD,[4] Our design is related to other methodologies that harness noise in judgements more measurements as a research tool (e.g., Cowgill, 2018; Eckles et al., 2020). Two particularly related papers are Angrist and Rokkanen (2015) and Rokkanen (2015). These papers require at least two noisy measurements of the same latent variable (in addition to the running variable). The papers provide conditions for identification of treatment effects away from the cutoff, based on extrapolation. This is conceptually similar to our approach; however, their approach uses parametric assumptions for the extrapolation. Our paper is, to the best of our knowledge, the first to find an instrument for differences between the predicted and actual realizations of running variable, and develop a strategy for incorporating this instrument into identification.

The paper is divided as follows. Section 2 describes the context. Section 3 covers or methodology conceptually. In Section 4 we explain in detail how the strategy is applied to our setting, and Section 5 contains empirical results. A discussion in Section 6 concludes. The paper is accompanied by an Online Appendix.

2 Empirical Setting

About 2 million Venezuelans live in Colombia today, or about 3.6% of Colombia's population. Most of these immigrants arrived after 2016, as a result of the political, economic, and humanitarian crisis in Venezuela following two decades of the *Chavismo* regimes of Hugo Chávez and Nicolás Maduro.

As is typical in crisis-driven migration, a large share of the Venezuelans who migrated to Colombia fled their homes, often bypassing the formal migratory process. Venezuelans often cross the border on foot and without a passport, since passports had become increasingly difficult to obtain for middle-class citizens in Venezuela.

[4]See, for example, Cattaneo et al. (2016); AbdulkadIroğlu et al. (2017b); Abdulkadiroğlu et al. (2017a); Sekhon and Titiunik (2017); Frandsen (2017); Eckles et al. (2020).

The migration wave created a significant policy challenge in Colombia. The Colombian government reacted quickly, though sometimes unsystematically, as they tried to manage the boom in undocumented migrants. The policy responses from the Colombian government to the first migration wave are the focus of our paper.

2.1 RAMV: Census of Undocumented Immigrants

To better understand the migration problem, the Colombian government implemented the *Registro Administrativo de Migrantes Venezolanos* (RAMV), a census of Venezuelan immigrants, which ran from April to June of 2018. The primary goal of the census was to collect data useful for informing future policy. Although participating in the census would later become a prerequisite for the PEP visa we study, this connection had not been made (even in the planning of policymakers). At its inception, the goal of the RAMV census was to understand immigration and have data for the design of government activities.[5]

The RAMV used a massive public advertisement campaign to attract Venezuelan immigrants to voluntarily provide personal information.[6] The Colombian government explicitly stated that registering will *not* result in deportations or negative legal consequences. This statement was credible, Colombia has traditionally enjoyed a welcoming relationship with Venezuelans, and there is no history of mass deportations. Furthermore, the census was not advertised as a platform to receive work permits or any other legal benefit that would facilitate the migrants' stay.

To participate in the RAMV, migrants needed to appear in person at one of 1,109 authorized points in 413 municipalities geographically spread across Colombia, visualized in Figure 1. Most respondents were in large cities, such as Bogotá, Medellin, or Cali, and locations alongside the Venezuelan-Colombian border. Respondents required to show evidence of nationality, such as a Venezuelan national ID (*cédula de identidad*), a document that is much more common than a valid passport, and could be used to take the census.

The census successfully surveyed 442,462 undocumented Venezuelans belonging to 253,575 different households. This is approximately 75% of the undocumented migrants resided in Colombia at the time according to official government estimates (although the exact number could not be known). The RAMV census officially terminated on June 9, 2018.

[5]We confirmed in conversations with government officials who oversaw the process.

[6]The information sought included names, dates of birth, current addresses, municipalities of origin in Venezuela, dates of crossing, education levels, and job statuses, among other details.

Descriptive Statistics. The starting point of our data is the 331,646 immigrants that have a valid Venezuelan identification in our RAMV data, which is necessary to match to the Colombia business registry. Table 1 contains descriptive statistics of our sample. Seventy-five percent of RAMV migrants are between 15 and 64 years of age, and over 83% of this group has completed at least secondary education. Compared with the Colombian labor force, this group is younger and more educated.[7] At the time of the survey, 46.3 percent of working-age migrants were engaged in some level of employment in the informal sector. Our large sample provides statistical power to detect economically tiny differences when comparing these migrants. As such, we assess differences throughout this paper for practical or economic significance in addition to statistical significance.

2.2 Permiso Especial de Permanencia (PEP)

In July 2018 — one month after closing the RAMV census — outgoing President Juan Manuel Santos unexpectedly shifted immigration policy in his final days in office. Under a new decree, all Venezuelans who took the RAMV census would be eligible for an official visa authorizing their presence in Colombia. Specifically, any undocumented migrant was eligible to apply who both i) had previously registered in the RAMV (which had closed and did not re-open), and ii) had no criminal records or pending deportation orders. The policy was implemented as part of Colombia's *Permiso Especial de Permanencia* ("PEP") program.[8]

The initial permit was for two years. Although the government did not explicitly guarantee indefinite renewal, it left this possibility open and has not attempted to dissuade PEP holders from settling. In fact, renewals were processed for PEP holders whose permit had expired.[9]

Legal Rights Created by PEP. The PEP visa was a *de facto* a resident visa. Among other things, this granted PEP visa holders the following legal rights comparable to Colombian

[7] According to 2018 population estimates, 66 percent of the Colombian population is between the ages of 15 and 64, and 61.5 percent of the active labor force in 2017 had completed at least basic secondary education.

[8] The PEP, a special visa created for Venezuelans, was previously provided to *documented* immigrants in two prior waves. Registration for *PEP1* was August 2017- October 2017, and registration for *PEP2* was February 2018- June 2018. By contrast, the program launched in August 2018 (the focus of this paper, "*PEP3*" or "*PEP-RAMV*") focused on *undocumented* immigrants.

[9] In March of 2021, the Colombian government announced it will roll out a new visa valid for ten years, named Estatuto Temporal de Protección (Temporal Protected Status) for all Venezuelans in the country who need it, including those with expired PEP visas. After 2021, expired PEP visas were replaced by the ten-year TPS visa. The TPS visa is also renewable.

citizens:

a) Access to the formal labor market as a worker.
b) Full constitutional and civil rights including standing in the Colombian courts as an individual (both criminal and civil).
c) Individual access to the banking and borrowing systems.
d) Freedom from potential deportation and the right to remain in Colombia (physically).
e) Access to social services including national healthcare, education and welfare.[10]

The PEP visa did not create a new legal right to create and register a business. Foreign citizens in Colombia were *already* able to register a new business in Colombia before the introduction of PEP and continued to be. Rather, the bundle of rights granted by PEP was to individuals, covering individual-level freedoms as described above.[11]

Expanding the rights of individuals can affect their incentives to invest in businesses. Because there were no changes to business rights, the effects in this paper arrive through individuals' incentives to invest. Many of these incentives flow from legal rights through direct and indirect channels. For example: Having a PEP facilitates the opening of bank accounts and, with it, getting credit to invest. Similarly, by granting freedom from deportation, PEP provides a sense of stability to its recipients, which might translate into higher certainty about reaping the rewards of long-term investments (such as entrepreneurship). Similarly, many business partners in Colombia seek government identification from individuals as a precondition of doing business. Even if a business is legal and registered, a counterparty may desire documentation about the specific individuals involved in the transaction. Although this is not necessarily required by law, it provides comfort and protection for the counterparty. PEP facilitated these transactions for migrants by providing the informally sought documentation.[12]

Takeup and PEP Applications. Among the 442K respondents to the RAMV Census, 100% were eligible to apply for PEP. About 280K applied and subsequently obtained the PEP visa (64%). This magnitude is similar to that of other migratory amnesty programs around the world, where takeup has regularly been below 100% (even with longer time

[10] The visa also allowed immigrants to be scored by SISBEN, the test of means used to target social programs in Colombia. Low-scoring immigrant families with the PEP became eligible for Colombian government assistance.

[11] By granting individuals the freedom to work, PEP also granted businesses the right to hire them. However, this applied equally to all businesses in Colombia, irrespective of domestic vs foreign ownership.

[12] In informal conversations with some of the entrepreneurs in our sample, we verified that some of these mechanisms played a role in their decision to invest in their new businesses. At the same time, while PEP granted several formal legal rights, sometimes migrants did face difficulty exercising these rights. For example, some Colombian employers were not always familiar with the rights granted by PEP.

horizons to apply).[13]

2.3 Outcome Data: Colombian Formal Firm Registry (RUES)

To study business formation, we match all RAMV immigrants to the Colombia firm registry *Registro Unico Empresarial* (RUES). The RUES is a comprehensive firm registry of *formal* firms in Colombia. Note that in contrast to more developed economies, most self-employed Colombians do not register their firm as a formal firm in RUES and instead remain in the informal economy. For example, while the self-employed population is higher than 10% in Colombia, the rate of firm registration is less than 1%. This implies that any registration in the Colombia firm registry is at the higher end of the quality distribution of firms.

The Colombia firm registry includes several types of firms. About 75% of them are sole proprietorships (*personas naturales*). These are new companies managed by an individual that has not established a separate legal entity to create a business. Creating a separate legal entity in Colombia represents a higher cost than establishing an LLC or corporation in the United States since it requires working with an officially designed notary to set up a corporate contract and entails capital requirements.

The remaining 25% of firms in the RUES are independent legal entities called *sociedades* (societies), as in most countries belonging to French legal tradition. The Colombian government offers several types of *sociedades*, with two main ones. *Corporaciones*, are the strongest legal entity, similar to corporations in the U.S. They provide shareholder rights, they are better set up for complex corporate contracts, can list in the stock market, and must be created through a public notary. *Sociedad por Acciones Simplificada* (SAS) are a simpler legal entity somewhere between sole proprietorships and corporations. SASs cannot list in the stock market, can be created through a private agreement, and have simpler governance requirements that are ideal for a simpler type of firm. Roughly speaking, these different types of legal forms map to different levels of the underlying potential of companies. Sole proprietorships are less growth-oriented than SAS, which are in turn less growth-oriented than corporations.

We matched the Colombian firm registry to the PEP registration using ID numbers for

[13]For example: The estimated takeup rate for the Immigration Reform and Control Act in 1986 (a Regan-era program) has been estimated at 44% -77% (link). A history of the program notes "the number of applicants fell considerably below expectations" (Jasso, 1993). Similarly, the takeup rate for the Obama 2012 DACA program was approximately 70% of the 1.3 million young adults eligible for DACA (as of September 2018, five years after the program began Patler et al., 2019).

migrants. For corporations and SASs, we match only to the public legal representative of the firm. This is not the firm's lawyer, but the CEO or otherwise chief person ultimately responsible for the firm's management. For all firms in our data, we also obtained the number of assets and employment reported in the RUES each year, allowing us to consider the levels of these variables at founding.[14]

2.4 Batch Access to PEP using Census Numbers

Although all individuals in our sample became *eligible* to apply for the PEP upon the decree, there was variation in the *timing* in which they were allowed to apply. The PEP required an online application. To avoid overwhelming Colombia's immigration bureaucracy, the government split the migrants into batches. Each batch was assigned a starting date for migrants to register for PEP between August 15 and October 15 of 2018. Because the PEP applications were entirely online, applicants did not face congestion or queues (as they had earlier for RAMV). However, they had to apply within their designated window. All access to the PEP application system closed for RAMV migrants on December 21st, 2018. Migrants in earlier batches enjoyed not only earlier access, but also a longer time horizon to apply for PEP.

The Colombian government divided migrants into batches based on their RAMV census numbers. This method was chosen as a simple method to ensure that each new migrant family had a unique ID; it was *not* developed as an administrative tool to index the timing of migration or census participation.[15] A census participant occurring in Bogota may be assigned a number of N, while a separate family in Medellin could receive registration number $N+1$ even though they completed their census seconds later in a different city.

To assign a census number to a batch, the numbers were assigned to a sequence of 22 cutoffs. These cutoffs were nationally advertised so that immigrants in RAMV could use their census number to determine the date they could begin applying for the PEP visa. Online Appendix Figure A1 shows an example of these advertisements containing a table of cutoff numbers and start dates. Figure 2 visualizes the distribution of census numbers per week of registration to RAMV (each color symbolizes a week), and the 22 thresholds (dashed line) that were defined ex-post to link census numbers to PEP application windows.

[14] Online Appendix Section B describes the main characteristics of the firms in our sample.

[15] Although the number process was sequential, we do observe some gaps in our data when a census response is cancelled mid-way through the process. In addition, there was a software upgrade that fast-forwarded the census numbers, and resumed sequential assignment on the other side. Our specifications normalize, remove or otherwise control for this jump where necessary.

The figure shows how within a week individuals could be assigned to different application dates.

2.5 Preview of Identification

Our identification comes from the combination of the cutoffs and the census numbers. Before delving into the details of the natural experiment, we show some intuition and tests behind the strategy.

Surprise. The August 2018 announcement – both the program to award legal status to undocumented immigrants, and the choice to use the RAMV census as an eligibility criteria – were a surprise.[16] We present visual evidence of this in Figure A2. Google searches for RAMV in Colombia were rare (and flat) during the actual RAMV census period. During this time, the population viewed RAMV only as a census for migrants. When the link between RAMV and authorization was decided and announced, searches for RAMV spiked – even though RAMV had been closed for a month.

Our design compares outcomes among migrants who completed the RAMV. The surprise is an attractive feature, but not strictly necessary for our threshold-based design. Our identification is based on the noisiness of one's census number. Had PEP been widely known as an outcome of RAMV, registrants would still need to know the cutoff numbers in order to game the identification strategy. Ultimately, all RAMV registrants had the opportunity to obtain a PEP visa, but some had a longer window. Even if some migrants foresaw the future perfectly and strategized to obtain a long window, these would be classified as "always takers" in our setup. Our strategy is based on "compliers" whose PEP status was sensitive to the timing and length of the window.

Census Numbers. Census numbers are the first component of our identification strategy. Over the whole sample, a family's census number is loosely correlated with observable characteristics, as is the likelihood they get the PEP. There are possibly several factors that cause families to register early that may also cause them to be more (or less) entrepreneurial (and thus introduce a confound). In this sense, the census numbers are not random across the whole sample.

[16] A team of qualitative researchers for *Innovations for Poverty Action* similarly describe announcement as unexpected (Romero et al., 2021). Two prior papers study the impact of the *PEP-RAMV* policy on other outcomes using the surprise as an identification strategy (Bahar et al., 2021; Ibanez et al., 2020); both papers examine municipality level- outcomes and thus cannot use our family-level RD design. Bahar et al. (2021) studies effects on native (non-migrant) Colombian workers, and Ibanez et al. (2020) studies the crime-reporting behavior of the Venezuelan migrants.

However, the registration number also contains noise coming from a variety of sources. For example, a migrant who leaves at 8 AM to arrive in line may enjoy a higher or low number, depending on the number of other migrants registering on the same day, the random component of traffic, the weather (which is a central factor in our empirical setting) or other sources. In addition, the nationwide sequencing of the registration numbers also created noise. A migrant registering in Bogota could have a higher (or lower) registration number, depending on the number of Medellin migrants who decided to register around the same time, and/or the rain in Cali, or a bureaucrat's speed at processing migrants in Barranquilla. In our empirical study, we exploit this noise as an identification strategy for obtaining the PEP visa. Table A1 reports raw descriptive statistics for those who got the visa versus those who didn't.

Cutoffs. The cutoffs were set by the Colombian government *ex-post* in order to space out the number of individuals in each bin. Each bin contained approximately 4.5% of the sample. The average cutoff was relatively close to the previous cutoff, usually under three calendar dates from the previous cutoff. 28% of the cutoffs were on a day immediately following another previous cutoff. The cutoffs did not take into account boundaries in the calendar days in which a migrant registered for the RAMV. Most of the cutoffs appear in the middle of a day so that two migrants completing the census on the same date could later be assigned to different batches. Figures 2 and A3 show the distribution of census numbers across weeks and days, along with threshold markers. The cutoffs divide days and weeks at arbitrary points, and there is no visible bunching before or after cutoffs.

Figure 3 advances this analysis further by plotting observables across a stacked model of all cutoffs. We observe a smooth histogram of census numbers on both sides of the cutoff, and we do not see any differences in the observables we plot, including demographics, such as gender and age, the marital status of the person filing, or, crucially, their likelihood to identify as self-employed in the RAMV registration.

Precipitation. Finally, we introduce randomization into our approach by taking advantage of precipitation. We use data about the number of hours of rain each day and location in Colombia, and we link this to the timing of RAMV registration (and thus their census number). Then we simulate how different weather would have affected each migrant's timing (and thus each migrant's census number). For "different weather," we use weather at the same location and the same week that the migrant registered, but on a different day. This generates a distribution of potential census numbers for each migrant.

