Lessons from the Application of the Instrumental Variable (IV) Research Design to Sudden Immigration Shocks[1]

Maria Noelia Romero

Universidad de San Andres, CEDH

Abstract

The Instrumental Variable (IV) approach is commonly used to solve the endogeneity of migrants' decisions on where to locate. Using the Venezuelan migration shock to Peru between 2016 and 2018 as an example, I compare the first-stage estimates of different shift-share instruments commonly used in the IV approach. I show that in this example a sudden migration shock cannot be predicted by past migration simply because past migration is little or insignificant. In a globalized world, when there is no historical previous influx of migrants to the destination location and mass migration, I encourage contributors to the migration literature to report detailed statistical analyses of the first stage using the instrumental variable approach or pursue other identification strategies.

Keywords: Shift-share, immigration, Instrumental Variable.

JEL Codes: C26, J61

[1] I thank Nicolas Gatti for his helpful comments and encouragement and the helpful feedback from Rebecca Thornton, Mary Paula Arends-Kuenning, Alexander W. Bartik, Darren Lubotsky, Marieke Kleemans, and Andrew Garin. I also acknowledge the assistance of the National Institute of Statistics and Information (INEI) with the Peruvian Labor Force Survey (LFS) and the Venezuelan Survey (ENPOVE). In particular, the expertise shared by Yohnny Campana in helping to access the 2007 Census at the district level. I also appreciate the comments from my Ph.D. program peers Anahid Bauer, Felipe Diaz Klaassen, Wilfried Adohinzin, Alok Ranjan, Anna Kyriazis, Priyasmita Gosh, and Christian Valencia Rivamontan.

Resubmitted: October 9, 2024

Contact information: Vito Dumas 284 (B1644BID) Victoria, Pcia. de Bs. As., Argentina/Tel: (54-11) 7078-0400. Email: mromero@udesa.edu.ar , m.n.romero91@gmail.com

1. Introduction

One of the main econometric challenges for identifying the causal effect of immigration shocks on labor market outcomes is the endogeneity of where migrants are located. In the immigration literature, it is common to use the networking idea where the immigrants in the destination country influence the location of the new immigrants (Adão et al., 2019; Altonji & Card, 1991; Borusyak & Hull, 2020; Goldsmith-Pinkham et al., 2020; Jaeger et al., 2018). A common approach is to use the spatial distribution of past settlements of immigrants to predict the current inflows, also known as the *shift-share* instrument (Altonji & Card, 1991; Jaeger et al., 2018; Monras, 2020).[2] While this instrument has been widely used, there is little discussion on its performance in predicting future inflows of migrants when there was no previous relevant influx to the destination location. The lack of predictive power of the instrument in that scenario might be due to pushing factors at the country of origin and not necessarily related to attractive conditions in the country of destination. The cases of humanitarian crises such as Syria, Haiti, or Venezuela fit under the previous description.

The purpose of this paper is to discuss potential problems of the network instrument under (1) mass migration and (2) lack of history of migration between local and destination countries (Jaeger et al., 2018). I use the example of the Venezuelan migration to Peru that combines both attributes previously described. Venezuela's economic and political crisis in the 2010s has caused an unprecedented inflow of migrants to Peru (Restuccia, 2018; UNHCR & IOM, 2022). The arrival of Venezuelan-born immigrants to the metropolitan area of Lima and Callao in Peru -where 85 % of immigrants are located – represented a 7.5% increase in their 2017 working-age population, see Figure B1 (OGPP, 2018).

I show how the past share of Venezuelans inflow to Peru does not predict the more recent influx of Venezuelan immigrants between 2016 and 2018. I closely follow and compare the different measures of the Bartik shocks instruments: (1) the distribution of Venezuelans in the past in a more simple version following Monras (2020), (2) the traditional immigrant enclave with two types of definition. One measure uses information on the city of origin of immigrants following Card's (2009), and another uses the classic shift-share definition as in Altonji & Card (1991) and Card (2001).

To compute the different shift shares, I use two sources of information to measure the Venezuelan influx: a recent Venezuelan Survey and Census data. I calculate the immigration flows from the "*Survey of the Venezuelan population that lives in Peru*" (ENPOVE) conducted in December 2018. This survey contains information about the arrival date to Peru to construct the total number of Venezuelan migrants by quarter and municipality using survey weights.[3] The

[2] For a recent survey on forced migration and identification strategies in the literature, see Becker & Ferrara (2019). Also, Table A.1. in Jaeger et al. (2018) shows an extensive list of selected papers using the *Bartik* shock instrument (or *Enclave* instrument) in the immigration literature. In this paper, I use these terms as synonyms to minimize repetition.

[3] The first wave of the survey in 2018 covers the five most populated provinces of Peru, with a sample size of 3,713 households, and 9,868 individuals. In Figure B.1 shows the spatial distribution of Peruvian and Venezuelan populations in the whole country in 2019. Although the Venezuelan survey includes other provinces such as Tumbes,

Census data from 2007 and 2017 helps to show the lack of predictive power to predict the current mass immigration of Venezuelans when Peru was not a typical destination country for Venezuelans. Overall, the candidate instruments have a first-stage F-statist ranging from 0.6 to 70, with two weak instruments with an F-statistic lower than 10 - the rule of thumb in the IV approach (Andrews et al., 2019). The F-statistic of the first stage drops below 10 even when considering two critical aspects: (1) a past settlement of migrants two years before the large and sharp migration, and (2) the main destination area with the largest enclave of immigrants. Therefore, they do not predict the unprecedented diaspora of Venezuelan-born immigrants in Peru.

