RESEARCH REPORT SERIES
(Statistics #2007-11)

Small Area Estimation of School District Child Population and Poverty: Studying Use of IRS Income Tax Data

Jerry J. Maples
William R. Bell*

Director's Office*

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

Report Issued: September 26, 2007

Summary

The Small Area Income and Poverty Estimates (SAIPE) program provides estimates for selected income and poverty statistics for states, counties, and school districts. The main objective of this program is to provide updated estimates of income and poverty statistics for the administration of federal programs and the allocation of federal funds to local jurisdictions. In particular, the No Child Left Behind Act of 2001 directs the Department of Education to distribute Title I basic and concentration grants directly to school districts on the basis of the most recent Census Bureau estimates of school-age children in poverty in each school district in the U.S.

While direct poverty estimates for school districts were produced from the 1990 and 2000 censuses, producing updated (post-censal) estimates at the school district level presents some challenging problems. Until recently there were no available data sources tabulated for school districts that provided reliable, updated information related to poverty. Therefore, the current official SAIPE method for producing updated school district child poverty estimates relies on a synthetic approach. The previous census results are used to estimate the proportions (shares) of the numbers of poor school-age children in each county who were in each school district that was wholly or partially contained in that county. (We refer to the intersections of counties and school districts as "school district pieces.") These shares are then carried forward through the decade and, to produce school district piece poverty estimates for a given year, they are multiplied by the SAIPE model-based county estimates of the numbers of poor school-age children in that year. The poverty estimates for the pieces are then appropriately summed to produce the poverty estimates for the school districts. This process also requires retabulating the previous census results each year to account for school district boundary changes. Apart from dealing with the boundary changes, this approach provides updated information on poverty only down to the county level – there is no updated information on the distribution of poverty within counties since the school district piece to county poverty shares remain constant.

Recent tabulations of IRS income tax data for school districts provide potentially useful sources of updated information on the distribution of poverty across school districts within counties. Analogous tabulations of IRS data for states and counties have been used successfully for years to provide regression variables for the SAIPE state and county poverty models. Tabulations of total exemptions (all exemptions on tax returns assigned to a given geographic area) provide a variable related to population, while tabulations of poor exemptions (exemptions on returns with adjusted gross income below the poverty threshold) provide a variable related to poverty. The idea explored here is to let the tabulations of poor exemptions for a given year define the school district piece to county poverty shares for that year. More formal statistical models relating the IRS data to poverty for school districts were explored by Maples and Bell (2004), but these models did not seem to have any advantages over the simpler approach considered here, and indeed sometimes these more formal models effectively reduced to the simpler approach.

Before the IRS data can be used in producing school district estimates, two issues need to be addressed: age relevancy and non-geocoded exemptions. The former refers to age ranges of the child population for any school district (which SAIPE assumes is 5-17 for unified districts, and which is a specific subset of this for separate elementary and secondary districts). We address the fact that income tax returns do not contain age information for the child exemptions by using demographic population estimates at the county level to provide approximate age distributions of the child exemptions for each school district. The second issue arises because significant numbers of tax returns cannot be assigned to school districts under current procedures, which require first geocoding the address of the tax return to a census block. (Virtually all addresses on tax returns can be assigned directly to a county, however.) To deal with this issue, two different methods are proposed for allocating the exemptions from each county's non-geocoded tax returns to its school district pieces: allocation proportional to population (PA), and allocation to minimize change (in the school district

piece to county shares) from the previous census (MC). In counties with few non-geocoded exemptions the allocation method matters little, but for counties with a significant number of non-geocoded exemptions the allocation method is important. For these counties the MC method of allocation will generally lead to either duplicating the previous census school district piece to county shares, or very closely approximating them.

To evaluate the predictive properties of the alternative approach to estimating school district poverty, we made such estimates for income year (IY) 1999 using the IY 1999 tabulations of income tax exemptions for school district pieces. Various alternative estimates were produced which differed according to the choice of method for allocating non-geocoded tax exemptions (PA versus MC), and from a choice among alternative updated school-age child population estimates for school district pieces. (The population estimates are obtained analogously to the updated poverty estimates, but using total age-adjusted child tax exemptions.) The alternative school district poverty estimates for IY 1999, along with those obtained from the official method (using the 1990 census as the "previous census"), were evaluated by comparing them to corresponding long form poverty estimates from Census 2000. To validate the results, this process was reversed in time and predictions made for IY 1989 using the IY 1989 tax data, and the estimates were compared to the 1990 census long form results.

For predicting Census 2000 results, an approach that estimated poverty shares based on pseudo poverty rates derived from the income tax data (a variant approach to using shares obtained directly from the tax data), and that used the PA method for non-geocoded exemptions, had the best predictions. Estimating shares of poor children using the MC allocation also performed better than estimating shares using the previous census data (the current official method). However, when we attempted to predict the 1990 census results, estimates using the PA method did not perform better than the current official method (reversed in time), whereas estimates using the MC allocation method did show some improvement over the official method. As the rate of non-geocoding of tax exemptions was higher in 1989 than in 1999, this suggests that the conservative nature of the MC method (reverting towards the previous census shares when non-geocoding rates are high) may be of benefit in counties with high non-geocoding rates.

The evaluations performed have certain limitations. One is that the 1990 and 2000 census estimates are not truth, both sets of estimates being subject to sampling and nonsampling errors. In particular, since many school districts are small the sampling variances for their long form poverty estimates can be substantial. While it may be desirable to allow for sampling variances in the census estimates, Bell and Kramer (1999) found, in an earlier evaluation of the current official method of school district poverty estimates, that doing so changed the numerical values of their evaluation measures but did not change the overall conclusions drawn from the results. Another limitation of the evaluations reported here is that, in evaluating the accuracy of the official method, its poverty shares from the previous census are ten years old. This is the worst case for the official method, while the methods using IRS data for the current year are unaffected by this 10-year horizon. Thus, the evaluation results presented here somewhat tend to favor the alternative methods. However, since we are now entering the second half of this decade (current production will provide estimates for IY 2004), we think the evaluation results do have relevance.

We also report here on similar evaluations of the alternative updated estimates of school district population (of relevant children) that were mentioned above. These results show significant improvements over the current official method when non-geocoded (child) tax exemptions are allocated by the MC method, but worse results than the official method when non-geocoded exemptions are allocated by the PA method. Further work in this area will be coordinated with staff of the Census Bureau's Population Division, who are also examining alternative approaches to estimating population for school districts.

1 Introduction

The U.S. Census Bureau, with support from other Federal agencies, created the Small Area Income and Poverty Estimates (SAIPE) program to provide more current estimates of selected income and poverty statistics than are available from the most recent decennial census. Estimates are created for states, counties, and school districts. The main objective of this program is to provide updated estimates of income and poverty statistics for the administration of federal programs and the allocation of federal funds to local jurisdictions. For example, the No Child Left Behind Act of 2001 directs the Department of Education to distribute Title I basic and concentration grants directly to school districts on the basis of the most recent Census Bureau estimates of school-age children in poverty in each school district in the U.S. For this application the SAIPE program is also responsible for creating child population estimates for each school district. In addition to the federal programs, there are hundreds of state and local programs that depend on income and poverty estimates for distributing funds and managing programs (U.S. Census Bureau, 2005a).

SAIPE estimates for states and counties make use of data from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). However, the CPS ASEC sample sizes for individual states and counties are not large enough to produce sufficiently reliable direct estimates for all states and counties for a single year of data. For this reason, CPS direct state estimates use a 3-year average. Hence, SAIPE uses small area models like that of Fay and Herriot (1979) to improve the estimates. The state and county models relate true poverty to other variables obtained from administrative records sources including IRS income tax data and food stamp program participation data, as well as corresponding estimates from the previous decennial census. Additionally, all estimates are raked to be consistent at aggregate levels. For example, the state estimates are raked to be consistent with the direct national poverty estimates given by the CPS ASEC, and the SAIPE county poverty estimates are raked to agree with the state estimates.

Estimation for school districts presents more severe data problems. First, CPS ASEC data is much too sparse to use at the school district level, the problem being more prevalent than that for counties due to the much larger number of school districts (14,334 school districts versus 3,141 counties in 1999-2000). The large number of school districts with very small CPS samples or no sample at all would make it difficult even to use CPS ASEC data in school district level models. School district estimates from the American Community Survey (ACS) are some years away until the ACS can accumulate up to five years of data to produce estimates for the small districts. The first set of five year ACS estimates are not scheduled to be produced until 2010. Data from the administrative records sources used in the SAIPE state and county models have previously not been tabulated down to the level of school districts. Administrative data that have previously been available for school districts (free and reduced price lunch data and school enrollment data) have presented data quality problems and other issues. (E.g., school enrollment data cover only children in public school and not children who are in private or parochial schools, or who are home-schooled.)

One geographical feature unique to the school district and county relationship, as compared to county and state or state and nation, is that school districts are not always neatly nested within counties. In some states school districts and counties are geographically identical. In other states there may be multiple school districts within a county and some school districts may cross county boundaries. In order to maintain consistency between school district and county level estimates for all of these scenarios, we split up school districts that cross over county boundaries into *school district pieces* corresponding to the parts of the district that overlap each county. We form estimates for the school district pieces and then aggregate these results across pieces within each school district. Breaking the school districts into these pieces facilitates controlling the school district estimates to agree with SAIPE county level estimates for the number of poor school-age children. For this research, we are using school districts as defined by the 1999-2000 school district boundaries. There are 14,334 school districts which are split up into 20,177 school district pieces. Of the 14,334 school districts, 896 districts are coterminous with a county, 9,296 districts are completely contained within a single county but do not compose the entire county, and the remaining 4,142 districts account for 9,985 pieces that cross over county boundaries.

Given these data limitations, SAIPE school district estimates for post-censal years have used a crude updating scheme which assumes that the ratio of poor school-age children in a district piece to the county number of poor school-age children remains constant over time, where this ratio is estimated from results of the previous census. These ratios (which we call the "school district piece to county poverty shares") are then carried forward for the post-censal years and multiplied by updated county estimates of the number of poor school-age children obtained from the SAIPE county model. School district piece estimates are then aggregated up to give estimates for school districts. (Section 4 gives a mathematical formulation of this procedure.) While this procedure uses updated information about poverty at the county level (via the county model), it does not account for any changes in poverty that differentially affect school districts within counties. (This is apart from school district boundary changes, which are accounted for by retabulating the previous census results each year using updated boundaries.) A similar update scheme is used to estimate the school-age child population in post-censal years for each school district piece by using the school district piece to county child population shares from the previous census.

Recently, however, IRS income tax data have been tabulated for school districts and school district pieces. This presents possibilities for using information from these data to provide current information about population and poverty within counties in formulating population and poverty estimates at the school district level. Maples and Bell (2004) investigated this possibility by fitting models relating census school district poverty estimates to school district tabulations of income tax data. The current report discusses an extension of that previous work to an evaluation of some school district population and poverty estimators that make use of the school district piece tabulations of the IRS income tax data. Various school district population estimates (for 2000) and poverty estimates (for income year 1999) are compared against corresponding estimates from the 2000 census to assess the accuracy of the various estimates in comparison to that of the official SAIPE school district estimates. The various estimation procedures are then reversed in time to "predict" school district population (in 1990) and poverty (in income year 1989), and these predictions are compared to corresponding estimates from the 1990 census to provide a second evaluation of the alternative estimators. This additional evaluation against 1990 census results provides a check on whether the methods are robust over time as opposed to just fitting well to the idiosyncrasies of one particular year. The main question to be answered by these evaluations is, "To what extent, if any, do the alternative estimators that make use of the IRS data improve the school district population and poverty estimates?"

Notice that the population estimates we consider (including the census count) refer to the years 2000 and 1990, while the poverty estimates and the tabulations of the IRS data refer to the income years 1999 and 1989, which are one year earlier. This is because income reported in the census or on IRS tax returns refers to the previous calendar year, which we call the income year (IY), but it is reported on by people resident (in given places) in the year when the income is reported (here 2000 or 1990).

Section 2 of this paper discusses the IRS data used in the models and how we handle the issues that arise in tabulating these data for school district pieces. Then, Section 3 discusses the alternative methods that use IRS data in constructing school district child population estimates, evaluates these and the official estimates against the corresponding 2000 and 1990 census population counts, and compares the performance of the alternative methods to that of the current official SAIPE method. Section 4 similarly evaluates several different school district child poverty estimates. Some of the alternative poverty estimates in Section 4 use one of the new population estimates from Section 3. Finally, Section 5 summarizes the results and conclusions.

2 Dealing with Difficulties in Tabulating IRS Income Tax Data for School District Pieces

Recently, IRS income tax data have been tabulated at the level of school district pieces to investigate the possible benefits of using such data for constructing population and poverty estimates for school districts.

