Endogeneity in the Decision to Migrate: Changes in the Self-Selection of Puerto Rican Migrants before, during, and after the Great Recession

Aasha Reddy

Dr. William A. Darity Jr., Faculty Advisor

Dr. Michelle Connolly, Honors Seminar Professor

Acknowledgements

I would like to thank both my thesis advisor, Dr. William A. Darity Jr., and my honors seminar instructor, Dr. Michelle Connolly. Both have given me invaluable advice and feedback during the research process, and continuous support and mentorship over the past year. Without their input, writing this thesis would not have been possible. I would also like to thank my family, friends, and peers for their enduring support and encouraging me to see this project to the end.

Abstract

Migrants self-select on characteristics such as income. We use the U.S. Census' ACS and PRCS to study changes in selection patterns of Puerto Rican migrants to the to the U.S. mainland (50 states) before, during, and after the Great Recession (2005 to 2016). We construct counterfactual income densities to compare incomes of Puerto Rican migrants to the mainland versus incomes of island residents under equivalent returns to skill. We examine where Puerto Rican migrants to the mainland tend to fall in the island's income distribution and find that Puerto Rican migrants tend to come from the top 20% of the island's income distribution. This pattern remained stable with little to no effect of the Great Recession on selectivity patterns.

Introduction

Migrants comprise a substantial demographic of the United States population and are thus important to study. Since 1980, the number of immigrants in the U.S. has increased by more than 29 million and the immigrant share of the total U.S. population has nearly doubled.[1] Among political rhetoric that aims to instill fear of immigrants, it is important to obtain a clear picture of those who decide to come to the U.S.

Migrants are not a random cross-section of their origin country's population; rather, migrants tend to self-select on observable and unobservable characteristics such as income, wealth, educational attainment, and ability. In this way, the decision to migrate is endogenous and certain individuals are more likely to migrate than others. Migrants are a self-selected population, but little research has been conducted on the characteristics of self-selected migrants and how their characteristics relate to those who decide to stay in their home countries. It is also plausible composition of migrants also changes during times of economic stress, as individuals with certain characteristics may be more likely to leave during times of financial hardship. In this study, we examine selectivity of Puerto Rican migrants to the U.S. mainland to obtain an idea of who decides to leave versus who decides to stay, specifically in times of economic stress. Throughout the paper we distinguish between the "mainland" and the "island". We use the term "mainland" to refer to the 50 states in the United States of America excluding Puerto Rico. We also use the term "island" to refer to the island of Puerto Rico.

---

[1] Immigrant share of the total U.S. population increased from 6.2% in 1980 to 13.5% in 2015 (Migration Policy Institute, 2016).

In order to understand who is most likely to leave their home country, we compare Puerto Rican migrants to the mainland with Puerto Rican island residents. We examine where in the island income distribution Puerto Rican migrants to the mainland fall. In this way, we can understand where in the home "country" income distribution migrants tend to come from. More specifically we look for whether migrants tend to be those with the highest, lowest, or intermediate incomes from their home "country". To account for differing returns to skill on the mainland and in Puerto Rico, we use methods developed by DiNardo, Lemieux, and Fontin (1995) to construct counterfactual income densities. These densities estimate the readjusted income of migrants on the mainland had these migrants been paid under the same returns to skill as Puerto Rican island residents. This allows for a direct comparison between income densities of Puerto Ricans who migrated to the mainland and Puerto Rican residents who remain on the island.

In previous studies, a positive self-selection bias is defined to exist when migrants come from the upper portion of their source country's income and skill distributions. A negative self-selection bias exists when migrants come from the lower portion of their source country's skill and income distributions (Borjas, 1987; Chiquiar and Hanson, 2005). We aim to analyze the direction of this self-selection bias throughout the period 2005 to 2016. We hypothesize that positive self-selection bias increased after the Great Recession affording only those with the highest incomes the ability to migrate from Puerto Rico to the mainland. We predict that this trend may be due to two factors: (1) increased cost of migration for those at the bottom of the income distribution, and (2) decreased labor demand for low skill workers on the mainland.

We choose to examine changes in the self-selection of Puerto Rican migrants specifically due to the large availability of data. Puerto Rican migrants make up a substantial portion of the population, comprising about 10 percent of the total Latino population on the mainland (Acosta-

Belén and Santiago, 2006).[2] The U.S. Census' American Community Survey (ACS) and Puerto Rican Community Survey (PRCS) provide detailed accounts of Puerto Rican migrants to the mainland and Puerto Rican island residents respectively. Both surveys are conducted and sampled in the same way and are conducted annually which also allows us to study changes in self-selection patterns year over year.[3] In 1917, passage of the Jones Act granted U.S. citizenship to all Puerto Ricans allowing them to move freely between the island of Puerto Rico and the mainland United States absent legal barriers. As such, Puerto Rican migration is useful to study as migration is uninhibited by legal barriers to entry and thus must be inhibited by other costs related to individual characteristics. From 2005 to 2016 we find that Puerto Rican migrants to the U.S. mainland were positively self-selected, coming from the top 20% of the Puerto Rican income distribution. This pattern remained stable over the time period under consideration, with little to no effect of the Great Recession on selectivity patterns.

This paper is organized as follows: first we provide a review of the literature relevant to changes in migrant selectivity. Second, we discuss ACS and PRCS data used in the study. Third, we introduce counterfactual income density methodology used in the study. Fourth, we address empirical findings and finally we conclude.

---

[2] Mexican Americans make up the largest portion of Latino migrants to the U.S. (Collazo et al., 2010).
[3] Individuals in the population are stratified into counties, municipalities, and county equivalents, and then into either housing units (HU) or group quarters (GQ). Housing units includes houses, apartments, mobile homes, and trailers. Group quarters include places such as nursing facilities, adult correctional facilities, psychiatric hospitals, college housing, military barracks, group homes, etc. County/county equivalents are sampled at pre-specified sampling rates and then HUs and GQs are sampled within the chosen county/county equivalents also at pre-specified sampling rates (Design and Methodology: American Community Survey, 2009).

Literature Review: Self-Selection of Migrants

In this study we examine self-selection bias on income. Previous research has studied income self-selection by comparing incomes of migrants to the United States with those of home country residents. For instance, Chiquiar and Hanson (2005) examine male Mexican migrants to the United States in the 1990 and 2000 Censes and find that if these migrants were paid according to current skill prices in Mexico, they would be located in the upper-middle portion of Mexico's income distribution. This reflects a positive self-selection bias. Ramos (1992) finds negative self-selection bias of Puerto Ricans migrants to the mainland. Using 1980 Census data on male workers ages 20-64, he finds that Puerto Rican migrants to the mainland in 1980 are less educated and have lower earnings than those on the island. Sotomayor (2009) finds contrasting results that Puerto Rican migrants to the mainland during 1990 and 2000 are a positively self-selected population. Using 1990 and 2000 U.S. and Puerto Rico Censes data on male workers ages 25-64, Sotomayor finds that if Puerto Ricans in the United states were paid according to current skill prices in Puerto Rico, they would be located in the upper portion of the island's educational attainment and income distributions.

