# Multiregional cohort enrollment projections: Matching methods to enrollment policies 

Stuart H. Sweeney ${ }^{1}$<br>and<br>Erin J. Middleton<br>Department of Geography<br>University of California<br>Santa Barbara, CA 93106-4060<br>email: sweeney@geog.ucsb.edu<br>phone: 805-893-5647<br>fax: 805-893-3146<br>forthcoming - Population, Space, and Place

February 18, 2005

[^0]
#### Abstract

School districts throughout California are struggling with uncertainty in primary school enrollment levels. Some areas of the state are witnessing soaring enrollment levels due to immigration and internal migration while other areas with expensive housing stock and aging populations are expected to decline. State level budget problems, the increasing popularity of home/private schooling, and school choice laws at the local level have resulted in extreme uncertainty about the expected enrollment levels at individual schools in some districts. Santa Barbara city schools provide an excellent example of enrollment instability and the costs associated with poor enrollment forecasts. The associated costs include consternation among employees due to last minute hires or reassignments, community unrest as school closures are threatened or carried out, and ineffective capital planning/allocation. This research uses student level records from the Santa Barbara city schools to evaluate the nature of the instability in past enrollment forecasts. The primary objective of the paper is to propose and to evaluate a Multiregional Cohort-Enrollment projection model that we believe is better suited to prevailing enrollment policies. Though the model as proposed still requires further refinements it appears promising as a forecasting and policy evaluation tool in districts that are increasing the level of school choice.


Key Words: School enrollment projections, student records, multistate demographic models.

## 1 Introduction

The quality of enrollment forecasts is critically important in most communities. Near-term forecasts are used as an input for determining staffing levels while both near- and long-term forecasts are used for capital improvement and facilities planning. Accurate forecasts are especially important in times of change; both increasing and decreasing numbers of students create challenges for school officials that affect not only planning decisions, but also the real and perceived quality of instruction. Indeed, the perceived quality of school facilities and instruction is probably only second to crime in its ability to foster strong citizen awareness and public participation. If school crowding results from poor facilities planning in a growing community, parents will forcefully express their discontent at school board meetings. Similarly, if parents perceive school closures and staffing reductions as premature or unnecessary they can quickly organize and emerge as a formidable political movement.

Yet, even given the widespread awareness of the consequences of inaccurate forecasts, enrollment projections methodology has remained relatively constant over the past fifty years. In many communities, the underlying demographic and economic forces determining enrollment levels can be extremely complex. Ideally, school planners would be able to choose projection models tailored to capture key aspects of community change and local policy. While existing methods do work well for some demo-economic growth regimes, they are a poor match for others. This paper proposes a sub-district enrollment projection model that is particularly suited for an open enrollment system in a district with substantial ethnic residential segregation. The prospect of students enrolling outside of their assigned attendance area - in search of better schools, specific programs, or for cultural comfort - complicates the projections task by increasing uncertainty in school-specific enrollment.

One significant reason for the lack of methodological innovation is due to data availability. The two most typically used sources of data, school records and censuses, both have shortcomings that translate into model deficiencies. The availability of more detailed school records is critical in developing more sophisticated models, especially ones that take into account intradistrict flows of students between schools as well as interdistrict connectivity. This is important information for schools with stable enrollment below school capacity and schools experiencing declining enrollment. In both cases, a system of open enrollment where an enrollment gap exists between actual enrollment and school capacity can lead to fewer students than expected in some schools and more in others. As school choice laws continue to be promoted at the national level, we suspect our work will find application in other communities throughout the U.S. We also suspect that the approach would be useful in other countries since school choice naturally arises out of concern for equity in educational opportunity when ethnic groups (or other sub-populations) are residentially segregated and there are real or perceived differences in the quality of education tied to place of residence.

The plan of the paper is to propose and to evaluate an enrollment projection model that is designed to accommodate complications resulting from an open enrollment system. Based on a multiregional model developed in demographic studies, the multiregional cohort-enrollment model used here is compared to the traditional cohort component model. Section 2 of the paper explores the complex social, economic, and policy interactions that can complicate projections in some districts. Section 3 introduces the multiregional model and contrasts it with existing enrollment projections methodologies. Sections 4 through 6 comprise the application of the method to the Santa Barbara Elementary School District (SBESD). Sections 7 contains our concluding remarks.

## 2 Enrollment Demographics: Simple or Complex?

Our point of departure is the California public education system since the data we use in our application are from the Santa Barbara Elementary School District (SBESD). Last year, over six million students were enrolled in the California public schools,and approximately 390,000 fulltime administrators, teachers and support staff were employed to meet their needs (Educational Demographics Office 2004). In California, each school district receives state money based on the number of students in attendance in a given year; consequently, any changes in student enrollment are directly linked to the size of a school's annual budget. Although enrollment is expected to increase for California as a whole through 2006, there are regional and sub-regional differences in where that growth will occur. Some areas will experience explosive growth of the school-aged population while other areas will experience stable or even declining enrollment. Differing growth regimes can occur even within a single county. For example, northern Santa Barbara county has experienced rapid population growth and plans to add four new schools whereas southern Santa Barbara county, with a stable overall population but declining student population, has been struggling with budget cuts and actually closed an elementary school last year (Kuznia 2004).

The demographics of enrollment projections appear simple. The overall size and age structure of the population are sufficient to determine near term forecasts whereas fertility rates are also needed for longer term forecasts. More formally, enrollment accounts are typically summarized by,

$$
\begin{equation*}
N(g+1)^{\{t+1\}}=N(g)^{\{t\}}-D(g)^{\{t, t+1\}}+A(g)^{\{t, t+1\}}-W(g)^{\{t, t+1\}} \tag{1}
\end{equation*}
$$

where $N(g+1)^{\{t+1\}}$ is the student cohort population in grade $g$ in year $t+1 ; N(g)^{\{t\}}$ is the number of students enrolled in grade $g$ at time $t, D(g)^{\{t, t+1\}}$ are the number of deaths among grade cohort $g$ between periods $t$ and $t+1^{1} ; A(g)^{\{, t+1\}}$ (ascensions) is the number of students that enter into grade $g$ after the start of the school year; and $W(g)^{\{t, t+1\}}$ (withdrawals) is the number of students that leave the school after the start of the year. For the first year of formal schooling (kindergarten; $g=0$ ),

$$
\begin{equation*}
N(0)^{\{t\}}=\rho_{1} P(5)^{\{t\}}+\rho_{2} P(6)^{\{t\}}+\rho_{3} P(7)^{\{t\}} \tag{2}
\end{equation*}
$$

where $P(X)$ is the population of age $X$ and $\rho_{i}$ is the fraction of that age group that are enrolled in kindergarten. If projections go beyond 5 years then a fertility model is needed to project the population at young ages. In a stable district, the ratio $N(g)^{\{t\}} / N(g-1)^{\{t-1\}}$ is often taken as an adequate basis for enrollment projections. This is essentially the core of the cohort-component model. ${ }^{2}$

A central assumption allowing for simple enrollment demographics is that place of residence determines place of schooling. Even without an open enrollment policy, which negates that assumption, the increased popularity of private schooling and home schooling erode the residence-school coupling. There will usually be some flow of students between the public schools, private schools, and home schools. As public school policies change or educational programs are introduced or terminated, we would expect some degree of repulsion or attraction from the alternatives to public schooling. These changes will be manifest in the terms,

[^1]$A(g)^{\{t, t+1\}}$ and $W(g)^{\{t, t+1\}}$ that are typically assumed fixed and negligible. Although there are no public records that indicate the flows of students between public and private education, in California there are records that indicate school enrollment for private schools greater than five students. It is estimated that $10 \%$ of students in California are enrolled in private schools, a figure that has remained relatively constant for the last several years (California Department of Education 2004). Home schooling is much more difficult to estimate. Although in general the proportion of students enrolled in non-public education remains relatively constant, changes in education policies, either at the state or local level, may lead to parents altering the education plans of their children.