By focusing within the week and location, we condition on other factors including the overall weather in that week$\times$ location. Identification simply relies on the assumption that the distribution of rain with days of each week$\times$ location was as good as random. Taken together, these features of our setting have the following implication: because cutoffs are relatively close together in time, and are agnostic to calendar boundaries, small changes in a migrant's queuing order could randomly change the date in which they can apply to PEP. The noise in the queue numbers induced by rain is large enough that many migrants have a non-zero probability of landing in more than one batch, and may even cross multiple cutoffs, depending on the noise draw.

3 Identification Strategy: Running Variable Instrument

We now formalize our strategy for estimating a causal effect. While our strategy is motivated by our setting, we present it as a more general empirical problem where treatment assignment depends on a threshold (as in a regression discontinuity), but the running variable is affected by noise. At the end of this section, we discuss potential applications in other settings.

Our approach builds from the notion of a *composite treatment*. A composite treatment is a treatment computed from multiple sources of variation, according to a known formula. Borusyak and Hull (2021) propose design-based theory and methods for composite treatments in which some – but not all – of their determinants are generated by a true or natural experiment. These new methods specifically address empirical settings where some inputs to the composite treatment are highly endogenous, and other inputs may be influenced by quasi-random variation.

Our approach can be seen as an application of these ideas on a policy discontinuity.[17] We provide one approach to this adaptation in which shocks arise from measurement noise along the running variable. Strictly speaking, our estimator is not a regression discontinuity (RD) in the traditional sense (Lee and Lemieux, 2010) because we are not comparing subjects strictly within a narrow window around the threshold. Nonetheless, we still use term RD, consistent with other papers that use the thresholds for treatment assignment, including those that use observations outside a narrow window (Angrist and Rokkanen, 2015; Rokkanen, 2015). One attractive feature of our approach is that although our approach places higher weight on observations closer to the threshold, it also allows

[17] In fact, the Borusyak and Hull (2021) write, "Policy discontinuities, as commonly used in regression discontinuity designs, can similarly justify local permutations of shocks."

observations farther from the threshold to have some weight as well (depending on how the running variable is affected by noise).

3.1 Data-Generating Assumptions

We begin by laying out a set of assumptions from which we build a strategy for estimating treatment effects. Ultimately, these assumptions will generate intermediate results that map to other, well-known approaches (in particular, propensity score adjustment and IV). At that point, we simply argue that our setup inherits the properties of these estimators.

Assumption 1 (Preliminaries)**.** *There are $i = 1, ..., n$ realizations. Each realization is*

$$\{Y_i(0), Y_i(1), Z_i, C_i^1, ..., C_i^k, U_i^1, ..., U_i^j\} \in \mathbb{R}^{2+k+j}.$$

Each realization includes k observable variables $\vec{C}_i = \{C_i^1, ..., C_i^k\}$ and j unobservable variables $\vec{U}_i = \{U_i^1, ..., U_i^j\}$. Z_i is an instrument and $Y_i(\cdot)$ is a function that takes the value of $Y_i(1)$ in the treated state of the world and $Y_i(0)$ in the untreated state of the world.

Assumption 2 (Running Variable Construction)**.** *Nature constructs a running variable X_i through a function M, mapping Z_i, observables, and unobservables. $X_i = M(Z_i, \vec{C}_i, \vec{U}_i)$, where $M : \mathbb{R}^{1+j+k} \rightarrow \mathbb{R}$.*

One interpretation for Z_i is that it is random measurement error. Running variables are often a measurement — e.g., estimates of scholastic aptitude (Jacob and Lefgren, 2004), or biometric measurements such as a blood test (Hansen, 2015). In these settings, measurement technology has finite precision and may therefore incorporate a degree of noise from the environment, calibration, equipment, measurement technique or individual administering the measurement. Z_i may capture noise in the measurement X_i coming through these sources. We do not require M take any particular functional form. But we do impose some conditions on Z_i.

Assumption 3 (Monotonicity)**.** *M is monotonic in Z_i, i.e., $\partial M / \partial Z_i \geq 0$.*

Assumption 4 (Sharp Treatment Discontinuity)**.** *Units are assigned a treatment D_i through a function $D(X_i) = \mathbb{1}(X_i > b)$, where b is a known boundary along the running variable X_i.*

3.2 Exogeneity Assumptions

We now add two exogeneity requirements.

Assumption 5 (Treatment Exogeneity). *Z_i is exogenous to D_i conditional on X_i; i.e.* $[D_i \perp\!\!\!\perp \mathcal{Z}_i]|X_i\}$.

Assumption 6 (Outcome Exogeneity). *Z_i is exogenous to Y_i conditional on X_i and D_i; i.e.* $[\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp \mathcal{Z}_i]|X_i, D_i\}$.

Assumptions 5 and 6 require that Z_i does not affect treatment assignment (or outcomes, 6) — except through the impact on the running variable X_i.

Definition 1 (Running Variable Instrument). *A variable Z_i is a* running variable instrument *for X_i if it meets Assumptions 2-6.*

A running variable instrument ("RVI") is different than a normal instrument, but meets similar exogeneity requirements. Rather than instrumenting for treatment, it instruments one of the inputs to treatment assignment (the running variable). Like typical instruments, RVIs will also necessarily be local. Only a subset of units (RVI compliers) may have their running variables altered by the RVI. The treatment effects revealed by RVIs will necessarily be local to this change.

For any given X_i running variable, there may be many possible instruments. The approach we describe here simply requires that the researcher know a single RVI. Because we know the treatment-assignment rule (the threshold), we can later map how the RVI affects treatment assignments.

3.3 Observability Assumptions.

We now add three additional observability assumptions.

Assumption 7 (Observable Variables). *For each observation, researchers observe* $\{Y_i, X_i, D_i, Z_i, C_1, ..., C_k\}$ *where* $Y_i = Y(D_i)$.

Assumption 8 allows each observation to have a Z_i drawn from its own distribution. The running variables for some observations could be more affected by noise than others.

Assumption 8 (Known Distribution of Z). *The researcher knows the joint distribution from which all Zs were drawn, F_Z.*

Assumption 8 requires that researcher knows the distribution of Z_i. It could be estimated from data by the researcher, or known from prior research. For many measurement technologies – including standardized tests, biological specimen tests (e.g., blood tests), and many physical measurement tools – the developer documents the levels of precision and measurement noise.[18] In other settings, repeat measurements may allow researchers to estimate the distribution of Z_i themselves. Lacking such prior knowledge, researchers may also build redundant measurements into a research design to generate knowledge of measurement errors or other Zs.

Assumption 9 (Known M Map)**.** *The researcher can map all possible realizations of Z_i in the support of F_Z to X_is.*

In some cases, it may be sufficient to add or subtract measurement noise to each unit's fixed average. Using knowledge of how each potential Z_i maps into a potential X_i, the researcher can compute the entire distribution of potential X_is for each observation. We can now turn to results.

3.4 Results

Our assumptions so far raise the possibility of *counterfactual* X_i running variable realizations. Had the Z_i draw been different for any of the n possible realizations, the resulting observation of the running variable X_i would be different.

Lemma 1 (Potential Running Variables)**.** *The distribution of potential running variable for each observation i has the CDF of $F_i^X = F_i^Z(M^{-1}(z; ...))$.*

Using F_i^X, we can compute the average X realization for each i, which we denote as $\bar{X}_i$. This is potentially useful as a control variable insofar as it proxies for non-exogenous reasons that an observation i lies (on average) close or far from the threshold. In some settings, the mean value $\bar{X}_i$ can could be interpreted as the "true value" of X_i (as distinct from X_i, which is inherently measured with noise).

Definition 2 (Local Propensity Score)**.** *For each observation i, the probability of treatment is $P_i^D = \Pr(D_i = 1) = 1 - F_i^X(b)$.*

Definition 2 essentially creates a design-based propensity score that is inherently local to the instrument. P_i^D can be interpreted as probability that Z (the RVI) places the running

[18] The creators of standardized tests (and other survey based psychometric scales) use item response theory (Embretson and Reise, 2013) as a principled way to assess the precision of their metric through, and most scales or standardized tests publish diagnostics for precision.

variable above the cutoff. The F_i^Z distribution in Definition 2 is calculated from Lemma 1, and the known threshold is b.

This setup differs from other approaches of estimating p-scores from data that lack a source of random variation or a structured model of treatment assignment. Our setup uses the structure of the treatment (cutoffs, Assumption 4) to estimate a distribution of running variables (Lemma 1) and the resulting probability of being treated (Definition 2). This estimated distribution is based on more primitive assumptions (1 through 9).[19] We label this propensity score "local" because depends on the instrument Z_i. Different running-variable instruments may induce different propensities.

Proposition 1 (Sharp Discontinuity)**.** *Given the assumptions above, the treatment effect is* $\tau_{Sharp} = \mathbb{E}[Y_i(1) - Y_i(0)|P_i^D, \bar{X}_i]$.

We call this version "sharp" because the treatment goes from zero to one when the realized X_i crosses the threshold. We can use the distribution of F_i^X to reason about the probability of this happening (Definition 2). The sharp version can be implemented using a regression that controls for the propensity score P_i^D and $\bar{X}_i$ (and possibly other pre-determined controls $\vec{C}_i$).[20] Controlling for the propensity score P_i^D in addition to $\bar{X}_i$ and C_i is particularly helpful when the distribution of F_i^X are not identical, and thus observations with the same $\bar{X}_i$ and other controls could have different propensities P_i^D.[21] In our measurement error interpretation of Z_i, this would apply if measurement error varied across different observations (i.e., some units are able to be measured more precisely).

As in other research designs, conditioning on the propensity score eliminates selection bias coming from the conditioning variables and potential outcomes. However, in our case, the propensity score does more than address selection bias. It also allows us to use all observations for evaluation whose treatment propensity lies between zero and one. That is, use the largest set of observations for which we have partially randomized assignment (coming through the running-variable instrument).

[19] This approach is similar to Abdulkadiroğlu et al. (2017a), which estimating the propensity scores using a known structure (the deferred acceptance algorithm) and random tiebreaking for indifferences.

[20] Controlling for propensity scores by regression adjustment differs from how propensity scores were originally used (e.g., to weight observations, Rosenbaum and Rubin 1983, or in matching, Abadie and Imbens 2016; King and Nielsen 2019); a classic exception of is Robins et al. (1992). However several recent papers have documented attractive theoretical properties of propensity scores as controls (Abdulkadiroğlu et al., 2017a; AbdulkadIroğlu et al., 2017b; Borusyak and Hull, 2021). In addition, a research note by Hull (2018) explicitly addresses propensity score controls by extending earlier results from Angrist (1998) where propensity scores are linear in other controls.

[21] it may also be useful if the propensity score is a non-linear function of other controls C_i, and thus controls for C_i would not be colinear with the propensity.

This addresses two key weaknesses in typical regression discontinuity designs. First, limiting the sample to this window significantly reduces the sample size and statistical power of the analysis. This is a chronic challenge with RD designs (Schochet, 2009; Deke and Dragoset, 2012). Second, in typical RD, the use of windows limits the external validity of the estimates since the treatment effect is identified at $X_i = c$. In our setup, the treatment effect is measured with respect to a broader population; specifically, it estimates a convex average of conditional causal effects across observations whose propensity scores are between zero and one.

The estimate coming out our approach is still not a population-wide average treatment effect. The resulting coefficient would still be variance-weighted (Angrist, 1998), which in this context would mean placing more weight on observations with p-scores closer to $\frac{1}{2}$, likely closer to the threshold. However, it is a more broad and potentially diverse population than typically inside the RD window, generating potential improvements in precision. It also grants researchers greater flexibility to examine sub-samples of the data, including sub-samples of populations whose average X_i is farther from the cutoff (but with propensity scores between 0 and 1).[22]

Visualization. Figure 4 illustrates our approach. Each observation i is represented as a vertical bar placed over the observation's realized running variable (on the X axis). The lower and upper bounds of the vertical bar represent the 95% confidence interval for the *potential* running variable for each observation (i.e., F_i^X as defined in Lemma 1). The horizontal line at a value of 140 represents the location of the threshold. The portion of each subject's line above this bar is their local propensity score P_i^D as defined in Definition 2. As the figure demonstrates, many observations have a positive probability of being treated (including those who aren't).

A visualization like Figure 4 is meant be used alongside other typical RD visualizations such as the treatment discontinuity at the boundary. As the expected running variable increases, the probability of treatment increases smoothly – even as treatment status changes discontinuously at the threshold based on the realized number.

[22]Concerns about the sensitivity to the bandwidth are also emphasized in Calonico et al. (2014), among others.

3.5 Fuzzy Discontinuities

Running variable instruments can also be used for fuzzy thresholds. This is the version used in our empirical application. In the fuzzy version, treatment does not go from zero to one when the noisy measurement (X_i) crosses the threshold. Instead, crossing this threshold simply increases the *probability* of being treated.

As in our sharp version, we can reason about the probability of crossing the threshold by applying our primitive assumptions. Some additional modifications to the setup are required. We also use the resulting probability in a different way. The following assumptions supersede Assumption 4.

Assumption 10 (Fuzzy Treatment Discontinuity)**.** *Units are assigned a* treatment instrument $\mathcal{Z}_i = \mathbb{1}(X_i > b)$*, where b is the known cutoff along a running variable X.*

We will call $\mathcal{Z}$ the treatment instrument or TI, in order to differentiate it from the running variable instrument (Definition 1). Note that $\mathcal{Z}$ is downstream/induced by the RVI. We also need an additional restriction about the TI:

Assumption 11 (Treatment Instrument Monotonicity)**.** *For a subset of the n observations (compliers), $D_i = 1$ if $\mathcal{Z}_i = 1$ but $D_i = 0$ if $\mathcal{Z}_i = 0$. There are no observations (defiers) for whom $D_i = 0$ if $\mathcal{Z}_i = 1$ but $D_i = 1$ if $\mathcal{Z}_i = 0$.*

The remainder of the assumptions remain. In the sharp design, the probability that the X_i crosses the threshold was the propensity score (i.e., propensity of treatment). In the fuzzy design, this probability is the "expected instrument" (Borusyak and Hull, 2021).

Definition 3 (Expected Instrument)**.** *For each observation i, the expected instrument is the local probability of crossing the threshold. $P_i^Z = \Pr(Z_i = 1) = 1 - F_i^Z(b)$.*

As before, the F_i^Z in Definition 3 is calculated from Lemma 1, and the known threshold is b. The expected instrument captures the idea certain observations have a greater average likelihood of an X_i landing above (or below) the threshold. In the sharp setup, this is the propensity of being treated. However, because of imperfect compliance (Assumption 11 rather than 4), observations with a high probability of exceeding the threshold may not always be treated.[23] In a fuzzy setup, the expected instrument is the propensity of falling above the threshold (and thus the propensity of getting a $\mathcal{Z} = 1$ treatment instrument).

[23]Under the monotonicity requirement of IV, observations with a high P_i^D will be more likely to be treated than those with a low P_i^D.

From here, we can compute an estimand by using the usual IV assumptions (Angrist et al., 1996). In settings like ours in which the treatment is partially randomly assigned (via $\mathcal{E}_i$ realizations) but not entirely, Borusyak and Hull (2021) discuss the need to control for the expected instrument. Given this, we can estimate a treatment effect on the compliers.

Proposition 2 (Fuzzy Discontinuity)**.** *Given the assumptions above, the treatment effect on the compliers can be estimated by:*

$$\tau_{Fuzzy} = \frac{E[Y_i|\mathcal{Z}_i = 1, P_i^D, \bar{X}_i] - E[Y_i|\mathcal{Z}_i = 0, P_i^D, \bar{X}_i]}{E[D_i|\mathcal{Z}_i = 1, P_i^D, \bar{X}_i] - E[D_i|\mathcal{Z}_i = 0, P_i^D, \bar{X}_i]} \tag{1}$$

This is simply the IV estimator using 2SLS, conditioning on P_i^D and $\bar{X}_i$. As with before, this enlarges the set of observations that can be included in estimation beyond the window typically appear in fuzzy RDs. This is not a population-wide causal effect, but a convex average of conditional causal effects for compliers.

3.6 Moving Towards Applications

Before we discuss our implementation of this design in Colombia, we summarize a few remarks about this methodological approach for a broader context.

First, our setup also allows discrete running variables. Indeed, our applied application is technically a discrete variable (census numbers, or natural numbers). Second, this setup could be applied in settings where subjects can manipulate the running variable – assuming they cannot manipulate it precisely enough to eliminate all possible instruments Z_i.