These studies use Census data to focus on a population of migrants during a few specific years, without considering short-term changes in selectivity over time. The empirical studies on Puerto Rican migrants account only account for every tenth year from 1980-2010. Instead, this study uses yearly data from the ACS and PRCS from 2005 to 2016 to understand how migrant selectivity changed before, during and after the Great Recession. We aim to understand if those with the highest or lowest incomes were more likely to leave Puerto Rico for the mainland during the period.

The decision to migrate from a country of origin depends on a variety of political and economic factors; however, immigration decisions are predicted on personal characteristics and the relative demand for those characteristics at home versus abroad. Labor demand on the mainland and in Puerto Rico changed after the Great Recession in ways that may have affected migrant selectivity. For over a decade Puerto Rico has been in almost continuous recession. The island experienced serious financial decline starting in 2006, as the government ended corporate tax breaks leading to business shutdowns and increased unemployment (Krogstad et. al, 2017). Puerto Rico experienced a sharper downturn than the mainland after the Great Recession and the number of manufacturing jobs on the island was halved by 2014.[4] These factors potentially increased the cost of migration, specifically for those with the lowest incomes and highest liquidity constraints. On the mainland, labor demand for low skill and low paying jobs drastically decreased after the Great Recession, with a significant decrease in construction and manufacturing jobs (Rothstein, 2017).[5] For these reasons we hypothesize that only those with the highest incomes were able to migrate from Puerto Rico to the mainland after the recession – those with lower incomes potentially faced an increase cost of migration as well as a decreased demand for labor on the mainland.

---

[4] The Puerto Rican economy has shrunk by more than 10% since 2006. Employment on the island has fallen by more than 14%. In 2015, the island's unemployment rate was at 11.6% which was more than twice the national average. The large scale of the downturn in Puerto Rico was a result of many factors including a large debt, a boom and bust cycle in real estate and credit, and fiscal crisis (U.S. Treasury, 2017).

[5] Since 2007, the manufacturing sector on the mainland has continued to decrease, losing roughly 1.3 million jobs since the start of the recession (Pew Research Center, 2017). Contrastingly, the professional and business services and food services industries have gained roughly 2.8 and 2.1 million jobs respectively since the start of the recession (Pew Research Center, 2017). The healthcare sector added 3.6 million jobs since the start of the recession (Pew Research Center, 2017).

Data

This study uses ACS and PRCS data from 2006 to 2015 drawn from integrated public use microdata samples (IPUMS). The ACS surveys individuals on the mainland while the PRCS surveys island residents. Analysis is restricted to individuals ages 18-65 born in Puerto Rico both in the PRCS and ACS, as they are the most likely to be employed. We combine the two surveys to form a substantive population of Puerto Rican island residents (in the PRCS), and Puerto Rican migrants to the U.S. mainland (in the ACS). Puerto Rico is a useful starting point for the study of migrant selectivity due to a large population of migrants in the destination country as well as a substantial population of island residents. Table 1 below shows the population of the island and

Table 1. *Population Tabulations (millions)*		
Year	Puerto Rico	Puerto Ricans on Mainland U.S.
2005	3.821	1.339
2006	3.805	1.426
2007	3.783	1.470
2008	3.761	1.398
2009	3.740	1.498
2010	3.722	1.510
2011	3.769	1.602
2012	3.634	1.547
2013	3.593	1.621
2014	3.535	1.634
2015	3.473	1.686
2016	3.411	1.755

Source: ACS and PRCS. Note that the right column is Puerto Rican migrants on the mainland, not the total population of Puerto Ricans on the mainland. The right column encompasses only those who were born in Puerto Rico and moved to the U.S. mainland.

the population of Puerto Rican migrants on the mainland in each year. The population of Puerto Rico has steadily declined by 10.7% from 2005 to 2016. Contrastingly, the number of Puerto Rican

migrants in the United States increased by 31% from 2005 to 2016, from about 1.4 to 1.75 million

people. Table 2 shows differences in characteristics of island residents and Puerto Rican migrants

to the mainland in select years. On average, Puerto Rican migrants to the mainland have lower

rates of educational attainment but higher annual incomes than island residents. Characteristics of

island residents and Puerto Rican migrants are relatively stable from 2005 to 2016. Annual income

declined by 4.5% for Puerto Ricans on the mainland and 0.86% for island residents from 2008 to

2009. We cannot take this as evidence that Puerto Rican migrants self-select from the lower or

upper portion of the Puerto Rican skill distribution until we adjust for differing returns to skill on

the mainland and in Puerto Rico.

Table 2. *Characteristics of Puerto Rican (PR) residents and Puerto Rican migrants to the U.S. (mainland)*

Characteristic	2005			2008			2009			2016		
	U.S.	PR	Total	U.S.	PR	Total	U.S.	PR	Total	U.S.	PR	Total
Total Income (mean)	21778	11861	14794	21891	10730	14011	20901	10638	13803	20560	10667	14308
Age (mean)	43.8	39.9	41.1	43.7	41.3	41.2	43.1	40.3	41.2	42.7	41.1	41.7
Male (%)	47.1	47.2	47.2	49.8	47.8	48.4	49.4	47.9	48.4	48.5	47.6	47.9
Speaks English (%)	81.3	38.5	51.2	82.5	36.0	49.7	81.8	35.4	49.7	81.2	40.3	55.3
Less than high school (%)	14.3	13.7	13.9	12.3	13.9	13.5	12.2	12.2	12.2	8.10	7.70	7.90
Some high school (%)	18.0	12.2	13.9	17.3	10.7	12.7	17.5	10.7	12.8	13.6	7.40	9.70
Graduated high school (%)	29.5	26.4	27.3	29.0	25.6	26.6	29.3	27.2	27.8	30.7	28.5	29.3
Some college (%)	23.4	27.5	26.3	25.5	28.7	27.8	25.3	28.2	27.3	29.2	30.8	30.1
Graduated college or higher (%)	14.9	20.1	18.6	16.0	19.4	19.4	15.2	21.0	19.3	18.6	25.4	23.0

Note. – Tabulated using ACS and PRCS and weighted to represent entire populations

Empirical Methods: Counterfactual Income Density Estimation

This study uses counterfactual income density estimation to study changes in Puerto Rican migrant selectivity over time. Our goal is to identify characteristics, specifically income, that make migrants to the mainland different from the rest of the island's population. Counterfactual density estimation allows for the comparison of mainland migrant incomes versus island resident incomes under equivalent returns to skill. Counterfactual income densities estimate the income of Puerto Rican migrants on the mainland in each year, adjusted for returns to skill in Puerto Rico. In other words, the income distribution of migrants on the mainland is readjusted to represent income had migrants been paid under the same returns to skill as island residents. This allows for a direct comparison of income densities of Puerto Ricans who migrated to the mainland versus those who remained on the island.