The presence of residence-school coupling also promotes demographic order, thus increasing stability and simplifying projects, at a deeper level. Economic theories of neighborhood choice suggest that school quality is partially capitalized into the house value (Roback 1982, Tiebout 1956). In other words, home buyers with school-aged children (or planning for children) are willing to pay more for a house, ceteris paribus, if the local school is of higher quality. As such, income segregated neighborhoods will evolve naturally and will roughly align with variation in school quality and the quality of other local public goods and services (Epple and Romer 1991, Epple and Sieg 1999). Yet that sorting result depends on rigid attendance boundaries that your children must attend the local school - and to some extent on school funding being dependent on local property taxes. In California the second condition is almost always violated ${ }^{3}$ (due to Proposition 13) and the first will not hold under open enrollment. A policy of open enrollment will allow for school choice to be decoupled from neighborhood choice. Over time, this would promote increased levels of intradistrict mobility of students.

Instability can also arise from the complex interactions among local economic factors, aspects of household and family demography, local cultures, and local policies. For example, the age structure and the household structure interact strongly with local economic conditions. Some communities in California have housing that is so expensive that families with children, even with good two-income salaries, cannot afford to purchase homes. Although the family may work and live in the community as young adult renters, the childbearing decision will often precipitate a move to a more affordable community. Over time this skews the age distribution towards older age groups and causes declining enrollment. It turns out, however, that leaving the community is partly a culturally determined outcome. Hispanic households, especially recent immigrants, react to the higher home prices by increasing the number of families per household. Thus, in SBESD, though the median home price is one of the most most expensive within California, the elementary schools have witnessed an increasing number of students qualifying for the free lunch program (California Department of Education 2004). With ethnically segregated neighborhoods, these student increases will be spatially concentrated. Further complicating the issue, even if some families leave the area in search of affordable housing, their children may return as interdistrict transfers if the parents continue to work in the community.

In general, even without an open enrollment policy there may be complex interacting social, economic, and demographic dynamics that erode the assumption of simple enrollment demographics. Most school planners are aware of the complications and attempt to incorporate them into the projections, often having to rely on more informal revisions when hard data are not available. We would argue that once an open enrollment policy is introduced then

[^2]both intradistrict and interdistrict student flows emerge as a fundamental feature of the enrollment demographics and need to be formally integrated into the projections model. Although residence-school coupling has been the norm in the past, open enrollment is re-emerging as a common policy. During the tumultuous period of desegregation in the 1960s, some districts relaxed mandatory enrollment based on location as a way to encourage voluntary desegregation, and later replaced this policy of school choice with busing. As this policy of open enrollment is again gaining popularity, in part as a way to combat increased segregation in schools and differentials between under- and over-achieving schools within the same district, there is a need for developing enrollment models that take into account student flows between schools.

## 3 Multiregional cohort-enrollment projection

The multiregional cohort-enrollment (MCE) projection model is derived by altering the accounts in equation (1) to reflect intra- and interdistrict student flows. For a district with two schools ( $s=1,2$ ) and connections to all other districts $(s=R)$ enrollment accounts are summarized by,

$$
\begin{align*}
& N_{1}^{\{t+1\}}=N_{1}^{\{t\}}+M_{21}^{\{t, t+1\}}-M_{12}^{\{t, t+1\}}+M_{R 1}^{\{t, t+1\}}-M_{1 R}^{\{t, t+1\}}  \tag{3}\\
& N_{2}^{\{t+1\}}=N_{2}^{\{t\}}+M_{12}^{\{t, t+1\}}-M_{21}^{\{t, t+1\}}+M_{R 2}^{\{t, t+1\}}-M_{2 R}^{\{t, t+1\}} \tag{4}
\end{align*}
$$

where the subscripts on the student stocks, $N_{i}$, and flows, $M_{i j}$, indicate origin (i) and destination $(j)$. Given access to student level records indicating the school attended and the home address, the intradistrict student flows ( $M_{12}$ and $M_{21}$ ) as well as inflows from outside the district ( $M_{R j}$ ) are observable and can be decoupled from the more general categories of ascensions $(A)$ and withdrawals ( $W$ ) in equation (1). In our application, the flows exiting the district are essentially a residual category. That is, the composition of $M_{i R}$ contains students that do not change residence but choose to attend home schools or private schools and students that move their residence outside the district. Given more detailed attendance data for home schools and private schools, or data for multiple interacting districts, it would be possible to isolate the components in $M_{i R}$.

The accounts in (3) and (4) only capture the aggregate flows of students whereas we also need to account for the cohort progression of students within each school. Matrix expressions for the cohort-specific model can be derived following the standard approach for multiregional projections (Rogers 1975, Rogers 1985). Expanding the model from (3) and (4) to include three grades $(g=0,1,2)$ the matrix expression for the MCE model is,

$$
\left[\begin{array}{c}
0  \tag{5}\\
N_{1}^{\{t+1\}}(0) \\
N_{1}^{\{t+1\}}(1) \\
N_{1}^{\{t+1\}}(2) \\
---- \\
0 \\
C_{1} \\
N_{2}^{\{t+1\}}(0) \\
N_{2}^{\{t+1\}}(1) \\
N_{2}^{\{t+1\}}(2)
\end{array}\right]=\left[\begin{array}{c:c} 
& C_{12} \\
--- & --- \\
& C_{21} \\
& C_{22} \\
& \\
&
\end{array}\right]\left[\begin{array}{c}
B_{1}^{\{t\}} \\
N_{1}^{\{t\}}(0) \\
N_{1}^{\{t\}}(1) \\
N_{1}^{\{t\}}(2) \\
--- \\
B_{2}^{\{t\}} \\
N_{2}^{\{t\}}(0) \\
N_{2}^{\{t\}}(1) \\
N_{2}^{\{t\}}(2)
\end{array}\right]+\left[\begin{array}{c}
0 \\
M_{R 1}^{\{t, t+1\}}(0) \\
M_{R 1}^{\{t, t+1\}}(1) \\
M_{R 1}^{\{t, t+1\}}(2) \\
---- \\
0 \\
M_{R 2}^{\{t, t+1\}}(0) \\
M_{R 2}^{\{t, t+1\}}(1) \\
M_{R 2}^{\{t, t+1\}}(2)
\end{array}\right]
$$

where the diagonal and off-diagonal blocks, $C_{i i}$ and $C_{i j}$, of the partitioned survivorship matrix C are of the form ${ }^{4}$,

$$
C_{i i}=\left[\begin{array}{cccc}
0 & \left(p_{b, i i}-m_{b, i R}\right) & a_{b, i i} & 0  \tag{6}\\
0 & r_{0, i i} & \left(p_{0, i i}-m_{0, i R}\right) & a_{0, i i} \\
0 & 0 & r_{1, i i} & \left(p_{1, i i}-m_{1, i R}\right) \\
0 & 0 & 0 & r_{2, i i}
\end{array}\right]
$$

and

$$
C_{i j}=\left[\begin{array}{cccc}
0 & p_{b, i j} & a_{b, i j} & 0  \tag{7}\\
0 & r_{0, i j} & p_{0, i j} & a_{0, i j} \\
0 & 0 & r_{1, i j} & p_{1, i j} \\
0 & 0 & 0 & r_{2, i j}
\end{array}\right]
$$

