Abstract

Recent research finds that native-born Americans avoid settling in immigrant neighborhoods. We examine whether sorting stems from reductions in native demand for public education. Our analysis focuses on Mexican immigration to California, addresses endogeneity of immigrant inflows using established settlement patterns, and uses a comparison group to account for the effects of immigration on other district attributes. We find that between 1970 and 2000, the average metropolitan school district in California lost 16 non-Hispanic households with children to other school districts for every 10 additional households enrolling low-English Hispanics in public schools. Our findings suggest that the native reaction to immigration disproportionally affects children, and thus may have longer-run consequences than previously thought.
I. Introduction

Americans have long revealed distaste for racial and socioeconomic diversity in social interactions by sorting residencially. In recent years, this distaste has been manifested in the native response to a large wave of low-skilled immigration to the U.S. from developing countries, particularly from Central America. These new immigrants are residentially isolated (Cutler, Glaeser, and Vigdor, 2008a), due in part to relocation of natives from the neighborhoods where they settle (Saiz and Wachter, 2006). Their physical separation from natives may reduce their access to social networks, economic opportunities, and public services, like education, that weigh heavily in their future economic prospects.

But what drives “native flight”? The potential consequences for immigrants depend in part on the answer to this question. This paper explores the possible role of native demand for public schools in the residential isolation of recent low-skilled Hispanic immigrants. The hypothesis is a compelling one. Analyses of past policies like court-ordered school desegregation (Reber, 2005, Baum-Snow and Lutz, 2009) and school redistricting (Weinstein, 2009) suggest that shocks to minority presence in neighborhood schools can generate considerable losses of white families with children. The children of low-skilled Hispanic immigrants may also be poorly prepared academically, pulling down average test scores in the schools they attend. Declines in performance of a neighborhood school can also lead to population losses, particularly of families with higher achieving children (Figlio and Lucas, 2004).1

We build off of this previous research. In broad terms, we investigate whether school districts that saw greater increases in low-skilled Hispanic enrollment shares in recent decades have also experienced slower increases in the settlement rate of non-Hispanic households with children.

---

1 A considerable hedonic literature also speaks to these issues. For example, Black (1999) estimates willingness to pay for a neighborhood school with higher average test scores, while Kane, Riegg, and Staiger (2006) and Boustan (2010a) estimate willingness to pay for schools with lower minority shares in enrollment.
Unlike previous research, however, these changes in school demographics are tied to immigrant settlement, not to policy. This presents two challenges for our analysis. First, increases in low-skilled Hispanic enrollment shares will be associated with an increased low-skilled Hispanic presence in a community overall, which could have independent effects on non-Hispanic location decisions by changing other district attributes. We account for these other impacts by having a comparison group. We use childless non-Hispanic heads under age 50 as a comparison group because they are similar to non-Hispanic households with children in most respects, and hence may react similarly these other changes in community characteristics.2

Second, changes in school demographics may be non-random with respect to other district characteristics that non-Hispanic parents value when choosing a residence. We address this by borrowing an instrument from research on the labor market impact of immigration (e.g., Card, 2001), which is based on the idea that locations with larger initial immigrant communities are relatively attractive for new immigrant settlement but on average similar to other locations in their appeal to natives. We exploit variation in the initial size of Mexican communities in our analysis, as Mexican immigration has been the most important driver of recent growth in the low-skilled Hispanic U.S. population.

Our analysis focuses on California over 1970 to 2000. Child demographics in California were dramatically affected by Mexican immigration over this period: as shown in Figure I, first- and second-generation Mexicans accounted for over a quarter of children in the state by 2000, up from 5.4 percent in 1970. A focus on Mexican settlement in California also makes for compelling application of the identification strategy outlined above. The historical roots of Mexican immigration

---

2 In practice, we compare changes in the population of non-Hispanic 0-19 year olds (“children”) to changes in the population of non-Hispanic 20-49 year olds, since no publicly available data give counts of households in cells detailed by age of the householder, detailed geography, and ethnicity. However, relative outflows of children are sufficient to demonstrate relative outflows of households with children. In addition, it is possible to convert our estimates to an approximation of what we would observe at the household level, which we do.
meant that, by 1970, Mexicans were spread throughout California, not concentrated in few places where they would have already been affecting the residential choices of non-Hispanic parents.\(^3\)

Consistent with this, our instrument for growth in the low-English Hispanic enrollment share is uncorrelated with both initial levels and prior trends in the distribution of non-Hispanic households with children across school districts within California metropolitan areas. Many observed correlates of trends in district choice of non-Hispanic parents are also uncorrelated with the instrument.

Our data come from numerous sources. Federal administrative data offer counts of Hispanic students with low English proficiency, a hallmark of low skill among recent immigrants and a salient marker of immigrant status for non-Hispanics. These data were collected to monitor compliance with federal civil rights law and have previously been used for research on school desegregation (e.g., Cascio et al., 2010). District population is drawn from Census tabulations.

Our estimates imply that, between 1970 and 2000, the average metropolitan school district in California lost 16 non-Hispanic households with children for every 10 additional households enrolling low-English Hispanic children in public schools. Auxiliary evidence supports that this effect is driven by reductions in non-Hispanic demand for public schools. For example, districts experiencing more gains in low-English Hispanic public enrollment share also saw larger increases in non-Hispanic resident private school enrollment rates over the same period.\(^4\) Our estimates are also comparable in magnitude to the white enrollment and population losses associated with increased exposure to blacks after court-ordered school desegregation in the 1960s and 1970s (Reber, 2005; Baum-Snow and Lutz, 2009).\(^5\) The comparison is helpful in placing our estimates into historical context, suggesting shocks to school demographics continue to shape residential segregation today.

---

\(^3\) For example, the Bracero Program drew Mexicans to the California countryside as farm labor starting in the 1940s.

\(^4\) This complements evidence of such an effect at the metropolitan area level (Betts and Fairlie, 2003).

\(^5\) Also, the fact that the household displacement rate is larger than one-for-one says that it cannot be entirely accounted for by an inelastic supply of housing suitable for households with children, and other evidence presented below suggests that crowding in the housing market may account for little of the estimate.
More generally, our estimates imply that recent low-skilled Hispanic immigration has led to more relocation of non-Hispanics with children than without. The children of these immigrants thus may be exceptionally isolated. Residential sorting by natives may therefore have longer-run consequences for immigrants than previously thought, potentially limiting their acquisition of U.S. specific human capital for generations.\(^6\) Investigating these consequences is an important topic for continued research, one made even more pressing by the fact that immigration is likely to drive growth in child population in the United States into the foreseeable future.\(^7\)

II. Immigration and District Choice in Theory

Our analysis examines whether there have been changes in the residential choices of non-Hispanic families with children as a result of rising low-English Hispanic enrollments in California public schools. The framework presented in this section highlights the conditions under which we might interpret such choices as isolating reduced demand for public education.

Suppose that households are exogenously assigned to different metropolitan areas by job opportunities, and within metropolitan area, choose a school district in which to reside.\(^8\) In equilibrium, the utility, \(V\), associated with the school district of settlement for a household must be at least as large as the highest available to it elsewhere in the metropolitan area, \(v\):

\[
V(p, q, k, \bar{i}) \geq v.
\]

Utility is decreasing in housing costs, \(p\), and (weakly) increasing in the output of public schools, \(q\).\(^9\) Utility is also (weakly) decreasing in both the low-English Hispanic share in public school enrollment, \(k\), and the overall low-English Hispanic share in the population, \(\bar{i}\). Both capture distaste

\(^6\) There is evidence that immigrants', particularly low-skilled immigrants', skill acquisition is significantly reduced by residential isolation (Cutler, Glaeser, and Vigdor, 2008b).

\(^7\) Park (2009) shows that low-English enrollment accounts for almost all of recent enrollment growth in public schools.

\(^8\) Metropolitan areas are constructed to capture labor markets. It is common to assume that metropolitan areas also define markets for schools (e.g., Hoxby, 2000; Urquiola, 2005; Rothstein, 2006). The assumption of exogeneity in metropolitan area of settlement is a simplification and is not required for identification in our analysis.

\(^9\) Housing costs include both the (rental) price of housing and taxes. Utility would generally be modeled as increasing in income less these housing costs, or potential consumption. We abstract from the effects of income here since cross-district moves within a labor market are not likely to change it.
for diversity, but on different dimensions: in schools \((k)\) and elsewhere in the community \((\dot{i})\).\(^{10}\) \(k\) is increasing in \(i\).

Now suppose that a district receives an influx of low-skilled Hispanic students, all else constant. Non-Hispanic demand for district residence may decline because of a distaste for diversity in public schools \((\partial V / \partial k < 0)\), or because low-skilled Hispanic students lower school output, e.g., through reductions in school resources or peer effects in the classroom \((\partial V / \partial q)(\partial q / \partial k) < 0)\).\(^{11}\) The reduced-form effect of interest could work through either channel. But non-Hispanic demand for district residence may also fall if low-skilled Hispanic neighbors are undesirable for other reasons \((\partial V / \partial i < 0)\). If the housing supply is not perfectly elastic, the underlying population growth will also lower non-Hispanic demand for district residence by raising housing costs \((\partial V / \partial p)(\partial p / \partial i) < 0)\) (Saiz, 2003, 2007).

Thus, non-Hispanic households with children may move away from districts with growing low-skilled Hispanic enrollments for other reasons. Our approach to this identification problem is to use a comparison group. Exactly how we do this is detailed in the next section, but the basic idea is that there exists some other set of non-Hispanic households without children (group 0), that on average reacts in the same way as households with children (group 1) to changes in other district attributes that accompany a rising the low-skilled Hispanic enrollment share (i.e.,

\[
\frac{\partial V_1}{\partial i} = \frac{\partial V_0}{\partial i} \quad \text{and} \quad \frac{\partial V_1}{\partial p} = \frac{\partial V_0}{\partial p}.
\]

In this case, the \textit{difference} in the two groups’

\(^{10}\) \(i\) also captures the effects of immigration on local public goods besides education.

\(^{11}\) California institutions may have limited the impacts of low-English Hispanics on school output. For most of the period of study, California maintained systems of bilingual education, whereby students could spend an extended period in separate classes taught in Spanish, and highly egalitarian school finance. Thus, spillovers to native students would have been attenuated, and reductions in property values as a result of immigration would not have manifested in lower school spending. On the other hand, since California’s system equalizes current funding and not facilities, increases in low-English Hispanic enrollment may have exacerbated crowding in schools. These students may have also indirectly reduced already limited resources for native students, if bilingual education is more expensive and inadequately funded.

\(^{12}\) We assume one housing market, which returns to equilibrium when \(p\) adjusts sufficiently to restore (1) for both household types \(j\). If Mexican immigrants and non-Hispanic households with children tend compete for the same types of houses (e.g., detached single-family residences), immigration may raise house prices relatively more for type 1.
probabilities of choosing the district reveals whether the presence of low-English Hispanics in public schools has made it less attractive. Differencing also yields a lower bound on the effect of interest, since changes in \(k\) (and \(q\)) may also affect the utility of households without children.