The type of counterfactual income construction used in this study was initially developed by Lemieux, Fontin, and DiNardo (1995) and applied to the study of migrant selection by Chiquiar and Hanson (2005). Chiquiar and Hanson (2005) use the 1999 and 2000 Mexican and U.S. Censuses to construct counterfactual income densities that Mexican immigrants in the U.S. would be paid according to skill prices in Mexico and compare this to observed income densities for residents of Mexico. Their findings suggest evidence of intermediate or positive selection of Mexican immigrants to the United States in 1999 and 2000. While Chiquiar and Hanson (2005) use counterfactual income density estimation to obtain a snapshot of migrant selectivity at a specific point in time, this study will apply the counterfactual density estimation used in Chiquiar and Hanson (2005) to a longer period of time in order to understand changes in migrant selectivity from 2005 to 2016.

To construct the counterfactual income densities of Puerto Rican migrants to the mainland adjusted for Puerto Rican returns to skill, we let $F(I, x, c)$ denote a joint distribution of income I, observable individual characteristics x, and countries c, where each observation in the distribution corresponds to an individual migrant or non-migrant. The universe of observations includes individuals in Puerto Rico and Puerto Ricans in the United States, all of whom were born in Puerto Rico. Then, $f_c(I)$ is the marginal distribution of income in country c and is a function of individual characteristics. It can be denoted as follows:

$$f(I|c) = \int_{x \in \Omega_x} f(I, x|c)dx = \int_{x \in \Omega_x} f(I|x, c)h(x|c)dx \ . \qquad (1)$$
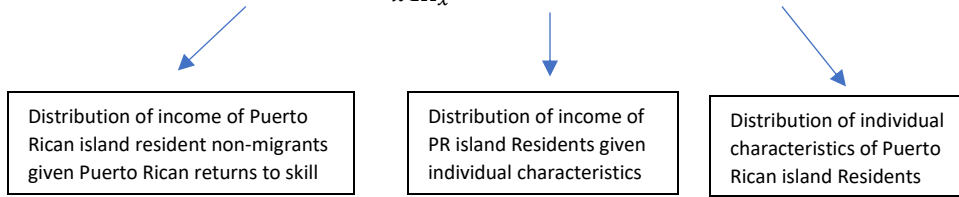
The above expression follows from properties of joint distributions, and Ω_x is the domain of observable characteristics in a given country.

There is a general difficulty associated with the examination of migrant self-selection to the mainland. Income data are only observed for migrants in a destination country and non-migrants of a source country. There is a lack of data that track labor market performance of migrants prior to migration in their country of origin, as countries tend not to catalog or track emigrants once they have left their country of origin.[6] In this case, we can only observe the income distribution of resident non-migrants in Puerto Rico and the income distribution of Puerto Rican migrants on the mainland, as opposed to the income distribution of the same migrants to the mainland before and after they left Puerto Rico. These observed distributions of income are shown below both mathematically and intuitively:
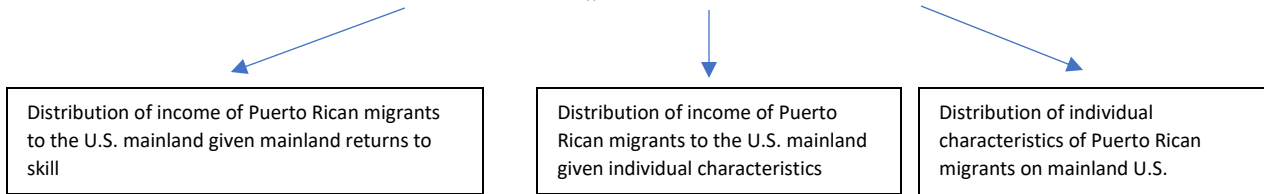
---

[6] One exception to this lack of data is the *Survey on the Federated States of Micronesia Migrants in the United States including Guam and Hawaii*. This survey tracks Micronesian emigrants and their labor market conditions both prior to migration and after arrival in the United States.

$$f(I|c = PR) = \int_{x \in \Omega_x} f(I|x, c = PR)h(x|c = PR)dx \qquad (2)$$

| Distribution of income of Puerto Rican island resident non-migrants given Puerto Rican returns to skill | Distribution of income of PR island Residents given individual characteristics | Distribution of individual characteristics of Puerto Rican island Residents |

$$f(I|c = US) = \int_{x \in \Omega_x} f(I|x, c = US)h(x|c = US)dx . \qquad (3)$$

| Distribution of income of Puerto Rican migrants to the U.S. mainland given mainland returns to skill | Distribution of income of Puerto Rican migrants to the U.S. mainland given individual characteristics | Distribution of individual characteristics of Puerto Rican migrants on mainland U.S. |

Deciles for the empirically observed income distributions are averaged over the period from 2005 to 2016 and shown in Table 2 below.[7] On average, the deciles of the yearly distributions of annual income for Puerto Rican migrants to the mainland is higher than those of island residents by

Table 3. *Average deciles of observed log annual income distributions for employed Puerto Rican Populations (Age 18-65)*

Decile	Income of Migrants to Mainland U.S. (USD Year 2000 Standardized)	Income of Puerto Rican Residents (USD Year 2000 Standardized)	Difference
10%	8.86	8.27	0.59
20%	9.34	8.81	0.53
30%	9.63	9.12	0.51
40%	9.84	9.31	0.53
50%	10.03	9.50	0.54
60%	10.21	9.70	0.52
70%	10.41	9.91	0.50
80%	10.64	10.15	0.49
90%	10.96	10.50	0.46
100%	13.54	13.01	0.53

Note. – tabulated using ACS and PRCS and weighted to represent entire populations

---

[7] We average the deciles over the period because both distributions are relatively stable over the 12-year period. See Appendix Table A.1 and A.2 for the deciles of each distribution in each year.

approximately \$0.50 (Year 2000 Standardized).[8] However, these observed distributions cannot be directly compared as Puerto Rico and the mainland face differing returns to skill.

To understand differences between migrants to the mainland and island residents, we need to examine a distribution of income of Puerto Rican migrants on the mainland under Puerto Rican island returns to skill. This allows for a direct observation of the labor market performance of migrants and island residents and a clear picture of who leaves. However, this is an unobserved distribution of income. This unobserved distribution of income of Puerto Ricans on the mainland had they been paid according to island returns to skill is represented by:

$$\int_{x\in\Omega_x} f(I|x, c = PR)h(x|c = US)dx .$$ (4)

> = (**Observed** distribution of Puerto Rican island resident income) * (**Unobserved** distribution of characteristics of Puerto Ricans on mainland)

Estimation of a counterfactual income distribution allows us to get around this inability to see the distribution of income of migrants in the mainland under Puerto Rican returns to skill. The above distribution is unobserved, but can be expressed in terms of an observed distribution as follows:

$$\int_{x\in\Omega_x} \theta_x(x)f(I|x, c = PR)h(x|c = PR)dx$$ (5)

> = (reweighting factor) *(**Observed** distribution of income of Puerto Ricans in the PR)* (**Observed** distribution of individual characteristics of Puerto Ricans in the PR)

$$\text{Where } \theta_x(x) = \frac{h(x|c=US)}{h(x|c=PR)} = \frac{P(c=PR)}{P(c=US)} * \frac{f(c=US|x)}{f(c=PR|x)} \text{ (By Bayes' rule).}$$ (6)

$\theta_x(x)$ can be thought of as a re-weighting factor that allows for the derivation of a counterfactual income distribution of Puerto Ricans in the mainland had they remained in Puerto Rico based on

---

[8] The log of annual income is taken to normalize the distribution.

the observed distribution of characteristics of Puerto Rican migrants on the mainland. Thus eq. (5) represents the distribution of income that would have prevailed if observable characteristics were valued in Puerto Rico and migrants had been paid according to income observed on the mainland.

