The Fallacy of Crowding-Out: A Note on "Native Internal Migration and the Labor Market Impact of Immigration"*

Giovanni Peri (University of California, Davis and NBER)

Chad Sparber (Colgate University)

January 2008

Abstract

In "Native Internal Migration and the Labor Market Impact of Immigration," George Borjas (2006) identifies a strong negative correlation between immigration and native-born employment in the US using local area data. This relationship is particularly strong at the metropolitan area level, weaker but still significant at the state level, and weakest at the Census region level. In this note, we show that Borjas's negative correlation arises due to the construction of the dependent and explanatory variables rather than from any true negative association between the employment growth of immigrants and natives. Borjas regresses log native employment, $\ln(N_t)$, on the *share* of foreign-born employment, $p_t = M_t/(M_t + N_t)$, across skill-state-year cells. The specification therefore includes native employment in the numerator of the dependent variable and in the denominator of the explanatory variable. This builds a negative correlation into the model that is particularly strong if the variance of N_t relative to M_t is large. To illustrate, we first show that state and city level regressions of the standardized native employment change, $(N_{t+10}-N_t)/(M_t+N_t)$, on standardized immigration, $(M_{t+10} - M_t)/(M_t + N_t)$, always find a positive and mostly significant correlation between the two. Second, we randomly simulate changes in the native and foreign-born workforce with a data generating process that has zero or positive correlation between the shocks ΔM_t and ΔN_t (i.e., so that immigration is associated with either no change or an increase in native employment). Borjas specifications employing this simulated data estimate large and significantly negative coefficients as long as the variance of ΔN_t is larger than the variance of ΔM_t, which is true in observed state-level and city-level data.

Key Words: Immigrants, Crowding Out, Employment Effects.

JEL Classification Codes: J61, R23

*Addresses: Giovanni Peri, Department of Economics, UC Davis, One Shields Avenue, Davis, CA, 95616. email: gperi@ucdavis.edu. Chad Sparber, Department of Economics, Colgate University, 13 Oak Drive, Hamilton, NY, 13346. email:csparber@mail.colgate.edu.

1 Introduction

There is a long-standing debate among academics on whether immigration reduces the employment opportunities of natives. Economic analyses often exploit the wide variation in immigration rates across US regions and skill groups to identify whether immigration is associated with low native employment growth (due to internal migration or job displacement) across skill-state (or skill-city) cells. Though this correlation cannot definitively identify the effects of immigration (since causality is unclear and there may be omitted variables bias), researchers often cite such results as *prima facie* evidence for or against the crowding-out theory.

Most of the literature on US immigration across local labor markets finds little to no impact of immigration on wages, with even smaller effects on the employment of native workers.[1] That is, immigration at the state and city level does not seem to reduce employment growth of native workers, even when looking within narrowly defined skill groups or after controlling for demand shocks. However, Borjas's (2006) "Native Internal Migration and the Labor Market Impact of Immigration" seems to contradict this view by finding a strong and significantly negative correlation between immigration and native employment across US states and metropolitan areas. Moreover, the paper argues that this native out-migration, which becomes stronger for smaller geographical units, explains the attenuated effects of immigration on wages observed at the city and state level but not at the national level. The paper claims that by absorbing confounding factors and pre-determined trends with an array of skill-region-time fixed effects, a negative correlation between immigration and native employment does emerge, thereby demonstrating a crowding out effect.[2] For every ten immigrants in a metropolitan area, six natives leave (or loose their job). For every ten immigrants in a state, three natives leave (or loose their job).[3]

In this note, we demonstrate that the very large crowding out effects in Borjas (2006) are a consequence of model misspecification. Regressions of native employment (within a skill-region group) on the share of foreign born (in the same group) build a strong negative correlation into the model since native employment enters in the numerator of the dependent variable and the denominator of the explanatory variable. Even if there is a positive correlation between the inflow of new immigrants and the change in native employment, the Borjas specification is likely to find a negative coefficient as long as there is variation of native employment for any other reason. We employ two strategies to illustrate this point.

First, regressions of the change of native employment on the change of foreign-born employment fail to uncover any negative effect of immigration. In most specifications, the estimated coefficient is actually positive and significant. This result holds in models with (or without) controls for the thorough set of dummy variables, common trends, and lagged economic conditions advocated by Borjas (2006).

[1] See Card (2001), Card and Lewis (2007), and Ottaviano and Peri (2007).

[2] Note that Borjas (2006) is based simply on correlations (no exogenous immigration shock is identified), but it employs fixed effects to control for many potential labor demand shifters.

[3] See page 243 and Panels II and III of Table 3 in Borjas (2006).

Second, we generate hypothetical native and foreign-born employment growth data in which the two variables are, in turn, uncorrelated, positively, or negatively correlated. We then show that the Borjas empirical specification continues to find a false negative effect even in the case of zero or *positive* correlation in the data, while our specification estimates either a zero, positive, or negative effect, depending on the actual correlation in the data-generating process. Importantly, the negative coefficient estimated using the Borjas method increases in absolute value as the variance of native employment rises relative to the variance of immigration, no matter what the actual correlation between native and immigrant employment is. Thus, if cities exhibit greater relative variance than states do (as might be expected in the data), then the magnitude of Borjas's negative association will be larger when analyzing metropolitan areas than it will be for states.

