The Labor Market Effects of a Refugee Wave: Applying the Synthetic Control Method to the Mariel Boatlift*

Giovanni Peri (University of California, Davis)†
Vasil Yasenov (University of California, Davis)‡

December, 7th, 2015

Abstract

We apply the synthetic control method to re-examine the wage and employment effect of the Mariel Boatlift in Miami. We focus exclusively on workers with no high school degree. They are the group competing more closely in the labor market with the newly arrived. We compare Miami's labor market outcomes with those in a control group of cities chosen using the synthetic control method so as to match Miami's wages and other labor market features in the period 1972 to 1979. Using most samples and different outcomes we find *no departure* between Miami and its control between 1979 and 1983. Significant noise exists in many samples but we never find significant negative effects especially right after the Boatlift, when they should have been the strongest. We point out that the very different conclusions in a recent reappraisal by George Borjas (2015) stem from the use of a small sub-sample of high school dropouts in the already very small March-CPS sample. That sample is subject to substantial measurement error and no other sample provides the same findings. Being imprecise about the timing of the data and the choice and validation of the control sample further contribute to the impression of an effect from the boatlift in Borjas (2015). We also revisit the non-Boatlift of 1994, considered by Angrist and Krueger (1999) and we do not find consistent deviations of Miami outcomes from the appropriate control that could be mistaken for labor market effects of a Cuban inflow.

JEL codes: J3, J61

Key Words: Immigration, Wages, Mariel Boatlift, Synthetic Control Method, Measurement Error.

*We started working on the application of the synthetic control method to the Mariel Boatlift in January 2015. In October 2015 we were surprised by Borjas (2015) that found results very different from ours. Part of this paper was developed before that paper. Section 5 and some other parts were written in response to that paper.

†Giovanni Peri: Department of Economics, UC Davis, One Shields Avenue, Davis CA, 95616; email: gperi@ucdavis.edu.

‡Vasil Yasenov: Department of Economics, UC Davis, One Shields Avenue, Davis CA, 95616; email: viyasenov@ucdavis.edu.

1 Introduction

The refugee crisis in Europe during the Summer and Fall of 2015 has once again ignited the discussion about immigration and its economic effects. It has also rekindled the interest in looking at the past history of refugee waves in developed countries to learn from these. How did receiving countries absorb those immigrants, who often arrived in waves? What were their immediate effects on wages and employment and how did the adjustment take place? The US has experienced very large waves of immigrants during the 1990's and 2000's with a portion of them being less educated (many of them undocumented) and another, comparably large portion, of highly educated immigrants. They came in large part for economic reasons (and some also as political refugees) and have contributed to the economic growth of the country. Certainly, in terms of numbers and diffusion, immigration since 1990 is the more important flow to analyze in the US.

An early episode, however, the Mariel Boatlift of April-September 1980, has held a special place in the minds of Americans in terms of refugee arrivals. On one hand, the close to 125,000 Cubans, mostly with low education, arriving within a short period and unexpectedly into Miami constituted an event with high potential for short-run consequences, deeply affecting the local labor market in Miami. On the other hand, an early study by David Card (1990) analyzed this episode and concluded that the impact on employment and wages of low skilled non-Cubans in Miami, in the years immediately following the inflow, was insignificant, making this case a prominent example of how the predictions of a simplistic canonical model of labor demand and labor supply do not work well in analyzing the consequences of immigration. That study was influential and suggested that local economies could absorb a significant number of immigrants without major short-run consequences on wages and employment. Several subsequent studies suggested how different channels for absorbing the "Marielitos" might have worked. Lewis (2004) showed that less skilled Cubans were absorbed by industries that chose "unskilled-labor" intensive technology and less automation. Bodvardson et al. (2008) argued that the immigrants increased significantly local demand for services, and hence also labor demand and not only labor supply.

Prominent textbooks in Labor Economics, both at the undergraduate (Borjas 2012, Laing 2011) and at the graduate level (Cahuc et al. 2014), still use the Mariel Boatlift study by Card (1990) to illustrate "difference-in-difference" as an empirical method. That study is also a prominent example of the lack of effects of immigrants on native wages arguing for the need to go beyond the canonical model. The idea suggested in that study, namely that immigrants can be absorbed relatively quickly and without negative wage or employment effects on natives, has then been confirmed in several broader studies of immigration in the recent years (Card 2001, Card 2009, Ottaviano and Peri 2012, Foged and Peri (forthcoming) among others).

Since Card (1990) we have learned a lot about the impact of immigration by analyzing the effects of long and persistent inflows of immigrants over decades in the US and in other countries. In spite of the challenges to identify the causal effect of immigrants in those scenarios, the existing estimates suggests, on average, small effects on wages and, possibly, positive long-run impacts (see survey in Lewis and Peri, 2015). Hence, in hindsight, the results of the original paper by Card (1990) can be explained as a combination of different absorption mechanisms that also changed labor demand in Miami and of the fact that some noise in the data (from small CPS samples) may obscure small effects.

One reason to revisit the Mariel Boatlift today is that since 1990 we have improved our econometric methods to deal with "case-studies" such as this one. While Mariel qualifies as a quasi-experimental case, being a sudden, large and unexpected temporary shock to immigration, it has the limitation of being a single episode affecting only one labor market (Miami) and hence not very well suited for statistical analysis and inference. Card (1990) identified a group of "comparison cities" and he evaluated the wage and employment outcomes of low skilled people in Miami vis-a-vis those comparison cities, before and after 1979. His method was an application of a difference in difference analysis, but the "ad-hoc" choice of the comparison group and the fact that only one city was "treated" left doubts that accidental events might be affecting the results (as later pointed out by Angrist and Krueger, 1999).

While we cannot change the nature of the episode, we now have an econometric tool, the "synthetic control method", developed and used in a series of papers by Alberto Abadie and coauthors (Abadie and Gardeazabal 20003; Abadie, Diamond and Heinmueller 2010) which is better suited for addressing these type of case-studies. This method analyzes the effects of a treatment that is only given to one "unit" (market, region, country) by comparing the outcome of the treated unit after the treatment, with the outcome of a "synthetic" control group, constructed as to mimic as best possible a set of pre–treatment outcomes and characteristics of the treated unit. In our case the method weights a group of control cities, so as to minimize the pre–1979 difference between Miami and these cities for a set of variables and then uses these weights on cities from the "donor pool" (of all available cities) to construct the post-1979 outcome of a synthetic control to be compared to Miami's outcome.

This method has three advantages. First, it reduces the "ad-hoc" nature of the control group by choosing it as to minimize the distance between a set of variables in the treated and control unit before treatment. Second, it allows to validate the quality of the control group by checking the pre-treatment differences between the outcome variable in the treated unit and in the synthetic control. Finally, by using this method we can construct a p-value for how significant the post-treatment difference between Miami and control is, relative to the pre-treatment difference, and where Miami's statistics stand relative to all permutations of similar statistics for other possible (non treated) units in the donor pool. Applying such a method to the Mariel boatlift seems natural. The seminal paper by Abadie et al (2010) itself mentions the impact of the Mariel Boatlift as a classic example for an application of this method.

The Synthetic control method identifies a group of control cities (among 43 for which data are available) and of weights so that their combination best matches the 1972-1979 levels of low-skilled wages and few other variables in Miami. Then, by analyzing the post 1979 behavior of Miami relative to the control we can infer the potential effects of the Boatlift. An important limitation in the analysis is the small size of samples in the available datasets. We use, preferentially, the CPS May-ORG dataset as it provides the largest sample for Miami since 1979 and then we identify the group of pre-existing Miami workers, most likely to be affected by the boatlift. Non-Cuban workers, with *no high school degree* between 16 and 61 years of age, not self employed and in the labor force constitute such a group. We focus on high school dropouts because, as Card (1990) already noted, the Marielitos had larger percentages of high school dropouts than the pre-1980 Miami population. We consider log wages (annual, weekly and hourly) and unemployment rates of this group, relative to the control, as outcomes.

Our results show no significant difference in the post-1979 labor market outcomes between Miami and Synthetic Miami (the control group). Neither wages (annual, weekly or hourly) nor unemployment of high school dropouts differ between Miami and the control after 1979, up to 1983. Considering wages in the bottom 20th or 15th percentile of the distribution for non-Cuban workers also shows no effect. Our focus is particularly on identifying a sudden drop in wages in 1979-81 or an increase in non-employment in that period. This would be the short-term effect of the increase in low-skilled labor supply implied by the canonical model of labor supply and demand, keeping everything else fixed. While there is a fair amount of noise in the year-to-year variation of log wages and unemployment, we never identify any significant departure of Miami from its control group between 1979 and 1982. We then run "difference in difference" type of regressions including Miami and the Synthetic control to validate the pre-trend match of the synthetic control and to check the post-trend differences. We do not find any systematic deviation post 1979 with this method. Occasionally, some significant differences exist but they are both positive and negative and the only arise after 1983 (three years after the inflow). We also perform statistical inference, by constructing the p-value of Miami versus control relative to all the possible permutations of other 43 cities (that did not receive the Boatlift) and we confirm that the change in wages and unemployment of Miami high school dropouts relative to the control group in 1979-1982 was, by no means, unusual and well within the distribution of other cities' idiosyncratic variation.

Our method, therefore, confirms Card (1990) results. We then invest some time in understanding why a recent paper by Borjas (2015) claims very large negative effects of the boatlift on the wages of high school dropouts in Miami. We find that the main reason is the use of a small sub-sample within the group of the high school dropouts, obtained by eliminating from the sample women, non-Cuban Hispanics and selecting a short age range (25-59). All three of these restrictions are problematic and, in particular, the last two as they eliminate groups on which the effect of Mariel should have been particularly strong (Hispanic and young workers). We can replicate Borjas' results when using this small sub-sample and the smaller March CPS, rather than the larger May-ORG CPS used by all other studies of the Boatlift. The drastic sample restrictions described above leave Borjas with only 17 to 25 observations per year to calculate average wage of high school dropouts in Miami. This increases the measurement error so substantially that not much can be learned from the data. We show that the measurement error for average log wages across metropolitan areas in the Borjas sample from the March-CPS has a standard deviation of 0.15 log points. Hence differences of 15-30% in average wages can easily arise between two cities only because of measurement error. We also show that the ORG-CPS average wages have a much smaller variance across cities due to measurement error. Also, in Borjas (2015), the smoothing of data over time and a choice of the control group with no validation of the pre-1979 match for the outcome variable, contribute to produce the impression of an effect of the boatlift in 1980 and 1981. We show that by adjusting Borjas' sample in minor ways, to include sub-groups of high school dropouts who should be *more intensely* affected by Cuban competition and by being careful in keeping a clean timing (hence no time smoothing) for the 1979 to 1981 outcomes, no significant difference between wages in Miami and in the synthetic control appear around the boatlift even when using the very small March-CPS sample. In one specification, using the small March-CPS sample, some divergence arises after 1982. However, such divergence is not robust. The recession of 1982, other events or measurement error may be the reason why the small sub-group of high school dropouts chosen by Borjas experiences a negative change in wages (relative to the control) in 1982. Even this drop after 1982, however, is not robust to extending the sample to all high school dropouts, nor to using CPS-ORG data, nor to looking at other definitions of wages. No corresponding effect on unemployment is ever observed. Most likely the measurement error in March CPS is responsible for the deviation observed after 1982.

We finally show that, when using the synthetic control method, choosing the appropriate control group for Miami in 1994 and testing different samples and different outcomes, one would not identify a labor market effect for the "Mariel boatlift that did not happen" in 1994. Certainly there is enough year to year noise and measurement error between city and synthetic control outcomes that if we are determined to choose a specific outcome and a specific year, ignoring everything else, we could find non trivial deviations when no event took place, especially for the smaller March-CPS sample. However a more complete and balanced exercise would not mistake the few deviations emerging in 1994 for a labor market shock in Miami caused by Cuban immigrants.

The rest of the paper is organized as follows. Section 2 describes the data and the relevant group of workers in Miami and in the comparison cities that we analyze. We also present the timing of the events and the characteristics of the Marielitos relative to the 1980 non-Cuban Miami labor force and we show measures of the inflow of Cubans in Miami from the CPS. Section 3 describes the Synthetic control method and it discusses the period and variables matched and some details of the cities selected in the synthetic control. Section 4 describes the main results from the Synthetic control method, relative to the impact of the Boatlift on wages and unemployment. In this section we conduct inference using a regression approach (as in Borjas 2015) and a "permutation" approach (as in Abadie et al 2010) to see if the differences in wages (and unemployment) arising in

1979-1981 between Miami and the synthetic control are unusual. As we do not detect any significant difference between wages and employment of high school dropouts in Miami and synthetic control, arising in the 1979-1981 period we compare our method to the one used in Card (1990). We find that, while we improved somewhat in the methodology and precision, our core results are very consistent with Card's finding. Then, in section 5, we consider the reappraisal in Borjas (2015) and we account for what determines the estimates of the large and persistent negative wage effects in that paper. In section 6 we quantify the measurement error for average weekly wages across the metropolitan areas used in the March CPS and the ORG-CPS. In section 7 we also reconcile the odd result of some apparent labor market effects in 1994 with no Boatlift happening, pointed out by Angrist and Krueger (1999). Section 8 concludes the paper.

2 Measuring the Boatlift and Miami Labor Markets

2.1 Number and Demographics of the Mariel Cubans

In order to identify the workers most likely to be affected by the competition from the Mariel immigrants and the timing of their arrival corresponding to a sudden shift in labor supply for that group, we briefly summarize the relevant facts related to the boatlift. Both Card (1990) and Borjas (2015) describe in detail the timing of events, so this is a quick review. They also provide measures, from the Census 1980, 1990 and from the CPS 1985 that identify the characteristics of the Mariel population and of the Miami workers in 1980. Both Card (1990) and Borjas (2015) clearly recognized that the Mariel boatlift consisted of a rather large inflow of Cubans, between April and September 1980 equal to about 125,000 individuals. Most of them (75%) arrived by the end of May 1980[1] in a very concentrated wave. In order to identify the characteristics of the Mariel Cubans one may use the data from Census 1990 (Card used data from the March 1985 mobility supplement to the CPS) and identify the "Marielitos" as individuals born in Cuba who entered the US in 1980. Limiting our attention to those who were between 16 and 65 years old when they arrived, in 1980, and assuming that the percentage arriving in Miami was equal to the percentage of those Cubans still in Miami in 1990 we obtain a total of 57,299 working-age Marielitos. Card (1990) indicates a number of 59,800 working age Mariel Cubans; Borjas (2015) sets the number at 55,700. This would represent an increase of around 7% (Card)

[1]The number of arrivals of Mariel Cubans, each month between April and September 1980, can be seen at https://en.wikipedia.org/wiki/Mariel_boatlift

or 8.4% (Borjas) of the Miami labor force (our estimate puts the number at 8%). One thing that is not well known is what percentage of the Mariel Cuban settled in Miami. Borjas and Card say (without citing specific sources) that about half of them initially settled in Miami[2]. We checked from the census 1990 that 62% of these people, identified as Mariel Cubans, were in still in Miami as of 1990. Hence either at arrival or in the successive years 40% of them located in other places. As we show below there is some evidence in the (noisy) CPS data for Miami that some of the Cubans arrived in 1980 might have left the city in the following 2-3 years. The share of Cubans in the Miami population, in fact, peaks in 1981 and then declines between 1981 and 1985[3]. Overall this implies that the inflow of Cubans into Miami in 1980 was at least equal to the figures reported above, that assume about 60% of Mariel Cubans arriving in Miami. It could be larger if a significant larger percentage of them first settled there, although some clearly moved out by 1990.

So, using different data sources, we come to similar aggregate number for Mariel Cubans who landed in the US between April and September 1980 and one can reasonably think that those numbers are close to the change in immigrants in the Miami labor force during 1980. When looking at their educational structure there has also been agreement that the largest share of Marielitos had no high school degree. Based on the 1985 CPS, Card (1990) states that 56.5% of the Mariel Cubans in the labor force had no high school education. Based on various sources Borjas (2015) estimates a percentage of high school dropouts between 55 and 65% among the Mariel Cubans. Using the Census 1990 we calculate and show in Table 1 that 54% of Mariel Cubans had no high school degree. Relative to the pre-existing Miami labor force, where this percentage was only 28.5% as of 1980, this inflow constituted a significant relative increase of high school dropouts. The proportion of high school dropouts among Marielitos that we calculated on the 1990 Census population are quite close to those reported by Borjas (2015) in Table 2, who shows 58% of them with no high school degree.

Overall, according to these numbers, the Mariel boatlift produced a 15 to 18% increase in the number of high school dropouts in the Miami labor market, while for the other education groups the increase was only between 3 and 6%. This is a significant increase and we will focus on its impact on less educated (high school dropouts) in Miami, who, everything else constant, were more likely to experience labor market competition from

[2] Card (1990) reports that 54% of Mariel Cubans were still in Miami as of 1985.
[3] The share of Cubans in the Miami Labor force follows a very similar pattern.

the Mariel Cubans. Both Card (1990) and Borjas (2015) are very clear that the group potentially exposed to competition is the one of less educated. They analyze the effect either on the lowest quartile of the wage distribution (Card 1990) or on the group of high school dropouts (Borjas 2015) to establish the potential wage effects of the boatlift.

Table 1 summarizes the main aggregate numbers and demographic characteristics for the Mariel Cubans (that we have been already discussing) and for the labor force in Miami as of the 1980 census, hence right before the boatlift. In the first column we show the data for the total Miami labor force in 1980, in the second column we report the data on all Mariel Cubans, identified in the Census 1990 as Cubans who arrived in the US in 1980 and were 16 or older at arrival. The third column shows the number and demographics for Mariel Cubans from Census 1990 who were still living in Miami as of 1990. We can think of the total number shown in column 2 as the upper bound, while the numbers in column 3 are the more likely values for the inflow of Mariel Cubans into Miami in 1980.

We discussed already the educational distribution of Mariel Cubans. If we look at other demographic characteristics, namely gender and share of young individuals, we see that those are similar in the Mariel population and in the 1980 Miami labor force. Even when we focus on high school dropouts we see similar proportions of females (43% in Miami 1980, 40% among the Mariel Cubans) and individuals in the age range 16-25 (22% in Miami 1980 and 16% among Mariel Cubans). Hence, the change in supply for high school dropouts due to Mariel Cubans seems reasonably balanced between genders and age groups. It seems totally reasonable to pool all high school dropouts, male and female in the age range 16 to 61 (as done in Card 1990) to reduce the measurement error when we estimate the average effects on wages.

2.2 Sample Size

Our preferred sample focuses on the group potentially most affected by immigrants competition: the high school dropouts in the 16-61 age range. As we want to identify the impact on people who were already in Miami we exclude those who self-identified as Cubans (from the question on Hispanic origin in the CPS). In order to use the largest possible sample we prefer the May (Pre-1979) and then the ORG (since 1979) CPS sample. The March CPS sample includes only between one third and a half of the observations relative to the ORG sample and, since 1979, the sample for Miami becomes quite small. We will use the March sample as a check and, in section 5, when we discuss Borjas (2015) for which this smaller sample is the one preferred, we will point out its limitations. The sample used to construct wages includes only non self-employed individuals in the labor force with non missing earnings and with positive weight. These choices are very standard in the literature analyzing the effects of immigrations and they are analogous, for instance, to those used in Borjas 2003 or in Ottaviano and Peri 2012.

Table 2 shows the number of observations used to construct the Miami average logarithm of wage or unemployment for the analyzed group. We see that the March-CPS (first column) only includes 60-80 observations for Miami in each year for our sample. This could certainly raise concerns of significant measurement error. The May-ORG CPS is relatively small between 1973 and 1978 (comparable with the March CPS), but beginning in 1979 it has usually more than 150 observations. While still not too large, these numbers of observations are closer to comfort. In the last two columns of Table 2 we show the size of the samples used to calculate the Miami average wages for high school dropouts in Borjas (2015). Those number are so much smaller than ours because Borjas drops from the sample of high school dropouts three large sub-groups. First, all non-Cuban Hispanics are excluded. They constitute about one fourth of the sample and there is no explanation at all for their exclusion. They should be among the workers more strongly exposed to Mariel competition, as Hispanics should fill labor market niches similar to Cubans. Omitting them seems very problematic. Second, Borjas reduces the age range to 25-59 omitting another fourth of the labor force either young (16-24) or older (60-61). We have already described how Mariel Cubans had a share of individuals under 25 similar to that of the labor force in Miami. Moreover, economic theory suggests that young workers, more vulnerable and with less stability of employment, could be strongly affected by the competition of the Mariel Cubans. Hence, these two groups excluded by Borjas (2015) are as subject to Mariel competition as, and possibly more than, those in Borjas' sample and they should certainly be included. Finally, as already noted by Roodman (2015), Borjas excludes women. While there are reasons to believe that women labor market outcomes can be different from men's, when it comes to attachment and level of wages, this is not ground to exclude them here. The proportion of Mariel Cubans who were women among high school dropouts was high, and there is no reason to believe that women's response to immigrants competition was different from men's. Hence also their exclusion is problematic.