For Venezuelan migrants, the main *push factor* for this migration is the local economic and political crises in 2017, which drove many immigrants to other parts of Latin America and the rest of the world (Bull, 2020; Restuccia, 2018). Despite its geographic distance, Peru appeals to Venezuelan immigrants because it is the first Latin American country whose immigration policy offers near-automatic approval to Venezuelans applying for employment status (Acosta et al., 2019; Decreto Supremo N° 002-2017-IN, 2017). Further, even the lower-paid jobs that Peru's sizeable informal labor sector offers to immigrant workers pay better (and are more numerous) than those at home in Venezuela (Borjas, 1987, 1991; Buxton, 2005; INEI, 2018). Peru has received more than 630,000 Venezuelan-born immigrants from 2016 to 2018, representing a 2.4% increase in the 2017 country's population, see Figure B2.[4] Yet, unlike the U.S.'s history of massive migrations in the 19th century or more recent Mexican immigration, Peru is a country without a historic influx of immigrants (Abramitzky et al., 2014; Monras, 2020; Sequeira et al., 2020). Thus, one first approach to study the effect of Venezuelan immigration on the Peruvian labor market outcomes such as wages and employment is to use the shift-share instrument to address the location endogenity.

Other previous papers have done outstanding work studying the effects of Venezuelan immigration on Peruvian labor market outcomes using a research design different from the IV approach with a shift-share. First, Asencios & Caselares (2020) who also focus on the Lima and Callao metropolitan area as this paper, follow a probit model to estimate de effects of Venezuelan arrival on Peruvian employment by exploiting the longitudinal aspect of the Labor Force Survey (LFS). Second, Boruchowicz, et al (2024) propose a synthetic control for Lima and Callao to estimate the effect of the Venezuelan migration at the provincial level. Third, Vera & Jimenez (2022) follow the Card (2009) and Ottaviano & Peri (2012) approach, where they use the aggregated influx of Venezuela to define the immigration shocks across education-experience groups. In contrast, Denisse & Morales, (2020) follow the shift-share immigrant instrument of past settlement of Venezuelans at the province level to estimate the effect of migration on Peruvian employment using individual panel data from a Household Survey (ENAHO). Finally, Groeger et al., (2024) work is the closest paper to this discussion of the IV approach in this context. Besides studying the effect of informal labor markets on Venezuelan discrimination at the district level; in the second part of their paper, they estimate the effect of the Venezuelan

La Libertad, Arequipa, and Cusco; 60% of the sample belongs to Lima and Callao as shown in the largest orange circles in Figure A.3.

[4] From 2017 to June 2022, more than 1.2 million Venezuelans migrated to Peru (UNHCR & IOM, 2022). This represents a 4.1% increase from Peru's total 2017 population.

shock on Peruvian labor market outcomes at the province level using the shift-share instrument. Instead, I discuss the issue of the lack of predictive power of this instrument measuring past settlement of Venezuelans within the metropolitan area of Lima and Callao at the district level, which is a smaller geographic unit than the province. I conclude that when there are no previous historical flows of migrants to a specific destination, especially when there are factors in the country of origin that cause the inflow of migrants, the shit-share instrument lacks statistical power at a small geographic unit (such as municipality level) and requires exploring alternative approaches.

2. A Simple Measure of The Net Influx of Venezuelans in 2018 on the Past Distribution of Venezuelans

The traditional immigration literature is interested in understanding the effects of migration on the local labor markets. To estimate the effect of the immigration shock on labor market outcomes, one wants to estimate as a main specification: $y = \alpha + \beta \times Immshock + u$, where y is the variable of interest such as wages and employment. Yet, the net inflow of immigrants, $Immshock$, is endogenous to changes in the local labor demand. Following Monras (2020), I instrument the net inflow of Venezuelans in 2018 with the share of Venezuelans from previous years and different data sources. Formally, the first stage is as follows:

$$\Delta Share2018_d = \alpha + \beta \times Share(t-1)_d + X_d \times \gamma + \varepsilon_d \qquad (1)$$

Where $\Delta Share2018_d$ is the net inflow of Venezuelans in 2018, $Share(t-1)_d$ is the past share of Venezuelans in $t-1$ district d, X_d are control variables by district d, and ε_d is the error term. I denote the past share with $t-1$ because I can use the share in 2007, 2016, or 2017 from the different data sources.[5] Intuitively, the past share of Venezuelans should affect the current Peruvian labor market outcomes (such as wages and employment) only through the channel of the change in Venezuelan share.

To estimate the net inflow of immigrants in 2018 using the share of Venezuelans in the previous year as shown in Figure B2, I exploit two data sources: the Venezuelan Survey (ENPOVE) and the Peruvian Labor Force Survey (LFS), described in further detail in Appendix C. First, from the Venezuelan Survey, I calculate the Venezuelan-born population in 2016, 2017, and 2018. Second, from the Peruvian LFS, I computed the Peruvian working population in January 2017 before the significant immigration arrival from Venezuela. Hence, I combine both surveys to create the past settlement shares until 2017 to predict the inflow in 2018, see the geographic distribution in Figure B.2 panel b. Finally, I use the Peruvian Census data in 2007 and 2017 to have a second measure of past shares as shown in Figure B3. Therefore, I have four different past shares of Venezuelans from two data sources to use as an instrument to predict the net inflow in 2018.

[5] I measure the number of Venezuelans relative to the 2017 working-age population in the neighborhoods of Lima and Callao by year:

Sharedt=Venezuelan immigrantsdtWorking age Population d17

Where Venezuelan immigrantsdt is the number of Venezuelans of working age for district d year t and Working age Population d17 includes those between 14 and 65 years old in district d in January 2017.