The remaining elements of this note are organized as follows. Section 2 describes the empirical specification used in Borjas (2006) and our alternative specifications. Section 3 presents and discusses the estimated effect of immigration on native employment using these specifications. We reproduce Borjas's estimates and develop ours using skill-state and skill-city cells as units of observations over the period 1960-2000 and 1980-2000, respectively. Section 4 generates simulated employment data across skill-state cells with zero, positive, or negative correlation between immigration and native employment. We then present the estimated coefficients for regressions using the simulated data. Section 5 refers to recent studies analyzing the impact of immigration on native employment across US regions, and we argue that each uses a model similar to our preferred specification. Moreover, none finds significant evidence of displacement. Section 6 briefly concludes.

2 Empirical Specifications

Equation (1) reproduces the main empirical specification in Borjas (2006).

$$\ln(N_{ijt}) = s_i + r_j + \pi_t + (s_i \times r_j) + (s_i \times \pi_t) + (r_j \times \pi_t) + \theta_N p_{ijt} + \beta X_{ijt} + \varepsilon_{ijt} \tag{1}$$

The variable N_{ijt} measures the total employment of native workers in skill group j (32 education by experience groups), state i, and Census year t. The terms s_i, r_j, and π_t control for skill, state, and year fixed effects. The three subsequent terms control for any two-way interactions between these effects. The variable $p_{ijt} = M_{ijt}/(M_{ijt} + N_{ijt})$ is the share of foreign-born workers in skill group j, state i, and year t. Finally, X_{ijt} represents controls that include, depending on the specification, the lagged level or the growth rate of native employment.

Borjas's regression generated three important contributions to the crowding out literature. First, he loosely derives Equation (1) from a structural model in which the relevant feature in the empirical specification is the inclusion of the fixed effects and their interactions. These, along with the lagged variables in X_{ijt}, control for pre-existing conditions specific to skill-state groups and other employment determinants so that regressions isolate the direct impact of new immigrants on native employment growth. Second, as skill by state effects are introduced, the coefficient θ_N is identified by the variations over time within narrowly defined skill-region cells. This should directly identify the effect of immigrants on the group of natives most closely competing with them for jobs. Third, relative to previous analyses, it uses a longer panel of states and metropolitan areas over time.

One potential problem with Specification (1) is that the construction of the dependent and explanatory variables may mechanically force a negative correlation between the inflow of immigrants and the change in native employment when none exists. The explanatory variable p_{ijt} contains N_{ijt} in its denominator, and $\ln(N_{ijt})$ is a monotonic positive transformation of N_{ijt}. If we imagine a case in which the number of immigrants (M_{ijt}) does not vary much across observations and is totally uncorrelated with N_{ijt}, but N_{ijt} varies significantly across observations, then there would still be a negative correlation between $\ln(N_{ijt})$ and p_{ijt} purely driven by the variation of N_{ijt}. In fact (as we show systematically in Section 4), larger variation of N_{ijt} (relative to M_{ijt}) is associated with larger (in magnitude) negative estimates of θ_N. The presence of fixed effects reduces the variation used to identify the coefficient, but it does not eliminate this problem.[4] While the problem with specification (1) is somewhat reminiscent of the "division bias" (emphasized, by the way, in Borjas 1980) it is much more severe in this context. In fact the correlation that we would like to estimate is between the change in employment of immigrants $\Delta M_{ijt} = M_{ijt} - M_{ijt-10}$ and the change in employment of natives $\Delta N_{ijt} = N_{ijt} - N_{ijt-10}$. The presence of N_{ijt} in the denominator of the explanatory variable p_{ijt} serves only to standardize the change in immigrants and is not needed to compute the relevant variable. Moreover, it induces the bias in any case (not only when N_{ijt} is measured with error as for the 'division bias" in Borjas, 1980) as long as the variable N_{ijt} varies over time for reasons independent of immigration. One, however, can easily think of a specification that is exempt from this bias. Consider, for instance, that we identify the correlation between immigrants and native workers through the more commonly used regression in Equation (2).

$$\ln N_{ijt} = s_i + r_j + \pi_t + (s_i \times r_j) + (s_i \times \pi_t) + (r_j \times \pi_t) + \theta_N \ln M_{ijt} + \beta X_{ijt} + \varepsilon_{ijt} \tag{2}$$

The inclusion of skill by state fixed effects imply that θ_N is identified on the percentage (logarithmic) changes of natives and immigrants within cells over time. The presence of all the fixed effects and their interactions guarantees that this specification controls for pre-determined conditions and other determinants of native employment as was done in (1). Hence, a structural model similar to that in Borjas (2006) would also support Regression (2). In this case a negative θ_N would genuinely indicate that skill-state groups experiencing large percentage inflows of immigrants also encounter native employment decreases (or smaller increases).

[4]Similarly, the construction of the dependent variable as the net native migration in a state obtained from data on the self-reported state of residence 5 years ago (as done in Section V.B of Borjas (2006)), fails to solve the problem as long as the explanatory variable on the right hand side is equal to the share of immigrants in total employment.

