NBER WORKING PAPER SERIES

THE ECONOMIC EFFECTS OF IMMIGRATION PARDONS: EVIDENCE FROM VENEZUELAN ENTREPRENEURS

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

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. All errors and omissions are our own.

The Economic Effects of Immigration Pardons: Evidence from Venezuelan Entrepreneurs
Dany Bahar, Bo Cowgill, and Jorge Guzman
NBER Working Paper No. 30624
November 2022
JEL No. J26,K37,L26

ABSTRACT

This paper shows that providing undocumented immigrants with an immigration pardon, or amnesty, increases their economic activity in the form of higher entrepreneurship. Using administrative census data linked to the complete formal business registry, we study a 2018 policy shift in Colombia that made nearly half a million Venezuelan undocumented migrants eligible for a pardon. Our identification uses quasi-random variation in the amount of time available to get the pardon, introducing a novel regression discontinuity approach to study this policy. Receiving the pardon has small initial effects but raises formal firm formation to close to parity with native Colombians by 2022. This impact is concentrated on individuals active in the labor force, and on sole proprietorships rather than sociedades (limited liability entities). The new firms created include both employer and non-employer firms and are relatively low on assets. In panel data specifications, the effect of the pardon on firm formation is twice the effect of migration. Our heterogeneous effects suggest a mechanism whereby legalization induces greater investments of time in developing new firms.

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

Over 10 million U.S. residents are undocumented immigrants, with large quantities in many other countries. A recurrent policy proposal is to develop a path to full legal residency by offering a pardon or amnesty. Popular arguments in favor of this policy often focus on humanitarian issues (Gonzales, 2016).[1] However, pardons could also have economic effects. Although a literature about the economic effects of immigration exists, this literature often studies the bundled effects of physical relocation and legal status. Undocumented immigrants only experience one part of this bundle (physical relocation), unless a pardon is given to them. Some of the economic effects of immigrants (e.g., the impact on local consumption) could operate entirely through this physical relocation aspect. Understanding the economic benefits of pardons requires isolating legal status from physical relocation, both theoretically and empirically.

In this paper, we focus on a particular economic channel for pardons: local economic investments with positive spillovers. We ask why these outcomes are important to the *legal status* and documentation part of the traditional immigration bundle, and why they are thus central to policy choices about pardons.

Specifically, we study the effect of an immigration pardon on immigrant entrepreneurship. Founding a new business is a quintessential investment with positive economic spillovers. Startups are a primary driver of new job creation (Haltiwanger et al., 2013) and the commercialization of new ideas (Akcigit and Kerr, 2018). Prior research shows that immigrants start businesses at higher rates (Azoulay et al., 2020; Kerr and Kerr, 2020).

However, it is unclear what role a pardon itself would play in immigrant entrepreneurship (or other forms of investment). Prior work often focuses on legal immigration (Hunt,

[1] For example, in the Gang of Eight immigration bill of 2013, the rationale for amnesty was the "11 million stories of heartbreak and suffering" created by "living in the shadows" (Gomez and Davis, 2013). Similarly, President Barack Obama's support for the DREAM Act emphasized the importance of "mend[ing] our nation's immigration policy, to make it fairer, more efficient, and more just" (Obama, 2012).

2011), featuring a bundle of relocation and legal rights. Higher levels of entrepreneurship by immigrants is often attributed to self-selection (Borjas, 1987) based on characteristics that both pre-date and outlast the act of immigration (documented or otherwise). Fairlie and Lofstrom (2015) notes that the effect of legalization on entrepreneurship is theoretically ambiguous. In addition, the majority of work about immigrants and entrepreneurs has focused on *skilled* immigrants and their commercialization of new technology (Saxenian, 2002). Undocumented immigrants are often less skilled.

While there are several ways in which an immigration pardon could lead to more entrepreneurship, we hypothesize two key mechanisms. First: Legal status confers better property rights. Investment choices should depend on the appropriability of returns (Chand and Clemens, 2019). Immigrants with legal status can justify greater entrepreneurial effort, knowing they can enjoy the fruits of their work. Second: Legal status improves access to complementary inputs. Resources such as banking, borrowing, and court systems complement an immigrant's own efforts to develop a business (but require documentation to access). If these complements are available, immigrants can rationalize higher levels of entrepreneurial effort. Although these mechanisms create an avenue for higher entrepreneurship, legal status can also increase *opportunity costs* of entrepreneurship (for example, by increasing the returns to normal employment, Kossoudji and Cobb-Clark 2002).

The effect of pardons on investment is therefore ambiguous, and we turn to empirics to study these effects in the real-world. Our data comes from an immigration pardon on Venezuelan migrants living in Colombia without documentation. Colombia provides a unique policy experiment through the *Permiso Especial de Permanencia* (PEP), a pardon to about 300,000 undocumented Venezuelan immigrants in 2018. In order to manage high volumes, Colombian authorities assigned migrants different start dates to register for the pardon. This created (effectively) exogenous variation in the amount of time available to obtain access to a pardon. Start dates were based on cutoffs along the migrants' previously-assigned administrative IDs. Each migrant enjoyed more (or less) time, depending on whether they fell above, or below, one or several of the discontinuities created by the thresholds of each bracket.

To exploit the variation created by these discontinuities, we introduce a novel extension to the regression discontinuity design (RD, Imbens and Lemieux, 2008) to use variation across the (sometimes multiple) thresholds created by each bracket. Our design harnesses measurement error (and other "noise" in the running variable), and explicitly incorporates this measurement error into an identification strategy.[2] This improvement also provides us with two benefits. First, we can use a principled, design-based approach for using observations farther from the cutoffs and hence include a larger sample in our analysis. Second, by using a larger sample, we can better examine mechanisms and heterogeneous effects, particularly on sub-populations. This tends to be challenging in typical RD setups.

Using this approach, we uncover five interrelated findings about undocumented immigrants, pardons, and entrepreneurship.

First, the additional time substantially shapes the selection of immigrants into the pardon. A 15-day increase in the time to get the PEP is associated with a 7.5 percentage point increase in the likelihood of applying to it (11% of the mean). Differences of 15 days or more actually occur in our data. Even for something as high-stakes as an immigration pardon, the time available to file the paperwork makes a difference. This shows that regulatory complexity and policy design matter (Davis, 2017).

Second, receiving the PEP increases entrepreneurship. Migrants who randomly get the PEP through our instrument are more likely to start a company during our sample period by 1.6 percentage points. This effect is economically significant. It is 10$\times$ the mean level of entrepreneurship in our data, 0.16%. The effect increases over time: while the impact of PEP on starting a firm in 2018 is 0.2%, it raises to 0.58% by 2022, a value close to the

[2] The design is also related to methods for composite treatments (Borusyak and Hull, 2021) and to noise-induced designs in Eckles et al. (2020).

Colombian native rate, which we estimate at 0.7%. An immigration pardon increases immigrant investment in a region, as evidenced by entrepreneurship.

Third, these resulting 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. We find that the impact is larger for sole proprietorships rather than higher-quality *sociedades* (corporations and LLCs). However, we find a significant increase in both types of firm formation.

Fourth, we decompose the immigrant entrepreneurship effect into a physical relocation effect and a legalization effect. This decomposition uses data about the timing of new firm formation. Even if we conservatively attribute 100% of pre-pardon entrepreneurship to migration, the effect of legalization is over twice as large. This suggests that the legal rights of migrants, and not only their physical presence in a country, are key for entrepreneurial investment.

Finally, our heterogeneous results show higher effects in groups with greater time to spend on entrepreneurship. Because undocumented immigrants are mostly poor, investments would come in the form of time and effort (rather than money). Older people and heads of household likely have competing responsibilities. We find higher effects on younger people and non-head-of-households. Although these groups are traditionally less entrepreneurial, and do not tend to have access to additional capital, they may have greater access to time to invest in the venture. Consistent with this, we find that our effect is larger for firms with lower financial assets at founding. These results highlight the importance of time rather than access to capital as a vehicle for the pardon’s effect.

Together, these results present a new perspective on the economic effects of immigration pardons. In the next section, we summarize our contribution to related literature in Section

2. In the remainder of our paper, Section 3 describes our empirical setting, and Section 4 covers or methodology conceptually. In Section 5 we operationalize the strategy in our setting, and Section 6 contains empirical results. A discussion in Section 7 concludes.

2 Related Literature

Our results contribute to multiple areas of the literature on entrepreneurship, immigration, and econometric methods.

Institutions and Entrepreneurship. 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 a self-reported survey and synthetic control design to study whether U.S. immigration reforms in 1986 influenced self-employment.

Inequality of Entrepreneurship. The literature about inequality of entrepreneurship has recently moved beyond general treatment effects to emphasize the inclusion of groups, such as women or the poor, into firm formation (Guzman and Kacperczyk, 2019; Field et al., 2013, 2016). We add to this line of work by focusing on undocumented immigrants, showing that the *effective* legal rights faced by individuals (not only the general legal environment), are key to their inclusion. This is true for other marginalized groups as well.

Economics of Immigration. Our results also contribute to the broader literature about the economics of immigration. 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 regular migratory status due to the policy shock. While prior work has looked at effects of legalization of migratory status in areas such as crime (Baker, 2015; Pinotti, 2017; Ibáñez et al., 2021), on consumption (Dustmann et al., 2017), markets access (Bahar et al., 2021; Amuedo-Dorantes et al., 2007), health (Giuntella and Lonsky, 2020), and education (Kuka et al., 2020), our paper is the first to study the role of pardons in shaping the investment choices of immigrants.

Treatment Effects from Thresholds. Lastly, our novel regression discontinuity (RD) contributes to the methodological literature for estimating treatment effects using threshold-based designs. RD designs are usually praised for their transparency and reproducibility. However, researchers have noted that RD has key limitations (Eckles et al., 2020; Borusyak and Hull, 2021). By focusing on treatment effects on the threshold, the method limits the external validity of the estimand. In addition, windows (bandwidths) around the threshold limit sample sizes and precision, and researchers' ability to study heterogeneity. As classically imagined, RD is also conceptually different than many other causal inference strategies. In design-based causal inference, estimators for a treatment effects are based on a claim of randomness in the treatment assignment mechanism. By contrast, estimation strategies for RD have been conceptualized through continuity arguments; the continuity of the conditional response function is assumed to be continuous, and treatment effects are estimated as the difference between the limits approaching the threshold from the left and right. Although the intuition of local randomization is sometimes used to explain RDs, it is rarely formally incorporated into inference.

Our approach lets researchers achieve more with threshold-based treatment assignment. To do this, we exploit knowledge of how the running variable is constructed. We add measurement error, which is a common feature in many settings, and use knowledge of what observable variables enter the running variable. In doing so, we make RD design-based by creating a plausible source of exogenous variation arising from measurement error. Our design is related to other methodologies that harness noise in judgements as a research tool (Cowgill, 2018).

Two particularly related papers are Angrist and Rokkanen (2015) and Rokkanen (2015). Rather than using knowledge of measurement error, 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 strong parametric assumptions for the extrapolation.[3]

By contrast, our approach imposes no parametric restrictions. We use explicit knowledge about measurement error to avoid the assumptions of the extrapolation-based approaches. In our setup, the structure of the extrapolation is driven by knowledge of random shocks and a known map (a threshold) between shocks and treatment. Our approach does require the researcher to collect information about the distribution of measurement error.

[3]In particular, the assumptions require that the distribution of the latent variable, the running variable and the noisy measurements be jointly normal, and that the response functions for potential outcomes (conditional on the latent variable) be linear.

3 Empirical Setting

Nearly 2 million Venezuelans live in Colombia today, representing about 3.6% of Colombia's population. Most of these immigrants arrived starting in 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.

Migration from Venezuelans occurred in several waves. The first major wave occurred in 2018 as a result of hyperinflation when price controls caused shortages of key products and a drop of about 60% of GDP. This led to the departure of 2.3 million Venezuelans, with the majority ending in Colombia (La Nacion, 2020; Kurmanaev, 2019).

As is typical in crisis-driven migration, a large share of the Venezuelans who migrated to Colombia fled their homes, 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.

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.

3.1 RAMV: Census of Undocumented Immigrants

To better understand the growing 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 pardon 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 document all immigrants to understand the problem of immigration and have data for the design of government activities.[4]

The RAMV used a massive public advertisement campaign to attract Venezuelan immigrants to voluntarily provide personal information.[5] The Colombian government explicitly stated that registering will *not* result in deportations or negative legal consequences. This statement was credible, Colombia has traditionally had 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.