Table 1 shows the estimates of the first stage in equation (1) from the four potential instruments. Columns (1) and (2) show the past share of Venezuelans using the ENPOVE survey, while Columns (3) and (4) display the estimates using the Census data to calculate the past settlement of Venezuelan immigrants. At first sight, one may wonder why the estimates for the share of Venezuelans in 2016 from ENPOVE and 2017 from the Census than the measure of the past share of Venezuelans in 2017 (ENPOVE) and 2007 using the Census data. The answer is there are large differences in the range of share of Venezuelans for the four candidates depending on the data source as illustrated in Figure B4. The share using ENPOVE in 2016 goes from 0 to

VARIABLES	(1) Δ Share of Venezuelans 2018	(2) Δ Share of Venezuelans 2018	(3) Δ Share of Venezuelans 2018	(4) Δ Share of Venezuelans 2018
Share of Venezuelans, 2016 (ENPOVE)	4.266**			
	(1.720)			
Share of Venezuelans, 2017 (ENPOVE)		1.600***		
		(0.191)		
Share of Venezuelans, 2017 (Census)			3.360***	
			(0.870)	
Share of Venezuelans, 2007 (Census)				0.072
				(0.091)
Observations	35	40	43	42
R-squared	0.193	0.782	0.178	0.004
F-statistic	6.154	70.03	14.91	0.615

Note: This table shows the net influx of Venezuelans in 2018 on the simple past share of immigrant definitions using two data sources. The dependent variable is the change in the share of Venezuelans in 2018. Column (1) and column (2) use past shares calculated from the Venezuelan Survey (ENPOVE) and the Peruvian LFS. Columns (3) and (4) use past shares calculated from the Census in 2007 and 2017. Clustered standard errors at the district level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

1.2%, which is smaller than the share of Venezuelans in 2018 using the same source from 1 to 31%. In contrast, the past share of Venezuelans in 2007 from the Census only range from 0 to 3%.[6] Keeping these large differences in the estimation of the past settlement of Venezuelans in 2007, 2016, 2017, and 2018 from each source, there are two lessons about using the simple past share of Venezuelans in previous years.

Table 1. First-Stage Regression of the net Venezuelan Immigration using the past settlement instrument

[6] Table 1 also shows the data availability limits from both sources with the differences in observations. In total, there are 43 districts in Lima and six in Callao. The Venezuelan Survey (ENPOVE) has no respondents for all the 49 districts that migrated before 2016 nor even until 2018, resulting in many districts with mission value in the dependent variable of the net influx of Venezuelans in 2018. In the case of the Census data, one district in 2007 was split into two districts by 2017 because of the ten-year differences and changes in the administrative limit at the district level.

The first lesson is that the past share instrument measured in 2017 – close to the sudden immigration shock – shows evidence of strong predictive power for the first-stage estimates. Regardless of the data source, either the Venezuelan Survey (ENPOVE) or the Census, the F-statistic of the first stage is larger than 10. Following the Staiger and Stock (1997) criterion, a first-stage F-statistic of 70 is suggestive evidence of a strong and relevant instrument. Yet, the share of Venezuelans measured with the Census data shows an 80% decrease in the F-statistic compared to the one using ENPOVE. The change in the predictive power of the past share of Venezuelans in 2017 when using different data sources might be explained by the population coverage of each source. The Census data covers the population with minimum measurement error, while the ENPOVE is a survey designed to be representative at the province level.[7] Thus, the source of data to calculate the past share of immigrants one year before the large migration is key to the prediction of the influx of Venezuelans in 2018.

The estimates of the first-stage coefficients in Table 1 show a second lesson regarding the Batik shock instrument measures further in time from the immigration shock. This network instrument two years before the large mass migration to Peru is not enough or good to estimate the effect of immigration on labor market outcomes. In column (1), the F-statistic of the share of Venezuelans in 2016 is less than 10, and it is even smaller and close to zero when using the Census data in 2007 as shown in column (4). As mentioned above regarding the range of values of past immigrant settlements, the average share of Venezuelans in 2007 of 0.0003 using the Census is too small to predict the most recent influx of 0.0405 in 2018.[8] Thus, while the F-statistics of the first stage might seem promising for some instruments close to the mass migration phenomenon, the fundamental problem with this immigrant Enclave instrument is that it uses shares extremely close zero to predict a large range of non-zero positive numbers.

3. Two Shift-Share Instruments: Using the City Of Origin and the Enclave Shares From the Census

To exploit the information about the state of origin in Venezuela from the Venezuelan Survey (ENPOVE), I follow Card's (2009) notation and definition of the shift-share instrument using the city of origin of immigrants:

$$\tilde{m}_d = \sum_o \frac{M_{od}}{M_o} \frac{\Delta M_{o2018}}{L_d} \qquad (2)$$

Where $\frac{M_{od}}{M_o}$ is the share of immigrants from origin o in Venezuelan in Peruvian district d, ΔM_{o2018} is the number of new arrivals to Peru in 2018 at the national level, and L_d is the normalization of the local population at the district level in 2017.

The second Bartik shock measurement exploits the information from the Census in 2017 and 2007 differently. I follow the definition by Altonji & Card (1991) and Card (2001):

[7] Additionally, the outcome variable of change of share of Venezuelan in 2018 is also measure in ENPOVE, which it might be possible that the reason that column (2) shows an inflated F-statistic.

[8] To illustrate this measurement issue, I use two districts in Lima as a leading example of the first-stage prediction problem in Table A1. Lima and Surquillo had a ratio of Venezuelan to working-age population of 0.0004 in 2007 from the Census data. There is not enough variation between districts to predict a change in the share of immigrant of 0.09 and 0.13, respectively, in 2018.

$$Shiftshare_d = \frac{\left(\frac{Venezuelan_{dt}}{Venzuelan_t} \times \Delta Ven_{18}\right)}{L_{d17}} \qquad (3)$$

Where $Venezuelan_{dt}$ is the number of Venezuelans in district d year t, $Venezuelan_t$ is the total number of Venezuelan immigrants in year t, ΔVen_{18} is the net aggregate inflow of Venezuelan-born immigrants in 2018 to Peru, and L_{d17} is the working-age population in district d in 2017 to normalize as the other shift-share instrument.