The MCE projection model diverges from standard multiregional projection models in a few key respects. Since the cohort specification is grade-based instead of age-based there is a more complex structure in the survivorship matrix. Specifically, note that the survivorship matrix allows for remediation $\left(r_{g, .}\right)$, normal promotion ( $p_{g, .}$ ), and acceleration ( $a_{g, .}$ ). Another departure is that the 'birth' cohort enters from a separate sub-model. ${ }^{5}$ The $B_{i}$ elements represent the pre-school age children defined as,

$$
\begin{equation*}
B_{i}=\sum_{5 \leq x \leq 7} \rho_{i, x} P_{i}(x) \tag{8}
\end{equation*}
$$

where $P(x)$ is the population age $x$ as noted in the previous section. Also note that unlike multiregional population projections that have an ill-defined upper bound for each region's population, an enrollment model is capped by the physical capacity of a school and classroom specific maximums that may vary by grade. In the form presented here the MCE model can exceed such constraints but this is also true of the standard cohort-survival model. The primary difference is that our interest is focused on sub-district projections, ideally school-level projections, so the capacity constraint is more binding. We will present work on a constrained MCE model in a future paper. ${ }^{6}$

There are several other enrollment projection models in the literature, some of which are tailored to fit specific growth/policy regimes that differ from the regime targeted by the MCE model. Before turning to the application of the MCE model we offer a brief comparison of the MCE model to existing models. A summary of the characteristics of those alternatives is provided in Table 1. The review focuses on projection models used in district level planning efforts in the United States. ${ }^{7}$

The three most widely used methods are listed in the first three rows of Table 1 and share the common feature of relying on a set of fixed rates derived from past observations. In the case

[^3]of the cohort survival model the fixed rate is the cohort survival ratio $\left(N^{\{t+1\}}(g+1) / N^{\{t\}}(g)\right)$, for the ratio method the fixed rate is with respect to the population $\left(N^{\{t\}}(g) / P^{\{t\}}(x)\right)$, and for the housing unit method there are fixed student production rates per housing unit type $\left(N / H U_{k}\right)$. The cohort survival model is popular because it has minimal data requirements and generally produces accurate results for stable demo-economic regimes. The ratio method is attractive because it can be directly coupled with a population forecast to derive an enrollment forecast. The housing unit method is particularly suited to rapidly growing areas where one would expect that past observations of either of the first two types of fixed rates would be inaccurate. The last method is also frequently used to perform impact analysis of large scale residential developments. In each of the cases, though the rates derived are fixed they will frequently be estimated as the average of the most recent five years of observed rates.

The first two methods are typically applied at the district level and then total district enrollment by grade is allocated to schools. Thus, a primary difference with respect to the MCE model is that the sub-district projections can be derived directly. Also note that as with the MCE model, a birth model is needed to initiate the first year of attendance for the cohort survival model. The housing unit method provides a fundamentally different basis for projection than the MCE model; the two models target completely different demo-economic regimes and policy settings.

The regression and gradient cohort survival models are designed to capture and forecast the change in a series of CSRs (or student-population ratios) rather than relying on a fixed rate. In essence, both methods attempt to extrapolate change in CSRs and only differ in the way in which they weight past observations and derive those weights. The regression approach simply relies on least squares to fit either linear or non-linear (e.g. polynomial, exponential, or logistic) curves to the historical series, then uses the parameter estimates from the model to extrapolate rates to future periods. In contrast, gradient cohort survival uses a more general optimization approach to select optimal weights for past observations of CSRs such that they predict the most recently observed CSR. The MCE model is also a fixed rate, or stationary, model but neither of the methods for a univariate time series of rates would allow for updating since the survivorship matrix is a single multivariate observation. Still a non-stationary multiregional growth matrix could be derived using the methods in Sweeney and Konty (2002). That option may be explored in future work but is beyond the scope of the current paper.

The last three methods in Table 1 are all advanced approaches that attempt to capture more of the complexity underlying enrollment demographics. The modified regression approach increases complexity by including additional covariates to predict CSRs. The additional covariates could include net migration, birth data, or other features of the local community deemed important. The problem is that the cross-sectional results are by definition primarily suited to interpolation - within the time period - rather than extrapolation since the latter would require an observed future time series of the covariates. Although a time series of the covariates may be available as output from other models, the compounding of errors generally make this approach an ill-advised alternative. Multivariate time series methods could, in theory, be used but both the extreme data requirement and model sophistication have precluded their use.

The modified spatial filter is specifically targeted at capturing the spatial variation of CSRs within a district. It is particulary well-suited for identifying the impact of a school closing or informing school siting decisions. As a forecasting tool, however, the primary strength - a spatial cross-section - is not particularly useful as a basis for projection. Still the spatial filtering approach could be very useful in downscaling , from district to school, the model output from other methods. The spatial filtering approach is similar in spirit to the MCE model. Both models attempt to increase the spatial resolution of the model results. The difference is that the spatial filter relies on proximity based on physical distances whereas the MCE model specifies
connectivity as observed student flow interactions.
Campbell's (1997) extended demographic model is the most directly related to our work since his is also based on Rogers' (1985) multiregional projection model. The core of the extended demographic model is a standard multiregional population projection model that is disaggregated by age, race, sex, and place of residence. The model is calibrated for the greater Chicago region and surrounding suburbs using the 5 -year retrospective migration data from the decennial census. Also, the place of residence is operationalized using Census defined sampling areas which are spatially extensive and do not provide a one-to-one mapping to school district boundaries. The model projects the population of students in the city of Chicago and its surrounding suburbs and simultaneously produces enrollment projections using an age-tograde allocation matrix. While the extended demographic model is a novel approach and a good match for long-term projections in highly populated cities with large school districts, both the spatial scale and the time scale of the model will generally be too coarse for application to most enrollment forecasting problems.

Both the time scale and spatial scale are extremely important for the open enrollment context. The MCE model is designed for short-term forecasts at the sub-district level. This requires a 1 -year time scale since school choice decisions are generally made at the start of the year. The sub-district spatial scale substantially exceeds the scale that could be approached with the 5 -year retrospective migration questions. The MCE model and the extended demographic model also diverge in other respects beyond the spatial scale and the time scale. The MCE model only maps age to grade at the first time step. After students enter the system the grade-cohort method of the MCE model is based on the assumption that grade-based rates of remediation, promotion, acceleration, and movement will be relatively homogeneous among the age groups within a grade. Clearly there may be some tendency for higher remediation rates among younger students in a grade and higher acceleration rates among older students in a grade. Still, classification into either of those categories is relatively rare and is generally linked much more strongly to intellectual ability than age. Another difference is that the underlying dynamics of the MCE model predominantly are based on school choice rather than residential choice.

## 4 Application: MCE Projection for the Santa Barbara Elementary School District

The novelty of the MCE approach proposed in this paper is that it adapts multiregional projections methodology to an open enrollment policy setting. While multiregional projections methodology has a long history in demographic studies, it has rarely been applied to school enrollment ${ }^{8}$ and never as a way to account for intradistrict school transfers. As noted previously, the main benefit of the MCE model is that it isolates important demographic components of enrollment change. In contrast, aggregate cohort survival ratios reflect a composite of potentially disjoint processes - particularly the directional flows related to school choice. The conceptual benefits of a multiregional approach over a uniregional approach, and the known biases that can result from a uniregional model, are relatively well known in the literature (Rogers 1990).

[^4]One significant reason multiregional approaches have not been applied to school enrollment projections is due to their data requirement. Application of the model requires estimates of the parameters in the survivorship matrix (C), see (6) and (7). School planners typically use district-level data disseminated by the state; the aggregate data would not permit estimation of $\mathbf{C}$. Yet most districts maintain extensive administrative records that essentially function as a registry system. If the administrative records are augmented using GIS methods it becomes possible to create the necessary stocks and flows to derive occurrence-exposure estimates of the survivorship matrix parameters. The resulting data not only provides a basis for projection, but also are a rich source of information detailing individual student sojourns through the school system. Unlike simple cohort survival ratios, the multiregional growth matrix allows for interdependent evolution of the sources of growth and decline within each school.