The census drive successfully surveyed 442,462 undocumented Venezuelans belonging to 253,575 different households. This is approximately 75% of the undocumented migrants resided in Colombia according to official government estimates (although the exact number could not be known). The RAMV census officially terminated on June 9, 2018.[6]

Descriptive Statistics. The starting point of our data is the 331,646 immigrants that have a valid Venezuelan identification noted in the RAMV, 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 percent of this group has completed at least secondary education. Compared with the

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

[5]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.

[6]Evidence of nationality was required. A Venezuelan national ID (*documento nacional de identidad*), a document that is much more common than a valid passport, could be used to take the census.

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.

3.2 The Immigration Pardon: *Permiso Especial de Permanencia* (PEP)

In July 2018 — one month after the closing of 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, or pardon, authorizing their presence in Colombia. Any undocumented migrant who both i) had previously registered in the RAMV (which had by this point closed and did not re-open), and ii) had no criminal records or pending deportation orders, was eligible to apply. The pardon 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 that the visas could be renewed indefinitely, 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]

[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.

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

Importantly for our research topic (entrepreneurship), 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, all rights granted by PEP were to individuals, covering individual-level freedoms as described above.[11]

Expanding the rights of individuals can affect their incentives to invest in businesses. Because of the absence of changes to business' rights, the effects in this paper arrive through individuals' incentives to invest. Many of these incentives flow from legal rights through informal channels. For example: By granting freedom from deportation, PEP lowers the cost of informally recruiting customers to the business. Although PEP did not grant the explicit right to recruit customers, the freedom from deportation lowered the cost of this activity. 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 from the specific individuals involved

[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 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.

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.

Takeup and PEP Applications. Among the 442K respondents to the RAMV Census, 100% were eligible to apply for PEP. About 280K applied (64%). This magnitude is similar to that of other amnesty programs around the world, where takeup has regularly been below 100% (even with longer time horizons to apply).[12]

Given the potential for non-random self-selection into PEP, we develop an identification strategy to measure the effects of the PEP amnesty program using an instrument based on the PEP program's rollout. We outline this strategy in Section 3.4 and discuss it in greater detail later in the paper. Appendix Table A1 shows differences in observable characteristics between visa applicants and non-applicants.

3.3 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 containing nearly 8 million observations. In contrast to more developed economies, most self-employed Colombians do not register their firm as a formal firm 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

[12]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). Other sources report lower (44% as of 2020, *American Immigration Council*, 2021). Vox reported, "As big a difference as DACA has made for those who have it, it hasn't gotten as many applicants as many had hoped." `https://www.vox.com/2014/8/18/5999601/deferred-action-obama-immigration-daca`

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., providing 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. In addition, *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 migrants. For corporations and SASs, we match only to the public legal representative of the firm. This is not the lawyer, but the CEO or other top person ultimately responsible for the firm's management. For sole proprietorships, we also obtained the amount of assets and employment reported in the RUES each year, allowing us to consider the founding level of each for these startups.

3.4 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 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. Access to the *PEP* application system closed for all RAMV migrants on December 21st, 2018. Migrants in earlier batches therefore enjoyed not only early access, but also a longer horizon to apply for *PEP*'s quasi-permanent residence permit. Since the end of *PEP-RAMV* registration in December 2018, there have been no new programs to offer mass amnesty to undocumented immigrants in Colombia.[13]

The Colombian government divided migrants into batches based on their RAMV census numbers. Census numbers were assigned during the RAMV census process described above. The government's bureaucracy used a sequential numbering system (starting with 1) to assign each family a census number. 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.[14] These numbers were assigned in sequential order from a nationally centralized computer system.[15] A census participant occurring in Bogota may be assigned a number of N, while a separate family in Medellin could receive registration number $N + 1$ if they completed

[13]Future versions of *PEP* required applicants to be in Colombia legally, or obtain documentation from their home countries (Betts, 2019; Selee and Bolter, 2021).

[14]The census records contained the date of each migrant took the census, which already contains information about relative sequences.

[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.

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. Figure 2 shows an example of these advertisements containing a table of cutoff numbers and start dates.

3.5 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. As described above, the August 2018 announcement – including the existence of a pardon program, the extent of amnesty and access, and a choice to use the RAMV census as an eligibility criteria – were a surprise.[16] We present visual evidence of this in Figure A1. Google searches for RAMV in Columbia 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 design. Our identification is based on the noisiness of one's census number. Had *PEP* been widely known as an outcome of RAMV, there would be registrants who would need to know the cutoff numbers in order to

[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 Venezualan migrants.

game the identification strategy. Even then, efforts to game one's census number are likely to be noisy or imprecise. Ultimately, all RAMV registrants had the opportunity to obtain a *PEP* visa, but some had a longer window. Our identification is based on randomness in these window sizes. A long window is not necessary to obtain *PEP*; qualifying immigrants could sign up online on the first day of their 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. However, the registration number also contains noise coming from a variety of sources.[17] 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 cutoff choices also did not take into account boundaries in the calendar days in which a migrant registered for the RAMV. Most of

[17]For example, a migrant who leaves at 8AM 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, or other sources. The nation-wide 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 traffic delays in Cali, or a bureaucrat's speed at processing migrants in Barranquilla.

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 3 and A2 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 are no visible bunching before or after cutoffs.

Figure A5 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.

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 behavior could randomly change the date in which they can apply to *PEP*. The noise in the queue numbers 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.

4 Empirical Strategy: Regression Discontinuity with Measurement Error

We now present our strategy for utilizing the thresholds above to estimate a causal effect. While our strategy is motivated by our setting in Colombia, we present it as a more general empirical problem in which treatments are assigned using a threshold (as in an RD), but the running variable is measured with 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 others inputs may be influenced by quasi-random variation.

Our approach can be seen as an application of these ideas to the regression discontinuity.[18] Here we provide one approach to this adaptation in which shocks arise from measurement noise along the running variable. Below, we lay out four assumptions from which we build a strategy for estimating treatment effects.

Assumption 1 (Regression Discontinuity Preliminaries)**.** *There are $i = 1, ..., n$ observations $\{Y_i(0), Y_i(1), C_i, X_i\} \in \mathbb{R}^4$. Units are assigned a treatment $D = \mathbb{1}(X_i > k)$, where k is a known cutoff along a running variable X. For each observation, researchers observe $\{Y_i, C_i, X_i\}$ where $Y_i = Y(D_i)$.*

Assumption 2 (Latent Variable)**.** *A latent variable $X_i^\star$ exists as a function of C_i. Although $X_i^\star$ is not directly observed, observations with the same C have the same $\mathcal{E}[X^\star]$.*

C_i refers to observable characteristics. We do not require that $X_i^\star$ be monotonic in C or take any particular functional form, only that all C have the same underlying $E[X^\star]$.

Assumption 3 (Noisy Running Variable)**.** *The running variable X_i is noisily related to the latent variable $X_i^\star$. Specifically, $X_i = X_i^\star + \mathcal{E}_i$, where $\mathcal{E}_i$ is distributed F_i such that $\mathbb{E}[\mathcal{E}_i] = 0$ for all i.*

Natural interpretations of X_i are the "true value" of what is being measured, and X_i is

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

the noisy measurement of this.[19] While measurement error terms $\mathcal{E}_i$ is zero on average, the distribution of theses errors need not be identical, even among those with the same value of the running variable X_i. In fact, our notation allow each observation i to have a separate distribution of noise F_i. Some observations could be measured more noisily than others.

Assumption 4 (Exogeneity)**.** *The measurement error in X is exogenous, meaning that* $\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp \mathcal{E}_i$*, conditional on* C_i *and* $X_i^\star$*.*

Assumptions 3 and 4 raise the possibility of *counterfactual* X_i *running variable realizations.* Had the realization of $\mathcal{E}_i$ been different for any of the n observations, the resulting observation of the running variable X_i would be different despite the same $X_i^\star$. Because X_i would change under different draws of $\mathcal{E}_i$ (while the cutoff k remained the same), the treatment assignment $D_i = \mathbb{1}(X_i \geq k)$ could also change for exogenous, random reasons. In this sense, Assumptions 1-4 imply that the treatment assignments $D_i = \mathbb{1}(X_i \geq k)$ are exogenous, conditional on C_i.

Our identification strategy assumes the researcher knows the $\mathcal{E}_i$ distributions. It could be estimated from data by the researcher, or known from prior research. With this knowledge, the researcher can compute the entire distribution of potential X_is for each of the n observations.

Definition 1 (Propensity Score)**.** *For any* C_i*, the probability of treatment is* $\Pr(X_i \geq k)$*, the probability that noise observations place the value above the cutoff k.*

Definition 1 essentially creates a propensity score (Rosenbaum and Rubin, 1983), albeit a design-based score. This setup differs from the typical approach of directly estimating p-scores from data without a more structured model of treatment assignment. In particular, our setup uses the structure of the treatment (cutoffs) to estimate a distribution of shocks

[19]In cases where the meaning of a "true" measurement is unclear, X_i would refer to the average over infinite noisy unbiased measurements of the same latent quantity.

and the resulting probability of being treated. This estimated distribution is based on more primitive assumptions (1 and 2).[20]

Proposition 1 (Sharp RD)**.** *Given the assumptions above, the treatment effect is* $\tau_{Sharp} = \mathbb{E}[Y_i(1) - Y_i(0), \Pr(X_i \geq k)]$.

We call this version "sharp" because the treatment goes from zero to one when X_i crosses the threshold. However we can use the distribution of $\mathcal{E}_i$ to reason about the probability of this happening (given observables and measurement error). The sharp version can can be implemented using a regression that controls for the propensity score $\Pr(X_i \geq c)$ and C_i. Controlling for the propensity score $\Pr(X_i \geq c)$ in addition to C_i is particularly helpful when the distribution of $\mathcal{E}_i$s are not identical (e.g., if there is heteroskedasticity and the variance of $\mathcal{E}_i$ differs across the sample), or when the propensity score is otherwise a non-linear function of C_i. If C_i is unconfounded, only the propensity score is necessary (Rosenbaum and Rubin, 1983).

As in other research designs, conditioning on the propensity score eliminates selection bias coming from the conditioning variables and potential outcomes (Rosenbaum and Rubin, 1983). 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. The propensity score thus helps us use the largest set of observations for which we have partially randomized assignment (coming through measurement noise).

This addressees two key weaknesses in typical regression discontinuity designs that use a window surrounding the cutoff. 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, the use of windows limits the external validity of the sample. In typical RD designs, the treatment effect is

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

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 the $X_i^{\star}$s whose propensity scores are between zero and one.

This is 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, yielding precision improvements. It also allows 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 who had a propensity score between 0 and 1).[21]

Fuzzy Discontinuity. The noisy RD setup also permits a fuzzy version. This is the version used in our empirical application. Some slight modifications to the setup are required. 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 earlier "sharp version," we can reason about the probability of crossing the threshold at all (by using the distribution of measurement error $\mathcal{E}$).

Assumption 5 (Fuzzy Preliminaries)**.** *There are $i = 1, ..., n$ observations $\{Y_i(0), Y_i(1), C_i, X_i, D_i\} \in \mathbb{R}^5$. C_i is an observable characteristic. Let $Z_i = \mathbb{1}(X_i > k)$, where k is a known cutoff along a running variable X. For each observation, researchers observe $\{Y_i, C_i, X_i\}$ where $Y_i = Y(D_i)$.*

Assumption 6 (Monotonicity)**.** *For a subset of the n observations (compliers), $D_i = 1$ if $Z_i = 1$ but $D_i = 0$ if $Z_i = 0$. There are no observations (defiers) for whom $D_i = 0$ if $Z_i = 1$ but $D_i = 1$ if $Z_i = 0$.*

Assumptions 5 and 6 update Assumption 1, and the remainder of the assumptions remain. The following proposition updates Definition 1:

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

Definition 2 (Expected Instrument)**.** *For any C_i, the expected instrument is* $\Pr(X_i \geq k)$*, the probability that noise observations place the value above the cutoff k.*

In the fuzzy setup, $\Pr(X_i \geq k)$ is not the propensity score (i.e., propensity of treatment). Instead it is the "expected instrument." The expected instrument captures the idea certain C_i have a greater 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, observations with a high probability of exceeding the threshold may still have a relatively low probability of treatment.[22] In a fuzzy setup, the expected instrument is the propensity of falling above the threshold (and thus getting a $Z_i = 1$ instrument).