Some would argue that Regression (2) is not ideal either. It might find a positive correlation between $\ln M_{ijt}$ and $\ln N_{ijt}$ due to scale effects. Namely, some skill-state groups may be much larger than others (because of state size), and a positive correlation in the size of native and immigrant employment may produce a positive estimate of θ_N. Certainly skill by state fixed effects and the measurement of variables in logarithms should mitigate this problem, but it might not solve it altogether. Moreover, when estimating Equation (2) one needs to decide how to deal with observations in which $M_{ijt} = 0$ (or $N_{ijt} = 0$).[5] Hence, our preferred specification is obtained by considering employment changes over a decade within skill-state cells and standardizing them by the size of the skill-state group at the beginning of the period. One can express employment changes for natives and immigrants as percentages of the initial total employment in the group and run Regression (3).

$$\Delta n_{ijt} = s_i + r_j + \pi_t + (s_i \times r_j) + (s_i \times \pi_t) + (r_j \times \pi_t) + \theta_N \Delta m_{ijt} + +\beta X_{ijt} + \varepsilon_{ijt} \tag{3}$$

In this equation, $\Delta n_{ijt} = \Delta N_{ijt}/(N_{ijt-10} + M_{ijt-10})$ with $\Delta N_{ijt} = N_{ijt} - N_{ijt-10}$. Similarly, $\Delta m_{ijt} = \Delta M_{ijt}/(N_{ijt-10} + M_{ijt-10})$ with $\Delta M_{ijt} = M_{ijt} - M_{ijt-10}$. Specification (3) has several advantages. First, it directly measures the displacement effect of interest. That is, it measures how native employment responds to one extra immigrant worker in the group. Second, it is not affected by the size of the group (as the changes are standardized by its initial employment), and it does not build in any correlation between the dependent and explanatory variable. Third, it is even more demanding than (1) in controlling for pre-determined trends of native employment growth. The variables are expressed in differences and we control for skill by region effects. The identifying variation comes from deviations of growth rates from skill-state specific trend growth while accounting for region-time effects, skill-time effects, and past levels or growth rates. Finally, the regression avoids dropping observations with zero foreign-born workers that would be lost in Regression (2).[6]

3 Estimation Results

3.1 States

We begin by employing data with variable definitions and sample choices as close as possible to Borjas (2006) to estimate Specifications (1), (2), and (3) across skill-state groups over Census years 1960-2000. We analyze all 50 states plus the District of Columbia, and we use 32 skill groups representing four education groups (High School Dropouts, High School Graduates, Some College Experience, and College Graduates) by eight experience groups (from 0 to 40 years of experience in five-year intervals). Individuals in the sample include only 18-64 year old individuals not residing in group quarters who worked the previous year.

[5] In our empirical estimates of (2), we drop all the skill-state observations for which this occurs in (at least) one year.

[6] Note that θ_N also becomes a direct measure of $\partial M/\partial N$ that does not need the conversion formula (17) used on page 243 of Borjas (2006).

Column I of Table 1 reports the OLS estimates of θ_N according to the Borjas specification in (1). Columns II and III report estimates from the alternate specifications in (2) and (3). In Column IV we estimate our preferred specification (3), weighting each observation by the employment in the corresponding cell.[7] Each regression includes year, skill, and state fixed effects in addition to all possible two-way interactions of those effects. The first three rows report the estimates obtained when employment figures are based upon all workers, while rows 4 through 6 report the estimates when using only male workers.

The results in Column I of Table 1 should be close to those in Panel II of Table 3 in Borjas (2006), though some differences may arise due to slight deviations in the selection of people in the samples. The first row includes only the fixed effects as controls. The second row controls for lagged native employment, $\ln(N_{ijt-10})$, and the third row controls for lagged employment growth of native workers, Δm_{ijt-10}. Rows four, five, and six repeat the same specifications using data on men only. The estimated values of θ_N are negative, significant, and larger in magnitude than those reported in Borjas (2006). Our estimates of θ_N are between -0.52 and -0.60 when using men only, and between -0.40 and -0.58 for all workers. The corresponding values for Borjas range between -0.27 and -0.38 for the estimates based upon male workers, and between -0.21 and -0.30 for those including all individuals. Our standard errors are between 0.09 and 0.11, and Borjas's are around 0.09. In both papers, all estimates are highly statistically significant. The inclusion of lagged employment of natives or lagged employment growth does not produce relevant differences in the results.[8]

Important lessons emerge from Table 1 by comparing the stunning differences between the estimates of Column I and those in the other columns. Unlike the Borjas specification, the estimates of θ_N in Columns II to IV are always positive, almost always significant, and, especially for our preferred specification reported in Column IV, also quite stable across regressions (with estimates averaging around 0.35 in value). Neither the inclusion of fixed effects (compare the last row with others), nor the choice of specification (including or not lagged employment) nor the weighting, can explain the difference between the significantly negative estimates in Column I and the significantly positive ones in II, III, and IV. Instead, it is the variable definitions that matter. The coefficients in Column IV now, interpreted as done in Borjas (2006), show that an increase of ten new foreign-born workers in a skill-state group is associated with an additional 3 to 4 extra jobs for natives. While one may argue that the fixed effects included in the regressions still do not capture all demand shocks, and that

[7] The choice to perform unweighted least squares estimation in Columns I through III follows footnote 28 in Borjas (2006) in which he argues that OLS is preferable to weighted least squares. This results, however, in a disproportionately large influence of small cells, whose employment is much more volatile. Thus, we also estimate (in Column IV) a weighted least square regression of specification (3). This does not have the problem mentioned by Borjas because larger cells (that receive large weight) are not systematically associated with larger values of the dependent variable.

[8] We are currently unsure why sizeable differences between our estimates (in Column I) and Borjas's exist; these should be identical specifications. The differences are not qualitatively substantial, but are quantitatively large. To be sure that coding errors did not affect our estimates, the two coauthors have independently performed regressions on independently extracted IPUMS data using different Stata commands (xtreg for panel estimation with fixed effects, or simple reg including all the dummies manually). The estimated coefficients only differ by very small amounts from each other. Access to the original Borjas data would be necessary to identify equivalent effects. Our main point, however, does not require the use of that dataset. Our qualitative findings using the Borjas method are similar to his and suggestive, if anything, of an even larger crowding out effect.

the correlations are driven by demand shifts, we still systematically find better employment performance for natives in skill-state groups with larger inflows of immigrants.