Most importantly, as Borjas (2015) sample ends up with only very few observations in the group of high school dropouts in Miami, it seems that the need to have a minimally representative sample prevails on all those restrictions. As a result of his selection in the March CPS sample, Borjas (2015) only includes between 17 and 24 observations per year in the period 1979-89. This number is incredibly small and it will produce very substantial measurement error in any statistic calculated for Miami. The restrictions also induce very small samples in those cities used by Borjas as controls. As we will see in section 5 most cities in Borjas' control group (e.g. San Jose and Anaheim) have samples with number of observations smaller than 20 and often in the single digits for several years. What in Borjas (2015) is called the average wage for (non-immigrants) dropouts in Miami, which is a group that includes about 120,000 individuals as of 1980 (our calculations from Table 1) is calculated in that paper using 17 to 24 individuals! The ORG CPS sample affords to the restricted Borjas (2015) sample between 26 and 72 observations per year in the same period. This is still somewhat small for comfort but certainly better. We will come back to the great importance of sample selection in section 5 and on the size of measurement error in section 6. Important for our method is also the fact that both the March and ORG-CPS allow us to identify the same labor market outcomes for 43 other metropolitan areas in the US, between the year 1972 and 1991, so that we can include all those in the group of potential controls (the so-called donor pool), when applying the synthetic control method.

2.3 Time Profile of the Labor Supply Shift

In order to validate the historical account of the Mariel Boatlift in Miami and to identify its "mark" in the labor supply, we look at the increase in the share of population of Cuban origin between the pre-Boatlift and the post-Boatlift period in the March-CPS and ORG-CPS. To be very precise with the timing of our data relative to the event, which all historical sources put between April and September of 1980 (with the overwhelming majority of arrivals in May and June), we produce figures with a common and clear time-convention that we describe below.

Identifying exactly the conditions (on labor market outcomes) right before the boatlift, and right after it, helps to maximize the chance of identifying the largest possible short-run effect. Remember that we are considering a one-time, exogenous, unexpected shift in supply, not a persistent event or a policy change. Hence, the largest effect on wages and on employment should be during and right after it. The adjustment dynamics, then, would determine its persistence after 1981, however the bulk of the effects should be detected in 1980 and 1981. We adopt the convention in all figures, to call "1979-Pre" the data relative to the last period before, and as close as possible to, the Mariel boatlift. This is usually year 1979[4]. We call "1980" the data relative to the period during the Boatlift in which already some effects can be seen and we call "1981-Post" the data after the whole shock unraveled. We also take the convention, in each figure, of showing a vertical bar exactly on the *last "pre-period", right before the shock* (hence on "1979-Pre"). This convention departs somewhat from that of showing a bar on the first "treatment" period and helps to visually identify the last period of the status quo, right before the one-time shock. To the immediate right of the bar we can see the impact of the sudden shock and we expect this to be the strongest spike we observe. To to the left of the bar we can see the trend and variation during the pre-treatment period.

Following the described notation we show in Figure 1 the Cubans in Miami as percentage of the total population between 16 and 61 (dark lines) and as percentage of high school dropouts, age 16-61 (light lines), between 1973 and 1985, hence from seven years before to five years after the Boatlift. In Panel A of Figure 1, we show the shares obtained from the March CPS. In Panel B we show the shares obtained from the May-ORG CPS. Notice that for the March CPS, as the data on demographics are relative to the month of March, the last Pre-treatment observation is the one collected in March 1980, and it is called ("1979- Pre"), and it is differentiated from the 1979 (March 1979). The 1981-POST is the observation for March 1981, while the "1980-shock" is simply the interpolation of 1979 and 1981[5]. For Panel B, the year corresponds to when the ORG survey (and before 1978 the May survey) was done. Paying attention to the pre-post details around 1980 allows us to align our data precisely around the Miami Boatlift shock. As with small samples such as those in the May and March-CPS, the noise can be substantial, identifying a clear jump upwards of the Cuban share from " Pre" to "Post" would be the "mark" of the Boatlift on the CPS data.

From Figure 1 three important facts emerge. First, the two surveys (March and May-ORG) show values generally close to each other (especially for the overall population shares and for the post-1979 shares) and they both show a high point at 1981, a sig-

[4] For wage and unemployment in the March CPS we use data collected in 1980, which is relative to the previous year. For the ORG-CPS we use data collected in year 1979.

[5] This is done only for this graph, due to the timing of the March CPS enumeration, that in 1980 was just before the Boatlift. The wage and employment data, however, are relative to the whole year and will be attributed to the relevant year in the other graphs.

nificant increase between "1979-Pre" and "1981-Post" and a decline after the 1981 high point. We see, also, however, significant noise, especially before 1980, when the samples are rather small. Notice an isolated spike in the share of Cubans among dropouts in 1978, only shown in the May-ORG-CPS. This is likely measurement error due to the fact that in 1978 the sample for Miami in the May-CPS survey was the smallest for the whole considered period. We will have to live throughout the analysis with these type of measurement errors. Second, putting some more weight on the more precise population shares, while the pre-1980 share of Cubans might have been growing, we observe a discernible increase between 1979-Pre and 1981-Post. Omitting the likely spurious spike in 1978 the "79-Pre" to "81-Post" increase is the most noticeable increase in the 1972-1985 period. Considering the March-CPS figures, the increase as percentage of the population equals about 6 points and as percentage of high school dropouts the increase is around 11 points. These percentages are roughly consistent with those of Mariel Cubans obtained from the Census and described in section 2.1 above. The third fact shown quite consistently in the two graphs is that, after the increase in the share of Cubans between 1979 and 1981, in the following 4 years that share decreased according to all the samples and especially as share of high school dropouts[6]. This could be because some of them left the city or because more non-Cubans arrived. The fact is that in 1985 Cubans as share of high school dropouts seem to be back at percentages comparable to those of 1979. This emphasizes strongly the temporary nature of the shock that happened suddenly in 1980 and reversed so as to be completely indiscernible in the CPS data *even in its effects on the Cuban presence in Miami* by 1985.

So Figure 1 suggests that the Boatlift was a sizeable temporary shock to the labor force of Miami. Our representation "stretches" the shock from 1979 to 1981, as we do not have more precise data, but if we think that all of that change took place essentially in 5 months of 1980 (and most of it in 2), the jump would appear even more significant. In a perspective that spans 1972 to 1985, however, the shock does not look formidable, relative to the year-to-year variations before and after 1980, some of which is due to measurement error.

One very important corollary is the following. For the "shock" identified in Figure 1, the canonical model predicts the strongest impact on Miami labor market between

[6] The sudden 1979-81 increase, the peak reached in 1981 and the following decline are also features of the Cuban share of the *labor force* in Miami (rather than population, as represented here). The full inflow of Mariel people into the Miami labor market happened in the 1979-81 period.

1979 and 1981, with not much residual impact as of 1984-85. If adjustment of capital/investment took place at all, the dynamic adjustment should ensue after 1981 and wages and employment would return to pre-shock level, certainly as of 1985, when even the population shock has disappeared. The profile we expect from the canonical model for wages of high school dropouts, relative to their trajectory without the boatlift, therefore, is as we represent in Figure A1 of the Appendix. The deepest depressive effect is shown between 1979 and 1981 and after 1981 dynamic adjustment should set in, if nothing else because the share of Cubans returned to the pre-1979 level as of 1985. We can therefore consider the behavior of Miami wages and employment of high school dropouts, in comparison to a group of control cities between 1979 and 1981 as a way to test these implications of the canonical model.

3 Synthetic Control Method

The main empirical method used in this paper is the synthetic control method, first introduced by Abadie and Gardeazabal (2003) and then perfected in Abadie et al (2010). That method provides a systematic way to analyze the impact of an event in case-studies such as the Mariel Boatlift, in which one unit, in this case one metropolitan area, experiences an event ("treatment" or "intervention") while others do not. In order to evaluate whether such event had an impact on some specific outcomes in the treated unit, relative to what that outcome would have been in absence of the treatment, the method allows us to identify a reasonable control group and hence a control outcome for this group (called the synthetic control) and to conduct a difference in difference analysis comparing the treated unit and the synthetic control. To do this in our case, consider $J+1$ metropolitan areas indexed by $j = 0, 1, 2...J$ [7] and denote Miami as 0. Then define a vector G_0 of dimension $k \times 1$ whose elements are equal to the values of variables that help predict wages of high school dropouts (such as the share of high school dropouts in the labor force, the share of Hispanic and the share of manufacturing workers) including the values of the wage variable itself for the city of Miami, in the year from 1972 to 1979, before the treatment begins. Then we define, similarly, a $k \times J$ Matrix, G_J in which row j is the sequence of values for the same variables and years described above, relative to city j in the "donor pool".

The Synthetic control method identifies the weights that produce a convex combination

[7] In our case $J = 43$ as we have 43 metropolitan areas in our sample, plus Miami.

of variables in cities in the donor pool so as to approximate as close as possible, in terms of quadratic error, the pre-treatment vector of variables chosen for metro area 0, in this case Miami.

Namely, define a $J \times 1$ vector of weights $W = (w_1, w_{2,...} w_J)$ that sum to one ($\sum_{j=1}^{J} w_j = 1$). This method will select those weights so as to minimize the difference between $G_0 - G_J W$:

$$W^* = \arg\min(G_0 - G_J W)' V (G_0 - G_J W) \text{ s.to} \sum_{j=1}^{J} w_j = 1, w_j \geq 0 \quad (1)$$

In (1) V is a $k \times k$ diagonal, positive definite Matrix that determines the weight for the contribution of each element of the vector in the objective function[8]. Once we have identified W^* from this "distance minimization" of the pre-treatment variables we can use those weights (one per city in the donor pool) to calculate the post-treatment outcome variables for the "synthetic control". Comparing the pre-post 1979 change in the outcome variable for Miami, relative to the pre-post change for the synthetic control is the basis to evaluate if the treatment has had any effect on Miami, relative to the synthetic control.

As there is some discretion in choosing what variables to match in the pre-treatment period, it is also important to validate the choice of the control group, and to check if it is reasonable. To do so we can check the pre-intervention (1972-1979) levels and trends of the outcome variable to see how closely the treated unit and the synthetic control group track each other before the event. This is prima-facie evidence that we have chosen a control group in which the wage of high school dropouts (and other variables) did not systematically differ from that in Miami in the pre-1980 period. A clear divergence and large differences in the pre-treatment trend of the treated and synthetic control would cast doubts on the validity of the chosen group as control.

As more formal test we can also check in a regression environment whether the pre-1979 and post-1979 difference between Miami and the synthetic control are significant. We will perform these analyses in the next section and discuss our validation strategy. First, however, let us describe what variables we included in the 1972-1979 minimization and what are the cities, from the donor pool, that get positive weights as members of the synthetic control. Table 3 shows this information for three different outcomes in our analysis and in comparison with the control group chosen by Card (1990) and Borjas

[8] In our estimation we use STATA's default option for the matrix V which is chosen among all diagonal and positive definite matrices to minimize the average squared prediction error of the outcome variable during the pre-shock period.

(2015).

Few important things are to be noted. First, we tried several different combinations of pre-treatment variables to be included in the pre-treatment distance minimization. We have finally selected variables that capture important features of the low skilled labor market, namely the share of dropouts, the share of Hispanics and the share of manufacturing workers in the labor force, besides also including the outcome variable for several pre-1979 years[9]. These variables provide a reasonable control group for wages and unemployment rates. We included the 1972-1979 period when Using the March-CPS and the 1973-79 period when using the May-ORG, to allow for a reasonably long pre-treatment. Most studies (including, for instance, Abadie et al 2010, Bohn, Lofstrom and Raphael 2014) use at least ten pre-treatment years and have less noisy data. We do the best we can, given our data constraints, by extending the analysis back to 1972 (or 1973). Before that the number of Metropolitan areas sampled was much smaller and, as a requirement of the Synthetic control method is to have a balanced sample of units from the "donor pool" we are unable to extend further. Using these variables the synthetic control group has three to four cities with non-zero weight. Philadelphia, Tampa, Birmingham and Anaheim are included in the control group for weekly wages from the May-ORG sample. Los Angeles, Dallas and New York in the Log wage control for the March sample. Each group share one city with the Card (1990) control (made of Atlanta, Houston, Los Angeles and Tampa) but it also includes other cities. We will check in the next section how well the synthetic controls match the pre-trend behavior of the outcome variable. In the last column of the table we show the approach by Borjas (2015) who chooses a synthetic control to minimize the distance for employment growth and employment growth of high school dropouts for the time interval between 1977-78 (pooled) and 1979-80 (pooled). In the employment placebo Borjas only chooses cities with similar employment growth in the 77-78 to 79-80 period. He ends up with a different set of cities, including mainly smaller cities such as Kansas City, Anaheim and Sacramento. Borjas does not perform a systematic validation to check that outcome variables for Miami and the control group are similar before 1979[10].

[9] In order to get closer to the choice in Borjas 2015 we have also selected the synthetic sample including employment growth rates and employment growth rates for high school dropouts which are the variables he uses. Synthetic control analysis for that case is shown in Appendix Figure A3 and it is similar to the Preferred case.

[10] In the figures 2-5 reported in Borjas (2015) the values and trends of wages for Miami and for the Control group in 77-79 seems usually quite different.

4 Empirical Estimates

4.1 Main Results

We begin with showing, in Figure 2, the main results from the Synthetic control method, using the preferred sample of non-Cuban high school dropouts, not self-employed in the labor force aged 16 to 61. The four panels of this figure tell the basic story that will be then confirmed, time and again, in the robustness checks. Panel A shows log weekly wages (gross wage and salary) for our preferred May-ORG sample for Miami (solid line) and for the "Synthetic Miami" which is the control group (dashed line). Panel C, below A, shows the same figure for log hourly earning as outcome. These are calculated dividing the weekly earning by hours worked per week. One may argue that this measure is closer to capturing the marginal productivity (hence price) of labor. The two panels to the right, B and D, show similar variables and the same definition of workers, but they use the March CPS sample that, as we documented in Table 2, is significantly smaller than the ORG one from 1979 (one third to one half of its size). Panel B shows the behavior of log weekly wages for Miami and Synthetic Miami, from the March CPS, and Panel D shows the behavior of log yearly wages, which is the variable directly measured in the CPS (while the weekly wages are constructed by dividing the yearly wages by the number of weeks reported).

Let us first comment on the pre-1980 time path for the considered wage outcomes in Miami and in the synthetic control. A close correspondence is needed to consider the synthetic control as a good placebo group. Overall the synthetic control does a reasonably good job in tracking the dropout wages and matching the 1972-1979 trend in each of the Panels. Certainly one can see noise and deviations between Miami and Synthetic control in the year-to-year variations before 1979. Deviations are in the order of 0.01-0.04 logarithmic points. This implies that such level of noise could make it hard to identify deviations of average wages between Miami and the Control in the order of 1 to 4%. Nevertheless, in spite of the noise, we should be able to see if some especially large wage differences (as large as 10% for instance) suddenly arise between Miami and its control in the immediate aftermath of the Cuban inflow between 1979 and 1981 (anything larger than 5-6 percent would be noticeable). The second important thing to notice is that, with some year-to-year noise, all graphs show a clear long-term downward trend for wages of high school dropouts in Miami and in the synthetic control since 1972 (and perhaps earlier, but we do not go further back) all the way to 1991. Matching and showing clearly this pre-existing time-trend both in Miami and its synthetic control is crucial to claim that we have identified a good control group. Remember that the 1980's were a period of large increase in wage inequality and poor performance of the wages of unskilled workers. Hence identifying these features as common to Miami and control and not as specific to Miami is important. The downward wage trend, both for the synthetic control group and for Miami does not seem to have any break or jump between 1979 and 1981. Nor there is any significant downward deviation of Miami relative to Synthetic control, arising between 1979 and 1981. If anything in the Panel A and C (from the ORG sample) the Miami wage rises from slightly below to slightly above its value in the "synthetic control" between 1979 and 1981, while in the other two panels (B and D) it seems to track closely the synthetic control. In fact inspecting the whole post-1979 period there seem to be only two instances in which the difference between Miami and synthetic control is not completely negligible. The first is a deviation of Miami *above* its synthetic control for weekly wages in the ORG sample, in 1984-85 and then 87-89 (Panel A). The second is a deviation of Miami *below* its synthetic control in 1984-87 for weekly wages in the CPS sample. The fact that the same variable measured on exactly the same sample in two different surveys (March and ORG) reveals such different behavior certainly confirms our worries about measurement error and noise. However the fact that such deviations occurs four or more years after the Boatlift, while no significant deviation occurred during it, implies, in our view, that they have absolutely nothing to do with that episode. Panel D (for yearly wages from March CPS) is the more noisy measure. Even this figure, however, shows a consistent trend down of Miami and Synthetic control wages, no significant deviation starting in 1979-1981 and hence no hint of the "canonical effects" of the Mariel Boatlift on the Miami labor market.

One more check that we perform, in order to use the larger sample (CPS-ORG) but to reduce the potential confounding effect of age, gender and ethnic heterogeneity across cities is to first "adjust" the average log wages by regressing those in the whole sample on five years age-dummies and on hispanic-by-year and female-by-year dummies. Then we take the residuals, average them by metropolitan areas and we implement the synthetic control method on those minimizing the distance between Miami and control 1972-79 for the same variables listed in Table 3, plus log wage residual. The resulting graph is in Figure A2 in the Appendix. It shows significant noise before 1979, but also an upward departure of Miami from control in the 1979-1982 period (if we align the 1979 observations). The post 1979 Miami-Control departures do not appear in any way unusually large or negative.

To summarize the findings of this section, vis-a-vis the predictions of the canonical model, the sudden nature of the shock would imply maximal deviation of Miami's high school dropout wages from the Synthetic control in the 1979-81 period. After that, transitional dynamics should have reduced the effect so that adjustment would have been almost full when the share of Cubans in Miami (shown in Figure 1) was back to pre-1979 levels, as of 1985. Instead what we observe for all samples and wage measures is no deviation of Miami from control in 1979-81, and also no deviation if we consider the longer period 1979-1983. In none of the Panels of Figure 2 do we observe a systematic large deviation and subsequent adjustment between Miami and synthetic control beginning in the 1979-1981 period. The only noticeable departure between Miami and synthetic control arises in 1984, when the share of Cuban among high school dropouts in Miami is almost at its pre-1979 levels. None of these facts conform with the predictions that a canonical model would have for the labor market in response to the Mariel boatlift.

4.2 Other Labor Market Outcomes

Some papers on the effect of immigration of native labor markets have defined potential competing workers in the receiving market not in terms of education but in terms of their wage potential. For instance Dustmann, Frattini and Preston (2013) argue that the strongest labor market competition from immigrants is on the group of workers who are in the same part of the wage distribution as the immigrants themselves and not necessarily on natives with the same level of schooling. Also Card (1990) defined the workers more likely to compete with immigrants based on their position along the wage distribution (bottom quartile) based on estimated returns to observable characteristics. In Panel A and B of Figure 3 we show the synthetic control results considering as outcome the wage at the 15th (Panel A) or at the 20th (Panel B) percentile of the distribution and its time evolution in Miami and in the Synthetic control between 1972 and 1991[11].

The sample we choose, consistently with the previous analysis, includes all non-Cuban individuals between 16 and 61 years of age, not self employed and in the labor force. We use the larger and preferred sample, the May-ORG CPS, to analyze these outcomes. One advantage in choosing the wage percentile, rather than the average wage of a small group

[11] The variables used to minimize distance in the pre-1980 period and produce the synthetic control are the share of hispanic, the share of dropouts and the share of manufacturing workers in the labor force.

(such as the high school dropouts), is that the sample used is larger (the whole non-Cuban labor force) in Miami and this statistic should be less sensitive to extreme values of wages in the city and hence less volatile. This is possibly reflected in a somewhat smaller year-to-year volatility, observed in Panel A and B (especially before 1980) relative to the Panels of Figure 2 and in an slightly improved match of the pre-1980 trend between Miami and the Synthetic control. In this case the during the 1979-1981 period both Panel A and B show a relative *increase* in the Miami wage relative to the Synthetic Miami, with small differences between Miami and control also in all other years after 1981. This is true both for the wage at the 15th percentile (shown in Panel A) as well as for that at the 20th percentile (shown in Panel B).