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 RD Estimand)**.** *Given the assumptions above, the treatment effect on the compliers can be estimated by:*

$$\tau_{Fuzzy} = \frac{E[Y_i|C_i, \Pr(X_i \geq c)] - E[Y_i|C_i, \Pr(X_i \geq c)]}{E[D_i|C_i, \Pr(X_i \geq c)] - E[D_i|C_i, \Pr(X_i \geq c)]} \tag{1}$$

The above can be estimated with 2SLS. 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.

[22]Under the monotonicity requirement of IV, observations with a high $\Pr(X_i \geq k)$ will be more likely to be treated than those with a low $\Pr(X_i \geq k)$.

4.1 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.

Incorporating Noise. In practice, implementing Definition 1 to create propensity scores requires information about the amount of noise. Researchers can obtain this data from other papers or prior sources. For some running variables, measurement noise may be documented by the creators of the running variable metric. For example, standardized test scores were designed by researchers to measure a latent variable. These researchers have principled ways to diagnose their metric through the literature on item response theory (Embretson and Reise, 2013), and student assessments contain published diagnostics around measurement error.

Lacking such prior knowledge, researchers may also build redundant measurements into a proactive research design in order to exploit knowledge of measurement errors. Assumption 3 allows $\mathcal{E}_i$ to be distributed differently across X_i (e.g., heteroskedasticity in measurement error). The requirement of knowing the noise distribution may pose a challenge in some applications. In these cases, our approach works if researchers use conservative assumptions about $\mathcal{E}_i$ (i.e., assumptions that underestimate the level of randomness).

Discrete Running Variables. This setup allows discrete running variables. Indeed, our applied application is technically a discrete variable (census numbers, or natural numbers).

Algorithms. 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 sampling error in the training set. Had the training sample been different, then the algorithm would use different weights or coefficients, and may ultimately score inputs differently. If researchers can quantify this measurement error, either through analytic prediction intervals, through cross-validation, or something else, then our strategy can be applied to settings featuring algorithmic running variables.

5 Operationalizing our Design in Colombia

In this section, we adopt the strategy above for using measurement noise to study Colombia's *Permiso Especial de Permanencia* (*PEP*) program (described in Section 2). The running variable in our setup is the census number. We use an adaptation of the fuzzy RD: Having a census number below a threshold more likely to obtain the treatment (the pardon), because the imigrant has more time to apply. However, crossing the threshold does not guarantee the pardon/treatment, because some migrants still do not apply for PEP given extra time.

In our implementation, the main outcome variable Y_i is the choice to become an entrepreneur and registering a firm. All migrants, including non-pardon holders before and after the PEP program, are eligible to create and register a new firm. In Colombia, as in the U.S., new firms can be created and registered by foreign citizens and do not require legal permanent residence in a country. Laws permitting business ownership and creation by foreign citizen exist around the globe to facilitate foreign investment in the domestic economy. All migrants in our sample are eligible to start and register a firm even without a PEP visa.

The instrument Z_i is the *window length* described in Section 3.5. This length refers to the amount of time each migrant was given to register for PEP documentation. As described earlier, the length of each migrant's PEP registration window was assigned through a noisy process in which census numbers (X_i, the running variable) are assigned relative to arbitrary cutoffs (k). Given this assignment process, there is likely a high degree of randomness in the window lengths. However, there may also be some degree of endogeneity to the window lengths if (say) systematically different people take the census early vs late.

Our implementation requires handling two issues. First, it must justify a model of noise in the assignment of census numbers. 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. Second, we need a strategy for using multiple discontinuities. Rather than one propensity score, we have twenty-two. The remainder of this section overviews our implementation details and validations for each of these points.

5.1 Noise Model

To model the distribution of noise realizations $\mathcal{E}_i$, we predict each family's census number as a function of observable characteristics and obtain the standard error of prediction for each family. Building on our knowledge of the institutional details of the PEP policy shock, we assume that deviations from these point estimates are independently drawn from a prediction distribution for each point, using the point estimate as the mean and the standard error of prediction to compute the variance.[23] To build this model, of noise, we begin by estimating the following linear model:

$$\begin{aligned} CensusNumber_i = &\beta_0 + \beta_1 Age_i + \beta_2 MorningRegistration_i + \beta_3 Female_i + \beta_4 Pregnant_i + \\ &\beta_5 NumRegistered_i + \beta_6 FamilySize_i + WeekOfCensusFEs_i + \mu_i \end{aligned} \tag{2}$$

[23]This imposes some parametric assumptions onto our noise distributions. An alternative approach would be to examine permutations in the residuals from a model of census numbers, conditional on the same observables. We have implemented this approach as well, and obtained very similar results to the parametric approach which we report below. Either way, draws of the measurement error term across different observations are not drawn from an identical distribution, but they are drawn independently.

where μ_i is an error term. Since census numbers were assigned to entire families (not to individuals), the unit of analysis is the family. The covariates in Eq. 2 relating to age, gender, and others are averages across all family members. About 64% of families had exactly one member. Standard errors of the model are clustered at the date the census (RAMV) is taken. This allows for a common shock to affect everyone who arrived on a particular day.[24] To aid in interpreting coefficients, we have standardized the census numbers (the outcome variable) so that coefficients can be read as standard deviations in Table 2.

Noise realizations represent deviations from average census numbers of migrants with similar characteristics, including the week they took the census. To measure the variance of this noise, we calculate the standard error of the fitted value ($\sigma_{est,i}$) for each observation, using the coefficients in Table 2 and their covariances (from Equation 2). For each observation, we then assume that counterfactual census numbers are drawn from a distribution with mean $\widehat{CensusNumber}_i$ and standard deviation of $\sigma_{est,i}$. In this approach, $\sigma^2_{est,i}$ represents the variance of noise around $\widehat{CensusNumber}_i$. Using this distribution, we can estimate the probability that any migrant i could have fallen into any of the twenty-two batches, using the cutoffs described in Section 3.4 (visualized in Figure 2).

Our estimation of Equation 2 in Table 2 has an R^2 of 0.98. We are able to explain a large amount of the variance by including fixed effects for the week that the migrant completed the RAMV census. Despite this high R^2, the residual of these models suggests a modest degree of noise in the census order. The average residual was about one tenth of a standard deviation of the distribution of census numbers. Standard errors of prediction were correspondingly small, but varied across our sample (indicating different F_i distributions of noise terms $\mathcal{E}_i$ across our observation

[24]The practical effect of clustering standard errors is to widen prediction intervals of the census number. Clustering errors makes it more possible that a large shock affected many people on the same day and thus greater measurement error.

As Table 2 shows, census numbers were very slightly correlated with several observable characteristics. Families are more likely to have low (early) census numbers if they have a pregnant member, have an older average age, or have a larger size. However, these differences are quantitatively small. For example, changing the age from 17 to 70 changes the census number by only 1 percent of its standard deviation, and being female changes the census number by 0.1% of its standard deviation.

One possibility for these small differences, speculated by our research partners in Colombia is that groups such as pregnant women and the elderly may have been directly moved to the front of the line by migrants as a show of courtesy. Alternatively, these groups may have simply been more organized, and/or busy and arrived earlier. Our data does not include the time of day that the migrant registered. However, to approximate the idea that people who arrive earlier in the day are different, we include a dummy variable for whether the family was below the median value census number for the day they arrived (*Registered in the Morning*).

Furthermore, while we have suggested a few reasons for why age and pregnancy may be correlated with slightly lower census numbers, our identifying assumptions are agnostic about these reasons. By controlling for age, pregnancy, and other factors, we account for their role in explaining these differences.

Figure 5 plots the distribution of differences in noise across census numbers thresholds in stacked regression discontinuity using data-driven parameters (Calonico et al., 2015; Cattaneo and Titiunik, 2021). Consistent with our model, we observe the largest differences in noise realizations are right at the threshold cutoff. For these migrants, falling below the cutoff leads to 2.5 more days to register for the PEP than above the cutoff, on average.

Identifying Assumptions. The identifying assumption is that, conditional on the terms used in the model, there are no such confounding variables that correlate with both census number noise and the probability of starting a firm.

5.2 Specifications

We perform our main estimates using two-stage least squares regressions (2SLS). The instrumented variable is PEP_i which identifies which migrants had a visa. We instrument PEP_i with the amount of delay (in days) that a migrant i experienced between the first day an RAMV taker was eligible to apply for a visa, and when i was eligible. This was based on i's census number and the thresholds. We call this instrument $ActualDelay_i$. However, as we have already discussed, the $ActualDelay_i$ is not entirely random. As such we control for the "expected instrument" – the expected amount of time based on the observables, particularly the week chosen to complete the census.

Let j represent a batch of migrants ($j \in J$) based on the cutoffs in census numbers described in Section 3.4, and let $p(i \text{ had delay } j)$ represent the probability that migrant i was assigned to batch j. We calculate this using the prediction distribution parameters for each point, derived from Equation 2 (described in Section 5.1). Using these intervals, we can estimate the distribution of potential census numbers under different noise realizations. The first stage equation is:

$$PEP_i = \beta_0 + \beta_1 \underbrace{ActualDelay_i}_{\text{Instrument}} + \beta_2 \underbrace{ExpectedDelay_i}_{\text{Expected Instrument}} + OtherControls_i + \epsilon_i \tag{3}$$

where ϵ_i is an error term. The general idea behind the first stage is: The actual delay is random, conditional on the expected delay. We vary the presence of $OtherControls_{ij}$ across specifications for transparency. When used, we control for the migrant's family role (head of household or not), marital status, occupation, gender, level of education and week of census completion (fixed effects). When controlling for our expected instrument (expected delays), the $OtherControls_{ij}$ terms do not significantly change our visa treatment coefficient.

Because of our understanding of the noise of the census numbering process, we can calculate the probability of each of the $J = 22$ possible delays. We call these probabilities $p(i \text{ had delay } j)$. Jointly, these are collinear with the expected delay, thus serve the same function as the expected instrument.

$$PEP_i = \beta_0 + \beta_1 \underbrace{ActualDelay_i}_{\text{Instrument}} + \underbrace{\sum_{j \in J} \delta_j p(i \text{ had delay } j)}_{\substack{\text{Collinear with } ExpectedDelay_i, \\ \text{the expected instrument}}} + OtherControls_i + \epsilon_i \tag{4}$$

This adds precision in some of our results, and we include both specifications in our results (the former in an appendix). Our second stage equation is:

$$Y_i = \beta_0 + \beta_1 \cdot \underbrace{PEP_i}_{\substack{\text{Instrumented} \\ \text{by Eq. 4}}} + \underbrace{\sum_{j \in J} \delta_j p(i \text{ had delay } j)}_{\substack{\text{Collinear with } ExpectedDelay_i, \\ \text{the expected instrument}}} + OtherControls_i + \eta_i \tag{5}$$

where η_i is an error term, and the $p(i \text{ had delay } j)$ are again collinear with the expected delay and serve the same function.

Finally, for completeness, we also consider a probit instrument, as recommended in Wooldridge (2010) to provide more precision in the presence of binary treatments. To do so, we first run a probit model of equation (3), then, we use the predicted value of this probit regression as the first stage instrument in our 2SLS model.

5.3 Balance and Profile of Compliers

We consider the instrument balance in Table 3 by comparing migrants that have a lower than expected delay (positive luck), to those that have a higher than expected delay. The two groups appear observably similar. All differences are economically very small and for the most part not statistically significant. This includes critical characteristics such as gender, age, marital status, family size, and whether they register in the morning. We conclude our instrument is balanced.

Next, we perform a better understanding of who are likely to be compliers vs always-takers in our setup. In Table 4, we report an implementation of the method of Marbach and Hangartner (2020) for profiling compliers. In essence, it uses the means of always-takers (migrants getting the PEP even if unlucky), never-takes (migrants not getting the PEP even if lucky), and the whole sample, to back out the mean of compliers. To apply this method, 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.