The fundamental pieces of information that we use from each tax return are the exemptions (total and child) and the adjusted gross income (AGI). The assumption underlying our approaches to using IRS data in estimating population for school district pieces is that tabulations of total and child exemptions for school district pieces are related to total and school-age (5-17) population, respectively. If a return reports an AGI below the official poverty threshold for a family of the size implied by the total number of exemptions on the return, then all of the exemptions on that return are considered to be "in poverty," and we label these as "poor exemptions." The assumption underlying our approaches to using IRS data in estimating child poverty for school district pieces is that tabulations of poor child exemptions for school district pieces are related to the numbers of children in poverty.

To tabulate exemptions (total or child, poor or all) at the level of school district pieces we must first assign each income tax return to an individual school district piece. This process runs into two problems. The first is that not all tax returns can be geographically assigned to a specific school district piece. The second is that some school districts are defined by grade ranges (e.g., for separate elementary and secondary school districts), but IRS income tax returns do not provide specific ages, let alone grades, for child exemptions.

The geographic assignment of tax returns to school district pieces first goes through an intermediate step where we attempt to *geocode* the address of the tax return to a census block. Then we tabulate results over all census blocks within each given school district piece. Not all tax returns can be successfully geocoded to a census block, so we are left with a pool of non-geocoded exemptions. Methods to handle these non-geocoded exemptions are discussed later in this section. Future research may examine whether many of the exemptions currently not geocoded to census blocks could instead be assigned directly to school district pieces to reduce the number of non-geocoded exemptions that we need to deal with.

The national non-geocoding rates, actual and effective, from the current geocoding method are given in Table 1, and the median county level non-geocoding rates are given in Table 2. The "effective" non-geocoding rates (which are lower than the "actual" rates) account for the fact that we know with 100% certainty in

Table 1 - Overall non-geocoding rates

	Exemption Class	Income Year 89	Income Year 99
Actual	Poor Children	28.20%	15.65%
	All Children	20.68%	11.23%
Effective	Poor Children	19.96%	11.12%
	All Children	15.02%	8.26%

which school district piece to put an exemption that does not geocode to a block if it falls in a county that only contains one school district piece, i.e., if the school district piece is coterminous with the county. Table 1 shows that, nationally, in 1999 about 11 percent of all child exemptions and about 16 percent of poor child exemptions could not be assigned to a block (the actual non-geocoding rates), but allowing for assignment of exemptions to counties coterminous with school districts, the rates of non-assignment to school district pieces (effective non-geocoding rates) were about 8 percent and 11 percent, respectively. Much higher percentages of child and poor child exemptions could not be assigned to school district pieces in 1989. For Table 2 we took the non-geocoding rate for each county, for child and poor child exemptions, and computed the medians of these rates across all U.S. counties, for each of the two years shown. These median rates are much higher than the overall non-geocoding rates shown in Table 1, reflecting the fact that many rural counties have high non-geocoding rates and small populations. The overall non-geocoding rates in Table 1 can be thought of as weighted averages of the county non-geocoding rates with weights proportional to the number of exemptions, both of which are strongly related to population size.

Non-geocoding rates for school districts can also be computed. The school district's non-geocoding rate is the weighted average over the school district pieces of the respective county level non-geocoding rate weighted by the census' relevant 5-17 population count for that school district piece. This procedure assumes that the non-geocoding rate is constant within a county because there is no information about non-geocoding rates at

Table 2 - Medians of county non-geocoding rates

	Exemption Class	Income Year 89	Income Year 99
Actual	Poor Children	55.09%	33.84%
	All Children	52.01%	31.07%
Effective	Poor Children	36.62%	22.27%
	All Children	34.26%	20.57%

Table 3 - Non-geocoding rates by size of school district

Size	1st Quartile	Median	3rd Quartile
1-199	22.3%	39.6%	51.5%
200-499	18.1%	34.3%	47.1%
500-999	13.1%	27.7%	38.3%
1000-1999	8.3%	20.6%	28.1%
2000-4999	4.4%	14.8%	20.7%
5000+	3.5%	9.2%	11.6%

the subcounty level. Table 3 provides data on the relationship between the Census 2000 5-17 year old child population counts and the non-geocoding rates. It shows the median, along with the 25% and 75% quartiles, of the distribution of school district non-geocoding rates for the various population size groups. We see that larger school districts tend to have smaller effective non-geocoding rates. There is, however, much overlap in the distribution of rates between the size groups – not all small school districts have high non-geocoding rates, and not all of the districts from the largest size group have very low non-geocoding rates.

The second problem mentioned above arises when some school districts have overlapping geographical boundaries, as when an area is serviced by separate elementary and secondary school districts. The process for geocoding tax exemptions to school districts is only able to determine that the exemption is in a location serviced by a school district. Due to the lack of age or information about school grade for child exemptions on tax returns, it is unclear in this situation to which of the overlapping school districts the exemption should be assigned. The next section presents a method to address this issue.

2.1 Adjusting Income Tax Child Exemptions for School District Grade Ranges

In some areas of 17 states there are separate elementary and secondary school districts, each exclusively responsible for providing education in a subset of the grades in their shared territory. We obtain the grade range for each school district from the NCES (National Center for Education Statistics) Common Core of Data. The most typical grade ranges are unified (K-12), elementary (K-8), and secondary (9-12). In these 17 states, the census school district estimates of poor and total children (referred to as "relevant" children) reflect the grade ranges of the school districts. Therefore, to use tax data in a model at the school district level, we need to modify the tabulated numbers of child tax exemptions and poor child tax exemptions to reflect the grade ranges of the school districts.

One way to modify tabulations of child tax exemptions (and poor child tax exemptions) to reflect the grade range of a school district would be to assume a uniform distribution of children over age. Consider a unified school district, i.e., one that includes grades K through 12, which we assume corresponds to children ages 5-17. Although no age is recorded for child exemptions on tax returns, we will assume that child exemptions refer to the population of children ages 17 years old and under. Thus, for a unified district we could multiply the tabulated number of child tax exemptions by 13/18 to convert this to a figure for school-age child tax exemptions. We could similarly convert tabulated values of child tax exemptions to reflect the restricted grade ranges of non-unified districts, e.g., for an elementary (K-8) district we could multiply by 9/18.

However, demographic estimates of the number of children by single years of age are available each year at the county level, and this information can be used to refine these adjustment factors to account for differences in age distributions between counties. Most counties (2,874 out of 3,141) only contain school districts that are unified and for districts in these counties the adjustments for age corresponding to grade range ultimately will cancel out (of the shares to be constructed later), since all districts within the county will be adjusted by the same proportional amount. Only for the remaining 267 counties with school districts that do not contain the full (K though 12) grade range do the age-grade adjustments differentially affect school district pieces within the counties. To construct the age-grade range adjustment we multiply the number of geocoded child tax exemptions in school district piece j within county i, denoted $T_{g,ij}$, by the proportion from the demographic population estimates of the 0-17 year-old children in the county who are in that school district age-grade range. We denote the adjustment factor by π_{ij}, so we have an age adjusted geocoded child exemption count, $T^*_{g,ij}$, of

$$T^*_{g,ij} = \pi_{ij} \times T_{g,ij}.$$

for school district piece j in county i. Note that the data (population estimates) determining π_{ij} are county level, and so depend on j only through the grade range of school district piece j.

2.2 Allocating Non-Geocoded Tax Exemptions to School District Pieces

The second issue that we need to address concerns the pool of non-geocoded child tax exemptions in each county, the number of which we denote by $T_{ng,i}$. The non-geocoded exemptions are first adjusted as just discussed to reflect the target population of age 5 to 17 year old children, i.e., we form

$$T^*_{ng,i} = \pi^c_i \times T_{ng,i},$$

where π^c_i is the county i ratio of 5-17 year-old children to 0-17 year-old children from the demographic population estimates. This assumes that the age distribution of the non-geocoded exemptions is the same as the age distribution of the geocoded exemptions. We then must make an additional assumption on how to distribute the non-geocoded exemptions among the school district pieces within a county.

There are many ways one could attempt to allocate the non-geocoded exemptions. One obvious choice is to allocate them proportional to some estimates of the school district piece school age populations. However, in order to not create circular estimates for the number of school-age children, we cannot use the current school district piece population estimates for this purpose. Conceptually, we could use the previous year population estimates, but when the school district boundaries are updated (which is currently done every two years), this would require reconstructing the school district piece population estimates for all years from the most recent census up to the year prior to that for which we are trying to form estimates. For methods that require tabulating the IRS income tax data, this may prove too expensive of an endeavor. A more feasible alternative is to allocate proportional to the previous census school district piece population counts, and this is one of the methods that we consider here. (Note that since we will evaluate school district estimates of child population and poverty by comparing them to estimates from Census 2000, for this purpose we should not use any information from Census 2000 to allocate the non-geocoded exemptions. We should only use information from the 1990 census, which here is the previous census.)

Other approaches can be used to allocate the non-geocoded exemptions to the school district pieces, and no one method is clearly optimal. While modeling data from Census 2000 may provide some insight into this, we do not have data of the breadth and scope of the decennial census every year to apply. Also, we have no reasonable way of using the census counts to disentangle differences in the observed geocodable exemptions that are due to differences across school districts in geocoding rates versus differences across school districts in tax nonfiling rates. In the remainder of this section, we will present two different methods for distributing the non-geocoded exemptions: Proportional Allocation (alluded to above) and Minimum Change.

For the Proportional Allocation method let p_{ij} denote the estimated proportion of the county i school-age population that is the appropiate age and resides in school district piece j as given by the previous census:

$$p_{ij} = \frac{[\text{Previous census relevant child population}]_{ij}}{[\text{Previous census child population 5-17}]_i}.$$

We assign $p_{ij} \times T^*_{ng,i}$ of the non-geocoded exemptions to school district piece j. Combining the grade range adjustment and this non-geocoded exemption adjustment together, we get the final adjusted number of poor child tax exemptions or adjusted number of all child tax exemptions for the school district piece, denoted T_{ij}, and given by

$$\begin{aligned} T_{ij} &= T^*_{g,ij} + (p_{ij} \times T^*_{ng,i}) \\ &= (\pi_{ij} \times T_{g,ij}) + (p_{ij} \times \pi^c_i \times T_{ng,i}) \end{aligned}$$

The Proportional Allocation method is a synthetic approach to distributing the non-geocoded exemptions that can be derived from underlying assumptions that (i) each child in the county has the same probability of being represented on a tax return (i.e., of being a child exemption), and (ii) each child exemption in the county has the same probability of not being geocoded.

The second method of allocating non-geocoded tax exemptions, the Minimum Change method, is motivated by a concern that neither population estimates nor numbers of geocoded exemptions may be closely related to the numbers of non-geocoded exemptions that truly belong to the various school district pieces. If this is true, then if the number of non-geocoded exemptions in a county is large, we may be better off falling back on the previous census population or poverty shares, unless the geocoded tax exemptions for the county provide definite evidence in contradiction to the previous census results. This can be achieved by allocating the non-geocoded exemptions to make the resulting within county shares of total or poor exemptions as close as possible to the corresponding shares from the most recent census (where "closeness" could be defined by some underlying loss function). If the number of non-geocoded exemptions is large then it is likely that they can be allocated in a way that the shares of the tax exemptions will exactly reproduce the corresponding census shares. With a small number of non-geocoded exemptions this may not be possible, and then the tax exemptions provide definite evidence to contradict the previous census results. We now explain this approach in more detail.

For the Minimum Change method, we first create shares of total (geocoded plus non-geocoded) exemptions for each school district piece from the age adjusted geocoded exemptions for each piece and treat the age adjusted non-geocoded exemptions for the county as an additional category. These are defined by

$$T^*_{sij} = \frac{T^*_{g,ij}}{\sum_j T^*_{g,ij} + T^*_{ng,i}}$$

where j ranges from 1 to the number of pieces. We then also define the non-geocoded exemption share as $T^*_{s,ng,i} = T^*_{ng,i}/(\sum_j T^*_{g,ij} + T^*_{ng,i})$. Denote the census share of child population or poverty in school district piece j by C_{sij}.

The general idea for the Minimum Change method as we have implemented it is to keep the relative distribution between as many school district pieces as possible, within a county, the same as from the previous census. To do this, we first fix the shares of the pieces whose geocoded tax share is higher than the corresponding census share, i.e. $(T^*_{sij} > C_{sij})$. To these pieces we do not allocate any of the non-geocoded exemptions. Next, the non-geocoded exemptions are allocated to the remaining pieces so that the relative distribution among the tax shares for those pieces is the same as the relative distribution among the corresponding census shares. If one of the pieces cannot maintain this relative distribution, because its geocoded share is too large, then we fix that share, allocate none of the non-geocoded exemptions to that piece, and attempt to reconstruct the relative distribution of the previous census shares of the remaining pieces by allocating non-geocoded exemptions only to these pieces. We continue this process until all non-geocoded exemptions have been allocated.