Simple summary statistics of the data reveal an additional important fact. The standard deviation of employment changes for natives (Δn_{ijt}) over the whole sample is 0.7 (the weighted standard deviation equals 0.6). The standard deviation for employment changes of immigrants (Δm_{ijt}) is 0.1. Similarly, the standard deviation of native employment across cells (N_{ijt}) is around 61,000, while for immigrant employment (M_{ijt}) it is around 15,000. Whether in levels, changes, or standardized changes, the variation of natives across cells largely dominates that of immigrants. Section 4 will demonstrate that a consequence of this phenomenon is that the built-in negative correlation from the Borjas specification will be very large.

3.2 Metropolitan Areas

The metropolitan-level data description in Borjas (2006) is brief, so it is impossible for us to reproduce his sample. In particular, the paper never indicates which metropolitan areas are included. Area definitions (in terms of counties included) and Census geographic identifiers vary over time. The IPUMS dataset variable METAREA identifies a subset of large metro-areas that is consistent across years, but it only covers 90 metropolitan areas consistently in 1980, 1990, and 2000 (see Table A1). Borjas uses the same three Census years but many more metropolitan areas (from his number of observations, we can infer that he uses 232 of them).

Table 2 reports the results paralleling those in Table 1, now run on a panel of 32 skill groups over 90 cities in three Census years (1980-2000). Each regression includes metro area, skill, and year fixed effects, plus all two-way interactions. While the short panel structure prevents us from running all the regressions with lagged variables, the overall results are very clear. Again the Borjas method produces negative, large, and significant estimates of θ_N, while the other three methods produce positive and significant results. Moreover, our preferred specification (IV) produces an estimate of θ_N equal to 0.48 that is similar to the figures (between 0.32 and 0.38) identified when using the state panel.

Also note that the coefficient magnitudes from our city-level Borjas specifications (Column I of Table 2) are not very different from those of their state-level counterparts (Column I of Table 1). This is not surprising. Our dataset includes only large metropolitan areas. As a consequence, the standard deviation of Δn_{ijt} (0.4) in our metro sample is similar to that of our state sample (0.7), while the standard deviation of Δm_{ijt} is near 0.10 in both. This implies that the estimates of θ_N in Column I should be close to each other as well. Borjas's sample, in contrast, includes many small cities. Thus, it is likely that Δn_{ijt} varies more in his metro sample than in his state sample, which would generate more sizeable negative estimates of θ_N. We believe this may explain why his metropolitan-level results are negative and large in absolute value.[9]

[9]See Table 4 of Borjas (2006).

3.3 Employment Rate as the Dependent Variable

Though not discussed in Borjas (2006), other variants of Regression (1) in the literature[10] use the employment rate rather than employment as the dependent variable. In some specifications, the explanatory variable is the immigrant share of the population (rather than the share of employment). Equations (4) and (5) express these two common modifications of (1).

$$\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}} = s_i + r_j + \pi_t + (s_i \times r_j) + (s_i \times \pi_t) + (r_j \times \pi_t) + \theta_N \frac{M_{ijt}^{Empl}}{M_{ijt}^{Empl} + N_{ijt}^{Empl}} + \beta X_{ijt} + \varepsilon_{ijt} \tag{4}$$

$$\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}} = s_i + r_j + \pi_t + (s_i \times r_j) + (s_i \times \pi_t) + (r_j \times \pi_t) + \theta_N \frac{M_{ijt}^{Pop}}{M_{ijt}^{Pop} + N_{ijt}^{Pop}} + \beta X_{ijt} + \varepsilon_{ijt} \tag{5}$$

N_{ijt}^{Empl} and N_{ijt}^{Pop} are the employment and population of natives (aged 16 to 65) in skill group i, state j, and period t. $M_{ijt}^{Empl}$and M_{ijt}^{Pop} represent the same figures for immigrants. The native variables (N_{ijt}^{Empl} or N_{ijt}^{Pop}) form both part of the dependent and explanatory variables and have the potential to induce spurious correlations. Though the direction of the bias is not immediately obvious, careful examination reveals that these specifications are also subject to negatively biased coefficients. This problem arises because there is a significant positive contemporaneous correlation between $\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}}$ and both N_{ijt}^{Empl} and N_{ijt}^{Pop}.[11] and it could generate negative estimates of θ_N even if $\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}}$ is uncorrelated with M_{ijt}^{Empl} and M_{ijt}^{Pop}. Specification (5) should be less subject to this problem than (4) will, as the former does not mechanically include the numerator of the dependent variable in the denominator of the explanatory variable.

In order eliminate the negative bias, we propose regressions of the change in the employment rate of natives $\left(\Delta\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}} = \frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}} - \frac{N_{ijt-10}^{Empl}}{N_{ijt-10}^{Pop}}\right)$ on the change in the immigrant population (or employment) standardized by its initial level, as in Equation (6).

$$\Delta\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pgop}} = s_i + r_j + \pi_t + (s_i \times r_j) + (s_i \times \pi_t) + (r_j \times \pi_t) + \theta_N \Delta m_{ijt} + \beta X_{ijt} + \varepsilon_{ijt} \tag{6}$$

Table 3 reports the estimates of θ_N employing 1960-2000 state-skill data and using Borjas's and our methods. The explanatory variable in the first row measures the share of immigrants in the population; the second row uses employment figures. In the first column of Table 3, we see that the specification measuring immigrants as a share of the population shows only a small negative coefficient estimate (-0.04), while the one with employment

[10] The employment regressions reported in Table III of Borjas (2003) are examples of such regressions. They use, in fact, "fraction of time worked" as a measure of employment which is calculated relative to the total population in a cell, (and is therefore a proxy of the employment-population ratio in a cell) and share of foreign-born in employment as explanatory variable.