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

6 Results

6.1 First Stage Regressions

We now move to the centerpiece of our analysis, the impact of receiving the pardon on the probability of starting a formal firm. We report in Table 5 several versions of the first stage OLS regressions of our instrument's relationship to the getting the pardon. Standard errors are clustered by the date the census is taken.

Column (1) reports the raw correlation between getting the PEP pardon and *Actual Delay*—the difference between the date this migrant can begin registration and the earliest available registration date for all migrants. The coefficient is noisy and small.

Column (2) introduces an *Expected Delay* as a control. In this case *Actual Delay* turns negative and significant, with a coefficient of -.0056. Controlling for the expected delay, an additional week of actual delay to get the PEP translates to a 4% higher probability of getting it; an additional month of time to 17% higher probability. The partial F-statistic of this regression is 34. In short, the role of random delays in time to get a PEP appears meaningful for migrants.

Column (3) is our preferred specification. We use *Predicted Delay Gap*, defined as the actual minus the expected delay, as the instrument. This allows us to focus explicitly on the random variation introduced institutionally to our sample. The coefficient and first stage F-statistic are the same as Column (2).

Columns (4) and (5) are robustness tests for our specification. Column (4) introduces a range of additional controls into the regression. These include fixed-effects for week of registration, role in the family, gender, marital status, occupation, level of education, and the actual date in which the migrant registration period starts. The F-statistic of this regression and coefficients remain very similar. Column (5) uses *Prob. PEP* as an instrument, which is the fitted value of a probit regression replicating column (3). This approach is known to add additional power (Wooldridge, 2010). In our case, the F-statistic increases to 36.

Finally, column (6) shows a reduced form regression of our instrument on starting a firm. The coefficient is negative and significant, with a value of -.00008.

6.2 Main Results

We consider our main results in Table 6 through a linear probability model of the impact of getting the PEP visa on starting a firm.

Column (1) is the reduced form OLS of getting the pardon on starting a formal firm.

The coefficient of *PEP Pardon* is 0.0013 and significant. This estimate may underestimate the local average treatment effect of the PEP visa on the entrepreneurship of migrants if the majority of people registering for the PEP do it for reasons that are uncorrelated to entrepreneurship. For example, highly risk-adverse individuals may be likely to register for PEP, but risk aversion is also known to be negatively correlated with entrepreneurship. This is likely the case in our setting, where compliers actually influenced by our treatment are the minority of migrants.

Columns (2) through (4) replicate the 2SLS specifications of Table 5 using our instrument. The preferred specification is column (3), which includes a series of controls. The coefficient is 0.016 and significant. Getting the PEP leads to an increase of 1.6 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%.[25]

6.3 The Cumulative Impact of PEP Over Time

In Figure 6, we study heterogeneity over time on the impact of the PEP. To do so, we run regressions using our RD approach where the dependent variable is starting a firm in each specific year, then report all coefficients from these regressions. We use our data on firm formation and estimates of the Colombian population to also estimate the native firm formation rate.

Panel A reports the role of PEP in starting any type of formal firm. The impact of PEP on starting a firm in 2017 or 2016, before PEP, is zero using our RD approach. This makes sense and serves as a placebo test. The effect then begins on a positive trend that appears cumulatively larger. While PEP only increases 2018 business formation by 0.2 percentage

[25] 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 the 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.

points in 2018, this number raises to 0.5 by 2021, and 0.55 percentage points by 2022. These last two coefficients appear close to the rate of native formal firm formation in Colombia, which we estimate at 0.7% in 2022.

Panels B and C split the outcome variable depending on whether the firms are sole proprietorships or *sociedades* (limited liability companies). The effects are positive for both, though the bulk of the effect is clustered around sole proprietorships, which raise up to the Colombian native rate of formation, while *sociedades* are still only a fourth of the native rate. As we emphasized in Section 2, formal sole proprietorships still represent a relatively higher level of quality of entrepreneurship compared to the U.S., since the vast majority of startups are instead informal firms.

6.4 Comparing the Benefit of Moving into Colombia to the PEP

Next, we evaluate the economic importance of the PEP (legalization of the migrant) to the benefit of physical migration itself in Table 7. To do so, we go beyond our RD setup to instead consider our data in a panel format. For all individuals, we create a quarterly panel that begins in Q1 of 2014 up to Q4 of 2021. We 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 their impact on starting a firm in a difference-in-differences setup.

We consider several specifications that may incorporate controls, pre-trends, and individual and time fixed effects. Our preferred one is Column (2), which includes fixed effects for the week of registration and the municipality of registration, but does not include 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 magnitude 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. It appears the legal rights of the migrants are twice as important as the physical act of migration for entrepreneurship.

6.5 Heterogeneity and Mechanisms

Finally, we develop a better understanding of the mechanisms by studying heterogeneity in our results across both migrant characteristics and the types of firms created in Tables 8 through 10. Four key themes emerge.

First, the entrepreneurs randomly treated by the pardon created firms that are economically meaningful. About half of the firms in our sample have employees. When we split our dependent variable by employer and non-employer firms in Table 8, we see effects that are positive and significant for both. The coefficient is also higher, in absolute and relative terms, for employer firms. While these are still small firms (five employees or less), the PEP does seem to lead to relevant job creation.

Second, the individuals that are induced to start a new firm by the pardon are entrepreneurs strongly attached to the labor force. When we consider, in Table 9, the type of labor force status recorded by the migrant in the original RAMV census, we see that the effects are small and not significant for the unemployed and noisy for students and homemakers. In contrast, these coefficients are positive for those that are either informal workers or previously self-employed. Because these individuals are already employed somehow, this evidence is consistent with the impact of the pardon leading to an increasing benefit of entrepreneurship for those active in the labor force rather than remedial startups created out of necessity.

Third, the fact that the coefficient is not concentrated only on the self-employed suggests that the PEP caused the creation of new ventures, rather than simply making informal entrepreneurs formalize their firms. This is also evident in the results in columns (4) and (5) of Figure 6 that reported a larger effect of PEP for registering a new firm in the years 2020 and 2021, rather than 2018 and 2019. Was our effect simply the formalization of existing firms, we would have expected most of the new firm formation to occur quickly after the migrants receive PEP and then reduce over time. Yet, we see the opposite. The firms created by the pardon appear to be net-new entrepreneurship.

Finally, fourth, the treatment effects suggest that those that receive the pardon are able to spend more time developing a business, rather than it leading to higher financial capital. For example, in Table 10 we observe the effect is clustered around individuals below 30 years old and on those registered as spouses rather than as head of household. Younger individuals tend to have more time in their hands even though—at least in the U.S.—entrepreneurs are more typically between 35 and 50 years old (Azoulay et al., 2020). Similarly, individuals who identify as spouses may also have more time, but most entrepreneurship is usually undertaken by heads of household. This mechanism is also consistent with columns (3) and (4) of Table 8, which show our effect is larger for low-asset firms. Together, we interpret this evidence as consistent with a mechanism where PEP receivers invest additional time, rather than money, to develop new firms.

6.5.1 Regional Variation

As a final piece of analysis, we study geographical variation based on the location of the RAMV immigrants to consider potential differences in the regions in which migrants are located and the local institutions they face. Table A5 replicates our main specification adding an interaction (and an instrumented interaction) across three main areas—access to markets, property rights, and outside options. To assess these, we take advantage of variables that vary at the municipal or departmental level downloaded from the Centro de Estudios de Desarrollo Economico at the Universidad de los Andes. These include the distance to the state capital, distance to a wholesale market, or distance to the country capital, Bogota; whether the location was occupied by the Spanish in 1510, whether it had an indigenous population in 1535, and the total level of FARC violence in the final decade of the FARC, 2000-2010; and the formal employment rate.

In general (and while we do have one significant coefficient) we do not observe a consistent pattern of regional characteristics predicting differences in our treatment effects. At least within the relatively limited variation available across Colombian regions, we do not see a clear relationship of regional heterogeneity moderating the treatment effect of PEP on entrepreneurship.

7 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.

Our paper is focused on the *benefits* of the pathways to citizenship. Without doubt, these pathways may feature costs as well. One example of costs may be a potential surge in unauthorized immigration that is encouraged by amnesty. We do not directly study this, or other, effects.

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 to drive this conversation.

References

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.

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.

Akcigit, Ufuk and William R Kerr, "Growth through heterogeneous innovations," *Journal of Political Economy*, 2018, *126* (4), 1374–1443.

American Immigration Council, "Immigrants in the United States1," 2021.

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

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.

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.

Betts, Alexander, "Venezuelan Survival Migration as a Development Opportunity," 2019.

Borjas, George J, "Self-selection and the earnings of immigrants," *The American economic review*, 1987, pp. 531–553.

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.

_ , _ , and _ , "Optimal data-driven regression discontinuity plots," *Journal of the American Statistical Association*, 2015, *110* (512), 1753–1769.

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.

Chand, Satish and Michael A Clemens, "Human capital investment under exit options: Evidence from a natural quasi-experiment," 2019.

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.*

Davis, Steven J, "Regulatory complexity and policy uncertainty: headwinds of our own making," *Becker Friedman Institute for Research in economics working paper*, 2017, (2723980).

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.

Field, Erica, Rohini Pande, John Papp, and Natalia Rigol, "Does the classic microfinance model discourage entrepreneurship among the poor? Experimental evidence from India," *American Economic Review*, 2013, *103* (6), 2196–2226.

_ , Seema Jayachandran, Rohini Pande, and Natalia Rigol, "Friendship at work: Can peer effects catalyze female entrepreneurship?," *American Economic Journal: Economic Policy*, 2016, *8* (2), 125–53.

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, Alan and Susan Davis, "'Gang of Eight' immigration bill clears Senate hurdle," 2013.

Gonzales, Roberto G, *Lives in limbo: Undocumented and coming of age in America*, Univ of California Press, 2016.

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

_ and Aleksandra Olenka Kacperczyk, "Gender gap in entrepreneurship," *Research Policy*, 2019, *48* (7), 1666–1680.

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

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

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

Kerr, Sari Pekkala and William Kerr, "Immigrant entrepreneurship in America: Evidence from the survey of business owners 2007 & 2012," *Research Policy*, 2020, *49* (3), 103918.

_ , _ , 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.

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.

Kurmanaev, Anatoly, "Venezuela's Collapse Is the Worst Outside of War in Decades, Economists Say," *The New York Times*, 2019.

La Nacion, "El exodo incomparable: ya son 2,3 millones de venezolanos que huyeron por la crisis," 2020.

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

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

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.

Obama, Barack, 2012.

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.

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.

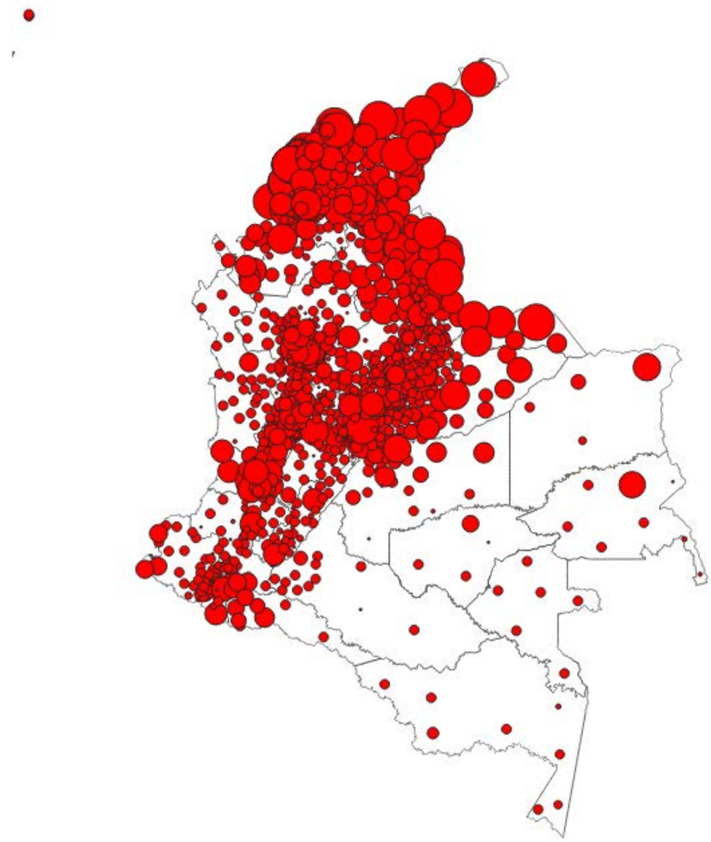
Saxenian, AnnaLee, "Silicon Valley's new immigrant high-growth entrepreneurs," *Economic development quarterly*, 2002, *16* (1), 20–31.