Estimation of the counterfactual income density then follows through a reweighted kernel density estimation, with weights given by $\theta_x(x)$. Both $P(c = US)$ and $P(c = PR)$ are given in the data as the probability of an observation (individual) being on the mainland or the island. This is simply the weighted number of observations in the mainland and the island respectively. The conditionals *f(c=US | x)* and *f(c=PR | x)* are the probabilities of an individual of being on the mainland or the island given certain individual characteristics and can be estimated through logistic regression. We note that log annual income is used in density estimation as it produces a normal distribution which satisfies kernel density estimation assumptions. [9] This study estimates counterfactual log annual income densities of Puerto Rican migrants on the mainland under island returns to skill each year from 2005-2016. These densities are then compared to island resident income densities over time to understand changes in migrant selectivity during this period. We predict that estimated counterfactual income densities will reveal some sort of change in the selectivity over time of Puerto Rican migrants to the mainland.

---

[9] Kernel density estimation is a non-parametric way to estimate the PDF of a random variable. This type of estimation takes input of a Kernel (non-negative function that integrates to 1), and a bandwidth (smoothing parameter). Our estimates of counterfactual densities are based on the Gaussian kernel function, with log annual income being normally distributed. Bandwidth selection is based on optimal bandwidth estimation proposed by Silverman (1986).

Results

We apply the empirical methodology to the combined sample of Puerto Rican island residents and Puerto Rican migrants to the mainland. An example of logit results used to construct the reweights for density estimation are shown in Appendix Table A.3. Figure 1 shows an example of results from the log annual income density estimates for Puerto Rican migrants to the mainland reweighted to capture returns to skill in Puerto Rico (counterfactual densities), compared with income densities for Puerto Rican island residents. To illustrate, graphical results from 2005 are shown below. Graphs for 2006-2016 are found in appendix Figure A.4.

Figure 1: (Top) Counterfactual income density of migrants on the mainland paid according to island returns to skill and Observed income density of Puerto Rican island residents in 2005. (Bottom) Difference in Densities = counterfactual (US) – Observed (PR). Straight dotted line indicates the median income.

Counterfactual and Observed Densities (2005)

Difference in Densities = Line 1 - Line 2
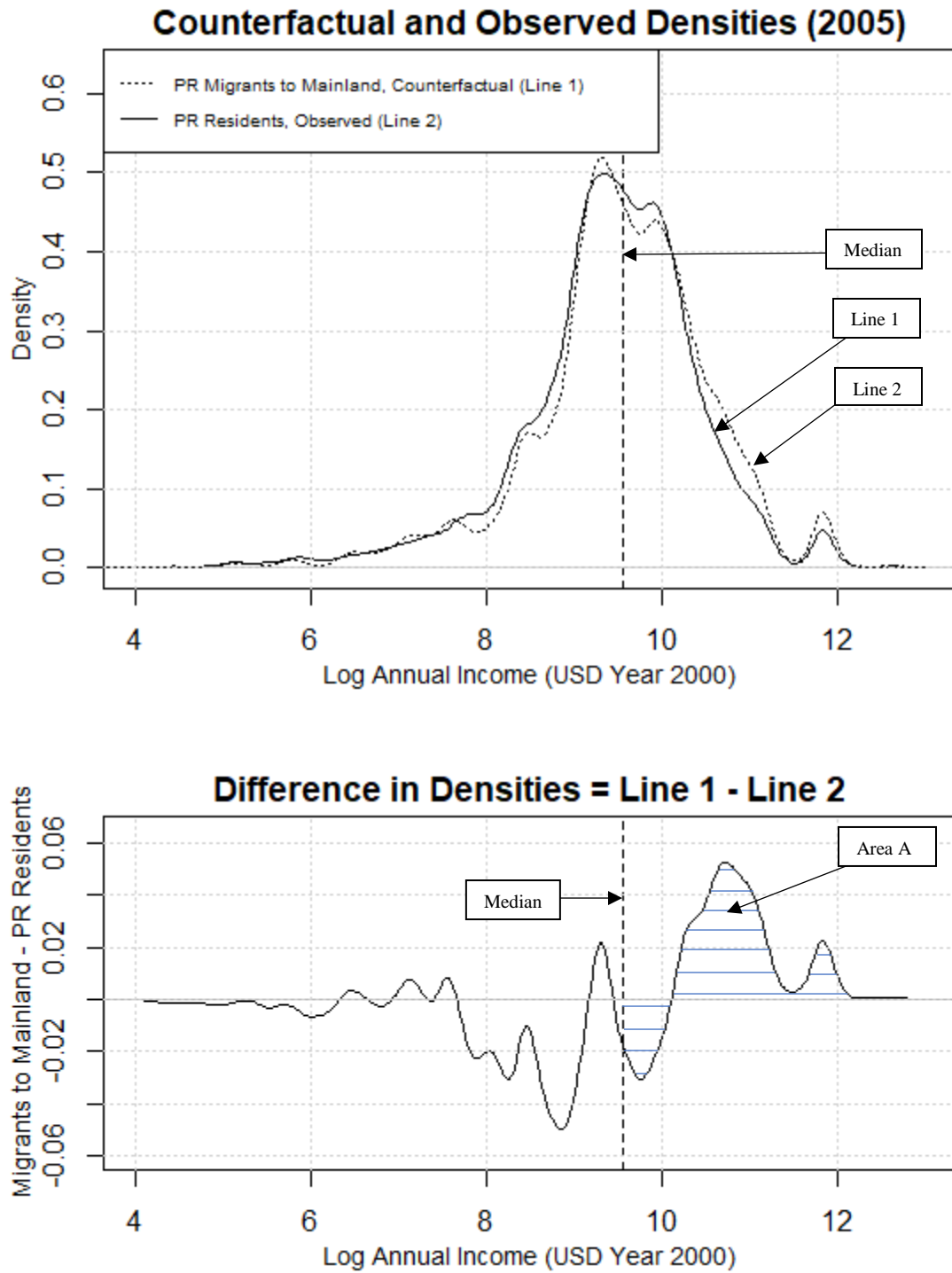
The upper graph shows the estimated counterfactual income density of migrants to the mainland adjusted for Puerto Rican returns to skill (Line 1) compared to the income density of island residents (Line 2). The counterfactual density is slightly to the right of the island resident density in all years suggesting that migrants on the mainland would earn slightly higher incomes than island residents had they been observed under returns to skill on the island (See Appendix Figure A.4 for density estimates in all years). In other words, if migrants to the mainland had stayed in Puerto Rico, they would have earned slightly more than other Puerto Rican island residents. This is evidence of positive self-selection bias in all years. The two densities are relatively close together each year, but Kolmogorov-Smirnov tests in each year confirm that estimated counterfactual densities and are significantly different from observed densities at the 95% significance level.[10]