In Table 4, we present three examples for a fictitious county to illustrate how the Minimum Change (MC) method works, and how it compares to the Proportional Allocation (PA) method. This county contains 3 school district pieces, for which the previous census shares of school-age child population were 60%, 30% and 10%. For simplicity we assume there are 100 tax exemptions for the county, with varying numbers of non-geocoded exemptions in the three examples as shown. (Having 100 exemptions implies that the numbers of exemptions in the pieces are numerically the same as their shares expressed as percentages.) In Example 1, all of the geocoded exemption shares are less than their census share counterparts, so the MC method can allocate the 29 non-geocoded exemptions to the three pieces to perfectly reconstruct the previous census

Table 4 - Examples of the Proportional Allocation and Minimum Change methods

		Example 1			Example 2			Example 3		
Piece	Cen Share	T^*_{sij}	PA	MC	T^*_{sij}	PA	MC	T^*_{sij}	PA	MC
1	60	47	64.4	60	68	74.6	68	68	72.2	68
2	30	23	31.7	30	14	17.3	24	15	17.1	23
3	10	1	3.9	10	7	8.1	8	9	9.7	9
Non-geocoded		29			11			8		

shares. In Example 2, the geocoded share of the first piece is larger than the census share, so we will only allocate the 11 non-geocoded exemptions to pieces 2 and 3. To keep the same relative distribution (3:1) in these pieces as in the previous census we allocate 10 non-geocoded exemptions to piece 2 and 1 to piece 3, resulting in 24 total tax exemptions for piece 2 and 8 for piece 3. Since neither of these shares exceed the corresponding census shares, we are done. Example 3 is similar to Example 2 in that the first piece's geocoded share is higher than its census counterpart. However, when attempting to allocate the 8 non-geocoded exemptions to maintain the 3:1 relative distribution between pieces 2 and 3 we hit the constraint that the geocoded share of the third piece is 9% which is larger than the 8% needed to maintain the 3:1 distribution. Thus, we allocate no non-geocoded exemptions to the third piece, holding it at its 9 geocoded exemptions, and restart this algorithm on the remaining pieces. With only one piece left we simply allocate all the non-geocoded exemptions to the second piece, keeping the exemptions in the first and third pieces fixed.

The idea behind the Minimum Change method could be implemented more formally by using an explicit loss function approach. This would allocate the $T^*_{ng,i}$ among the school district pieces to minimize some loss function,

$$Loss(\mathbf{T}^*_{si}, \mathbf{C}_{si}) = \sum_j f(T^*_{sij}, C_{sij})$$

where $f(x, y)$ is the individual loss between the scalars x and y. Some common choices of loss functions are the quadratic loss, $f(x, y) = (x - y)^2$, and squared log loss, $f(x, y) = (\log x - \log y)^2$. Whether we use the MC method or make a specific choice of loss function is probably not a big issue for our application, however. The main concern involves doing something for the counties with large numbers of non-geocoded exemptions to bring their shares closer to the census shares so as to reduce the need for making assumptions about how to allocate the non-geocoded exemptions.

For the MC method or any choice of loss function, if all of the geocoded shares are less than their corresponding census shares, then we can allocate the age-adjusted non-geocoded exemptions to perfectly match the census shares. In this case the tax exemption data alone do not definitively contradict the previous census shares, and so changing from the previous census shares can be seen as requiring some additional assumptions. Table 5 shows the number of counties, and the number of affected school districts, for which we were able to perfectly recreate the 1990 and 2000 census shares from the 1989 and 1999 tax exemptions. (Recall that the census income and poverty estimates also refer to 1989 and 1999.) In cases where at least one of the school district pieces has a geocoded tax share larger than its census counterpart, then we have evidence that there has been a change in the distribution of population or poverty within the county since the last census. We see from the table that, due to the non-geocoded exemptions, for less than half of the counties does the IRS exemption data give us definitive evidence that there has been a change in the underlying distribution of shares of poor children or of all children from the previous census. However, these counties where the income tax data is informative contain a greater number of school districts than the counties where the income tax data is non-informative.

By using either the Proportional Allocation or the Minimum Change method for the non-geocoded exemptions, we create an income tax exemption variable that is practical to use for population and poverty estimation. Note that we do not round our resulting T_{ij}'s to be integers. To distinguish between total child tax exemptions and poor child tax exemptions we will denote them by $T^{\text{tot}}_{(pa)\text{IY},ij}$ and $T^{\text{poor}}_{(pa)\text{IY},ij}$, respectively,

Table 5 - Number of counties for which the Minimum Change method shares equal the census shares and the corresponding number of affected school districts

		Counties		School Districts	
Year	Exemption Class	Equal	Different	Equal	Different
1989	Poor Children	2256	885	5176	9158
	All Children	2545	596	7087	7247
1999	Poor Children	1700	1441	2735	11599
	All Children	2059	1082	4319	10015

note: 935 of the 3141 counties only had 1 piece for the 1999-2000 school district boundaries

for proportional allocation of non-geocoded exemptions, and similarly $T^{\text{tot}}_{(mc)\text{IY},ij}$ and $T^{\text{poor}}_{(mc)\text{IY},ij}$, respectively, for the minimum change allocation. In this notation IY is the income year, 89 or 99, which is one less than the census year.

3 Models for School District Population Estimates

In this section we examine school district population estimators that use the adjusted tabulations of the IRS total child exemptions for school district pieces to provide updated information related to population for post-censal years. Empirical evaluations using school district population counts from the 1990 and 2000 censuses are performed to determine the extent to which, if any, these estimators improve on the official method for updating the school district population estimates. For this section, all references to child population are specifically for relevant children between the ages of 5 and 17.

The estimators are constructed using data for school district pieces. We have a collection of school district pieces ($j = 1, ..., J_i$) in each county i ($i = 1, ..., I$). Preliminary work by Maples and Bell (2004) using the 2000 Census and 1999 IY IRS income tax data has shown that modeling the shares of poor children in school district pieces (relative to the whole county) works better than modeling the number of poor children in school district pieces directly. We will build on this share methodology and apply it to child population estimates.

The current method used to estimate the number of children in school district pieces for a post-censal year uses a synthetic approach. The previous census data are used to estimate school district piece to county shares of population, and these estimated shares are then carried forward and multiplied by the current year county intercensal population estimates for children aged 5-17. For example, in estimating child population in school district pieces for the year 2000 (without using the 2000 census results) we have:

$$\text{Child}\,\widehat{\text{pop}}_{2K,ij} = \frac{C^{\text{tot}}_{90,ij}}{\sum_j C^{\text{tot}}_{90,ij}} \times \text{CNTY-POP}_{2K,i} \tag{1}$$

where $C^{\text{tot}}_{90,ij}$ is the 1990 census count of related 5-17 children in families in school district piece j of county i, and $\text{CNTY-POP}_{2K,i}$ is the intercensal child population estimate (for ages 5-17) for year 2000 in county i. The underlying assumption with this approach is that the distribution of the school-age child population among school district pieces within a county does not change over time. We want to explore estimation methods that use the current-year IRS data to reflect changes over time in the distribution of school-age child population within the county.

3.1 Using IRS Data in Share Models for Population Estimates

Since all estimates of child population for school district pieces will be controlled to the official county population estimates of children aged 5-17 to maintain consistency, it is sufficient to estimate the school district piece to county population shares. Share models attempt to describe the distribution of 5-17 year old children among the school district pieces within a county. Note that within a county, the estimated shares must add up to 100 percent. By construction, these share models will automatically be consistent with the county estimates of child population.

We will consider four competing models (the current official method and three experimental models that incorporate IRS income tax data) for estimating the current year school district piece to county population shares of 5-17 related children. The experimental models are described below in terms of producing population estimates for the year 2000 (without making use of the Census 2000 results).

POP-CEN uses population shares from the previous census (official method) – see Eq (1)

POP-TAXPA uses current year shares of the age adjusted child tax exemptions with non-geocoded exemptions allocated proportional to the previous census counts:

$$\widehat{\text{POP-TAXPA}}_{2K,ij} = \frac{T^{\text{tot}}_{(pa)99,ij}}{\sum_j T^{\text{tot}}_{(pa)99,ij}} \times \text{CNTY-POP}_{2K,i} \tag{2}$$

POP-TAXHYB is a hybrid of POP-CEN and POP-TAXPA that averages the shares from the previous census (POP-CEN) and current year shares from the POP-TAXPA method:

$$\widehat{\text{POP-TAXHYB}}_{2K,ij} = \frac{1}{2}\left(\frac{C^{\text{tot}}_{90,ij}}{\sum_j C^{\text{tot}}_{90,ij}} + \frac{T^{\text{tot}}_{(pa)99,ij}}{\sum_j T^{\text{tot}}_{(pa)99,ij}}\right) \text{CNTY-POP}_{2K,i} \tag{3}$$

POP-TAXMC uses current year shares of the adjusted child tax exemptions with non-geocoded exemptions allocated based on the minimum change method:

$$\widehat{\text{POP-TAXMC}}_{2K,ij} = \frac{T^{\text{tot}}_{(mc)99,ij}}{\sum_j T^{\text{tot}}_{(mc)99,ij}} \times \text{CNTY-POP}_{2K,i} \tag{4}$$

The hybrid estimator (POP-TAXHYB) and the minimum change income tax share estimator (POP-TAXMC) both combine information from income tax and census data in different ways. The final step to create population estimates for school district sd is to sum over corresponding school district pieces. For example, the school district population estimate using the POP-CEN method would be

$$\widehat{\text{POP-CEN}}_{sd} = \sum_{(i,j)\in sd} \widehat{\text{POP-CEN}}_{ij}.$$

All four of these estimators will be evaluated against census results in the next section.

3.2 Evaluation of Population Estimators

Our goal is to develop a school district population estimator that uses the IRS income tax data and performs better than the current official method which is based on the previous census shares. We will evaluate the accuracy of the four estimators presented in the previous section by using them to construct population estimates for 2000 and comparing these estimates to the Census 2000 school district population counts. For the official method the 1990 Census provides the previous census share estimates. Note that this is the worst case scenario for the official method because it uses census data that is 10 years old. The question is basically whether the more timely information in the IRS income tax data, even with the need to allocate non-geocoded tax exemptions, are more predictive about current population shares than the previous census data.

In order to increase confidence that our results will generalize to years other than 2000, we ran a validation study by reversing the official estimator in time, taking Census 2000 as the "previous census," and thus estimating the number of school age children in school districts in 1990. We also constructed school district child population estimates for 1990 by the methods that use the IY 1989 IRS data. We then compared all these 1990 school district population estimates against the corresponding 1990 Census child population counts for school districts. This gives us a second time point to compare the performance of the alternative school district child population estimators.

For the purposes of evaluation we replace the updated county population estimates in the year of evaluation by the corresponding census counts. Since, for these evaluations, the census counts are standing in for the true population, this means that all the school district population estimators benefit from using the "true" county population figures. While this understates the error in the school district population estimates, it means that the evaluations are really assessing the accuracy of the estimates of the school district piece to county population shares, at least in regard to the effects of the share estimates on the accuracy of the school district population estimates. This is what is really of interest since all the estimators would, in practice, make use of the same county population estimates.

Although the four estimation methods all make estimates for school district pieces, our concern is with estimates for whole school districts. Thus, our unit of analysis for evaluation will be whole school districts. To compare estimators we examine the mean squared difference over school districts of the log number of 5-17 children estimated by our models compared to the corresponding Census 2000 count:

$$\text{MSDiff} = 1/N \sum_{sd} [\log(C^{\text{tot}}_{2K,sd} + 1) - \log(ESTPOP_{sd} + 1)]^2 \tag{5}$$

where $ESTPOP_{sd}$ is the relevant population estimate for school district sd from any of the four estimators listed in the previous section. We use $\log(x + 1)$ to deal with the occasional zeros in the data. This slightly distorts the comparisons for the smallest of school districts but has virtually no effect on the mid-sized and larger districts. In addition to looking at the overall MSDiff for the estimators, we also examine MSDiffs for categories of school districts defined by the sizes of their child populations (as given by the Census 2000 population counts) as shown in Table 6. We will also examine MSDiffs for categories of school districts defined by their non-geocoding rates as shown in Table 7. In addition, for the three estimators using information from the IRS income tax data, we calculate their relative accuracy against that of the official estimator by, for example,

$$\text{RelMSDiff(POP-TAXPA)} = \frac{\text{MSDiff(POP-TAXPA)}}{\text{MSDiff(POP-CEN)}}. \tag{6}$$

Examining the MSDiffs within the population size or non-geocoding rate categories provides an indication of how the error in the estimates (taking the census counts as truth) depends on size or non-geocoding rate. To study this in more detail we could plot the absolute errors corresponding to the individual terms in eq. (5), that is, plot $|\log(C^{\text{tot}}_{2K,sd} + 1) - \log(ESTPOP_{sd} + 1)|$ against, say, the non-geocoding rate. But doing this with the results for all the 14,000 plus school districts would yield a very noisy scatter plot. Instead, we smooth this data using a Loess local regression line (Cleveland 1979). This allows us to see how the relative accuracy of the alternative estimators generally depends on the non-geocoding rate with more detail than is permitted by the aggregate MSDiffs. One qualification of these plots is that the results at the very

Table 6 - School District Size Categories

Child Pop Size	frequency	Child Pop Size	frequency
1-199	2132	200-499	2292
500-999	2275	1000-1999	2602
2000-4999	2905	5000+	2102

Table 7 - Non-geocoding Rate Categories

IY 1999 Non-geocoding rate	frequency	IY 1989 Non-geocoding rate	frequency
0-.08	2754	0-.125	2541
.08-.15	2618	.125-.30	2791
.15-.225	2085	.30-.50	2720
.225-40	3093	.50-.70	3006
>.40	2858	>.70	2344

extreme ends of the range of the variable used on the x-axis (e.g., the non-geocoding rate) are affected by how the Loess smoothing deals with end effects. We did not include the school districts with zero effective non-geocoding rates in the Loess smoothing, because many of these districts are coterminous with their county and will have zero prediction error after control to county totals. If we had included these school districts, it would have distorted the left side of the plots for Figures 1, 2, 5 and 6.