[11] Regressing the change in $\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}}$ on the change in N_{ijt}^{Pop} across state-skill-time groups reveals an OLS coefficient of 0.72 with a standard error of 0.08. Regressing $\frac{N_{ijt}^{Empl}}{N_{ijt}^{Pop}}$ on N_{ijt}^{Emp} generates a coefficient of 0.69 with a standard error of 0.08. Intuitively, state-skill groups that receive many natives (for any demand-driven reason) also experience growing employment, population, and employment rates of natives.

figures still exhibits a large negative bias (-0.14). In contrast, the estimates of our preferred alternative exhibit a positive and significant coefficient when using the employment change of immigrants (as a share of initial employment), and an insignificant estimate when using the change in the immigrant population. We believe that the negative estimates of Column I arise due to the construction of the dependent and explanatory variables in Specifications (4) and (5).

4 Simulated Results

In this section we perform a simulation exercise to illustrate how the systematic differences in the estimates of Specifications I and IV (in Table 1 and 2) depend upon the construction of the variables and the variance of the shocks to native employment, ΔN_{ijt}.

First, we generate an identical initial value of native employment (equal to 93) and immigrant employment (equal to 7) for 1632 skill-state cells to mirror the average share of foreign-workers in 1960 (7% of employment). Then we generate random values for the employment growth of natives and immigrants over four subsequent decades using a mean-zero normal distribution. The random changes in immigrant employment have a standard deviation (standardized by total initial employment) equal to 0.10, which corresponds to the standard deviation of the immigrant employment change in the state sample. On the other hand, we simulate ΔN_{ijt} from a mean-zero normal distribution but with values for its standard deviation (when standardized by initial employment) ranging from 0.10 to 0.80 (reported in the first column of Table 4).

As we alter the data generating process, we consider four cases. In the first, there is no effect of immigration on native employment so that the correlation between ΔN_{ijt} and ΔM_{ijt} equals zero. The second case assumes a small positive effect of immigration and sets the correlation between ΔN_{ijt} and ΔM_{ijt} equal to 0.10. In the third scenario, the correlation equals 0.30 so that immigration has a large positive effect on native employment. Finally, in the fourth case we simulate data with a large and negative correlation between ΔN_{ijt} and ΔM_{ijt} equal to -0.30. Thus, we can think of the simulated skill-state groups as receiving random immigration shocks, with standard deviations similar to those measured in the decade-changes over 1960-2000, that also generate either a zero, positive, or negative native employment response.[12] To focus on our coefficient of interest (the relation between immigration and native employment in the same skill-state group) we use purely white noise immigration shocks, only allowing them (in Cases 2, 3, and 4) to induce a potential native employment response within that corresponding cell.[13]

[12] Note that we do not include any correlation in the shocks across skill groups, within a state, or over time. Those would be confounding effects, controlled for by the fixed effects and their interactions.

[13] We considered other data generating possibilities as well. In one case, we generated data with the actual initial employment distribution of natives and immigrants by cell in 1960, multiplied by random shocks for subsequent decades. In another, we allowed for systematic deterministic differences in growth rates of employment across groups. These features do not change the regression results as they are absorbed by controlling for an array of fixed effects.

Once we have generated these changes, we can construct the simulated variables M_{ijt} and N_{ijt} for each skill-state group in each year (1960-2000). We then use this simulated data to estimate Specifications (1) and (3) – Table 4 reports the values of θ_N obtained using Specification (1), labeled "Borjas Method," and Specification (3), labeled "Our Method." Case 1, reported in the first two columns, uses data generated with no correlation between ΔN_{ijt} and ΔM_{ijt}. We clearly see that while our method consistently delivers an estimate of θ_N insignificantly different from zero and quite precise (except for the case of the huge standard deviation of ΔN_{ijt} relative to ΔM_{ijt} in the last row), the Borjas method always uncovers negative and significant results. Furthermore, the Borjas estimates become larger in absolute value as the standard deviation of ΔN_{ijt} rises relative to the standard deviation of ΔM_{ijt} (which can be seen by proceeding downward in the third column of Table 4). In contrast, our method's estimates do not depart much from zero in any case. Even when ΔN_{ijt} has a small standard deviation (equal to 0.10 in the first row), the Borjas method estimates a negative correlation when there should be none. This confirms our intuition that the negative correlation between p_{ijt} and $\ln(N_{ijt})$ increases with the variance of N_{ijt} even when employment changes of natives and immigrants are uncorrelated.

Cases 2 and 3 are even more telling. In Case 2 when there is a small positive effect of immigration flows on native employment, the Borjas method finds a negative coefficient in all instances, most of which are significant (third and fourth columns of Table 4). As the variance of ΔN_{ijt} increases, the built-in negative correlation in the Borjas method further obscures the true positive correlation between N_{ijt} and M_{ijt}. On the other hand, our method correctly identifies a positive and very precisely estimated coefficient.[14] Case 3 shows that even when the positive correlation between ΔN_{ijt} and ΔM_{ijt} is large and equal to 0.30, the Borjas methodology continues to generate negative and significant estimates of θ_N (between -0.8 and -4.42) when the standard deviation of $\Delta N_{ijt}/(N_{ijt} + M_{ijt})$ is larger than 0.30. Since the empirically observed weighted standard deviation of $\Delta N_{ijt}/(N_{ijt} + M_{ijt})$ is 0.6 for states and 0.4 for cities, the results suggest important limitations of the Borjas specification in identifying the relevant correlation.