One explanation for the small effects observed is that wages in Miami were rigid downward in the years 1979-1981 and hence a negative demand shock for native workers did not translate into lower wages. Or it may be that the decrease in demand for native workers did not materialize because the Mariel Cubans displaced one for one native workers in the short run. In both these cases the effects on wages could be attenuated. However, if the inflow of Mariel Cubans decreased demand for other workers, one would observe an increase in the unemployment rate of the group of non-Cuban high school dropouts in Miami. Panel C and D shows the unemployment rate (number of weeks unemployed during the year as fraction of total weeks) of the non-Cuban high school dropouts 16-61 for Miami and for the Synthetic control. Panel C uses the preferred May-ORG CPS sample and Panel D uses the March-CPS. First, let us notice that the year-to-year volatility of unemployment in Miami before 1979 was quite large. In particular, both figures show a spike in unemployment Miami in 1975 that is not matched by the synthetic control. This should make us cautious, as other factors differentiating Miami from its control, existed in the pre-1979 period. With this caveat in mind we observe that the 1979-81 behavior of Miami unemployment rate relative to the Synthetic control does not show any clear relative departure both in the May-ORG and in the March sample. Even in 1982 and 1983 no significant difference between the unemployment rate in Miami and Synthetic control arises. The lines cross each other several times after 1980 confirming some volatility but no systematic deviation in 1979-1981 or later.

4.3 Subsamples

As we discussed in Section 2, there is the risk of introducing very substantial noise by restricting the sample too much. However, using the May-ORG sample, which is somewhat larger and counts up to 200 observations per year in Miami after 1979, we will consider two subgroups of the Miami high school dropouts. First, while mariel Cubans were divided between men and women roughly in the same proportion as the pre-existing labor force, one may think that their impact was different by gender. This may be the case if Mariel Cubans mainly took jobs in competition with the male labor force.[12] Hence, we show the synthetic control analysis, restricting our preferred sample to only males. Second, one may think that the non-Cuban Hispanics in Miami (included in our sample, as they were not the Marielitos) were also likely to be immigrants and they should not be considered in the impact of Mariel Cubans on natives. On the other hand the US-born non-Cuban Hispanics were likely the workers in closer competition with the Marielitos, so including them should maximize the chances of finding some competition effects. We exclude this group in Panel B.

Figure 4 shows the logarithm of weekly wages in the two sub-samples for Miami and Synthetic control, between 1973 and 1991. In Panel A we only include males, (non-Cuban high school dropouts, age 16-61). In Panel B we include non-Hispanic male and females (high school dropouts, age 16-61). The non-Hispanic only (male and female) sample shows a good pre-1979 fit and a somewhat persistent *positive* deviation of Miami from the Synthetic control beginning right in 1980. One may be tempted to argue that a *positive temporary* effect of Mariel Cubans on native non-Hispanic existed, based on this case. In this sample a divergence suddenly appears between 1979 and 1981, it persists for 4 years and then disappears by 1986 (roughly when the Cuban share of dropouts in Miami is back to trend). If one is *really* determined to find an effect that matches the temporary characteristics of a shock and the timing of the Mariel Boatlift, this could be it. Except that it is positive! As this is an isolated case, not confirmed in other subsamples, and we are aware of the measurement error present in these samples, we rather consider it as by-product of noise. Also, as we will argue below, this effect is the opposite of what appears in the March male non-Hispanic sample, and it does not appear in any other

[12] Borjas (2015) argues that women should be left out of the sample as their labor force participation increased in the 1980's. We fail to see why this should matter unless it happened differentially between Miami and the control cities and it had a discontinuity in 1980. Neither of those is likely. Still, we show what happens for males only.

sample, strengthening our idea that it is mainly noise. The male only sample on the other hand, shown in Panel A, is more noisy in tracking the pre-1980 trend, but it shows a really tight correspondence of Miami and the Synthetic control between 1979 and 1983. Absolutely no deviation appears in those years. Some divergence arising after that also seems to be noise.

4.4 College Educated

The results presented so far suggest that some noise likely produced by the measurement error exists, but there is no evidence of a significant difference in wages and unemployment of high school dropouts between Miami and its control group arising in 1980 and 1981 (nor later, in most cases). Let's now consider a falsification of our exercise. We choose a group that should have experienced no significant competition with the Mariel Cubans, that of college educated workers in Miami. The Mariel boatlift brought very few college educated to Miami, while, as argued above, it brought a large number of high school dropouts. College educated, therefore, might have benefitted from complementarity with less educated or, if technology and production was adjusted as to absorb low skilled (as argued for instance in Lewis, 2013), they may have not experienced any effect at all. Certainly, they should not have experienced a wage decline from the Boatlift, according to the canonical model, as they are not highly substitutable with high school dropouts. On the other hand, college educated workers could be subject to some of the other local productivity shocks and business cycle forces present in Miami and affecting also high school dropout wages. Hence, analyzing college educated in Miami versus the synthetic control is a way of checking if the deviations for this group look very different from those of high school dropouts after 1979 indicating that something might have generated a relative gap. Figure 5 shows the log weekly wage for non-Cuban college educated workers in Miami and in the synthetic control. The pre-trend matching is reasonable, with some measurement error. After 1980 the synthetic control matches remarkably well Miami until 1982. Then there is a negative departure of Miami college-educated wages from the synthetic control for the period 1983-1986. Interestingly, this is the period in which some of the high school dropouts graphs (e.g. Figure 2 Panel B) also show a negative deviation of Miami wages. While this dip appearing in some graphs has certainly little to do with the Boatlift (as it affects college educated and takes place three years after it), Figure 5 is a bit of a cautionary note on the fact that common unobserved labor market shocks in Miami, that affected more and less educated, can play a role after 1984. It would be unwise to attribute deviations arising after 1983 to consequences of the Mariel inflow.

4.5 Regression Analysis

There are two types of uncertainty that affect inference with the synthetic control method and it is not immediately clear how to produce standard errors and confidence intervals that account for both. The first type is due to measurement error, stemming from the fact that we are measuring the average outcome (wage, unemployment) with error in Miami and in the Control group. This type of error would be eliminated if we had data on every single worker in Miami (and in the control cities) to construct the wages. In our case, given the small size of the sample, this error can be large. We can assume that each year we observe a different realization of this measurement error, in the average Miami outcome, that has a certain distribution and hence we can use observations over many years to estimate the uncertainty on the difference in mean values between Miami and Synthetic control, accounting for this error in a regression framework. A second type of uncertainty is due to the fact that we do not think that even if we could measure the average outcome exactly, the pre-post 1980 differences in wages would determine without uncertainty the impact of the Mariel Boatlift. Several unobserved factors, potentially affecting Miami differently from the control could still generate variation. In order to deal with this second type of uncertainty we produce a simulated test of the significance of the difference between the post-1980 outcome in Miami relative to its synthetic control vis-a-vis the distribution of that statistics for all other cities in the sample. We will address this issue in section 4.6 below.

Let's consider the following regression that includes observations for Miami and its synthetic control between 1972 and 1991 and will provide coefficients akin to those produced in an event study. It will also be a way to validate the choice of the comparison group in the Synthetic control method.

$$\begin{aligned} y_{it} &= Miami_i + \sum_{P \in PRE-79} \alpha_P D_P + \sum_{P \in POST-79} \alpha_P D_P + \\ &+ \sum_{P \in PRE-79} \beta_P (D_P * Miami_i) + \sum_{P \in POST-79} \beta_P (D_P * Miami_i) + \varepsilon_{it} \end{aligned} \tag{2}$$

In (2) the variable y_{it} is the outcome of interest (e.g. average log of weekly wages of high school dropouts) in unit i which can take only two values, either "Miami" or its "Synthetic control" and in year t, between 1972 and 1991. The variable $Miami_i$ is a dummy equal to one for Miami and 0 for the Synthetic control. D_P is a set of 3-year dummies that span the whole period but omit 1979, which is absorbed in the constant and hence serves as reference year, right before the shock took place. In the pre-1979 period the dummies are D_{72-75} and D_{76-78} in the post-1979 period they are $D_{80-82}, D_{83-85}, D_{86-88}$ and D_{89-91} and they equal one in the years indicated in the subscript and 0 otherwise. α_P is the set of coefficients corresponding to the period dummies and β_P is the set of coefficients associated to the interaction between the dummy $Miami_i$ and the period dummies. The term ε_{it} captures a classical error term, uncorrelated with the observables and with 0 average that we interpret as residual measurement error for each metropolitan area. The coefficients of interest are β_P. In particular, if the Mariel shock had any effect, it should be the largest in the years during and right after it, and this should be captured by the estimates of β_{80-82}. That coefficient captures the average difference between Miami and Synthetic control arising in 1980, 1981 and 1982 once the 1979 difference is standardized to 0. The subsequent coefficients $\beta_{83-85}, \beta_{86-88}$ and β_{89-91} are there to complete the picture. Several economic shocks (including an important recession in 1982) took place during the decade and Miami could have responded differently from the control, hence the farther we go from 1979 the more likely it is that other factors affected the difference between Miami and its synthetic control. Let's be very clear on the timing. If we identify a large effect on labor market outcomes in $1980, 1981$ and 1982 (a significant value for β_{80-82}), then one could take that as evidence of an impact of the Boatlift. Some persistence could be expected to still linger in the period $83-85$ and possibly even later (see the graph in Appendix figure A1, showing the impact effect in 79-81 and then adjustment). However if no effect is observed from 1980 to 1982 (and even more precisely in $79-81$), then it is certainly implausible and not in line with the canonical model to attribute later divergence in labor market conditions between Miami and control to the effects of the Boatlift.

Our framework allows us also to estimate another set of coefficients. They are just as important as the post-shock ones, and they are the values for the pre-1979 differences between Miami and its synthetic control. The estimates of β_{73-75} and β_{76-78}, are a more formal validation of the synthetic control, quantifying the significance of the deviations between Miami and synthetic control, pre-1979. In particular, by testing these differences in 2 periods of time before 1979, we can check if there were systematic deviations or a pre-1979 differential trend. Positive and significant pre-1979 differences would cast doubts on our control group as they will imply systematic deviations between the Miami and control, even before the treatment (Boatlift). Ultimately, these estimates should mirror closely the differences between Miami and Control seen in Figure 2 and 3.

Table 4 shows the estimates obtained by non weighted OLS, for the same outcomes and sample specifications that we used in Figure2 (all Panels) and in Figure 3 (Panels A, C and D). Each column shows the coefficients of all the pre- and post 1979 period-Miami interactions. These are the estimates of the differences in outcome between Miami and Synthetic control relative to its value in 1979. The dependent variable analyzed is the log of weekly wages in Column (1) and (2) using the May-ORG CPS in the first case and the March CPS in the second. In column (3) the variable is the log of hourly wages from the May-ORG CPS, in column 4 it is log yearly wages from the march CPS. In column (5) we analyze the effects on the 15th percentile of the wage distribution, and in column (7) and (8) the unemployment rate from the May-ORG or the March CPS respectively. We should keep in mind that each regression is estimated on 38 or 40 observations[13], and hence any result should be taken with a grain of salt. The estimates are simply a quantification of what is represented in figures 2 and 3 with the provision that the regression standardizes to 0 the difference between Miami and Synthetic control as of 1979, while the graph minimizes the whole pre-1980 distance without setting it to 0 in 1979.

Some features of the results are quite clear. First, Miami and Synthetic control move together, to a reasonable extent, in the pre-1980 period so that the coefficients of the pre-period interactions (β_P) are never significant. This is a validation of the control group and an indicator of good fit, although the standard errors for the wage regressions (between 0.02 and 0.07 log points) in Columns 1 to 5. The unemployment regressions (column 6 and 7) are even more noisy but even in this case no pre-1979 interaction is significantly different from 0. Notice that the coefficient for the 15th wage percentile are significantly more precisely estimated. In this case the pre–1980 coefficients are quite tight and essentially equal to 0. Second, and most importantly, the basic result from Table 4 is that there is no significantly negative coefficient in 1980-1982, in any of the variables. Not only all estimates are mostly insignificantly different from 0, their point-estimates are sometimes positive for wages and negative for unemployment. The only significant coefficient in 1980-82 is a positive value for the wage at the 15th percentile of the distribution. Among the point estimates of the 80-82 interactions for the wage outcomes,

[13] The May-ORG CPS does not include year 1972 and has, therefore 38 observations.

two of them are positive (0.045 and 0.045 from columns 1 and 3) one is negative -0.038 (in column 2) and one is essentially 0 (0.002 in Column 4). None of them is significant. Similarly, the 80-82 interaction coefficients for unemployment are one positive (column 7) and one negative (column 6). Let us emphasize once more that, as the Mariel inflow happened suddenly between May and September 1980, the strongest impact should have been in 1980 and 1981, possibly still persisting in 1982. As we consider outcomes three or more years after the shock, a series of other events may have produced divergence between Miami and its synthetic control. After 1985 even the share of Cuban among Miami high school graduates was back to its pre-1979 level (see Figure 1 Panel B). The results shown above of no impact in the 1980-1982 period seem to reveal no effect of the Mariel boatlift on labor market outcomes for natives in Miami.

Let's then consider the estimates of the coefficients for 1983-85 interaction coefficient. We see a few significant and *positive* values, in columns (1), (3) and (5). These deviations are still not far from the range of pre-1979 deviations estimated for this sample. Most importantly, in specification (1) and (3) they become significant and larger only after 1983, as no significant effect was estimated in 80-82. Hence, such deviations have likely nothing to do with the Mariel boatlift. A similarly statistically significant and negative effect arises in 1986-88 for Column (2) and for unemployment in the March CPS (column 7). It would be misleading to connect these deviations to the Boatlift.

4.6 Inference Using Permutations

While the regression approach has its appeal as it applies to the synthetic control analysis some of the checks done in the difference-in-difference and in event-study literature, the small number of time series observations and the imprecision of the estimates limits such exercise. A more common way of doing inference in the synthetic controls method is to simulate a distribution of deviations between each unit included in the donor pool and its synthetic control and see if the unit actually treated shows a post-treatment deviation from its synthetic control that is large, in relation to all the others. To do this we consider all the possible post-1979 differences in a specific outcome (say log of weekly wages) between the unit and its synthetic control for each of the metropolitan areas in the donor pool relative to their pre-1979 unit-control differences. This implies that we estimate a synthetic control for each of the other 43 cities. First, we show in a graph the pre and post-1979 deviation of the unit from its synthetic control. Those deviations for all the permutations of the 43 metropolitan areas identified in the CPS are shown in Figure 7. Panel A uses log weekly wages as outcome, Panel B uses log hourly wages, and Panel C unemployment rates. The sample chosen is always the May-ORG, preferred sample and the dark line corresponds to Miami, while the lighter lines correspond to the other 43 cities. All figures reveal that Miami is a rather average city in the behavior of its high school dropouts wages (weekly and hourly) and unemployment rates vis-a-vis the other cities. Both in the pre-1979 fit and in the post-1979 deviations Miami performs in the middle of the distribution. In particular the Miami-synthetic control deviations post 1979, and specifically in 1980-1982 seem to be well within the distribution of responses for cities that did not receive the treatment. Looking at the pre-1979 fit, we see that the wage samples achieve a reasonably good fit, as many other cities show larger deviations in their pre-1980 trend. The unemployment sample, instead, is rather noisy in the pre-1979 period with large deviations in the early years, revealing a certain volatility of unemployment of high school dropouts in Miami, relative to other cities. Looking at the post-1979 deviations for the more relevant period 1980-82, or even for the whole 1980-1991 period, Miami seems to be well within the sample distribution[14].

To test this formally we do two exercises. First we calculate the Pre-Post difference in the average deviation of Miami from its placebo, considering 1980-82 as the post-period and, alternatively, either the 1972-79 or the more recent 1977-79 interval as pre-period. Notice that by taking the deviation of Miami from the synthetic control after the Boatlift (1980-82) we are considering if any significant deviation arises, and subtracting the pre-1979 average deviation we "clean" this value of the idiosyncratic deviation generated by pre-1979 existing factors. We then do the same for all other cities in the sample of 44 and we produce the same statistics. In Table 5 we show the value of this Difference-in-Difference statistics calculated using 72-79 as " Pre" period (upper panel of figure 5), or 77-79 as " Pre" period (lower panel of figure 5). We also show the rank of Miami statistics in the distribution of 44 cities (1 being the lowest value and 44 the highest) and the p-value of a one-sided test that uses the distribution of these statistics for 44 cities, for the probability of a city in the distribution having a statistics smaller (in column 1 or 2) or larger (in column 3) than Miami. A very low value of the rank (1 or 2) and a value

[14] In Figure A4, in the Appendix we show the distribution of average 1980-1982 deviations of all cities from their synthetic control, for log weekly wages (Panel A), hourly wages (Panel B) and unemployment (Panel C). Miami is the vertical line and we report the p-value for a city deviation being smaller (for weekly and hourly wages) or higher (for unemployment) than Miami.

of the p-statistics of 0.05 or smaller in column (1) or (2) would indicate that high school dropout wages in Miami had an unusual decline in 1980-1982. High values of the rank (and a p-value of 0.05 or smaller) in column (3) would imply a spike in Miami in high school dropout unemployment in 1980-1982, relative to before. What we find, instead, is that Miami is well within the distribution for all variables and statistics. Its rank is between 29 and 36 out of 44 for wages (hence rather large and certainly not a low outlier), with a p-value never smaller than 0.66. For unemployment the ranking is between 16 and 19 with a p-value of at least 0.54. This means that idiosyncratic variability produced by many other factors and by measurement error, and likely to exist in any unit-control pairing, fully explains the post-1979 behavior for Miami as that city does not appear to be an outlier[15].

4.7 Did We Improve on Card (1990)?

Our analysis uses the synthetic control approach which imposes discipline and validation in the choice of a control group and hence make it less "ad-hoc". It also allows to check how good the choice of a control group is. In terms of results, however, our analysis squares very well with the earlier findings of David Card (1990). There is no evidence that Mariel affected labor market outcomes of high school dropouts using three different measures of their wage (weekly, hourly and yearly) and two different samples (May-ORG and March CPS) or restricting the analysis to men only or to non-Hispanic only. Moreover we find no evidence of any effect on log wages of workers at the 15th or 20th percentile of the distribution which are measures of low skilled wages preferred by some and closer to the one chosen by Card (1990). Also we do not find any evidence that the boatlift affected the unemployment rate of high school dropouts. All these results confirm Card (1990).

Table 3 showed that the synthetic control method ends up including for average wages at least one of the cities that Card also choose as control in his study. Hence it is natural to check whether this more sophisticated synthetic control method improved at all on the choice of Card's control. Figure 6 shows the same comparisons, in terms of variables

[15] Alternatively, one can construct the mean prediction square error (MPSE) post-1979 and pre-1979 and take the Post/Pre ratio. A large value of that statsitic relative to the simulated distribution would indicate deviation of Miami from the synthetic control after 1980. The statistic for Miami, its ranking and p-value are presented in Table A1 in the appendix. The upper part of the table uses 72-79 as reference, and the lower part uses 77-79. The statistics has always a quite low value, indicating no significant post-1979 deviation.

and samples as Figure 2, but it substitutes the synthetic control with a control made combining the four cities chosen by Card (1990). We standardize the control to the level of Miami in 1979. The correspondence of Card's control with the pre-1979 Miami trend is a check of the validity of this control group. The departure of Miami outcome from the Card control after 1979 is a measure of the effect of the Mariel Boatlift as detected using Card's control. The visual impression is that the Card (1990) control is less precise in matching the pre-1979 trend relative to the synthetic control used in Figure 2. It certainly misses most of the year-to-year fluctuations before 1979, and as it exhibits an increase in 1978-79, while Miami has a decrease (in all wages) when we align them on 1979 the match is not great. However, it does a reasonable job in matching the downward sloping trend before 1979. One cannot discern any significant decline in the Miami-Control difference in 1979-81 (if anything the ORG samples show some increase of Miami relative to control in 1979-81). Certainly, there is a lot of noise and the imperfect pre-1979 matching translates into imprecise post-1979 matching too. The regression results of Table 6 contribute to illustrate the same facts. They correspond to the coefficients on the period-Miami interactions for regressions (2) estimated on Miami and the Card (1990) control in the same specifications and samples as those described in panels A-D of Figure 6 above using non-weighted OLS. The estimates can be compared to those of the first four columns of Table 4, which use Miami and the synthetic control.

Three facts emerge. First, the Card control does not perform as well in the pre-trend, as there are some significant departures, even in the period right before the Boatlift (1976-78) which signal departure of labor markets outcomes between Miami and its control before 1979. Second, the standard errors (pre- and post trend) are large, in the order of 0.03-0.06 (but not significantly larger than in the Synthetic control regressions). This is another indicator of noise. Third and most importantly, however, even in this case we do not find any significant negative estimate of the interaction for the 1980-1982 period in any specification. Most importantly, the deviation 1980-82 are right in the same range as the deviations before 1979. In one case we find a *positive* and significant deviation (Column 3), but in that case we also find several positive and significant deviations from the control in the pre-1979 period and we cannot take these positive deviations as implying an effect of the Boatlift.