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

Selee, Andrew and Jessica Bolter, "Colombia's open-door policy: An innovative approach to displacement?," *International Migration*, 2021.

Wooldridge, Jeffrey M, *Econometric analysis of cross section and panel data*, MIT press, 2010.

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

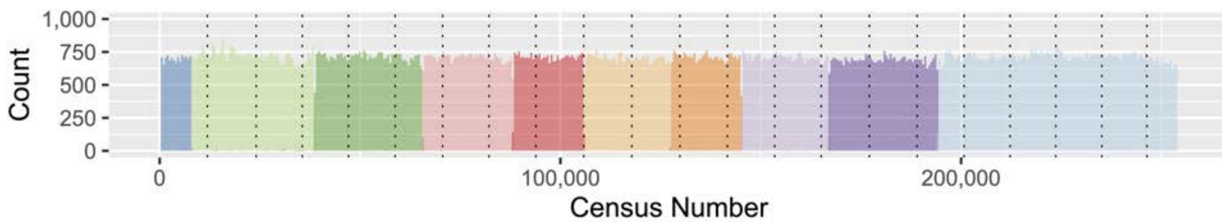


Notes: This figure visualizes all the municipalities in Colombia where there was a RAMV registration point. The markers are scaled using as weight the total number of people who registered in each municipality.

Figure 2: **Nationally Syndicated Ad for PEP Registration Cutoff Dates**

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.

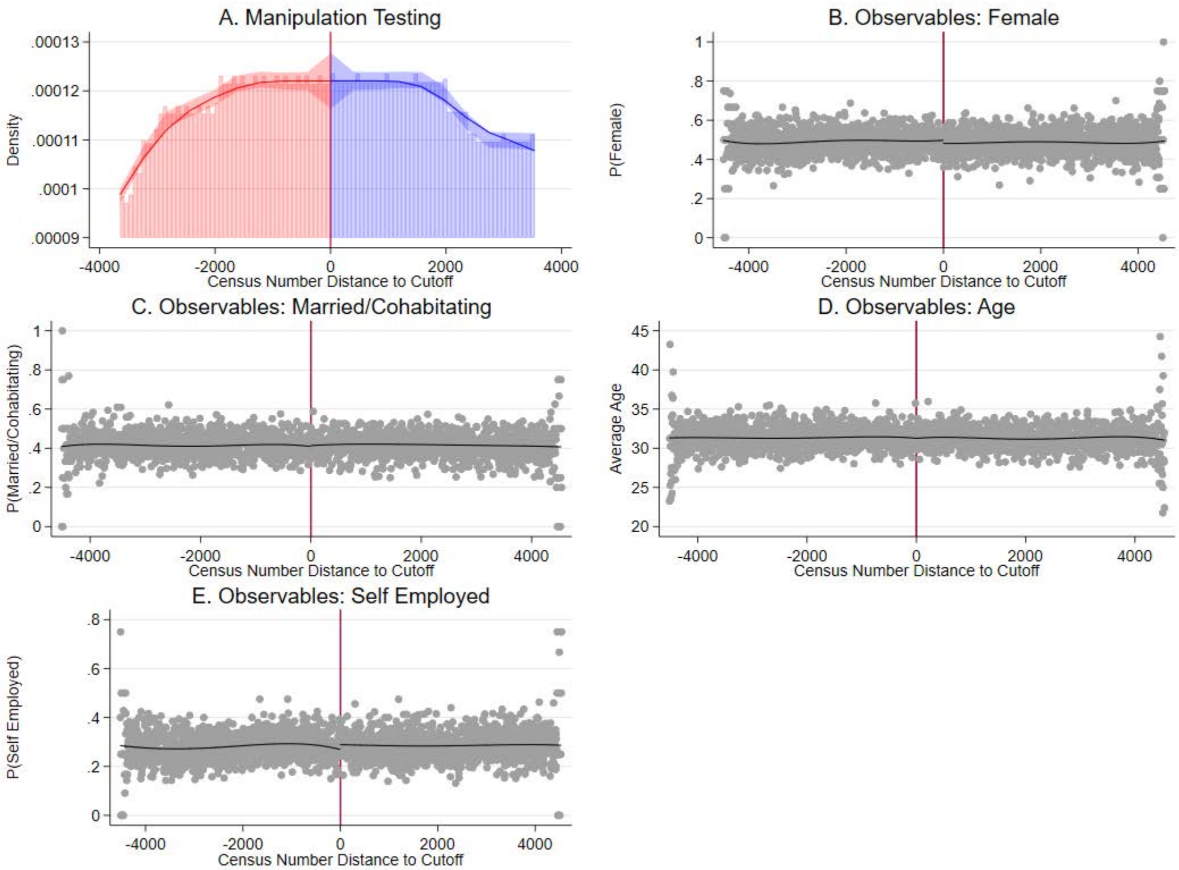
Figure 3: **Histogram of RAMV Census Numbers by Week (with Thresholds)**



Notes: Each color represents a week's worth of RAMV census numbers, assigned zero to ≈ 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 Figure 2 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. *Sociedades* represent formal separate limited liability entities created by migrants.

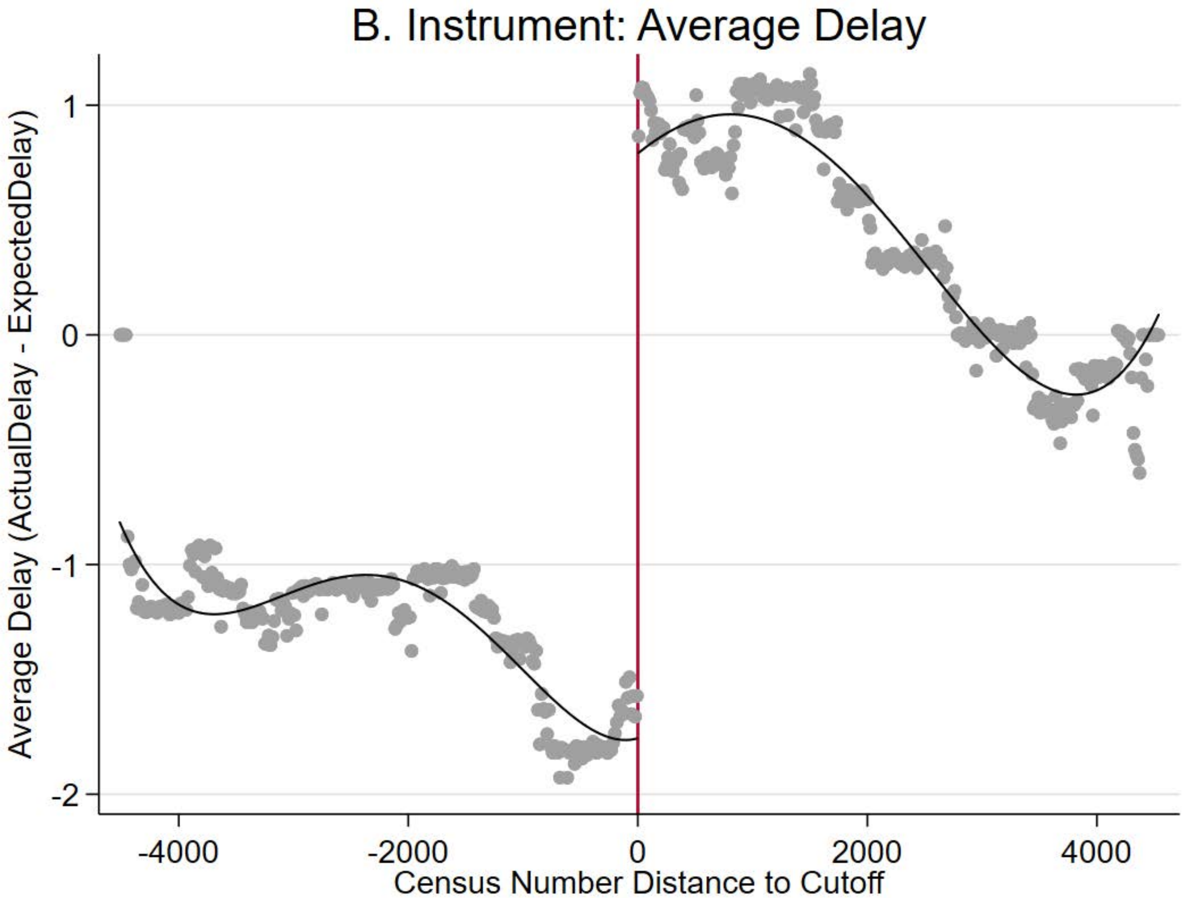
To have each bin be approximately 4.5% of the sample, the thresholds did not respect calendar boundaries by giving all migrants who took the census in the same week the same bin. The figure above shows that migrants who took the RAMV census in the second week (the light green area on the left) could potentially fall into one of four different batches, depending on randomness in the census-taking order. In Figure A2, we see that the the thresholds did not respect daily boundaries either.

Figure 4: **Distribution of Observables across the RD Threshold in Stacked Setup**



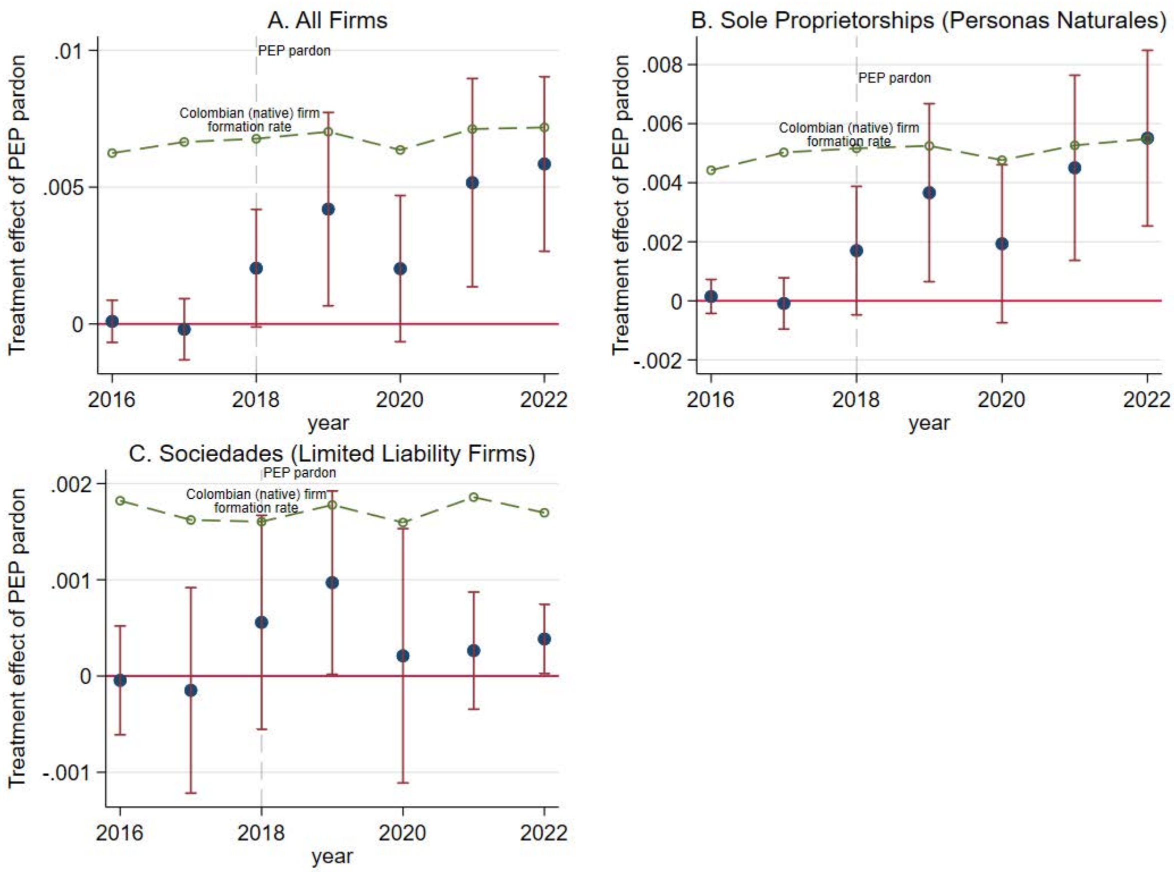
Notes: We plot the distribution of observables across the thresholds in a stacked setup. Panel A is a manipulation test on the density of census numbers away from the threshold as in Cattaneo et al. (2018). Panels B through E plot observables in a regression discontinuity setup following the approach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

Figure 5: **Distribution of Noise across the RD Threshold in Stacked Setup**



Notes: We plot a binned scatterplot using the regression discontinuity approach in Cattaneo and Titiunik (2021) of the noise variable realizations in a stacked RD setup. Consistent with the assumptions of our method, noise realizations are largest for those migrants right at the cutoff, who are moved to a different bin. For these migrants, the difference between being on the right vs the left of the cutoff implies a different of 2.5 days in time to register for the *Permiso Especial de Permanencia* (migrant pardon). This noise measure is the instrumental variable in our noisy RD setup.