This section reports on an application of the MCE projection model to the Santa Barbara Elementary School District (SBESD). After providing a brief history of the district's open enrollment policy in section 4.1, we describe the construction of the data sources for the MCE model (Section 4.2) and then present the model results (Section 4.3). The model results section contrasts the MCE projections with uniregional cohort-survival projections.

### 4.1 SBESD: An overview

Enrollment planning in Santa Barbara has always been complicated by strong residential segregation patterns. For example, the residential segregation patterns are unambiguous in Figure 1 ; the shading is over census blocks where $60 \%$ or more of the population under 18 is Hispanic. The map also indicates the locations of the district elementary schools and the attendance boundaries for each school. Given the residential segregation, it is not surprising that the school population reflects similar patterns. Tables 2 and 3 provide summary characteristics for each elementary school including the proportion of students in attendance with Spanish as their first language. ${ }^{9}$ Notice that Cleveland, Franklin, and McKinley all have student populations with more than $80 \%$ having Spanish as their first language and pull from predominantly Hispanic neighborhood enrollment areas. Also note that both the Santa Barbara Academy and the Open Alternative schools do not have attendance boundaries.

Starting in the mid-1960s, the Santa Barbara School District initiated a series of policies to promote ethnic balance among district schools. Early policies attempted busing and also resulted in the closure of two majority-hispanic schools. The planning context has always been contentious, marked by legal battles at both the state and local level. Two other significant processes that shaped local policies were the increasing pace of suburbanization during the 1970s and fears that policy changes would result in "white flight" from the public school system.

While open enrollment was proposed in the late 1970s, it was only enacted recently in its current form. The current policy, starting in 1997, allows students to attend any school within the district if there is space in the school. A key element of that policy is that transfer wavers have to be assigned randomly; schools cannot use transfers to actively promote ethnic balance. Still, the implicit goal of the policy is promote ethnic balance by allowing self-sorting through interschool mobility. Also note that busing is still used within the district. There are three high minority attendance areas which are bused to otherwise low minority enrollment schools (see Figure 1). Over the past decade, the district has opened new schools (Open Alternative and S.B. Academy) and allowed one to become a charter school (Peabody) to increase the breadth of educational philosophies among the district elementary schools.

[^5]
### 4.2 Data Sources

The data needed for the MCE projection model include: (1) information on student sojourns through the school system conditioned by their place of residence each year, and (2) information on place-specific births to construct the kindergarten cohort. The student-level administrative records were provided by the Santa Barbara School District. The records span five years starting with 1997/98 and ending with 2001/02. For each academic year the records include a unique student identifier, home address, school attended, date of birth, race/ethnicity, free or reduced lunch status, level of English proficiency and primary language. The district also provided a list of streets included within each school boundary. Our birth data are from the California Department of Health. The birth data contains the number of birth per year by zip code and ethnicity for the years 1992-2001.

These two raw data sources were then used to create the necessary input data for the MCE projection model. The transformation of the raw data included three steps: (1) generating the transfer status of each student, (2) spatially interpolating school attendance boundary births from the zip code births, and (3) modifying the data to define a consistent school-grade system over the observation period. These steps are described below.

1. Generating Transfer Status. The transaction data underlying $\mathbf{C}$ reflects attendance at the local school assigned by your address (diagonal blocks of $\mathbf{C}$ ) or attendance at schools outside of your assignment area (off-diagonal blocks of $\mathbf{C}$ ). The two sources of information needed for each student-year are the assignment area, based on the home address, and the school attended. The assignment area was derived by using ArcGIS to geocode the home addresses of each student and then assign each address to a the school assignment boundaries provided by the district. This was not an entirely routine task because of the inaccuracies that are common in administrative records. The student records were cleaned of spelling errors, P.O. boxes and zip codes that are not in the areas immediately surrounding school boundaries. Using these cleaned data files, address matching was performed in ArcMap. The geo-coded address database used was GDT Dynamap/2000. The result of address-matching yielded less than $0.5 \%$ unmatched addresses. Students were assigned to their "local" school using a point-in-polygon routine where the point was the home address and the polygon was the school attendance boundary.
Some adjustments were necessary for unmatched addresses and students with only post office boxes. In both cases, the school of origin was assumed to be the school that they first attended. If a student's first school of record is a school that does not have an attendance boundary (two schools within this data set do not have attendance boundaries), their school of origin is recorded as a missing value. The number of P.O. boxes and unmatched addresses relative to matched addresses is relatively small, with the former equal to approximately $0.7 \%$ to $1.0 \%$ of the latter.
2. Distributing Births. As noted already, the kindergarten cohort for each school is defined by (8) which, in turn, requires information on births for projections beyond five years. As such, the birth data by zip code need to be interpolated to define births by school boundary. We used a simple two-step spatial interpolation method. First, we used GIS operations to identify the proportion of the physical area of each census block zone, $b$, within school attendance zone, $s$, and zip code zone, $z$. In most cases the census blocks are completely contained within a school attendance zone and zip code zone. At the boundary of the school and zip zones, the block zones could potentially be partitioned into four subsets. The areal weights for each block zone were then used to assign census block populations by age class ( $\leq 18$ ) and ethnicity (Hispanic or Non-Hispanic) to the school
zone by zip code zone geography. Summing over the school zone by zip code zone label resulted in estimates of the proportion of each zip code population by age and ethnicity within a school attendance zone. Second, we used the school by zip code proportions to allocate the zip code births to school attendance zones. The two-step interpolation provided an estimate, $\hat{B}_{s, t}$, of the births by ethnicity for each school attendance zone from 1992 to 2001.
Given the estimate $\hat{B}_{s, t}$ it would be possible to survive the population forward using standard uniregional or multiregional population projections. This would provide the future 5 to 7 year old population as input to equation 8. A common short cut used in enrollment forecasting is to relate the birth population directly to the appropriately lagged kindergarten entry thus defining birth to kindergarten (BK) survival ratios. Like a grade cohort survival ratio, a birth to kindergarten ratio defines the ratio of kindergartners in year $t$ to births 5 and 6 years earlier. Estimates of the BK ratio from historical data are then used to project the future kindergarten cohort from observed or projected birth counts. This is essentially the method we used. First, given the birth year of the students in kindergarten, it was possible to find the proportion of students born 5 and 6 years previously. Next, this BK ratio was applied to the historical births by school to determine the possible pool of kindergartners available if all children born in the area attended their local school. Finally, the ratio of actual kindergartners to possible kindergartners predicted by births was determined at the school, regional and district level. In the latter two cases, the population was then disaggregated to the school level using the share of the 2002 school population in the region or district.

There were obviously several assumptions made in order to distribute births to schools. First, it was assumed that the under-18 population was representative of the distribution of new births. Second, an assumption was made that the block level populations are uniformly spatially distributed over each census block, when a much better assumption might have been to distribute this by, for example, street or housing densities. Finally, although population density varies from year to year, only the 1990 census block population densities were used. This creates a more static population dynamic than actually occurs.