An additional complication for estimation is that growth in low-skilled Hispanic students in a district is not random: the same basic model in (1) underlies the residential choices of their families. They may be attracted to declining school districts by lower rents, in which case non-Hispanic departures from a district might generate increases in \(i\) (and \(k\)), not vice versa.\(^{13}\) Or a positive shock might attract households with children, regardless of origin, possibly generating a positive correlation between \(i\) (and \(k\)) and the presence of non-Hispanic children in a district. We approach this identification problem using the instrumental variables approach described below. The basic idea is that an initial Mexican presence in a district will attract substantial new Mexican settlement, but be too small to appreciably affect the residential decisions of non-Hispanics.

III. Econometric Specification and Data

III.A. Basic Model

Our empirical specification generalizes the comparative static described in the previous section as a differences-in-differences model. In particular, we would like to test whether the change over time in the relative rate at which non-Hispanic households with children chose a district is related to the change over time in its low-English Hispanic enrollment share. Such a model is:

\[
\Delta \frac{N_{dun}^1}{N_{w}^1} - \Delta \frac{N_{dun}^0}{N_{w}^0} = \theta \Delta k_d + x_d' \beta + \epsilon_{de},
\]

where \(\Delta k_d\) represents the change over time in the low-English Hispanic share in enrollment in district \(d\), and \(\Delta \left( N_{dun}^1 / N_{w}^1 \right)\) represents the change over time in the share of metropolitan area \(m\)'s households, prompting greater losses of non-Hispanic households with children even absent changes in schools. We explore this alternative hypothesis below.

\(^{13}\) For example, Boustan (2010b) shows that the foreign-born were attracted to center cities that whites had earlier fled in response to black in-migration.
non-Hispanic households of type \( j \) living in \( d \). The parameter of interest is \( \theta \), which gives how much larger a proportion of a metropolitan area’s non-Hispanic households with children \( (j=1) \) – the “treatment” group – than households in the comparison group \( (j=0) \) departed a district with a rise in its low-skilled Hispanic enrollment share, holding constant other observed factors that may have made the district less desirable to non-Hispanic parents, captured in \( x_{\text{dist}} \).\(^{14}\)

For the moment, ignore endogeneity, or assume that \( \Delta k_d \) is uncorrelated with the error term, \( \varepsilon_{\text{dist}} \). That is, assume that changes in low-skilled Hispanic enrollment share are not systematically higher (or lower) in districts that might be losing (or gaining) non-Hispanic families with children for other, unobserved reasons. Ordinary least squares (OLS) will then produce unbiased and consistent estimates of \( \theta \). But for estimates of \( \theta \) to have the desired interpretation – capturing residential choices tied to demand for public schools – it must be the case that the comparison group reacts in the same way as the treatment group to all other changes in district attributes that might accompany a rise in low-skill Hispanic enrollment.

To achieve this interpretation, we use a comparison group that appears to be the closest match to the treatment group in available data – non-Hispanic household heads under age 50 without children. This group is a good match to the treatment group on several grounds. First, almost all non-Hispanic children are in households with an adult under age 50 (93.5% in 2000 data). Second, non-Hispanics under age 50 express similar general views about immigration regardless of the presence of children in their household; older non-Hispanics express relatively negative views.\(^{15}\)

\(^{14}\) \( \theta \) can be interpreted as the difference across groups in the sensitivity of location decisions to low-English Hispanic enrollment share, \( k \), in a household-level linear model of district choice that is motivated by the theoretical model presented in Section II (see Appendix).

\(^{15}\) The General Social Survey (Davis and Smith, 2009) asks a variety of questions about views on immigration. We examined the following questions, available in the 1996 and 2004 waves: “How much do you agree or disagree with the statement[s]:” (1) “Immigrants take jobs away from people who were born in America,” (2) “Immigrants increase crime rates,” (3) “America should take stronger measures to exclude illegal immigrants,” and (4) “Immigrants are generally good for America's economy.” We computed the first principal component of the 16 dummies for all possible responses (“neither agree nor disagree” excluded), and also a simpler measure based on the sum of negative responses. Both variables show significantly more negative views about immigrants for those over 50 than under age 50. In
Third, older Californians face considerably higher tax costs of moving than younger Californians as a result of Proposition 13. In a matching estimator with individual-level data, were such an approach feasible, the comparison group would thus rely heavily on childless households of parenting age, just as we are.

In practice, counts of households by presence of children are not broken out by the age of the householder at the school district level. Instead, we use available data on non-Hispanic population by age at the district level from the 1970 and 2000 school district tabulations (SDT) of the U.S. Census of Population. Our baseline model considers the treatment group to be non-Hispanic 0 to 19 year olds (“children”) and the comparison group to be non-Hispanic 20 to 49 year olds (“adults of parenting age”). That is, our estimating equation is:

$$\Delta \frac{N_{a}^{0-19}}{N_{m}^{0-19}} - \Delta \frac{N_{a}^{20-49}}{N_{m}^{20-49}} = \tilde{\theta} \Delta k_{d} + \frac{\alpha}{\varphi_{m}} \tilde{\beta} + u_{d},$$

where $\Delta \left( \frac{N_{a}^{0-19}}{N_{m}^{0-19}} \right)$ represents the change in the share of non-Hispanics in age group $a$ in metropolitan area $m$ living in district $d$. The parameter $\tilde{\theta}$ is similar in interpretation to $\theta$, but now gives how much larger a proportion of a metropolitan area’s non-Hispanic children than adults of parenting age departed a district with a rise in the low-English Hispanic enrollment share. In the Appendix, we show that $\tilde{\theta}$ in equation (2') understates $\theta$ in equation (2) by a factor proportional to the fraction of adults of parenting age who are parents.

---

16 Proposition 13 (1978) effectively locked in property taxes for existing homeowners by establishing a statewide property tax rate of 1 percent and setting assessed valuations of property at 1975 levels, with a maximum increase of 2 percent per year and no re-assessment. Propositions 60 and 90 in 1986 and 1988, respectively, allowed individuals aged 55 and over to transfer this tax benefit to a new home, but only one of equal or lesser value (Ferreira, 2009).

17 Because school districts are sometimes missing in these data sets (see below), we obtain non-Hispanic population at the MSA level using separately-reported county level aggregates downloaded from the National Historical Geographic Information System (Minnesota Population Center, 2004), rather than by aggregating across school districts.

18 In particular, $\tilde{\theta} \approx (1 - \varphi_{20-49}) \frac{\varphi_{20-49}}{\varphi_{0-19}} \theta$ where $\varphi_{0-19}$ represents the fraction of non-Hispanic 0 to 19 year olds who are parents. In 2000 Census microdata for metropolitan California $\varphi_{20-49} \approx 0.55$, suggesting that our estimates should be...
We construct the share low-English Hispanic in enrollment using district public school enrollment by ethnicity and need for accommodation for poor English language skills from the 1976 and 2000 Elementary and Secondary School Civil Rights Surveys, which were conducted by the Office for Civil Rights (OCR) in the Department of Health, Education, and Welfare (HEW, later the Department of Education) to monitor compliance with federal civil rights law. Non-native English speakers became protected under federal law as a result of the Equal Educational Opportunity Act of 1974, and HEW set forth guidelines for accommodation and began monitoring district compliance in 1975. The first year in which data are available for all districts is 1976.19

III.B. Instrument

OLS estimates of $\tilde{\theta}$ in equation (2) will be biased if low-skilled Hispanics choose to settle in districts that non-Hispanic households with children are already departing, or if both groups are attracted to the same places.20 Measurement error in $\Delta k_d$ will also lead to attenuation bias. Our instrument for $\Delta k_d$ is constructed in two steps using data from multiple sources.

The first step is to generate a prediction of the change in the number of low-English Hispanic students enrolled in a district’s public schools over 1976 to 2000, given Mexican immigration over the period and initial patterns of Mexican settlement. This prediction is given by:

$$\hat{\Delta K}_d = \frac{M_{d}^{1970}}{M^{1970}} \Delta K^{1976-2000},$$

where $M_{d}^{1970}/M^{1970}$ is the share of the Mexican-born population (of all ages) in the U.S. residing in district $d$ in 1970 (from the 1970 SDT), and $\Delta K^{1976-2000}$ represents the number of low-English Hispanic children of school age in the 2000 Census (Ruggles, et al., 2010) who were either born in

scaled up by about 2.2 ($=1/0.45$) to represent the outflows of non-Hispanic households with children. It is also possible to directly estimate $\theta$ using population data and a more complex estimation procedure, also described in the Appendix. 19 Most of the 1970 to 2000 increase in California’s Mexican population appears to have occurred after 1976 (Figure I), so the lack of data prior to 1976 likely has little effect on our findings. The 1980 SDT lacks sufficient disaggregation of population counts by age and ethnicity to apply our empirical strategy.

20 Purely mechanically, departures of non-Hispanic children will also inflate $\Delta k_d$. 


Mexico or born in the U.S. to at least one Mexican-born parent (or if both parents are foreign-born, to a Mexican-born mother) who arrived in the U.S. in 1976 or later. We define as “low-English” children who are neither fluent in English nor speak English very well, since this definition yields similar low-English Hispanic shares in aggregate enrollment in the Census and the OCR. Our two-stage least squares (TSLS) estimates are, however, not sensitive to this definition.

In the second step, we obtain the instrument itself, \( Z_d \), which gives the predicted change in \( \Delta k_d \) under the assumption that \( \Delta K_d \) represents the only source of enrollment change in a district:

\[
Z_d = \frac{K_{d,1976} + \Delta K_{d,1976}}{Enr_{d,1976} + \Delta K_{d,1976}} - \frac{K_{d,1976}}{Enr_{d,1976}}.
\]

\( K_{d,1976} \) and \( Enr_{d,1976} \) represent the initial (1976) enrollment of low-English Hispanics and total enrollment, respectively, in district \( d \). The second component of \( Z_d \) is thus the initial share low-English Hispanic in the district, while the first represents what that share would have been in 2000 had nothing else changed. A first stage coefficient on \( Z_d \) of one would be expected if, on average, the settlement patterns of Mexican immigrants in the U.S. had not changed since 1970, and enrollments were otherwise changing little.