I construct the first enclave instrument using mainly information from the Venezuelan survey, which has data on the Venezuelan immigrants' origin. In Table A2, I show the percentage of the sample from the Venezuelan survey (ENPOVE) by the state of origin in their country. Notice that 50% of the total sample from the Venezuelan Survey is from only five of the wealthiest states: Caracas (capital), Carabobo, Lara, Zuli, and Aragua. With this information, I construct the enclave shares $\frac{M_{od}}{M_o}$ in equation (2) for Venezuelans that arrived in Peru in 2016 and 2017. I define the aggregate inflow of Venezuelans by city of origin ΔM_{o2018} as the difference in arrivals between 2017 and 2018. For the second definition of Bartik shock instrument, I use the Census data to calculate the Venezuelan shares $\frac{Venezuelan_{dt}}{Venzuelan_t}$. I compute the aggregate inflow of Venezuelans to Peru in 2018, ΔVen_{18}, as the difference between the stock of immigrant-born in 2017 and 2018 from the Venezuelan Survey. Finally, I use the Peruvian Labor Force Survey to calculate the working wage population by district L_d to normalize both shift-share definitions with the same baseline as in Section 2.

Table 2. First-Stage Regression of the net Venezuelan Immigration using the past settlement instrument two shift-share definitions

VARIABLES	(1) Δ Share of Venezuelans 2018	(2) Δ Share of Venezuelans 2018	(3) Δ Share of Venezuelans 2018	(4) Δ Share of Venezuelans 2018
Shift-share 2016 (ENPOVE)	0.257**			
	(0.111)			
Shift-share 2017 (ENPOVE)		0.668***		
		(0.085)		
Shift-share 2017 (Census)			0.380***	
			(0.099)	
Shift-share 2007 (Census)				0.043
				(0.055)
Observations	35	40	43	42
R-squared	0.214	0.780	0.178	0.004
F	5.399	61.79	14.91	0.616

Note: This table shows the first stage estimates of the net influx of Venezuelan immigrants in 2018 following the two definitions of shift shares with both data sources. The dependent variable is the change in the share of Venezuelans in 2018.

Column (1) and column (2) use a shift-share definition based on the city of origin from the Venezuelan Survey (ENPOVE) and the Peruvian LFS. Columns (3) and (4) follow the standard Enclave definition from the Census in 2007 and 2017, respectively. Clustered standard errors at the district level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2 presents the first-stage estimates of the net inflow of Venezuelan immigrants in 2018 on the two shift-share definitions measured in different periods. First, notice the coefficients do not vary as much as using the simple past share of immigrants as in Section 2. The estimates of these four correlations are also illustrated in Figure B5. Second, notice again that the information from the Census 2007 reveals that few Venezuelan-born immigrants cannot predict their location decision in 2018. The enclave instruments with a high correlation between the 2018 immigrant inflow are the ones that exploit the spatial variation of the stock of immigrants measured in 2017 as in the previous case. Thus, regardless of the Bartik shock definition - one that could rely on the shares of the city of origin of Venezuelans or the share of Venezuelans by the district of destination for identification in an IV approach – are both weak instruments when they are measured far in time (even two years prior) to the mass immigration influx in 2018.

4. Conclusions

In this article, I raise the issue of the lack of relevance of network instruments when using an IV research design in the context of mass immigration. The relevance condition of the shift-share instrument can suffer predictive power of the endogenous immigration shock when there is almost no historical influx of immigrants to the destination country. To illustrate this empirical concern, I instrumented contemporaneous migrant inflows using different measures of the past settlement of migrants. The identification assumption of the IV research design is that the share of Venezuelans in the past – either in the first simple approach or the shift-share instruments- is not correlated with changes in the local labor market that affect the native labor market outcome such as wages and employment and predicts the current immigrants location decision (Altonji & Card, 1991; Monras, 2020). I show that the immigrant enclave approach is not a suitable instrumental variable when there is nearly no past settlement of migrants at the destination location at the municipality level. Only when the network instrument is measured one year prior to the sudden immigration, we find a strong instrument to predict the influx of Venezuelans. The different instrument candidates have a weak first stage when calculated two years or more before the mass immigration. More importantly, when using Census data, I confirm there were few Venezuelans 10 years ago to predict a large range of arrivals to various neighborhoods in the metropolitan area of Lima and Callao. This fact also shows that the district spatial variation from the past is not enough, even in the provinces with the highest immigration influx relative to other locations in Peru. When the migration influx has no previous history between locations, alternative approaches to deal with the endogeneity of location should be explored (Asencios & Caselares, 2020; Vera & Jimenez 2022; Boruchowicz, et al, 2024). Further, my work raises a word of caution and encourages future contributors to carefully examine and report the statistical analysis of instrumental variables in this setting.

References

Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A nation of immigrants: Assimilation and

economic outcomes in the age of mass migration. *Journal of Political Economy*, *122*(3), 467–506. https://doi.org/10.1086/675805
Acosta, D., Blouin, C., & Freier, L. F. (2019). La emigración venezolana: respuestas latinoamericanas. *Fundación Carolina*, *March*. https://doi.org/10.13140/RG.2.2.33921.71523
Adão, R., Kolesár, M., & Morales, E. (2019). Shift-Share Designs: Theory and Inference. *Quarterly Journal of Economics*, 134(4), 1949–2010.
Altonji, J. G., & Card, D. (1991). Chp 7 - The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives. In *Immigration, Trade, and the Labor Market* Volume (pp. 201–234).
Andrews, I., Stock, J. H., & Sun, L. (2019). Weak Instruments in Instrumental Variables Regression: Theory and Practice. *Annual Review of Economics*, 11, 727–753.
Asencios, R., & Castellares, R. (2020). Impacto de la Inmigración Venezolana en el Empleo y los Salarios: el Caso Peruano. In *Banco Central de Reserva del Perú*. https://www.migraciones.gob.pe/index.php/estadisticas/
Becker, S. O., & Ferrara, A. (2019). Consequences of forced migration: A survey of recent findings. *Labour Economics*, 59(February), 1–16.
Borjas, G. J. (1987). Self-Selection and the Earnings of Immigrants. *The American Economic Review*, *77*(4), 531–553.
Borjas, G. J. (1991). Immigration and Self-Selection. In *Immigration, Trade and the Labor Market Volume* (pp. 29–76). http://www.nber.org/books/abow91-1
Boruchowicz, C., Martinelli, C., & Parker, S. W. (2024). Economic Consequences of Mass Migration: The Venezuelan Exodus in Peru. *Economía LACEA Journal*, *23*(1), 107–136. https://doi.org/10.31389/eco.436
Borusyak, K., & Hull, P. (2020). Non-Random Exposure to Exogenous Shocks: Theory and Applications. In *NBER Working Paper* No. 27845.
Bull, B. (2020). The crisis in Venezuela. *European Review of Latin American and Caribbean Studies*, *109*(109), 1–20.
Buxton, J. (2005). Venezuela's contemporary political crisis in historical context. *Bulletin of Latin American Research*, *24*(3), 328–347. https://doi.org/10.1111/j.0261-3050.2005.00138.x
Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics.*
Card, D. (2009). Immigration and Inequality. In *NBER Working Paper No*. 14683.
Decreto supremo N° 002-2017-IN, 8 (2017). https://diariooficial.elperuano.pe/Normas
Denisse, M., & Morales, F. (2020). *Venezuelan migration in Peru: Adjustments in the Labor Market* (IDB-WP-1146 Provided; IDB Working Paper Series).
Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, when, why, and how. *American Economic Review*, 110(8), 2586–2624.
Groeger, A., León-Ciliotta, G., & Stillman, S. (2024). Immigration, labor markets and discrimination: Evidence from the Venezuelan Exodus in Perú. *World Development*, *174*, 106437.
INEI. (2018). Producción y Empleo Informal en el Perú. In *Instituto Nacional de Estadística e Informática*. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1589 /