The results of our evaluations of the school district child population estimates for 2000 are given in Tables 8-9 and Figure 1. The results in Table 8 show that the proportional allocation share estimator, POP-TAXPA, performed worse overall than the official method (using previous census shares), and improved on the results from the official method only for the category with the largest school districts. The hybrid estimator, POP-TAXHYB, performed much better, improving on the official method overall and for all but two of the smaller school district categories. The minimum change share estimator, POP-TAXMC,

Table 8 - Mean squared differences between Census 2000 and alternative estimates of child population for school districts by categories of school district size

Size	POP-CEN	POP-TAXPA	POP-TAXHYB	POP-TAXMC
overall	.049	.091 (1.84)	.037 (.75)	.035 (.72)
1-199	.169	.276 (1.63)	.136 (.80)	.152 (.90)
200-499	.035	.153 (4.33)	.041 (1.17)	.029 (.83)
500-999	.019	.072 (3.82)	.020 (1.05)	.013 (.70)
1000-1999	.037	.038 (1.03)	.012 (.34)	.010 (.26)
2000-4999	.018	.018 (1.03)	.009 (.50)	.007 (.39)
5000+	.018	.011 (.59)	.006 (.34)	.005 (.25)

($\cdot$) MSdiff divided by that of official method, POP-CEN

POP-TAXPA, POP-TAXHYB and POP-TAXMC make use of the IY 1999 IRS income tax data

performed still better, beating the official method overall and for all size categories. While it showed only modest improvements on the official method in the three smaller size categories, for the largest three size categories it cut the MSDiff by more than half.

Evaluating the estimates as a function of the non-geocoding rate, we see from Table 9 that POP-TAXPA again performed worse than the official method across all the non-geocoding rate categories, while POP-TAXHYB and POP-TAXMC performed uniformly better than the official method. Figure 1 provides more detail on these comparisons (omitting POP-TAXPA since it performed so much worse than the other methods.) It shows that the relative performance of the POP-TAXHYB estimator deteriorated as the non-geocoding rate exceeded 20 percent, yielding somewhat larger errors than the official method. Additionally, we see that as the non-geocoding rate increased, the relative errors of POP-TAXHYB and POP-TAXMC tended to approach one. For POP-TAXMC this is expected since, as the non-geocoding rate increases, the minimum change shares will more often duplicate the shares from the previous census.

Table 9 - Mean squared differences between Census 2000 and alternative estimates of child population for school districts by categories of county non-geocoding rates

Non-geo rate	POP-CEN	POP-TAXPA	POP-TAXHYB	POP-TAXMC
overall	.049	.091 (1.84)	.037 (.75)	.035 (.72)
0-.08	.039	.048 (1.23)	.019 (.48)	.019 (.48)
.08-.15	.041	.102 (2.49)	.027 (.66)	.026 (.63)
.15-.225	.033	.093 (2.81)	.026 (.79)	.024 (.72)
.225-.40	.064	.130 (2.03)	.047 (.74)	.041 (.64)
>.40	.068	.103 (1.51)	.065 (.95)	.064 (.94)

(·) MSDiff divided by that of official method, POP-CEN

POP-TAXPA, POP-TAXHYB and POP-TAXMC make use of the IY 1999 IRS income tax data

The results from the evaluations of the population estimators for 1990 are shown in Tables 10-11 and Figure 2. The POP-TAXPA estimator again was almost uniformly worse than the official method, and the hybrid estimator (POP-TAXHYB) was worse than the official method except for the largest school districts and the districts with the lowest non-geocoding rates. The POP-TAXMC estimator was again better than the official method, both overall and for all the population size and non-geocoding rate categories. However, the overall relative MSDiff for the POP-TAXMC estimator was higher in 1990 than it was in 2000, which may be due to the higher non-geocoding rates for the tax exemptions in IY 1989. The minimum change allocation method will thus more often default back to using the census shares and will more often agree with the official method for estimating the population of school district pieces in 1990 versus 2000. The more detailed results in Figure 2 on error as a function of the non-geocoding rate are broadly consistent with the results in Figure 1, showing improvements on the official method by POP-TAXMC and POP-TAXHYB, though with the performance of the latter deteriorating some as the non-geocoding rate exceeds 20 percent.

It is somewhat surprising that the POP-TAXPA estimator does so poorly compared to the official method

Table 10 - Mean squared differences between 1990 Census and alternative estimates of child population for school districts by categories of school district size

Size	POP-CEN	POP-TAXPA	POP-TAXHYB	POP-TAXMC
overall	.049	.170 (3.47)	.061 (1.25)	.044 (.89)
1-199	.169	.498 (2.93)	.211 (1.24)	.152 (.89)
200-499	.035	.246 (7.00)	.062 (1.78)	.033 (.95)
500-999	.019	.142 (7.51)	.035 (1.84)	.017 (.94)
1000-1999	.037	.109 (2.95)	.044 (1.19)	.035 (.96)
2000-4999	.018	.043 (2.40)	.015 (.87)	.014 (.81)
5000+	.018	.018 (.99)	.010 (.57)	.012 (.68)

($\cdot$) MSDiff divided by that of official method, POP-CEN

POP-TAXPA, POP-TAXHYB and POP-TAXMC make use of the IY 1989 IRS income tax data

in 1990 and 2000, while the POP-TAXHYB does relatively well in 2000 and not so badly in 1990. To investigate this further we considered a general hybrid estimator that takes a weighted average of the POP-TAXPA population shares and the previous census shares, and we let the weight on the POP-TAXPA shares vary from 0 to 1. Thus, setting the weight to 1 gives the POP-TAXPA method, setting the weight to 0 gives the POP-CEN method, and setting the weight to .5 gives the POP-TAXHYB method. Figures 3 and 4 show plots of the overall MSDiffs for this weighted hybrid estimator in 2000 and 1990 as the weight varies from 0 to 1. We note that the optimal value of the weight in 2000 is around 0.4, so the hybrid estimator actually used (with weight equal to 0.5) did fairly well in 2000. For 1990 the optimal weight is closer to 0.1 to 0.2 so the hybrid estimator used doesn't do quite as well, and does a little worse than POP-CEN. We also notice that the MSDiff climbs sharply as the weight approaches 1 in both 2000 and 1990, consistent with the observed result that POP-TAXPA does poorly.

Table 11 - Mean squared differences between 1990 Census and alternative estimates of child population for school districts by categories of county non-geocoding rates

Non-geo rate	POP-CEN	POP-TAXPA	POP-TAXHYB	POP-TAXMC
overall	.049	.170 (3.47)	.061 (1.25)	.044 (.89)
0-.125	.038	.102 (2.68)	.031 (0.81)	.029 (.76)
.125-.30	.057	.265 (4.65)	.069 (1.21)	.043 (.75)
.30-.50	.036	.258 (7.16)	.070 (1.94)	.035 (.97)
.50-.70	.048	.153 (3.18)	.066 (1.37)	.045 (.93)
>.70	.054	.087 (1.61)	.060 (1.11)	.052 (.96)

$(\cdot)$ MSDiff divided by that of official method, POP-CEN

POP-TAXPA, POP-TAXHYB and POP-TAXMC make use of the IY 1989 IRS income tax data

4 Models for School District Poverty Estimates

In this section we examine estimators that use the adjusted tabulations of the IRS child and poor child tax exemptions to provide, during post-censual years, updated information related to child poverty for school district pieces. As with estimating population, the new models all produce estimates for the school district pieces ($j = 1, ..., J_i$) in each county i ($i = 1, ..., I$), which are then appropriately aggregated to provide estimates for the school districts. Also, the new models all predict school district to county shares of school-age child poverty, and these estimated shares are then multiplied by a county estimate of poverty. Empirical evaluations compare the resulting school district poverty estimates for IYs 1989 and 1999 against poverty estimates from the 1990 and 2000 censuses. We also provide analogous evaluation results for the official method of producing school district poverty estimates, and we examine the extent to which the experimental estimates may improve on the official poverty estimates.

The current method used to estimate the number of poor children in school district pieces for a post-censal year is based on a synthetic approach analogous to the approach used for producing post-censal school district population estimates. The previous census data are used to estimate school district piece to county shares of poverty, and these estimated shares are then multiplied by the SAIPE model-based county child poverty estimate for the current year. Thus, to estimate the number of poor children in school district piece j of county i in IY 1999 (without using the 2000 census results) we have:

$$\text{Poor }\widehat{\text{Children}}_{99,ij} = \frac{C^{\text{poor}}_{90,ij}}{\sum_j C^{\text{poor}}_{90,ij}} \times \text{CNTY-POV}_{99,i} \tag{7}$$

where $C^{\text{poor}}_{90,ij}$ is the smoothed 1990 census long form estimate of related 5-17 children in families in poverty for school district piece j of county i, and $\text{CNTY-POV}_{99,i}$ is the IY 1999 SAIPE model-based estimate for county i of the number of poor children ages 5-17. The underlying assumption with this approach is that the distribution of child poverty among school district pieces within a county does not change over time. We want to explore estimation methods that use the current-year IRS data to reflect changes over time in the within county distribution of poor children.

(Note: In constructing the official school district poverty estimates the census long form estimates are "smoothed." See U.S. Census Bureau (2005a).)

4.1 Using IRS Data in Share Models for Poverty Estimates

Since all estimates for school district pieces will be controlled to the official county estimate of poor children age 5-17 to maintain consistency, it is sufficient to estimate the school district piece to county poverty shares. Share models attempt to describe the distribution of poor 5-17 children among the school district pieces within a county. Note that within a county the estimated shares must add up to 100 percent.

We will present results for nine methods for estimating the school district piece to county poverty shares of 5-17 related children. These include the current official method and six new methods that could be implemented in practice. We will also examine two benchmark models that could not be used in practice because they make use of the true population, which is unknown. However, for evaluation purposes the

1990 and 2000 census counts are standing in for the true population, and these are used by the benchmark models.

The first three methods that we consider directly estimate school district piece to county poverty shares (for school-aged children).