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.

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: **Model of Noise**

	(1) Census Number	(2) Census Number	(3) Census Number
Age		-.0058***	-.00022***
		(.00019)	(.000028)
Registered in the Morning		-.1***	-.053***
		(.004)	(.00058)
Female		.082***	.0011*
		(.0047)	(.00069)
Pregnant		-.077***	-.012***
		(.016)	(.0023)
Number of People Registered		-.048***	-.0015***
		(.0018)	(.00027)
Family Size		.03***	-.000037
		(.001)	(.00015)
Week of Registration FEs	Y		Y
Main Model			Y
Observations	254,201	251,053	251,053
R^2	.98	.011	.98

Notes: In this table we model the RAMV census numbers using Eq. 2. To aide in interpreting coefficients, we have standardized the outcome variable. Note that registry numbers were assigned to entire families (not to individuals), and so the unit of analysis is the family. The covariates relating to age, gender and others are averages across all family members. About 64% of families had exactly one member. Standard errors clustered at the date of census completion are reported in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 3: **Instrument Balance**

	Negative Luck	Positive Luck	Difference
Age	30.75	30.62	0.13
Female	0.50	0.50	0.00
married	0.12	0.12	-0.00
Education is High School or Lower	0.81	0.80	0.01***
Informal Labor	0.29	0.29	0.00
Self Employed	0.27	0.26	0.01
Unemployed	0.21	0.23	-0.02**
No Occupation Reported	0.09	0.08	0.01*
Census # (Standardized)	0.39	0.05	0.34
Labor Certificate	0.14	0.16	-0.02***
Head of Household	0.65	0.65	0.00
Family Size	3.40	3.35	0.05
Has Family in Colombia	0.41	0.43	-0.03
Has Family in Venezuela	0.69	0.70	-0.01
Had No Food in Last 3 Months	0.36	0.36	-0.00
Expects to Stay in Colombia 1 Year or More	0.90	0.91	-0.00
Registered in the Morning	0.51	0.47	0.05

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.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 4: **Compliers vs Always-Takers**

	Compliers	Always Takers	Difference
Age	30.91	30.70	0.21***
Female	0.50	0.50	0.00**
married	0.20	0.13	0.07***
Education is High School or Lower	0.75	0.79	-0.04***
Informal Labor	0.26	0.31	-0.05***
Self Employed	0.21	0.26	-0.05***
Unemployed	0.38	0.21	0.18***
No Occupation Reported	0.07	0.08	-0.01***
Queue # (Standardized)	-1.82	0.38	-2.20***
Labor Certificate	0.25	0.16	0.09***
Head of Household	0.64	0.64	-0.00***
Family Size	3.03	3.45	-0.43***
Has Family in Colombia	0.71	0.39	0.32***
Has Family in Venezuela	0.73	0.68	0.04***
Had No Food in Last 3 Months	0.51	0.34	0.17***
Expects to Stay in Colombia 1 Year or More	0.92	0.91	0.01***
Registered in the Morning	0.15	0.52	-0.37***

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.
(i) Add number of people.
(ii) Report Difference / mean of compliers as fourth column.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 5: **First Stage Regressions**.
The role of delay noise in predicting receiving the PEP pardon.

	(1) *Dep Var:* PEP Pardon	(2) *Dep Var:* PEP Pardon	(3) *Dep Var:* PEP Pardon	(4) *Dep Var:* PEP Pardon	(5) *Dep Var:* PEP Pardon	(6) *Dep Var:* Business Created
Actual Delay	-.000089	-.0056***				
	(.00024)	(.00096)				
Predicted Delay Gap			-.0056***	-.0053***		-.000084***
			(.00097)	(.00094)		(.000016)
Running Expected Delay		.0056***			-9.6e-06	
		(.001)			(.00011)	
Prob. PEP					1***	
					(.025)	
F-Statistic		34	33	31	36	
Method		First Stage (No Controls)	First Stage (Centered)	First Stage (Controls)	First Stage Probit (Controls)	Reduced Form (Controls)
Observations	331,639	331,639	331,639	331,639	331,639	331,639
R^2	.000013	.0012	.0012	.02	.02	.0018

Notes: This table presents results for the first stage of our 2SLS estimation as part of our empirical strategy. Standard errors clustered at the date of census completion are reported in parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 6: **Main Results.** 2SLS estimates.
Dep. Var.: Business Created.

	(1)	(2)	(3)	(4)
PEP Pardon	.00131***	.0162***	.0159***	.0101***
	(.00015)	(.00408)	(.00323)	(.00281)
F Statistic	.	32.7	31.4	35.9
Method	OLS	IV (Centered)	IV (Controls)	IV (Probit)
Observations	331,639	331,639	331,639	331,639

Notes: This table estimates Equation 5 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. For a specification that only controls for 'Expected Delay', see Table A2
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 7: **Panel Results.**
Dep. Var: 1[Has started a company]

	(1)	(2)	(3)	(4)	(5)
After Entry into Colombia	0.000333***	0.000423***	0.0000506	0.000250***	0.000249***
	(0.0000386)	(0.0000412)	(0.0000318)	(0.0000480)	(0.0000480)
After Receiving PEP	0.00104***	0.000907***	0.000842***	0.000887***	0.000363***
	(0.0000776)	(0.0000743)	(0.0000974)	(0.0000877)	(0.0000571)
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

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 8: **Heterogeneity by Firm Employment and Assets.**

	(1) *Dep Var:* Creates Employer Firm	(2) *Dep Var.:* Creates Non Employer Firm	(3) *Dep Var:* Creates Low Asset Firm	(4) *Dep Var:* Creates High Assets Firm
PEP Pardon	.0064***	.0051***	.0067***	.0039***
	(.0022)	(.0019)	(.0023)	(.0014)
F Statistic	31	31	31	31
Method	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)
Outcome Mean	.00062	.00057	.00057	.00047
Relative Effect	10	9	12	8.3
Observations	331,620	331,620	331,620	331,620

Notes: This table estimates Equation 5 using as dependent variable an indicator on starting a business that responds to different characteristics of a firms. That is, we modify the dependent variable such that it is 1 (or zero otherwise) if the business started reports having employees (Column 1); or reports not having employees (Column 2); or was registered in 2018 or 2019 (Column 3) or registered in 2020 or 2021 (Column 4); or if it reports having assets below the median value (Column 5); above it (Column 6).Standard errors clustered at the date of census completion are reported in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 9: **Heterogeneity by Labor Force Status**

	(1) *Subsample:* Formal Labor	(2) *Subsample:* Informal Labor	(3) *Subsample:* Self Employed	(4) *Subsample:* Unemployed	(5) *Subsample:* Student	(6) *Subsample:* Homemaker
PEP Pardon	.15	.019**	.016**	-.0036	-.051	.012
	(.22)	(.0084)	(.0068)	(.0052)	(.14)	(.0084)
F Statistic	1	15	23	25	.18	21
Method	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)
Outcome Mean	.0041	.002	.0012	.0011	.00016	.00054
Relative Effect	37	9.6	13	-3.3	-324	23
Observations	2,930	96,589	89,397	73,272	12,706	29,406

Notes: This table estimates Equation 5 using sub-samples based on different self-reported status of labor participation by the individuals in our sample. Each column reports 2SLS estimates of the effect of having received the PEP visa on starting a formal business for the different sub-samples as described in the label for each column. Standard errors clustered at the date of census completion are reported in parenthesis.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table 10: **Heterogeneity by Demographics**

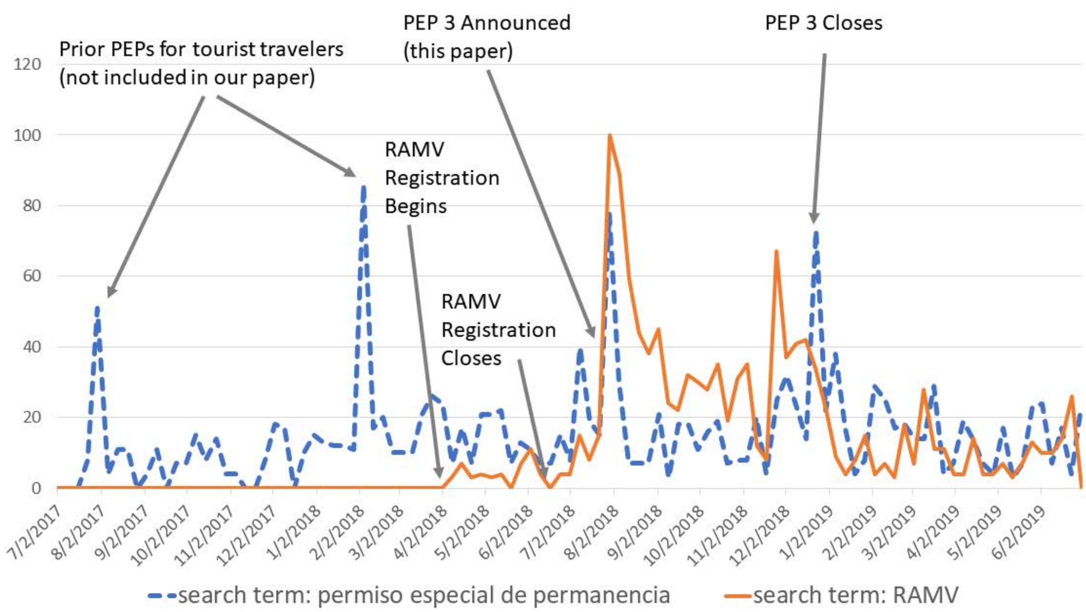
	(1) *Subsample:* Men	(2) *Subsample:* Women	(3) *Subsample:* Head of Household	(4) *Subsample:* Spouse	(5) *Subsample:* Age Below 30	(6) *Subsample:* Age Over 30	(7) *Subsample:* Secondary Education	(8) *Subsample:* Tertiary Education
PEP Pardon	.013***	.0084**	.0092***	.036***	.012**	.01	.0066**	.031**
	(.0039)	(.0038)	(.0031)	(.011)	(.0057)	(.0073)	(.0028)	(.013)
F Statistic	20	42	20	72	26	24	31	21
Method	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)
Outcome Mean	.0014	.0011	.0015	.0014	.0013	.0014	.00074	.0033
Relative Effect	9.2	7.7	6.3	26	9.2	7.4	9	9.3
Observations	166,262	165,107	214,755	43,787	153,976	145,158	265,917	65,722

Notes: This table estimates Equation 5 using sub-samples based on different demographic characteristics of the individuals in our sample. Each column reports 2SLS estimates of the effect of having received the PEP visa on starting a formal business for the different sub-samples as described in the label for each column. Standard errors clustered at the date of census completion are reported in parenthesis. Note head of household is self-reported in the RAMV census. Secondary education indicates people who have completed high school. Tertiary education includes all types of post secondary school, including technical school and partial college credits. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Appendix: For Online Publication Only