The most important features of enrollment decline in Santa Barbara are its spatial variability; because Hispanic fertility rates are higher than non-Hispanic rates and the settlement patterns of the two ethnic groups are not evenly distributed, certain schools could experience much more rapid decline than others. Because births are available by zip code, at least a rough approximation of how spatial differences in births translate into school boundaries can be created. Given this information, modelling kindergarten enrollment only on past counts, a method typically employed in enrollment projections, does not adequately take advantage of all data available. Due to the availability of women of childbearing age in only 10 -year census intervals, using the number of women in an area and their fertility rate would need to be either interpolated or made static, neither of which provided a particular advantage over distributing actual births to a particular school and region based on population weights by ethnicity. In addition, because all of the births needed to project the school enrollment to 2007 have already occurred, there is no need to project births, only distribute them.
3. Distributing 6th Grade. The state space of the MCE projection model is defined over a fixed set of schools and grades within each school. In practice, the SBESD administration had adopted the strategy of assigning some 6th grades to junior high school in the early 1990s. Because this was not a permanent arrangement - it began in 1993 and ended
in 2002 - for the purpose of enrollment forecasting it is necessary to reassign 6th grade students back to the elementary school they would have attended. To do this, the students in the middle school were assigned to 6th grade in the school that they attended the year prior to changing to the middle school. If there was no previous school of record, the students were assigned to the school in their attendance zone.

As with any applied project, the required data and assumptions of idealized models are frequently at odds with the reality of administrative practices and imperfect record keeping. Still, the data derived in the above three steps accords well with the primary dynamics of the MCE model. The primary strength is in the transaction flows derived from the student level records and our primary interest is to investigate the impact of allowing for directional student flows under open enrollment. As such, we are less concerned with the inadequacy of the birth data since it will be a common input to either the MCE projection or the standard cohort survival projection.

### 4.3 MCE Projection Results

Given the recent instability in yearly school enrollment, the recent opening of two schools (Open Alternative and S.B. Academy), and the recent change in the open enrollment policy, a standard projection competition between competing methods is unlikely to demonstrate the superiority of one model over the others. Instead, since there is no precedent for the MCE projection model, the main purposes of this application is to contrast its results with those of the standard cohort survival model. The tables of the mean absolute percent error (MAPE) and mean percent error (MPE) reported throughout this section are defined in reference to a baseline cohort survival model. The "errors" are simply deviations from that baseline. ${ }^{10}$ We selected the baseline model to reflect the model that is most commonly used by school planners - a district-level cohort survival model with the aggregate enrollment results allocated to individual schools using a fixed share of enrollment per school derived from the most recent year available.

Although our basic MCE projection model is defined above, the "regions" that define the basis of the multiregional interactions are not necessarily individual schools. Schools are certainly the preferred level of aggregation given that our stated intent is to improve school level staffing and capital facilities planning. Yet, as a rule the more disaggregate the system the lower the signal-to-noise ratio. With top-down uniregional (district) projections there is less annual variation in enrollment thus providing a more robust basis for forward projection. While this can mean more accurate district projections, yearly fluctuations at the school level will be masked as will the intradistrict transfers of interest in this study. Conversely, in a bottom-up approach, there is a problem of small samples sizes at the school level, which are compounded when aggregating to the district level. If there is homogeneity in the interaction structure of a subset of schools with other subsets of schools, then a multiregional model can be defined that still captures open enrollment dynamics while stabilizing the estimates in the transition probability matrix. As such, we report results for the MCE model and the cohort-survival (CS) model using for each a pure bottom-up approach (schools are the regions), an intermediate approach using an a priori assignment of schools to more aggregate regions, and a third that takes the aggregate region results and allocates the projections to schools using the fixed share

[^6]of the 2002 school population in the region. The baseline model is the pure top-down approach; a uniregional district cohort-survival models with school enrollment derived from fixed shares.

Before turning to the projection results it is instructive to also consider some descriptive results in the context of the open enrollment policy aims. If the number of students that choose to go to non-local schools is small, then there is clearly no reason to pursue the MCE projection model. It is also of interest to ask whether the interactions that do occur mitigate the effects of ethnic and socio-economic residential segregation and improve ethnic and socioeconomic school balance. The results in Table 3 shed some light on that question. Both panels contain the percentage assigned by residence, and percentage actually attending, each of the elementary schools for (A) students whose first language is spanish (our ethnicity proxy) and (B) students qualifying for free lunch (our proxy for low income students). ${ }^{11}$ If the open enrollment system were working as intended - that is, to promote diversity - the schools with higher assignment percentages should have lower attendance percentages. In fact, the opposite is generally true in Panel A. The four schools with the highest percentages assigned (Cleveland, Franklin, Harding, and McKinley) have even higher percentages attending. The converse is true for Roosevelt, Washington, and Monroe. The pattern is similar for Panel B with the two schools with the lowest percentages assigned having even lower attending and the two with the highest percentages assigned having even higher percentages attending. Thus, interest in the interaction dynamics certainly seems warranted from the standpoint of policy effectiveness.

What about the absolute level of interaction? That is, are large numbers of students choosing to enroll in schools outside of their attendance areas? Table 4 confirms that the level of interaction is indeed very high. In the 2001/02 academic year, $19 \%$ to $37 \%$ attended schools outside of their assigned area; five of the eight schools had more than $30 \%$ attending outside of their assigned area. The structure of the interaction pattern for 2001/02 is more easily grasped in Figure 2. In the figure, the schools are located to promote visual clarity rather than geographic location, arrows are darker and wider for larger flows, and the numbers in the circles indicate the number attending each school from outside the SBESD district. Notice that two of the schools, Franklin and Cleveland, with the highest proportion of first-language spanish students actually exchange a very large number of students. Also, from Table 2 and Figure 1 note that the two schools are geographically proximate and that Franklin requires school uniforms and has other policies that are intended to reflect the culture of Mexican grade schools. Harding and Adams are also exchanging a large number of students. Overall then, the volume of student exchanges is very large and the pattern seems strongly structured. We undertake a more complete study of the description of student flows using event history analysis in another paper.

The MCE and CS model comparison results are contained in Tables 5-7. The three levels of aggregation are as defined above: $\mathrm{R}=$ regional, $\mathrm{RS}=$ regional downscaled to school using a fixed share, and $\mathrm{S}=$ school (no disaggregation). Under the baseline model and all of the other models we use a common assumption to derive the kindergarten cohort from the birth data. A more complete comparison of results under alternative constructions of the B-K ratios is available in (Middleton 2004). The pattern of deviations in Table 5 suggests that the MCE model does not produce substantially larger absolute deviations from the baseline model than the CS model, yet the deviations do vary over a larger range. The main outlier is the MAPE of 61.4 for the Open Alternative for the MCE school projection. The deviations increase, as expected, for the more disaggregate regions under both models. A more interesting result is that the MCE model

[^7]always produces higher enrollment estimates than the baseline. Again the Open Alternative school is the outlier with an almost $50 \%$ increase over the baseline model for the MCE school projection. It is important to note again that large "errors" are not pejorative in this context since they simply indicate the deviation from a very simple baseline model. Thus the MCE projections do indeed produce different results, but with the exception of the Open Alternative results, the projected values are not so different as to be dismissed. Table 6 provides deviations by grade and again provides a very plausible contrast for the MCE projections.

Table 7 is provided to identify, and isolate, the contribution from students attending from outside the district; the last term in (5). In the MCE projections we assumed a stationary transition matrix, $\mathbf{C}$, and a fixed in-migration vector. Given the structure of the MCE projection model we can examine how the district enrollment would proceed if the local schools were closed to students outside the district. Recall from the above discussion that school funding is closely tied to student enrollment. The results in Table 7 suggest that out-of-district students are vital to maintaining enrollment levels. Without in-migration, district enrollment almost universally decreases, relative to the baseline model, for each school. Only three schools benefit under the MCE school-level model without in-migration (SM-m).