POV-CEN : use poverty shares from the previous census (official method) see Eq. (7)

POV-TAXPA : use current year shares of the adjusted poor child tax exemptions with non-geocoded tax exemptions allocated proportional to the previous census population counts.

$$\text{POV-T}\widehat{\text{AX}}\text{PA}_{99,ij} = \frac{T^{\text{poor}}_{(pa)99,ij}}{\sum_j T^{\text{poor}}_{(pa)99,ij}} \times \text{CNTY-POV}_{99,i} \tag{8}$$

POV-TAXMC : use current year shares of the adjusted poor child tax exemptions with non-geocoded tax exemptions allocated based on the minimum change method.

$$\text{POV-T}\widehat{\text{AX}}\text{MC}_{99,ij} = \frac{T^{\text{poor}}_{(mc)99,ij}}{\sum_j T^{\text{poor}}_{(mc)99,ij}} \times \text{CNTY-POV}_{99,i} \tag{9}$$

The next six estimators (which include the two benchmark estimators) use current year tax poverty rates and population estimates to estimate the number of poor children in school district pieces, and these results are used to estimate the school district piece to county poverty shares. We call these "tax poverty rate share models." Doing this with proportional allocation of non-geocoded tax exemptions we would proceed as follows:

1. For whole school district sd, compute the IRS tax poverty rate for children as the adjusted number of poor child tax exemptions divided by the adjusted number of all child tax exemptions:

$$\text{POVRT(PA)}_{sd} = \frac{\sum_{(i,j)\in sd} T^{\text{poor}}_{(pa)IY,ij}}{\sum_{(i,j)\in sd} T^{\text{tot}}_{(pa)IY,ij}}$$

2. Create a preliminary estimate of the number of poor 5-17 children in school district piece j of county i by multiplying the tax poverty rate for its school district ($sd(ij)$), computed in step 1, by the estimated

number of school-age children in the piece.

$$\text{Preliminary(PA)}_{ij} = \text{POVRT(PA)}_{sd(ij)} \times \text{Est-}\widehat{\text{Child}}\text{-Pop}_{ij}$$

3. Form the school district piece to county poverty shares from these preliminary estimates and multiply them by the model based county poverty estimate, CNTY-POV_i.

$$\text{Est-}\widehat{\text{Child}}\text{-Pov}_{ij} = \frac{\text{Preliminary(PA)}_{ij}}{\sum_j \text{Preliminary(PA)}_{ij}} \times \text{CNTY-POV}_i$$

4. Sum the estimates from the pieces to obtain the estimated number of poor school-age children for the whole school district:

$$\text{Est-}\widehat{\text{Child}}\text{-Pov}_{sd} = \sum_{(i,j)\in sd} \text{Est-}\widehat{\text{Child}}\text{-Pov}_{ij}$$

The minimum change allocation for non-geocoded exemptions can also be used in the above procedure. The procedure can also use the different estimators from Section 3 of the population of children ages 5-17 in school district pieces. We will use two of the population estimates as noted below. In addition, we will create two benchmark estimators, one for each of the two methods of allocating non-geocoded poor child tax exemptions, but using the true school district piece child population figures (census counts) at Step 2. These benchmark estimators represent the best case for the tax poverty rate share estimators since they eliminate error from having to estimate the school district piece populations. We will thus consider the following six poverty rate share estimators:

Estimator	non-geocoded allocation	population estimate
POV-RT-PA-HYB	Proportional to Pop.	POP-HYB
POV-RT-PA-CEN	Proportional to Pop.	POP-CEN
POV-RT-MC-TAXMC	Minimum Change	POP-TAXMC
POV-RT-MC-CEN	Minimum Change	POP-CEN
POV-RT-PA-TRUE	Proportional to Pop.	true pop (from census)
POV-RT-MC-TRUE	Minimum Change	true pop (from census)

For both tax exemption allocation methods, we use the corresponding population estimator (from Section 3), the population estimator based on the official method, and the true population count. Comparing evaluation results for the first four estimators against those for the benchmark estimators will give us an indication of the effect of errors in the population estimates on the accuracy of the poverty estimates.

4.2 Evaluation of Poverty Estimators

In this section we apply the alternative estimators just discussed to produce school district poverty estimates for IY 1999, and we compare these against Census 2000 poverty estimates. For the official method we will use the 1990 census as the previous census and take its estimated school district piece to county poverty shares as estimates of the IY 1999 shares. Analogous to the evaluation of the population estimates we will multiply all the various estimated shares by the census county poverty estimates because our interest is in evaluating the alternative share estimates. We will also reverse the process and apply the estimators to produce school district poverty estimates for IY 1989, evaluating these against the 1990 census results.

One difference from the evaluation of the population estimates in the previous section is that the 1990 and 2000 census long form estimates contain sampling error. This affects both the previous census estimates used in the official method and the census estimates used as targets in the evaluations. The sampling error variance in the targets could be accounted for in the evaluation measures, but Bell and Kramer (1999) found that while doing so lowered the mean squared differences in the log scale (since the sampling relative variances are subtracted from the mean squared differences) this did not change any overall conclusions about the accuracy of the estimators. (They evaluated the official estimator against an estimator that assumed uniform change in poverty across the country over the decade.) Also, there is current research going on to produce improved variance estimates for the Census 2000 long form estimates used by SAIPE (such as school district 5-17 poverty estimates.) Refinement of the comparisons presented here to take sampling variances into account may be done when improved sampling variance estimates become available.

Our evaluation will make use of two metrics to compare estimators against the Census 2000 (or 1990 census) results: mean squared difference in the log scale (MSDiff-1) and mean squared difference of poverty rates weighted by the square root of population (MSDiff-2). More specifically, the first metric compares the mean squared difference of the log number of poor 5-17 children estimated by our models to the corresponding Census 2000 long form estimate averaging over all school districts subscripted by sd:

$$\text{MSDiff-1} = 1/N \sum_{sd} [\log(C^{\text{poor}}_{2K,sd} + 1) - \log(ESTPOV_{sd} + 1)]^2 \tag{10}$$

where $ESTPOV_{sd}$ is any of the seven usable estimators or the two benchmarks listed in the previous section. We use $\log(x+1)$ to deal with the occasional zeros in the data, which occurs more frequently with the census estimates of number of poor children than with the population estimates evaluated in Section 3. The second metric compares the estimated school district poverty rate to the census poverty rate and weights the squared difference by the square root of the census population:

$$\text{MSDiff-2} = 1/N \sum_{sd} C^{\text{tot}}_{2K,sd} \left[\frac{C^{\text{poor}}_{2K,sd}}{C^{\text{tot}}_{2K,sd}} - \frac{ESTPOV_{sd}}{ESTPOP_{sd}} \right]^2 \tag{11}$$

where $ESTPOP_{sd}$ matches the methodology use for $ESTPOV_{sd}$. This second metric gives larger school districts more weight compared to MSDiff-1, but it is not as skewed towards the largest districts as would be the squared error on the number of estimated poor children. We will use the same size and non-geocoding rate categories as listed in Table 6 and Table 7 and also compute ratios of the MSDiffs for the alternative estimators to those for POV-CEN (analogously to Eq(6)) for both metrics.

We will also plot the absolute difference between the log estimated number of poor children and the log number of poor children estimated by the census. We plot these individual "errors" as a function of non-geocoding rates and school district population sizes, using Loess local regression lines to smooth out what would otherwise be very noisy scatterplots.

The evaluation results for IY 1999 are given in Tables 12-13 and Figures 5 and 7 for MSDiff-1, and in Tables 14-15 for MSDiff-2. In Table 12 all of the estimators are presented, while in Tables 13-15 and the plots we focus on the POV-RT-PA-HYB and POV-TAXMC estimators. The former is the best of the PA group of estimators, while the three usable MC estimators all perform similarly for most evaluations. Under MSDiff-1 (Tables 12 and 13), overall, as well as for each size group and for each non-geocoding rate group, all of the estimators using information from the IY 1999 IRS income tax data gave a better prediction of the Census 2000 estimates compared to using the smoothed census shares from 1990. Figure 5 suggests, however, that the POV-RT-PA-HYB estimator performed poorly as the non-geocoding rate became severe (> 80 percent). The PA method for allocating non-geocoded exemptions gave slightly better results than MC under MSDiff-1. As expected, the MSDiff-1's relative to those of the official estimator were better for larger school districts and for school districts with lower non-geocoding rates. Under MSDiff-2 (Tables 14 and 15), the PA method performed somewhat better than MC for all groups except the largest school districts and the districts with the highest and lowest non-geocoding rates. From Table 12, examining the benchmark estimators shows that there is room for about a 7-8% improvement overall on the MSDiff-1 error metric if we could eliminate error in the population estimates for the school district pieces.

To validate the results, we used our models and the IY 1989 IRS income tax data to estimate school district poverty for IY 1989 and compared these results against the long form estimates from the 1990 Census. We also applied the official method with the data reversed in time, taking school district piece to county poverty shares from Census 2000 as the "previous census shares." The mean squared differences and relative

Table 12 - MSDiff-1 for Census 2000 and alternative estimates of poor 5-17 children in school districts by population size

Size	POV-CEN	POV-TAXPA	POV-RT-PA-HYB	POV-RT-PA-CEN	POV-RT-PA-TRUE
overall	.398	.345 (.85)	.315 (.79)	.329 (.82)	.294 (.73)
1-199	.969	.846 (.94)	.838 (.86)	.913 (.94)	.751 (.77)
200-499	.477	.432 (.84)	.350 (.73)	.357 (.74)	.328 (.68)
500-999	.432	.405 (.87)	.368 (.85)	.369 (.85)	.361 (.83)
1000-1999	.306	.252 (.79)	.227 (.74)	.234 (.76)	.224 (.73)
2000-4999	.194	.151 (.77)	.129 (.66)	.125 (.64)	.121 (.62)
5000+	.091	.067 (.56)	.054 (.58)	.057 (.62)	.047 (.51)
Size	POV-CEN	POV-TAXMC	POV-RT-MC-TAXMC	POV-RT-MC-CEN	POV-RT-MC-TRUE
overall	.398	.339 (.86)	.341 (.85)	.342 (.85)	.315 (.79)
1-199	.969	.915 (.94)	.926 (.95)	.932 (.96)	.809 (.83)
200-499	.477	.405 (.85)	.404 (.84)	.401 (.84)	.381 (.79)
500-999	.432	.376 (.87)	.378 (.87)	.377 (.87)	.369 (.85)
1000-1999	.306	.244 (.80)	.245 (.80)	.247 (.80)	.240 (.78)
2000-4999	.194	.128 (.66)	.128 (.66)	.127 (.65)	.123 (.63)
5000+	.091	.052 (.56)	.052 (.56)	.056 (.61)	.044 (.48)

(·) MSDiff-1 divided by that of official method, POV-CEN

Table 13 - MSDiff-1 for Census 2000 and alternative estimates of poor 5-17 children in school districts by non-geocoding rate

Non-geo rate	POV-CEN	POV-RT-PA-HYB	POV-TAXMC
overall	.398	.315 (.79)	.339 (.85)
0-.08	.529	.362 (.68)	.381 (.72)
.08-.15	.463	.341 (.73)	.375 (.81)
.15-.225	.371	.289 (.77)	.315 (.85)
.225-.40	.369	.309 (.83)	.342 (.92)
>.40	.390	.371 (.95)	.387 (.99)

(·) MSDiff-1 divided by that of official method, POV-CEN

Table 14 - MSDiff-2 for Census 2000 and alternative estimates of poor 5-17 children in school districts by population size

Size	POV-CEN	POV-RT-PA-HYB	POV-TAXMC
overall	3.386	2.406 (.71)	2.306 (.68)
1-199	1.195	.975 (.81)	1.255 (1.05)
200-499	1.665	1.269 (.76)	1.644 (.99)
500-999	2.027	1.656 (.82)	1.750 (.86)
1000-1999	2.544	1.941 (.76)	2.018 (.79)
2000-4999	3.538	2.280 (.64)	2.257 (.64)
5000+	9.832	6.693 (.68)	5.147 (.52)

(·) MSDiff-2 divided by that of official method, POV-CEN

Table 15 - MSDiff-2 for Census 2000 and alternative estimates of poor 5-17 children in school districts by non-geocoding rate

Non-geo rate	POV-CEN	POV-RT-PA-HYB	POV-TAXMC
overall	3.386	2.406 (.71)	2.306 (.68)
0-.08	6.569	4.142 (.63)	3.442 (.52)
.08-.15	3.770	2.551 (.68)	2.431 (.64)
.15-.225	2.841	1.775 (.62)	1.811 (.64)
.225-.40	2.847	2.012 (.71)	2.405 (.84)
>.40	1.889	2.112 (1.12)	1.940 (1.03)

$(\cdot)$ MSDiff-2 divided by that of official method, POV-CEN

MSDiffs of the alternative estimators are given in Tables 16-19 and Figures 6 and 8. Similarly to the results of the validation study done for the population estimators in Section 3, the Census 1990 results (here, the estimated number of poor school age children) were harder to predict than the Census 2000 results. Under MSDiff-1, the PA based estimators struggled with districts with more than 2000 children, while the MC based estimators only struggled with the largest districts with more than 5000 children. Upon further review, the largest errors in POV-TAXMC for large school districts were in districts with very small poverty rates, hence in those with a small number of poor children. In these cases a small difference in the estimated number of poor children results in a large relative difference. Using the "previous census shares" from Census 2000 (official method reversed in time) worked better in this case. Additionally, the POV-RT-PA-HYB estimator struggled a bit for the districts with the lowest non-geocoding rates (Table 17). This is actually where we would expect this estimator to be at its best. The accuracy of the POV-TAXMC estimator hovers closely around the accuracy of the official method for all groups of non-geocoding rates, except for doing somewhat better for the second category. This is also borne out in Figure 6.

Under MSDiff-2 (Tables 18-19), a slightly different set of results occurs. First, both estimators perform about the same or better than the official method across all age groups with the exception of POV-RT-PA-HYB for the school districts with the largest non-geocoding rates. Given the much higher non-geocoding rates in the IY 1989 IRS income tax data, this shows how the MC allocation method is a conservative approach that defaults back to the official method when the IRS data is not able to contribute much information. Among the school district size groups, only for the largest school districts (5000 or more children) did both of the two new estimators show marked improvement over the official method when predicting the 1990 Census results. POV-TAXMC also showed fairly substantial improvement for the 2000-4999 size category. For the geocoding rate categories both of the alternative estimators showed marked improvements for the lowest non-geocoding rates, and POV-TAXMC also showed improvement for the next highest non-geocoding rate category. As noted above, POV-RT-PA-HYB did poorly for the highest category of non-geocoding rates.