The bottom graph shows the difference between the counterfactual income density of migrants on the mainland and Puerto Rican island residents (i.e. the difference between the Line 1 and Line 2 values). Positive mass indicates a higher prevalence of migrants in that portion of the log annual income density. In all years, we see a positive mass in Area A, or the upper portion of the distribution above the median, meaning that there are more migrants to the mainland than island residents above the median island income – migrants on the mainland are more likely to come from this upper section of the distribution (See Appendix Figure A.4 for density differences

---

[10] The Kolmogorov-Smirnov test is a test of the equality of two probability distributions by quantifying the distance between two empirical distributions, which is exactly what is shown in the bottom graph of Figure 1. The test operates under the null hypothesis that that two samples are drawn from the same continuous distribution. We find p-values below 0.01 in each year, supporting that the counterfactual density of log income of migrants to the U.S. mainland is significantly different from the observed density of log income of Puerto Rican island residents.

in all years). This is evidence of a positive self-selection bias. Results of this exercise are shown

for each year from 2005 to 2016 in Table 4 below.

Table 4. *Area under the difference curve and above median*	
Year	Area A
2005	0.0385
2006	0.0525
2007	0.0385
2008	0.0474
2009	0.0441
2010	0.0359
2011	0.0288
2012	0.0199
2013	0.0463
2014	0.0503
2015	0.0437
2016	0.0380

Note. – Area A is positive in all years indicating positive self-selection bias of migrants to the mainland throughout the time period under consideration.

Positive values of A indicate that counterfactual densities of migrants on the mainland adjusted for

Puerto Rican returns to skill are slightly above of the observed densities of island residents in every

year. This is evidence of a positive self-selection bias of migrants to the mainland, with a higher

prevalence of those in the upper portion of the income density deciding to migrate.

We apply a test introduced by Chiquiar and Hanson (2005) to quantify differences between

the counterfactual migrant densities and the observed Puerto Rican island resident densities within

each year to see if self-selection has significantly changed over the period. Chiquiar and Hanson

(2005) quantify differences in migrant and non-migrant densities by examining differences in

deciles of each density. We see how many individuals in the migrant counterfactual density would

fall within the income bounds of each decile of the observed income distribution of island residents. By definition, each decile of the observed island resident income density has 10% of the total population. First, we sort observations in each year of the Puerto Rican island resident income density by decile. Then we reweight those observations using the reweighting factor proposed in eq. (6) to adjust the Puerto Rican island resident population as if they had the characteristics of Puerto Rican migrants to the mainland. This is the same procedure used to produce counterfactual density estimations. We then test whether the weighted share of the migrant counterfactual population within the given decile is significantly different from 10%. Positive selection is indicated if the share of migrant counterfactual population is above 10% in the deciles above the median. This would mean that there is a higher proportion of migrants than non-migrants in those upper deciles of the income density. Results are shown in Table 5.

Table 5. *Results of the decile test – Proportion of Puerto Rican migrants to the U.S. mainland in each decile of the Puerto Rican resident income distribution*

Decile	Share of Residents	Year											
		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
0-10%	0.10	0.093	0.095	0.082	0.095	0.093	0.089	0.091	0.091	0.081	0.089	0.092	0.086
10-20%	0.10	0.083	0.087	0.089	0.105	0.080	0.086	0.096	0.099	0.086	0.080	0.088	0.079
20-30%	0.10	0.080	0.086	0.085	0.092	0.101	0.086	0.089	0.102	0.098	0.090	0.086	0.094
30-40%	0.10	0.096	0.094	0.101	0.083	0.091	0.095	0.076	0.101	0.091	0.090	0.085	0.093
40-50%	0.10	0.107	0.085	0.098	0.086	0.093	0.101	0.085	0.088	0.103	0.088	0.102	0.108
50-60%	0.10	0.096	0.096	0.096	0.081	0.093	0.090	0.100	0.085	0.089	0.104	0.083	0.098
60-70%	0.10	0.091	0.098	0.099	0.091	0.108	0.105	0.095	0.085	0.107	0.101	0.109	0.098
70-80%	0.10	0.100	0.107	0.099	0.110	0.100	0.089	0.102	0.100	0.099	0.101	0.104	0.097
80-90%	0.10	0.113	0.120	0.118	0.120	0.100	0.117	0.101	0.119	0.110	0.110	0.108	0.121
90-100%	0.10	0.141	0.131	0.143	0.138	0.142	0.140	0.145	0.130	0.137	0.148	0.142	0.124

Note. – All colored cells are significantly different than 0.10 at the 95% significance level. Green cells indicate a higher proportion of migrants (adjusted for Puerto Rican returns to skill) within that decile, while red cells indicate a lower proportion. Green cells are mainly concentrated above the median, indicating positive self-selection bias.

Green cells indicate deciles with a significantly greater number of migrants than island residents, and red cells indicate deciles with a significantly lower number of migrants than island

residents. The test shows that proportions significantly greater than 0.10 are generally above the median, concentrated mainly in the top 20% of the Puerto Rican income distribution in each year. In other words, migrants to the mainland consistently came from the top 20% of the Puerto Rican income distribution. The top 10% of the income distribution consistently contains the largest number of migrants to the mainland. This is evidence of a positive self-selection bias throughout 2005-2016. Results suggest that individuals in the middle and bottom portion of the income distribution are least likely to migrate.

These results contradict the hypothesis that after serious financial decline in 2006, most migrants to the mainland were less educated and more likely to hold low skill jobs than the island population (Mora, Davila and Rodriguez, 2014). [11] Instead, we observe a consistent pattern of positive self-selection bias. This trend may appear because migration costs are decreasing in income – migration costs become lower as one's income increases, only affording those with the highest incomes the ability to migrate (Chiquiar and Hanson, 2005). Migration costs may also be sufficiently high in an absolute sense such than lower income cannot cover the cost.

The pattern of self-selection bias only changes slightly throughout the period, with the range of positive self-selection bias following a slight U-shaped pattern: in 2005 migrants are most likely in the top 60% of the island income distribution, by 2010 they are most likely in the top 20%, and by 2016 they are most likely from the top 60% again. It is possible that economic stress may have afforded only the migrants with the highest incomes the ability to migrate to the mainland; however, the pattern of self-selection changes little over the period and is unlikely

---

[11] In 2006, the U.S. government ended corporate tax breaks for corporations that operated in Puerto Rico, leading to business shutdowns and increasing unemployment.

associated with the Great Recession. This is inconsistent with previous literature seen in census data on village households in Indonesia; Bazzi (2017) finds that negative agricultural income shocks decrease labor emigration flows in villages with small landholders with lower incomes and increase labor emigration flows in the most developed villages with higher incomes. In this case, a large income shock did not seem to have an effect on selection patterns. This is consistent with initial findings that characteristics of migrants and island residents remain relatively stable from 2005 to 2016.

