Supplementary analysis for the blog post, 'What Economists Can Learn from the Mariel Boatlift: Answering Questions about Our Research'

Michael A. Clemens, May 30, 2017

This supplement contains a more detailed discussion of the issues in the accompanying blog post, for economists familiar with the debate. Code to replicate these results is provided in the blog post.

Begin by considering what is not in dispute: 1) The fraction of blacks in the sample skyrocketed: it doubled between the control period and the treatment period, while nothing like this happened in the control cities. 2) Blacks' wages were much lower than non-blacks' wages. These two unchallenged facts require, arithmetically, that some portion of the measured decline in wages is spurious. The question for research is not whether Borjas's paper was spuriously attributing some decline in wages to the Mariel Boatlift, but how large that spurious decline was.

Reproducing our results in the Borjas regression framework

The substantive disagreement in earlier blogging on our results was how to control for the difference between blacks' wages and non-blacks' wages. If one uses the average black-nonblack wage gap *across all cities and education levels*, this removes about one third of the original treatment effect. This is what Borjas showed, and Tables 1 and 2 below replicate exactly. This by itself implies large bias in the original results.

But the assumption that the black-nonblack wage gap is identical across all cities and education levels is a very strong one. Empirically, that gap varies greatly in the CPS data across cities, and more so at some education levels. (We show this in Table 2 of our paper.) There are also theoretical reasons not to impose a nationally-homogeneous wage gap. The racial composition of less-than-high-school workers varies drastically across cities. In Miami, the black fraction of the Borjas sample of less-than-high-school workers is 63%, whereas in the Borjas control cities it is under 10% (Table 3, below). A high concentration of low-skill black workers has important implications for neighborhood segregation and occupational segregation of black workers, which would tend to affect the black-nonblack wage gap.

If one allows the black-nonblack wage gap to differ across cities (but not across education groups), the treatment effect is severely attentuated and no longer statistically significant in the ORG data (Table 2 below, columns 5–6). The ORG extract samples are much larger and more reliable than the March CPS extract (Table 1 of our paper). Even in the March CPS, the treatment effect is no longer statistically significant relative to the Card control cities (Table 1 below, columns 5–6). So at this stage, the treatment effect is only statistically distinguishable from zero in one of the two representative datasets (the smaller one), and in that dataset, relative to control cities selected by Borjas but not by Card.

Because the massive shift in race composition of the sample only happens in Miami and only among workers with less-than-high-school, the appropriate way to control for the wage effect of that shift is to allow the black-nonblack wage to vary by city *and* education level. This is done in the final two columns of Tables 1 and 2 below. Those results cannot reject a –5% to –8% wage effect of the Mariel Boatlift on this small subpopulation, but they cannot reject a zero effect either. This is what we show in Figure 4d of our paper; the only substantive difference between that figure and the regressions here is that here we residualize wages including age-range dummies, and in our Figure 4d we do not. But this discussion substantively harmonizes our result with the regression framework in Borjas's paper.

Additional evidence

As we discuss at length in our paper, there is an important caveat to the above findings. It is hypothetically possible that the entire effect of the Mariel Boatlift fell upon black men with less than high school, not on Hispanics and not on whites or Asians at that education level, or any workers who finished high school. This is unlikely, but cannot be definitively ruled out. If that is the case, columns 7 and 8 in the above tables control away the effect of interest. By allowing the black indicator to take a Miami-specific value for black workers with less than high school, columns 7 and 8 are effectively measuring the effect of the Boatlift on non-Hispanic whites and Asians. If the effect on blacks in that skill group was far greater, that differential effect would be missed by those regressions. In weighing the likelihood of this scenario, consider several points.

First, as mentioned above, the significance of the treatment effect disappears *without* an education-specific black indicator, in the larger of the two datasets (Table 2, cols. 5–6). In the other dataset, the effect remains significant relative to control cities picked by Borjas, not by Card (Table 1, cols. 5–6).