The instrument’s power thus derives from the tendency of new immigrants to settle in existing Mexican communities. Its validity hinges on the size of these existing Mexican communities being unrelated to other future changes in a district’s desirability for non-Hispanic families with children. We assess this assumption in multiple ways, testing whether \( Z_d \) is uncorrelated with relative outflows of non-Hispanic children from a district over the 1960s – before the big wave of Mexican immigration, but when the district may already have been in decline – or with other district

---

21 Appendix Table I, Panel A lists the top ten districts in our sample ranked by the share of Mexicans in the U.S. residing in the district in 1970. Los Angeles Unified alone was home to over 16 percent of Mexicans in the U.S. in 1970.
22 This is to be expected given that \( \Delta K_{1976-2000} \) depends in no way on the district. For the same reason, using 1976 data to calculate \( \Delta K_{1976-2000} \) has no impact on our estimates.
characteristics observed in the early 1970s that may predict whether a district will later become less attractive for non-Hispanic children, captured in $x_{de}$. For example, a series of California Supreme Court decisions in the 1970s led to an equalization of school funding across California school districts, which may have made initially high-spending districts less attractive over the period we study.\(^{23}\) We capture the possible intensity of this effect with district property tax revenues and expenditures per pupil prior to these court decisions using data from the 1972 Census of Governments. Information on a number of other district characteristics used to evaluate our identification strategy is drawn from various historical sources described in the Appendix.

### III.C. Estimation Sample

We restrict our analysis to California’s 23 metropolitan statistical areas (MSAs), by their 1990 definition. We define “districts” to serve all grade levels and to have constant boundaries between 1970 and 2000 and aggregate key variables accordingly. School districts occasionally reorganize, and we want to ensure that we follow consistent geographic units over time.\(^{24}\) Including both secondary districts (which operate high schools) and elementary districts (which operate primary and middle schools) would be redundant, as elementary districts feed secondary districts and thus cover an overlapping geographic area. We focus on secondary district boundaries, since secondary districts represent elementary districts that are not directly observed in the 1970 SDT due to their small size (see Appendix). Our findings are not substantively changed when we drop observations that rely on aggregated data. In our main sample, there are a total of 40 combined elementary-secondary districts and 193 unified districts, for a total of 233 observations.\(^{25}\)

---

\(^{23}\) School finance equalization led to outflows to private schools (Downes and Schoeman, 1998) and affected the income heterogeneity of populations within districts (Aaronson, 1999). See Brunner and Sonstelie (2006) for a discussion of school finance in California.

\(^{24}\) For example, if districts A and B in 1970 merge to form C by 2000, we aggregate A and B to create an observation for C in 1970. See Appendix for details.

\(^{25}\) There are 290 total secondary district boundaries (after accounting for reorganizations) in the 23 California metropolitan areas under observation. We lose 25 of these boundaries because they are not represented in the 1970 SDT due to their small size and another 32 boundaries, primarily due to poor data quality in the 2000 OCR.
IV. Exploratory Analysis

IV.A. Descriptive Statistics

Table I shows summary statistics for our estimation sample. Panel A gives statistics on the key variables in our analysis. The low-English Hispanic share in public school enrollment on average rose a substantial 10.7 percentage points between 1976 and 2000, on a base of 2.6 percent (Panel C). The standard deviation on this variable is also 10 percentage points, suggesting that there was quite a bit of variation in this variable across districts. The instrument has a comparable mean and standard deviation. The components of the main outcome – changes in the proportions of an MSA’s non-Hispanic children and adults of parenting age residing in a district – are on average close to zero. This is expected by definition; deviations from zero come about because our coverage of metropolitan areas’ districts is at times incomplete. The private school enrollment rate of non-Hispanics, also calculated using the SDT, rose by a considerable 10.3 percentage points between 1970 and 2000, more than a seven-fold increase over its 1970 level of 1.4 percent (Panel B). This could be a consequence of school finance equalization or other factors, not Mexican immigration. Equalization is a potential confounder that we hope to rule out through our identification strategy.

Descriptive statistics on control variables, including per-pupil expenditures and property tax revenues, are shown in Panel C. The average district in our sample enrolled about 14,849 students in 1976 and raised $636 per-pupil in revenues through the property tax and spent $1084 per student in 1971-72, with a standard deviation of around $300 in each. About 8 percent of districts in our sample are in center cities, and 71 percent had a private school serving any grade level as of 1976, the first year in which such information is available.

IV.B. Evaluating the Identification Strategy

Our identification strategy relies on the assumptions that the instrument is related to changes in the actual presence of low-English Hispanics in a district’s public schools, but is unrelated to
other factors that might have compelled non-Hispanic households with children to locate elsewhere, like a loss in the district’s fiscal autonomy as a result of school finance equalization or declining school quality. It is useful to begin by examining the credibility of these assumptions.

The first row of Table II speaks to the former, showing that there is a strong first-stage relationship between the instrument and the change in a district’s low-English Hispanic enrollment share in our baseline specification, which includes fixed effects for MSA by district type.26 Figure II shows this relationship graphically. The first-stage coefficient is statistically distinguishable from zero, at the one percent level, but also significantly less than one. Fewer than the expected number of low-English Hispanic students actually show up in districts, possibly because Mexican immigrants have spread out geographically over time (Card and Lewis, 2007).

To explore the second identifying assumption, the remainder of Table II shows slope estimates from comparable reduced-form regressions of many potential correlates of outcomes on the instrument. By and large, the findings support the credibility of our research design. While the instrument is negatively correlated with the likelihood of having a private school by 1976, it is unrelated to myriad other initial district characteristics, as shown in Panel C. It is not significantly related to initial enrollment in logs and levels (not shown) or to center city status.27 It is also not correlated with the density of public schools within district, which we capture with a dummy for whether it had an above average number of schools (in 1972) given its enrollment (in the 1976 OCR) and land area.28 Finally, relationships between the instrument and the initial school finance

26 These effects account for the fact that districts are sometimes missing from the sample, so that district shares of the MSA non-Hispanic population do not always sum to one within MSA. Nearly 80 percent of the variation in $Z_d$ in our sample is within metropolitan areas. Standard errors are clustered on metropolitan area, and t-statistics are compared to critical values in a t-distribution with 21 degrees of freedom (the number of clusters less two, a rule of thumb explored in Cameron et al., 2008).

27 Related, Appendix Table I shows that the largest districts in our sample are not those with the largest predicted increases in low-English Hispanic share.

28 This could also be interpreted as showing that the instrument is uncorrelated the median voter’s preferences for segregation. We arrive at this prediction by regressing the natural log number of schools on the natural log in enrollment, the natural log of land area, their interaction, and all of the other pre-existing district characteristics listed in
variables are statistically insignificant and small, amounting to, for example, only about $20
differences in per-pupil property tax revenues and per-pupil expenditures for a one standard
device (0.10) change in the instrument. Still, confidence intervals on some of these estimates are
fairly wide, so we include all of these characteristics as controls in some specifications below.

The instrument is nevertheless significantly related to the low-English Hispanic share in
public school enrollment in 1976. Mexican immigration may have already been changing the
demographics of children in California in the early 1970s (see Figure I). Even if this were not the
case, this relationship is not surprising to see given that the instrument derives from the district’s
Mexican-born population in 1970. The initial share of enrollment that was low-English Hispanic was
small, however, compared to the change that occurred over the next 30 years, making any changes in
district choice among non-Hispanics plausibly related to the predicted change rather than the initial
level. Put differently, if the initial low-English Hispanic share represents a meaningful source of
variation in the instrument, then the instrument should predict initial levels and prior trends in non-
Hispanic demand for schools, invalidating our preferred interpretation.

Table II Panel B shows that this is not the case. The instrument is unrelated to 1970 levels of
our outcome variables; it also unrelated to prior trends, shown in a separate table below. Despite
having fewer private schools by 1976, the 1970 private school enrollment rate of non-Hispanics was
also not significantly lower in districts with higher values of the instrument. And while the
instrument is negatively related, though not significantly so, to the share of an MSA’s non-Hispanics
residing in a district in 1970, the difference between the estimates for children and adults of
parenting age – the 1970 level version of our dependent variable – is not statistically significant and
fairly precisely estimated.

Table I, Panel B (except the initial low-English Hispanic enrollment share), and classifying as “high choice” those with
non-negative residuals. We explored alternative regression models (e.g., in levels, not logs), as well as considered schools
per square mile and schools per child enrolled. All yielded insignificant relationships with the instrument.
V. Choice of School District

V.A. Reduced-Form Estimates by Age Group

To make our estimation strategy transparent, we begin the presentation of our findings by decomposing the reduced-form specification of model (2') into its constituent parts – separate estimates of the relationship between the instrument and the change in the rate at which non-Hispanic children and non-Hispanic adults of parenting age chose a district. We do this in the first two columns of Table III, again conditioning on metropolitan area by district type fixed effects. The difference between these estimates, presented in column (3), is the reduced-form of the model of interest.

Panel A presents these estimates for the full sample. The instrument is negatively related to the 1970 to 2000 change in the proportion of an MSA's non-Hispanic children residing in a district: a 0.1 increase in the expected increase in low-English Hispanic share – again, roughly the standard deviation of the instrument – is on average associated with a population loss amounting to 0.729 percent of an MSA's children (column (1)). For a district of average initial size, this figure represents about 9.2 percent of non-Hispanic children in 1970 (0.729/7.9) – a sizable effect. 29

However, a one standard deviation increase in the instrument is associated with a loss of about 0.425 percent of an MSA's non-Hispanic adults of parenting age as well (column (2)). If this estimate accounts for all other reasons that Mexican immigration might have prompted moves from a district, the parameter of interest becomes the difference in these coefficients, which is -0.0304 (column (3)). This suggests that, for every standard deviation increase in the instrument, a district of average initial size would have lost about 3.8 percent (0.3/7.9) of its non-Hispanic children due to changing school demographics. Since many adults of parenting age are indeed parents, this figure understates the loss of non-Hispanic households with children, as noted above.

---

29 The average district in our sample had 7.9 percent of its MSA's non-Hispanic children in 1970 (Table I, Panel B).
Figure III is a graphical representation of the regression in column (3). With the exception of large districts like Los Angeles Unified, San Francisco Unified, and San Diego Unified – which naturally account for larger shares of their respective MSA’s non-Hispanics at baseline – districts are tightly clustered around the downward sloping line. Below, we show that our findings continue to hold when large center city school districts are dropped from the sample – which is not surprising given that the instrument is uncorrelated with district enrollment and center city status – and under alternative specifications of the dependent variable that are less sensitive to scale. We also present the TSLS estimate that corresponds to this reduced-form model, explore the sensitivity of this model to inclusion of controls, and discuss ways to interpret the magnitude of the estimate.

The remainder of Table III presents a falsification exercise that complements that performed above. In particular, we test whether there was an “effect” on relocation in the 1960s – prior to the big wave of Mexican immigration (see Figure I). If so, it would suggest that it is not changes in school demographics stemming from immigration that drove the “outflows” of non-Hispanic children, but some other correlated factor. Limiting the sample to the 13 MSAs observable in 1960 (see Appendix), the 1970 to 2000 finding is still negative, significant, and quite similar in magnitude to that found for the full sample, at -0.0353 (Panel B, column (3)). Nevertheless, the age-group-specific relationships with the instrument are different in this subsample of cities. The sensitivity of the estimates in column (1) but not (3) reinforces the utility of using a comparison group, as differencing accounts for all unobservables district desirability for non-Hispanics regardless of age.