Overall we think that the exercise was worthwhile, as we have brought some more rigor, robustness and replicability to the results in Card (1990). Replicability and standardization in the choice of the control is particularly important and it will allow us to address (in section 7) one issue raised by Angrist and Krueger (1994) related to the risk of finding some effects simply by accident. The essential message of this study, however, is exactly the same as Card (1990): there is no evidence of any significant labor market effect of the Miami Boatlift on less educated natives in Miami up to three years after that event of April-October 1980. After 1982 things are noisy but no clear evidence of a decline in Miami wages relative to control exist either. We think that it would be unreasonable to consider newly arising effects after 1982 as consequence of the Boatlift.

5 Explaining the Results in Borjas (2015)

An important question to address is how to reconcile our results with the very different findings of Borjas (2015). In that paper not only are there very strong claims of negative wage effects from the Mariel Boatlift, but the estimates of these effects are very large. A decline of up to 30% in the wages of high school dropouts is attributed to the Boatlift. This would be the largest effect ever found in the literature, with an elasticity of wages in the order of -2 or -3 (when most estimates in the literature are at most as low as -0.3). The effects are also incredibly long lived, as they last for a whole decade. Panel A of Figure 8 shows exactly Figure 2A in Borjas (2015), that is the essence of his findings, extended backwards to 1972, as we replicated it. The panel shows the high school dropout log wages in Miami and in Borjas' "Employment Control" (those cities whose employment growth in 1977-80 was closest to Miami). We simply "align" Miami and the employment control in 1979, to give a cleaner visual impression of the difference between Miami and control after 1979. The picture conveys a very strong idea of significant, large and protracted divergence starting in 1980. But a few features are surprising. First, the picture, although it uses a very small sample of 17-24 observations per year, has much less year-to-year variation than all the previous we produced. Second, Miami and Control diverge from 1980 in a progression that peaks in 1985 and continues up to 1987, and this is a feature that no other previous figure showed. While one can see that the "employment control" chosen by Borjas does not match very well Miami in the pre-1979 period, being flatter in the 1977-79 period and steeper before, the impression from the graph is that something very major started in 1980 and continued for 7 years affecting negatively Miami dropout wages. We also notice, differently from graphs in Figures 2 and 3, that the high school dropouts wage do not seem to have much of a pre-1979 negative trend in this graph.

Let's now introduce two completely non-controversial modifications to this picture, in the sense that they should actually *increase* the chances of isolating an effect of the Boatlift, if there is one. First, Borjas smooths the time series using a 3-year moving average. This is not a good idea when we rely for identification on the suddenness and exact timing of a temporary shock that happened mainly within 2 months (May-June 1980). By using a moving average we confound data of the pre-shock observation of 1979 with 1980 data. We also include event that took place in 1982, in the "post" observation of 1981, which should capture just the response to the shock. So we will not use the moving average. Borjas claims that this is "to smooth the noise" from the data. Hence, we will introduce a much better way to reduce the noise, which is that of including more individuals in the sample. We do this by adding to the sample individuals who should be even more likely to be affected by the boatlift and are excluded in Borjas sample with minimal explanation. Borjas only includes non-Hispanic prime-age (25-59) males. We add to these Hispanic non-Cuban individuals and we extend the age range to include the more often used working-age period for high school dropouts (16-61), as was done in Card (1990). Hispanic workers were more similar in their jobs and occupations to the newly arrived Cubans, because of their language and skills and hence this broader choice should improve precision and detect a stronger effect. Individuals with weaker job protection and shorter labor market attachment (young) could also be more vulnerable to new immigrant competition. Both extensions together afford us a sample that is more than double the size of the one used by Borjas and they go in the direction of increasing the labor market competition with Mariel people.

If the Mariel event worked as the canonical model predicts, this second figure should show an sharper drop of Miami wages relative to its control exactly between 1979 and 1981, a larger effect due to the stronger competition with Cubans and some adjustment dynamic after that. Instead, we get the figure reported in Panel B. In this figure, we have also included a third modification relative to Panel A, namely the synthetic control is constructed by matching values of 1972-1979 variables as in our previous analysis (not just employment growth in 77-80 as in Borjas). The figure has changed substantially and now it looks significantly more like the panels in Figure 2. First, the downward trend of dropout wages 1972-1986 common to both Miami and Synthetic control, is now visible, albeit with noise. This trend has clearly nothing to do with the Boatlift. Second, the changes in dropout wages in 1979-81, when the shock took place and should have produced the largest wage effects, show no deviation of Miami from its synthetic control. Third, we notice how the synthetic control group matches much better the pre-1979 behavior of Miami than the "employment control" did. In this frame, the fact that we observe no departure between Miami and control up to 1981 is the strongest hint that the Boatlift did not have any significant effect on male high school dropouts[16]. We notice in Panel B that a departure of Miami from the synthetic control arises in 1982 and lasts till 1986. To explore whether this feature may be due to something real happening in Miami in 1982 (not in 1980!) or if it is mainly noise from the March CPS sample, which is pretty small in spite of the extension (ranging between 35 and 59 observations), we take the exact same specification, sample and outcome variable (log weekly earnings) of Panel B and we simply use the larger May-ORG CPS sample. Doing this we obtain figure 8 panel D, showed right below Panel B. The departure of Miami log wages from the control between 1979 and 1983 becomes negligible, and in 1984 and 1985 a *positive* departure appears. This is a sign that the residual departure observed in 1982 for Panel B was likely to be simply measurement error. The larger sample does not show it. Finally, let's take a step back and keep the exact sample definition of the dropouts sample as in Borjas (2015), but let's simply not smooth the data, use the ORG-CPS sample and use the synthetic control matching 1972-79 variables. This is reported in Panel C of Figure 8. Even in this case of minimal difference with Borjas the divergence between Miami and synthetic control between 1979 and 1983 disappears once we align Miami and synthetic control in 1979.

The regressions of Table 7 show the estimated coefficient of the time-period-Miami interactions in regression (2) for the Miami-control specifications as presented in Panel A-D. These estimates illustrate in a slightly more formal way how the findings in Borjas (2015) quickly disappear, and sometimes reverse, when we move from specification of Panel A (in column 1 of Table 7) to Panel B and C (specification 3 and 4). They also show that, albeit with relatively large standard errors, the specifications using the synthetic control pass the validation of no significant pre-1979 coefficients, while the employment control of Borjas (of Panel A) fails it very significantly in both pre-1979 coefficients. Focussing on the most relevant deviation, arising in 1980-82, specification 1 shows a deviation of -0.17 log points, significant and goes on to find a -0.44 log points deviation in

[16] We are still omitting women from this sample. The inclusion of women would produce the graph shown in Panel A of Table 3. We firmly believe that women should be in the sample, and if some worries about their differences in national trends with men exist they should be addressed by regression cleaning before averaging. In this section, however, we omit them as we want to stay as close to Borjas (2015) sample as we can.

1983-85, extremely significant. All the other specifications find no significant coefficients in 1980-82 and sometimes positive and sometimes negative (never significant) coefficients in 1983-85. Clearly, the smoothing, the extremely small size of sample and the choice of a control that is not validated against the pre-1979 trend produce the appearance of very large negative deviations between Miami and control, that do no appear in any other case. Lets' also notice, that in no specification of Figure 8 and Table 7 we have included women. Roodman (2015) has already noted the drastic reduction in observations imposed by Borjas (2015) because of omitting women from the sample. While we agree that women should be included, and doing that produces the results of our Figure 2 and 3, we can see from this section that their inclusion is not crucial in explaining away the negative effects.

6 Measurement Error in March- and ORG-CPS

To show with more confidence that the apparently large deviation of log weekly wages in Miami post-1982 (or post 1979 in the smoothed figure) shown in panel A and B of Figure 8, but completely absent in Panel C and D of that figure, can be due to the variation of measurement error across cities in the March CPS sample relative to the May-ORG sample, we produce a simple estimate of the measurement error in each survey. As for year 1979 we have available wage data to construct city averages from the March CPS data (from 1980) the ORG-CPS data (from 1979) as well as from the Census 1980 we can calculate average log wages for the Borjas (2015) sample definition, of non-Hispanic, male, high school dropouts in the labor force in the age range 25-59, in each of the 41 metropolitan areas available in all three datatsets[17]. Then, we can calculate the difference of average log wage in each metropolitan area, between the March-CPS and the census, and we call this difference "measurement error" in the March-CPS (measured on 1979 data), and do the same between the ORG-CPS and the Census and call this "measurement error" in the ORG-CPS (in 1979 data). This assumes that the Census measure has no (or negligible) measurement error (this is a 5% sample of all the population and in each group it includes thousands of observations). We are also assuming that the case of 1979 is illustrative of the typical measurement error in the log wage statistics for the group of cities analyzed over the period 1979-1991. Figure 9 shows the distribution (kernel density)

[17] As usual, when calculating wages relative to 1979 we use the March 1980. The Census 1980 also reports wages for 1979.

of this measurement error (deviation from Census) for average log wages in the Borjas sample across the 41 cities used in the donor pool, using the March CPS (grey kernel density) or using the May-ORG (black kernel density). It is evident from the picture that the variance of the measurement error is much larger in the March-CPS which implies that two cities' average log wages may differ from each other by a very significant amount in this sample, simply because of their different measurement error. Remarkably, the March CPS measurement error has a standard deviation of 0.148 logarithmic points (about 15 %) while the ORG-CPS measurement error, has only a standard deviation 0.065 log points (7%). A difference in average wages of 25-30% between two cities could easily arise by pure error in the March-CPS. That difference, in fact, is well within the range of the distribution of errors that, for the March CPS ranges from -0.30 to +0.49 log points. The noise of the data in the March CPS sample seems simply too large to say anything significant, when using only one city relative to another small group of cities as comparison. Let us notice once more that not only the sample for Miami is very small in the Borjas CPS analysis (15 to 25 observations) but the samples for the cities in the "employment control"(e.g. Anaheim, San Jose and Rochester) are even smaller and for several years they only have a number of observations in the single-digit. This confirms our skepticism in using the March-CPS for such a small sample, when focussing only on an event in one city. While the ORG-CPS has a substantial standard deviation of the measurement error, that value is still less than half the one for the March CPS.

7 The Boatlift That Did Not Happen

A relevant methodological caveat to Card (1990) was raised by Angrist and Krueger (1999). While illustrating the difference in difference methodology used by Card (1990) they showed that relying on a treatment-control comparison of only few units (labor markets) before and after a shock can be subject to error, not just because we are measuring an average statistics imprecisely but due to unobserved labor market factors that may affect the treated and control units differently. While Card (1990) finds no effect of the Boatlift in 1979-1981 they show that using the same method one could find a "false positive" result in 1993-95 by comparing unemployment rate for black workers in Miami and in a control group made of the same four cities used by Card (1990): Atlanta, Houston, Los Angeles and Tampa-St. Petersburg. The year 1994 is interesting because a smaller version of a boatlift from Cuba happened, but it was not allowed to reach the US and it was diverted to Guantanamo. So this "non-shock" should not have had consequences on Miami labor market[18]. However, Angrist and Krueger (1999) show that between 1993 (pre non-shock) and 1995 (post non-shock) the unemployment rate for Black workers in Miami increased by 3.6 percentage points, while in the control group it decreased by 2.7 percentage point producing a difference in difference of +6.3 percentage points. Angrist and Krueger (1999) argue that this "false positive" is a cautionary tale for using difference in differences with small number of units. Just as the coincidence of the non-event and the change in unemployment was a fluke, so Card (1990) findings of no effect of the real Boatlift could be a fluke as well.

While their story is a very appropriate cautionary tale for using mechanically a Difference-in-Difference approach, and idiosyncratic deviations between treatment and control groups, especially in the short run, can be pervasive and need to be kept in mind, this is only a partial tale. Analyzing more variables and checking the quality of the control group one realizes that several things do not add up in considering the unemployment change as an effect of a potential (non-existent) Cuban immigration shock.

First, one should look at several groups and more outcomes. Second, one should check the pre-event match between treated unit and control group. In so doing we already see that even using the data reported in the paper Angrist and Krueger (1999) the unemployment rate for white workers shows a Difference-in-Difference change for Miami relative to control equal to only 0.3 percentage points from 1993 to 1995. The unemployment of Hispanics, the group that could have been most affected for their similarity shows a Difference-in-Difference change of +1.4 percentage points. Both these differences are within two standard errors for the average Miami unemployment rate reported by Angrist and Krueger (1999). Hence neither of them confirms the presence of an effect. Second, looking at the deviations of Miami relative to the control group, for the values of unemployment rates of black workers, one realizes that there are many instances of quite large deviations before 1994 (e.g. +4.7 percentage points in 88-89, or -2.7 percentage points in 1991-1992). These are symptoms that the pre-1994 match between Miami and control is not very good for such outcome, and also that there are relatively large deviations

[18] Borjas (2015) argues that some of these Cubans eventually arrived in the US. However, we do not know if they reached Miami and when and, by looking at Miami CPS-ORG data, the share of Cubans in Miami actually declined between 1993 and 1995, or remained unchanged as share of high school dropouts, (see figure A5 in the Appendix). Hence we maintain the assumption of Angrist and Krueger (1999) that this boatlift did not change or changed only minimally, the supply of Cubans in Miami in the period 1993-1995.

between Miami and control in several years, which affect the precision of inference. So, from that exercise we mostly learn about the need to stay cautious and possibly to require an array of results to converge in one direction, before claiming we identify an effect. Let us also notice that in 1994 the CPS underwent a major redesign and several measures of employment and unemployment, especially for males and subgroups were significantly affected (see Polivka and Miller 1995). Hence focussing on changes exactly around 1994 can be very risky.

Nevertheless, does the use of synthetic control method help in this case? After all it eliminates the arbitrary choice of control and it allows the control cities for Miami in 1994 to be different from those chosen by Card (1990). It also allow us a validation, checking how good is the pre-1994 fit of Miami and control. Hence, we apply the synthetic method to high school dropouts weekly wages and hourly wages and to wages at the 20th percentile, for the preferred sample using the ORG-CPS, which are among the variables we considered in our previous analysis[19]. Was there a significant and sudden deviation down of Miami wages relative to the control between 1993 and 1995?

Figure 10 Panel A-B shows the behavior of hourly and weekly wages of high school dropouts in Miami and synthetic control between 1989 and 2001. In order to have a balanced panel of control cities (consistently defined over the whole period) we keep the pre-1994 period to 6 years only. The relevant period to identify an effect of the non-existent Boatlift is 1993-1995. In Panel A, using hourly wages, we do not observe any departure in 93-94 and a small one in 1995-96, however the size of the departure is within the typical variation between Miami and control along the pre-period. The deviation disappears in 1997. In Panel B we observe a somewhat larger deviation in 94-95 that disappears in 96 and reverse and becomes a positive one for post-1996. Using the 20th percentile as alternative measure of low skill (non-Cuban) wages, in Panel C and D, we do not observe any significant difference between Miami and control arising between 93 and 95 neither in Hourly nor in Weekly wages. We observe, however, significant noise, both in Miami year to year variation and in its control so that a deviation of 0.05 between Miami and placebo is not unusual, which is unfortunate as this makes the precision of inference is limited. Overall these pictures do not produce any consistent evidence of a downward wage movement for low skilled in Miami and in 93-95. As for the unemployment rates of

[19] In this case, to keep computational time within a reasonable amount, we limit the "donor pool" for the control group to cities with at least 20 observations in the relevant group of high school dropouts. This produces a pool of about 40 cities.

minorities (Black and Hispanics) that are the only variable shown in Angrist and Krueger (1999), we show their behavior vis-a-vis the synthetic control in Figure A6 of the appendix. While unemployment of Black individuals still shows an increase relative to the synthetic control in 93-94, that difference is much less dramatic and it is actually reversed by 1996. The unemployment of Hispanic individuals experiences actually a decline relative to the synthetic control in 1993-1995. The unemployment of Hispanics in the synthetic control grew significantly more than that of Miami in that period. While the synthetic control cannot solve the problem of significant noise in the data, it can certainly provide tools to account for such noise. Overall, we think that the synthetic control analysis done on wages and unemployment of Miami 1993-95 would bring the researcher to recognize significant noise of the data, some discontinuous jump of Miami and of the control variables in 1994, possibly related to the deep re-design of CPS, but it would not bring him/her to identify consistent signs of an effects on wage and employment of the non-existent 1994 Boatlift.

8 Conclusions

This paper applies the synthetic control method to the Miami boatlift episode, that increased the labor supply of Cubans in Miami possibly up to 18% of the high school dropout group, between April and September 1980. We use a wide variety of labor market outcomes for high school dropouts non-Cubans in Miami and for several sub-groups of those and we look for a significant and sudden impact of this shock on Miami labor markets in the period 1979-1981. We do not find any consistent evidence of a depressing effect on low-skilled labor demand between 1979 and 1981 nor any protracted dynamics after that. In applying the method we learned that noise is significant in all CPS samples, and year to year variation can be non trivial, so that matching exactly outcomes in the pre-event period is important, and sometimes less than perfect. Noise remains large enough that log wage deviations in the order of 0.05 between Miami and control cannot be distinguished from error. This is a significant limitation of the method in this application.

We also learned that applying the synthetic control method, plus including extensive checks (for different samples, variables and groups), improve somewhat on Card (1990). The matching of the pre-event trend by the control group is improved and we are given a less ad-hoc method to choose it. We have run many more tests and regressions that we did not report. However we reported all the representative and robust results that together reveal a reasonable picture. We also understood that by choosing a small enough sample and varying controls one can produce quite different representation of the data, some of them quite far from the overall picture that we produced.

However, running many reasonable scenarios conveys the picture that nothing significant happened to the wage and to the unemployment of Miami high school dropouts between 1979 and 1981, relative to any reasonable control groups. Certainly, the noise and standard errors that we face make it hard to find effects on wages in the order of 2-3 logarithmic points, which would still be non-trivial. We think that with these data it is not possible to get to that level of precision. However we think that one can certainly rule out big negative effects (in the order of 10 percentage points or larger) that have been claimed by Borjas (2015).

Overall we encourage other people to convince themselves of the validity of our results, and of the fact that we are not hiding or taking shortcuts, by accessing all our data and codes, available at our websites, and using them for replicating and producing variations on our results[20]. We think the final goal of the economic profession should be to agree that, even using the more current econometric methods, we do not find any significant evidence of a negative wage and employment effect of the Miami boatlift and move to analyze other cases, to think of adjustment mechanisms that allow absorption of immigrants and to go beyond the naive canonical model that does not provide a good understanding of the labor market effects of immigrants.

[20] All the codes, a readme file and the data to replicate the tables and figures of the paper are available at the website:
http://giovanniperi.ucdavis.edu/data-and-codes.html

References

[1] Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, 2010 “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program,” *Journal of the American Statistical Association* 105 (2010), 493–505.

[2] Abadie, Alberto, and Javier Gardeazabal, 2003“The Economic Costs of Conflict: A Case Study of the Basque Country,” *American Economic Review* 93 (2003), 113–132.

[3] Angrist Joshua and Alan Krueger (1999) "Empirical Strategies in Labor Economics" in O. Ashenfelter & D. Card (ed.), 1999. "Handbook of Labor Economics," Handbook of Labor Economics, Elsevier, edition 1, volume 3, number 3.

[4] Bodvarson Orn, H.F. Van der Berg and J.J. Lewer (2008) "Measuring Immigration’s effect on Labor Demand: A Reexamination of the Mariel Boatlift" *Labor Economics*, 15:4.

[5] Borjas George (2012) "Labor Economics" 6th Edition, McGraw-Hill Companies.

[6] Borjas George, (2015) "The Wage Impact of the Marielitos: a Reappraisal" National Bureau of Economic Research, Working Paper # 21588, Cambridge Mass.

[7] Bohn Sara, Magnus Loefstrom and Steven Rapahel (2014) "Did the 2007 Legal Arizona Workers Act Reduce the State’s Unauthorized Immigrant Population?," *The Review of Economics and Statistics*, MIT Press, vol. 96(2), pages 258-269, May.

[8] Card David (1990) "The Impact of the Mariel Boatlift on The Miami Labor Market" *Industrial and Labor Relation*, Vol 43 n. 2.

[9] Card David (2001) "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, University of Chicago Press, vol. 19(1), pages 22-64, January.

[10] Card, David (2009) "Immigration and Inequality," *American Economic Review*, American Economic Association, vol. 99(2), pages 1-21, May.

[11] Cahuc, Pierre, Stephan Carcillo, A. Zylberberg and Andre’ Zylberberg (2014) "Labor Economics" Second Edition, MIT Press, Cambridge Mass.