Overall, there isn't a single poverty estimator that gives better predictions than the official method uniformly over the different size categories and non-geocoding rates for both years 1990 and 2000. All the estimators using the IRS data predict the Census 2000 results better than the official method. In predicting the 1990 Census results, the POV-TAXMC estimator performed about the same as the official method for most groups, and did show marked improvement compared to the official method in a few instances under MSDiff-2 (for large school districts or school districts with low non-geocoding rates). POV-RT-PA-HYB, in contrast, did poorly in a few of the 1990 comparisons.

5 Discussion

In this report we presented various models that make use of the IRS income tax data to estimate the number of school-aged children and the number of school-aged children in poverty for school districts. Estimates were constructed and their accuracy assessed by comparing them to 100% population counts and long form poverty estimates from the 1990 and 2000 censuses.

Table 16 - MSDiff-1 for 1990 Census and alternative estimates of poor 5-17 children in school districts by population size

Size	POV-CEN	POV-TAXPA	POV-RT-PA-HYB	POV-RT-PA-CEN	POV-RT-PA-TRUE
overall	.507	.535 (1.06)	.508 (1.01)	.538 (1.07)	.498 (.99)
1-199	1.338	1.321 (.98)	1.322 (.98)	1.412 (1.05)	1.430 (1.06)
200-499	.663	.706 (1.06)	.637 (.96)	.672 (1.01)	.629 (.94)
500-999	.474	.500 (1.05)	.457 (.96)	.486 (1.02)	.439 (.92)
1000-1999	.327	.361 (1.10)	.332 (1.01)	.347 (1.06)	.292 (.89)
2000-4999	.230	.291 (1.26)	.274 (1.18)	.282 (1.22)	.227 (.98)
5000+	.100	.147 (1.46)	.145 (1.44)	.156 (1.55)	.102 (1.01)
Size	POV-CEN	POV-TAXMC	POV-RT-MC-TAXMC	POP-RT-MC-CEN	POP-RT-MC-TRUE
overall	.507	.495 (.98)	.497 (.99)	.496 (.98)	.469 (.93)
1-199	1.338	1.339(1.00)	1.349(1.01)	1.342 (1.00)	1.351 (1.01)
200-499	.663	.647 (.97)	.646 (.97)	.645 (.97)	.608 (.91)
500-999	.474	.456 (.96)	.458 (.96)	.458 (.96)	.426 (.90)
1000-1999	.327	.308 (.94)	.308 (.94)	.307 (.93)	.273 (.83)
2000-4999	.230	.235 (1.02)	.237 (1.02)	.236 (1.02)	.201 (.87)
5000+	.100	.111 (1.10)	.111 (1.11)	.113 (1.13)	.080 (.79)

(·) MSDiff-1 divided by that of official method, POV-CEN

Table 17 - MSDiff-1 for 1990 Census and alternative estimates of poor 5-17 children in school districts by non-geocoding rate

Non-geo rate	POV-CEN	POV-RT-PA-HYB	POV-TAXMC
overall	.507	.509 (1.00)	.495 (.97)
0-.125	.624	.669 (1.07)	.602 (.96)
.125-.30	.618	.604 (.98)	.573 (.92)
.30-.50	.538	.525 (.98)	.521 (.97)
.50-.70	.482	.473 (.98)	.492 (1.02)
>.70	.436	.434 (.98)	.446 (1.02)

(·) MSDiff-1 divided by that of official method, POV-CEN

Table 18 - MSDiff-2 for 1990 Census and alternative estimates of poor 5-17 children in school districts by population size

Size	POV-CEN	POV-RT-PA-HYB	POV-TAXMC
overall	3.941	3.525 (.89)	3.040 (.77)
1-199	2.145	2.062 (.96)	2.282 (1.06)
200-499	2.571	2.473 (.96)	2.615 (1.02)
500-999	2.495	2.289 (.92)	2.325 (.93)
1000-1999	2.948	2.967 (1.01)	2.652 (.90)
2000-4999	3.584	3.543 (.99)	3.036 (.85)
5000+	10.595	8.202 (.77)	5.570 (.52)

(·) MSDiff-2 divided by that of official method, POV-CEN

Table 19 - MSDiff-2 for 1990 Census and alternative estimates of poor 5-17 children in school districts by non-geocoding rate

Non-geo rate	POV-CEN	POV-RT-PA-HYB	POV-TAXMC
overall	3.941	3.525 (.89)	3.040 (.77)
0-.125	7.774	5.143 (.66)	4.327 (.56)
.125-.30	4.328	4.416 (1.02)	3.334 (.77)
.30-.50	3.721	3.382 (.91)	3.227 (.87)
.50-.70	2.887	2.788 (.96)	2.859 (.99)
>.70	2.374	3.099 (1.30)	2.393 (1.01)

$(\cdot)$ MSDiff-2 divided by that of official method, POV-CEN

The evaluation results presented here in Sections 3 and 4 show improvements in the estimators when incorporating the IRS income tax data in both the school age child population estimates for 2000 and the school age child poverty estimates for IY 1999, and also for the school age child population estimates for 1990. However, the results are not as clear when evaluating the school age child poverty estimates for IY 1989. Overall, using the Minimum Change method to allocate the non-geocoded tax exemptions resulted in estimates that more consistently performed either better or about the same as the official method across both 1990 and 2000. The estimates using Proportional Allocation were worse than those from the Minimum Change method for estimating population. While PA in some instances did better than MC for estimating school district poverty, in a few cases it performed poorly.

There are still issues that need to be addressed regarding use of the IRS tax data at the school district level. Clearly, from the results here, the main issue appears to be dealing with the non-geocoded tax exemptions. This issue is important since some counties have a large percentage of non-geocoded tax exemptions. The new estimators tended to show less gain over the official method when non-geocoding rates were higher, particularly for the poverty estimates. While alternative models for allocating the non-geocoded exemptions could be investigated, a better long-term solution is to assign more of the tax returns to school districts. Geocoding rates should continue to improve over time as more rural areas adopt city-style addresses. But the situation might also be improved by directly assigning some tax returns to school districts that cannot be assigned under the current scheme that requires first geocoding the return to a census block. Investigation of this possibility involves significant efforts in computer processing, resources for which are not currently available, but it is hoped that this alternative can be explored in coming years.

Another issue with the tax data is that we do not know the non-filing rates (proportion of population not represented as exemptions on tax returns) for school districts or school district pieces within counties. Currently we implicitly assume that the non-filing rate is constant across pieces within a county. Data to estimate non-filing rates for school district pieces are not available. Thus, differences for school districts between the true population count and the number of tax exemptions can be due to any combination of non-geocoded exemptions and non-filing.

The official and the alternative population and poverty estimators considered here are all forms of synthetic estimators. More formal statistical models were considered early in the research process but, after raking school district piece estimates to agree with official county estimates, some of these more formal models produced estimates nearly identical to the share models POV-TAXMC and POV-TAXPA (Maples and Bell 2004). For this reason, and because the difficult data issues that must be addressed with the IRS income tax data seem more important to the accuracy of the estimators, these more formal models were not explored here.

One issue not addressed in this paper is obtaining error variances for the school district estimators considered here. This would be somewhat complicated because the estimators considered here are subject to multiple sources of error. First, there is error in the model-based county estimates for population or poverty. Second, there is error associated with having to allocate the non-geocoded exemptions from the IRS income tax data. Third, there is model error in that the IRS income tax data does not perfectly estimate the school district piece to county shares for population and poverty. Fourth, for the official method of poverty estimates there is sampling error in the previous census long form estimates that are used.

Finally, data from the American Community Survey (ACS), which has recently gone to full implementation, will eventually be a rich source of information about the total number of children and the number of children in poverty at the school district level. Even with its large sample size (about 3 million addresses per year nationally), ACS estimates for many small school districts will have substantial levels of sampling error (as do the census long form poverty estimates). To help address this issue, ACS estimates for smaller places will be based on 3-year or 5-year data collections, which means that such estimates will not be available for some years. Using the ACS data to provide updated information about school district population and poverty, and possibly combining this with information from the IRS tax data, is an important topic for future research.

References

Bell, William R. and Kramer, Matthew (1999). "On School District Error Decomposition (7/8/99 draft)," unpublished report, Statistical Research Division, U.S. Census Bureau.

Cleveland, W.S. (1979) "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, **74**, 828-836.

Fay, R. E. and Herriot, R. A. (1979) "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, **74**, 269-277.

Maples, J. and Bell, W. (2004), "Investigating the use of IRS Tax Data in the SAIPE School District Poverty Estimates," Proceedings of the American Statistical Association.

U.S. Census Bureau [a] (2005), "Small Area and Income Estimates - School District Estimates," http://www.census.gov/hhes/www/saipe/, accessed on Oct 11, 2005.

U.S. Census Bureau [b] (2005), "Population Estimates," http://www.census.gov/popest/estimates.php, accessed on Nov 15, 2005.

Appendix A: Creating Minimum Change Shares

Algorithm for finding the minimum change shares (MC).

Let j denote the $1, .., J_i$ school district pieces within a given county i and T^*_{sij} denote the age-adjusted geocoded IRS income tax share, i.e. the proportion of age-adjusted geocoded child (or poor child) exemptions for school district piece (i, j) to the age-adjusted child (or poor child) exemptions for the county, where $(\sum_j T^*_{sij} \leq 1)$, (see Section 2.2). Let $T^*_{s,ng,i}$ be the share of age-adjusted child (or poor child) non-geocoded exemptions which gives $\sum_j T^*_{sij} + T^*_{s,ng,i} = 1$.

Define C_{sij} as the census share for school district piece (i, j). For population, this share is based school district piece to county proportion of the 100% count for relevant 5-17 year olds. For poverty, this share is based on the school district piece to county proportion of the smoothed estimate of relevant 5-17 year olds in families from the census long form. Note that $\sum_j C_{sij} = 1$.

Let M_{sij} denote the MC shares we are creating. Also, define I_{ij} as an indicator denoting which pieces have already had their minimum change share determined.

1. Begin: Set I_{ij}=0 for $j = 1, .., J_i$
2. If all IRS income tax shares are less than or equal to census shares ($T^*_{sij} \leq C_{sij}$ for all j) then $M_{sij} = C_{sij}$ for $j = 1, , .J_i$. Goto step 5.
3. For pieces where $T^*_{sij} \geq C_{sij}$, $M_{sij} = T^*_{sij}$ and $I_{ij} = 1$.
4. For pieces where $I_{ij} = 0$:
   (a) Let $M_{tot} = 1 - \sum_j I_{ij} M_{sij}$. This is the total amount left to be allocated among the remaining pieces.

(b) Let $C_{tot} = 1 - \sum_j I_{ij} C_{sij}$. This is the total of census shares corresponding to the remaining pieces.

(c) Let $m^*_{sij} = \frac{C_{sij}}{C_{tot}} M_{tot}$. let m^*_{sij} be the trial allocation of the remaining shares proportional to the corresponding census shares.

(d) For pieces where $m^*_{sij} < T^*_{sij}$: Set $M_{sij} = T^*_{sij}$ and $I_{ij} = 1$

(e) Go back to Step (a) until no change in M_{tot}

(f) Set $M_{sij} = m^*_{sij}$ where $I_{ij} = 0$.