The final panel shows the findings from the falsification exercise. There is little evidence that districts with larger values of the instrument were already experiencing greater losses of children in the 1960s: the coefficient in column (3) is a small -0.0066 and is not statistically significant.30

Figure IV shows scatter plots that correspond to this model (Panel B) and its 1970 to 2000

---

30 Even if we multiply this coefficient by three, to scale up to a change over a 30 year period instead of over a 10 year period, it remains lower in magnitude than the effect we estimate over 1970 to 2000.
counterpart for the restricted sample (Panel A). The graph shows that the lack of a pre-trend is not
driven by some outlying observation. Thus, a significant negative relationship between the
instrument and relative loss of children from a district appears limited to the period in which
Mexicans were arriving in California in large numbers.

V.B. Main Findings

Table IV shows our main results – TSLS (Panel A) and OLS (Panel B) estimates of model
(2). The TSLS estimate in column (1) corresponds to the reduced form estimates in column (3) of
Table III, Panel A, and is thus based on a specification that includes fixed effects for MSA by district
type. The statistically significant point estimate of -0.0705 implies that a 10 percentage point
increase in the share of a district's public school enrollment that is low-English Hispanic induced
0.705 percent of an MSA’s non-Hispanic children to locate elsewhere, or about 8.9 percent
(0.705/7.9) of non-Hispanic children in a district of average size initially. This estimate is larger than
the corresponding reduced-form estimate because the first-stage coefficient is less than one.31

As anticipated, the TSLS estimate changes little with the addition of controls for initial
district characteristics listed in Table I, Panel C (column (2)) and for the initial (1970) level of the
dependent variable (column (3)).32 In fact, they are slightly larger. By contrast, OLS estimates fall in
magnitude with the addition of controls (Panel B). They are also considerably smaller in magnitude
at the outset (-0.0215 in column (1)) and not statistically significant in subsequent specifications.
The OLS estimates may be attenuated by measurement error or by unobserved positive shocks that
attract families with children, regardless of origin, to the same districts.33

31 See Table II, Panel A. The first-stage coefficient estimates on the instrument are very stable across specifications.
32 We do not include the 1976 low-English Hispanic share in public school enrollment among these controls. Adding
this control has little impact on the OLS estimates, and increases the TSLS estimates in magnitude. Also, adding only
the control for the presence of a private school in 1976, which was correlated with the instrument (Table II), has no
substantive effect on the TSLS estimates. (Estimates available on request.)
33 Using an instrumental variable with similar motivations to that used here, Saiz and Wachter (2006) also obtain OLS
estimates that are much less negative than TSLS when examining the response of native and white non-Hispanic
population to changes in the foreign-born population at the tract level in the entire U.S.
We present estimates for different subsamples in the remaining columns. Column (4) drops center city districts from the sample. The TSLS point estimate is slightly smaller than that in column (1), but remains highly significant, confirming that our main estimates are not driven by the sensitivity of the outcome measure to district size. Further, it shows that our main findings are not being driven by non-Hispanic families with children moving out of center cities. The final columns break out the estimates by district type. We cannot reject that the estimates are identical, but the findings suggest that our main estimates are driven by unified districts. Estimates for the combined elementary-high school districts are much smaller and not statistically significant.\(^{34}\)

**V.C. Additional Sensitivity Analysis**

Table V examines various alternative formulations of the dependent variable. For comparison purposes, row (a) of Panel A repeats the TSLS (column (1)) and OLS (column (2)) estimates that appeared in column (1) of Table IV. Row (b) replaces the district’s share of the MSA’s non-Hispanic population, \(s_{d,s} = N_{d,s} / N_s\), with the log odds, \(\ln(s_{d,s} / (1 - s_{d,s}))\). The dependent variable is therefore the treatment-comparison difference in the 1970 to 2000 change in log odds, or approximately the treatment-comparison difference in the growth rates of \(s_{d,s}\). Logit is an attractive specification for this application, where the distribution of district size (and hence, \(s_{d,s}\)) is right skewed.\(^{35}\) Both OLS and TSLS estimates are statistically significant in this specification, and the marginal effects, evaluated for a district of average initial size (coefficient\(\times 0.079\times (1-0.079)\)), are larger than were estimated in the linear model. Moreover, we see the same pattern as in Table IV: TSLS is larger in magnitude than OLS, and OLS is more sensitive to controls (latter not shown).

We also estimate these two models weighted by the district’s 1970 population of non-Hispanic children in Panel B. Weighting provides insight into whether the effects of changing

\(^{34}\) The TSLS estimate for elementary districts only is slightly larger in magnitude than that in column (6) and is more precisely estimated (coefficient= -0.0193, se=0.0126).

\(^{35}\) This specification would be appropriate if the household-level model underlying district choice were logit.
school demographics for the typical non-Hispanic of school age differ from those for the typical
district. In the main specification (row (a)), the weighted TSLS coefficient is larger than the
unweighted TSLS coefficient. Weighted estimation of the logit model may be more insightful, since
the logit transformation is theoretically less sensitive to scale. The weighted TSLS logit coefficient in
Panel B (row (b)) is a bit smaller than its unweighted counterpart in Panel A. Roughly speaking, this
implies that larger districts lost on average fewer non-Hispanic children as a percent of their baseline
population than smaller districts in response to increasing low-English Hispanic enrollment. There
may be greater capacity to sort within larger districts, or less centralized administration may allow
larger districts to mitigate some of the potential adverse effects of changing school demographics for
non-Hispanic students. Still, it is again the case that the marginal effect, now evaluated using
weighted means, is larger than in the linear model.

Returning to the bottom two rows of Panel A, we present (unweighted) estimates for
dependent variables that may be easier to interpret. In TSLS, an increase of 0.10 in the low-English
Hispanic share in enrollment is associated with an 11 percentage point decline in the relative
population growth rate of non-Hispanic children (row (c)). This is similar to estimates of “white
flight” after school desegregation. Using a similar growth rate specification, for example, Baum-
Snow and Lutz (2009) find that court orders to desegregate center city districts in the 1960s and
1970s on average led to an increase in white exposure to black peers of 0.09 and declines in white
enrollment and population of 12 percent and 6 percent, respectively. 36 This supports that demand
for public schools drives these estimates, a point to which we return below. TSLS estimates also
imply that a 10 percentage point increase in low-English Hispanic enrollment share is associated
with a 2.5 percentage point decline in the child share of the non-Hispanic population under 50 (row
(d)).

36 See also Reber (2005). A 10 percentage point increase in low-English Hispanic share will lead to a 10 percentage point
increase in the exposure of non-Hispanics to low-English Hispanics if there is no sorting response within the district.
V.D. Displacement Rates

Our substantive conclusions are therefore unaffected by changes in functional form or weighting. But are the quantitative conclusions affected? One way to make the estimates in Table V comparable to each other is to restate them as “displacement rates” – how many non-Hispanic children “left” the average district for each low-English Hispanic arrival into public schools. Displacement rates based on our TSLS estimates are given in column (3) of Table V. To be consistent with earlier discussion, we base our calculations on a district of average initial size and assume a 10 percentage point increase in the low-English Hispanic public enrollment share over 1970 to 2000. The displacement rates are similar if we consider larger or smaller changes or a district of median initial size.

To illustrate, it is useful to walk through calculation of the displacement rate for our main specification, given in Panel A, row (a). In a district of average initial size, it would have required about 1,690 additional low-English Hispanics to increase their enrollment share by 10 percentage points.37 We calculated above that this 10 percentage point increase led to an 8.9 percent reduction in the population of non-Hispanic children – about 1,946 non-Hispanic kids. So, the TSLS coefficient implies a displacement rate of 1.15 (=1,946/1,690), or that 11 to 12 non-Hispanic children located elsewhere for every 10 additional low-English Hispanic arrivals in a district's public schools.

It is also useful to calculate displacement rates in terms of households. We present such estimates in column (4). To arrive at these numbers, we rescaled the TSLS coefficients upward, to adjust for the fact that slightly over half of young adults are in households with children (see Section III and Appendix), and converted the number of low-English Hispanic public enrollees needed to

---

37 The average district in our sample had 14,850 students in 1976, 313 (about 2 percent) of whom were low-English Hispanics. To raise the low-English Hispanic share by 10 percentage points, to 12 percent, required the addition of about 1,690 low-English Hispanic students, since (313 + 1,690)/(14,850 + 1,690) ≈ 0.12.
deliver a 0.1 increase in their enrollment share into the number of households they represent.\(^{38}\)

These household-level displacement rates are uniformly larger than the child displacement rates – 1.6 in our main specification, for example.

The other estimates in Panel A imply larger displacement rates. The estimated marginal effect from the logit specification implies displacement rates of 1.44 and 2 at the child and household levels, respectively (row (b)). The 11 percent loss of non-Hispanics in the growth specification with a 10 percentage point increase in low-English Hispanic share implies similar displacement rates (row (c)). When the dependent variable is the change in the share of non-Hispanics under age 50 who are children, the displacement rates are also slightly larger than in the main specification (row (d)). The weighted estimates in Panel B imply lower displacement rates, on the basis of weighted means of all key variables, but they remain above one at the household level.

VI. Discussion

Our estimates suggest that the children of low-skilled Hispanic immigrants are left exceptionally isolated by non-Hispanic residential choices. This finding is interesting in its own right. But can we go further to infer that this reflects a reduction in non-Hispanic demand for public schools? As noted above, the answer is “yes” if the comparison group values changes in other district attributes the same as families with children. While we attempted to justify our choice of comparison group on these grounds, it is useful to discuss several competing hypotheses and to provide complementary evidence that rising low-English Hispanic enrollments reduced non-Hispanic demand for public schools over the period of interest.

VI.A. Crowding in the Housing Market

One competing hypothesis for the effects we estimate is crowding in the housing market. To see this, suppose that that housing costs, \(p\), in equation (1) differ across household types \(j\), and

\[^{38}\text{Details on our displacement rate calculations are available upon request.}\]
that Mexican immigration puts greater pressure on housing costs for households with children (i.e., by bidding up prices on detached single-family residences), or that \( \frac{\partial p_1}{\partial i} > \frac{\partial p_0}{\partial i} \). Then even if \( p \) is valued in the same way regardless of household type, changes in housing costs resulting from immigration will have a relatively large impact on the settlement decisions of non-Hispanic households with children.

Several pieces of evidence suggest that crowding in the housing market cannot fully account for – and may not even account for much of – our findings. First, as shown in column (4) of Table V, the average district lost more than one non-Hispanic household for each low-English Hispanic household arrival. Thus, even if households with children displace each another one-for-one, at least some of our estimate must be driven by other factors. Second, non-Hispanic children and low-English Hispanic children in metropolitan California live in very different types of houses. Our tabulations from the 2000 Census, for example, show that the houses in which low-English Hispanic children reside were much more likely to be rented (35 percent versus 62 percent for non-Hispanic kids), much less likely to be detached single-family homes (41 percent versus 66 percent), and on average had two fewer rooms (3.5 versus 5.5). In fact, the housing choices of low-English Hispanic families with children looked more like non-Hispanic adults of parenting age without children. This suggests that, if anything, Mexican immigration may have driven up housing costs for the comparison group by relatively more, possibly biasing us against finding any effect.