## 5 Conclusions

Given the past enrollment history and local birth records, the results of both the cohort survival and multiregional cohort model indicate continued enrollment declines. Since the available data was restricted to only a few time periods, it is impossible to make a definitive statement about the relative improvement in forecast accuracy due to the multiregional specification. The focus on intradistrict mobility has revealed many interesting trends in the Santa Barbara schools. Whether the data availability and technical expertise of a school system allows for this kind of data to be utilized in forecasts, the value of understanding student flows has many applications. In this case, tracing student progression through grades and among schools reveals the highly variable spatial and temporal relationship between place of residence and school attendance that ultimately give rise to observed enrollment patterns. While a cohort survival method may suggest that 96 percent of the students present in the previous year will return the following year, the matrix model reveals that only 87 percent actually attended the school system in the previous year. Patterns revealed at the school-level potentially could be used to develop targeted policies. For example, in knowing that 24 percent of students in Adams did not attend the previous year (excluding kindergartners), it is possible either to develop programs that encourage those students to stay or to provide outreach to parents unfamiliar with the school's programs. In general, enrollment projections and school policies can benefit from careful analysis of intradistrict student flows.

Several issues complicate further interpretation of the projections. During the model calibration period there was both a major policy change and an entirely new school added to the choice set for parents. The policy change was the complete relaxation of ethnicity or socioeconomic constraints as part of the open enrollment policy. From 1994 to 1997, student transfers were conditional on available space and that the transfer would directly promote ethnic balance among the schools. After 1997 the condition of promoting ethnic balance was removed from the policy. As such, it is likely that intradistrict flows would increase after 1997 due to the more lenient policy. The new school is Santa Barbara Community Academy (SBCA). Its opening allowed for more students to move to a school that has no attendance boundaries associated with it. Not only has this school opened, attracting students away from their previous school of attendance, but a recent newspaper article reported that although enrollment is declining in SBESD, SBCA will be expanding due to its popularity. In the case of both of these school
policy changes, the timing of the events - just at the beginning of the study period - make it unclear whether the flows observed during the study period are a response to new opportunities that will soon stabilize or a lasting feature of enrollment change in Santa Barbara. As a rule, we suspect that model calibration often will be problematic for local districts due to the types of problems just listed and to the inherent instability of small population systems.

The quality of the input data for the MCE model is still a potentially large impediment to implementation. While school administrative records offer the most complete and timely data collection mechanism, the data will suffer from known sources of measurement error. Specifically, student entries into the system are recorded with more precision than student exits. Entries are recorded as a specific date even for mid-year transfers whereas exits are inferred when the student identification number is not present in a subsequent year. As such, measures of student duration in a school or the district cannot be measured exactly. There are also problems with jointly measuring grade progression and school attendance related to geocoding and the reliability of the address entries in the student records. Still, the only alternative is to use state records, which rely on data collected on a prearranged day each year. This only provides aggregate, not individual, data records, and may result in an undercount of students.

The results presented in this paper are clearly preliminary but are suggestive of the potential utility of the MCE projection model. The large positive deviations for the the Open Alternative school reflects the need for a constrained MCE model. Such a model is currently under development with constraints on both class size (since laws stipulate the maximum student-teacher ratio) and for each school (given a fixed number of classrooms). The extension to a constrained model is not trivial. There is also a need to check for origin dependence among the intradistrict flows and to isolate the intradistrict mobility of students entering from outside the district. Both of these extensions are possible but are beyond the scope of the current paper. Finally, for policy purposes the administrative records, while providing excellent details on intradistrict mobility, do not provide much of a basis to understand why the mobility patterns appear as they do. We are currently implementing a household survey to directly gather information related to the joint interaction between school choice and neighborhood choice.

In districts throughout the U.S., there is an acute need for reliable enrollment forecasts and for enrollment impact analysis of changes in capital facilities, in program offerings, or to other policies. To date, there is little research that describes enrollment patterns under complex policy schemes or how parents choose among a set of alternative educational environments for their children. Without knowledge of the attractiveness of a new school program, the average duration in which a parent will place a child in a school before changing to another school, or the features of a school that parents and children actively seek, school enrollment projections could be under- and over-estimating the number of students that attend individual schools even if the overall district projections are adequate. At a more general level, school enrollment dynamics and forecasting models do not appear to get the research attention they deserve given the obviously important role of the education sector in modern economies. From the systems perspective, it seems long overdue to treat enrollment dynamics as an open and interacting system. Indeed, our work here is very much in the spirit of Richard Stone's (1965, 1972) recommendations from more than 30 years ago. In our application we examine intradistrict mobility, but the arguments and modelling framework apply equally to multi-district systems or to incorporate the impacts of international migration (immigration and emigration) on a system. The U.S. and international education policy contexts and the local demographics process impacting schools are complex and continually evolving. Our work here is simply one of several alternative modelling strategies that attempts to go beyond the simple, and widely used, models currently employed in the education sector.

## References

California Department of Education: 2004, District reports, http://www.eddata.k12.ca.us/welcome.asp.

Campbell, H. S.: 1997, School enrollment in an extended demographic model, Journal of Economic and Social Measurement 23(1), 27-44.

EdSource: 2004, Basic Aid, http://www.edsource.org/edu_fin_basicaid.cfm.
Educational Demographics Office: 2004, State Profile, Education Data Partnership, http://www.ed-data.k12.ca.us/state.

Epple, D. and Romer, T.: 1991, Mobility and redistribution, Journal of Political Economy 99(4), 828-858.

Epple, D. and Sieg, H.: 1999, Estimating equilibrium models of local jurisdictions, Journal of Political Economy 107(4), 645-681.

Fabricant, R. and Weinman, J.: 1972, Forecasting first grade public school enrollment by neighborhood, Demography 9(4), 625-633.

Grip, R. S. and Young, J. W.: 1999, The modified regression technique: A new method for public school enrollment projections, Planning and Changing 30(3-4), 232-248.

Jaffe, A.: 1969, Handbook of statistical procedures for long-range projections of public school enrollment, Technical report, Office of Program Planning and Evaluation, US Department of Health, Education, and Welfare.

Kuznia, R.: 2004, 4 new schools to open in S.M., Santa Barbara News-Press.
Middleton, E. J.: 2004, Alternative enrollment forecasts for the santa barbara elementary school district, Master's thesis, Department of Geography, University of California at Santa Barbara.

Pullum, T. W., Graham, S. N. and Herting, J. R.: 1986, How to forecast public school enrollments, American Demographics 8(10), 52-54.

Rives, N. W.: 1977, Forecasting public school enrollment, Socio-Economic Planning Science 11, 313-318.

Roback, J.: 1982, Wages, rents, and the quality of life, Journal of Political Economy 90(6), 1257-1278.

Rogers, A.: 1975, Introduction to Multiregional Mathematical Demography, Wiley, New York.
Rogers, A.: 1985, Regional Population Projection Models, Vol. 4 of Scientific Geography, Sage, Beverly Hills, CA.

Rogers, A.: 1990, Requiem for the net migrant, Geographical Analysis 22(4), 283-300.
Rushton, G., Armstrong, M. P. and Lolonis, P.: 1995, Small area student enrollment projections based on a modifiable spatial filter, Socio-Economic Planning Sciences 29(3), 169-185.

Shaw, R. C.: 1984, Enrollment forecasting: What methods work best?, NAASP Bulletin p. 52?58.

Stone, R.: 1965, A model of the educational system, Minerva III(2), 172-186.
Stone, R.: 1971, Demographic Accounting and Model-Building, Organisation for Economic Coorperation and Development, Paris.

Stone, R.: 1972, A markovian education model and other examples linking social behaviour to the economy, Journal of the Royal Statistical Society A 135(4), 511-544.

Stone, R.: 1975, Towards a System of Social and Demographic Statistics, Organisation for Economic Coorperation and Development, Paris.

Sweeney, S. H. and Konty, K. J.: 2002, Population forecasting with nonstationary multiregional growth matrices, Geographical Analysis 34(4), 289-312.

Taylor, R. G., Vasu, M. L. and Causby, J. F.: 1999, Integrated planning for schol and community: The case of Johnston County, North Carolina, Interfaces 29(1), 67-89.