Jaeger, D., Ruist, J., & Stuhler, J. (2018). Shift-Share Instruments and the Impact of Immigration. In *NBER Working Paper* No. 24285.

Monras, J. (2020). Immigration and wage dynamics: Evidence from the Mexican peso crisis. *Journal of Political Economy*, 128(8), 3017–3089.

OGPP. (2018). *Informe Migratorio. Caracteristicas sociodemograficas de los ciudadanos venezolanos que tramitaron el Permiso Temporal de Permanencia-PTP en el Peru 2017-2018.*

Ottaviano, G. I. P., & Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European Economic Association*, *10*(1), 152–197. https://doi.org/10.1111/j.1542-4774.2011.01052.x

Restuccia, D. (2018). The Monetary and Fiscal History of Venezuela: 1960-2016. In *University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2018-59*. https://doi.org/10.2139/ssrn.3238188

Sequeira, S., Nunn, N., & Qian, N. (2020). Immigrants and the Making of America. *The Review of Economic Studies*, *87*(1), 382–419. https://doi.org/10.1093/restud/rdz003

Staiger, D. and J.H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557 – 586.

UNHCR, & IOM. (2022). *Regional Inter-Agency Coordination Platform*. https://www.r4v.info/en/aboutus

Vera, C. P., & Jimenez, B. (2022). *Do immigrants take or create natives' jobs? Evidence of Venezuelan immigration in Peru* (No. 009; Lima School of Economics Working Papers).

A. Appendix Tables

Table A1. Two districts in Lima as leading examples of the first-stage estimation problem

District	Share of Venezuelan 2017	Share of Venezuelan 2007	Share of Venezuelans 2016	Share of Venezuelans 2017	Share of Venezuelans 2018	Δ share of Venezuelan 2018
LIMA	0.0086	0.0004	0.005	0.039	0.130	0.090
SURQUILLO	0.0185	0.0004	0.003	0.047	0.180	0.131
Data source	Census	Census	ENPOVE	ENPOVE	ENPOVE	ENPOVE

Note: This table shows the share of Venezuelan estimates by year and data source for the districts of Lima and Surquillo as leading examples of the range of numbers of share of immigrants.

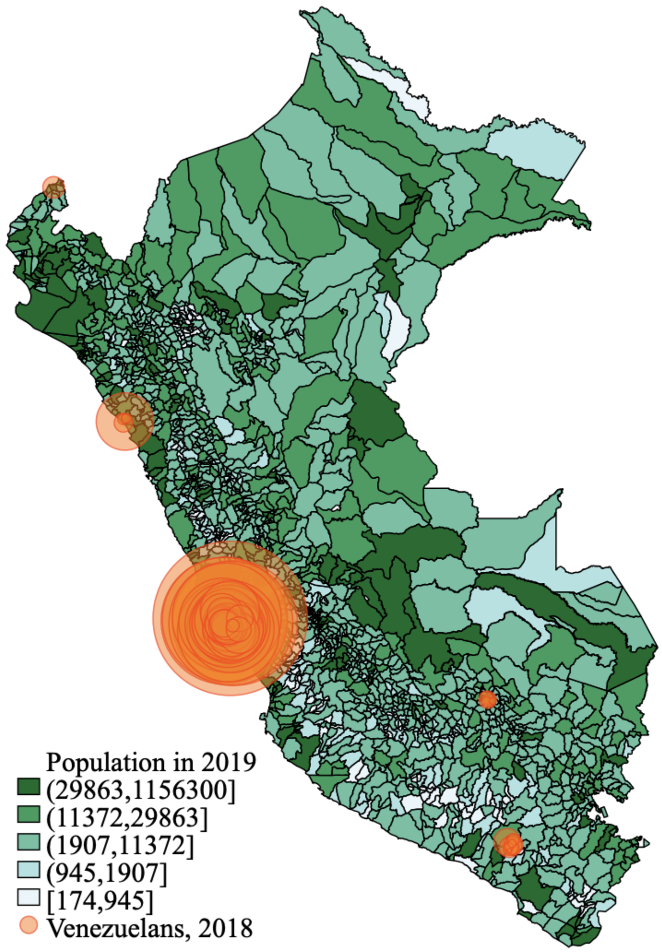
Table A2. Percentage of the Venezuelan sample by state of origin