Algorithms. Finally in many RD papers, a cutoff is applied to algorithmic prediction or evaluation serving as a running variable (e.g., Narayanan and Kalyanam, 2015; Barach et al., 2019). The algorithms may themselves be noisy measurements. If these algorithms were trained on a finite, randomly selected sample, then they would contain "measurement noise" arising from finite sampling. If researchers can quantify this measurement error, either analytically, through cross-validation, or something else, then our strategy can be applied to settings featuring algorithmic running variables.

4 Operationalizing our Design

We now apply the estimation strategy above to study Colombia's *Permiso Especial de Permanencia* (PEP) program (described in Section 2). We are interested in measuring the effect of the PEP visa on migrants' entrepreneurship choices. We begin by mapping the key elements of our research question into the language of our design.

- Units: Each observation i is a migrant. Our main results are cross-sectional.
- Treatment: The treatment D_i is having a PEP visa before December 21, 2018.
- Outcome Y_i. The outcome is migrant i's choice to become an entrepreneur and registering a firm in the formal business registry by July 2022.

We now outline the components of treatment assignment by the threshold:

- Running Variable X_i: RAMV census number.
- Thresholds b: The 22 thresholds in the census numbers granting more (or less) time.
- Estimator: We use the fuzzy discontinuity (Proposition 2). Falling below a threshold makes it easier to obtain a PEP, but the migrant still has to apply.

Finally, we outline our strategy for using the RVI.

- Running Variable Instrument Z_i: Local weather conditions around the time of the migrant's registration Z_i.
- Treatment Instrument $\mathcal{Z}_i$: The treatment instrument is the *window length*, set by the realization of census numbers. $\mathcal{Z}_i$ is the amount of time a migrant was given to register for PEP.

In the remainder of this section, we describe how we operationalize our design and meet the assumptions laid out in Section 3.

4.1 Running Variable Instrument

In our context, we need an instrument that affects the assignment of census numbers. Our instrument consists of rainfall in the local context of each migrant's registration week$\times$ location. Because census-taking required in-person registration at specific locations, the timing of participation may be sensitive to weather conditions. Other papers have used local weather shocks as instruments for participation in in-person events such as elections (Gomez et al., 2007), protests (Madestam et al., 2013; Wasow, 2020), or other outdoor activities (Shenoy et al., 2022).

We measure rainfall using data from Visual Crossing, a leading provider of weather data for both commercial and scientific use. The data measures, within each 24-hour period, a value that spans 0 and 100 in a particular Columbian city, and represents the percentage of time during the day that precipitation occurred.[24]

Table A2 shows the relationship between rain during registration dates and census numbers. The key regressions control for date fixed effects (which controls for the overall level of precipitation throughout Columbia on each day) as well as municipality fixed effects (which absorbs the average weather each city). When there is rainfall in a municipality, census numbers are higher (i.e., subjects on that day are further back in the national queue for that day).

We speculate that this relationship is driven by delays in arrival, possibly as a result of slower and/or more congested transportation options to the RAMV locations. Census takers in rainy cities are delayed, and thus arrive at their local census locations after those from sunny locations. On average, a one standard-deviation increase in precipitation coverage corresponds to about 450-466 spots later in the RAMV census number. However, we also find significant heterogeneity in how responsive subjects are along demographic and other dimensions.

Although the census numbers contain a random component, the migrants' exposure to this shock is not entirely random. Migrants may have chosen a particular time to register for endogenous reasons that correlate with later behavior. To capture this aspect of the running variable, we need to use the next parts of the setup.

Distribution of the Running Variable Instrument (Z_i). Assumption 8 requires the researcher to know and use F_i^Z, the distribution of the running variable instrument Z_i. In our context, this means we need to know the distribution of potential weather for a given day at a potential location.

To this distribution of potential weather realizations, we simply permute the observed weather within a small local area and window of time. Specifically, we develop potential rain realizations by drawing 200 random samples. In each sample, the realizations of rain within a week$\times$ city cell are permuted. For example, if a city had two days of rain in week #2 (Monday 50% and Tuesday 75%, and 0% on all other days), we would randomly permute which days of the week the 50% and 75% occurred (within the same calendar

[24]We obtain our data from `https://www.visualcrossing.com/resources/blog/what-is-precipitation-coverage/`. We used this variable because it seemed to be a good proxy for transportation difficulty for a given day$\times$location.

week and location).

The permutation process produces a set of 200 counterfactual weather realizations, each containing a separate permutation for 8,151 city-by-week groups.[25] In all permutations, the average weather is the same within each week$\times$city cell. However, the order of the rain dates have been swapped.

4.2 Mapping Potential Rainfall to Census Numbers

Assumption 9 requires that we know the map between potential weather realizations (above) and the running variable X_i. To create this map, we proceed in two steps. First, we fit a model of Census numbers from the realized local weather — i.e., the actual weather that occurred in the data.

Then, we apply this fitted model to a series of new input data to generate prediction from the model using the new data. The new input data is identical to the original training data, but the local weather realization is replaced with one draw of the 200 counterfactual weather permutations (described above). Through this procedure, we generate a distribution of 200 potential census numbers that each migrant could have taken, depending on different potential weather realizations. This is F_i^X in our notation above.

Model Details. The model we fit takes the following form. k indexes each registering migrant family in our sample.

$$CensusNumber_k = \beta_0 + \beta_1 Precipitation_k + \gamma' \mathbf{X}_k + \alpha_{m(k)} + \rho_{d(k)} + \mu_k \tag{2}$$

where $CensusNumber_k$ represents the family's RAMV census number; $Precipitation$ represents the intensity of rain — as measured by hours rained in the day of registration. $\alpha_{m(k)}$ are municipality fixed effects, $\rho_{d(k)}$ are date fixed effects, X_i is a vector including average family age, share female, share pregnant, number of people in the unit, and number that are direct family members. 64% of families had exactly one member, and μ_i is random noise. The observables and date fixed effects are also interacted with precipitation in that day to better incorporate heterogeneity across families. By controlling for day and city fixed effects, we focus only on within-day, within-city variation in RAMV numbers (i.e.,

[25] Our study contained 11 weeks in 741 municipalities. Weather data was unavailable for six municipalities.

earlier/later within the same day).[26]

Robustness and Alternative Specifications. We also make a few adjustments as robustness checks. Equation 2 is a linear model, but census numbers are an ordinal variable. As such, we re-ordered the predicted census numbers within each simulation so that they would have the same values as the original census numbers (but mapped to families differently). The results presented use the re-ordered version, and our results were similar either way.

In addition, the *residual* between the prediction from Equation 2 and the migrant's actual census number main contain may contain relevant unobserved heterogeneity. As such, we saved these residuals — and in some specifications, we added them back onto the census number prediction from Equation 2. The results below include the residuals, but their inclusion did not affect the findings.

Finally, we add some additional interactions. We call Equation 2 the *simple* model. We repeat the same process for two expanded variations of Equation 2. The first, which we call the *Main* model, also incorporates the square of precipitation in each day as both an independent term and interacted with other observables (including the city fixed effects). This captures the idea that rain may impose convex costs that vary across different types of migrant families (including those in different locations). In our *Expanded* model, we additionally include a binary term indicating whether there was any precipitation at all that day, which is also interacted with observables. This accounts for the possibility of having any rain (at all) introduces fixed costs that could vary by family type.

Visualization. Figure 5 implements our visualization in our context, displaying the idea that each migrant has a distribution of potential census numbers in. Migrant families are indexed by the X-axis (listed left to right based on their actual census number). Each vertical line spans potential census numbers (the inter-quartile range) from the *Main* model based on the 200 simulations. The horizontal lines represent cutoffs. As the figure shows, the census number can vary significantly across local weather shocks. In addition, the weather shocks affect some subjects more than others. Each subject has a F_i^X with different variance.

[26]Weather shocks could have also induced some subjects to register in different days or locations. As such, the within-day noise in the running variable we leverage is a lower bound for the total amount of weather-induced noise.

4.3 Expected Instrument

We used the approach above to generate a distribution of potential census numbers. We now convert the counterfactual census numbers into to a time window for each migrant. The window lengths are important because they give migrants more (or less) time to become documented (and thus affect PEP applications).

Within each of the 200 permutations, we map each migrant's predicted census number census numbers to the cutoffs defined by the Colombian government. For each migrant, we take the average window length across the 200 permutations. This average is the "expected instrument" in Definition 3, and needs to be calculated and controlled for in our final IV estimation (Proposition 2). As Figure 5 shows, the rain-induced variation sometimes causes migrants to cross multiple thresholds. In this case, the expected instrument will average over more than two possible window lengths.

We now have all the ingredients for our specifications. Our main specifications estimate the impact of getting the PEP on entrepreneurship. Per Proposition 2, we use an IV and two stage least squares (2SLS).

4.4 Specifications

The first stage of our 2SLS specification is below. Migrants are indexed by i.

$$PEP_i = \gamma_0 + \gamma_1 \underbrace{ActualDelay_i}_{\text{Treatment Instrument } \mathcal{Z}_i} + \gamma_2 \underbrace{ExpectedDelay_i}_{\text{Expected Instrument}} + \bar{X}_i + OtherControls_i + \eta_i \tag{3}$$

Each migrant i's treatment status (PEP_i) is instrumented with the delay (in days) experienced by a migrant i. However, as we have already discussed, the $ActualDelay_i$ is not entirely random. As such we control for the "expected instrument" (Borusyak and Hull, 2021), which is the $ExpectedDelay_i$ (generated by the procedure above).

$OtherControls_i$ are the migrant's family role (head of household or not), marital status, occupation, gender, level of education and fixed effects for week of census completion and location of registry for the migrant.[27] $\bar{X}_i$ is the average census number across the 200 permutations, and η_i is an error term.

[27] After controlling for the expected instrument, the $OtherControls_i$ terms do not significantly change our visa treatment coefficient.

The identifying assumption is that the actual window length ($ActualDelay_i$) is a random draw from larger distribution of potential window lengths (based on local weather shocks) whose mean is $ExpectedDelay_i$. Both can be calculated using the policy cutoffs discussed in Section 2.4 (and shown in the public advertisement in Online Appendix Figure A1).

Our second stage equation is:

$$StartBusiness_i = \beta_0 + \beta_1 \underbrace{PEP_i}_{\substack{\text{Instrumented} \\ \text{by Eq. 3}}} + \underbrace{ExpectedDelay_i}_{\text{Expected Instrument}} + \bar{X}_i + OtherControls_i + \eta_i \tag{4}$$

where ϵ_i is an error term, and the $ExpectedDelay_i$ is again the expected instrument.

4.5 Balance and Profile of Compliers

Table 2 compares migrants that have a lower-than-expected delay, which we denote as those having "positive luck", to those that have a higher-than-expected delay (e.g., "negative luck"). The instrument appears balanced. The two groups are observably similar across all their characteristics, and all differences are economically very small and statistically insignificant.[28]

Next, we characterize the compliers to our instrument compared to others, particularly the always-takers. The compliers are the people who would obtain the residency visa – but only if they have enough time. The always-takers are the migrants who get the PEP visa, irrespective of the shocks to their timing.

In Table 3, we implement method of Marbach and Hangartner (2020) for profiling compliers. In essence, this procedure uses the means of always-takers (migrants getting the PEP even if unlucky), never-takers (migrants not getting the PEP even if lucky), and the whole sample, to back out the mean of compliers. We use a binary version of our treatment, where the individuals with above-median positive luck (within those with positive luck) are considered treated, and the rest are not. Table 3 reports the comparisons for the *Main* version of our instrument, but the results are similar for *Simple* and *Expanded*.

Compliers in this binary instrument are observably different from always-takers across many dimensions. Always-takers are more likely to be single, higher educated, employed,

[28] To compute the differences we follow Imbens and Rubin (2015) and Mckenzie (n.d.). In particular, the fourth column of Table 2 reports the difference in means between the "lucky" (treatment) and "unlucky" (control) groups, divided by the square root of half the sum of the treatment and control group variances.

and have family in Venezuela, among others. There are no differences across age or gender. The differences are meaningful but still leave substantial variation in observables to fall within different sides of the distribution.

5 Results

5.1 First Stage Regressions

We now move to the centerpiece of our analysis, the impact of receiving the PEP visa on the probability of starting a formal firm. Table 4 contains the first stage regressions of equation 3. Standard errors are clustered by family.

Column (1) reports the raw correlation between getting the PEP and *Actual Delay*—the difference between the date this migrant can begin registration and the earliest available registration date for all migrants. The point estimate is negative, and its value is -0.0053. This is consistent with our intuition: the longer the delay, the lower the chances are the immigrant will get the PEP visa.

Columns (2) to (4) are the first stage regression for three different versions of our instrument explained in Section 4. We now examine the actual delay conditional on the expected delay. Column (2) presents expected delay calculated by the *Simple* model Column (3) uses expected delay based on the *Main* model, and Column (4) presents the *Expanded* version. All columns include fixed effects for the week of registration, the role of the migrant in the family (e.g., spouse, head of household, grandmother, etc.), marital status, occupation, level of education, and the municipality of registration are included.

The partial F-statistic of these regressions is significant with values is higher for the models using more complex models. The coefficient for *Actual Delay* is positive and significant with a value (in the main estimate) of -.0031. Controlling for the expected delay, having one day less of time to get the PEP translates to a 0.31 percentage points lower probability of getting it; one week less to a lower 2.2 p.p. probability; and one month less to 9.3 p.p. lower probability. Relative to the mean of 68%, these changes are meaningful, at 0.5%, 3.2%, and 14%, respectively.

These first-stage regressions also represent our first substantive result: increasing the available time to do a regulatory transaction matters.

5.2 Main Results

Table 5 contains our main results using a linear probability model. Standard errors clustered at the family level.

Before discussing our IV results, Column (1) contains a reduced-form regression predicting $StartBusiness_i$ from getting the PEP (with no instrument). The coefficient of PEP is 0.0013 and significant. As we shall see, this underestimates the effect we get from our IV strategy.

This difference exhibits some of the challenges of estimating the effect of legal status without an IV. Suppose that highly risk-averse individuals may be drawn to register for PEP. However, risk aversion is also known to be negatively correlated with entrepreneurship (Puri and Robinson, 2013). Similarly, mothers with children or senior citizens may place a higher value on getting the PEP due to a desire for a stronger social safety net. However, these groups are not naturally the most likely to start firms.

More generally, if the majority of migrants register for PEP for reasons that are *negatively* correlated to entrepreneurship, then OLS could produce a downward bias. These possibilities are likely in our setting, where the compliers to our instrument are a smaller percentage of the whole population.

Columns (2) through (4) report 2SLS regressions using each version of our strategy. Column (2) reports a coefficient of 0.016, and Column (3) reports 0.018, both significant. Column (4), the expanded instrument, has a slightly lower point value of 0.012 and is more precise, with the standard error halved. Getting the PEP leads to an increase of 1.2 to 1.8 percentage points in the probability of starting a firm. This effect is substantial. The effect is over ten times higher than the mean of the outcome variable, at 0.16%.[29]

5.3 The Impact of PEP by Year of Registration

We now consider the evolution of the benefit of PEP over time in Figure 6. We report the coefficient of independent regressions using our approach where the dependent variable is 1 only if a firm is started in each specific year. We compare this to the Colombia native

[29] A natural question that arises in this context is what is happening in formal labor markets for these Venezuelan immigrants who receive the PEP, who now can also work as formal employees. Bahar et al. (2021) provides evidence that following amnesty, Venezuelan immigrants increase their participation in formal labor markets, but the effects are economically negligible. Based on this evidence, we believe access to formal employment is not significant enough to play a role in confounding our results.

population rate, which we estimate using our firm formation registry and annual public estimates of the Colombian population from the World Bank Data Commons.

Three facts are apparent in the figure. First, the impact of PEP on starting a firm in 2017 or 2016, before PEP, is zero using our approach. This makes sense and serves as a placebo test. The receipt of the PEP using our instrument should not predict any entrepreneurship before the RAMV registration (and therefore the rain-induced variation) has occurred.

Second, the effect begins on a positive trend that appears larger as each year passes. While PEP only increases 2018 business formation by 0.2 percentage points in 2018, this number raises to 0.5 by 2021, and 0.75 percentage points by 2022. These last two coefficients are meaningful relative to the rate of native formal firm formation in Colombia, which we estimate at 0.7% in 2022.

Finally, the results shown in this figure also deal with a potential concern in our setting: that immigrants might be simply formalizing firms that already exist but are not registered. In such cases getting the PEP would not have boosted entrepreneurship, only formalized it. However, if our results were driven only by formalization, we would expect to see a sharp increase in firm registration in the year of the PEP roll-out that gets smaller over time. Instead, we see an effect that grows over time.