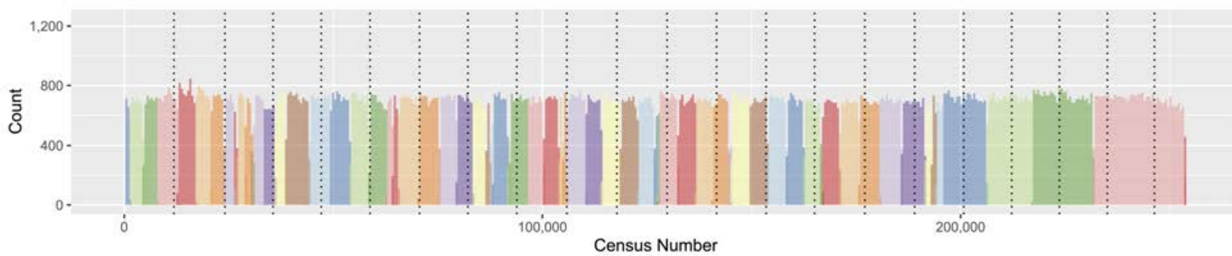
A Appendix Figures

Figure A1: **Google Trends of Search in Colombia for 'Permiso Especial de Permanencia' and 'RAMV' around PEP period**



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

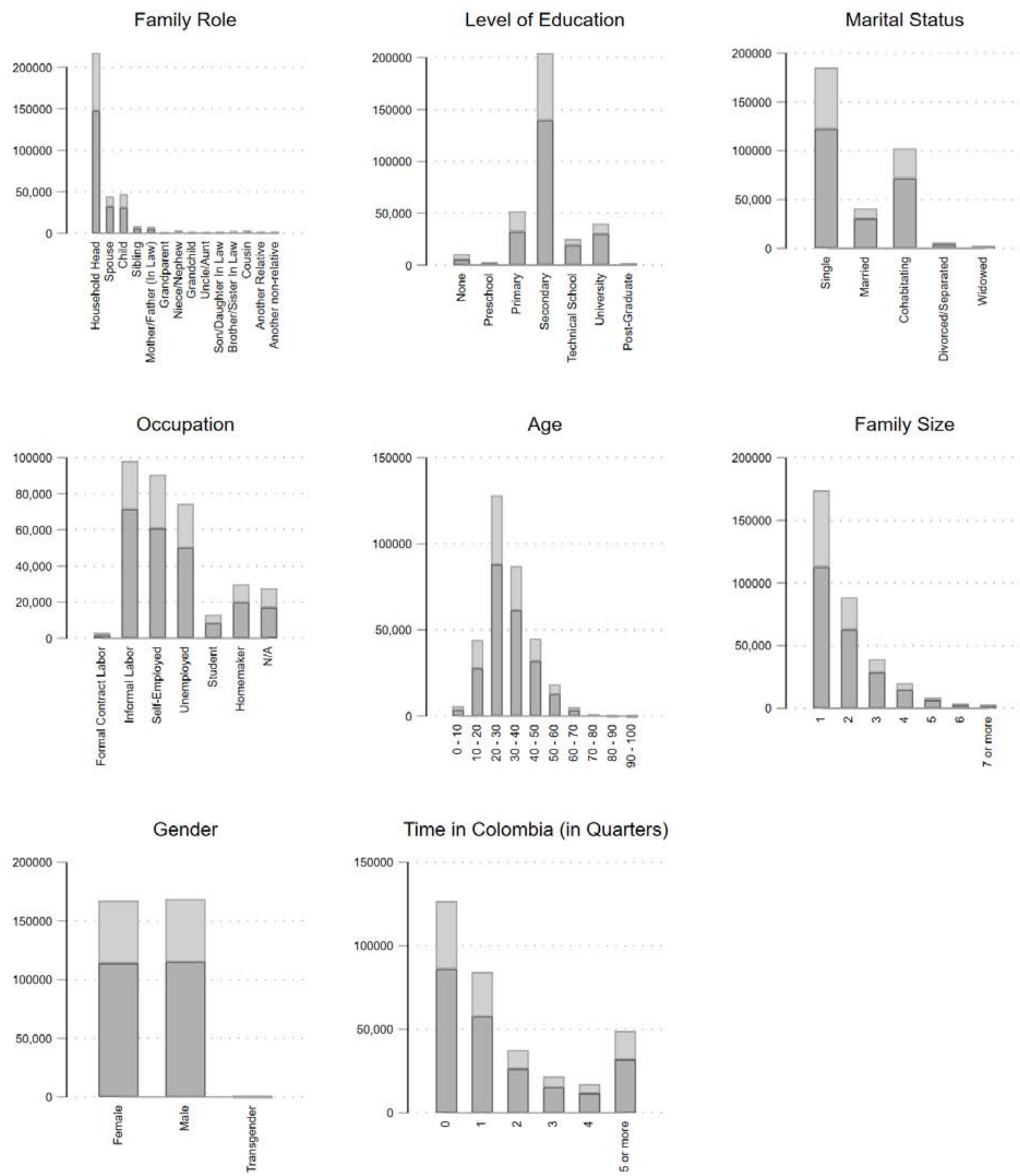
Figure A2: **Histogram of Census Numbers by Day (with Thresholds)**



Notes: Each color represents a day's worth of 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 numbers represent the thresholds in the Figure 2 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 3, we see that the the thresholds did not respect weekly boundaries either.

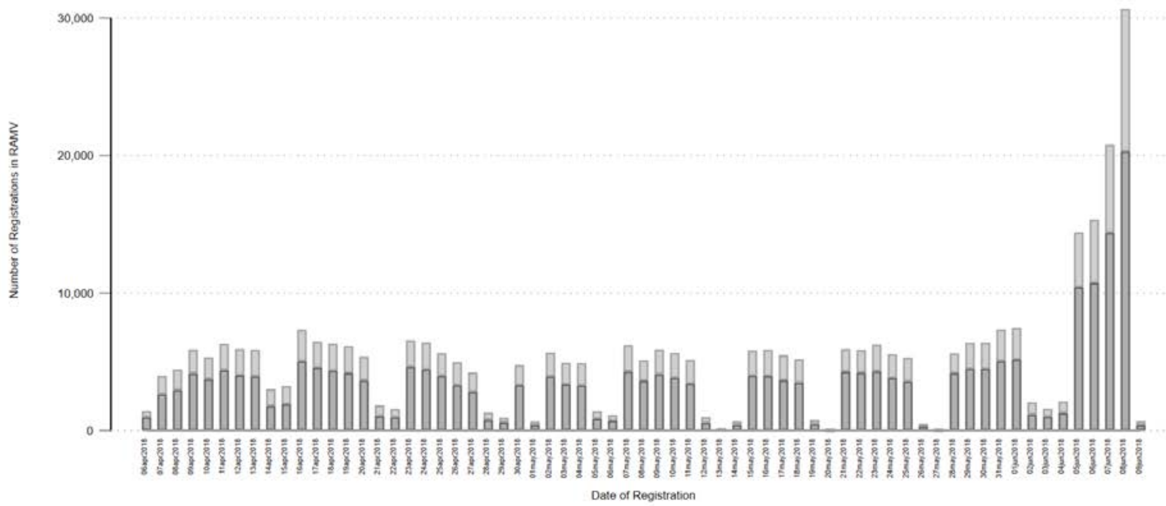
To achieve approximately 4.5% of the sample each bin, the thresholds did not respect calendar boundaries by (say) giving all migrants who took the census in the same day the same bin.

Figure A3: **Observables of RAMV registrants with and without PEP**



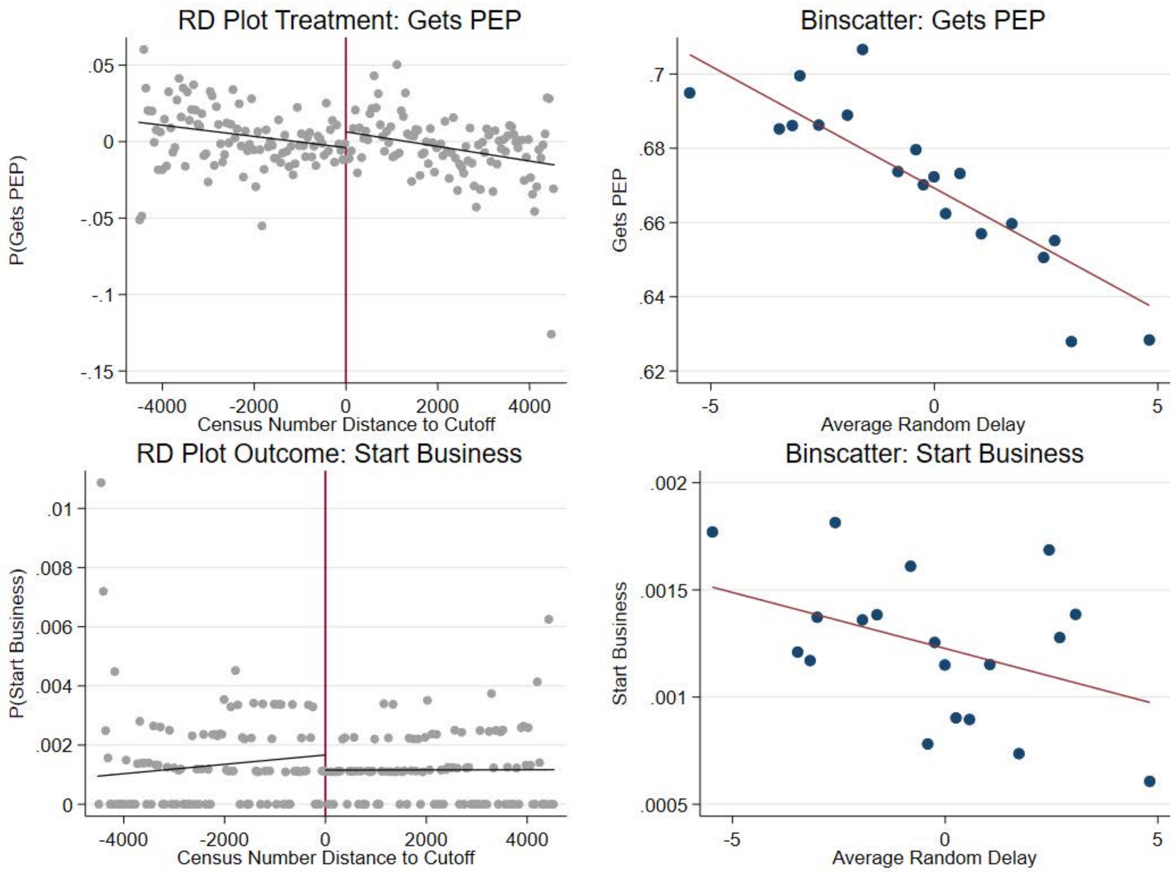
Notes: These figures plot number of people registered in RAMV across several observable characteristics, each one with defined categories. Within each characteristic, the bars show the amount of registrants within a category that received a PEP visa (in dark grey) and who did not receive a PEP visa (light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

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



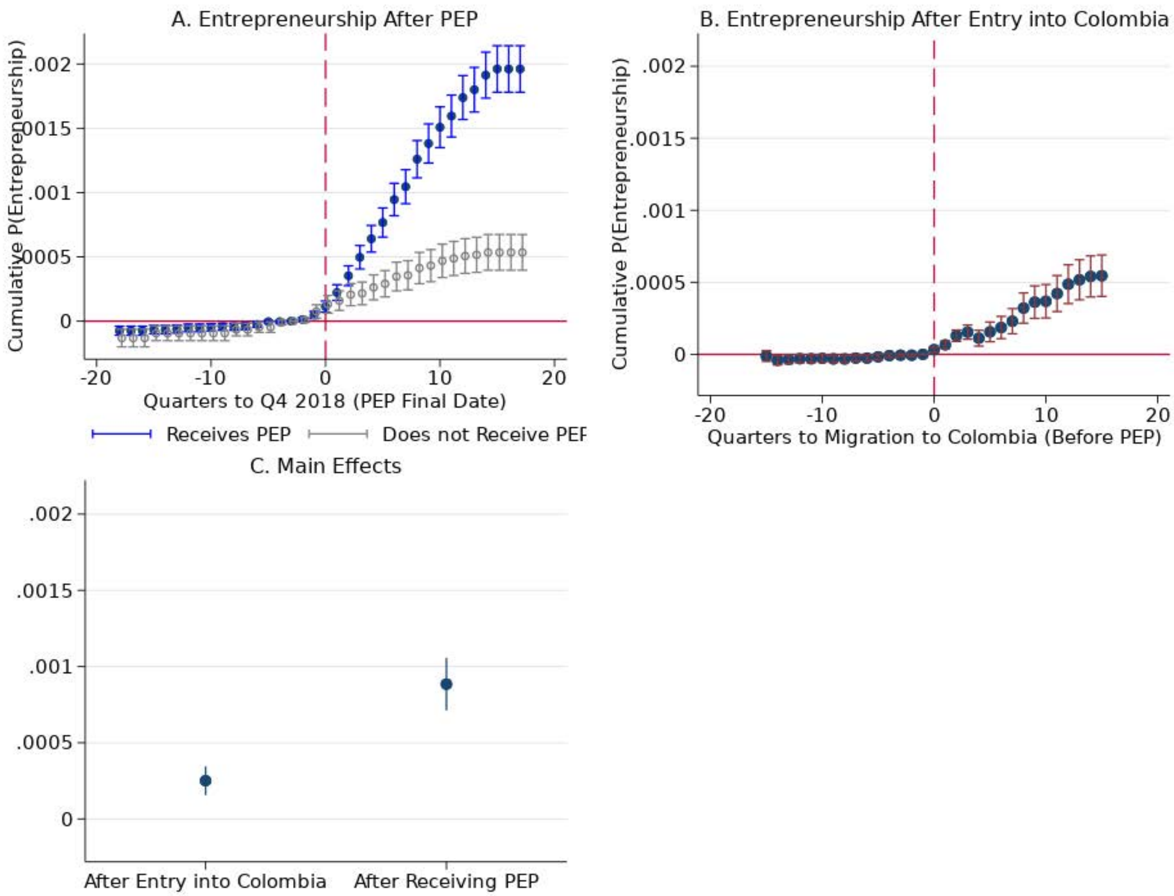
Notes: The figure plots the number of registrants per date of registration in the RAMV census. Each bar shows the amount of registrants that received a PEP visa (in dark grey) and who did not receive a PEP visa (light grey).