Results also suggest that those below the median did not see large changes in self-selection patterns over the period. These individuals did not leave Puerto Rico for the mainland and the Great recession may not have altered migration costs for this group. A potential limitation is that many Puerto Ricans decide to migrate between the island and the mainland multiple times, which may tend to overstate self-selection bias.[12] Overall, results of positive self-selection are relatively stable over the time period, and no clear pattern caused by the Great Recession emerges.

Conclusion

Selectivity of migration is important to study as certain individuals are more likely to migrate than others, and by studying selectivity a clearer picture of who decides to leave a source country can be established. Migration is often seen as a source of opportunity, but certain groups are often unable to participate in that decision due to the high costs of migration. Thus, it is also important to study who is *able* to migrate, and why. Few studies examine who is able to migrate

---

[12] If there is much back and forth migration between a source and destination country, those who tend to earn the highest incomes after migration will stay in the destination country while others will migrate back to the source country. This may overstate positive self-selection bias because only the most able migrants will remain in the destination country (Borjas, 1987).

due to lack of data on the individual economic characteristics of migrants and migration costs in particular.

In this study we use data on island residents and Puerto Rican migrants on the mainland to estimate counterfactual income densities of Puerto Rican migrants to the mainland under Puerto Rican island returns to skill. We compare this to the income distribution of island residents. Puerto Rico is particularly useful for studying selectivity given its unique migration history with the mainland – island residents have U.S. citizenship and thus face no legal barriers to migration. Empirical results are consistent with positive self-selection bias of migrants from Puerto Rico to mainland United States throughout the period 2005 to 2016. We find that those in the top 20% of the island's income distribution are most likely to migrate to the mainland.

These results are consistent with the idea that migration costs are decreasing in skill: those with the highest incomes (and likely highest skill levels) will have the lowest migration costs and thus the lowest barriers to migration (Chiquiar and Hanson, 2005). Those with the highest incomes may be subject to fewer credit constraints when financing migration and the lowest risk aversion for migration. Empirical results also show that selectivity patterns did not change substantially before, during, or after the Great Recession. This is surprising, as labor demand for low skilled jobs decreased drastically on the mainland, and depressed conditions on the island caused increased migration flows. These results point to the idea that when immigration is uninhibited by legal barriers, it is still those with the highest incomes who will tend to migrate. Though we cannot extrapolate results to migrants from other countries, future research and policy should consider the idea that increasing immigration barriers may be an unproductive way to ensure migrants are the those with the highest skill.

Appendix

Table A.1: Quintiles of log annual income distributions for employed Puerto Rican residents age 18-65

Quintile	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
min	-0.37	-0.38	-0.34	0.99	3.28	-0.40	-0.54	-0.51	1.98	-0.54	-0.59	-0.64
10%	8.23	8.28	8.23	8.25	8.25	8.42	8.28	8.33	8.29	8.26	8.21	8.24
20%	8.79	8.86	8.79	8.80	8.78	8.93	8.83	8.84	8.83	8.79	8.77	8.75
30%	9.08	9.12	9.08	9.08	9.08	9.22	9.15	9.17	9.14	9.11	9.10	9.07
40%	9.29	9.31	9.29	9.27	9.28	9.38	9.33	9.33	9.32	9.30	9.29	9.28
median	9.51	9.52	9.51	9.47	9.45	9.57	9.51	9.52	9.48	9.49	9.49	9.44
60%	9.72	9.70	9.72	9.68	9.67	9.76	9.70	9.71	9.68	9.67	9.68	9.66
70%	9.93	9.91	9.93	9.89	9.89	9.97	9.91	9.93	9.91	9.89	9.90	9.87
80%	10.17	10.15	10.17	10.13	10.12	10.19	10.14	10.14	10.15	10.13	10.15	10.11
90%	10.54	10.51	10.54	10.47	10.47	10.56	10.49	10.49	10.51	10.50	10.48	10.46
max	12.97	12.88	12.97	12.94	13.05	13.06	12.69	13.26	13.01	12.87	13.39	13.04

NOTE - Tabulated using American Community Survey and Puerto Rican Community Surveys and weighted to represent entire populations

Table A.2: Quintiles of log annual income distribution for employed Puerto Rican migrants to the mainland ages 18-65.

Quintile	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
min	0.89	1.93	0.97	0.88	2.56	4.33	3.39	-0.50	-0.63	2.29	-0.66	0.74
10%	8.97	8.92	8.97	8.95	8.95	8.86	8.80	8.83	8.73	8.73	8.83	8.80
20%	9.42	9.40	9.42	9.42	9.40	9.35	9.31	9.33	9.26	9.25	9.32	9.25
30%	9.70	9.68	9.70	9.70	9.66	9.63	9.60	9.61	9.57	9.53	9.58	9.55
40%	9.92	9.87	9.92	9.90	9.87	9.84	9.81	9.80	9.80	9.75	9.81	9.78
median	10.10	10.05	10.10	10.08	10.07	10.05	10.01	10.00	9.99	9.95	10.01	9.98
60%	10.25	10.21	10.25	10.26	10.25	10.24	10.20	10.19	10.19	10.15	10.18	10.18
70%	10.43	10.40	10.43	10.43	10.45	10.43	10.40	10.39	10.41	10.36	10.38	10.40
80%	10.65	10.63	10.65	10.66	10.65	10.65	10.63	10.64	10.64	10.60	10.64	10.63
90%	10.96	10.93	10.96	10.96	10.96	10.97	10.97	10.97	10.98	10.95	10.99	10.98
max	13.60	13.45	13.62	13.66	13.64	13.69	13.15	13.51	13.50	13.44	13.90	13.29

NOTE - Tabulated using American Community Survey and Puerto Rican Community Surveys and weighted to represent entire populations