Venezuelan State name (in Spanish)	**Percent of the Venezuelan Survey sample**
Distrito Capital	13.86
Estado Carabobo	11.27
Estado Lara	10.28
Estado Zulia	7.35
Estado Aragua	7.33
Estado Anzoategui	6.37
Estado Miranda	5.52
Estado Tachira	4.92
Estado Monagas	3.97
Estado Barinas	3.91
Estado Merida	3.72
Estado Portuguesa	3.55
Estado Bolivar	2.98
Estado Trujillo	2.77
Estado Falcon	2.12
Estado Yaracuy	2.09
Estado Nueva Esparta	1.71
Estado Vargas	1.62
Estado Sucre	1.45
Estado Cojedes	1.18
Estado Guarico	1.15
Estado Apure	0.71
Estado Delta Amacuro	0.13
Estado Amazonas	0.04

Note: This table shows the percentage of Venezuelan respondents between 15 and 65 years old by state of origin using the Venezuelan Survey (ENPOVE)

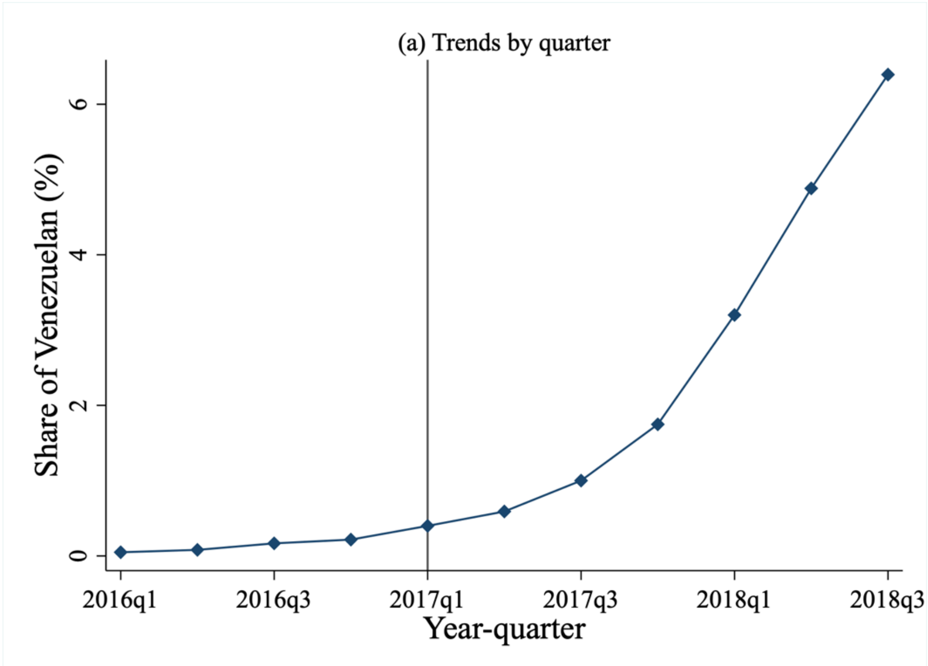
B. Appendix Figures

Figure B1. Peruvian and Venezuelan geographic distribution, 2018 and 2019



Note: This map shows the distribution of the Peruvian and Venezuelan populations in 2019. Sources: latest INEI update on the Peruvian population estimates in 2019, and the Venezuelan Survey (ENPOVE) in 2018 for the Venezuelan population. The circles' sizes are proportional to the number of Venezuelans.

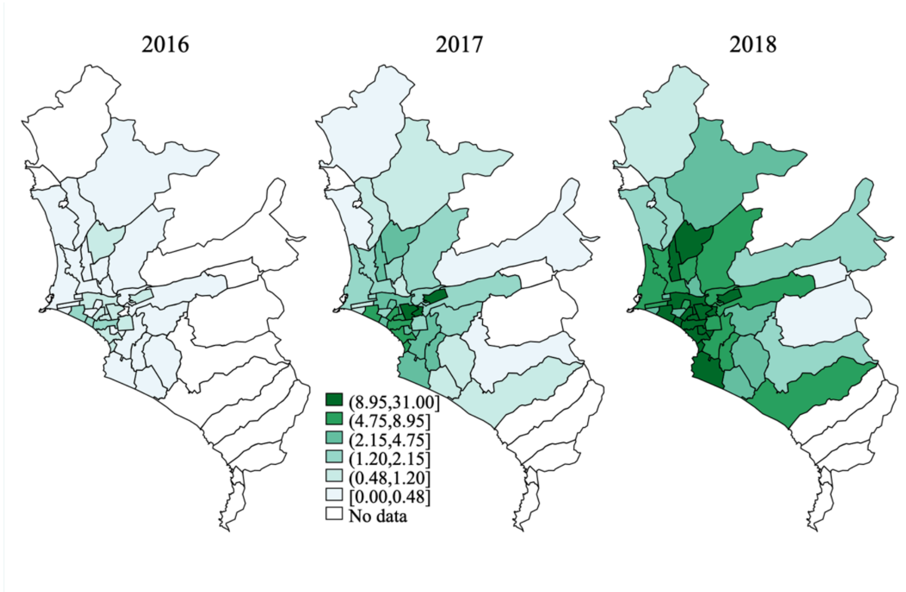
Figure B2. Increase of the share of Venezuelan immigrants (in %) in Lima and Callao, 2016-2018