In Case 4, we set the correlation of ΔN_{ijt} and ΔM_{ijt} equal to -0.30 to check that our preferred method has no "positive bias", that is it would identify a negative θ_N in the presence of negative correlation in the data. The values of θ_N from our regression specification are always negative and precise. The figures from Borjas's method are negative as well.[15] Importantly, however, when the variance of the native employment change (ΔN_{ijt}) is large (last two rows), the built-in negative effect in the Borjas method becomes so dominant that it estimates large and negative coefficients regardless of the true correlation between ΔN_{ijt} and ΔM_{ijt}. To the contrary, our method gives the correct sign in each case. While our simulations are very simple and do not capture many features of real employment data (e.g., they do not allow persistence or skill-state specific factors and do not

[14] The coefficient increases with $\sigma_{\Delta N}$, the standard deviation of ΔN, because the OLS estimator of θ_N is equal to $\rho_{\Delta N \Delta M}(\sigma_{\Delta N}/\sigma_{\Delta M})$, where $\rho_{\Delta N \Delta M}$ is the correlation coefficient between ΔN_{ijt} and ΔM_{ijt}, while $(\sigma_{\Delta N}/\sigma_{\Delta M})$ is the ratio of their standard deviations.

[15] Note that the Borjas estimates of θ_N when ΔN_{ijt} and ΔM_{ijt} are negatively correlated (Case 4) are also larger in absolute value than when compared to the scenario in which the shocks exhibited zero correlation (Case 1).

have the skewness exhibited in the data), they clearly demonstrate that the construction of the dependent and explanatory variables has the potential of producing the negative estimates of θ_N in Borjas (2006) even when no negative correlation exists in the data. The controls for common trends and predetermined conditions in a complicated dynamic process do not add much to this basic story.

5 Literature

Most regional analyses find that immigration generates little to no native employment effects. It is instructive to note that those previous articles employ specifications similar to (2) or (3). For example, Cortes (2006) uses variants of (2) to analyze the link between immigration and employment of less educated workers across 75 metropolitan areas. She finds a positive OLS estimate around 0.20 and an IV value near 0.05.

Models similar to the specification in (3) begin with Card (2001), who uses population growth in a skill-city group as the dependent variable and the inflow rate of immigrants in the same city and skill group as the explanatory variable.[16] He always finds positive and sometimes significant effects on the native population (around 0.10). His subsequent IV estimates (using the shift-share instrument to impute the number of immigrants in a cell) often find results similar to those of his OLS regressions.[17]

Other examples of studies using a specification akin to (3) include Ottaviano and Peri (2007). They aggregate individuals from all skill levels within a state and estimate an impact of immigration on native employment between -0.3 and 0.3 that is never significant (standard errors around 0.3). Card and Lewis (2007) estimate the effect of low skilled Mexican immigrants on native employment. Their Table 5 results find an effect of low skilled immigrants on natives between 0 and 0.5 that is rarely significant. Card's (2007) Specification (2) adopts the total (immigrant and native) change in the less educated population (or employment) as the dependent variable. His estimated coefficient on Δm_{ijt} implies a θ_N value slightly larger than zero.

The previous literature, therefore, exhibits a clear preference for Specifications (2) and (3) over (1). Furthermore, these studies often find positive or no native employment effects of immigration. The present note simply applies those more common specifications to a five Census year panel of skill-region employment data and confirms that no evidence of a negative impact exists.

6 Conclusions

The debate on the labor market effects of immigration is important from an academic and policy point of view. It is essential that we have the most accurate understanding of the potential crowding-out of native workers

[16] See Equation (8), page 39, in Card (2001).
[17] All the articles described in this section produce IV estimates as well.

that immigrants might cause. Analysis of the correlation between immigrant and native employment within skill-state (or skill-city) groups over time might be able to provide or deny support to the crowding out theory. "Native Internal Migration and the Labor Market Impact of Immigration" by Borjas (2006) claims to find extremely large crowding out effects that are particularly sizeable at the metropolitan area level, but still very significant at the state level. This finding was used as an important piece of evidence in explaining the small wage effect of immigration in states and cities. It also purported to correct several previous studies by claiming that the crowding out effect is large once regressions control for long-run trends and predetermined conditions.

This note shows that controls for trends and predetermined conditions are not responsible for the large negative effects of immigration on native employment estimated in Borjas (2006). Instead, the construction of the explanatory variable as $M_{ijt}/(M_{ijt}+N_{ijt})$ coupled with the use of $\ln(N_{ijt})$ as the dependent variable builds a significant negative correlation into the model that is particularly strong when the variance of N_{ijt} is large relative to the variance of M_{ijt}. This bias, therefore, tends to grow as cells become small. We then show that alternative specifications popular in the literature mostly reveal a positive and significant correlation between immigration and native employment across skill groups in US states and metropolitan areas from 1960 through 2000, even when regressions include the rich set of controls.

References

Borjas, George. 1980. "The relationship between Wages and weekly Hours of Work: The Role of division Bias" *Journal of Human Resources*, 15 no3: 409-423.

Borjas, George. 2003. The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *Quarterly Journal of Economics* 118, no. 4: 1335-1374.

Borjas, George. 2006. "Native Internal Migration and the Labor Market Impact of Immigration." *Journal of Human Resources* 41, no.2: 221-258.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19, no. 1: 22-64.

Card, David. 2007. "How Immigration Affects U.S. Cities." CReAM Discussion Paper, no. 11/07.