Figure A5: **Distribution of Treatment and Outcome across the RD Threshold in Stacked Setup and in Binned Scatterplot**



Notes:

Figure A6: **Panel Analysis**



Notes:

B Appendix Tables

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 receiving 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: **Main Results: Alternate Expected Delay Control**

	(1)	(2)	(3)	(4)
PEP Pardon	.0013***	.016***	.015***	.01***
	(.00015)	(.0033)	(.0031)	(.0028)
F Statistic	.	34	32	36
Method	OLS	IV (Centered)	IV (Controls)	IV (Probit)
Observations	331,639	331,639	331,639	331,639

Notes: This uses Equation 3 first stage and Equation 5. The one will the full distribution. For a specification that only controls for the full distribution of possible delays, see Table 6.
*** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

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: (Ommitted 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.

Table A4: **Heterogeneity by Registration Form Information**

	(1)	(2)	(3)	(4)	(5)
	Dep. Var.: Any Business	*Dep. Var.:* Sole Prop. (Personas Naturales)	*Dep. Var.:* Corp. or LLC (Sociedades)	*Dep. Var.:* Founded 2018-2019	*Dep. Var.:* Founded 2020-2022
PEP Pardon	.016***	.014***	.0018	.0034*	.0071***
	(.0032)	(.0031)	(.0016)	(.002)	(.0025)
F Statistic	31	31	31	31	31
Method	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)	2SLS (Controls)
Outcome Mean	.0016	.0014	.00023	.00049	.00074
Relative Effect	9.9	10	7.8	7	9.7
Observations	331,639	331,639	331,639	331,639	331,639

Notes: This table estimates Equation 5 using as dependent variable an indicator on starting a business that responds to different characteristics of a firms. That is, we modify the dependent variable such that it is 1 (or zero otherwise) if the business started reports having employees (Column 1); or reports not having employees (Column 2); or was registered in 2018 or 2019 (Column 3) or registered in 2020 or 2021 (Column 4); or if it reports having assets below the median value (Column 5); above it (Column 6).Standard errors clustered at the date of census completion are reported in parenthesis. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

Table A5: **Institutions.** Interactions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PEP Pardon	.016***	.02*	.025	.0038	.0094	.022**	.099
	(.0049)	(.012)	(.021)	(.0062)	(.0072)	(.0087)	(.082)
Distance to Markets							
PEP Pardon × Log(Distance to State Capital)	.00095						
	(.0025)						
PEP Pardon × Log(Distance to Market)		-.00067					
		(.0031)					
PEP Pardon × Log(Distance to Bogota)			-.0017				
			(.0038)				
Property Rights							
PEP Pardon × 1[Spanish Occupation 1510]				.019*			
				(.0099)			
PEP Pardon × 1[Indigenous Pop in 1535]					.011		
					(.011)		
PEP Pardon × Log(Total FARC Violence 2000 -2010)						-.0094	
						(.014)	
Job Opportunities							
PEP Pardon × Formal Employment Rate							-.00099
							(.00097)
F Statistic	8.7	15	17	6.6	6.9	14	18
Observations	331,347	331,347	331,347	331,319	331,319	331,347	331,347

Notes: Dependent variable is whether the immigrant starts a firm. 2SLS regressions. The main effect of the regional characteristic is also included as a control in all regressions, but not reported for simplicity. Standard errors clustered at the date of registration. Fixed-effects for week of registration, family role, gender, marital status, occupation status, and level of education included. *** $p < 0.01$ ** $p < 0.05$ * $p < 0.10$

C Firms Descriptives

This section describes the characteristics of the 442 firms in our sample, overall and by entrepreneurs who did and did not receive the PEP visa.

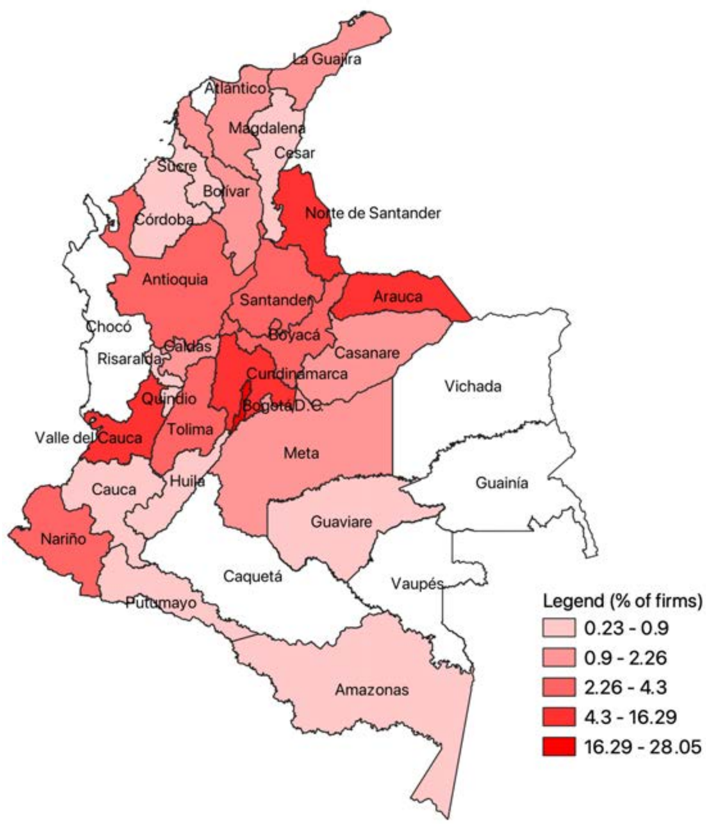
Figure A7 plots the geographic distribution of the firms in our sample across Colombian national territory. The departments with the highest share of firms are those in the border (Norte de Santander and Arauca) where there is a large number of Venezuelans, as well as the capital city, Bogota, and departments 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 A8 presents the same visualization but only for firms created by entrepreneurs who did get the PEP visa, which encompass 385 firms (out of 442). 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, A9 presents the geographic distribution of firms by entrepreneurs without the PEP visa, which corresponds to a much more limited sample of only 57 firms. 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).

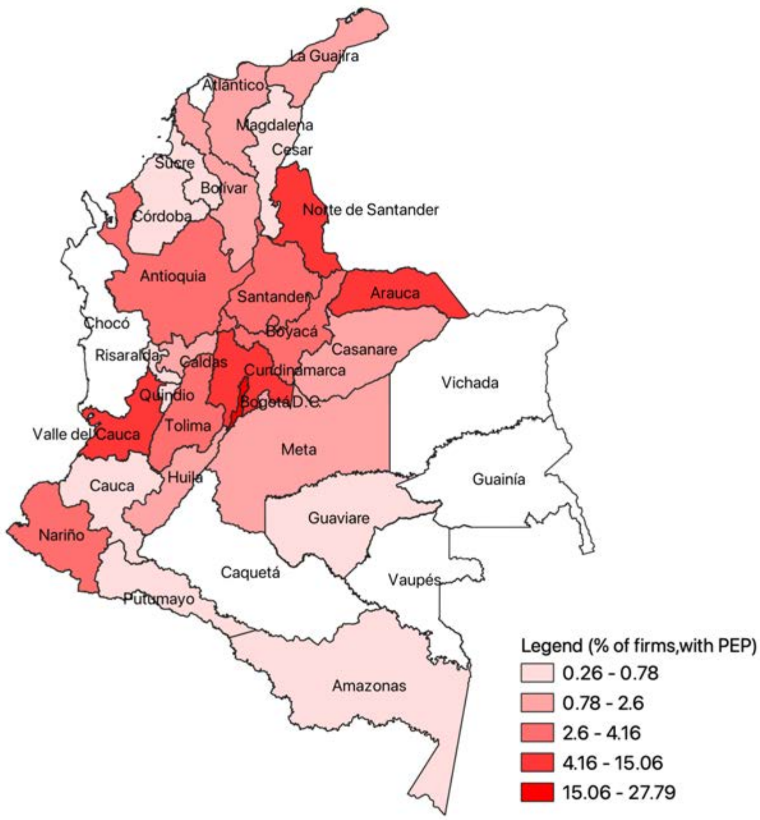
Thus, we find that the geographic distribution of firms across those individuals with and without PEP follows a similar pattern.

Figure A7: **Geographic Distribution New Firms**



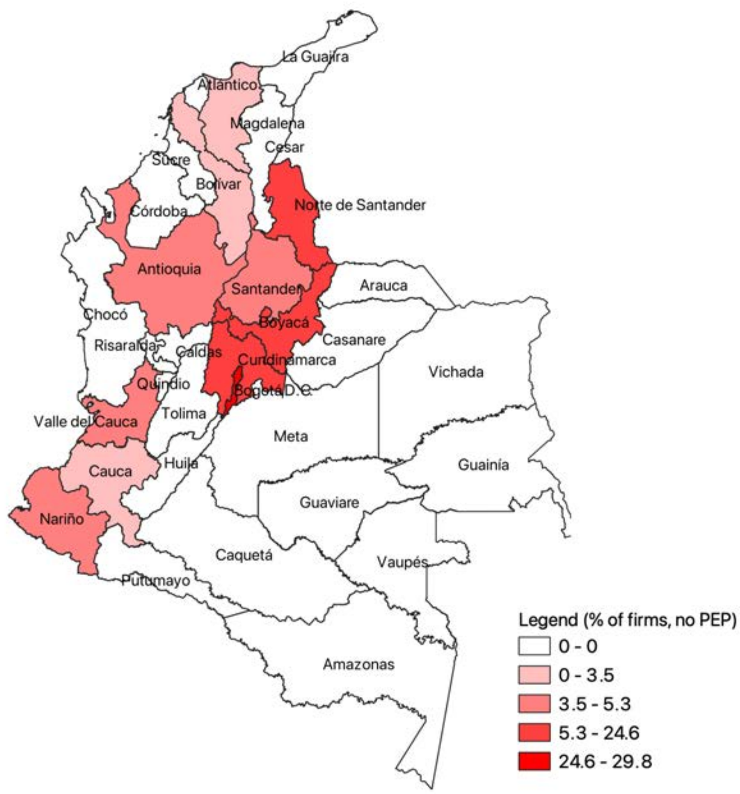
Notes: This figure visualizes the geographic distribution of new firms across the different departments of Colombia among entrepreneurs. Darker shades imply a larger share of firms being registered in that department, according to the legend.

Figure A8: **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 A9: **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.

We also present, in Table A6, 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 A7 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,

Table A6: 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.

are in personal services, the food business, and retail.

Table A7: 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 jewellery, costume jewelery 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.