Third, a direct investigation suggests that tight markets for family housing are not greatly affecting our estimates. Ideally, we would have some district-level measure of constraints on new construction, such as the initial share of land in a district that is developable (see, for example, Card, Mas, and Rothstein, 2008). Unfortunately, such a measure is not available for 1970. However, we can observe initial population densities at the district level, both overall and by age group, which are

---

39 Forty-six percent of non-Hispanic 20 to 49 year olds without kids lived in owner occupied units and 44 percent lived in detached single family homes in 2000. The average home of this group had 4.5 bedrooms in 2000.
strongly correlated with the more sophisticated measures available in later years, but not correlated with our instrument. We find no evidence that the parameter of interest is larger in districts that initially have higher population densities. However, consistent with initial population density measuring constraints on housing supply, initially denser districts see significantly less growth in the representation of non-Hispanic children between 1970 and 2000 (results available on request).

VI.B. Other Amenities

Another competing hypothesis for the effects we estimate is that households with children are more sensitive to immigration on the community outside of public schools. For example, even though non-Hispanics with and without children have the same propensity to express that immigration raises crime rates, as discussed above, those with children might be more apt to consider impacts of immigration on crime when choosing a residence. While we cannot entirely rule this out, the same wave of immigration studied here appears to have had no impact on crime rates at the MSA level (Butcher and Piehl, 1998). Further, Mexican immigration does not unambiguously make a district relatively less desirable for non-Hispanic households with children. For example, reductions in the cost of child care or housekeeping services associated with immigration (Cortes, 2008) may be valued relatively highly by households with children and may –if proximity of immigrants is important in acquiring these services – bias against finding an effect.

VI.C. Private Schooling

Unlike decisions about where to live, the decision to attend private school should be driven only by the characteristics of public schools themselves. Exploring the effect of rising low-English Hispanic enrollments on private school enrollment rate can therefore provide direct evidence that changes in school demographics associated with low-skilled immigration reduced non-Hispanic demand for public schools.
Table VI presents TSLS and OLS estimates of the effect of the low-English Hispanic enrollment share on the 1970 to 2000 change in the non-Hispanic private school enrollment rate of district residents; as above, the baseline model controls for MSA by district type fixed effects. Both the TSLS estimates (Panel A) and OLS estimates (Panel B) are positive. The TSLS coefficient of roughly 0.18 (column (1)) implies that a 10 percentage point increase in the low-English Hispanic share in public school enrollment prompted about 1.8 percent of a district's non-Hispanic children to enroll in private school.\textsuperscript{40} This estimate suggests that rising low-English Hispanic enrollments can account for about 17 percent of the 10.3 percentage point increase in California’s non-Hispanic private school enrollment rate over 1970 to 2000. Controlling for the availability of a private school in 1976 – which was correlated with instrument (Table II, Panel C) – raises this estimate somewhat (column (2)), as might be expected, but adding the remaining controls has little impact (column (3)).\textsuperscript{41}

VII. Conclusion

This paper has examined whether the large increase in the low-English Hispanic presence in California public schools between 1970 and 2000 reduced non-Hispanic demand for public schools, as revealed through changes in the relative rate at which non-Hispanic households with children chose schools districts within California MSAs. Our empirical approach accounts for endogeneity using 1970 settlement patterns of Mexican immigrants. Supporting the credibility of our research design, districts predicted on the basis of 1970 Mexican settlement to receive more low-English Hispanics in their schools in future years were not already losing relatively more non-Hispanic children in the 1960s, and were comparable to other districts along many observable dimensions.

\textsuperscript{40} This response is close to the effect that Betts and Fairlie (2003) found at the MSA level for secondary students – that two natives enroll in private school for every 10 immigrant arrivals. Our study differs from theirs in several ways beyond the unit of observation, so this comparison should be made cautiously.

\textsuperscript{41} We have also estimated the model for sub-samples defined by center city status and district type. Estimates were statistically indistinguishable across groups (available on request).
We find that districts with larger increases in their low-English Hispanic enrollment shares lost more non-Hispanic children than adults of parenting age between 1970 and 2000. Not all of this effect can be explained by crowding in the housing market. Districts with larger increases in low-English Hispanic shares saw larger increases in their non-Hispanic private school enrollment rates over 1970 to 2000 as well, suggesting that the loss of non-Hispanic children was brought about at least in part by reductions in demand for public education. Our approach nevertheless leaves open the question of whether there have in fact been real negative spillovers from immigration for native students.\textsuperscript{42} Our estimates may simply reflect distaste for ethnic or socioeconomic diversity in schools.

Regardless of the mechanism, our findings suggest that the native sorting leaves immigrant children especially isolated, and thus may have long-run and intergenerational consequences. More generally, our findings suggest that public services are an important determinant of the residential choices of natives in response to immigration, and in turn, the residential isolation of immigrants. Existing research on the residential isolation of immigrants has had little comment on the role of local public goods in shaping residential decisions (Saiz and Wachter, 2006; Cutler, Glaeser, and Vigdor, 2008a), and, for that matter, there has not been much research on the effects of immigration on public goods at all.\textsuperscript{43} This is a fruitful area for future research.

\textsuperscript{42} Betts (1998) and Gould, Lavy, and Paserman (2009) estimate the effects of immigration on the educational outcomes of native students.

\textsuperscript{43} The few existing studies of the impact of immigration on public goods provision almost exclusively use an “accounting” type of framework – that is, adding up immigrants’ contribution to taxes and government expenditures in a static framework which ignores behavioral responses (e.g., Smith and Edmonston, 1997). A recent exception is Card (2007), who uses a cross-market regression to examine the impact of immigration on local dependency ratios.
VIII. Appendix

School District Level Data: Sources and Construction of Key Variables

1. 1970 and 2000 School District Tabulations

For 1970, school district level data on total population, by age and ethnicity, and private school enrollment, by ethnicity and level, were drawn from the 1970 Fourth Count (Population) School District Data Tapes (U.S. Department of Education, 1970). These data permit identification of all school districts in the country with at least 300 students as of the 1969-70 school year. For 2000, school district level data on total population and private school enrollment, by age and ethnicity, were drawn from the Census 2000 School District Tabulation. All operating districts are included in the age-specific resident counts, but private enrollment counts are missing for districts with 49 or fewer children.

Presentation of the data differs across years. For consistency over time in the definition of our key dependent variable, we aggregate non-Hispanic resident counts to the 0-19 and 20-49 age groups and aggregate these counts to constant secondary district boundaries. To arrive at the dependent variable used in our analysis, we divide by the non-Hispanic MSA population for that age group, generated from county level Census data available at the National Historical Geographic Information System (Minnesota Population Center, 2004). We use a similar approach to calculating the non-Hispanic private school enrollment rate: we first create comparable counts of non-Hispanics enrolled in private school; we next aggregate these counts to consistent secondary district boundaries; and finally, we divide by the aggregated district’s 5 to 19 year old population.

From the 1970 data, we also collected information on the distribution of foreign-born Mexicans across school districts, used in construction of the instrument.

2. 1976 and 2000 Office for Civil Rights Data

For 1976, school district level data on the number of low-English students, by race/ethnicity, were drawn from the Fall 1976 Elementary and Secondary School Civil Rights Survey, fielded by the Office for Civil Rights in the Department of Health, Education, and Welfare and recently decoded from binary to Stata format by Denckla and Reber (2006). The 1976 OCR survey covered all school

---

44 For California residents, the Spanish Heritage population includes “persons of Spanish language or persons not of Spanish language but of Spanish surname identified by matching with a list of about 8,000 such names.”
45 These data are available at <http://nces.ed.gov/surveys/sdds/downloadmain.asp>. To avoid disclosure, cell values are rounded so that exact values cannot be inferred; generally, this rounding is to the nearest 5, or to 4, when the population count is under 5. On a few occasions, rounding leads to (small) negative values.
46 In 1970, counts of residents by gender were originally reported for the total population and for the “Spanish Heritage” (hereafter referred to as Hispanic) population in detailed age bins (Table 17). In 2000, counts of residents by age and gender were reported for the total population (Table P8 for Total – Population and Households (TT)) and for the Hispanic/Latino population (Table 145H for TT).
47 The 1970 data reports counts of residents aged 3 to 34 enrolled in private school, by level (kindergarten, elementary, and secondary), for the total population and for the Hispanic population (Table 38). The 2000 data report counts of residents in private school, by gender, separately for all children and for Hispanic/Latino children either enrolled in or of age to be enrolled in the grades served by the district (Tables P8 and 145H for Children (CO): Relevant Children – Enrolled Private). For consistency, we limit counts of non-Hispanic private school enrollees in 1970 to the grade levels served by the district, and aggregate the 2000 figures across age and gender.
districts in the United States. For 2000, school district level data on the number of low-English students by ethnicity were drawn from the 2000 Elementary and Secondary School Civil Rights Compliance Report District Survey, fielded by the OCR in the U.S. Department of Education. The 2000 OCR survey covered all school districts in the United States, with tabulations rounded to the nearest 5, to avoid disclosure.

The original data give counts of “pupils whose primary language is other than English” in total and by race/ethnicity. Our treatment variable is constructed using the number of Hispanics (of all races) with this designation and total enrollment. We drop districts for which either of the following held in either 1976 or 2000: (1) the sum of non-low-English enrollment by race was more than 10 percent above or below reported non-low-English enrollment; or (2) the sum of enrollment by race was more than 10 percent above or below reported enrollment.

3. Other data sources


Tract-level data on non-Hispanic population for 1960 are from NHGIS (Minnesota Population Center, 2004).

School District Geography

We identify school district reorganizations using data from the Elementary and Secondary Education General Information System and the Common Core of Data Public Agency Universe and internet searches. We drop “aggregated” districts involved in reorganizations over the period if any of the component districts are not present in years they should be or, in a few cases, if we were not able to ascertain the nature of the reorganization that occurred.

Matching of 1960 Tracts to School Districts

In many cases 1960 and 1970 tract boundaries were identical. In the cases where they were not, we used published Census tabulations of the correspondence between 1960 and 1970 tracts (Table B in US Bureau of the Census, 1972) to construct collections of tracts that could be used to identify the smallest possible identical geographic regions in each census. For example, if a 1960 tract was split into two pieces, we would use that tract in the 1960 data and the aggregate of the two corresponding

48 We downloaded these data from <http://www.ed.gov/about/offices/list/ocr/data.html>
tracts in the 1970 data. In some cases the overlap between tracts was more complex than this example, but it was almost always possible to construct an exact match by aggregating enough tracts in both years.\textsuperscript{49} We then used the School District Geographic Reference File, 1969-1970 (U.S. Department of Commerce, 1970) to determine the fraction of each tract aggregate’s total population inside the borders of each school district in 1970. We apportioned non-Hispanics in each “tract aggregate” to school districts with these weights – which were mostly 0 or 1 – in 1960.