5.4 Panel Results

We now turn to studying the impact of the PEP in a panel format. This approach has two benefits. The first is that it allows us to incorporate a significant number of fixed effects that may be a concern for endogeneity, including fixed effects for each person and for the time of entry into Colombia. This is a fundamentally different set of identification assumptions, and therefore an alternative approach for comparison with our IV/RD strategy.

Second, the panel approach allows us to consider a different margin — the entry into Colombia — and compare the difference in startup formation between two alternative channels. These two channels are the *physical relocation effect* and the *visa effect* (for the receipt of a legal residency permit). To separate these effects, we use a conservative assumption about the firms created before PEP. Our assumption is that 100% of firms that appear in the business registry before PEP (and created by RAMV migrants) are also net new firms. In other words, none of these firms is brought from Venezuela.

Of course, in reality some of the firms are not new. By adopting this assumption, we will generate an *overestimate* of the true effect of physical relocation and *underestimate* the effect of PEP by comparison. To the extent that this assumption is not true, and migrants are simply re-registering their Venezuelan firms in Colombia, the induced bias would make the effect of physical migration look larger than it truly is, and the effect of PEP, in consequence, relatively smaller.

We create a quarterly panel that begins in Q1 of 2014 up to Q4 of 2021. Our regressions include indicators for both their quarter of arrival into Colombia (which is heterogeneous across migrants) and the quarter in which they get the PEP (Q4 of 2018). We evaluate the impact of these changes impact on starting a firm using a difference-in-differences setup. We consider several specifications that may incorporate controls, pre-trends, and individual and time fixed effects.

Table 6 contains our results. Our preferred is column (2) which includes fixed effects for the week of registration and the municipality of registration, but does not individual or quarter fixed effects. We prefer this specification due to a concern that a binary outcome and treatment that only moves from 0 to 1 may not lend accurate estimates with individual fixed effects (even though results are similar after including them).

The estimate of physically migrating to Colombia on the probability of starting a firm is *half* of the magnitude of receiving the PEP, a difference that is statistically significant. The results suggest that legal rights of the migrants are about twice as important as the physical act of migration for entrepreneurship.

Similar results can also be visualized in Figure 7. Panel A plots the differential effect between immigrants who received and who did not receive the PEP in their propensity of creating firms quarter by quarter, with quarter 0 being the beginning of the PEP rollout, which is consistent with the main results and Figure 6. Panel B presents the trend of creating firms on immigrants, quarter by quarter, after entering the country, using the same scale in the vertical axis. The two graphs show that, indeed, the estimated effect of creating firms due to the PEP is significantly larger than due to relocation.

5.5 Heterogeneity and Mechanisms

Finally, we dig deeper into the mechanisms by studying heterogeneity in our results across migrant characteristics and the types of firms created.

Table 7 splits the dependent variable by the legal form of the type of firm created. In Colombia (as in the U.S.), there are several legal types of firms that can be created. The simplest one is a sole proprietorship (*persona natural*). This type of firm represents the independent economic activity of entrepreneurs without the legal formation of a new company. However, in contrast to the U.S., it tends to represent the establishment of a more formal business since these are legally registered sole proprietorships, while most firms are informal and not registered.

The more formal type of entrepreneurship—constituting the creation of a new legal entity to run a business with—are *sociedades*, which offer owners limited liability, shareholder rights, general operating agreements and bylaws, and allow multiple owners. Creating a *sociedad* is a practical requirement to take any external investment.

While noisier on some specifications, the relative effect for the more growth-oriented *sociedades* is as large as the effect for sole proprietorships. In the cross-sectional estimates, the relative effect is, if anything, slightly larger for *sociedades*. In the panel estimates, the effect sizes vary depending on which fixed effects are included. Getting the PEP appears to increase the registration of both less growth oriented and more growth-oriented firms.

Next, Table 8 splits the firms by whether they have founding employment. This is a different dimension of firm size focused more on initial investment than growth intention. It also may be economically relevant to assess the economic impact of the policy. At least in the U.S. the majority of employment created by startup cohorts occurs in the founding stages (Haltiwanger et al., 2013b). 47% of our firms have at least one founding employee, and 53% do not. The effects appear of similar magnitudes for both employer and non-employer firms, suggesting the impact of PEP equally created both types of companies.

Finally, Table 9 considers differences in the labor force status of the migrants when they initially registered for the census to assess whether our effect focuses mostly on individuals already employed, those already self-employed, or the unemployed. The effects are broadly similar across groups and not statistically different from each other. These results suggest that the effect of getting the PEP was not merely focused on formalizing existing businesses (which would be only for those self-employed), or on providing remedial options to those with fewer labor market opportunities (unemployed), but rather has an impact broadly across the spectrum of labor force status.

6 Conclusion

Migration is at the core of regional economic development. A large portion of international migration happens through undocumented immigration, but its consequences and potential benefits are poorly understood. Our paper studies the role of legal rights in changing the investment choices of immigrants, as observed in their entrepreneurship. To do so, we studied the introduction of the *Permiso Especial de Permanencia* (PEP) in Colombia in 2018, which provided about 300,000 Venezuelan immigrants with a de facto legal residency. We introduced a novel regression discontinuity approach that takes advantage of exogenous variation introduced by the Colombian government in the design of the program, by separating migrants into different brackets based on a previously provided number. We find receiving the PEP increases new legal firm formation. We show the most likely mechanism to drive this change is investment choices, and that (within only a few years), the regular migratory status alone brings Venezuelans close to locals in terms of entrepreneurial activity.

At a broader level, the role of immigrants and their economic benefits is one of the most common regulatory and policy discussions (Clemens, 2011; Kerr et al., 2016; Abramitzky and Boustan, 2017; Azoulay et al., 2020), but little work has been done studying how the design of individual institutions such as the legal framework promote the participation of immigrants in such investment and economic dynamism. Our paper hopes to provide an initial set of results in this conversation.

References

Abadie, Alberto and Guido W Imbens, "Matching on the estimated propensity score," *Econometrica*, 2016, *84* (2), 781–807.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak, "Research design meets market design: Using centralized assignment for impact evaluation," *Econometrica*, 2017, *85* (5), 1373–1432.

AbdulkadIroğlu, Atila, Joshua D Angrist, Yusuke Narita, Parag A Pathak, and Roman A Zarate, "Regression discontinuity in serial dictatorship: Achievement effects at Chicago's exam schools," *American Economic Review*, 2017, *107* (5), 240–245.

Abramitzky, Ran and Leah Boustan, "Immigration in American economic history," *Journal of economic literature*, 2017, *55* (4), 1311–45.

Acemoglu, Daron and James A Robinson, *Why nations fail: The origins of power, prosperity, and poverty*, Currency, 2012.

Amuedo-Dorantes, Catalina and Cynthia Bansak, "The Impact of Amnesty on Labor Market Outcomes: A Panel Study Using the Legalized Population Survey," *Industrial Relations: A Journal of Economy and Society*, 2011, *50* (3), 443–471.

_ , _ , **and Steven Raphael**, "Gender differences in the labor market: Impact of IRCA," *American Economic Review*, 2007, *97* (2), 412–416.

Andrade, Gustavo Henrique De, Miriam Bruhn, and David J McKenzie, "A helping hand or the long arm of the law? Experimental evidence on what governments can do to formalize firms," *Experimental Evidence on What Governments Can Do to Formalize Firms (May 1, 2013). World Bank Policy Research Working Paper*, 2013, (6435).

Angrist, Joshua D, "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 1998, *66* (2), 249–288.

_ **and Miikka Rokkanen**, "Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff," *Journal of the American Statistical Association*, 2015, *110* (512), 1331–1344.

_ **, Guido W Imbens, and Donald B Rubin**, "Identification of causal effects using instrumental variables," *Journal of the American statistical Association*, 1996, *91* (434), 444–455.

Arendt, Jacob Nielsen, Iben Bolvig, Mette Foged, Linea Hasager, and Giovanni Peri, "Language Training and Refugeesâ€™ Integration," CReAM Discussion Paper Series 2104, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London January 2021.

Azoulay, Pierre, Benjamin F Jones, J Daniel Kim, and Javier Miranda, "Age and high-growth entrepreneurship," *American Economic Review: Insights*, 2020, *2* (1), 65–82.

Bahar, Dany, Ana María Ibáñez, and Sandra V Rozo, "Give me your tired and your poor: Impact of a large-scale amnesty program for undocumented refugees," *Journal of Development Economics*, 2021, *151*, 102652.

Baker, Scott R, "Effects of immigrant legalization on crime," *American Economic Review*, 2015, *105* (5), 210–13.

Barach, Moshe A, Aseem Kaul, Ming D Leung, and Sibo Lu, "Strategic redundancy in the use of big data: Evidence from a two-sided labor market," *Strategy Science*, 2019, *4* (4), 298–322.

Baumol, William J, "Entrepreneurship: Productive, Unproductive, and Destructive," *The Journal of Political Economy*, 1990, *98* (5 Part 1), 893–921.

Borusyak, Kirill and Peter Hull, "Non-random exposure to exogenous shocks: Theory and applications," Technical Report, National Bureau of Economic Research 2021.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Cattaneo, Matias D and Rocio Titiunik, "Regression Discontinuity Designs," *arXiv preprint arXiv:2108.09400*, 2021.

_ **, Michael Jansson, and Xinwei Ma**, "Manipulation testing based on density discontinuity," *The Stata Journal*, 2018, *18* (1), 234–261.

_ **, Rocio Titiunik, and Gonzalo Vazquez-Bare**, "Inference in regression discontinuity designs under local randomization," *The Stata Journal*, 2016, *16* (2), 331–367.

Chatterji, Aaron, Edward Glaeser, and William Kerr, "Clusters of entrepreneurship and innovation," *Innovation policy and the economy*, 2014, *14* (1), 129–166.

Clemens, Michael A, "Economics and emigration: Trillion-dollar bills on the sidewalk?," *Journal of Economic perspectives*, 2011, *25* (3), 83–106.

Cowgill, Bo, "Bias and productivity in humans and algorithms: Theory and evidence from resume screening," *Columbia Business School, Columbia University*, 2018, *29*.

Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda, "The role of entrepreneurship in US job creation and economic dynamism," *Journal of Economic Perspectives*, 2014, *28* (3), 3–24.

Deke, John and Lisa Dragoset, "Statistical Power for Regression Discontinuity Designs in Education: Empirical Estimates of Design Effects Relative to Randomized Controlled Trials. Working Paper.," *Mathematica Policy Research, Inc.*, 2012.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The regulation of entry," *The quarterly Journal of economics*, 2002, *117* (1), 1–37.

Dustmann, Christian, Francesco Fasani, and Biagio Speciale, "Illegal migration and consumption behavior of immigrant households," *Journal of the European Economic Association*, 2017, *15* (3), 654–691.

Eckles, Dean, Nikolaos Ignatiadis, Stefan Wager, and Han Wu, "Noise-induced randomization in regression discontinuity designs," *arXiv preprint arXiv:2004.09458*, 2020.

Embretson, Susan E and Steven P Reise, *Item response theory*, Psychology Press, 2013.

Fairlie, Robert and Christopher M Woodruff, "Mexican-american entrepreneurship," *The BE Journal of Economic Analysis & Policy*, 2010, *10* (1).

Fairlie, Robert W and Magnus Lofstrom, "Immigration and entrepreneurship," in "Handbook of the economics of international migration," Vol. 1, Elsevier, 2015, pp. 877–911.

Foged, Mette, Janis Kreuder, and Giovanni Peri, "Integrating Refugees by Addressing Labor Shortages? A Policy Evaluation," NBER Working Papers 29781, National Bureau of Economic Research, Inc February 2022.

_ , Linea Hasager, and Giovanni Peri, "Comparing the effects of policies for the labor market integration of refugees," Technical Report, Cambridge, MA October 2022.

Frandsen, Brigham R, "Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete," in "Regression discontinuity designs," Vol. 38, Emerald Publishing Limited, 2017, pp. 281–315.

Galiani, Sebastian, Marcela Meléndez, and Camila Navajas Ahumada, "On the effect of the costs of operating formally: New experimental evidence," *Labour Economics*, 2017, *45*, 143–157.

Giuntella, Osea and Jakub Lonsky, "The effects of DACA on health insurance, access to care, and health outcomes," *Journal of Health Economics*, 2020, *72*, 102320.

Gomez, Brad T, Thomas G Hansford, and George A Krause, "The Republicans should pray for rain: Weather, turnout, and voting in US presidential elections," *The Journal of Politics*, 2007, *69* (3), 649–663.

Guzman, Jorge, "The Direct Effect of Corporate Law on Entrepreneurship," 2020.

Haltiwanger, John, Ron S. Jarmin, and Javier Miranda, "WHO CREATES JOBS? SMALL VERSUS LARGE VERSUS YOUNG," *The Review of Economics and Statistics*, 2013, *95* (2), 347–361.

_ , Ron S Jarmin, and Javier Miranda, "Who creates jobs? Small versus large versus young," *Review of Economics and Statistics*, 2013, *95* (2), 347–361.

Hansen, Benjamin, "Punishment and deterrence: Evidence from drunk driving," *American Economic Review*, 2015, *105* (4), 1581–1617.

Hull, Peter, "Subtracting the Propensity Score in Linear Models," *Unpublished Manuscript, MIT*, 2018.

Hunt, Jennifer, "Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa," *Journal of Labor Economics*, 2011, *29* (3), 417–457.

Ibanez, Ana, Sandra Rozo, and Dany Bahar, "Empowering Migrants: Impacts of a Migrant's Amnesty on Crime Reports," 2020.

Ibáñez, Ana, Sandra V. Rozo, and Dany Bahar, "Empowering Migrants: Impacts of a Migrant's Amnesty on Crime Reports," *World Bank Policy Research Working Papers*, 2021, (9833).

Imbens, Guido W. and Donald B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, 2015.

Imbens, Guido W and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of econometrics*, 2008, *142* (2), 615–635.

Jacob, Brian A and Lars Lefgren, "Remedial education and student achievement: A regression-discontinuity analysis," *Review of economics and statistics*, 2004, *86* (1), 226–244.

Jaramillo, Miguel, *Is there demand for formality among informal firms: evidence from microfirms in downtown Lima* number 12/2009, Discussion Paper, 2009.

Jasso, Guillermina, *Journal of Policy Analysis and Management*, 1993, *12* (2), 403–406.

Karlan, Dean and Jonathan Zinman, "Microcredit in theory and practice: Using randomized credit scoring for impact evaluation," *Science*, 2011, *332* (6035), 1278–1284.

Kaushal, Neeraj, "Amnesty programs and the labor market outcomes of undocumented workers," *J. Hum. Resour.*, 2006, *XLI* (3), 631–647.

Kerr, Sari Pekkala, William Kerr, Çağlar Özden, and Christopher Parsons, "Global talent flows," *Journal of Economic Perspectives*, 2016, *30* (4), 83–106.

Kerr, William R, "US high-skilled immigration, innovation, and entrepreneurship: Empirical approaches and evidence," Technical Report, National Bureau of Economic Research 2013.

King, Gary and Richard Nielsen, "Why propensity scores should not be used for matching," *Political analysis*, 2019, *27* (4), 435–454.

Kossoudji, Sherrie A and Deborah A Cobb-Clark, "Coming out of the shadows: Learning about legal status and wages from the legalized population," *Journal of Labor Economics*, 2002, *20* (3), 598–628.

Kossoudji, Sherrie A. and Deborah A. Cobb-Clark, "Coming out of the Shadows: Learning about Legal Status and Wages from the Legalized Population," *Journal of Labor Economics*, 2002, *20* (3), 598–628.

Kuka, Elira, Na'ama Shenhav, and Kevin Shih, "Do human capital decisions respond to the returns to education? Evidence from DACA," *American Economic Journal: Economic Policy*, 2020, *12* (1), 293–324.

Lee, David S and Thomas Lemieux, "Regression discontinuity designs in economics," *Journal of economic literature*, 2010, *48* (2), 281–355.

Lerner, Josh, *Boulevard of broken dreams*, Princeton University Press, 2009.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott, "Do political protests matter? evidence from the tea party movement," *The Quarterly Journal of Economics*, 2013, *128* (4), 1633–1685.

Marbach, Moritz and Dominik Hangartner, "Profiling compliers and noncompliers for instrumental-variable analysis," *Political Analysis*, 2020, *28* (3), 435–444.

Marx, Benjamin, Thomas Stoker, and Tavneet Suri, "The economics of slums in the developing world," *Journal of Economic perspectives*, 2013, *27* (4), 187–210.

Mckenzie, David, "Should we require balance t-tests of baseline observables in randomized experiments?"