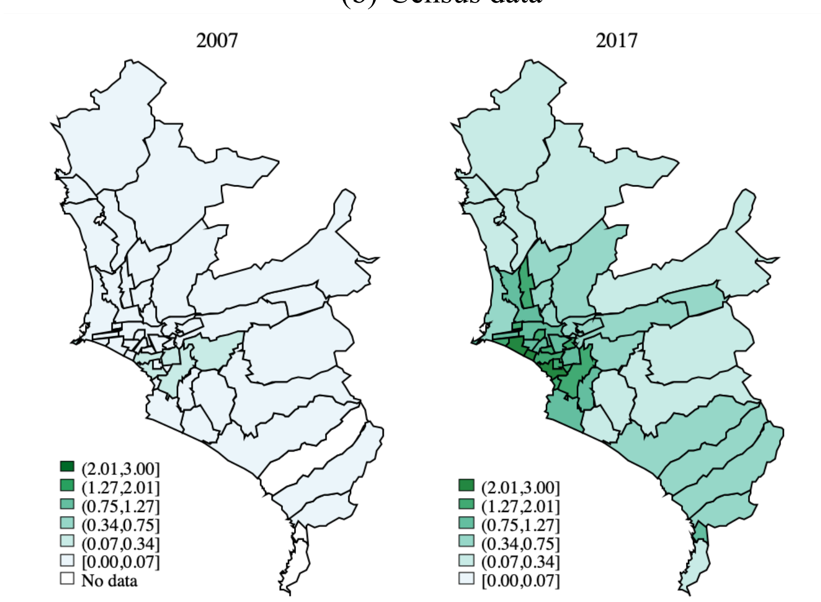
Note: This figure shows the time variation of the influx of Venezuelans to Lima and Callao between 2016 and 2018. It shows the cumulative share of Venezuelans, defined as the ratio of the stock of immigrant-born to the population between 14 and 65 years old, in January 2017 (in %). The vertical gray line indicates the introduction of the working permit in January 2017, which is a legal immigrant status for Venezuelans.

Figure B3. Share of Venezuelans using the Venezuelan Survey (ENPOVE) between 2016 to 2018 and the 2007 and 2017 Census
(in %)

(a) Venezuelan Suvery (ENPOVE)