Derivation of Estimation Equation

A reduced-form, linear expression of the model in section II is given by:

\[
Y^j_{ntdm} = \gamma^j_{dt} + \beta_j k^j_{dt} + \eta^j_{ntdm},
\]

where \( Y^j_{ntdm} = 1 \) if non-Hispanic household \( n \) of type \( j \) resides in school district \( d \), which is located in metropolitan area \( m \), at time \( t \textsuperscript{50} \). This decision is a function of the share of public school enrollees who are low-skilled Hispanics in \( d \) at time \( t \), \( k^j_{dt} \), the response to which varies by household type, with \( j=1 \) for non-Hispanic households with children (the treatment group), and \( j=0 \) for the comparison group. This decision is also a function of other district characteristics at time \( t \) not differentially valued across types, captured by \( \gamma^j_{dt} \). Conceptually, \( \gamma^j_{dt} \) captures reactions to all district amenities that the two household types hold in common (including the other effects of “\( i \)” in the model). The parameter of interest is \( \theta = \beta_1 - \beta_0 \), the difference across household types in the sensitivity of location decisions to \( k \).

We are not able to estimate this equation given a lack of household-level data with sufficient geographic detail. To get to the difference-in-differences model, we begin by summing this across all non-Hispanic households \( n \) of type \( j \) in metro area \( m \) at time \( t \):

\[
N^j_{ntdm} = \sum N^j_{ntd} (\gamma^j_{dt} + \beta_j k^j_{dt}) + \sum \eta^j_{ntdm}
\]

The ideal household-level difference-in-differences model at the beginning of section III, (2), divides through (5) by \( N^j_{ntd} \) and then differences over time and across types to eliminate the effects of all common factors affecting location at a point in time (the \( \gamma^j_{dt} \)):

\[
\frac{\Delta N^j_{nt}}{N^j_t} - \frac{\Delta N^0_{nt}}{N^0_t} = \theta \Delta k^j + \epsilon^j_{nt}
\]

As noted, public use data with counts of households are insufficiently detailed to estimate this equation directly with our comparison group, childless households with a householder of parenting

\textsuperscript{49} The only exception to this was there were a handful of 1970 tracts or parts of tracts on the edge of metro areas that were untracted in 1960.

\textsuperscript{50} \( Y \) in equation (2) might also be interpreted as the latent propensity to live in district \( d \), with the probability of living in the district a non-linear transformation of it. If this transformation is a logistic CDF, then equation (2) would be the same except that the share of type \( j \) households in \( d \) will be replaced with the log odds that a type \( j \) household is in \( d \) (equal to \( \ln(\text{share}/[1-\text{share}]) \)), i.e., the logit model.
age (under age 50). Due to this data constraint, we instead use population data to estimate \((2)\), in which the dependent variable is the difference between the district’s 0 to 19 year old and 20 to 49 year old metropolitan shares. As a result of including parents in the comparison group, we noted, our estimates understate the parameter of interest.

To see this, let \(\tau_{at}^{ja}\) represent the average number of individuals in age group \(a\) (=0-19 or 20-49) per household of type \(j\) in \(m\) at \(t\). Multiplying \((5)\) by \(\tau_{at}^{ja}\) and summing across household types within age group generates a model for the (approximate) number of individuals of age \(a\) in district \(d\):

\[
N_{dmt}^{a} = N_{m}^{a} (\gamma_{d} + \beta_{0} k_{d mt}) + N_{m}^{1a} (\gamma_{d} + \beta_{1} k_{d mt}) + \sum \left( \eta_{at}^{0a} + \eta_{at}^{1a} \right),
\]

where \(N_{m}^{ja} \equiv \tau_{m}^{ja} N_{dmt}^{ja}\) and \(N_{m}^{1ja} \equiv \tau_{m}^{ja} N_{dmt}^{1ja}\). Letting \(N_{m}^{a} \equiv N_{m}^{a} + N_{m}^{1a}\) and \(N_{m}^{a} \equiv N_{m}^{a} + N_{m}^{1a}\), and noting that population aged \(a\) in \(m\) at \(t\) is \(N_{m}^{a} = \sum_{d \in m} N_{dmt}^{a}\), this model can be rewritten as

\[
(6) \quad \frac{N_{dmt}^{a}}{N_{m}^{a}} = \gamma_{d} + (\beta_{0} + \varphi_{mt}^{a} (\beta_{1} - \beta_{0})) k_{d mt} + \mu_{dmt}^{a},
\]

where \(\varphi_{mt}^{a} \equiv \frac{N_{m}^{1a}}{N_{m}^{a}}\) is the fraction of individuals aged \(a\) in \(m\) at \(t\) who are in households with any 0-19 year olds. Thus, by definition, \(\varphi_{mt}^{0-19} = 1\) for all \(m\) and \(t\). To get to our estimation equation, we simplify further by assuming that there is no variation over time and across metropolitan areas in this parameter for 20-49 year olds, or that \(\varphi_{mt}^{20-49} = \varphi^{20-49}\). Differencing over time within age groups, then across age groups then yields:

\[
(7) \quad \Delta \frac{N_{dmt}^{0-19}}{N_{m}^{0-19}} - \Delta \frac{N_{dmt}^{20-49}}{N_{m}^{20-49}} = (1 - \varphi^{20-49}) \theta \Delta k_{d} + (\Delta \mu_{dmt}^{0-19} - \Delta \mu_{dmt}^{20-49}).
\]

The parameter of interest remains \(\theta\), but as \((7)\) shows, the slope parameter we actually estimate is \(\tilde{\theta} \approx (1 - \varphi^{20-49}) \theta\). So our estimate is attenuated by a factor roughly proportional to the share of 20 to 49 year olds in households with children.\(^{51}\)

An alternative approach would have been to estimate \((6)\) more directly by interacting the treatment with time x metro varying estimates of \(\varphi_{mt}^{20-49}\), rather than assuming away the variation in \(\varphi_{mt}^{20-49}\). Our view is that our simpler approach is more transparent, and therefore more convincing. Nevertheless, we have estimated (a transformed version of) \((6)\), described here.

\(^{51}\) The logit and growth specifications in Table V are also biased by approximately the same factor, and so their coefficients are also divided by this in the calculation of household-level displacement rates. Converting the estimate in row (d) of Table V to a household-level displacement rate is not as simple, but also depends on this factor. (This and other details of our displacement calculations are available on request.)
A more practical challenge in directly estimating (6) is that metropolitan-specific estimates of the fraction of families with kids are not available in the initial period (because the 1970 Census has less geographic detail than the 2000 Census). However, differencing (6) between age groups and over time reveals that a more general estimation strategy can be employed without this information:

$$\Delta \frac{N_{m}^{0-19}}{N_{m}^{0-19}} - \Delta \frac{N_{m}^{20-49}}{N_{m}^{20-49}} = (1 - \phi_{m,2000}^{20-49}) \theta \Delta k_{j} + \Delta \phi_{m}^{20-49} \theta k_{j,0} + (\Delta u_{d}^{0-19} - \Delta u_{d}^{20-49})$$

This expression differs from (7) in two ways. First, the $(1 - \phi_{m,2000}^{20-49})$ interacted with $\theta \Delta k_{j}$ varies across metropolitan areas according to its end year (2000) values. Second, a term capturing the impact of changes in $\phi_{m}^{20-49}$ is in the equation, i.e. $\Delta \phi_{m}^{20-49} k_{j,0}$. Note that this term is an omitted variable in the simplified approach we take, but we cannot measure or control for it directly. However, it can be absorbed by allowing initial low-English Hispanic share to have metro-specific effects, as in:

$$\Delta \frac{N_{m}^{0-19}}{N_{m}^{0-19}} - \Delta \frac{N_{m}^{20-49}}{N_{m}^{20-49}} = (1 - \phi_{m,2000}^{20-49}) \theta \Delta k_{j} + \gamma_{m} k_{j,0} + \varepsilon_{d}.$$

$\gamma_{m} k_{j,0}$ captures the effects of $\Delta \phi_{m}^{20-49} k_{j,0}$.

We have estimated (8). The $\gamma_{m} k_{j,0}$ control makes our estimate of $\theta$ larger in magnitude (just like the less general version of this control did in Table IV). Allowing for metro-specific variation in $\phi_{m}^{20-49}$ makes the estimate of $\theta$ a little larger still. The basic reason is that areas which received more immigrants relative to their population, such as in the Central Valley, tended to have a higher share of families with kids. Thus, the main approach we take in this paper is conservative: by assigning the average fertility rate in the state to these high treatment areas, we underadjust our coefficient estimates. The bottom line is that if we could actually obtain appropriate household count data to employ our estimation strategy, the estimated household displacement rates would likely be larger in magnitude than those we report in column (4) of Table V.
References


Davis, James Allan and Smith, Tom W. 2009. *General social surveys, 1972-2008* [machine-readable data file] /Principal Investigator, James A. Davis; Director and Co-Principal Investigator, Tom W. Smith; Co-Principal Investigator, Peter V. Marsden; Sponsored by National Science Foundation. --NORC ed.-- Chicago: National Opinion Research Center [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut [distributor].