Mel, Suresh De, David McKenzie, and Christopher Woodruff, "The demand for, and consequences of, formalization among informal firms in Sri Lanka," *American Economic Journal: Applied Economics*, 2013, *5* (2), 122–50.

Monras, Joan, Javier Vázquez-Grenno, and Ferran Elias, "Understanding the effects of legalizing undocumented immigrants," Technical Report February 2018.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny, "The allocation of talent: Implications for growth," *The quarterly journal of economics*, 1991, *106* (2), 503–530.

Murray, Fiona and Scott Stern, "Linking and leveraging," *Science*, 2015, *348* (6240), 1203–1203.

Narayanan, Sridhar and Kirthi Kalyanam, "Position effects in search advertising and their moderators: A regression discontinuity approach," *Marketing Science*, 2015, *34* (3), 388–407.

Patler, Caitlin, Erin Hamilton, Kelsey Meagher, and Robin Savinar, "Uncertainty about DACA may undermine its positive impact on health for recipients and their children," *Health Affairs*, 2019, *38* (5), 738–745.

Pinotti, Paolo, "Clicking on heaven's door: The effect of immigrant legalization on crime," *American Economic Review*, 2017, *107* (1), 138–68.

Porta, Rafael La and Andrei Shleifer, "The Unofficial Economy and Economic Development.," *Brookings Papers on Economic Activity*, 2008.

— **and** — , "Informality and development," *Journal of economic perspectives*, 2014, *28* (3), 109–26.

Puri, Manju and David T Robinson, "The economic psychology of entrepreneurship and family business," *Journal of Economics & Management Strategy*, 2013, *22* (2), 423–444.

Robins, James M, Steven D Mark, and Whitney K Newey, "Estimating exposure effects by modelling the expectation of exposure conditional on confounders," *Biometrics*, 1992, pp. 479–495.

Rokkanen, Miikka AT, "Exam schools, ability, and the effects of affirmative action: Latent factor extrapolation in the regression discontinuity design," 2015.

Romero, Nicolás, Laura Uribe, and Abraham Farfán, "Mechanisms that strengthen integration: Migrant networks and residence permits in migratory processes1," 2021.

Rosenbaum, Paul R and Donald B Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 1983, *70* (1), 41–55.

Saxenian, AnnaLee, "Silicon Valley's new immigrant high-growth entrepreneurs," *Economic development quarterly*, 2002, *16* (1), 20–31.

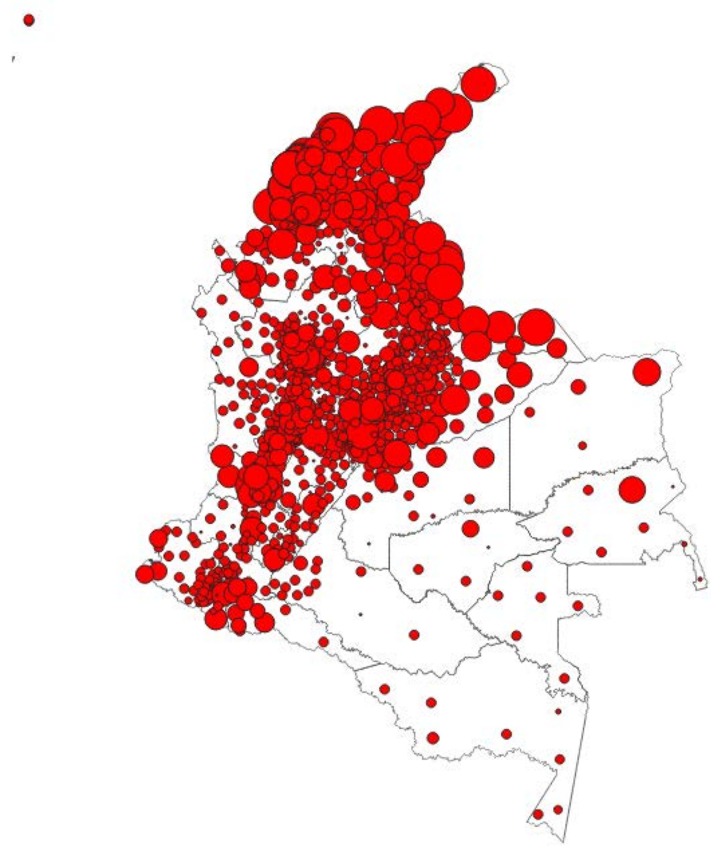
Schochet, Peter Z, "Statistical power for regression discontinuity designs in education evaluations," *Journal of Educational and Behavioral Statistics*, 2009, *34* (2), 238–266.

Sekhon, Jasjeet S and Rocío Titiunik, "On interpreting the regression discontinuity design as a local experiment," in "Regression Discontinuity Designs," Vol. 38, Emerald Publishing Limited, 2017, pp. 1–28.

Shenoy, Ajay, Bhavyaa Sharma, Guanghong Xu, Rolly Kapoor, Haedong Aiden Rho, and Kinpritma Sangha, "God is in the rain: the impact of rainfall-induced early social distancing on COVID-19 outbreaks," *Journal of Health Economics*, 2022, *81*, 102575.

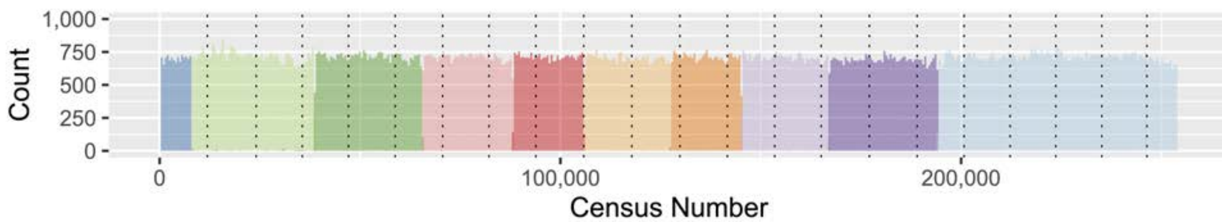
Wasow, Omar, "Agenda seeding: How 1960s black protests moved elites, public opinion and voting," *American Political Science Review*, 2020, *114* (3), 638–659.

Figure 1: **RAMV Registration Points (Municipalities)**



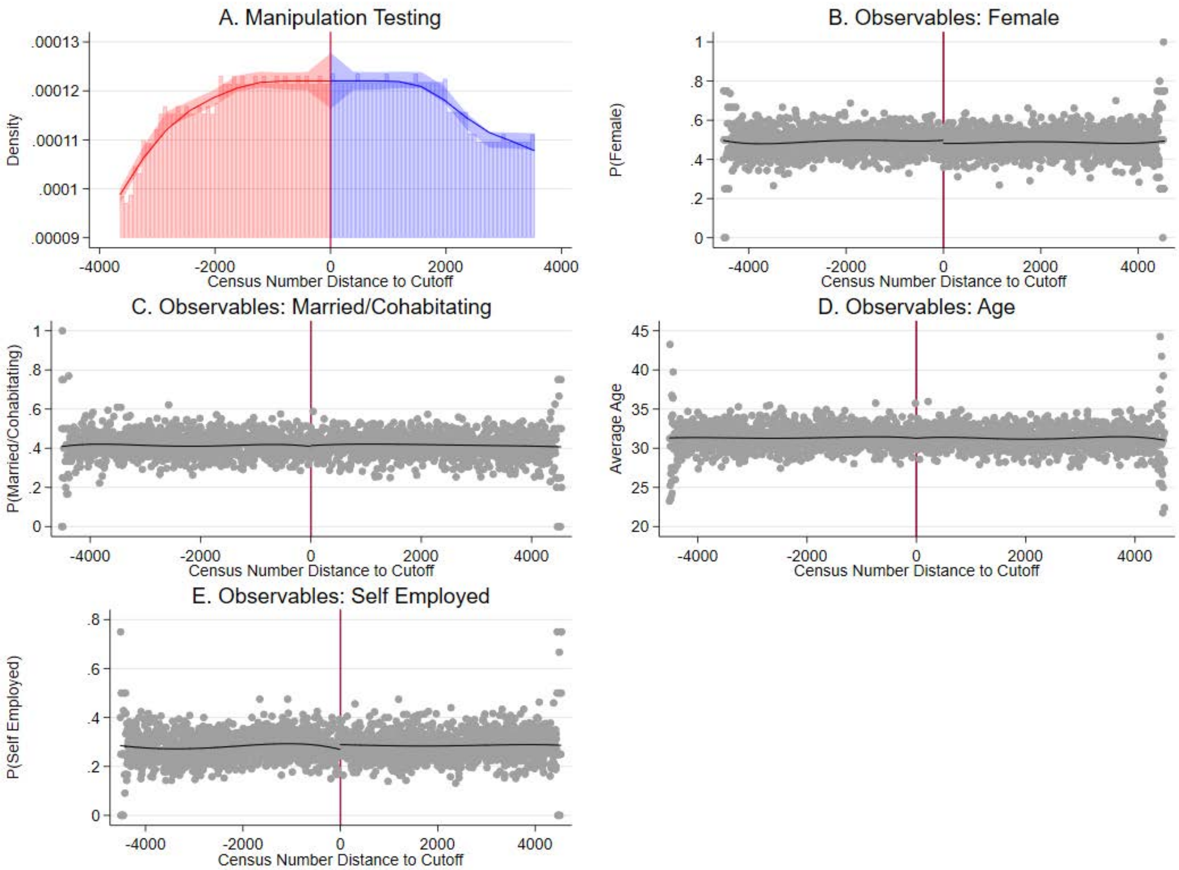
Notes: This figure visualizes all the municipalities in Colombia where there was a RAMV registration point. The markers are scaled using as weight the total number of people who registered in each municipality.

Figure 2: **Histogram of RAMV Census Numbers by Week (with Thresholds)**



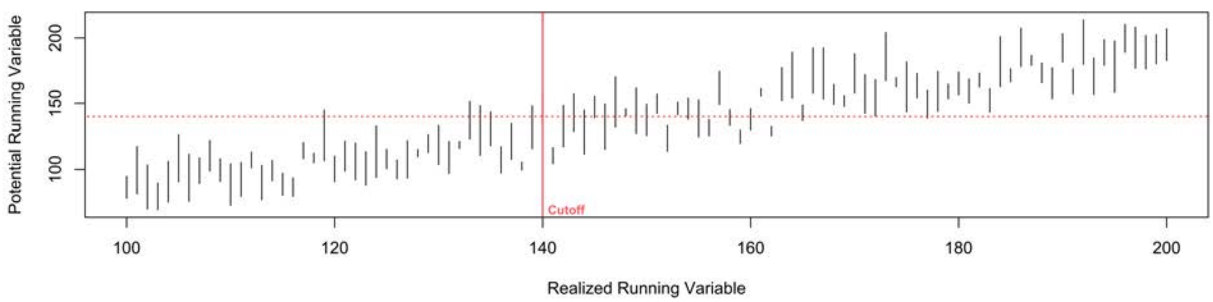
Notes: Each color represents a week's worth of RAMV census numbers, assigned zero to $\approx$ 250K (as described in the text, we have normalized a jump in these numbers thanks to a software upgrade). The black dashed lines represent the thresholds in the Online Appendix Figure A1 advertisement, adjusted for the aforementioned jumps. As these black lines show, the thresholds are evenly spaced, placing approximately 4.5% of the sample each bin. The bins were not exactly equal in the number of individuals, perhaps because individuals are clustered by families which were not broken apart. To have each bin be approximately 4.5% of the sample, the thresholds did not respect calendar boundaries by giving all migrants who took the census in the same week the same bin. The figure above shows that migrants who took the RAMV census in the second week (the light green area on the left) could potentially fall into one of four different batches, depending on randomness in the census-taking order.

Figure 3: **Distribution of Observables across the RD Threshold in Stacked Setup**



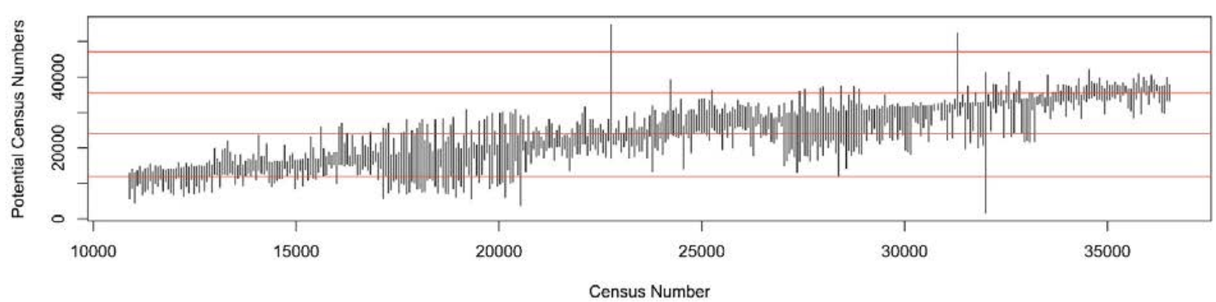
Notes: We plot the distribution of observables across the thresholds in a stacked setup. Panel A is a manipulation test on the density of census numbers away from the threshold as in Cattaneo et al. (2018). Panels B through E plot observables in a regression discontinuity setup following the approach in Cattaneo and Titiunik (2021). We document no differences across any observable.

Figure 4: **Visualization for Research Design**



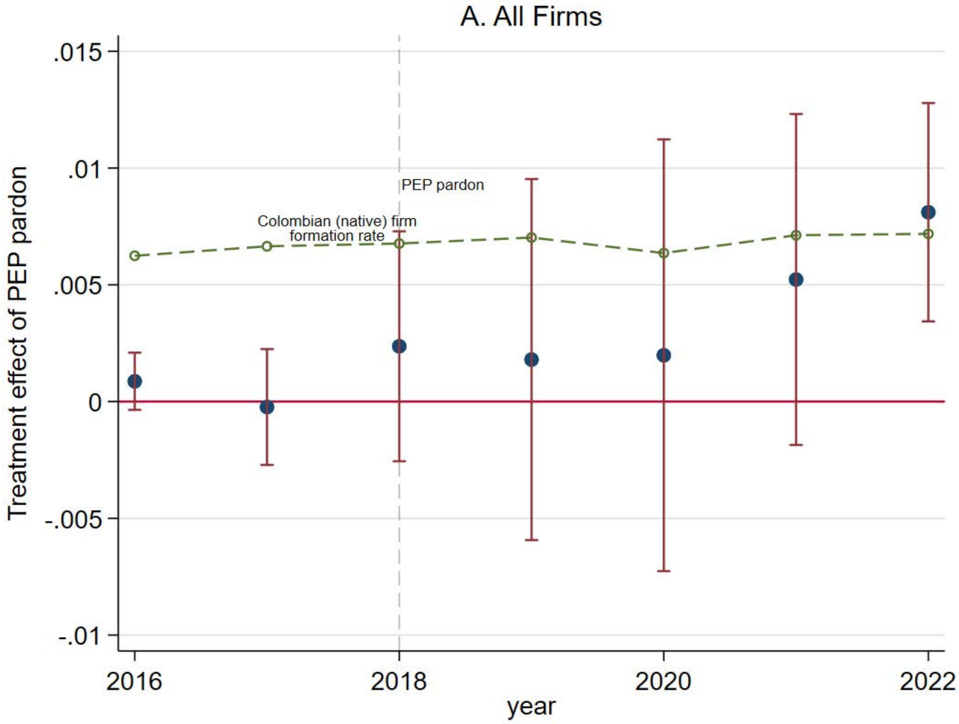
Notes: This visualizes the key elements of our empirical approach. Each observation is represented as a vertical bar placed over the observation's realized running variable (on the X axis). The lower and upper bounds of the vertical bar represent the 95% confidence interval for the *potential* running variable for each observation (i.e., F_i^X as defined in Lemma 1).