[12] Dustmann, Christian, Tommaso Frattini and Ian Preston (2013) "The Effect of Immigration along the Distribution of Wages," *Review of Economic Studies*, Oxford University Press, vol. 80(1), pages 145-173.

[13] Foged and Peri (forthcoming) "Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data," forthcoming in the *American Economic Journal, Applied Economics*.

[14] Laing Kevin (2011) "Labor Economics" W.W. Norton and Company, New York-London.

[15] Lewis, Ethan (2004) "How did the Miami labor market absorb the Mariel immigrants?," Working Papers 04-3, Federal Reserve Bank of Philadelphia.

[16] Lewis Ethan and Giovanni Peri (2015) "Immigration and the Economy of Cities and Regions," in Gilles Duranton, J. Vernon Henderson and William C. Strange Eds., Handbook of Regional and Urban Economics, Volume 5,

[17] Ottaviano Gianmarco I.P. and Giovanni Peri (2012) "Rethinking The Effect Of Immigration On Wages," *Journal of the European Economic Association, European Economic Association*, vol. 10(1), pages 152-197, 02.

[18] Peri (2014) "Do Immigrant Workers Depress the Wage of Native workers?", IZA World of Labor, Evidence-Based Policy Making, May 2014, available at http://wol.iza.org/articles/do-immigrant-workers-depress-the-wages-of-native-workers/long (accessed November 20th 2015).

[19] Polivka, Anne E. and Stephen M. Miller (1995) "The CPS After the Redesign: Refocusing the Economic Lens." US Bureau of Census, 1995; available at "http://www.bls.gov/ore/abstract/ec/ec950090.htm" (accessed on November 20th, 2015)

[20] Roodman David (2015) "Headwind for the Boatlift?" October 21, 2015, available at http://davidroodman.com/blog/2015/10/21/headwind-for-the-boatlift/ (accessed on November 20th 2015)

Tables and Figures

Table 1
Demographics of Mariel Immigrants and of Miami Labor Force in 1980

	Miami Labor Force in 1980	Mariel immigrants, measured from the 1990 Census	Mariel Immigrants still in Miami as of 1990
Total in Labor Force (16 to 65 years of age)	678, 500	93, 129	57,299
Share with no HS degree	28.59	54.41	54.99
Share with HS degree	31.69	25.74	24.92
Share with some college	21.43	13.29	13.04
Share with college	18.29	6.57	7.05
Share of female	45.80	38.60	42.84
Share of young (<25 years old)	20.50	21.53	18.66
Only individuals with no HS degree			
Total in labor force	193,960	50,670	31,509
Percentage female	43.32	39.88	44.59
Percentage young (<25 years old)	22.56	16.26	13.02

Note: The values for the Miami Labor force are obtained from the 1980 census. Those on the Mariel Immigrants are obtained from the 1990 census as people born in Cuba who arrived in the US in 1980 and 1981 and were at least 16 years of age at the time of arrival. Labor force is defined as individual 16-65, not in school, and working or looking for a job.

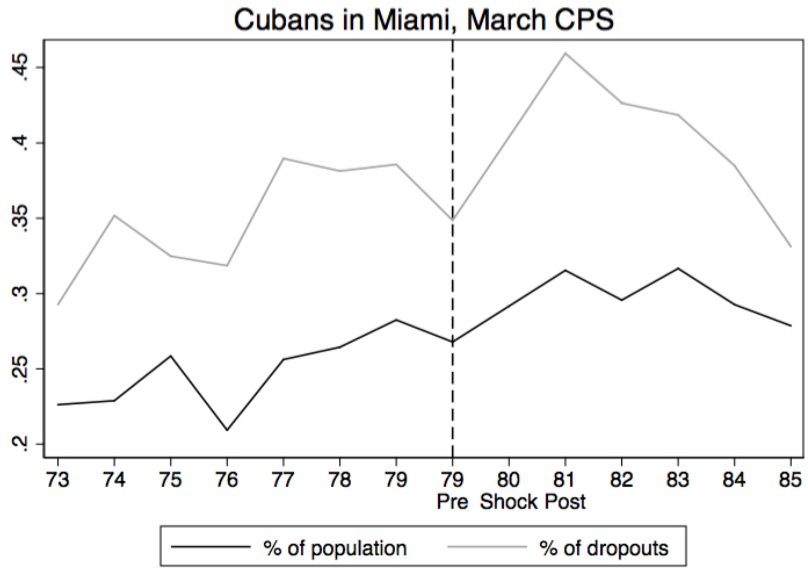
Table 2:
Number of Observations in the Miami sample of High School dropouts used in the analysis

Year	**Our Sample, March**	**Our Preferred Sample, May-ORG**	**Borjas' Sample, March CPS**	**Borjas' Sample, May-ORG**
1973	79	43	30	17
1974	71	38	25	11
1975	72	46	16	13
1976	61	49	22	18
1977	71	43	25	19
1978	61	39	21	11
1979-Pre	72	**156**	17	56
1980-Shock	76	**173**	16	55
1981-Post	72	**156**	18	51
1982	65	**144**	24	39
1983	57	**157**	17	50
1984	59	**150**	15	48
1985	62	**79**	17	26
1986	68	**219**	17	61
1987	79	**240**	17	78
1988	98	**264**	17	72
1989	95	**258**	17	64
1990	76	**272**	4	63
1991	76	**188**	9	41

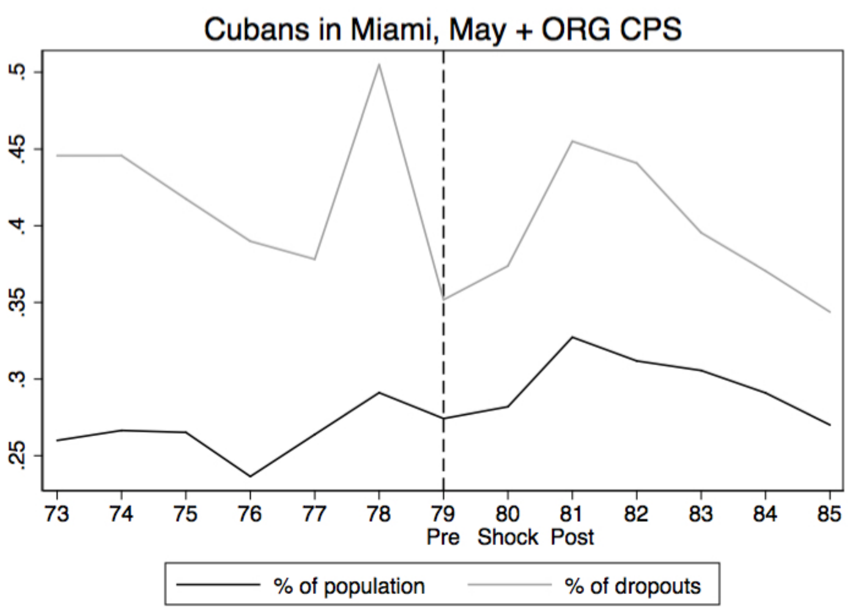
Note: Our sample includes individuals with no high school degree, non-Cuban, with positive earnings, not self-employed, in the labor force in the age range 16-61. The Borjas' sample includes individuals with no high school degree, with positive earnings, not self-employed, in the labor force, male, non-Hispanic in the age range 25-59. A one year adjustment is made to the March CPS numbers as previous year earnings are reported.

Figure 1: Cubans in Miami as share of overall population and of the high school dropout population

Panel A



Panel B



Note: We calculate the share of Cuban people as all those who define themselves as "Cuban" in the ethnicity question of the CPS. The population considered is the total number of individuals between 16 and 61. The high school dropout population is constituted by those who do not have a high school degree between 16 and 61. For the March CPS, we include the figure for March 1980 as "1979-Pre" and we interpolate the figure for "1980-Shock", between 1979-Pre and 1981-Post. The vertical dashed bar is drawn at the last observation before the Mariel Boatlift happened.

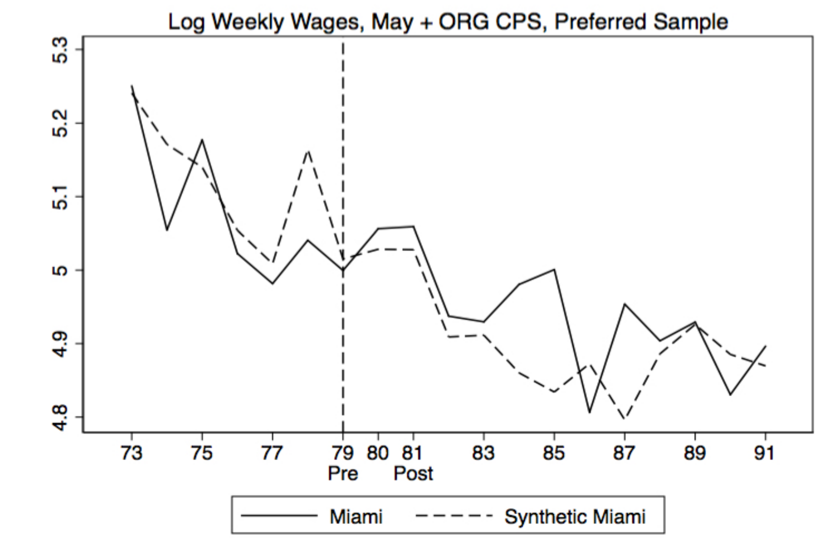
Table 3: "Synthetic Control" Variables, period and Cities-weights

Our Analysis			**Card (1991)**	**Borjas (2015)**
Outcome: Weekly wages for High school dropouts May-ORG CPS	**Outcome: Log weekly Wages for high school dropouts March-CPS**	**Outcome: Unemployment for high school dropouts May-ORG CPS**	**Outcome: Weekly wages overall and by quintile ORG CPS**	**Outcome: Log Weekly wages for High school dropouts March-CPS**
Variables used to minimize pre-treatment distance				
Log weekly wages Share of high school dropouts, Share of Hispanics Share of workers in manufacturing	Log weekly wages Share of high school dropouts, Share of Hispanics Share of workers in manufacturing	Unemployment Share of high school dropouts, Share of Hispanics Share of workers in manufacturing	(not systematically) Share of Black and Hispanics Employment growth rates	Employment growth Employment growth for high school dropouts
Period used to minimize pre-treatment distance				
1972-1979	1972-1979	1972-1979	(not systematically) Decade of the 70's and early 80's	1977-1979
Synthetic control cities and non-zero weights				
Philadelphia, 38.6% Tampa-St. Petersburg, 33.1% Birmingham, 18.9% Anaheim-Santa Ana 9.5%	Los Angeles, 80% Dallas, 14.4% New York, 5.6%	New Orleans, 51% New York, 29.3% Albany, 13.4% Cincinnati, 6.4%	Atlanta Huston Los Angeles Tampa-St. Petersburg	Kansas City, 56% Anaheim-Santa Ana, 20% Sacramento 4.1% San Diego: 1.5% (other cities have less than 1%)

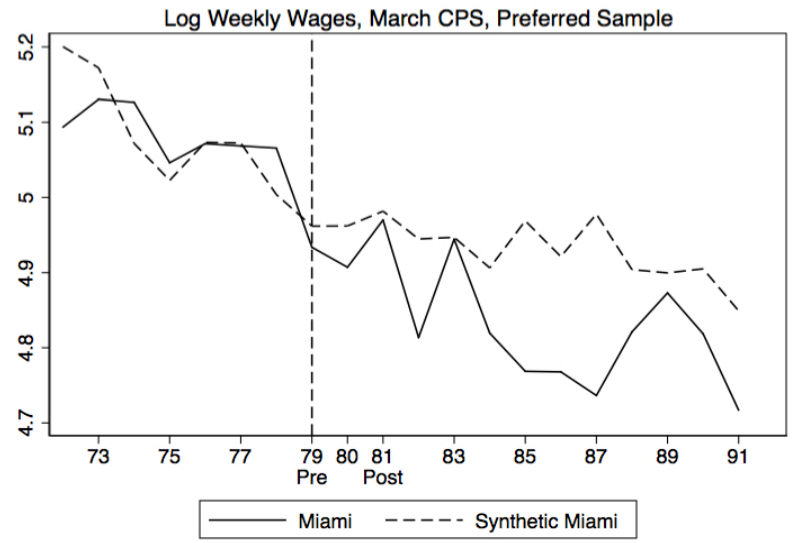
Notes: The choice of variables in our Synthetic control method is driven by the goal of matching as best as possible the pre-1980 trend and level of the outcome variable. Card (1990) only provides in words a description of how he chooses his comparison sample. Borjas (2015) minimizes the difference in employment growth and dropout employment growth between the average 1977-78 and the average 1979-80, independently of the outcome considered.

Figure 2: Log wage measures, Miami and Synthetic control, 1972-1991

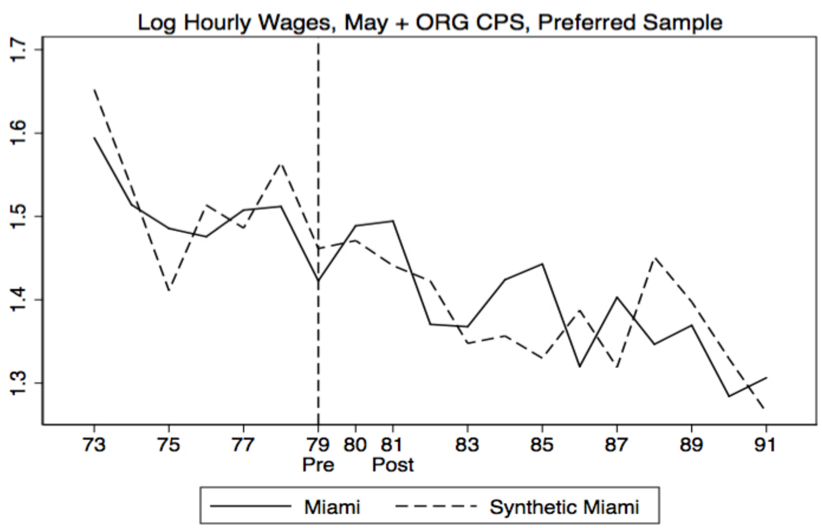
Panel A



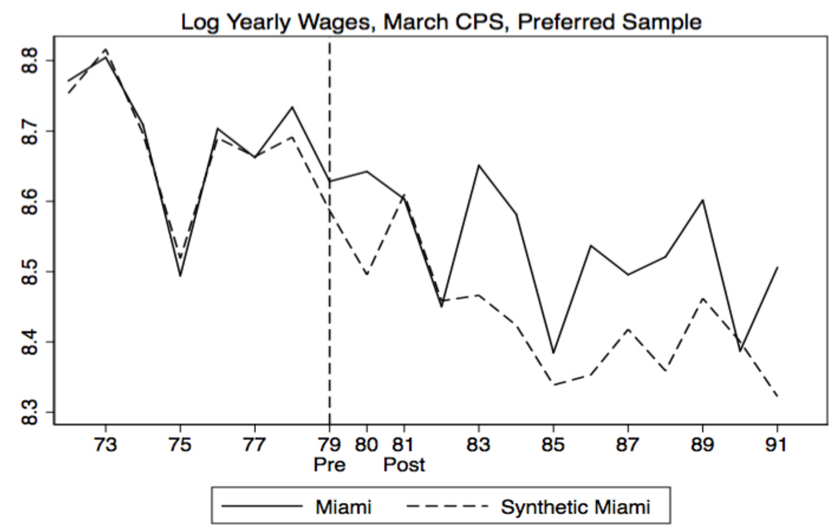
Panel B



Panel C

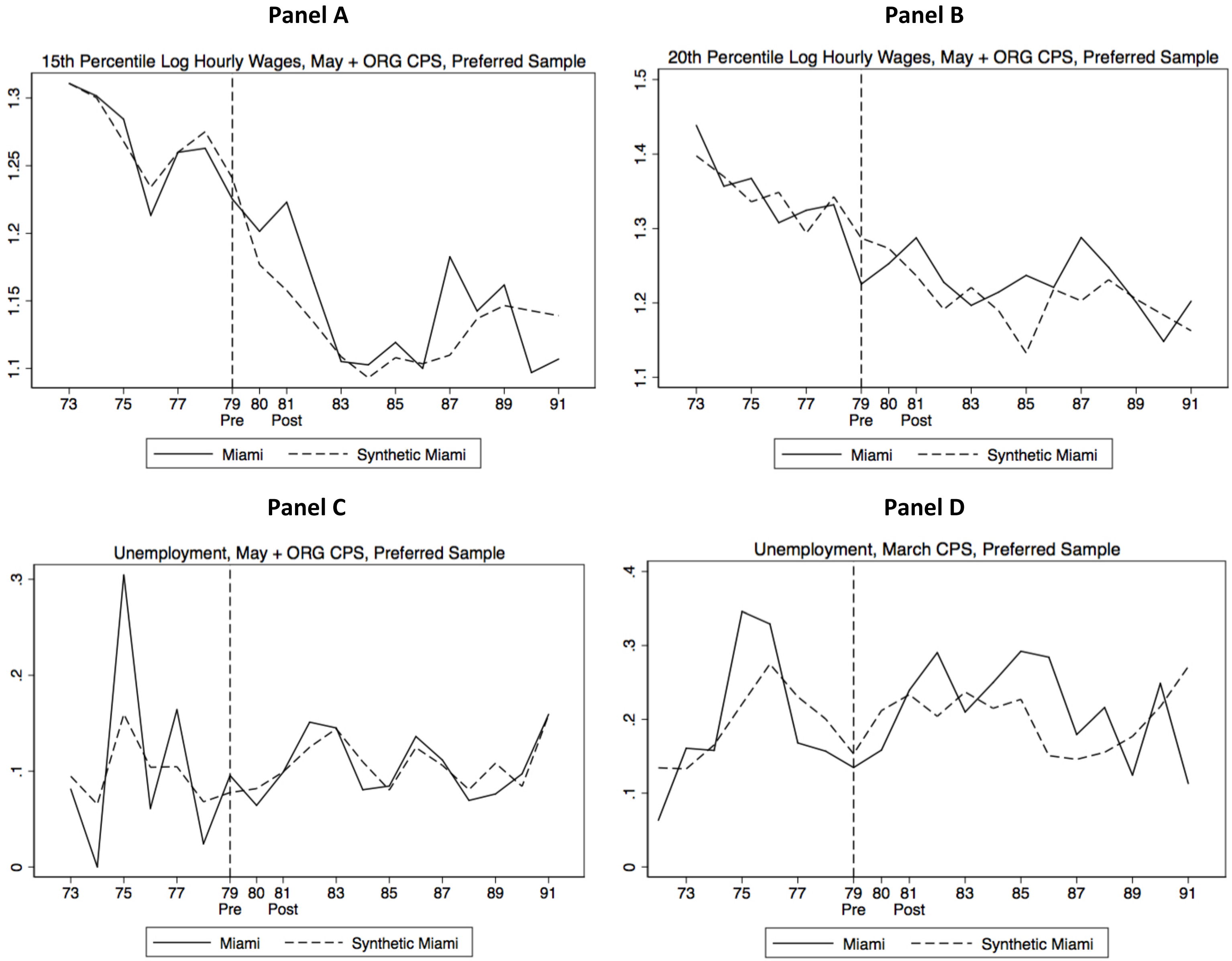


Panel D



Note: Each Panel shows the outcome variable for Miami (solid line) and Synthetic control (dashed line) in the period 1972-1991. The outcome shown and the sample used are noted in the title of each panel (A through D). Preferred sample means: non-Cubans, high school dropouts, not self-employed in the labor force, age 16-61. The vertical line is drawn for year 1979, the last observation before the immigration shock, and it is called “1979-pre”.