Table I. Descriptive Statistics for California School Districts

<table>
<thead>
<tr>
<th></th>
<th>Mean (1)</th>
<th>St. Dev. (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Key Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ public enr. share low-English Hispanic, 1976-2000</td>
<td>0.107</td>
<td>0.098</td>
</tr>
<tr>
<td>Instrument †</td>
<td>0.114</td>
<td>0.110</td>
</tr>
<tr>
<td>Δ share of MSA's non-Hispanics in district, 1970-2000:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-19 Year Olds (T)</td>
<td>-0.004</td>
<td>0.040</td>
</tr>
<tr>
<td>20-49 Year Olds (C)</td>
<td>-0.003</td>
<td>0.044</td>
</tr>
<tr>
<td>T-C Difference</td>
<td>-0.001</td>
<td>0.014</td>
</tr>
<tr>
<td>Δ Non-Hispanic private enrollment rate, 1970-2000</td>
<td>0.103</td>
<td>0.051</td>
</tr>
<tr>
<td><strong>B. Pre-existing levels of outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of MSA's non-Hispanics in district, 1970:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-19 Year Olds (T)</td>
<td>0.079</td>
<td>0.120</td>
</tr>
<tr>
<td>20-49 Year Olds (C)</td>
<td>0.079</td>
<td>0.125</td>
</tr>
<tr>
<td>T-C Difference</td>
<td>0.000</td>
<td>0.013</td>
</tr>
<tr>
<td>Non-hispanic private enrollment rate, 1970</td>
<td>0.014</td>
<td>0.015</td>
</tr>
<tr>
<td><strong>C. Other pre-existing district characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>public enr. share low-English Hispanic, 1976</td>
<td>0.026</td>
<td>0.044</td>
</tr>
<tr>
<td>public school enrollment, 1976</td>
<td>14849</td>
<td>41486</td>
</tr>
<tr>
<td>per-pupil property tax revenue, 1971-72</td>
<td>636</td>
<td>292</td>
</tr>
<tr>
<td>per-pupil total expenditures, 1971-72</td>
<td>1084</td>
<td>259</td>
</tr>
<tr>
<td>=1 if elementary-high organization</td>
<td>0.172</td>
<td></td>
</tr>
<tr>
<td>=1 if above-average #public schools</td>
<td>0.579</td>
<td></td>
</tr>
<tr>
<td>=1 if at least one private school, 1976</td>
<td>0.712</td>
<td></td>
</tr>
<tr>
<td>=1 if center city, 1972</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td><strong>Number of Observations</strong></td>
<td></td>
<td>233</td>
</tr>
<tr>
<td><strong>Number of MSAs</strong></td>
<td></td>
<td>23</td>
</tr>
</tbody>
</table>

*Note:* The unit of observation is either a unified school district or a combination of school districts that serve the same geographic area and all elementary and secondary grades (one secondary district plus a number of elementary districts). The sample includes all such observations with complete data on the characteristics listed. See text for more details on sample construction and description of data sources. † The instrument is the predicted 1976 to 2000 change in low-English Hispanic public school enrollment share arrived at by apportioning low-English Hispanic first- and second-generation immigrants from Mexico and enrolled in public school in 2000 to the school districts where immigrants of all ages from Mexico settled in 1970. An individual is classified as "low-English" if reported to speak English "not at all," "not well," or "well" and "second generation" if native-born to at least one Mexican-born parent (or if both parents are foreign-born, to a Mexican-born mother) who arrived in the U.S. in 1976 or later.
### Table II. Are the Identifying Assumptions Satisfied?
The Instrument and District Observables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Coefficient (standard error) on the Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Change in immigrant share</td>
<td></td>
</tr>
<tr>
<td>$\Delta$ public enr. share low-English Hispanic, 1976-2000</td>
<td>0.431*** (0.099)</td>
</tr>
<tr>
<td>B. Pre-existing levels of outcomes</td>
<td></td>
</tr>
<tr>
<td>Share of MSA's non-Hispanics in district, 1970:</td>
<td></td>
</tr>
<tr>
<td>0-19 Year Olds (T)</td>
<td>-0.238 (0.196)</td>
</tr>
<tr>
<td>20-49 Year Olds (C)</td>
<td>-0.242 (0.206)</td>
</tr>
<tr>
<td>T-C Difference</td>
<td>0.005 (0.017)</td>
</tr>
<tr>
<td>Non-hispanic private enrollment rate, 1970</td>
<td>0.000 (0.007)</td>
</tr>
<tr>
<td>C. Other pre-existing district characteristics</td>
<td></td>
</tr>
<tr>
<td>public enr. share low-English Hispanic, 1976</td>
<td>0.253*** (0.040)</td>
</tr>
<tr>
<td>ln(public school enrollment, 1976)</td>
<td>-1.69 (1.95)</td>
</tr>
<tr>
<td>per-pupil property tax revenue, 1971-72</td>
<td>-197.2 (188.8)</td>
</tr>
<tr>
<td>per-pupil total expenditures, 1971-72</td>
<td>-172.5 (167.6)</td>
</tr>
<tr>
<td>=1 if above-average #public schools</td>
<td>enrollment, land area, 1972</td>
</tr>
<tr>
<td>=1 if at least one private school, 1976</td>
<td>-1.350*** (0.416)</td>
</tr>
<tr>
<td>=1 if center city</td>
<td>0.001 (0.332)</td>
</tr>
</tbody>
</table>

Number of Observations: 233
Number of MSAs: 23

**Notes:** Each entry gives the coefficient (standard error) on the predicted 1976 to 2000 change in low-English Hispanic public school enrollment share arrived at by apportioning low-English Hispanic first- and second-generation immigrants from Mexico and enrolled in public school, in the 2000 Census of Population, to the school districts where immigrants of all ages from Mexico settled in 1970, according to the 1970 Fourth Count (Population) School District Data Tapes. An individual is classified as "low-English" if reported to speak English "not at all," "not well," or "well," and "second generation" if native-born to at least one Mexican-born parent (or if both parents are foreign-born, to a Mexican-born mother) who arrived in the U.S. in 1976 or later. All regressions are based on the full sample of California unified and combined elementary x high school districts (233 districts in 23 MSAs), and include fixed effects for MSA by district type. Standard errors (in parentheses) are calculated to be robust to arbitrary error correlation within MSA. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.
<table>
<thead>
<tr>
<th></th>
<th>Dep. Var.: Δ Share of MSA's non-Hispanic Population in District</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ages 0-19 (T)</td>
<td>Ages 20-49 (C)</td>
<td>Difference (T-C)</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Coefficient (standard error)</td>
<td>-0.0729</td>
<td>-0.0425</td>
<td>-0.0304**</td>
</tr>
<tr>
<td>on instrument</td>
<td>(0.0546)</td>
<td>(0.0548)</td>
<td>(0.0140)</td>
</tr>
<tr>
<td>Number of Districts</td>
<td>233</td>
<td>233</td>
<td>233</td>
</tr>
<tr>
<td>Number of MSAs</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

**A. Full Sample: 1970-2000**

<table>
<thead>
<tr>
<th></th>
<th>Coefficient (standard error)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>on instrument</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>-0.0041</td>
<td>0.0312</td>
<td>-0.0353*</td>
</tr>
<tr>
<td>Number of Districts</td>
<td>161</td>
<td>161</td>
<td>161</td>
</tr>
<tr>
<td>Number of MSAs</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>Coefficient (standard error)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>on instrument</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>-0.0217</td>
<td>-0.0151</td>
<td>-0.0066</td>
</tr>
<tr>
<td>Number of Districts</td>
<td>161</td>
<td>161</td>
<td>161</td>
</tr>
<tr>
<td>Number of MSAs</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

**C. 1960-1970**

**Notes:** The dependent variable in the first two columns is the change in the share of an MSA's non-Hispanic population residing in a district, for the age group specified in the column and over the period and for the sample of districts specified in the panel. The instrument is the predicted 1976 to 2000 change in low-English Hispanic public school enrollment share arrived at by apportioning low-English Hispanic first- and second-generation immigrants from Mexico and enrolled in public school to the school districts where immigrants of all ages from Mexico settled in 1970. An individual is classified as "low-English" if reported to speak English "not at all," "not well," or "well," and "second generation" if native-born to at least one Mexican-born parent (or if both parents are foreign-born, to a Mexican-born mother) who arrived in the U.S. in 1976 or later. All regressions are based on the full sample of California unified and combined elementary x high school districts, unless otherwise noted, and include fixed effects for MSA by district type. Standard errors (in parentheses) are calculated to be robust to arbitrary error correlation within MSA. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.
Table IV. TSLS and OLS Estimates for Non-Hispanic District Choice

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (C), 1970</td>
<td>0.079</td>
<td>0.060</td>
<td>0.073</td>
<td>0.111</td>
</tr>
</tbody>
</table>

### A. Two Stage Least Squares

**Coefficient (standard error) on:**

- Δ public school enrollment share
  - low-English Hispanic, 1976-2000
    - -0.0705**
    - (0.0318)
  - -0.0802*
  - (0.0441)
  - -0.0790*
  - (0.0438)
  - -0.0674**
  - (0.0288)
  - -0.0764*
  - (0.0385)
  - -0.0164
  - (0.0206)

- RMSE
  - 0.0147
  - 0.0147
  - 0.0147
  - 0.0119
  - 0.0153
  - 0.0101

- First stage partial F-stat on instrument
  - 19.0
  - 25.8
  - 27.7
  - 18.2
  - 12.4
  - 14.1

### B. Ordinary Least Squares

**Coefficient (standard error) on:**

- Δ public school enrollment share
  - low-English Hispanic, 1976-2000
    - -0.0215**
    - (0.0100)
  - -0.0035
  - (0.0107)
  - -0.0040
  - (0.0105)
  - -0.0183*
  - (0.0101)
  - -0.0218**
  - (0.0102)
  - -0.0068
  - (0.0100)

- RMSE
  - 0.014
  - 0.0132
  - 0.0133
  - 0.0111
  - 0.0144
  - 0.0101

- Number of Districts
  - 233
  - 233
  - 233
  - 211
  - 193
  - 40

- Number of MSAs
  - 23
  - 23
  - 23
  - 23
  - 23
  - 19

**Controls:**

- MSA fixed effects X X X X X X
- District type fixed effect X X X X
- MSA by district type fixed effects X X X
- Pre-existing district characteristics† X X
- T-C diff. in dep. var., 1970 X

### Notes:

"T" represents 0-19 year olds and "C" represents 20-49 year olds. The first row gives the share of an MSA's non-Hispanic 20-49 year olds residing in the average district in 1970. The instrument used in panel A is the predicted 1976 to 2000 change in low-English Hispanic enrollment share arrived at by apportioning low-English Hispanic first- and second-generation immigrants from Mexico and enrolled in public school in 2000 to the school districts where immigrants of all ages from Mexico settled in 1970. An individual is classified as "low-English" if reported to speak English "not at all," "not well," or "well," and "second generation" if native-born to at least one Mexican-born parent (or if both parents are foreign-born, to a Mexican-born mother) who arrived in the U.S. in 1976 or later. Standard errors (in parentheses) are calculated to be robust to arbitrary error correlation within MSA. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. † Pre-existing district characteristics include: natural log of 1976 total enrollment, per-pupil property tax revenue and per-pupil total expenditure (1971-72), and indicators for center city district (1969-70), above-average number of public schools given land area and enrollment (1972), and with at least one private school (1976).