Figure 5: **Distribution of Potential Census Numbers**



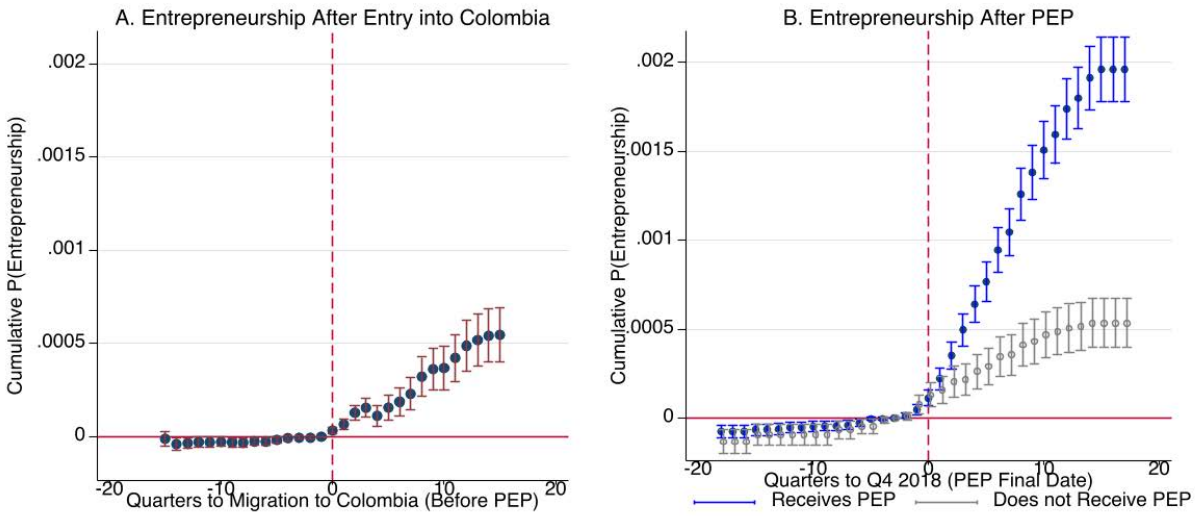
Notes: We plot the 95-percent interval of realizations of the predicted census numbers for a subset of our data, based on the possible rain distributions that week using our main model. For additional details, see Section 4.

Figure 6: **Treatment Effect by Year**



Notes: This figure plots, in each panel, the coefficients of seven instrumental variables regressions, with the dependent variable indicating whether a firm is created in each year, from 2016 to 2022.

Figure 7: **Panel Analysis**



Notes: We plot the quarterly coefficients of regression models documenting changes in the probability of starting a firm after migration changes. The dependent variable is a binary outcome measure equal to 1 if the migrant has started a firm at quarter t and 0 otherwise. Regressions include individual fixed effects. Panel A compares the probability of starting a firm before and after the migration event under the assumption that no firm created in Colombia by migrants previously existed in Venezuela. If there are firms simply moving from Venezuela, then our estimate would *overclaim* the effect of physical migration. Panel B compares the likelihood of starting a firm after the PEP reform was available for migrants that do get the PEP and those that do not. There appear to be very similar pre-trends before the PEP but the estimates separate significantly afterwards.

Table 1: **Descriptive Statistics: Full Sample**

	N	Mean	Std.Dev	Min	P25	P50	P75	Max
RAMV Observables								
Age	331115	31	12	0	22	29	38	118
Female	331376	.5	.5	0	0	0	1	1
married	331646	.12	.33	0	0	0	0	1
Education is High School or Lower	331646	.8	.4	0	1	1	1	1
Informal Labor	331646	.29	.45	0	0	0	1	1
Self Employed	331646	.27	.44	0	0	0	1	1
Unemployed	331646	.22	.41	0	0	0	0	1
No Occupation Reported	331646	.082	.28	0	0	0	0	1
Census # (Standardized)	331646	.019	.99	-1.7	-.83	.023	.88	1.7
Labor Certificate	331613	.16	.37	0	0	0	0	1
Head of Household	331646	.65	.48	0	0	1	1	1
Family Size	331639	3.3	2.1	0	2	3	5	10
Has Family in Colombia	331639	.42	.49	0	0	0	1	1
Has Family in Venezuela	331639	.69	.46	0	0	1	1	1
Had No Food in Last 3 Months	331639	.37	.48	0	0	0	1	1
Expects to Stay in Colombia 1 Year or More	331646	.9	.3	0	1	1	1	1
Registered in the Morning	331646	.48	.5	0	0	0	1	1
PEP Pardon	331646	.68	.46	0	0	1	1	1
Outcomes								
Company Created	331646	.0016	.04	0	0	0	0	1
Sole Proprietorship Created	331646	.0014	.037	0	0	0	0	1
Sociedad Created	331646	.0002	.014	0	0	0	0	1
Number of Employees	458	.79	1.2	0	0	0	1	14
Founding Assets (Colombian Pesos)	458	1.5e+07	1.4e+08	0	1000000	1500000	2000000	2.9e+09

Notes: This table presents descriptive statistics (number of observations, mean, standard deviation, minimum, percentiles 25, 50 and 75, as well as maximum values) for the used sample of Venezuelan immigrants registered in the RAMV census.

Table 2: **Instrument Balance**

Variable	*Lucky Migrants*	*Unlucky Migrants*	*Normalized Difference*
Age	30.116	29.973	.012
	(12.2)	(12.509)	
Female	.504	.525	-.04
	(.5)	(.499)	
Married	.129	.126	.01
	(.335)	(.332)	
Education is High School or Lower	.811	.819	-.02
	(.392)	(.385)	
Informal Labor	.298	.266	.071
	(.457)	(.442)	
Self Employed	.246	.26	-.032
	(.431)	(.438)	
Unemployed	.216	.204	.031
	(.412)	(.403)	
No Occupation Reported	.095	.102	-.022
	(.294)	(.302)	
Labor Certificate	.16	.138	.06
	(.366)	(.345)	
Head of Household	.538	.55	-.024
	(.499)	(.497)	
Family Size	3.633	3.595	.018
	(2.231)	(1.991)	
Has Family in Colombia	.431	.426	.01
	(.495)	(.494)	
Has Family in Venezuela	.773	.613	.352
	(.419)	(.487)	
Had No Food in Last 3 Months	.404	.37	.071
	(.491)	(.483)	
Expects to Stay in Colombia 1 Year or More	.909	.913	-.012
	(.287)	(.282)	

Notes: This table presents sample averages for different characteristics of the individuals in our sample based by levels of the instrument that results of the noise model: Those with "negative luck" (column 2) vs. those with "positive luck" (column 3). The fourth column presents mean differences with corresponding p-value levels with the usual notation. To compute the differences in the fourth column we follow Imbens and Rubin (2015) and Mckenzie (n.d.). In particular, we report the difference in means between the "lucky" (treatment) and "unlucky" (control) groups, divided by the square root of half the sum of the treatment and control group variances. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 3: **Compliers vs Always-Takers**

	Compliers	Always Takers	Difference
Age	31.51	30.19	1.32***
Female	0.64	0.52	0.12***
Married	0.16	0.14	0.02***
Education is High School or Lower	0.91	0.80	0.11***
Informal Labor	0.09	0.28	-0.19***
Self Employed	0.24	0.26	-0.02***
Unemployed	0.11	0.20	-0.09***
No Occupation Reported	0.26	0.09	0.17***
Census # (Standardized)	0.58	0.23	0.35***
Labor Certificate	-0.02	0.15	-0.17***
Head of Household	0.28	0.55	-0.27***
Family Size	3.53	3.61	-0.09***
Has Family in Colombia	0.39	0.41	-0.02***
Has Family in Venezuela	0.08	0.61	-0.53***
Had No Food in Last 3 Months	0.38	0.36	0.03***
Expects to Stay in Colombia 1 Year or More	0.89	0.92	-0.03***

Notes: This table presents sample averages for different characteristics of the individuals in our sample based on their characterization of 'compliers' (column 2) and for 'always-takers' (column 3). The fourth column presents mean differences with corresponding p-value levels with the usual notation. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 4: **First Stage Regressions**. Rain-based Noise.

	(1)	(2)	(3)	(4)
	Dep Var:	*Dep Var:*	*Dep Var:*	*Dep Var:*
	PEP Pardon	PEP Pardon	PEP Pardon	PEP Pardon
Actual Delay	-.0053***	-.0029***	-.0031***	-.0032***
	(.00033)	(.00041)	(.0004)	(.0004)
Running Expected Delay		-.0014	-.0022	-.0024***
		(.0012)	(.0016)	(.00024)
F-Statistic		50	61	263
Method		Simple	Main	Expanded
Observations	336,291	336,291	336,291	336,291
R^2	.089	.09	.09	.09

Notes: This table presents results for the first stage of our 2SLS estimation as part of our empirical strategy. Fixed effects for the week of registration, the role of migrant in the family (e.g., spouse, head of household, grandmother, etc.), marital status, occupation, level of education, and the municipality of registration included. Standard errors clustered by family ID. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 5: **Main Results.** 2SLS estimates.
Dep. Var.: Business Created.

	(1)	(2)	(3)	(4)
PEP Pardon	.0013***	.016*	.018**	.012***
	(.00013)	(.0095)	(.0088)	(.0043)
F Statistic	.	50	61	263
Method	OLS	Simple	Main	Expanded
Observations	336,419	336,291	336,291	336,291

Notes: This table estimates Equation 4 using the full sample, which estimates the effect of having received the PEP visa on starting a formal business. Column 1 presents OLS results, while Columns 2 to 4 present 2SLS results. Standard errors clustered at the date of census completion are reported in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 6: **Panel Results.**
Dep. Var: 1[Has started a company]

	(1)	(2)	(3)	(4)	(5)
After Receiving PEP	0.00104***	0.000907***	0.000842***	0.000887***	0.000363***
	(0.0000776)	(0.0000743)	(0.0000974)	(0.0000877)	(0.0000571)
After Entry into Colombia	0.000333***	0.000423***	0.0000506	0.000250***	0.000249***
	(0.0000386)	(0.0000412)	(0.0000318)	(0.0000480)	(0.0000480)
Linear Trend for Getting PEP					0.0000291***
					(0.00000319)
Individual F.E.	No	No	No	Yes	Yes
Year-Quarter F.E.	No	No	Yes	Yes	Yes
Week of Registration F.E.	No	Yes	No	No	No
RAMV Reg. Municipality	No	Yes	No	No	No
Difference in Effects	0.00071***	0.00048***	0.00079***	0.00064***	0.00011*
s.e. of Difference	(0.00010)	(0.00010)	(0.00011)	(0.00009)	(0.00007)
R-squared	0.000624	0.00388	0.000780	0.415	0.415

Notes: Robust standard errors clustered at the individual level in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 7: **The Impact of PEP on Depending on the Legal Form of Firm**

Panel A: Cross Section Regressions

	(1) Sole Prop.	(2) Sociedades	(3) Sole Prop.	(4) Sociedades
PEP Visa	0.0152*	0.00260	0.00940**	0.00240*
	(0.00808)	(0.00284)	(0.00402)	(0.00129)
Method	Main	Main	Expanded	Expanded
Outcome Mean	0.00139	0.000220	0.00139	0.000220
Relative Effect	10.90	11.79	6.758	10.91
F Statistic	60.87	60.54	264.7	264.4
Observations	336217	335824	336217	335824
R-squared	-0.0303	-0.00612	-0.0110	-0.00524

Panel B: Panel Regressions

	(1) Sole Prop.	(2) Sociedades	(3) Sole Prop.	(4) Sociedades
After Receiving PEP	0.000848***	0.0000520	0.000306***	0.0000579***
	(0.0000658)	(0.0000320)	(0.0000532)	(0.0000209)
After Entry into Colombia	0.000329***	0.000111***	0.000189***	0.0000601***
	(0.0000337)	(0.0000215)	(0.0000454)	(0.0000156)
Linear Trend for Getting PEP			0.0000283***	0.000000856
			(0.00000274)	(0.00000162)
Individual F.E.	No	No	Yes	Yes
Year-Quarter F.E.	No	No	Yes	Yes
Week of Registration F.E.	Yes	Yes	No	No
RAMV Reg. Municipality	Yes	Yes	No	No
Year-Month of Entry	Yes	Yes	No	No
Observations	12094056	12079944	12095640	12081528
R-squared	0.00505	0.00182	0.376	0.580

Notes: Panel A is cross-sectional regressions as in Table 5, including the same menu of fixed effects and using different versions of our predicted instrument. Panel B is panel data at the quarter and individual level with difference-in-differences models as in 6. Year-Quarter F.E. are fixed effects for the quarter of the observation, Year-Month of Entry is fixed effects for the month the migrant entered Colombia. Robust standard errors clustered at the individual level in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 8: **The Impact of PEP on Depending Firm Employment**

Panel A: Cross Section Regressions

	(1) Non Employer	(2) Employer	(3) Non Employer	(4) Employer
PEP Visa	0.00537	0.00663	0.00503*	0.00392
	(0.00534)	(0.00545)	(0.00273)	(0.00265)
Method	Main	Main	Expanded	Expanded
Outcome Mean	0.000702	0.000672	0.000702	0.000672
Relative Effect	7.645	9.854	7.159	5.830
F Statistic	60.79	60.55	264.9	264.1
Observations	335986	335975	335986	335975
R-squared	-0.00781	-0.0114	-0.00683	-0.00363

Panel B: Panel Regressions

	(1) Non Employer	(2) Employer	(3) Non Employer	(4) Employer
After Receiving PEP	0.000381***	0.000433***	0.000172***	0.000145***
	(0.0000485)	(0.0000487)	(0.0000407)	(0.0000349)
After Entry into Colombia	0.000226***	0.000162***	0.000142***	0.0000713**
	(0.0000275)	(0.0000251)	(0.0000336)	(0.0000314)
Linear Trend for Getting PEP			0.0000111***	0.0000148***
			(0.00000217)	(0.00000205)
Individual F.E.	No	No	Yes	Yes
Year-Quarter F.E.	No	No	Yes	Yes
Week of Registration F.E.	Yes	Yes	No	No
RAMV Reg. Municipality	Yes	Yes	No	No
Year-Month of Entry	Yes	Yes	Yes	Yes
Observations	12085704	12085344	12087288	12086928
R-squared	0.00373	0.00420	0.402	0.405

Notes: Panel A is cross sectional regressions as in Table 5, including the same menu of fixed effects and using different versions of our predicted instrument. Panel B is panel data at the quarter and individual level with difference-in-differences models as in 6. Year-Quarter F.E. are fixed effects for the quarter of the observation, Year-Month of Entry is fixed effects for the month the migrant entered Colombia. Robust standard errors clustered at the individual level in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 9: **The Impact of PEP on Across Prior Labor Force Status**

Panel A: Cross Section Regressions

	(1) Subsample: Employed	(2) Subsample: Self-employed	(3) Subsample: Unemployed	(4) Subsample: Employed	(5) Subsample: Self-employed	(6) Subsample: Unemployed
PEP Visa	0.0361	0.0280*	0.0135	0.0158	0.0180**	0.00382
	(0.0342)	(0.0162)	(0.0147)	(0.0105)	(0.00773)	(0.00807)
Method	Main	Main	Main	Expanded	Expanded	Expanded
Outcome Mean	0.00247	0.00145	0.00160	0.00247	0.00145	0.00160
Relative Effect	14.59	19.33	8.400	6.387	12.44	2.383
F Statistic	8.645	15.79	39.36	96.19	89.36	88.01
Observations	100996	90429	74211	100996	90429	74211
R-squared	-0.0936	-0.109	-0.0186	-0.0172	-0.0440	-0.000713

Panel B: Panel Regressions

	(1) Subsample: Employed	(2) Subsample: Self-employed	(3) Subsample: Unemployed	(4) Subsample: Employed	(5) Subsample: Self-employed	(6) Subsample: Unemployed
After Receiving PEP	0.00121***	0.000810***	0.00120***	0.000264*	0.000316***	0.000622***
	(0.000158)	(0.000140)	(0.000147)	(0.000142)	(0.000108)	(0.000112)
After Entry into Colombia	0.000675***	0.000420***	0.000328***	0.000342***	0.000149	0.000414***
	(0.0000951)	(0.0000771)	(0.0000613)	(0.000110)	(0.0000977)	(0.0000724)
Linear Trend for Getting PEP				0.0000405***	0.0000238***	0.0000376***
				(0.00000716)	(0.00000577)	(0.00000615)
Individual F.E.	No	No	No	Yes	Yes	Yes
Year-Quarter F.E.	No	No	No	Yes	Yes	Yes
Week of Registration F.E.	Yes	Yes	Yes	No	No	No
RAMV Reg. Municipality	Yes	Yes	Yes	No	No	No
Year-Month of Entry	Yes	Yes	Yes	No	No	No
Observations	3635028	3254112	2673648	3635316	3254436	2673720
R-squared	0.00948	0.00573	0.00348	0.389	0.410	0.422

Notes: Panel A is cross sectional regressions as in Table 5, including the same menu of fixed effects and using different versions of our predicted instrument. Panel B is panel data at the quarter and individual level with difference-in-differences models as in 6. Year-Quarter F.E. are fixed-effects for the quarter of the observation, Year-Month of Entry is fixed-effects for the month the migrant entered Colombia. Robust standard errors clustered at the individual level in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

0 0

Online Appendix for *Legalizing Entrepreneurs*

A Additional Descriptives

A.1 RAMV Census

Figure A1 visualizes a syndicated ad by the Colombian government announcing the dates for which Venezuelan immigrants registered in the RAMV would be eligible to apply for the PEP visa based on their census number. It shows that there were 22 different windows in which immigrants could apply for the PEP based on the exogenous allocated census number.