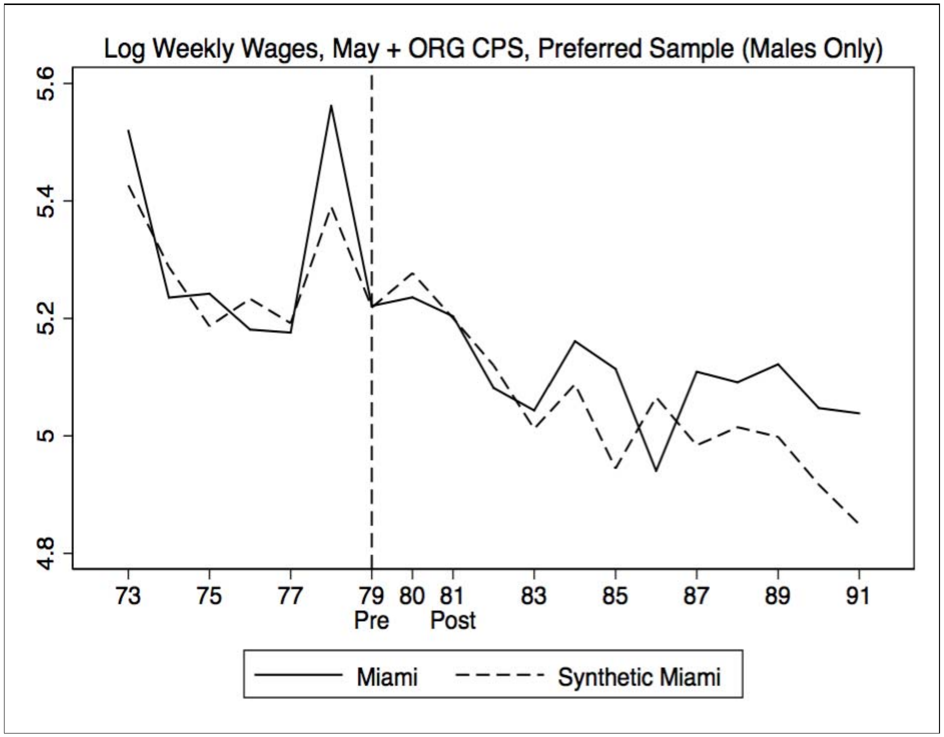
Figure 3: Other labor market outcomes, Miami and Synthetic control, 1972-1991



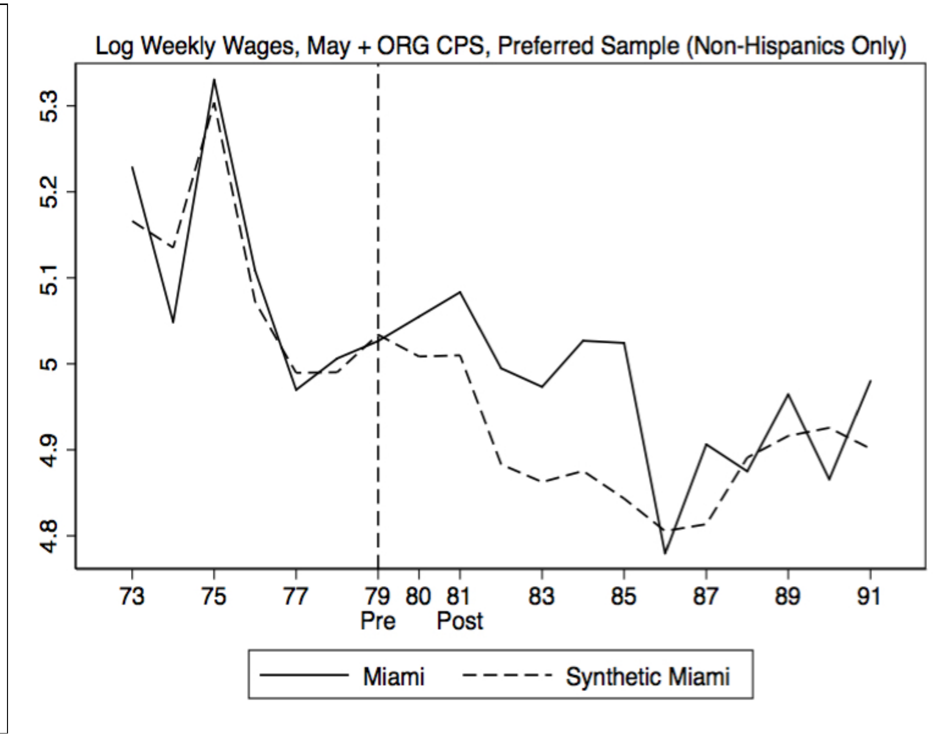
Note: Each Panel shows the outcome variable for Miami (solid line) and Synthetic control (dashed line) in the period 1972-1991. The outcome showed for each Panel and the sample used are noted in the title of each panel. Preferred sample means: non-Cubans, high school dropouts, not self-employed in the labor force, age 16-61. The vertical line is on year 1979, the last observation before the immigration shock, and it is called 1979-pre.

Figure 4: Subsamples of high school dropouts: Males only and non-Hispanics only

Panel A

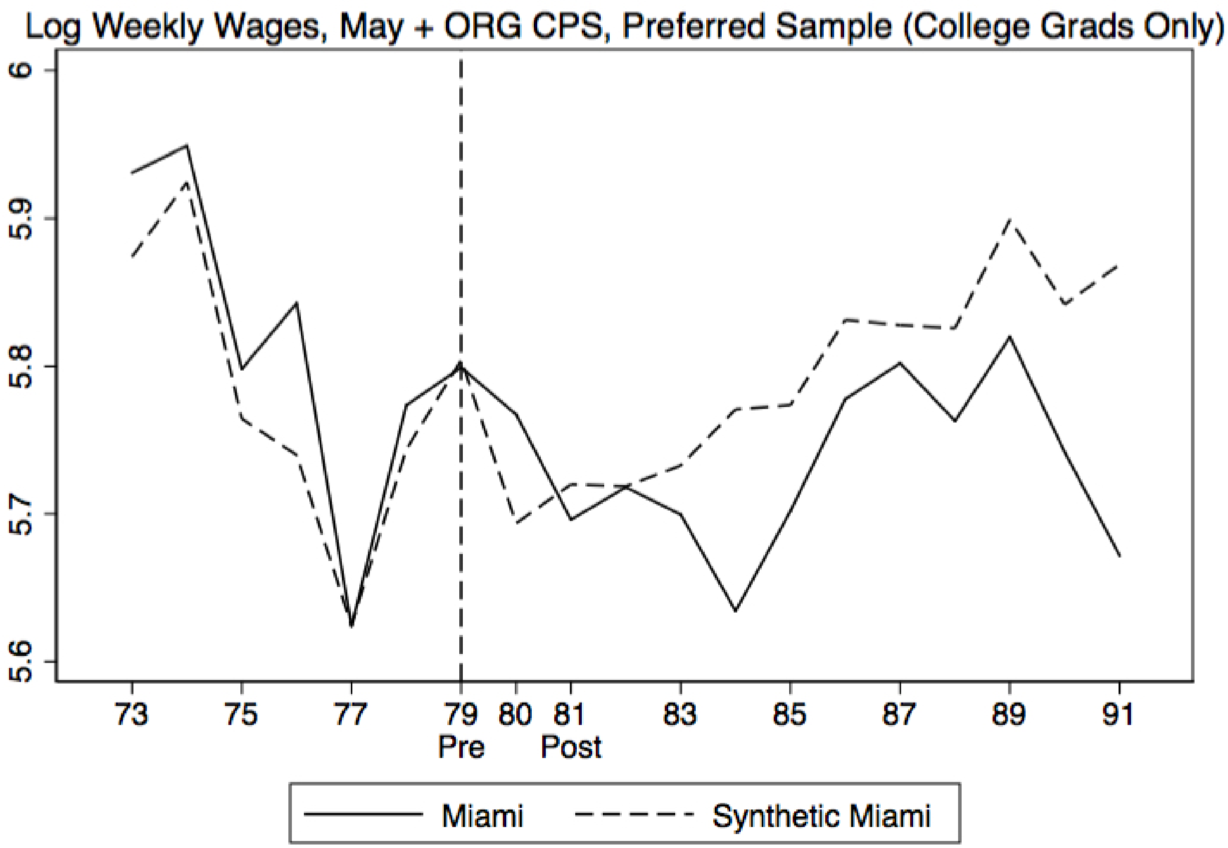


Panel B



Note: Each Panel shows the outcome variable for Miami (solid line) and Synthetic control (dashed line) in the period 1972-1991. The vertical line corresponds to year 1979, right before the Mariel Boatlift happened. The outcome variable and the sample are noted in the title of each panel. Preferred sample means: non-Cubans high school dropouts in the labor force, not self-employed, age 16-65. Panel A further restricts the preferred sample to males only. Panel B restricts the preferred sample to non-Hispanic only.

Figure 5: College educated log weekly wages, Miami and Synthetic control



Note: The figure shows the logarithm of weekly wages for Miami (solid line) and Synthetic control (dashed line) in the period 1972-1991. The vertical line corresponds to year 1979, when the Miami Boatlift happened in Miami. The sample includes college educated non-Cubans between 16 and 61 years of age, not self-employed in the labor force and it is taken from the May-ORG CPS.

Table 4: Difference in Difference Regressions using Miami and Synthetic Miami observations 1972-1991

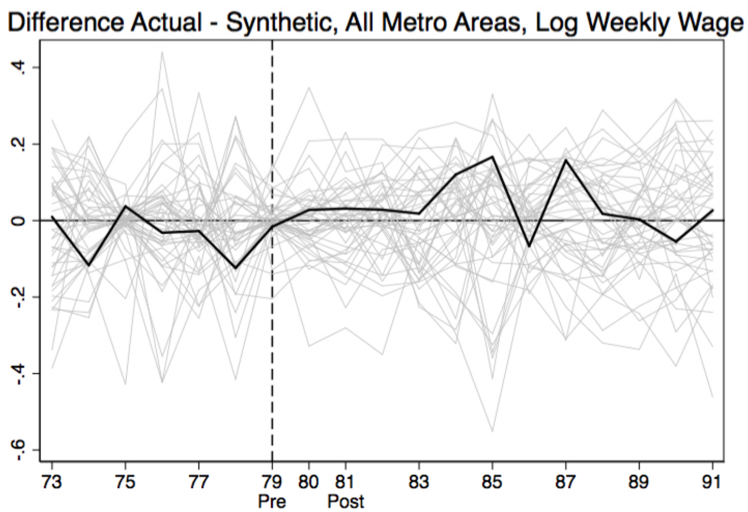
Preferred Sample, non-Cuban, high school dropouts 16-61.

	Dependent Variable and Sample						
	(1) Log Weekly Wages, HS dropouts, May-ORG	**(2) Log Weekly wages, HS dropouts, March**	**(3) Log Hourly Wages, HS dropouts, May-ORG**	**(4) Log Yearly wages, HS dropouts, March**	**(5) 15th percentile, May-ORG**	**(6) Unemployment rate, HS dropouts May-ORG**	**(7) Unemployment rate, HS dropouts March**
Miami X ('72-'75)	-0.008	0.010	0.038	-0.043	0.021	0.004	0.037
	(0.066)	(0.049)	(0.079)	(0.102)	(0.015)	(0.098)	(0.067)
Miami X('76-'78)	-0.045	0.047	0.016	-0.024	0.004	-0.027	0.001
	(0.051)	(0.023)	(0.026)	(0.023)	(0.021)	(0.045)	(0.060)
Miami X ('80-'82)	0.045	-0.038	0.045	0.002	0.055*	-0.015	0.032
	(0.058)	(0.047)	(0.044)	(0.075)	(0.022)	(0.029)	(0.040)
Miami X ('83-'85)	0.117**	-0.069	0.106***	0.088	0.021**	-0.026	0.043
	(0.032)	(0.065)	(0.025)	(0.089)	(0.007)	(0.029)	(0.025)
Miami X ('86-'88)	0.052	-0.131***	0.009	0.100***	0.057*	-0.016	0.094**
	(0.053)	(0.034)	(0.047)	(0.024)	(0.026)	(0.024)	(0.031)
Miami X ('89-'91)	0.007	-0.054	0.028	0.062	0.004	-0.024	-0.041
	(0.034)	(0.050)	(0.047)	(0.075)	(0.021)	(0.034)	(0.052)
R^2	0.829	0.856	0.732	0.700	0.934	0.057	0.309
Observations	38	40	38	40	38	38	40

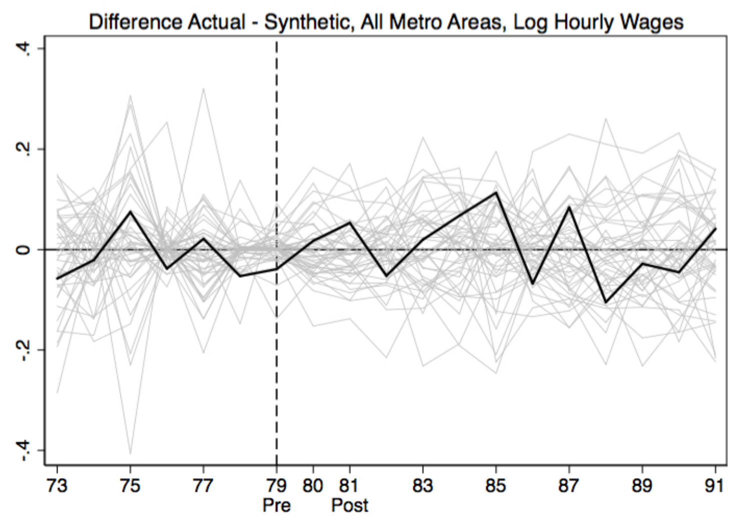
Note: Each column represents a regression of annual observations for Miami and the corresponding synthetic counterfactual between 1972 (1973 for May+ORG sample) and 1991. Each specification includes vectors of city and year bins dummies. Each period dummies extends for three years, except for the beginning of the period in which 4 years are included in each dummy. The bin for 1979 is excluded so as to standardize the value of that interaction to 0. The interaction coefficients between a dummy variable for Miami and a corresponding year bin are reported. Method of estimation is non-weighted OLS. Robust standard errors are reported in parenthesis. * p<0.05; ** p<0.01; *** p<0.001.

Figure 6: Inference, Simulated permutations, 44 metropolitan areas

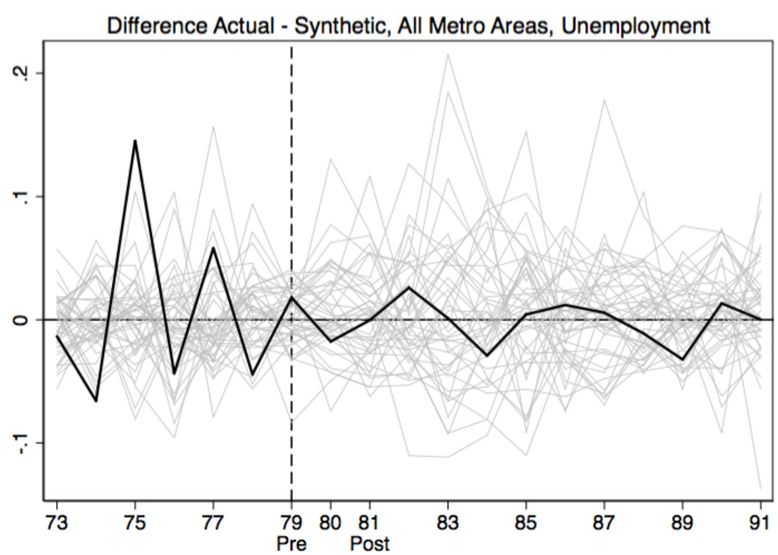
Panel A



Panel B



Panel C



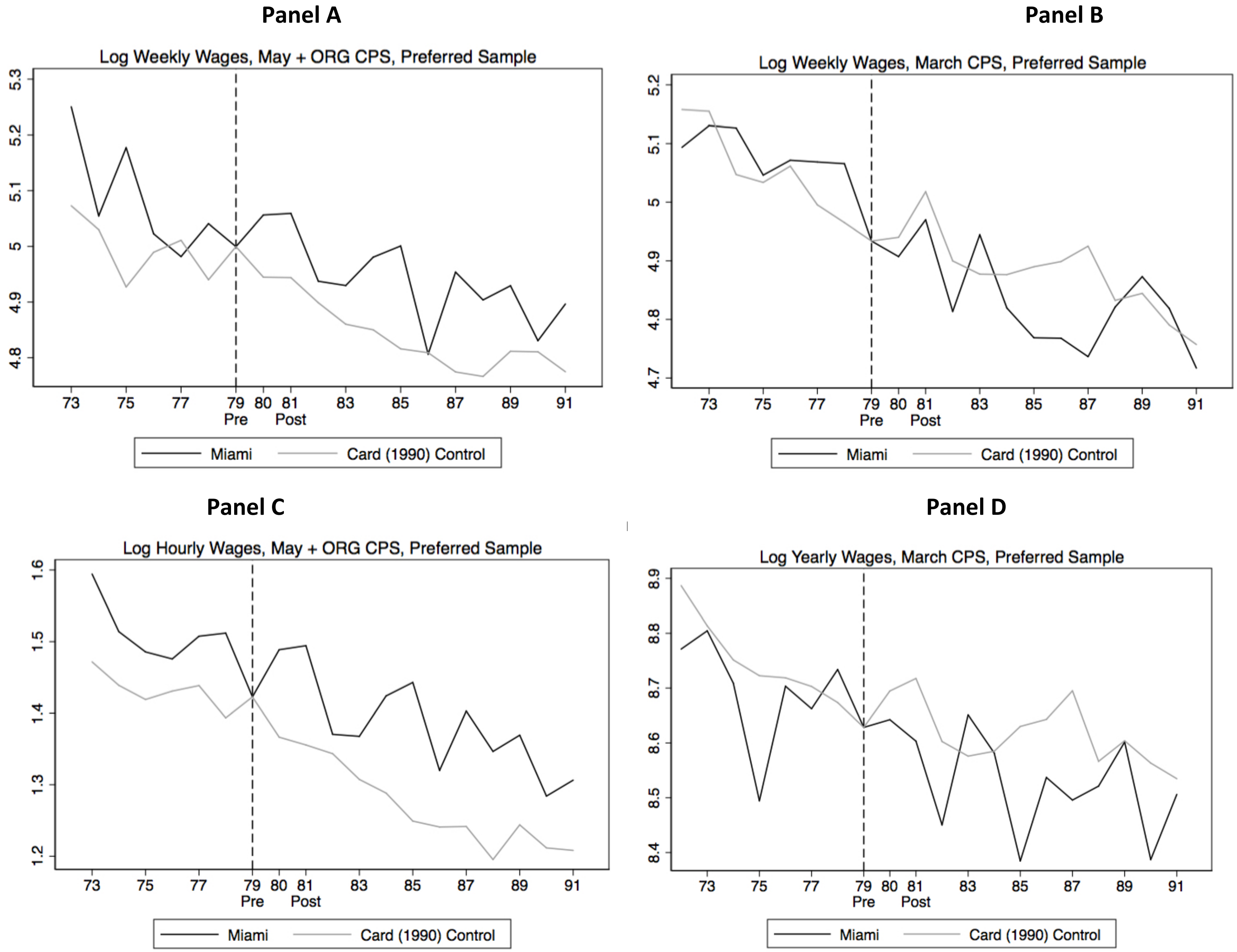
Note: Each graph reports the difference in the outcome variable between treated group and synthetic control, assuming a treatment in 1980, for 44 metropolitan areas. The bold line represents Miami. Panel A shows the graph for the logarithm of weekly wages, Panel B shows it for the logarithm of hourly wages. Panel C for the unemployment rate. The sample is always non-Cuban, high school dropouts, 16-61 years old from the May-ORG CPS.

Table 5: Inference Using Permutations of 44 metropolitan areas and their respective synthetic controls

	Outcome variable		
	(1) Log weekly Wages	**(2) Log Hourly wages**	**(3) Unemployment**
	Analysis relative to Pre-period 72-79		
Diff-in-Diff	0.067	0.022	-0.005
Rank, lowest to highest	29/44	30/44	19/44
P-value, one tailed test $P(\Delta>\Delta_{MIAMI})$ in (1) and (2) $P(\Delta<\Delta_{MIAMI})$ in (3)	0.66	0.68	0.57
	Analysis relative to Pre-period 77-79		
Diff-in-Diff	0.085	0.030	-0.008
Rank, lowest to highest	36/44	32/44	16/44
P-value, one tailed test $P(\Delta>\Delta_{MIAMI})$ in (1) and (2) $P(\Delta<\Delta_{MIAMI})$ in (3)	0.82	0.73	0.64

Note: The "Diff-in-Diff" equals the average Miami-Synthetic control deviation in 80-82 minus the average Miami-Synthetic control deviation in the pre-period. In the upper panel the pre-period is the whole period 72-79, in the lower panel it is the last two years 77-79. We also calculate the same difference for each city in the donor pool and we construct a distribution of the diff-in-diff statistics. The "rank" entry shows were Miami ranks in the distribution of 44 values (bottom to top) the p-value is a test of the probability that a random draw from the donor pool takes a value lower (for wages) or higher (for unemployment) than Miami.

Figure 7: Log wage measures, high school dropouts, Miami and Card (1990) control, 1972-1991



Note: Each Panel shows the behavior of the outcome variable for Miami (solid line) and for the Card (1990) control (dashed line) in the period 1972-1991. The variable and sample are noted in the title of each panel. Preferred sample means: non-Cubans, high school dropouts in the labor force, age 16-61. The vertical line is on 1979, Miami and control have been aligned exactly in 1979.