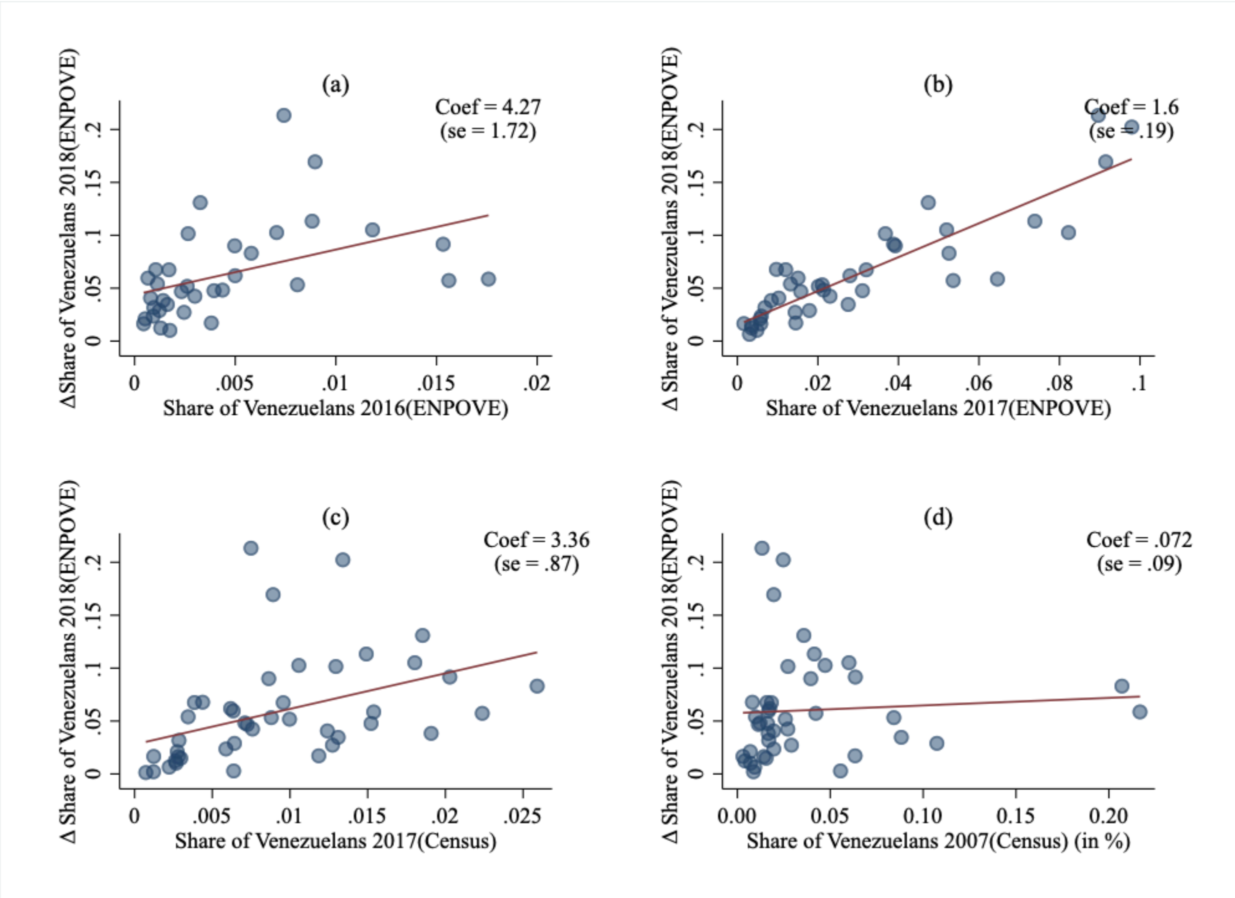


(b) Census data



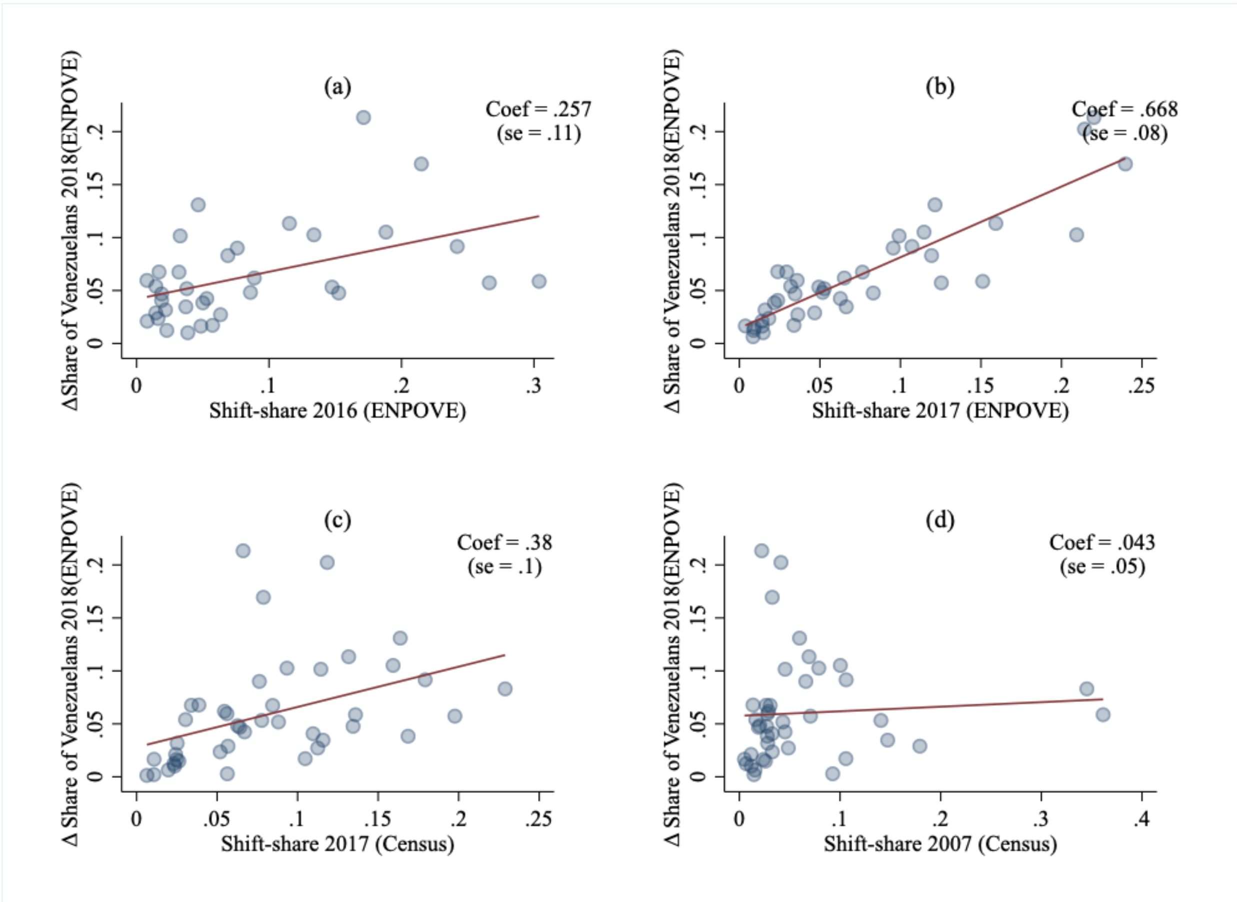
Note: These figures show the past settlement instrument using two data sources. Panel (a) shows the share of Venezuelans in the Lima and Callao districts by year (in %) using ENPOVE data. Panel (b) displays the share of Venezuelans in 2007 and 2017 calculated with the Census information. In both cases, the share is defined as the number of Venezuelans divided by the Peruvians and Venezuelan population between 15 and 65 years old, INEI.

Figure B4. Regression estimates between the net inflow of Venezuelans in 2018 and past-share using two sources



Note: This figure shows the first-stage estimates of the net influx of Venezuelans in 2018 on the simple past share of immigrant definitions using two data sources. The dependent variable is the change in the share of Venezuelans in 2018. Panel (a) and Panel (b) use past shares calculated from the Venezuelan Survey (ENPOVE) and the Peruvian LFS. Panel (c) and (d) use past shares calculated from the Census in 2007 and 2017. Clustered standard errors at the district level are in parentheses.

Figure B5. Partial correlations between the net inflow of Venezuelans in 2018 and two shift-share definitions from different data sources



Note: This figure shows the first-stage estimates of the net influx of Venezuelan immigrants in 2018 following the two definitions of shift shares with both data sources. Panel (a) and (b) use a shift-share definition based on the city of origin from the Venezuelan Survey (ENPOVE) and the Peruvian LFS in 2016 and 2017, respectively. Panel (c) and (d) use the shift-share definition in equation (4) as the standard Enclave definition from the Census in 2017 and 2007, respectively. The red line represents the linear projection from the OLS estimation.

C. Data

From the National Institute of Statistics and Information (INEI), the primary datasets used in the analysis and publicly available are:

- The Peruvian Labor Force Survey (*Encuesta Permanente de Empleo*, EPE, in Spanish). This Labor Force Survey is conducted in the metropolitan areas of Lima and Callao. This survey is quarterly, and the geographic unit of this survey is a conglomerate that contains between three to five blocks. Each quarter of the survey consists of a total of 400 conglomerates. The survey covers the 43 districts in Lima and six districts in Callao. The sample size is 4,800 households, with around 15,000 individuals each quarter. I use the last quarter of December 2016 and January February 2017 to measure the Peruvian labor force before the large influx of Venezuelan immigrants in 2017 and 2018.
- The 2018 Wave of Venezuelan Survey (*Encuesta Dirigida a la Poblacion Venezolana que reside en Peru, ENPOVE*): The first wave of the survey in 2018 covers the five most populated provinces of Peru: Lima, Callao, Tumbes, La Libertad, Arequipa, and Cusco. The sample size is 3,713 households and 9,868 individuals. I use the information on Peru's arrival date to estimate the Venezuelans' share by district and the net influx of Venezuelans by year.

These two microdata sets are publicly available on the following website: https://proyectos.inei.gob.pe/microdatos/. The crosswalk between the conglomerate from EPE to districts in ENPOVE is available under direct request to INEI.

To calculate the shit-share using a second data source at the district level, I use the information on the Peruvian and Venezuelan-born population in Lima and Callao from:

- Census 2017, available in the REDATAM system at the district level: https://censos2017.inei.gob.pe/redatam/
- Census 2007 is not publicly available at the district level. I got access under direct request to INEI.