Card, David, and Ethan Lewis. 2007. "The Diffusion of Mexican Immigrants During the 1990s: Explanations and Impacts." in Borjas, George , editor *Mexican Immigration to the United States*. National Bureau of Economic Research Conference Report, Cambridge Ma.

Cortes, Patricia. 2006. "The Effect of Low-Skilled Immigration on US Prices: Evidence from CPI Data." Manuscript. Massachusetts Institute of Technology, Cambridge, MA.

Ottaviano, Gianmarco, and Giovanni Peri. 2007. "The Effect of Immigration on U.S. Wages and Rents: A General Equilibrium Approach." CReAM Discussion Paper no 13/07 London, UK.

Table 1:
Estimated from US State Data, 1960-2000
(Table Entries are Estimates of θ_N)

	I Borjas Specification	II ln(N) regressed on ln(M)	III Δn regressed on Δm	IV Δn regressed on Δm (weighted)
All Workers	-0.58** (0.11)	0.03** (0.004)	1.30** (0.28)	0.38** (0.06)
All Workers, controlling for lagged native employment	-0.40** (0.09)	0.03** (0.005)	0.86** (0.21)	0.35** (0.05)
All Workers, controlling for lagged native employment growth	-0.41** (0.10)	0.02** (0.006)	1.13** (0.32)	0.36** (0.06)
Men	-0.60** (0.11)	0.02** (0.005)	2.16** (0.99)	0.36** (0.05)
Men, controlling for lagged native employment	-0.53** (0.11)	0.015** (0.006)	1.50* (0.80)	0.36** (0.07)
Men, controlling for lagged native employment growth	-0.52** (0.12)	0.004 (0.006)	2.25** (1.11)	0.32** (0.07)
Men, without skill by state fixed effects	-1.26** (0.10)	0.017** (0.007)	2.08** (1.01)	0.30** (0.05)

Note: Standard errors are heteroskedasticity robust and clustered by skill-state group. Each cell is from a separate regression. All regressions, except in the last row, include skill, state, and year fixed effects, plus skill by state, skill by year, and state by year interactions. The last column weights each observation by its employment. The total number of observations equals 32 skills X 51 states X 5 years = 8160 in Columns I and II, while we lose one year of observations when using differences in Columns III and IV. Moreover, when we include one lagged variable (Rows 2 and 4) we lose one further year of observations, and when we include one lagged growth rate we lose two additional years.

Table 2:
Estimated from 90 Large US Metropolitan Areas, 1980-2000 (Those in IPUMS)
(Table Entries are Estimates of θ_N)

	I Borjas Specification	II Ln(N) regressed on Ln(M)	III Δn regressed on Δm	IV Δn regressed on Δm (weighted)
All Workers	-0.35** (0.06)	0.02** (0.04)	0.14** (0.06)	0.48** (0.20)
All Workers, controlling for lagged native employment	-0.67** (0.06)	0.01* (0.005)	n.a.	n.a.
All Workers, controlling for lagged native employment growth	n.a.	n.a.	n.a.	n.a.

Note: Standard errors are heteroskedasticity robust and clustered by skill-city group. Each cell is from a separate regression. All regressions include skill, metro area, and year fixed effects, plus skill by metro area, skill by year, and metro area by year interactions. The last column weights each observation by its employment. As we only have 3 years of data and we include fixed effects, the regressions in levels cannot be run if we include lagged growth, and those in growth rates cannot be run if we include any lagged variables. The total number of observations equals 32 skills X 90 metropolitan areas X 3 years = 8640 in Columns I and II. In Columns III and IV we lose one year of observations as we take differences, and in Rows 2 and 3 one or two additional years are lost due to lagged variables. Thus, some estimates are not available due to insufficient observations.

Table 3:
Employment Rates as the Dependent Variable;
US State Data, 1960-2000
(Table Entries are Estimates of θ_N)

	I **Borjas Variants, Specification (4) and (5): Employment rates regressed on share of immigrants in employment (Specification (4)) or population (Specification (5)).**	**II** **Our Specification (6): Changes in employment rates regressed on changes in foreign employment (or population) as a share of initial group size.**
Immigrants in Employment	-0.14** (0.02)	0.30** (0.13)
Immigrants in Population	-0.047 (0.028)	0.19 (0.11)

Note: Standard errors are heteroskedasticity robust and clustered by skill-state group. Each cell is from a separate regression. All regressions include skill, state, and year fixed effects, plus skill by state, skill by year, and state by year interactions. Regressions weight each observation by its employment. The total number of observations equals 32 skills X 51 states X 5 years = 8160.

Table 4:
Estimates of θ_N from Randomly Generated Employment-Growth of Natives and Immigrants,

Standard Deviation of $(\Delta Mij_t)/(M_{ijt}+N_{ijt}) = 0.10$ (The Observed Value in the State and City Samples)