Tiebout, C.: 1956, A pure theory of local expenditures, Journal of Political Economy 64(5), 416-424.

Webster, W. J.: 1970, The cohort-survival ratio method in the projection of school attendance, The Journal of Experimental Education 39(1), 89-96.

## List of Tables

1 Alternative enrollment projection models ..... 18
2 Santa Barbara Elementary Schools ..... 19
3 Socio-cultural indicators by school, 2001-2002 ..... 20
4 Assignment-Attendance Mismatch, 2001-2002 ..... 21
5 Deviation from baseline by school ..... 22
6 Deviation from baseline by grade ..... 23
$7 \quad$ Multiregional with and without in-migration, by school ..... 24
List of Figures
1 Santa Barbara Elementary Schools ..... 25
2 Assignment-Attendance Mismatch ..... 26
Table 1: Alternative enrollment projection models

| Type of Model | Data Requirements | References | Type of School District Growth |
| :---: | :---: | :---: | :---: |
| Cohort Survival | Past enrollment data and birth records | Jaffe (1969), Fabricant and Weinman (1972), Rives (1977), Taylor et al. (1999), Webster (1970), Grip and Young (1999) | Best when change is constant; constant growth, constant decline, or stable. Also assumes constancy of transfers, migration, and entry and exit into and out of the school system. For growth and decline, long-term projections suffer if no maximum or minimum is imposed. |
| Ratio | Past enrollment data, population data and birth records | Webster (1970), Campbell (1997), Shaw (1984) | Stable ratio of population to number of children; feature of communities with constant enough migration to replace aging population with those of child-bearing age. |
| Housing Unit | Birth records, past and future housing unit data, with the number of students each housing type yields | Jaffe (1969), Campbell (1997) | Rapidly growing districts, especially when used with cohort survival to provide proposed enrollment caps. |
| Regression | Past enrollment data and birth records | Webster (1970), Rives (1977) | Growing or declining, but not constant, enrollment growth. |
| Gradient Cohort Survival | 6 years past enrollment data | McKnight, Taylor et al. (1999) | In growing or declining scenarios, optimal weights can be generated; in constant growth, same as cohort survival. |
| Modified Regression | Past enrollment data, birth records, migration data, student drop-outs, etc. depending on model | Grip and Young (1999), Fabricant and Weinman (1972) Shaw (1984) | In theory, any region experience growth or decline could benefit, though regression methods in enrollment projections have often failed to result in better projections than simple cohort models |
| Modifiable Spatial Filter | Address geo-coded enrollment data for three to four years | Rushton et al. (1995) | Areas experiencing growth, or areas that need to open or close schools to determine alternative enrollment scenarios. |
| Extended Demographic | Past enrollment data, birth records, and student dropouts, migration and grade retention records | Campbell (1997) | Informative in any enrollment scenario, but the migration requirement is difficult at the level of most school districts. |

Table 2: Santa Barbara Elementary Schools

| School | First Lang. Spanish | School Programs | Calif. <br> Dist. <br> School | Attendance Boundary |
| :---: | :---: | :---: | :---: | :---: |
| Adams | 0.58 | Gifted and Talented Education (GATE) | 2002 |  |
| Cleveland | 0.83 | Year round | 1989 |  |
| Franklin | 0.84 | School uniform |  |  |
| Harding | 0.73 |  |  | Students within attendance zone can walk to school. |
| McKinley | 0.83 | English as a second language program |  |  |
| Monroe | 0.45 | Multi-grade math and language arts, English emphasis program | $\begin{aligned} & \text { 1987, } \\ & \text { 1997, } 2002 \end{aligned}$ | Some students bused from downtown Santa Barbara |
| Open Alternative | 0.73 | Committee of parents provides school administration |  | No attendance boundary. Located outside SBESD boundary |
| Peabody | 0.41 | GATE, Charter School, Fine arts programs, English/Spanish language. | 1997, 2000 |  |
| Roosevelt | 0.19 | GATE |  |  |
| S.B. Academy | 0.10 | School uniforms, year round calendar, foreign language program. |  | No attendance boundary. Located in annex bused to Washington. |
| Washington | 0.28 | GATE, Science lab. | 2000 | Some students bused from downtown Santa Barbara |

Table 3: Socio-cultural indicators by school, 2001-2002

| A. First Language is Spanish |  |  |  |  |
| ---: | ---: | :---: | ---: | ---: |
| Region | School | Number | Percentage <br> attended |  |
|  |  |  | assigned |  |

B. Qualify for Free Lunch Program

| Region | School | Number | Percentage <br> attended |  |
| ---: | ---: | :---: | ---: | ---: |
|  |  |  | 48.9 | 49.1 |
| 1 | Adams | 347 | 0.0 | - |
| 2 | S.B. Academy | 0 | 44.2 | - |
| 2 | Open Alternative | 134 | 37.5 | 35.3 |
| 3 | Peabody | 279 | 30.1 | 26.7 |
| 3 | Roosevelt | 173 | 79.9 | 58.8 |
| 4 | Cleveland | 401 | 4.2 | 16.8 |
| 4 | Franklin | 35 | 66.7 | 52.0 |
| 4 | Harding | 448 | 9.6 | 24.0 |
| 4 | McKinley | 59 | 48.2 | 47.6 |
| 5 | Washington | 308 | 33.6 | 35.8 |


| чวұеust！u VV |  |  <br>  |
| :---: | :---: | :---: |
| ләчҰО <br> Кишррэәу ‘g＇s <br> －भt uәdO |  |  |
| uoł．8u！̣Se $M$ <br> әолиојN <br> Кәци！уフォ <br> su！̣рлен <br> и！̣yயext <br> риегәлә！О <br> Ұ［əләSOOY <br> Кроqеәд <br> surp |  |  <br> peusis！ss |

Table 5: Deviation from baseline by school

| School | Multiregional |  |  |  | Uniregional |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | :---: |
|  | R | RS | S | R | RS | S |  |
|  |  |  |  |  |  |  |  |
| Adams | 10.3 | 10.3 | 10.7 | 13.3 | 13.3 | 10.5 |  |
| Open Alternative | 22.2 | 20.6 | 61.4 | 24.3 | 22.2 | 28.7 |  |
| S.B. Academy |  | 24.4 | 15.1 |  | 27.4 | 18.3 |  |
| Peabody | 8.7 | 8.2 | 7.0 | 7.2 | 6.8 | 9.9 |  |
| Roosevelt |  | 9.6 | 14.0 |  | 8.2 | 9.6 |  |
| Cleveland | 4.2 | 11.3 | 6.1 | 5.8 | 4.7 | 9.2 |  |
| Franklin |  | 3.6 | 10.1 |  | 6.5 | 6.0 |  |
| Harding |  | 2.9 | 3.2 |  | 8.1 | 3.7 |  |
| McKinley |  | 3.8 | 10.3 |  | 6.2 | 8.0 |  |
| Monroe | 8.3 | 8.3 | 16.8 | 5.5 | 5.5 | 11.2 |  |
| Washington |  | 8.3 | 7.6 |  | 5.5 | 9.7 |  |
|  |  |  |  |  |  |  |  |
| Adams | 7.0 | 7.0 | 7.4 | -8.4 | -8.4 | -5.7 |  |
| Open Alternative | 17.0 | 14.7 | 49.1 | 22.9 | 20.5 | 17.8 |  |
| S.B. Academy |  | 20.2 | 14.7 |  | 26.3 | 17.4 |  |
| Peabody | 4.8 | 3.1 | -0.4 | 3.6 | 2.0 | -0.2 |  |
| Roosevelt |  | 7.0 | 12.1 |  | 5.8 | 6.2 |  |
| Cleveland | 3.9 | 11.3 | 3.4 | -4.6 | 2.2 | -6.6 |  |
| Franklin |  | 2.8 | 10.1 |  | -5.7 | 0.2 |  |
| Harding |  | 0.5 | 1.6 |  | -7.8 | -2.5 |  |
| McKinley |  | 3.3 | 8.5 |  | -5.2 | -7.8 |  |
| Monroe | 6.2 | 6.2 | 8.8 | -3.4 | -3.4 | 0.0 |  |
| Mashington |  | 6.3 | 4.4 |  | -3.4 | -5.2 |  |