5. End of Algorithm, all MC shares, M_{sij}, should be determined.

Figure 1: Abs Log Diff 2000 Pop by nongeocoding rate

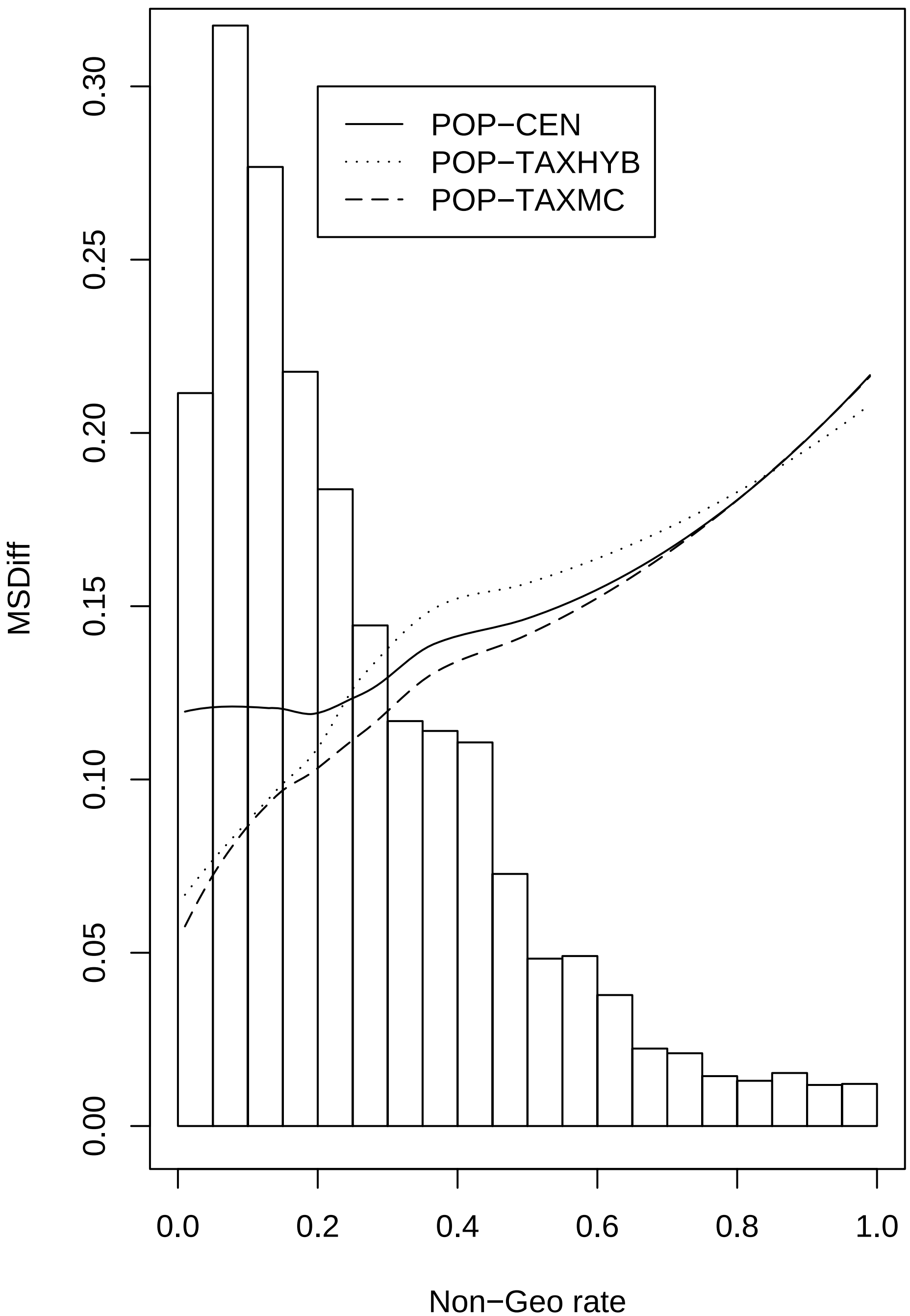

Figure 2: Abs Log Diff 1990 Pop by nongeocoding rate

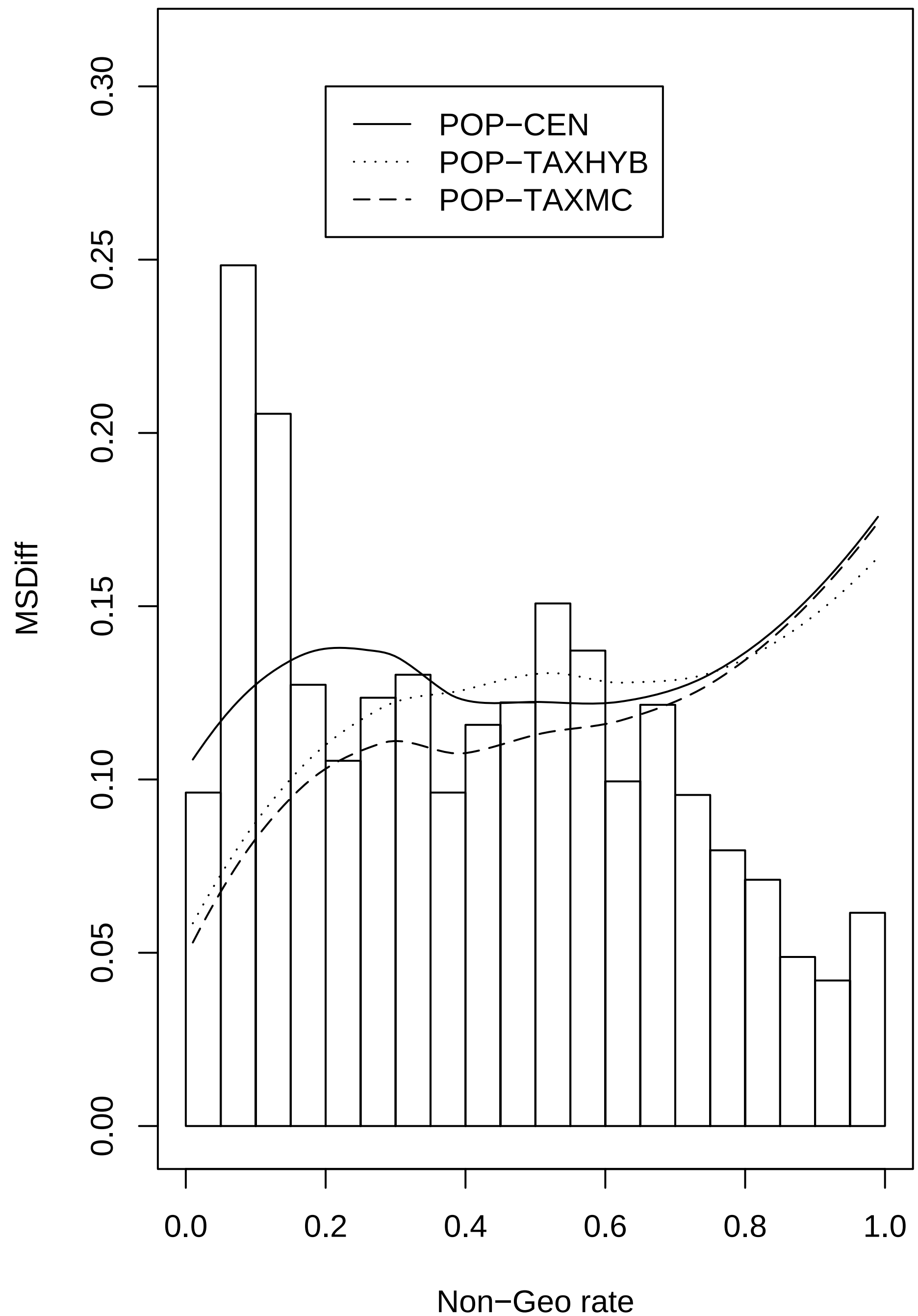

Figure 3: MSDiff for Generalized Hybrid Estimators − 2000

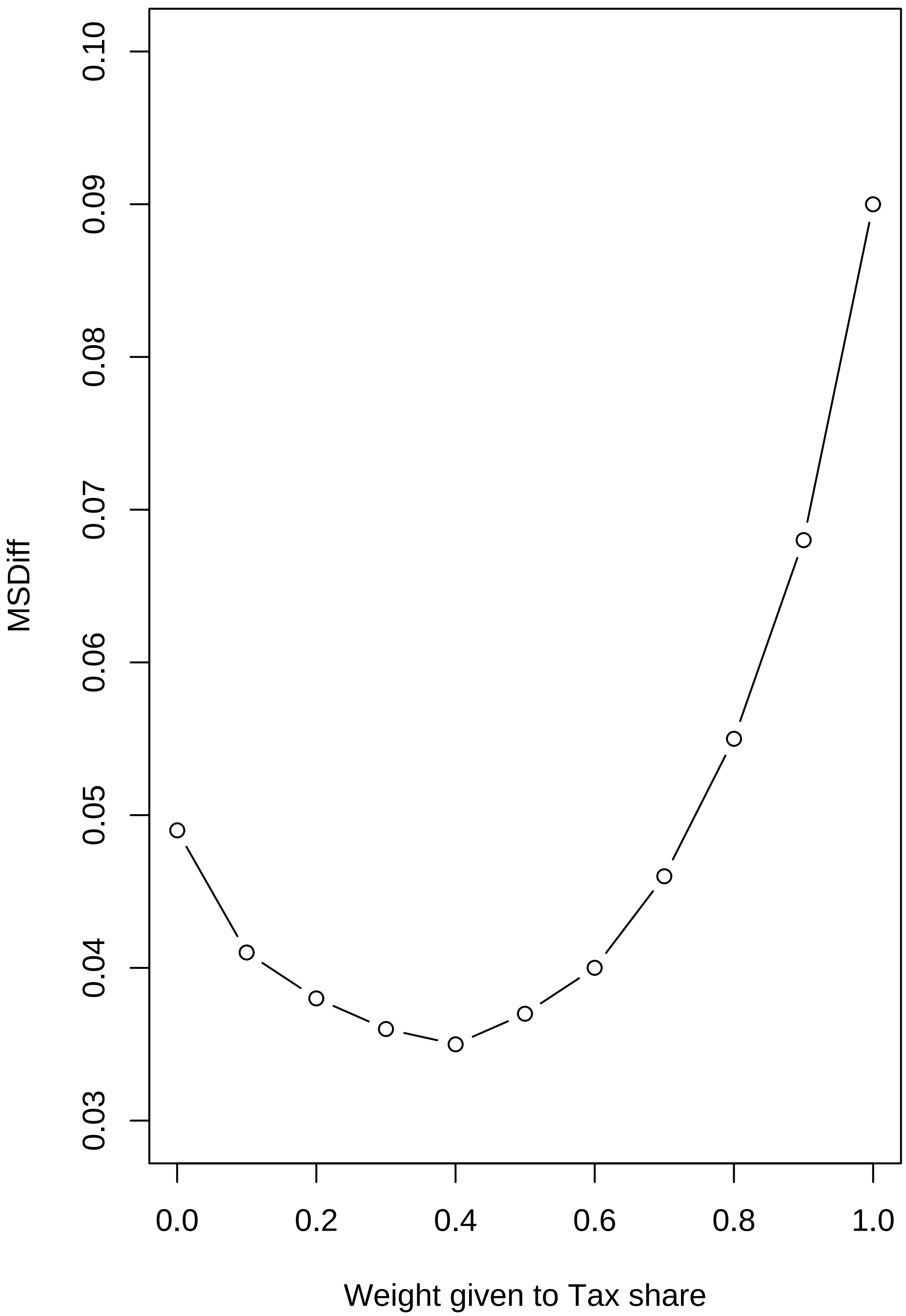

Figure 4: MSDiff for Generalized Hybrid Estimators − 1990