Table A.3: Sample logit regression for estimating re-weights of observed density to construct counterfactual density. Dependent variable is whether an individual is in Puerto Rico or mainland U.S. A new logit regression is used each year, to produce separate counterfactual densities for each year. Data: ACS and PRCS

```
Deviance Residuals:
    Min      1Q   Median      3Q      Max
-2.2329  -0.7225  -0.2997   0.7732   3.2666

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              -4.267774   0.833672  -5.119 3.07e-07 ***
FAMSIZE7                 -0.898697   0.051944 -17.301  < 2e-16 ***
FAMSIZE7SQ                0.098656   0.007038  14.019  < 2e-16 ***
NCHILD5                   0.401024   0.046011   8.716  < 2e-16 ***
NCHILD5SQ               -0.039177   0.013011  -3.011 0.002603 **
SEX                     -0.159701   0.031890  -5.008 5.50e-07 ***
logAGE                   0.792105   0.215455   3.676 0.000237 ***
married                  0.833196   0.430845   1.934 0.053130 .
englishnotwell           2.355846   0.763485   3.086 0.002031 **
englishwell              2.706239   0.751548   3.601 0.000317 ***
englishverywell          5.740855   0.731205   7.851 4.12e-15 ***
englishonly              5.963860   0.828384   7.199 6.05e-13 ***
someHS                   1.289817   0.774001   1.666 0.095628 .
gradHS                  -0.403051   0.730923  -0.551 0.581341
somecoll                -2.826744   0.750681  -3.766 0.000166 ***
gradcoll                -1.602067   0.853410  -1.877 0.060483 .
masters                 -3.125649   1.239428  -2.522 0.011674 *
profdegree              -2.234023   2.002828  -1.115 0.264664
doctoral                 0.475896   3.077132   0.155 0.877093
employed                 0.529305   0.035728  14.815  < 2e-16 ***
unemployed              -0.026202   0.069062  -0.379 0.704394
logAGE.someHS           -0.349225   0.205078  -1.703 0.088588 .
logAGE.gradHS           -0.091840   0.192551  -0.477 0.633387
logAGE.somecoll          0.381195   0.198816   1.917 0.055197 .
logAGE.gradcoll         -0.183445   0.226386  -0.810 0.417757
logAGE.masters           0.259061   0.326306   0.794 0.427241
logAGE.profdegree       -0.158981   0.539130  -0.295 0.768082
logAGE.doctoral         -0.947824   0.807504  -1.174 0.240487
logAGE.englishonly      -0.808416   0.222999  -3.625 0.000289 ***
logAGE.englishverywell  -0.464307   0.196677  -2.361 0.018237 *
logAGE.englishwell       0.086562   0.201435   0.430 0.667395
logAGE.englishnotwell   -0.070163   0.203128  -0.345 0.729785
married.logAGE          -0.302641   0.105483  -2.869 0.004117 **
married.someHS          -0.054812   0.133945  -0.409 0.682380
married.gradHS          -0.033862   0.118654  -0.285 0.775349
married.somecoll         0.120541   0.123639   0.975 0.329589
married.gradcoll         0.338079   0.134528   2.513 0.011968 *
married.masters          0.340039   0.178335   1.907 0.056553 .
married.profdegree       0.879531   0.309481   2.842 0.004484 **
```

```
married.doctoral            0.662667    0.400575    1.654 0.098069 .
married.englishonly         0.719283    0.149579    4.809 1.52e-06 ***
married.englishverywell     0.504794    0.132591    3.807 0.000141 ***
married.englishwell         0.391656    0.134393    2.914 0.003565 **
married.englishnotwell      0.262317    0.133389    1.967 0.049233 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35882  on 29113  degrees of freedom
Residual deviance: 25867  on 29070  degrees of freedom
AIC: 25955

Number of Fisher Scoring iterations: 6
```

Figure A.4: Counterfactual income density for all years (2005-2016). **Left axis:** Log Income Distributions of Puerto Rican Residents (observed) and Puerto Rican migrants to the U.S. mainland (reweighted counterfactual). Counterfactual distributions of migrants are slightly to the right of those of non-migrants indicating a positive self-selection bias. **Right Axis:** Counterfactual income densities of Puerto Rican migrants to the mainland minus observe income density of Puerto Rican residents. Positive mass indicates higher prevalence of migrants from that portion of the income distribution.

2011

2012

2013

2014

2015

2016

PR in PR (observed) — — — PR in US (counterfactual) ——— (PR in US) − (PR in PR)
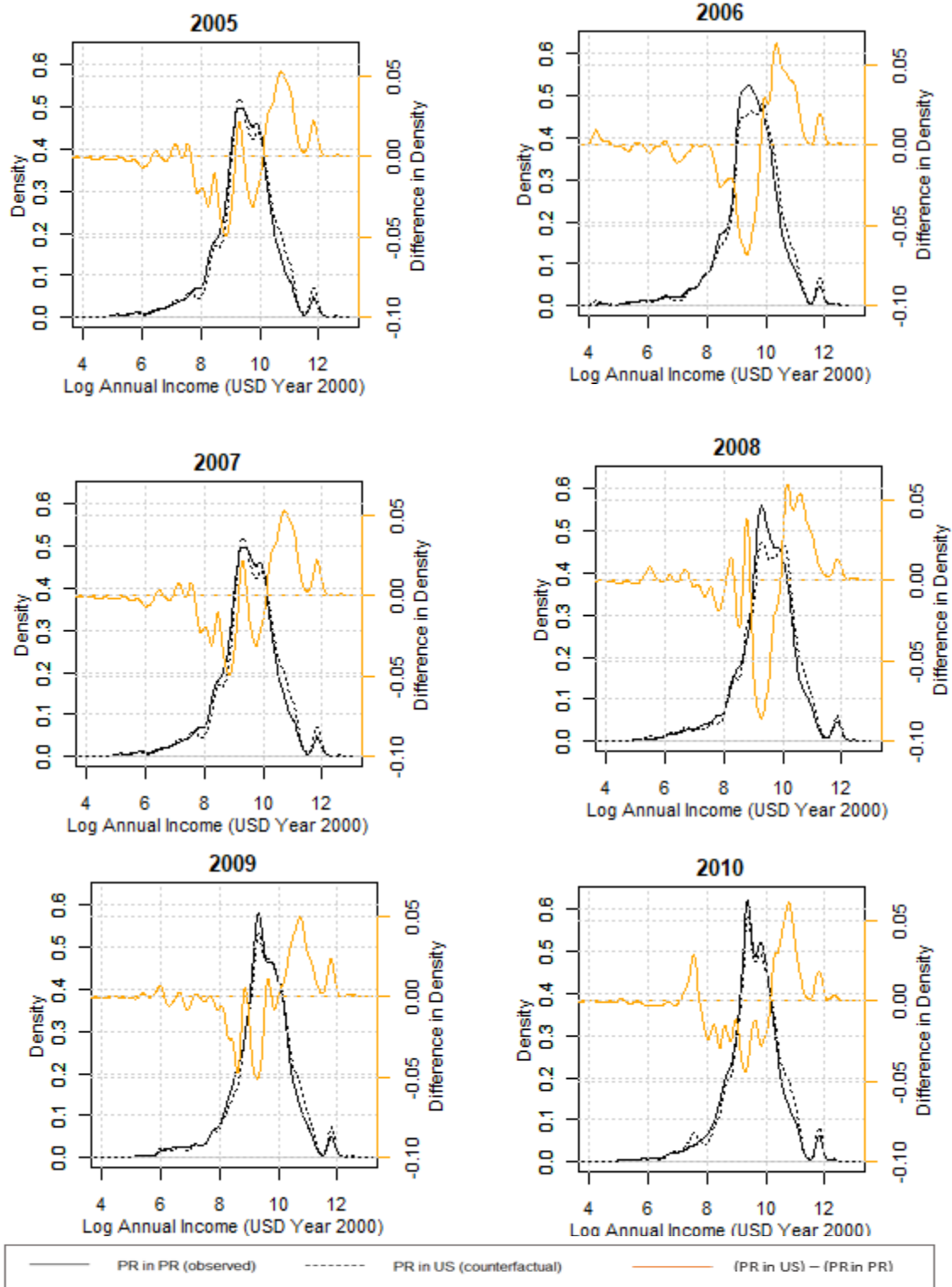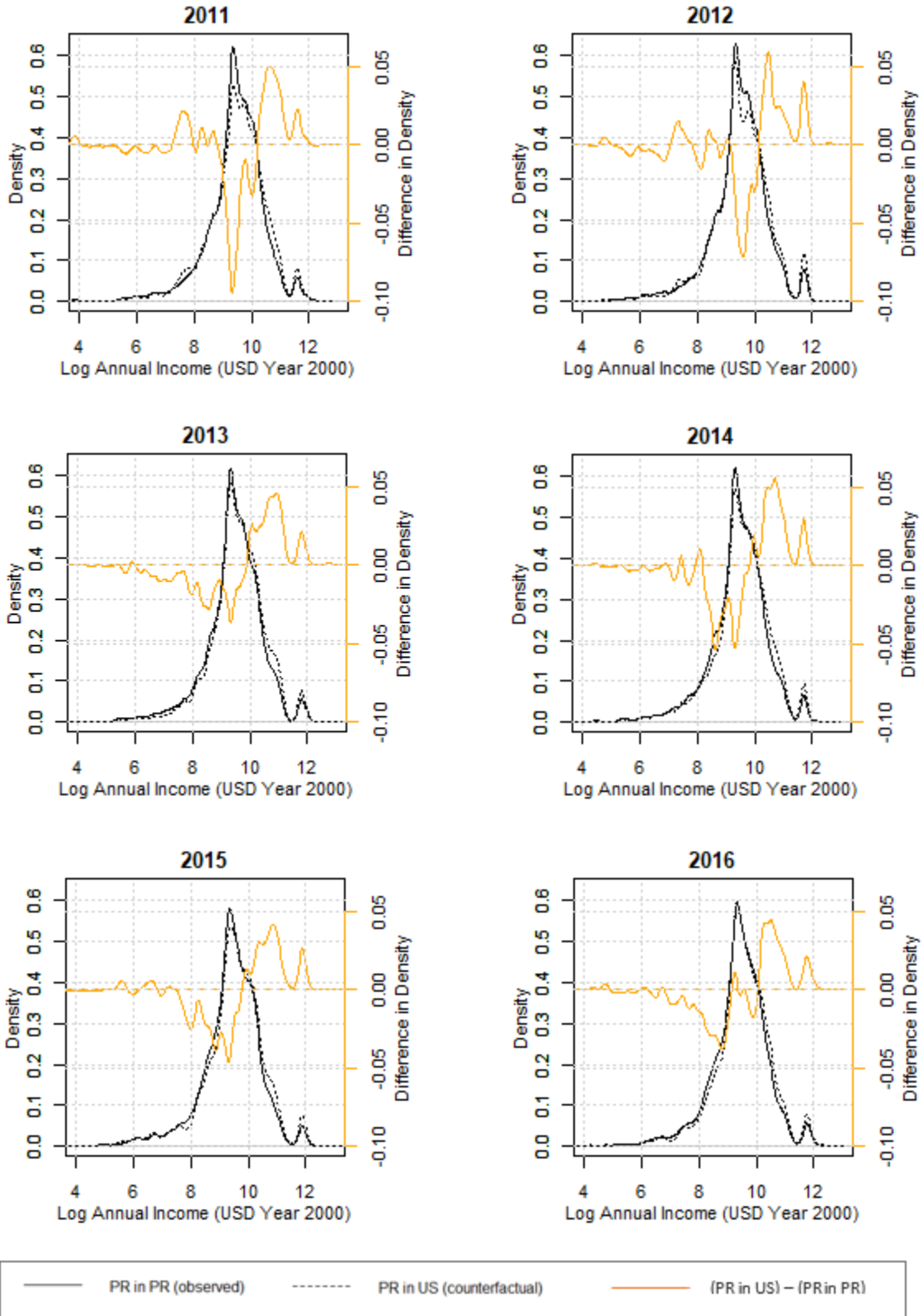
References