39
Table V. Sensitivity of the Estimates for Non-Hispanic District Choice and Implied Displacement Rates

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>TSLS (1)</th>
<th>OLS (2)</th>
<th>Children (3)</th>
<th>HHs (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) T-C Diff: Δt_de (Share of MSA non-Hispanic Pop. in District), 1970-2000</td>
<td>-0.0705***</td>
<td>-0.0215**</td>
<td>1.15</td>
<td>1.60</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0147</td>
<td>0.0140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) T-C Diff: Δln(t_de/1-t_de), 1970-2000</td>
<td>-1.213***</td>
<td>-0.281**</td>
<td>1.44</td>
<td>2.00</td>
</tr>
<tr>
<td>Marginal Effect (at mean t_de,1970)</td>
<td>-0.0880</td>
<td>-0.0204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>0.209</td>
<td>0.191</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) T-C Diff: Δln(pop_d), 1970-2000</td>
<td>-1.077***</td>
<td>-0.242**</td>
<td>1.39</td>
<td>1.92</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.195</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(d) Δpop_d,T/pop_d, 1970-2000</td>
<td>-0.250***</td>
<td>-0.062**</td>
<td>1.24</td>
<td>1.67</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0445</td>
<td>0.0409</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A. Choice of Outcome

B. Weighting

Weight: 1970 District Population of non-Hispanic 0-19 Year Olds

(a) T-C Diff: Δt_de (Share of MSA non-Hispanic Pop. in District), 1970-2000 | -0.138*** | -0.074** | 0.90 | 1.28 |
| RMSE                | 0.0195  | 0.0184  |             |        |
| (b) T-C Diff: Δln(t_de/1-t_de), 1970-2000 | -0.940*** | -0.351*** | 0.99 | 1.39 |
| Marginal Effect (at weighted mean t_de,1970) | -0.152 | -0.057 |             |        |
| RMSE                | 0.161   | 0.15    |             |        |

Notes: **“T” represents 0-19 year olds and “C” represents 20-49 year olds. Each entry in columns (1) and (2) represents a different regression. All regressions are based on the full sample of California unified and combined elementary-high school districts (233 districts in 23 MSAs). Column (3) gives an estimate of the number of non-Hispanic child departures for every low-English Hispanic arrival over 1970 to 2000 to a district of average initial (1970) size, based on the TSLS estimate in the row and the assumption that the low-English Hispanic share increased by 0.1. Column (4) gives an estimate of the number of departures of non-Hispanic households with children with the arrival of a household with at least one low-English child enrolled in public schools. To calculate this figure, we rescale the TSLS coefficients upward - to adjust for the fact that slightly over half of the comparison group of 20-49 year olds are in households with children (see text and Appendix) - and convert the number of low-English Hispanic public school enrollees needed to deliver a 0.1 increase in their public enrollment share into the number of households they represent. Specifications in Panel A are unweighted, and specifications in Panel B are weighted by the 1970 non-Hispanic population of 0-19 year olds. Throughout, the explanatory variable of interest is the 1976 to 2000 change in the share Hispanic low-English in public school enrollment, instrumented (in column (1)) with the predicted change arrived at by apportioning low-English Hispanic first- and second-generation immigrants from Mexico and enrolled in public school in 2000 to the school districts where immigrants of all ages from Mexico settled in 1970. (See notes to earlier tables for more details.) Marginal effects (row (b) of both panels) are calculated by multiplying the regression coefficient by s_nh*(1-s_nh). Standard errors (in parentheses) are clustered on MSA. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.
Table VI. TSLS and OLS Estimates for Non-Hispanic Private School Enrollment

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>All (Unified+Combined Elem/High)</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Mean, 1970</td>
<td>0.0137</td>
</tr>
</tbody>
</table>

### A. Two Stage Least Squares

<table>
<thead>
<tr>
<th>Coefficient (standard error) on:</th>
<th>Panel (1)</th>
<th>Panel (2)</th>
<th>Panel (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆ public school enrollment share</td>
<td>0.184*</td>
<td>0.260**</td>
<td>0.274**</td>
</tr>
<tr>
<td>low-English Hispanic, 1976-2000</td>
<td>(0.089)</td>
<td>(0.100)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0418</td>
<td>0.0433</td>
<td>0.0427</td>
</tr>
<tr>
<td>First stage partial F-stat on instrument</td>
<td>19.0</td>
<td>17.3</td>
<td>27.7</td>
</tr>
</tbody>
</table>

### B. Ordinary Least Squares

<table>
<thead>
<tr>
<th>Coefficient (standard error) on:</th>
<th>Panel (1)</th>
<th>Panel (2)</th>
<th>Panel (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆ public school enrollment share</td>
<td>0.061***</td>
<td>0.074***</td>
<td>0.069**</td>
</tr>
<tr>
<td>low-English Hispanic, 1976-2000</td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0403</td>
<td>0.0398</td>
<td>0.0391</td>
</tr>
<tr>
<td>Number of Districts</td>
<td>233</td>
<td>233</td>
<td>233</td>
</tr>
<tr>
<td>Number of MSAs</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

**Controls:**
- MSA by district type fixed effects: X X X
- Has private school, 1976 (=1): X X
- Other pre-existing district characteristics†: X

**Notes:** The first row gives the share non-Hispanics of school age attending private school in the average district in 1970. The dependent variable in Panels A and B is the 1970 to 2000 change in the district's non-Hispanic private school enrollment rate. The instrument used in panel A is the predicted 1976 to 2000 change in low-English Hispanic enrollment share arrived at by apportioning low-English Hispanic first- and second-generation immigrants from Mexico and enrolled in public school in 2000 to the school districts where immigrants of all ages from Mexico settled in 1970. See notes to earlier tables and text for more details. Standard errors (in parentheses) are calculated to be robust to arbitrary error correlation within MSA (23 MSAs).

***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively. † Other pre-existing district characteristics include: the 1970 non-Hispanic private school enrollment rate, the natural log of 1976 total enrollment, per-pupil property tax revenue and per-pupil total expenditure (1971-72), and indicators for center city district (1969-70) and above-average number of public schools given land area and enrollment (1972).
### Appendix Table I. Construction of the Instrument

#### Share of Mexicans in 1970:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Value</th>
<th>District</th>
<th>MSA</th>
<th>Rank</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1626</td>
<td>LOS ANGELES UNIF</td>
<td>Los Angeles, CA</td>
<td>27</td>
<td>0.2539</td>
</tr>
<tr>
<td>2</td>
<td>0.0124</td>
<td>SAN DIEGO CITY UNIF</td>
<td>San Diego, CA</td>
<td>84</td>
<td>0.1136</td>
</tr>
<tr>
<td>3</td>
<td>0.0094</td>
<td>SWEETWATER UNION HIGH</td>
<td>San Diego, CA</td>
<td>13</td>
<td>0.3028</td>
</tr>
<tr>
<td>4</td>
<td>0.0091</td>
<td>SAN FRANCISCO UNIF</td>
<td>San Francisco, CA</td>
<td>67</td>
<td>0.1435</td>
</tr>
<tr>
<td>5</td>
<td>0.0066</td>
<td>SANTA ANA UNIF</td>
<td>Anaheim, CA</td>
<td>42</td>
<td>0.2013</td>
</tr>
<tr>
<td>6</td>
<td>0.0064</td>
<td>EL MONTE UNION HIGH</td>
<td>Los Angeles, CA</td>
<td>4</td>
<td>0.5145</td>
</tr>
<tr>
<td>7</td>
<td>0.0058</td>
<td>ALHAMBRA CITY ELEM-HIGH</td>
<td>Los Angeles, CA</td>
<td>14</td>
<td>0.3015</td>
</tr>
<tr>
<td>8</td>
<td>0.0051</td>
<td>EAST SIDE UNION HIGH</td>
<td>San Jose, CA</td>
<td>33</td>
<td>0.2385</td>
</tr>
<tr>
<td>9</td>
<td>0.0050</td>
<td>SAN JOSE UNIF</td>
<td>San Jose, CA</td>
<td>69</td>
<td>0.1406</td>
</tr>
<tr>
<td>10</td>
<td>0.0045</td>
<td>OAKLAND CITY UNIF</td>
<td>Oakland, CA</td>
<td>95</td>
<td>0.0961</td>
</tr>
</tbody>
</table>

#### Instrument:

**A. CA Districts Ranked by Share of U.S. Mexican-Born Population, 1970**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Value</th>
<th>District</th>
<th>MSA</th>
<th>Rank</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>0.0006</td>
<td>LE GRAND UNION HIGH</td>
<td>Merced, CA</td>
<td>1</td>
<td>0.6363</td>
</tr>
<tr>
<td>26</td>
<td>0.0023</td>
<td>SANTA PAULA UNION HIGH</td>
<td>Oxnard-Ventura, CA</td>
<td>2</td>
<td>0.6147</td>
</tr>
<tr>
<td>36</td>
<td>0.0018</td>
<td>DELANO JOINT UNION HIGH</td>
<td>Bakersfield, CA</td>
<td>3</td>
<td>0.5401</td>
</tr>
<tr>
<td>6</td>
<td>0.0064</td>
<td>EL MONTE UNION HIGH</td>
<td>Los Angeles, CA</td>
<td>4</td>
<td>0.5145</td>
</tr>
<tr>
<td>77</td>
<td>0.0007</td>
<td>LIBERTY UNION HIGH</td>
<td>Oakland, CA</td>
<td>5</td>
<td>0.4169</td>
</tr>
<tr>
<td>61</td>
<td>0.0010</td>
<td>FALLBROOK UNION HIGH</td>
<td>San Diego, CA</td>
<td>6</td>
<td>0.3987</td>
</tr>
<tr>
<td>44</td>
<td>0.0015</td>
<td>DINUBA UNIFIED</td>
<td>Fresno, CA</td>
<td>7</td>
<td>0.3759</td>
</tr>
<tr>
<td>55</td>
<td>0.0012</td>
<td>CUTLER-OROS UNIF</td>
<td>Visalia, CA</td>
<td>8</td>
<td>0.3701</td>
</tr>
<tr>
<td>127</td>
<td>0.0003</td>
<td>KINGSBURG JT UNION HIGH</td>
<td>Fresno, CA</td>
<td>9</td>
<td>0.3363</td>
</tr>
<tr>
<td>54</td>
<td>0.0012</td>
<td>FILLMORE UNIF</td>
<td>Oxnard-Ventura, CA</td>
<td>10</td>
<td>0.3311</td>
</tr>
</tbody>
</table>

**B. CA Districts Ranked by Predicted Change in Hispanic-low-English Share in Public School Enrollment, 1976-2000**


Figure I. Trends in the Mexican Share of the Child Population in the United States

Note: Sources are SIE (1976) and Census PUMS. Sample includes 0-17 year olds and Mexicans include both the Mexican-born and the children of Mexican-born parents.
Figure II. First-Stage Relationship

1970 to 2000 Change in ELL Hispanic Pub. Enr. Share

Notes: Regression adjusted for MSA x district type fixed effects. Sample includes unified and combined HS-elem districts in 23 MSAs in CA.
Figure III. Reduced-Form Relationship
T-C: 1970-2000 Change in Share of MSA's non-Hispanics in District

Notes: T=0-19 year olds. C=20-49 year olds. Regression adjusted for MSA x district type fixed effects. Sample includes unified and combined HS-elem districts in 23 MSAs in CA.
Figure IV. Pre-Trends in Outcomes?
T-C Diff: Change in Share of MSA’s non-Hispanics in District

A. 1970-2000

B. 1960-1970

Notes: T=0-19 year olds. C=20-49 year olds. Regression adjusted for MSA x district type fixed effects. Sample includes 13 MSAs in CA with 1960 data available.