Table 6: Deviation from baseline by grade

|  | Grade | Multiregional |  |  | Uniregional |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | R | RS | S | R | RS | S |
| $\stackrel{r 1}{2}$ | K | 8.9 | 8.7 | 8.6 | 8.6 | 8.1 | 8.3 |
|  | 1 | 10.3 | 9.4 | 18.4 | 12.2 | 11.2 | 8.8 |
|  | 2 | 8.5 | 8.5 | 16.3 | 11.8 | 10.8 | 9.7 |
|  | 3 | 11.6 | 10.6 | 15.3 | 11.2 | 10.7 | 10.7 |
|  | 4 | 11.8 | 10.7 | 13.9 | 10.9 | 10.4 | 12.6 |
|  | 5 | 11.9 | 10.9 | 13.0 | 10.5 | 9.9 | 12.7 |
|  | 6 | 12.0 | 11.9 | 17.8 | 13.3 | 11.7 | 16.6 |
| $\frac{1}{2}$ | K | 3.9 | 4.3 | 3.6 | 2.9 | 2.9 | 1.3 |
|  | 1 | 7.7 | 7.0 | 14.6 | 1.4 | 2.0 | 2.2 |
|  | 2 | 7.6 | 7.4 | 13.6 | 1.3 | 1.4 | 1.2 |
|  | 3 | 11.3 | 9.6 | 14.2 | 2.0 | 1.7 | 1.6 |
|  | 4 | 11.1 | 9.7 | 12.9 | 2.8 | 2.1 | 1.9 |
|  | 5 | 7.1 | 6.7 | 8.7 | 1.6 | 1.0 | -0.2 |
|  | 6 | 5.8 | 7.8 | 8.6 | 2.2 | 3.4 | 0.6 |

Table 7: Multiregional with and without in-migration, by school

|  | School | Multiregional |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | RS | RS-m | S | S-m |
|  | Adams | 10.3 | 19.8 | 10.7 | 36.6 |
|  | Open Alternative | 20.6 | 16.4 | 61.4 | 89.1 |
|  | S.B. Academy | 24.4 | 17.8 | 15.1 | 98.8 |
|  | Peabody | 8.2 | 15.3 | 7.0 | 30.5 |
|  | Roosevelt | 9.6 | 12.7 | 14.0 | 18.6 |
|  | Cleveland | 11.3 | 11.4 | 6.1 | 54.0 |
|  | Franklin | 3.6 | 14.6 | 10.1 | 31.4 |
|  | Harding | 2.9 | 15.6 | 3.2 | 70.4 |
|  | McKinley | 3.8 | 14.4 | 10.3 | 10.9 |
|  | Monroe | 8.3 | 12.9 | 16.8 | 19.2 |
|  | Washington | 8.3 | 12.9 | 7.6 | 7.4 |
| $\stackrel{I}{1}$ | Adams | 7.0 | -15.2 | 7.4 | -36.6 |
|  | Open Alternative | 14.7 | -0.3 | 49.1 | 89.1 |
|  | S.B. Academy | 20.2 | 4.5 | 14.7 | 96.4 |
|  | Peabody | 3.1 | -15.2 | -0.4 | -30.5 |
|  | Roosevelt | 7.0 | -12.1 | 12.1 | -16.6 |
|  | Cleveland | 11.3 | -5.6 | 3.4 | 54.0 |
|  | Franklin | 2.8 | -12.9 | 10.1 | -31.4 |
|  | Harding | 0.5 | -14.9 | 1.6 | -70.4 |
|  | McKinley | 3.3 | -12.5 | 8.5 | -4.8 |
|  | Monroe | 6.2 | -11.4 | 8.8 | 9.2 |
|  | Washington | 6.3 | -11.4 | 4.4 | -2.6 |



Figure 1: Santa Barbara Elementary Schools


Figure 2: Assignment-Attendance Mismatch


[^0]:    ${ }^{1}$ This research was supported by a grant from the Doctor Pearl Chase Fund for Local Community Development Research Projects, UCSB Academic Senate. The paper benefitted from the insightful comments from two anonymous reviewers and from participants in the Population Projections in the 21st Century session at the 2004 annual meeting of the Population Association of America.

[^1]:    ${ }^{1}$ This number is usually very small and ignorable in grades 10 and lower; indeed, we hope that it is zero. Accidents (especially auto-related), suicides, and homicides (in some districts) do register a significant increase later in high school. We include the category here but will exclude deaths throughout the rest of the paper since our focus is on elementary school
    ${ }^{2}$ Note that the term cohort in this setting refers to year of entry into school rather than birth year which is the common usage in demography.

[^2]:    ${ }^{3}$ Each school district is guaranteed a certain amount of money by the state, called the revenue limit. The state contributes the difference between the local property taxes and the revenue limit for each district. Those districts whose local taxes are larger than the revenue limit are called basic aid districts, and the state contributes $\$ 120$ per average daily attendance or $\$ 2400$ per district, whichever is higher. Though the schools that qualify for this program change from year to year, there are approximately 60 of these basic aid districts in any given year. (EdSource 2004).

[^3]:    ${ }^{4}$ Note that (5) has a transpose operator on the partitioned matrix, $C$. This allows for a more natural origin (row) and destination (column) orientation for the submatrices defined in (6) and (7).
    ${ }^{5}$ In a multiregional projection model the birth rates are included in Leslie growth matrix (Rogers 1985).
    ${ }^{6}$ There are a few approaches to constrained models in the literature. Stone (1965) proposes the use of an epidemic process such that transition rates follow a logistic path towards some limit. Most other approaches suggest the use of constraints in a mathematical programming context. In both cases, the forward projection of the system becomes secondary to the a priori specification of system targets.
    ${ }^{7}$ Another well developed body of literature focuses on national education and manpower planning models and applications in different countries. The seminal work by Richard Stone (1965, 1971, 1972, 1975) provides a complete introduction to those large scale planning models. Since our interest is in district level models we do not review the literature any further here.

[^4]:    ${ }^{8}$ Campbell (1997) is the one known exception. Pullum et al. (1986) construct sub-district planning units, which can be used for planning decisions such as school boundary changes or school closures. The percentage of students in a planning unit that do not attend their local school is a parameter in the model, but the authors do not attempt to understand the flow patterns nor do they incorporate directional flows into the model. Also, Shelly Lapkoff noted at a recent Population Association of America meeting that modelling intra-district transfers was considered in some recent work she did for the Berkeley school district.

[^5]:    ${ }^{9}$ We use first language as a proxy for ethnicity and minority status because self-reporting of ethnicity tends to be irregular.

[^6]:    ${ }^{10}$ We would have preferred to use a more traditional model calibration and validation approach. This would have entailed estimating the elements of the survivorship matrix on a subset of the data, and then using the model for out of sample prediction. In our application we simply did not have enough of a time series to follow that protocol. Indeed, based on our reading of the literature, it appears that this is almost never done for district level projections.

[^7]:    ${ }^{11}$ The percentage attended in Table 3 is calculated as the percent of students in the school having a particular characteristic; such as, that their first language is Spanish. The percentage assigned is the percent of students that would have had that characteristic if all students attended the schools based on their place of residence, e.g. no transfers.