Figure A2 plots the intensity of Google searches for the terms 'Permiso Especial de Permanencia' and 'RAMV' from mid 2017 to mid 2019 within Colombia. It shows that indeed in terms of the RAMV keyword, google searches started in April of 2018, when the census was initiated. Yet, the search for the word PEP peaks in 3 moments of time. The first two are on July 2017 and February of 2018, which are the dates where prior versions of PEPs for documented immigrants (such as tourists) were enacted. The third one is after July of 2018 when the PEP-RAMV was announced.

Figure A3 replicates Figure 2, but where colors represent different days –as opposed to weeks– during the RAMV census registration period. We also notice in this graph that even people registering within the same day might have been assigned to different windows to apply for their PEP visa.

Figure A4 presents the number of registrations to RAMV per day during the April to June 2018 period who ultimately received the PEP visa (dark grey) and those who did not (light grey).

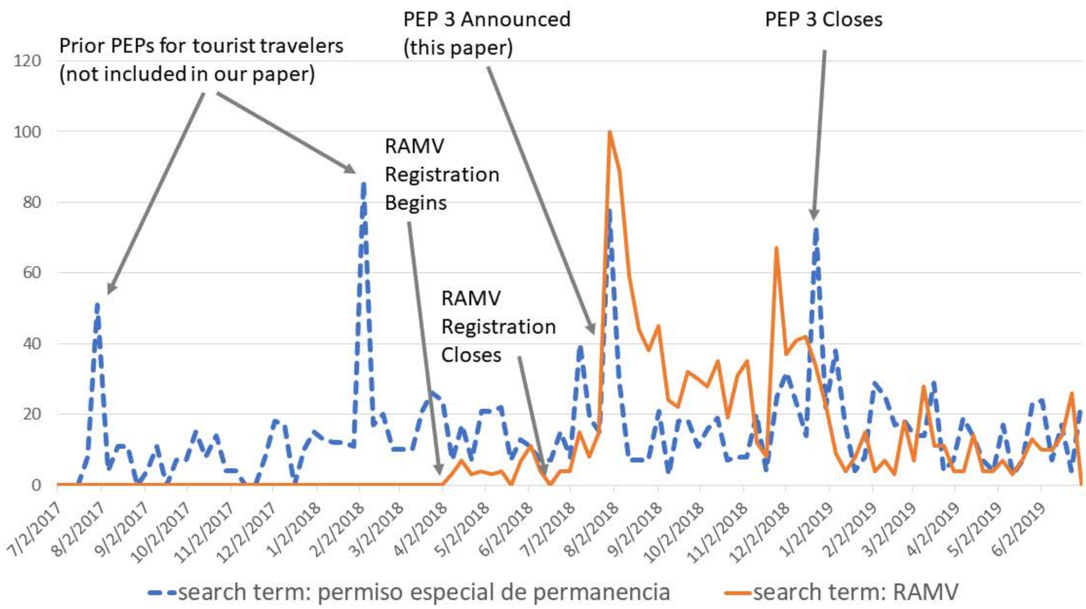
Table A1 summarizes the demographic and socioeconomic characteristics of immigrants registered in RAMV who got the PEP visa versus those that did not. Since ultimately the choice of getting the PEP is endogenous, we do see important differences. This is complemented by Figure A5 that presents the distribution of immigrants with and without PEP across several characteristics

Figure A1: **Nationally Syndicated Ad for PEP Registration Cutoff Dates**

NÚMERO DE FORMULARIO		PODRÁ SACAR SU PEP - RAMV A PARTIR DEL
DESDE	HASTA	
1	14.752	jueves, 2 de agosto de 2018
14.753	30.213	domingo, 5 de agosto de 2018
30.214	4.002.617	miércoles, 8 de agosto de 2018
4.002.618	4.014.997	sábado, 11 de agosto de 2018
4.014.998	4.027.640	martes, 14 de agosto de 2018
4.027.641	4.040.663	viernes, 17 de agosto de 2018
4.040.664	4.053.186	lunes, 20 de agosto de 2018
4.053.187	4.065.677	jueves, 23 de agosto de 2018
4.065.678	4.078.492	domingo, 26 de agosto de 2018
4.078.493	4.091.505	miércoles, 29 de agosto de 2018
4.091.506	4.104.531	sábado, 1 de septiembre de 2018
4.104.532	4.117.421	martes, 4 de septiembre de 2018
4.117.422	4.130.322	viernes, 7 de septiembre de 2018
4.130.323	4.142.976	lunes, 10 de septiembre de 2018
4.142.977	4.156.009	jueves, 13 de septiembre de 2018
4.156.010	4.168.922	domingo, 16 de septiembre de 2018
4.168.923	4.182.673	miércoles, 19 de septiembre de 2018
4.182.674	4.196.951	sábado, 22 de septiembre de 2018
4.196.952	4.209.778	martes, 25 de septiembre de 2018
4.209.779	4.222.027	viernes, 28 de septiembre de 2018
4.222.028	4.234.070	lunes, 1 de octubre de 2018
4.234.071	4.242.447	jueves, 4 de octubre de 2018

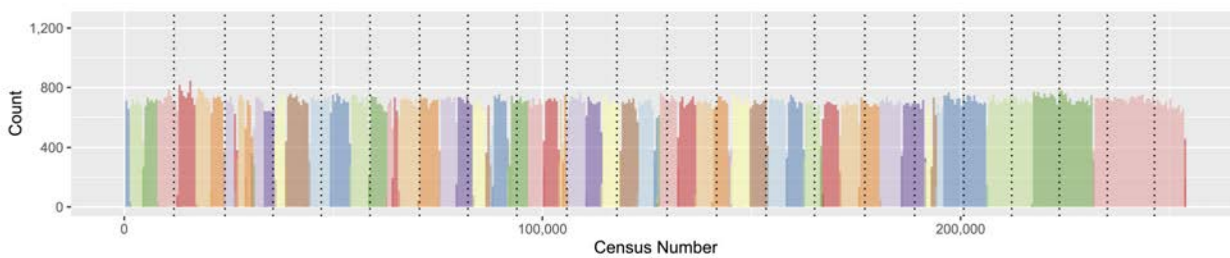
Notes: This figure shows the dates provided to those registered in the RAMV to issue their PEP visa, based on the census number they received. Note that the third row appears to contain a span of approximately four million census numbers. However, around this time, there was a jump in the sequential numbering system caused by a software upgrade. Sequential numbering proceeded before and after. Thus, the third bin did not contain a larger or smaller amount of registrants than the other. All our estimates normalize and/or control for this jump.

Figure A2: **Google Trends of Search in Colombia for 'Permiso Especial de Permanencia' and 'RAMV' around PEP period**



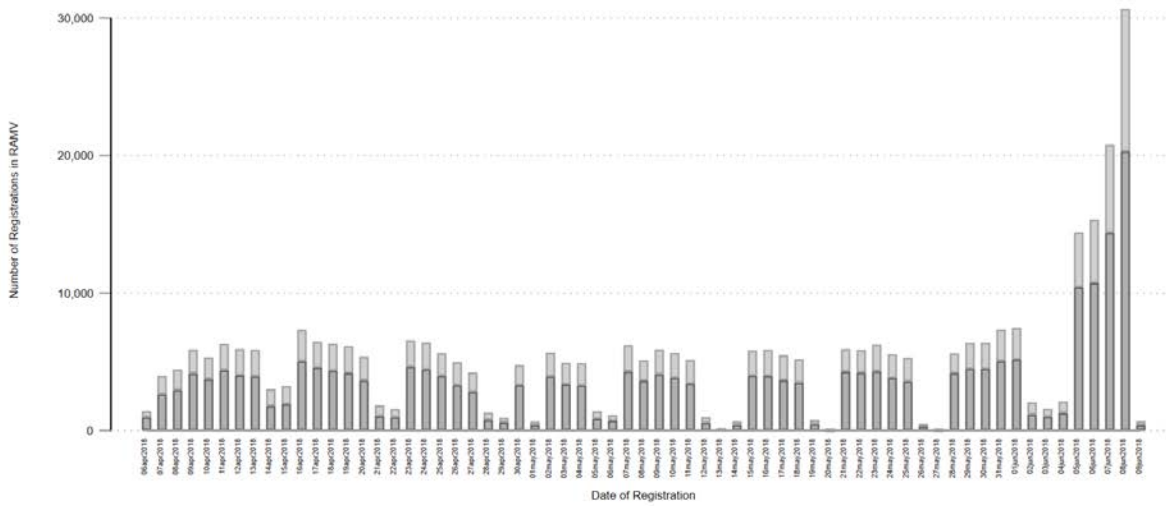
Notes: This figure plots intensity of Google searches for the terms 'Permiso Especial de Permanencia' (dashed line) and 'RAMV' (continuous line) from mid 2017 to mid 2019 in Colombia. The data is sourced from Google Trends.

Figure A3: **Histogram of Census Numbers by Day (with Thresholds)**



Notes: This figure replicates Figure 2 though each color represents a day's worth of census numbers (instead of a week). The black numbers represent the thresholds in the Figure A1 advertisement, adjusted for the aforementioned jumps. As this black lines show clearly shows, the thresholds are evenly spaced, placing approximately 4.5% of the sample each bin. The bins were not exactly equal in the number of individuals, perhaps because individuals are clustered by families which were not broken apart. In Figure 2, we see that the thresholds did not respect weekly boundaries either. To achieve approximately 4.5% of the sample each bin, the thresholds did not respect calendar boundaries by (say) giving all migrants who took the census in the same day the same bin.

Figure A4: **Number of Daily Registrations with and without PEP**



Notes: The figure plots the number of registrants per date of registration in the RAMV census. Each bar shows the amount of registrants that received a PEP visa (in dark grey) and who did not receive a PEP visa (light grey).

Table A1: **Descriptive Statistics: RAMV registrants with and without PEP visa**

	Gets PEP = 0	Gets PEP = 1	Difference
Age	29.95	30.78	-0.83***
Female	0.50	0.50	0.00
married	0.10	0.13	-0.04***
Education is High School or Lower	0.85	0.78	0.07***
Informal Labor	0.25	0.31	-0.06***
Self Employed	0.28	0.27	0.01***
Unemployed	0.23	0.22	0.01***
No Occupation Reported	0.10	0.07	0.03***
Census # (Standardized)	0.02	0.02	0.01
Labor Certificate	0.13	0.17	-0.04***
Head of Household	0.65	0.65	0.01**
Family Size	3.27	3.38	-0.11***
Has Family in Colombia	0.46	0.41	0.05***
Has Family in Venezuela	0.70	0.69	0.01
Had No Food in Last 3 Months	0.39	0.36	0.04***
Expects to Stay in Colombia 1 Year or More	0.89	0.91	-0.02***
Registered in the Morning	0.48	0.48	-0.00
Company Created	0.00	0.00	-0.00***

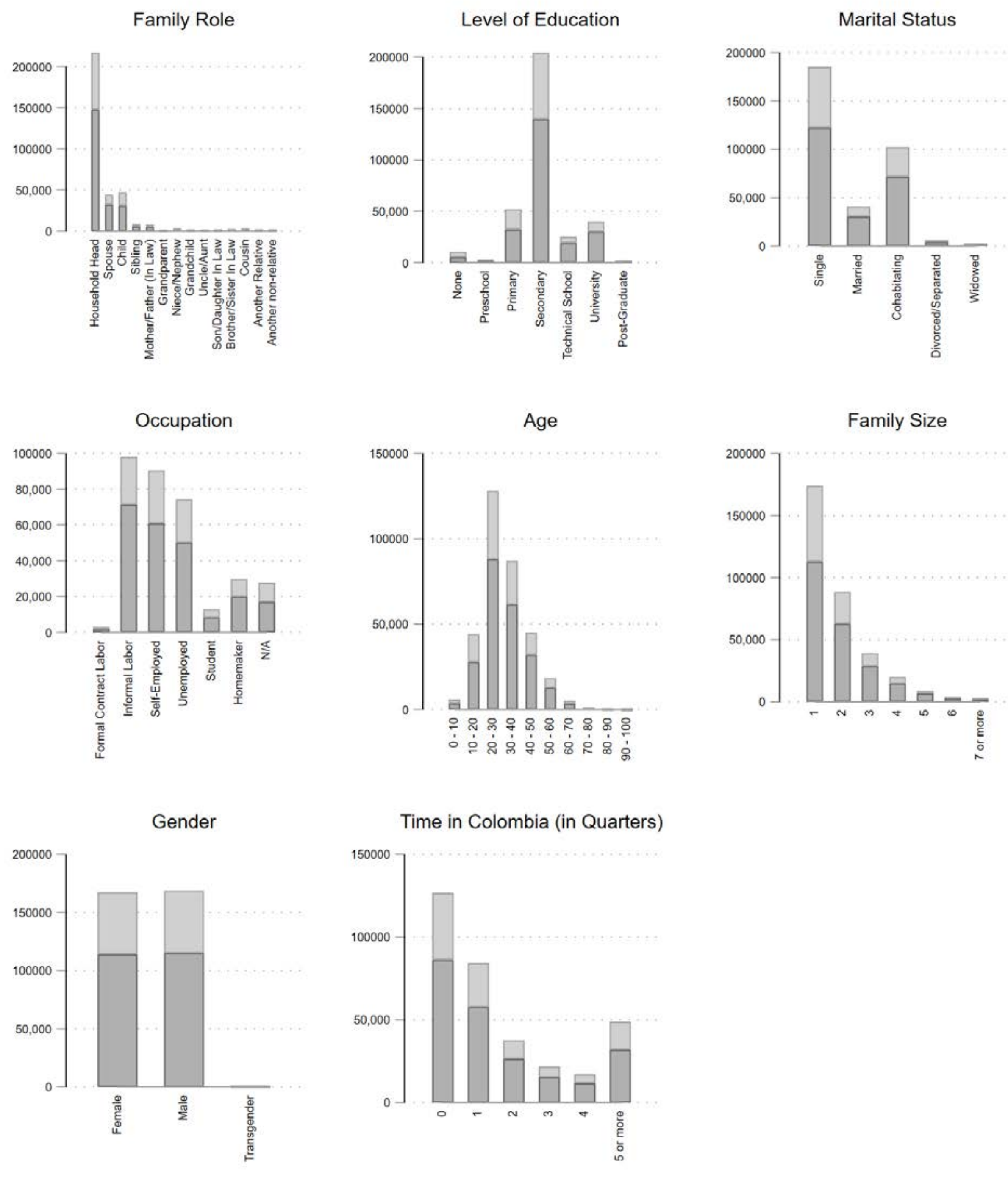
Notes: This table presents sample averages for different characteristics of the individuals in our sample, conditional on having received the PEP visa vs. not having received it. Mean differences are reported in the last column with corresponding p-value levels with the usual notation. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table A2: **Rainfall and Census Regressions**

	(1) Census No.	(2) Census No.	(3) Census No.
Rain	-5,576***	466***	447***
	(386)	(120)	(117)
Week of Census FEs		Y	Y
Municipality FEs		Y	Y
Family Characteristics Controls			Y
Observations	253,961	253,873	253,135
R^2	.00099	.95	.95

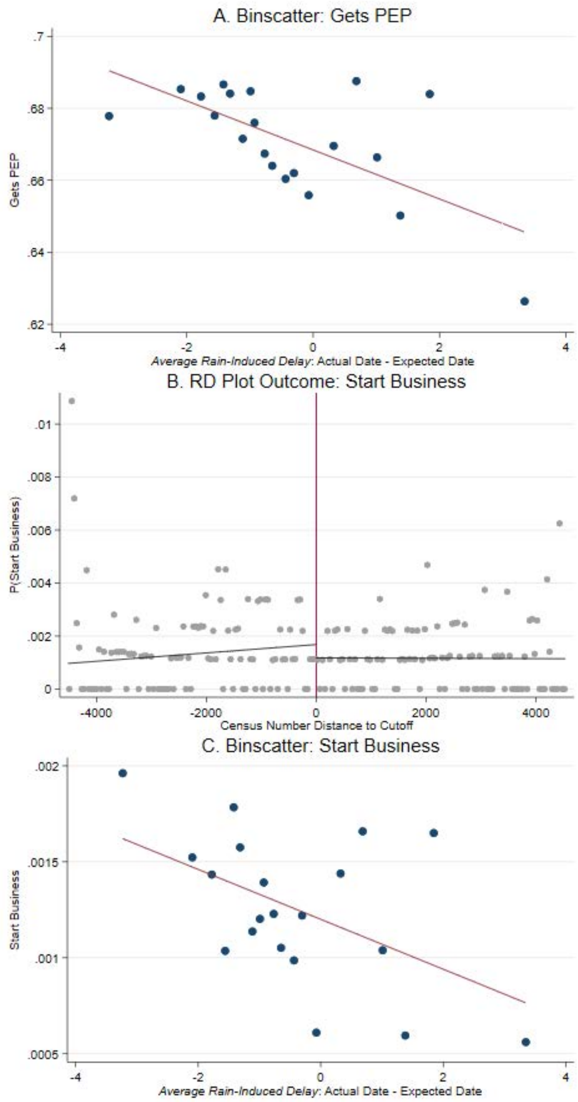
Notes: This table shows regressions predicting the running variable (Census number) from the level of precipitation on the day of the migrant's registration in his or her registration municipality. Family Controls include the percentage of female members, the percentage of pregnant members, the total members in the household, and the maximum, minimum and average ages. Standard errors are clustered by family. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Figure A5: **Observables of RAMV registrants with and without PEP**



Notes: These figures plot number of people registered in RAMV across several observable characteristics, each one with defined categories. Within each characteristic, the bars show the number of registrants within a category that received a PEP visa (in dark grey) and who did not receive a PEP visa (light grey).