Table 6: Difference in Difference Regressions using Miami and Card(1990) control as observations 1972-1991

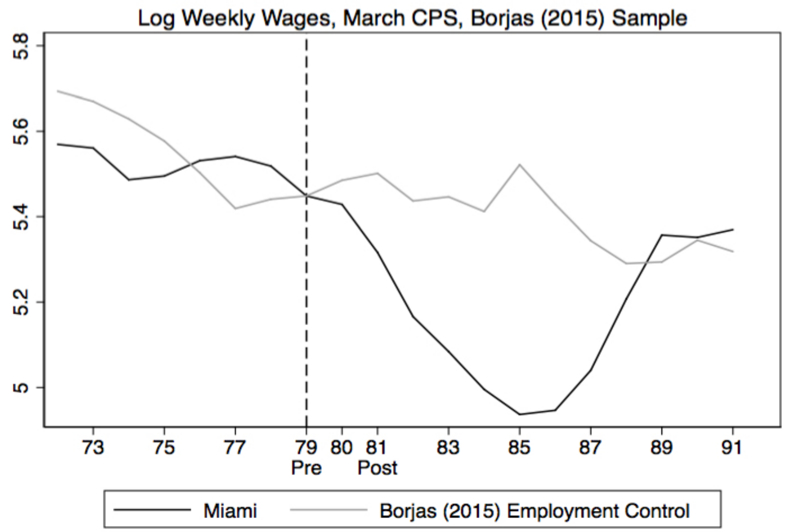
Our preferred Sample, non-Cuban high school dropouts 16-61.

	(1) Log Weekly Wages, HS dropouts, May-ORG	(2) Log Weekly wages, HS dropouts, March	(3) Log Hourly Wages, HS dropouts, May-ORG	(4) Log Yearly wages, HS dropouts, March
Miami X ('72-'75)	0.151 (0.074)	0.001 (0.042)	0.088* (0.037)	-0.099 (0.084)
Miami X('76-'78)	0.035 (0.028)	0.061* (0.029)	0.077*** (0.018)	0.001 (0.025)
Miami X ('80-'82)	0.089 (0.044)	-0.056 (0.058)	0.096* (0.042)	-0.106 (0.069)
Miami X ('83-'85)	0.128*** (0.026)	-0.037 (0.053)	0.130*** (0.029)	-0.058 (0.083)
Miami X ('86-'88)	0.105* (0.046)	-0.110** (0.037)	0.130*** (0.030)	-0.117** (0.040)
Miami X ('89-'91)	0.086* (0.032)	0.006 (0.053)	0.099** (0.029)	-0.069 (0.066)
R^2	0.851	0.860	0.908	0.624
Observations	38	40	38	40

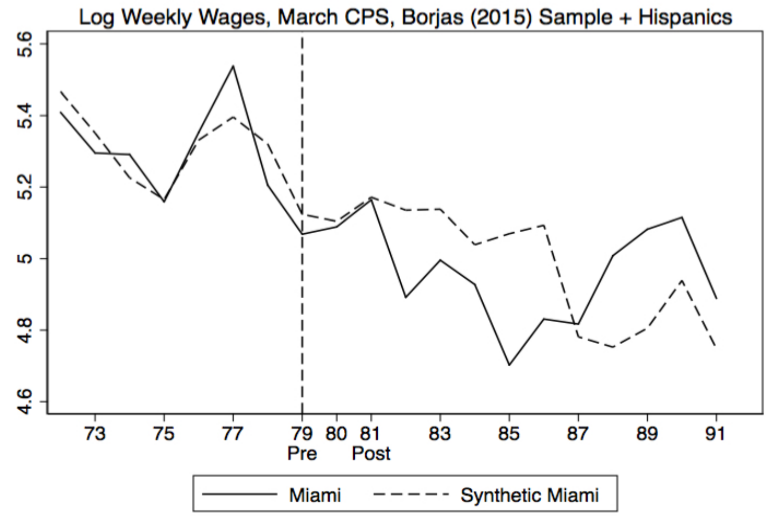
Note: Each column represents a regression of annual observations for Miami and the corresponding Card (1990) control between 1972 (1973 for May+ORG sample) and 1991. Each specification includes a Miami fixed effect, period dummies and the interaction between Miami dummy and period effects. Each period dummies extends for three years, except for the beginning of the period in which 4 years are included in each dummy. The bin for 1979 is excluded so as to standardize the value of that interaction to 0. The interaction coefficients between a dummy variable for Miami and a corresponding year bin are reported. Method of estimation is non-weighted OLS. Robust standard errors are reported in parenthesis. * p<0.05; ** p<0.01; *** p<0.001.

Figure 8: Explaining the Borjas (2015) result: Sample, timing and Control group

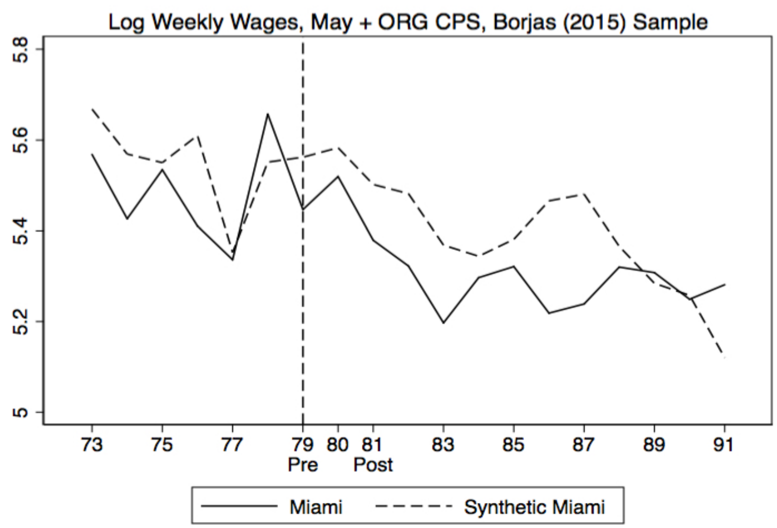
Panel A



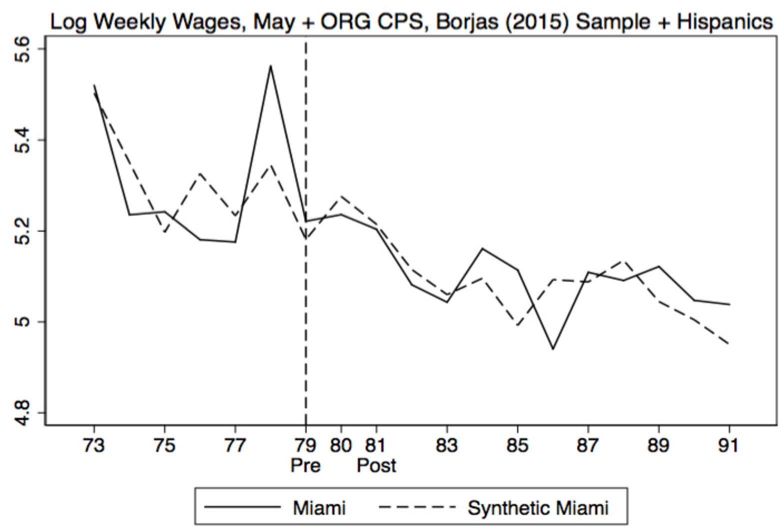
Panel B



Panel C



Panel D



Note: Panel A reproduces Figure 3A in Borjas (2015). The logarithm of weekly wages for males, high school dropouts, non-Hispanic age 25-59, smoothed with a 3-years moving average in the March CPS sample for Miami and the "Employment control" are shown. Panel B extends the sample to non-Cuban, males 16-61, drops the smoothing and introduces the synthetic control. Panel C goes back to Borjas original sample, drops the smoothing and uses ORG-CPS data. Panel D uses the same group definition of Panel B, on ORG data.

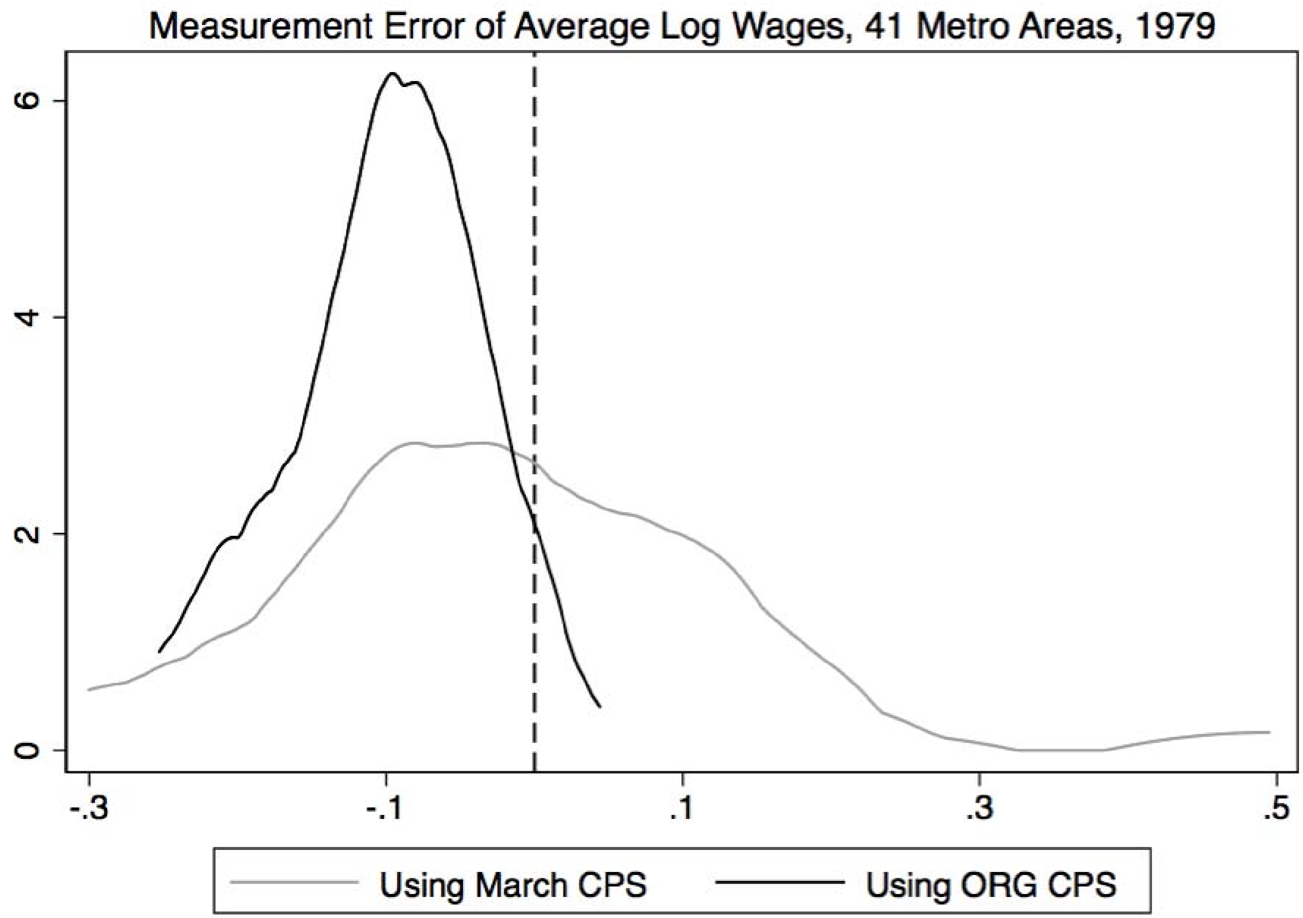
Table 7: Regression analysis for the Borjas Sample and Control Group: Miami and Control 1972-1989

	Dependent variable, Sample, Control			
	(1) **Log Weekly Wages, Borjas Sample, Borjas Employment Control, March CPS**	**(2)** **Log Weekly Wages, Borjas Sample + Hispanics, Our Control, March CPS**	**(3)** **Log Weekly Wages, Borjas Sample, Our Control, May + ORG CPS**	**(4)** **Log Weekly Wages, Borjas Sample + Hispanics, Our Control, May + ORG CPS**
Miami X (72-75)	-0.114**	0.042	0.029	-0.058
	(0.036)	(0.091)	(0.058)	(0.132)
Miami X (76-78)	0.076**	0.071	0.079	-0.035
	(0.026)	(0.100)	(0.128)	(0.136)
Miami X (80-82)	-0.171*	-0.033	0.000	-0.068
	(0.080)	(0.085)	(0.068)	(0.068)
Miami X (83-85)	-0.445***	-0.152	0.023	0.017
	(0.054)	(0.095)	(0.041)	(0.047)
Miami X (86-88)	-0.290**	0.065	-0.063	-0.098
	(0.087)	(0.127)	(0.049)	(0.057)
Miami X (89-91)	0.040*	-0.253*	0.174**	0.029
	(0.016)	(0.092)	(0.055)	(0.039)
R^2	**0.914**	**0.785**	**0.746**	**0.658**
Observations	**40**	**40**	**38**	**38**

Note: Each column represents a regression of annual observations for Miami and the corresponding Borjas (2015) control between 1972 (1973 for May+ORG sample) and 1991. Each specification includes a Miami fixed effect, period dummies and the interaction between Miami dummy and period effects. Each period dummies extends for three years, except for the beginning of the period in which 4 years are included in each dummy. The bin for 1979 is excluded so as to standardize the value of that interaction to 0. The interaction coefficients between a dummy variable for Miami and a corresponding year bin are reported. Robust standard errors are reported in parenthesis. * p<0.05; ** p<0.01; *** p<0.001.

Figure 9: Distribution of measurement error for the average log wage of high school dropouts across 41 cities

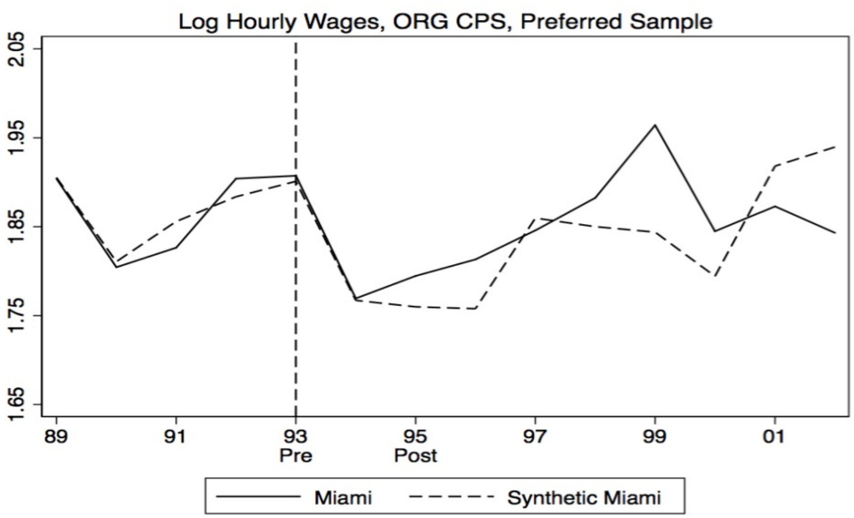
Borjas sample in March and ORG CPS



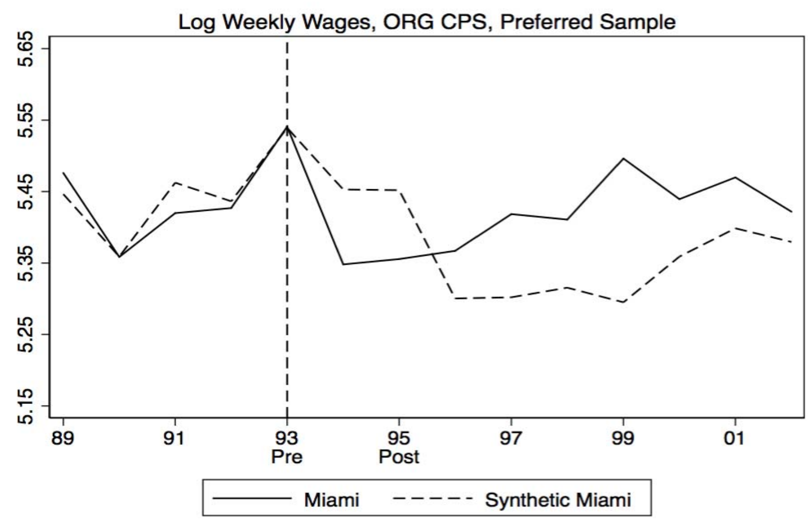
Note: Distribution of the measurement error for the average ln(wage) of the sample of high school dropouts, non-Hispanic, male, 29-59, across 41 cities in 1979 for the March-CPS (grey line) and the ORG-CPS (black line). The measurement error is defined as the deviation from the same statistics (average log wage) for the same groups and city calculated using the Census 1980.

Figure 10: The Non-Event of 1994, wage outcomes in Miami and in the Synthetic control 1989-2001

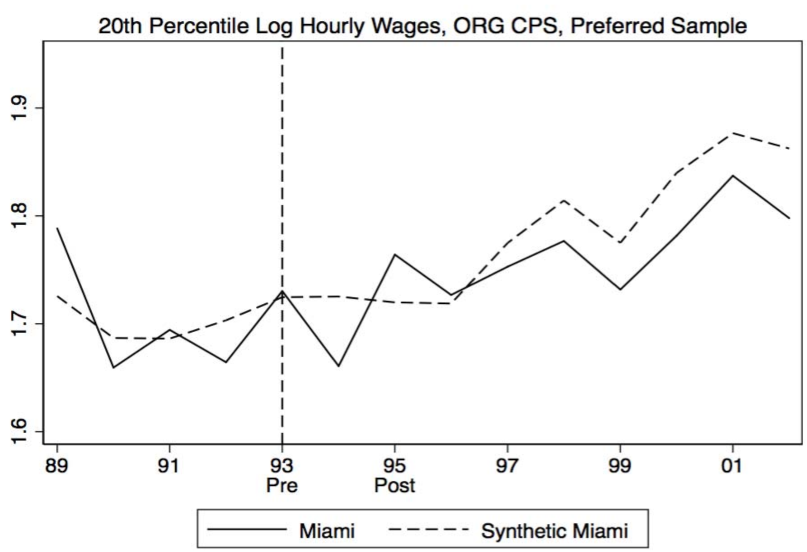
Panel A



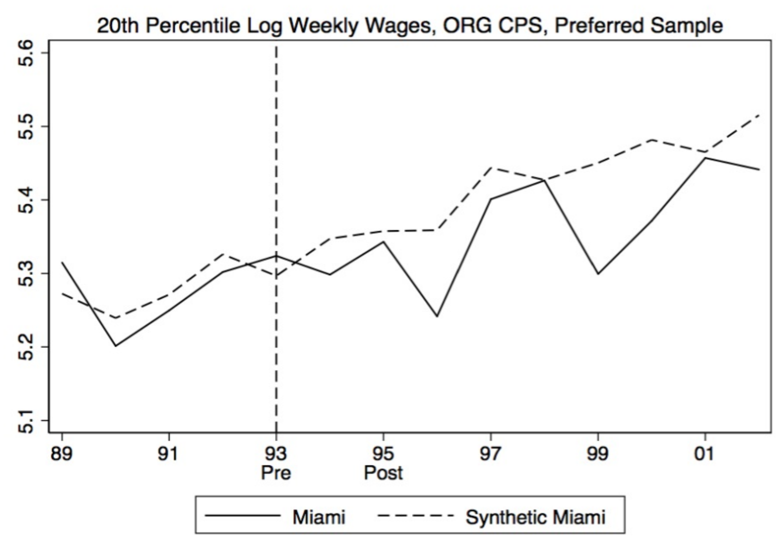
Panel B



Panel C



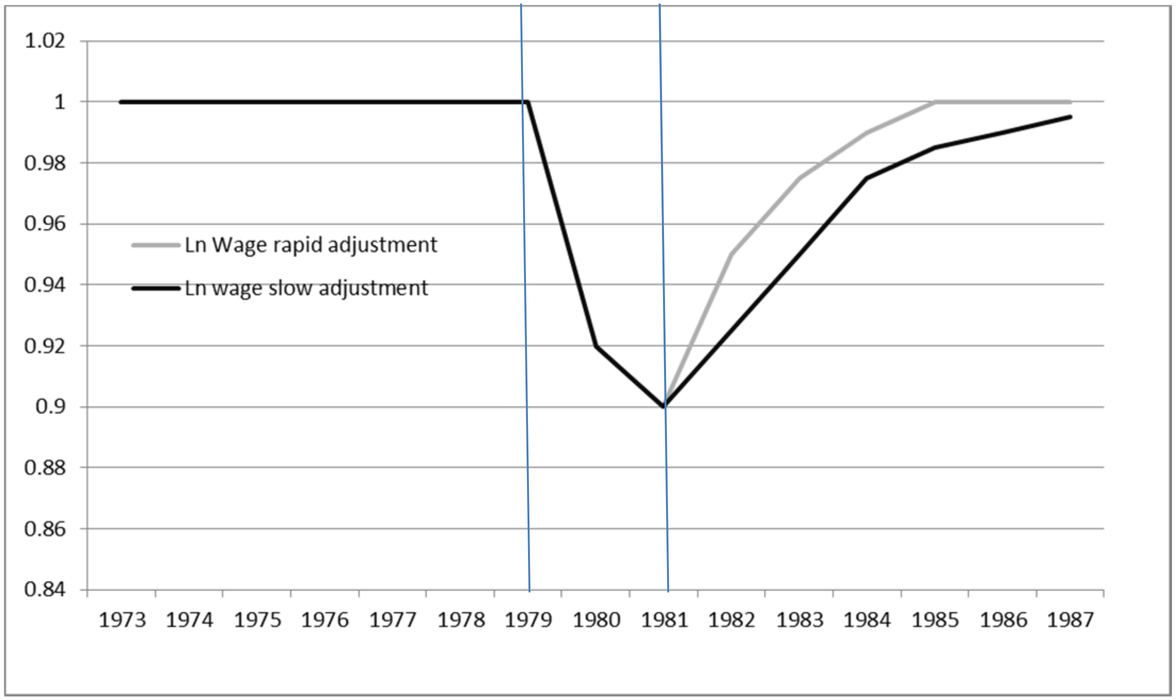
Panel D



Note: Each Panel shows the outcome variable for Miami (solid line) and for the synthetic control (dashed line) in the period 1989-2001. The vertical line corresponds to year 1993, immediately before the non-event of 1994. The variable and sample are noted in the title of each panel. Preferred sample means: non-Cubans, high school dropouts in the labor force, age 16-61.