Acosta-Belén, E., & Santiago, C. E. (2006). *Puerto Ricans in the United States: A contemporary portrait* (p. 84). Boulder, CO: Lynne Rienner Publishers.

Akee, R. (2010). Who leaves? Deciphering immigrant self-selection from a developing country. *Economic Development and Cultural Change*, *58*(2), 323-344.

Autor, D. (2009). MIT 14.662 Graduate Labor Economics II Spring 2009 Lecture Note 2: Educational Production and Wage Structure.

Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, *9*(2), 219-55.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants.

Browning, H. L., & Feindt, W. (1969). Selectivity of migrants to a metropolis in a developing country: A Mexican case study. *Demography*, *6*(4), 347-357.

Bustamante, J. A., Jasso, G., Taylor, J. E., & Legarreta, P. T. (1998). The selectivity of international labor migration and characteristics of Mexico-to-US migrants: Theoretical considerations. *Migration between Mexico and the United States*, *2*, 699-708.

Chiquiar, D., & Hanson, G. H. (2005). International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of political Economy*, *113*(2), 239-281.

Collazo, S. G., Ryan, C. L., & Bauman, K. J. (2010, April). Profile of the Puerto Rican Population in United States and Puerto Rico: 2008. In *Annual Meeting of the Population Association of America*.

DiNardo, J., Fortin, N. M., & Lemieux, T. (1995). *Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach* (No. w5093). National bureau of economic research.

Hirabayashi, L. R. (1993). *Cultural Capital: Mountain Zapotec Migrant Associations in Mexico City*. University of Arizona Press.

Krogstad, Jens Manuel, et al. (2017). Key Findings about Puerto Rico [Article]. *Pew Research Center.* Retrieved from www.pewresearch.org/fact-tank/2017/03/29/key-findings-about-puerto-rico/.

Lindstrom, D. P., & López Ramírez, A. (2010). Pioneers and followers: Migrant selectivity and the development of US migration streams in Latin America. *The Annals of the American Academy of Political and Social Science*, *630*(1), 53-77.

Lowell, B. L. (1987). Scandinavian exodus: demographic and social development of 19th-century rural communities.

Massey, D. S. (2012). Immigration and the great recession. *A Great Recession Brief*.

Massey, D. S., & Pren, K. A. (2012). Unintended consequences of US immigration policy: explaining the post-1965 surge from Latin America. *Population and development review*, *38*(1), 1-29.

Migration Policy Institute. (2016). U.S. Immigrant Population and Share over Time, 1850-Present. Retrieved from https://www.migrationpolicy.org/programs/data-hub/charts/immigrant-population-over-time.

Ramos, F. (1992). Out-migration and return migration of Puerto Ricans. In *Immigration and the workforce: Economic consequences for the United States and source areas* (pp. 49-66). University of Chicago Press.

Rothstein, J. (2017). The Great Recession and its Aftermath: What Role for Structural Changes?. *RSF*.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge

Torrieri, N. (2014). American Community Survey: Design and Methodology (January 2014). *Available on-line at: http://www. census. gov/acs/www/Downloads/survey_methodology/acs_design_metho dology_report_2014. pdf*.

Villarreal, A. (2014). Explaining the decline in Mexico-US migration: The effect of the Great Recession. *Demography*, *51*(6), 2203-2228.