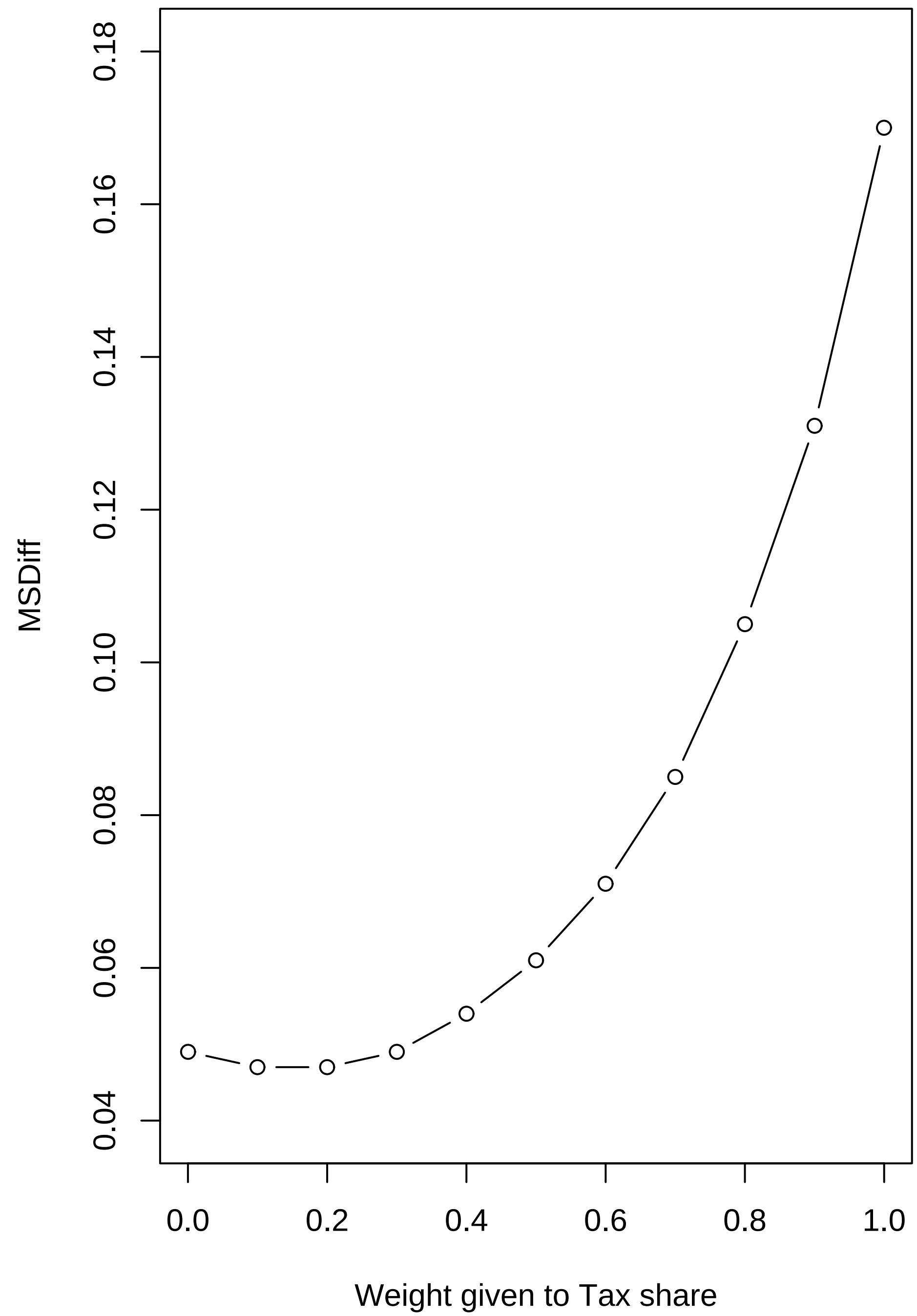

Figure 5: Abs Log Diff 2000 Pov by nongeocoding rate

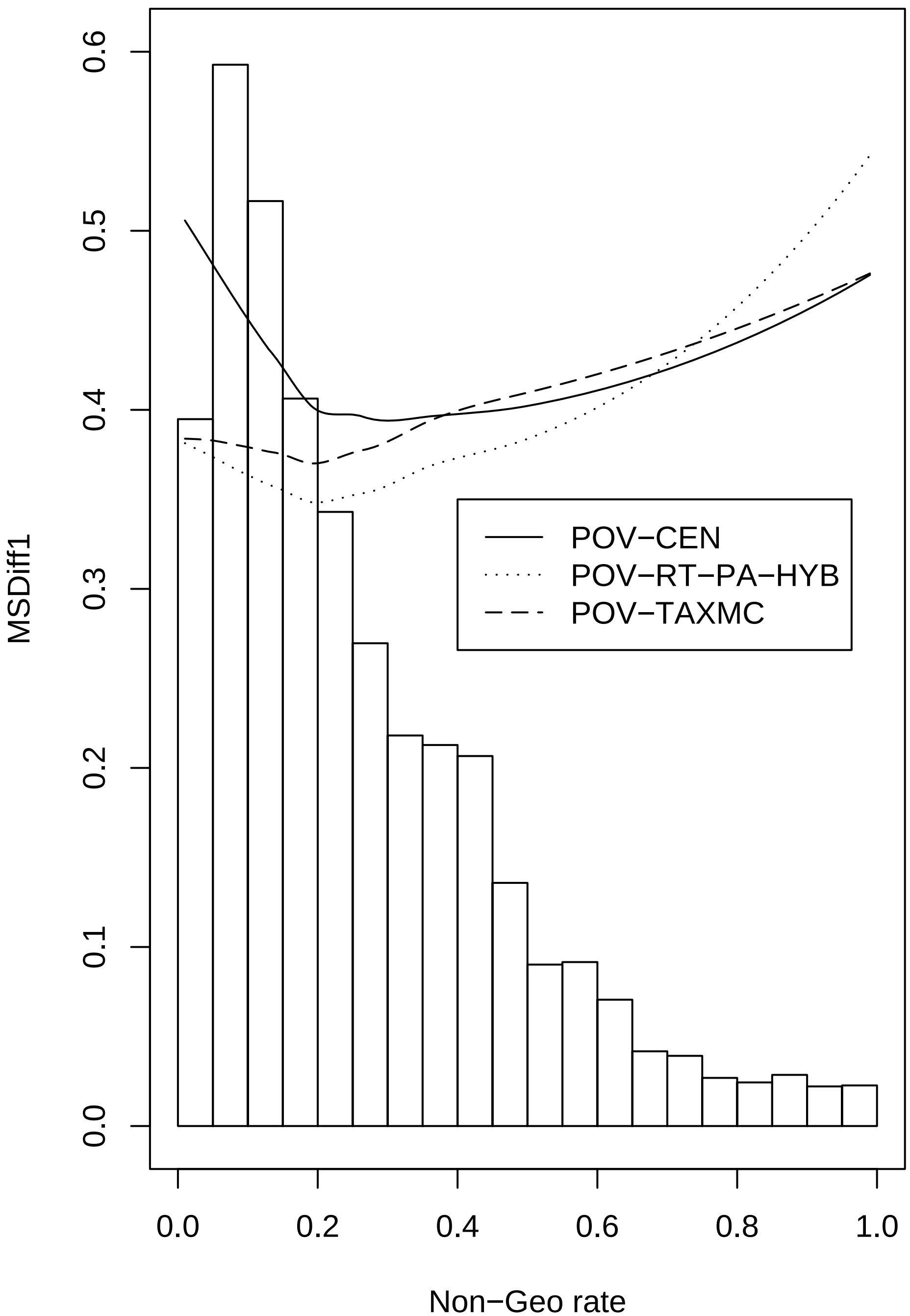

Figure 6: Abs Log Diff 1990 Pov by nongeocoding rate

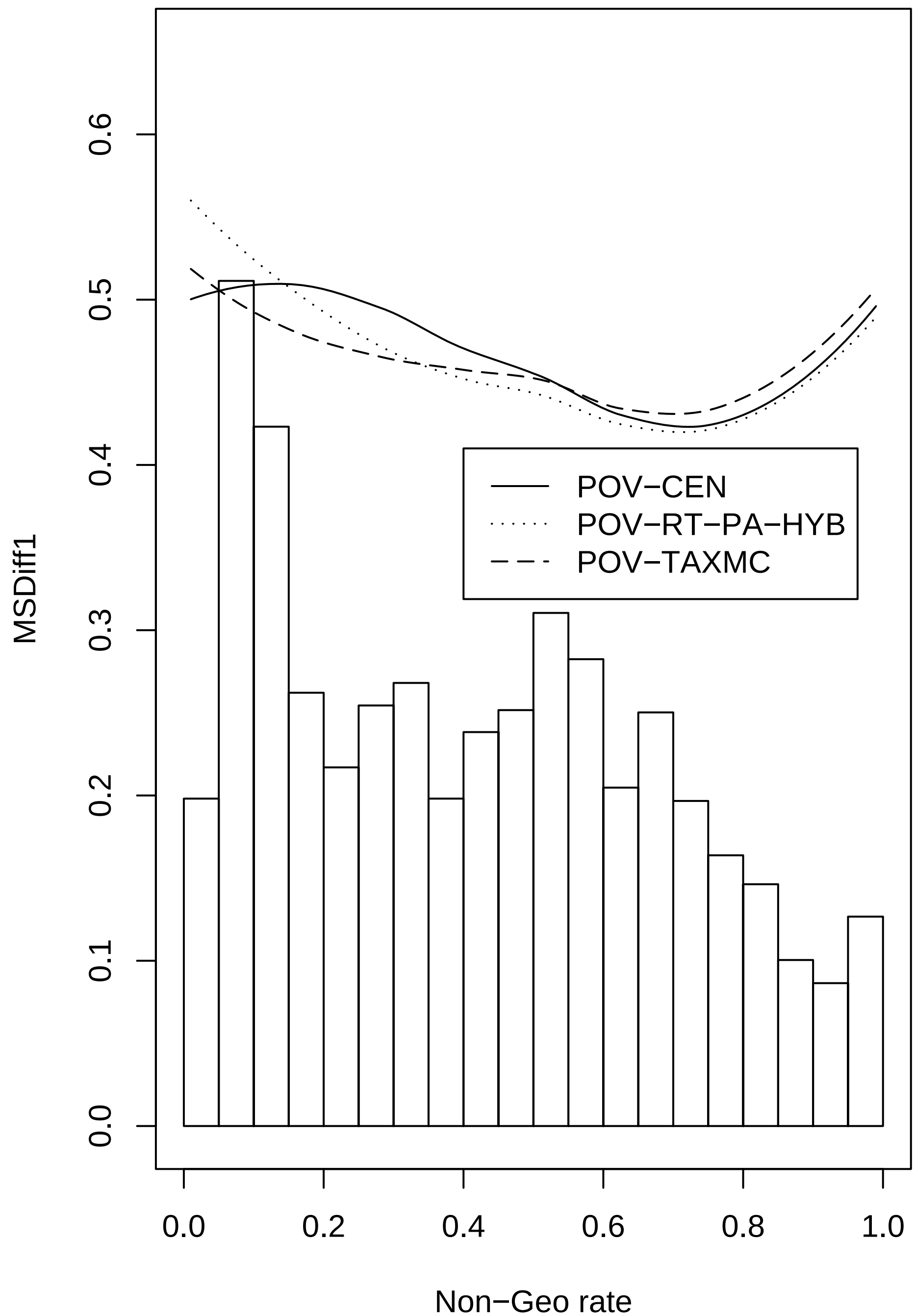

Figure 7: Abs Log Diff 2000 Pov by log pop

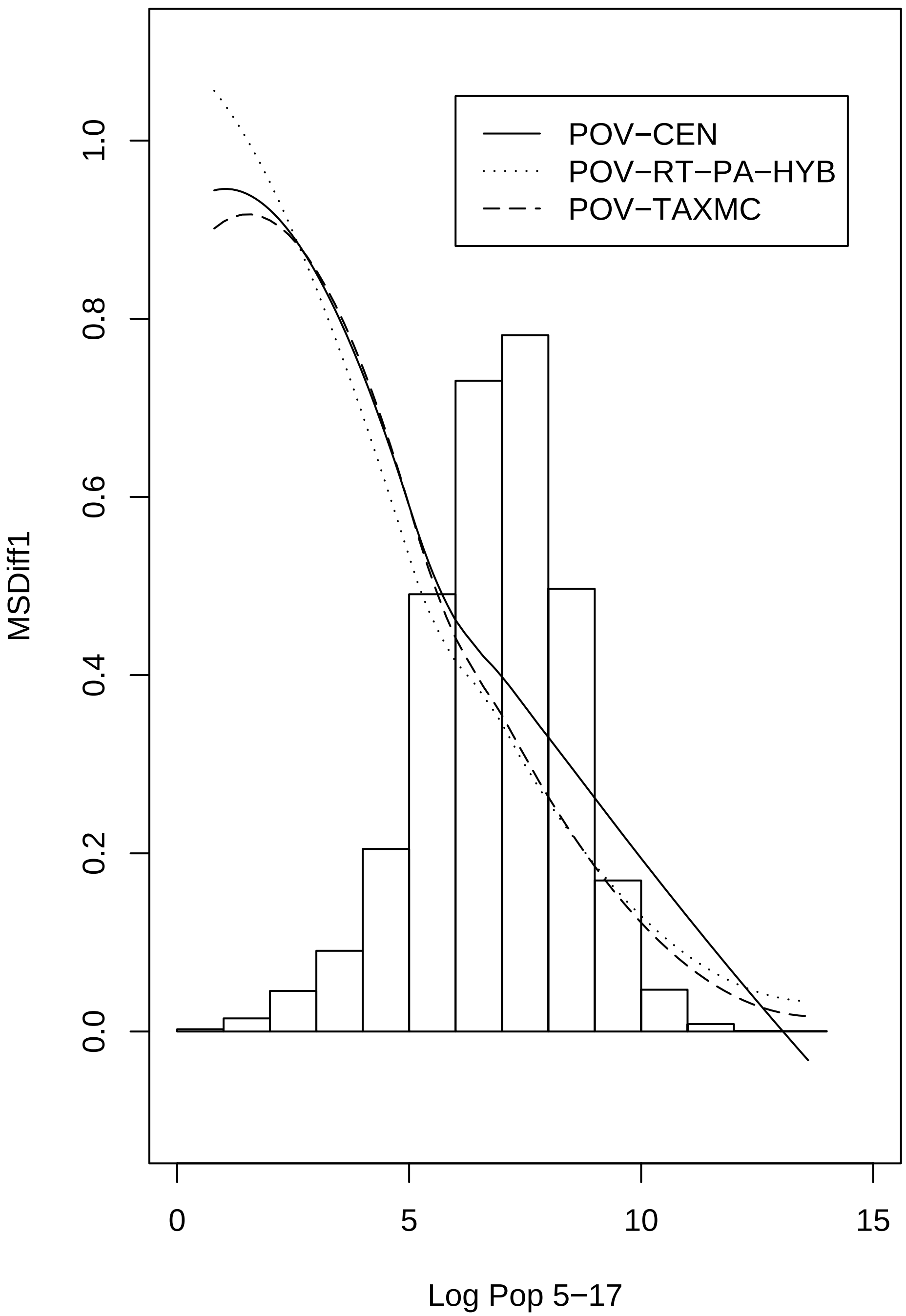

Figure 8: Abs Log Diff 1990 Pov by log pop