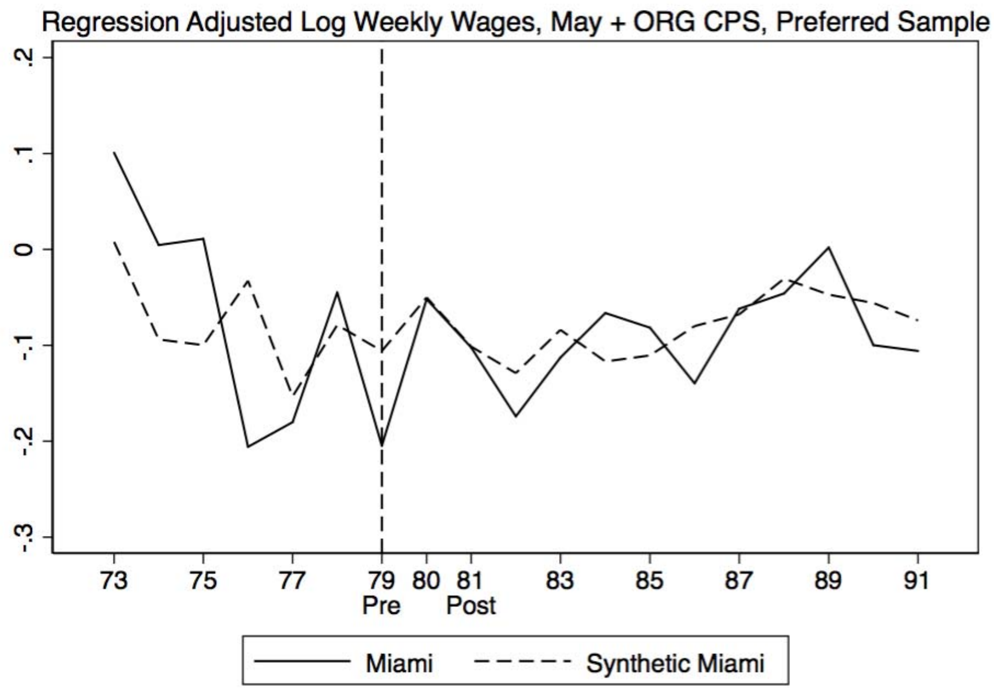
Table and Figure Appendix

Figure A1: Wage Dynamics according to the canonical model by the Boatlift shock on Miami Dropouts



Note: A hypothetical labor supply shock large enough to reduce wages by 10% on impact would have had the above illustrated dynamics, according to the canonical labor demand-labor supply model on wages of high school dropouts in Miami. The value of 10% is arbitrarily chosen. What matter are the shape and timing of the dynamics. The maximum effect is reached in 1981, and transition dynamics start in 1981 and reduce the effect. As the labor supply shock produced by the Boatlift is not discernible any longer as of 1985 in the CPS data we have illustrated the adjustment as full or almost full by that date. Even with slow capital adjustment, the Mariel supply shock (for Miami) was temporary, as the share of Cuban went back to pre-1979 levels in 1985 and hence so should be the response of wages.

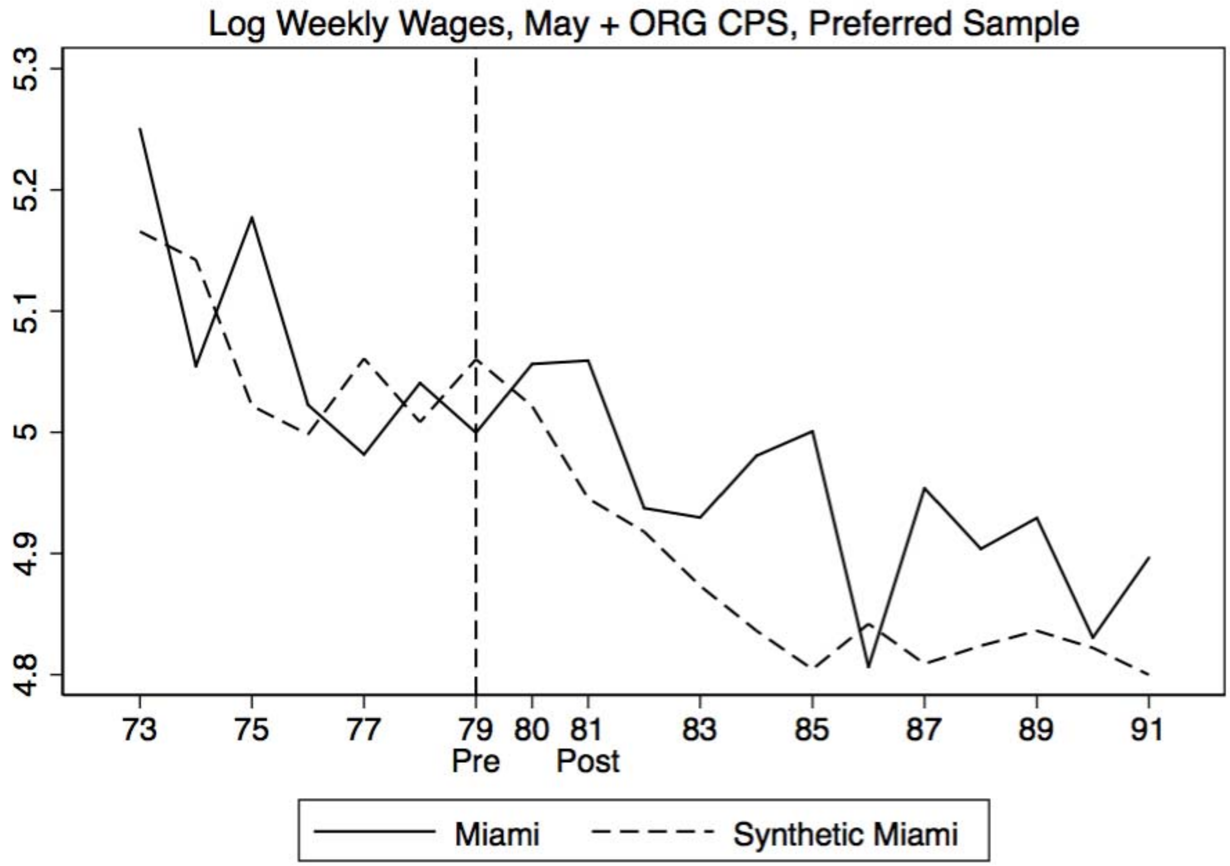
Figure A2: Log wage residuals, Miami and Synthetic control, 1972-1991



Note: The figure shows the outcome variable for Miami (solid line) and Synthetic control (dashed line) in the period 1972-1991. The outcome shown is the average residual of log weekly wages, after we have controlled in an aggregate regression for 5-years age dummies, Hispanic-by-year dummies and female-by-year dummies. Preferred sample means: non-Cubans, high school dropouts, not self-employed in the labor force, age 16-61. The vertical line is drawn for year 1979, the last observation before the immigration shock, and it is called "1979-pre".

Figure A3:

Synthetic control uses log weekly wages, employment growth and employment growth of high school dropouts



Note: The sample and specification as in Figure 2, Panel A, but the variables used to minimize distance are log weekly wages, growth rate of employment and growth of employment of high school dropouts. The synthetic control group is constituted by the following cities and weights in parenthesis:

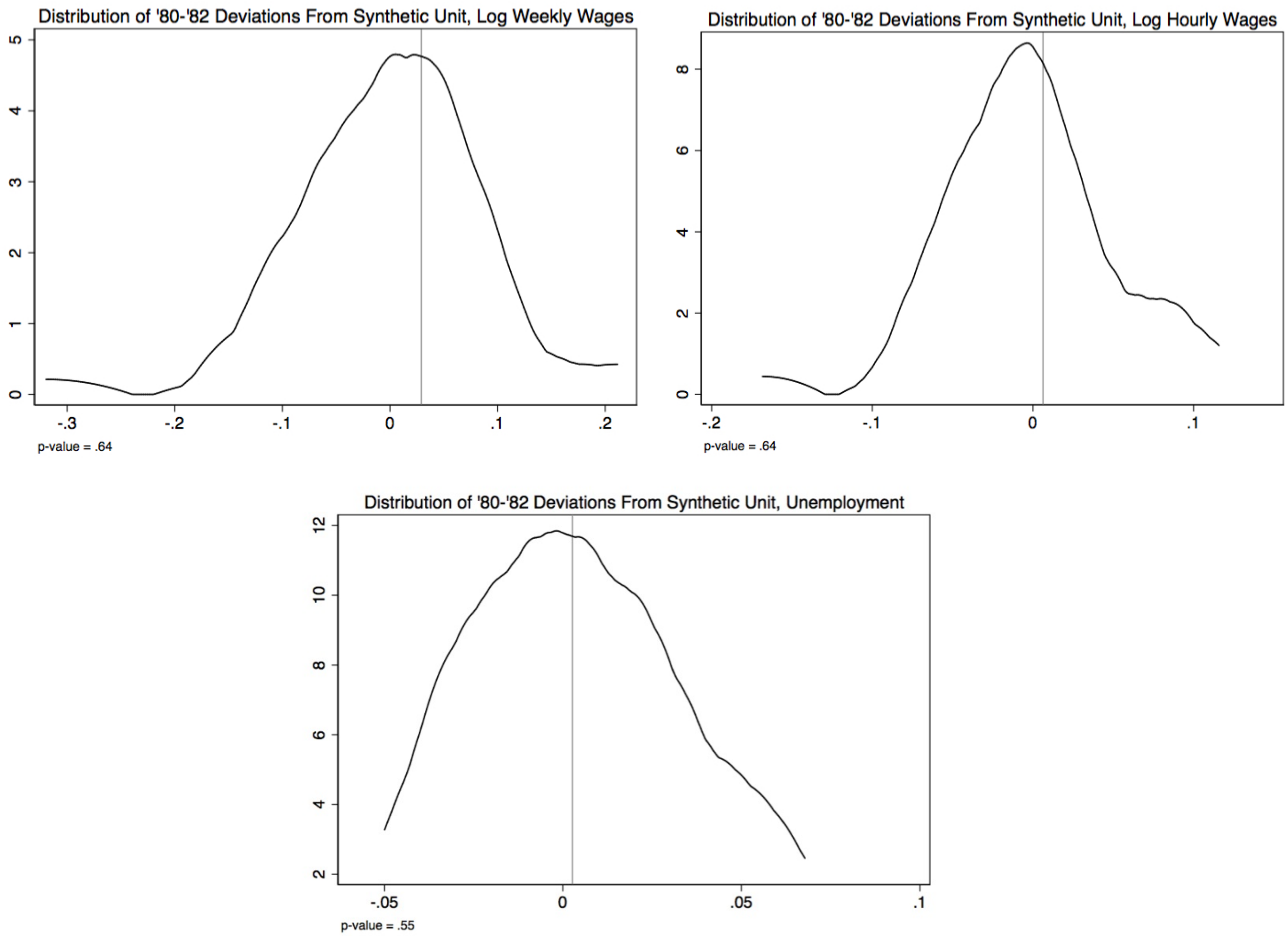
Table A1:
Distribution of 80-82 deviations of city outcomes from their synthetic control, relative to pre-1980 deviations

Permutations of 44 metropolitan areas

	Outcome variable		
	(1) Log weekly Wages	**(2) Log Hourly wages**	**(3) Unemployment**
	Analysis relative to Pre-period 72-79		
Ratio of Post-Pre MSPE	0.326	0.169	0.148
Rank, lowest to highest	15/44	9/44	1/44
P-value, one tailed test $P(\Delta>\Delta_{MIAMI})$	0.66	0.80	0.98
	Analysis relative to Pre-period 77-79		
Ratio of Post-Pre MSPE	0.525	0.269	0.253
Rank, lowest to highest	12/44	8/44	2/44
P-value, one tailed test $P(\Delta>\Delta_{MIAMI})$	0.73	0.82	0.96

Note: The "Ratio of Post-Pre" equals the absolute value of the ratio of the average Miami-Synthetic control square deviation in 80-82 divided the average Miami-Synthetic control square deviation in the pre-period. In the upper panel the pre-period is the whole period 72-79, in the lower panel it is the last two years 77-79. We also calculate the same ratio for each city in the donor pool and construct a distribution of the 44 ratio statistics. The "rank" entry shows were Miami ranks in the distribution of 44 values (bottom to top) the p-value is a test of the probability that a random draw from the donor pool takes a higher than Miami value.

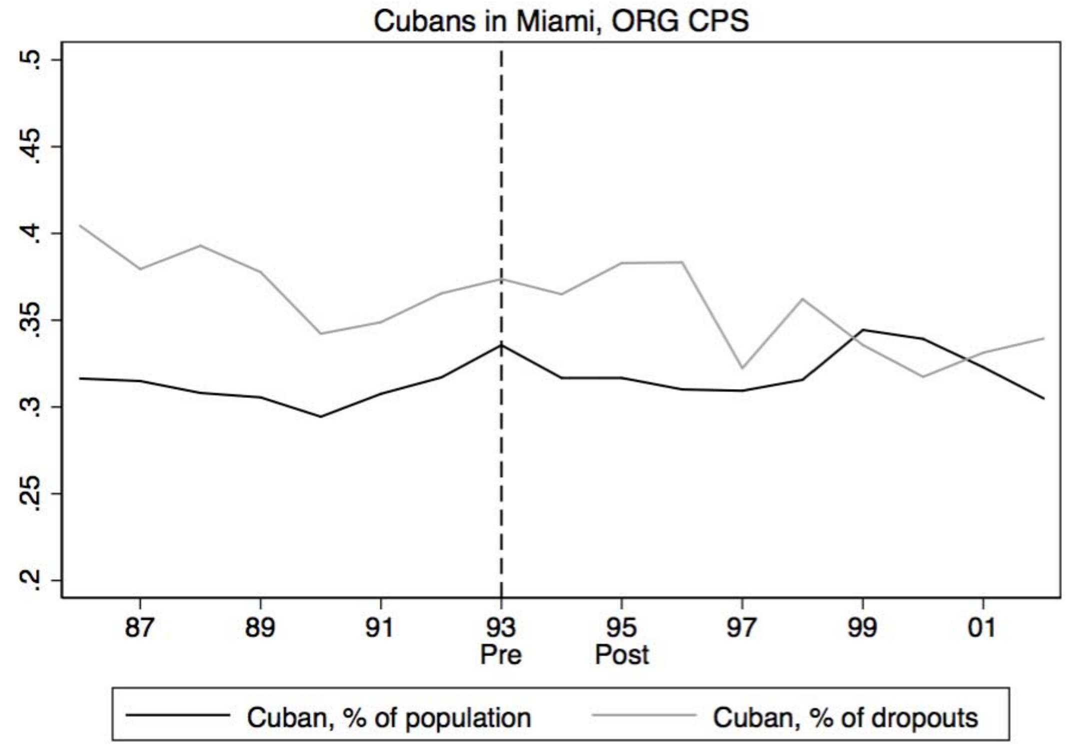
Figure A4:
Distribution of the Unit-synthetic control deviations 1980-1982,
44 metropolitan areas



Note: Each Panel shows the (kernel) distribution of the average deviation 1980-1982 between unit and synthetic control for 44 metropolitan areas. Outcome variable is in the title of each panel. Miami is represented by the vertical line. The p-value is for the one-sided test that a city's value is smaller (in Panel A and B) or higher (in Panel C) than Miami's.

Figure A5:

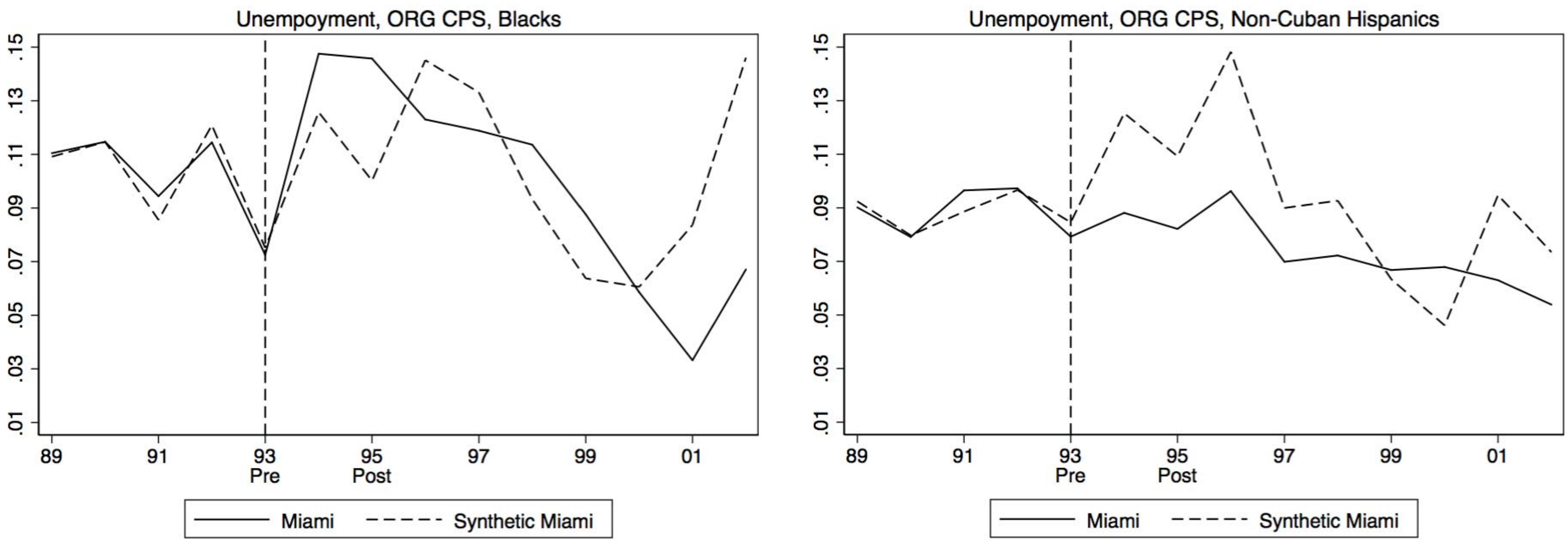
Non-Event of 1994, Cubans in Miami as Share of population and of the high school dropout population



Note: We calculate the share of Cuban people as all those who define themselves as "Cuban" in the ethnicity question of the CPS. The population considered is the total number of individuals between 16 and 61. The high school dropout population is constituted by those who do not have a high school degree between 16 and 61. The vertical dashed bar is drawn at 1993.

Figure A6:
The Non-Event of 1994, unemployment in Miami and in the Synthetic control

Panel A Panel B



Note: Each Panel shows the behavior of the outcome variable for Miami (solid line) and for the synthetic control (dashed line) in the period 1989-2001. The vertical line corresponds to year 1993, immediately before the non-event of 1994. The variables are noted in the title of each panel. The sample here includes all non-Cubans, in the labor force, age 16-61 either of Black ethnicity (Panel A) or of Hispanic ethnicity (Panel B).