Second, as we discuss in our paper, the race-composition change explains numerous anomalies in the original findings that have not been adequately explained. Why does there seem to be such a large wage effect years after the increased supply of Cuban workers had subsided (that is, after 1984), in Table 1 cols. 1–2? If Cubans competed for jobs so severely that wages collapsed 30–40%, why was there no effect on unemployment? Why do the same regressions show no effect on Hispanics, who would plausibly compete with Cubans at least as much as anyone else? Why do the same regressions show that the Boatlift *raised* the wages of workers with high-school only, when half the Mariel Cubans had a high school degree (Table 4 below)?

All of these can be explained by the large shift in racial composition of the sample. The shift in racial composition lasted long after 1984, which can explain why the wage decline similarly lasted, even though the increase in the supply of Cuban labor did not. There was no difference between black and nonblack unemployment in Miami at this time, so there would be no direct compositional effect on unemployment corresponding to the spurious wage effect. This can explain why there is a wage decline but no unemployment decline. It can explain why no effect is seen on closely-competing Hispanics; there is no evidence of a large change in sample coverage of Hispanics. And it can explain why the Boatlift, where half of the migrants has a high school degree, appears to cause a rise in the wages of workers with high school only: that may not reflect any real, beneficial effect of the Boatlift, but simply the fact that the number of blacks in the high-school-only sample falls at the same time. All of this is in our paper already.

Third, we discovered additional *direct* evidence of negative selection into the sample recently, thanks to a helpful reviewer of our work. In the March CPS data, the Boatlift appears to cause a decline in years of schooling within workers who have less than high school. Table 5 below runs exactly the Borjas differences-in-differences regression for less-than-high-school workers, on the same March CPS sample, but rather than wage the outcome is years-of-education. There appears to be a substantial negative 'effect' of the Boatlift. But years-of-schooling is predetermined among these workers (age 25+, with no high school), so this effect cannot be real. It could only arise because something is negatively selecting workers into this sample: workers with very low education and earnings. That would corroborate changes in survey coverage that include lower-wage workers at the margin, or the simultaneous immigration of Haitians. In our paper we call these 'mechanism *B*' and 'mechanism *C*'. Table 5 below constitutes direct but not conclusive evidence that those mechanisms were active. ■

TABLE 1: MARCH CPS, REANALYSIS OF BORJAS TABLE 5, WORKERS WITH LESS THAN HIGH SCHOOL

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Add indicator for black					
	Borjas replication		Nationwide		By city		By city-less than HS	
Control cities:	*Card*	*Borjas*	*Card*	*Borjas*	*Card*	*Borjas*	*Card*	*Borjas*
1981–1983	−0.204***	−0.290***	−0.121	−0.194***	−0.0964	−0.174**	0.000854	−0.0785
	(0.0758)	(0.0734)	(0.0776)	(0.0698)	(0.0815)	(0.0722)	(0.0620)	(0.0605)
1984–1986	−0.368***	−0.454***	−0.202***	−0.301***	−0.137	−0.227***	0.109*	−0.00131
	(0.0601)	(0.0587)	(0.0721)	(0.0527)	(0.0828)	(0.0580)	(0.0627)	(0.0591)
1987–1989	−0.329***	−0.303***	−0.202**	−0.237***	−0.135	−0.149**	−0.0247	−0.0492
	(0.0810)	(0.0725)	(0.0928)	(0.0708)	(0.0886)	(0.0658)	(0.0977)	(0.0705)
1990–1992	−0.0259	−0.0561	0.0938	0.0248	0.105	0.0375	0.220	0.121
	(0.0724)	(0.123)	(0.0859)	(0.117)	(0.0895)	(0.112)	(0.134)	(0.103)
N	75	75	75	75	75	75	75	75

Unit of observation is city-period. Dependent variable is log real wage. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample is working non-Hispanic males age 25–59 with less than high school. 'HS' is high school.