Figure A6: **Distribution of Treatment and Outcome across the RD Threshold in Stacked Setup and in Binned Scatterplot**



Notes: This figure plots relationships between two central variables in our analysis –getting a PEP and starting a business– and distance to the closest census number cutoff for RAMV registrants (upper left and bottom left panels, respectively). The panels on the right plot in a binscatter the relationship between getting a PEP (top) and starting a business (bottom) against the delay generated by the census numbers for getting a visa. It shows a clear negative correlation – the higher the delay, the less likely individuals are to both getting a visa and starting a business (the latter can be thought of as a visualization of our reduced form).

Table A3: Selection into residency permit (PEP) across migrants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Logit PEP	*Logit* PEP	*Logit* PEP	*Logit* PEP	*Logit* PEP	*OLS* Reg. Date	*OLS* Reg. Date Subsample: Has PEP
Is Single	-0.191***				-0.177***	0.511***	0.577***
	(0.0221)				(0.0215)	(0.190)	(0.212)
Is Male	0.00327				0.0200*	-0.568***	-0.467***
	(0.0118)				(0.0119)	(0.158)	(0.168)
Age (Omitted 18-29):							
18 or less	-0.392***	-0.464***			0.0530	1.151**	1.033
	(0.0556)	(0.0517)			(0.0697)	(0.559)	(0.648)
30-50	0.104***	0.137***			0.145***	-1.431***	-1.657***
	(0.0331)	(0.0321)			(0.0302)	(0.137)	(0.0939)
50-65	-0.0126	0.0290			0.135***	-1.617***	-1.873***
	(0.0612)	(0.0609)			(0.0425)	(0.356)	(0.356)
over 65	-1.359***	-1.336***			-1.079***	-2.881	0.444
	(0.163)	(0.172)			(0.146)	(1.816)	(1.180)
Education: (Omitted Primary)							
None			-0.299***		-0.171***	0.438**	0.629
			(0.0310)		(0.0377)	(0.186)	(0.403)
Preschool			-0.0601		0.0177	0.434	0.759**
			(0.0431)		(0.0295)	(0.270)	(0.324)
Secondary			0.364***		0.320***	0.987**	1.352***
			(0.0471)		(0.0468)	(0.470)	(0.466)
Technical School			0.757***		0.685***	0.775	1.374***
			(0.0688)		(0.0702)	(0.522)	(0.489)
University			0.732***		0.670***	1.163**	1.783***
			(0.0567)		(0.0526)	(0.474)	(0.469)
Post-Graduate			0.436***		0.370***	-1.468**	0.0972
			(0.0420)		(0.0390)	(0.739)	(0.991)
Family Role:							
Head of Household				0.364***	-0.0921***	0.758	0.363
				(0.0372)	(0.0227)	(0.768)	(0.623)
Spouse				0.564***	0.0803***	0.138	-0.128
				(0.0138)	(0.0148)	(0.259)	(0.218)
Occupation: (Omitted Informal Labor)							
Formal Contract Labor					-0.345***	-3.048*	-3.900**
					(0.120)	(1.815)	(1.664)
Self-Employed					-0.252***	0.351	0.0846
					(0.0771)	(1.121)	(1.081)
Unemployed					-0.274***	1.631	1.187
					(0.0391)	(1.256)	(1.058)
Student					-0.228***	0.752	-0.992
					(0.0660)	(0.803)	(0.715)
Homemaker					-0.289***	3.425***	2.735***
					(0.0607)	(0.884)	(0.826)
Quarters in Colombia					-0.00590	-0.132***	-0.160***
					(0.00492)	(0.0311)	(0.0411)
Observations	443018	443018	443014	443018	435538	435538	276684
R^2						0.006	0.007

Significance reported as: * p <0.10, ** p <0.05, *** p <0.01.

B Firm Characteristics

This section describes the characteristics of the firms in our sample, overall and by entrepreneurs who did and did not receive the PEP visa.

Figure B1 plots the geographic distribution of the firms in our sample across Colombian national territory. The departments with the highest share of firms are those in the border (Norte de Santander and Arauca) where there is a large number of Venezuelans, as well as the capital city, Bogota, and departments with large cities or near large cities, such as Cundinamarca, Valle del Cauca (home to Cali), and Antioquia (home to Medellin). We see, however, presence of migrant entrepreneurs all across the national territory.

Figure B2 presents the same visualization but only for firms created by entrepreneurs who did get the PEP visa. Here we see pretty much the same pattern as in the previous figure, with firms created all over the territory but the departments with the highest proportions are such in the border with Venezuela and that are home to the largest cities.

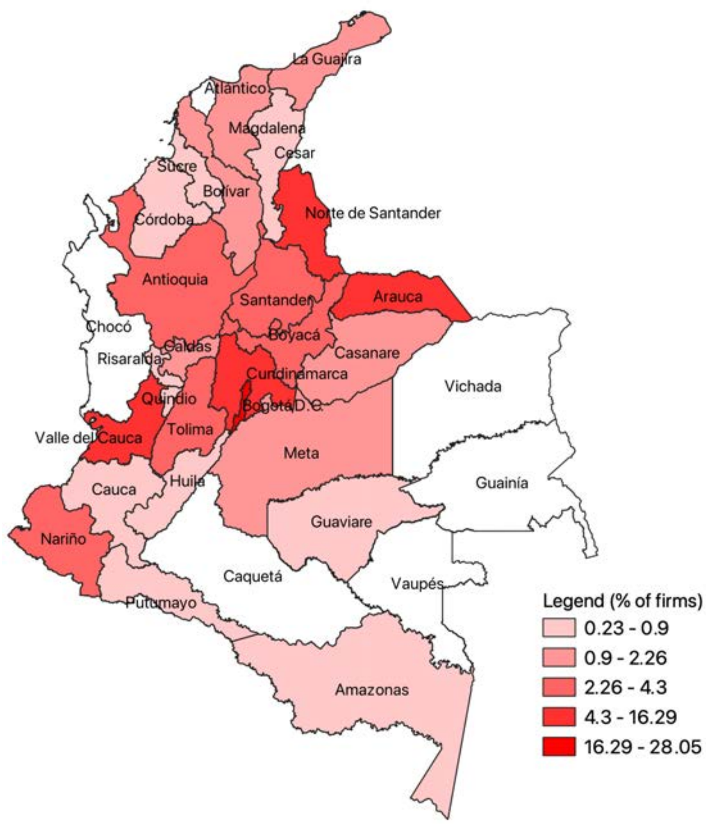
Finally, B3 presents the geographic distribution of firms by entrepreneurs without the PEP visa, which corresponds to a much more limited sample. As such, there is many more departments without firm creation at all, but yet, we see the same pattern: highest share of firms in departments that are in the border (Norte de Santander) and in departments that host large cities (Bogota, Cundinamarca, Antioquia and Valle del Cauca).

Thus, we find that the geographic distribution of firms across those individuals with and without PEP follows a similar pattern.

We also present, in Table B1, the distribution of sectors to which the 442 firms in our sample belong to, as defined by the ISIC 3-digit codes. The table shows that over 27% of the firms in our samples are in the "personal services" sector. The vast majority of firms in this category corresponds mostly to hairdressing and beauty treatment (115 firms). Over 18% of firms are in the prepared food industry, but there is also under over 10% of firms in retail of foods, including groceries (code 472), sale of alcoholic beverages (code 563) and manufacturing of food products (code 108). Another common economic activity is retail trade of all kinds (codes 471, 475, 477 and 479). All in all, the vast majority of firms in our sample are small service, food and retail establishments. See the table for the full list of industries.

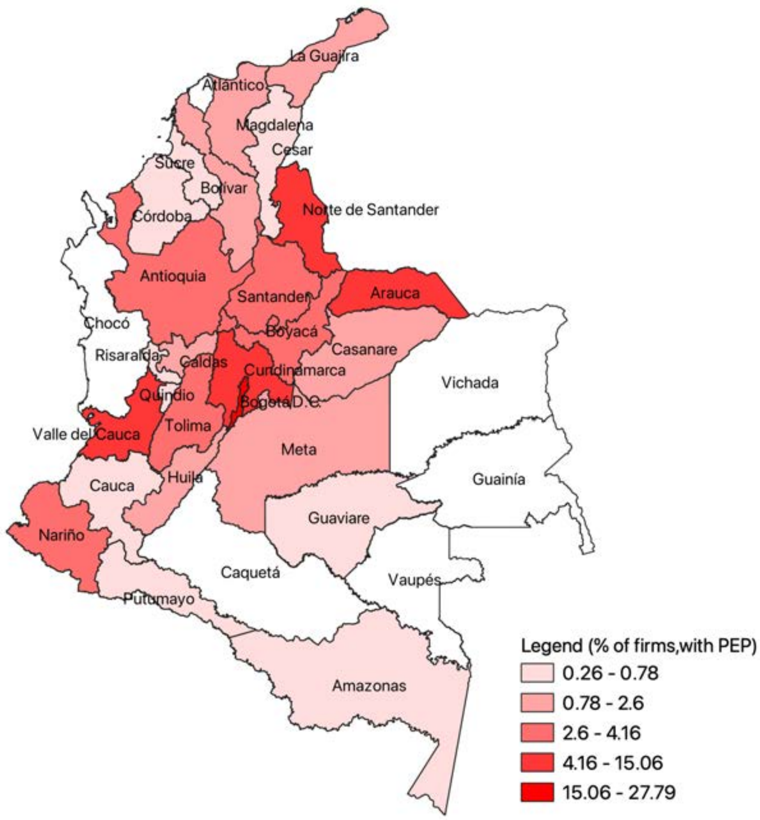
Table B2 present the share of firms by industry and by whether the owners or entrepreneurship received or not the PEP visa. Here we see, too, that the distribution of firms by industry or economic sector follows a similar pattern among individuals with and without the PEP visa. Most of the firms, regardless of the migratory status of the owners, are in personal services, the food business, and retail.

Figure B1: **Geographic Distribution New Firms**



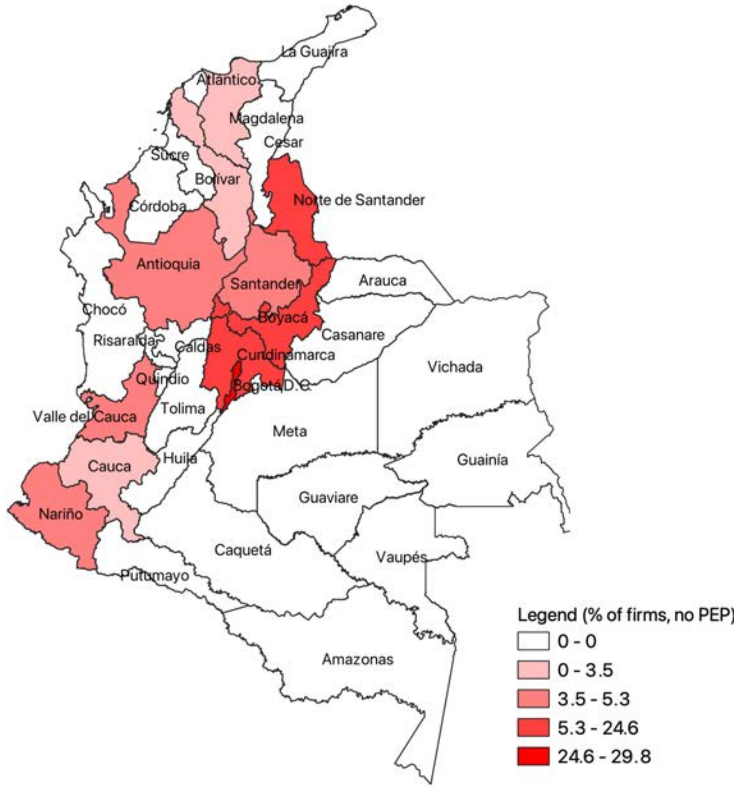
Notes: This figure visualizes the geographic distribution of new firms across the different departments of Colombia among entrepreneurs. Darker shades imply a larger share of firms being registered in that department, according to the legend.

Figure B2: **Geographic Distribution New Firms, owners with PEP**



Notes: This figure visualizes the geographic distribution of new firms across the different departments of Colombia among entrepreneurs who received the PEP visa. Darker shades imply a larger share of firms being registered in that department, according to the legend.

Figure B3: **Geographic Distribution New Firms, owners without PEP**



Notes: This figure visualizes the geographic distribution of new firms across the different departments of Colombia among entrepreneurs who did not receive the PEP visa. Darker shades imply a larger share of firms being registered in that department, according to the legend.

Table B1: Firms by ISIC Group

ISIC	Description	Firms	%
960	Other personal service activities	107	19.21
561	Restaurant, cafeteria and mobile food service activities	75	13.46
471	Retail trade in non-specialized establishments	24	4.31
108	Manufacture of other food products	16	2.87
472	Retail trade of food (groceries in general), beverages and tobacco, in specialized establishments	16	2.87
563	Sale of alcoholic beverages for consumption within the establishment	16	2.87
477	Retail sale of other products in specialized establishments	14	2.51
829	Business support service activities n.c.p.	14	2.51
452	Maintenance and repair of automobiles	13	2.33
475	Retail trade of other household goods in specialized establishments	11	1.97
479	Retail trade not carried out in establishments, stalls or markets	11	1.97
731	Advertising	5	0.90
952	Maintenance and repair of personal effects and household goods	5	0.90
951	Maintenance and repair of computers and communications equipment	4	0.72
522	Activities of stations, tracks and complementary services for transport	4	0.72
474	Retail trade of computer and communications equipment, in specialized establishments	4	0.72
855	Other types of education	4	0.72
141	Manufacture of garments, except leather garments	4	0.72
900	Creative, artistic and entertainment activities	3	0.54
532	Messaging activities	3	0.54
433	Completion and finishing of buildings and civil engineering works	3	0.54
202	Manufacture of other chemicals	3	0.54
321	Manufacture of jewellery, costume jewelery and related articles	3	0.54
-	All Others	195	35.01

This table presents the distribution of industries (3-digit ISIC codes) of the 442 firms in our sample.

Table B2: Firms by ISIC Group, by entrepreneurs with and without PEP

ISIC	Desc	% PEP	% No PEP
960	Other personal service activities	25.21	30.77
561	Restaurant, cafeteria and mobile food service activities	18.84	13.46
471	Retail trade in non-specialized establishments	5.54	7.69
108	Manufacture of other food products	4.16	1.92
472	Retail trade of food (groceries in general), beverages and tobacco, in specialized establishments	4.16	1.92
477	Retail sale of other products in specialized establishments	3.60	1.92
563	Sale of alcoholic beverages for consumption within the establishment	3.60	5.77
452	Maintenance and repair of automobiles	3.32	1.92
479	Retail trade not carried out in establishments, stalls or markets	2.77	1.92
475	Retail trade of other household goods in specialized establishments	1.94	7.69
141	Manufacture of garments, except leather garments	0.83	1.92
522	Activities of stations, tracks and complementary services for transport	0.55	3.85
855	Other types of education	0.55	3.85
202	Manufacture of other chemicals	0.55	1.92
321	Manufacture of jewelry, costume jewelry and related articles	0.55	1.92
-	All Others	23.82	11.54

This table presents the distribution of industries (3-digit ISIC codes) of the 442 firms in our sample across individuals with and without the PEP visa. Firms owned by individuals with PEP are 385 while firms owned by individuals without PEP are 57.