Standard Deviation of $(\Delta N_{ijt})/(M_{ijt}+N_{ijt})$	**Case 1** **No Correlation between** (ΔM_{ijt}) and (ΔN_{ijt})		**Case 2** **0.10 Positive Correlation between** (ΔM_{ijt}) and (ΔN_{ijt})		**Case 3** 0.30 **Positive Correlation between** (ΔM_{ijt}) and (ΔN_{ijt})		**Case 4** –0.30 **Negative Correlation between** (ΔM_{ijt}) and (ΔN_{ijt})	
	Our Method	Borjas Method	Our Method	Borjas Method	Our Method	Borjas Method	Our Method	Borjas Method
0.10	0.01 (0.02)	-0.26** (0.04)	0.11** (0.01)	-0.04 (0.03)	0.28** (0.04)	0.27** (0.02)	-0.30** (0.01)	-0.70** (0.02)
0.20	-0.04 (0.03)	-1.28** (0.07)	0.23** (0.03)	-0.64** (0.07)	0.57** (0.02)	0.20** (0.06)	-0.55** (0.03)	-1.67** (0.06)
0.30	0.01 (0.05)	-2.97** (0.13)	0.29** (0.04)	-2.56** (0.19)	0.77** (0.05)	-0.80** (0.16)	-0.98** (0.05)	-3.05** (0.10)
0.50	-0.01 (0.1)	-5.10** (0.05)	0.37** (0.06)	-4.88** (0.60)	0.62** (0.29)	-4.42** (0.11)	-0.83** (0.33)	-4.93** (0.04)
0.80	0.37 (0.29)	-5.70** (0.04)	1.68* (0.98)	-5.80** (0.04)	0.52** (0.19)	-5.80** (0.06)	-0.20** (0.08)	-5.57** (0.03)

Note: Number of observations generated is 1632 (to reproduce 32 skill groups by 51 states) in each of 5 Census years. The initial native employment equals 93, and the initial foreign employment equals 7 in each cell in the base year (this equals the average employment shares of natives and immigrants in 1960). We then generate random draws for (ΔM_{ijt}) and (ΔN_{ijt}) taken from mean-zero normal distributions with standard deviations and correlations reported in the table. The standard deviation of (ΔMij_t) standardized by initial employment in the group $(M_{ijt}+N_{ijt})$ is fixed at 0.10, which corresponds to the standard deviation of (ΔMij_t) in the actual US state sample from 1960-2000. The changes have no correlation across skills, states, and year groups; the only potential contemporary correlation is within groups between (ΔM_{ijt}) and (ΔN_{ijt}).
The estimates labeled "Our Method" are based on Specification (3) applied to the simulated data, while those labeled "Borjas Method" are based on Specification (1) and are also applied to simulated data. Each is an OLS regression that includes year, skill, and state fixed effects with all two-ways interactions but no further controls. Standard errors are clustered by skill-state group.

Appendix

Table A1
Metropolitan Areas Included in Our Sample

Name	METAREA Codes
Akron, OH	8
Albany-Schenectady-Troy, NY	16
Allentown-Bethlehem-Easton, PA/NJ	24
Atlanta, GA	52
Austin, TX	64
Bakersfield, CA	68
Baltimore, MD	72
Baton Rouge, LA	76
Birmingham, AL	100
Boston, MA-NH	112
Buffalo-Niagara Falls, NY	128
Canton, OH	132
Charleston-N.Charleston,SC	144
Charlotte-Gastonia-Rock Hill, NC-SC	152
Chicago, IL	160
Cincinnati-Hamilton, OH/KY/IN	164
Cleveland, OH	168
Columbia, SC	176
Columbus, OH	184
Dallas-Fort Worth, TX	192
Dayton-Springfield, OH	200
Denver-Boulder, CO	208
Detroit, MI	216
El Paso, TX	231
Fort Lauderdale-Hollywood-Pompano Beach, FL	268
Fort Wayne, IN	276
Grand Rapids, MI	300
Greensboro-Winston Salem-High Point, NC	312
Harrisburg-Lebanon--Carlisle, PA	324
Hartford-Bristol-Middleton- New Britain, CT	328
Honolulu, HI	332
Houston-Brazoria, TX	336
Indianapolis, IN	348
Jackson, MS	356
Jacksonville, FL	359
Kansas City, MO-KS	376
Knoxville, TN	384
Lancaster, PA	400
Las Vegas, NV	412
Little Rock--North Little Rock, AR	440
Los Angeles-Long Beach, CA	448
Louisville, KY/IN	452
Memphis, TN/AR/MS	492
Miami-Hialeah, FL	500
Milwaukee, WI	508
Minneapolis-St. Paul, MN	512
Nashville, TN	536
New York-Northeastern NJ	560
Memphis, TN/AR/MS	492
Miami-Hialeah, FL	500
Milwaukee, WI	508
Minneapolis-St. Paul, MN	512
Nashville, TN	536
New York-Northeastern NJ	560
Norfolk-VA Beach--Newport News, VA	572
Oklahoma City, OK	588
Omaha, NE/IA	592
Orlando, FL	596
Peoria, IL	612
Philadelphia, PA/NJ	616
Phoenix, AZ	620
Pittsburgh, PA	628
Portland, OR-WA	644
Providence-Fall River-Pawtucket, MA/RI	648
Richmond-Petersburg, VA	676
Riverside-San Bernadino, CA	678
Rochester, NY	684
Sacramento, CA	692
St. Louis, MO-IL	704
Salt Lake City-Ogden, UT	716
San Antonio, TX	724
San Diego, CA	732
San Francisco-Oakland-Vallejo, CA	736
San Jose, CA	740
Scranton-Wilkes-Barre, PA	756
Seattle-Everett, WA	760
Spokane, WA	784
Springfield-Holyoke-Chicopee, MA	800
Stockton, CA	812
Syracuse, NY	816
Tacoma, WA	820
Tampa-St. Petersburg-Clearwater, FL	828
Toledo, OH/MI	840
Tucson, AZ	852
Tulsa, OK	856
Ventura-Oxnard-Simi Valley, CA	873
Washington, DC/MD/VA	884
West Palm Beach-Boca Raton-Delray Beach, FL	896
Wichita, KS	904
Youngstown-Warren, OH-PA	932

Note: Metropolitan area definitions are consistent across census years and are described in the documentation for METAREA variable at www.ipums.org.