TABLE 2: CPS-ORG, REANALYSIS OF BORJAS TABLE 5, WORKERS WITH LESS THAN HIGH SCHOOL

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Add indicator for black					
	Borjas replication		Nationwide		By city		By city-less than HS	
Control cities:	*Card*	*Borjas*	*Card*	*Borjas*	*Card*	*Borjas*	*Card*	*Borjas*
1981–1983	−0.0753***	−0.140***	−0.0466	−0.104**	0.00490	−0.0561	0.0361	−0.0252
	(0.0262)	(0.0494)	(0.0288)	(0.0461)	(0.0339)	(0.0402)	(0.0341)	(0.0440)
1984–1986	−0.0690	−0.116*	−0.0240	−0.0792	−0.00149	−0.0533	0.0162	−0.0331
	(0.0568)	(0.0655)	(0.0666)	(0.0595)	(0.0630)	(0.0474)	(0.0514)	(0.0473)
1987–1989	−0.106***	−0.175***	−0.0744*	−0.137**	−0.0460	−0.101	0.0511	0.0152
	(0.0365)	(0.0639)	(0.0393)	(0.0635)	(0.0419)	(0.0666)	(0.0530)	(0.0739)
1990–1992	0.0191	−0.0700	0.0694	−0.0412	0.105*	−0.00450	0.162**	0.0394
	(0.0408)	(0.0620)	(0.0442)	(0.0699)	(0.0552)	(0.0756)	(0.0611)	(0.0780)
N	75	75	75	75	75	75	75	75

Unit of observation is city-period. Dependent variable is log real wage. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample is working non-Hispanic males age 25–59 with less than high school. 'HS' is high school.

TABLE 3: FRACTION BLACK, BY CITY, IN BORJAS SUBSAMPLE

Miami-Hialeah, FL	0.63
Card control cities	
Atlanta, GA	0.30
Houston-Brazoria,TX	0.29
Los Angeles-Long Beach, CA	0.19
Tampa-St. Petersburg-Clearwater, FL	0.21
Borjas control cities	
Anaheim-Santa Ana-Garden Grove, CA	0.03
Nassau-Suffolk, NY	0.12
Rochester, NY	0.13
San Jose, CA	0.05

Sample is working non-Hispanic males age 25–59 with less than high school.

TABLE 4: REANALYSIS OF BORJAS TABLE 5, WORKERS WITH HIGH SCHOOL ONLY

	(1)	(2)	(3)	(4)
	March CPS		CPS-ORG	
Control cities:	*Card*	*Borjas*	*Card*	*Borjas*
1981–1983	0.109***	0.0835	0.0517***	0.00504
	(0.0402)	(0.0662)	(0.0177)	(0.0161)
1984–1986	0.0531	0.0442	0.0453	0.0200
	(0.0761)	(0.0719)	(0.0327)	(0.0277)
1987–1989	0.0756	0.00264	0.0318	−0.0330
	(0.0461)	(0.0570)	(0.0355)	(0.0247)
1990–1992	0.106***	0.0356	0.0326	−0.0187
	(0.0342)	(0.0662)	(0.0244)	(0.0227)
N	75	75	75	75

Unit of observation is city-period. Dependent variable is log real wage. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample is working non-Hispanic males age 25–59 with high school only.

TABLE 5: REANALYSIS OF BORJAS TABLE 5, WORKERS WITH LESS THAN HIGH SCHOOL: FALSIFICATION TEST WITH YEARS OF EDUCATION AS OUTCOME VARIABLE

	(1)	(2)
	March CPS	
Control cities:	*Card*	*Borjas*
1981–1983	−0.348	−0.124
	(0.446)	(0.573)
1984–1986	−0.444	−0.142
	(0.267)	(0.514)
1987–1989	−0.562**	−0.422
	(0.230)	(0.418)
1990–1992	0.0578	0.384
	(0.964)	(0.936)
N	75	75

Unit of observation is city-period. Dependent variable is years of education. Robust standard errors in parentheses: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample is working non-Hispanic males age 25–59 with less than high school.