Income Sorting: Measurement and Decomposition

Thomas Davidoff *

May 14, 2005

*davidoff@haas.berkeley.edu, Haas School of Business, UC Berkeley. This paper is a revision of a chapter in my MIT dissertation. I thank William Wheaton, Peter Diamond, Sendhil Mullainathan, Robert Edelstein, Jesse Rothstein, Jan Brueckner, two referees, and seminar participants at MIT, University of Illinois, and the American Real Estate and Urban Economics Association for helpful comments.

Abstract

This paper addresses the measurement of income sorting across jurisdictions and the attribution of sorting to governmental differences. Measurement error and differences between transitory and permanent income bias variance decompositions sorting estimates downward by approximately 50 percent. Adjusted US Census data show an average across Metropolitan Areas (MSAs) of approximately eight percent of income variation explained by differences across jurisdictions; approximately 28 percent in the Boston MSA. There, politics' role in the sorting process seems small because boundaries between adjacent jurisdictions explain only approximately two percent of income variation.

JEL Classification Codes: D31, R21

1 Introduction

This paper addresses the measurement of income sorting and the attribution of observed sorting to different causes. The causes, extent and consequences of the segregation of demographically heterogeneous populations into relatively homogeneous neighborhoods and jurisdictions are objects of considerable interest among economists and social scientists generally.[1] Segregation on the dimension of income ("income sorting") at the jurisdictional level is particularly interesting because, under some conditions, it is an equilibrium condition in the political economy models of jurisdiction choice that follow from Tiebout [18].[2] Whether local political economy has a large or small effect on housing choice is important for social choices such as school financing mechanisms.

A natural way to estimate the extent of income sorting across jurisdictions is to decompose the variance of income within a region into two components: variance of income within jurisdictions and the variance of mean income across jurisdictions. As the ratio of within jurisdiction variance to variance of mean income across jurisdictions rises, the extent of income sorting falls. Equivalently, the measure of sorting increases with the R^2 arising from a regression of household income on a set of jurisdiction dummy variables.[3]

It is tempting to use such a variance decomposition to infer the extent to which differences in local government affect the distribution of household incomes across jurisdictions. That is, it is natural to interpret a large ratio of variance of mean income across jurisdictions to within jurisdiction income variance as indicative of a strong Tiebout mechanism drawing high income households to jurisdictions with high levels of locally provided public goods.

There are two problems with this approach. First, as demonstrated below, if income is measured with error, then the true degree of income sorting will be understated by a variance decomposition. Second, as acknowledged by Epple and Sieg [8], jurisdictions are

[1]See, for example, Wilson [20], Benabou [2], Kremer [14], Glaeser and Cutler [10].

[2]Examples include: Epple et al. [7] Fernandez and Rogerson [9] and Westhoff [19].

[3]One could imagine measures of income sorting for which this is not necessarily true. For a discussion of measures of group isolation, see Enchenique and Fryer [6]

differentiated by other attributes associated with location, such as amenity and accessibility. A large degree of income sorting across jurisdiction is thus consistent both with (a) a powerful Tiebout mechanism and with (b) a very weak Tiebout mechanism, but a strong relationship between income and willingness to pay for extra-governmental locational amenities.

This paper offers solutions to both of these problems. Section 2 shows that the true R^2, which would be found if income were perfectly measured, is proportional to the observed R^2 times one plus the noise to signal ratio of observed income. Using 2000 US Census data, I find that the adjustment for measurement error and differences between permanent and transitory income increases the estimated extent of sorting by approximately 50 percent.

Substantial income variation within jurisdictions, absent the measurement error correction, has been documented by Ioannides and Seslen [13], among others. Even with the correction, Section 4 reports that the extent of sorting is small. On average, across all US metropolitan areas (MSAs), less than eight percent of the variation in household income levels within MSAs can be explained by differences across jurisdictions. The extent of sorting varies widely across MSAs; the R^2 estimates vary from less than one percent to 33 percent.

In Section 5, I propose a solution to the problem that differences in local taxes and public goods are associated also with differences in extra-governmental locational amenity. The proposed solution is to compare R^2 estimates calculated in two types of subregions. The first type of subregion is a set of two adjacent zip codes (postal service areas) that are in the same jurisdiction. For these subregions, the average R^2 measures the average sorting across zip code boundaries when the zip code boundary lies within a jurisdiction. The second type of subregion is a set of two adjacent zipcodes that are separated by a jurisdiction boundary. For this latter set of subregions, the average R^2 measures average sorting across jurisdiction boundaries when the jurisdictions are next to each other. Suppose that (i) jurisdictions' governmental characteristics are spatially uncorrelated and (ii) zip code boundaries and jurisdiction boundaries are associated with equivalent levels of differentiation of extra-governmental amenity. Then the difference between the average R^2 for the latter type of zip code pair and the average R^2 for the first type of zip code pair provides an estimate of the contribution of governmental differentiation to the sorting process.

In the Boston MSA, I find that the extent of sorting across zip code boundaries within jurisdictions is very close (within two percent) to the extent of sorting across jurisdiction boundaries. This analysis, which borrows from the work of Black [3], suggests that differences in government play only a very small role in the generation of income sorting in the Boston MSA. Given the relative autonomy granted to jurisdictions in Massachusetts, we can expect this result to generalize to other regions.

Realistically, it is highly unlikely that government characteristics would be uncorrelated with location, as can be seen on inspection of maps of crime rates or test scores. As long as differences in location generate any differences in income across jurisdictions, if differences in income are associated with differences in preferences over governmental behavior, then differences in government will become correlated with location, even if preferences over government do not enter into jurisdiction choice. Even if preferences for government taxes and outputs are uncorrelated with income, differences in income based purely on locational characteristics, combined with peer effects and the the institution of property taxes will lead to differences in the quality of government output across locations. These considerations make it very difficult to estimate the role of public goods in housing choice, as has been noted by Bayer et al. [1] and Rothstein [16] among others. For these reasons, the average difference between any governmental characteristics across a jurisdiction boundary is likely to be smaller than the variation of characteristics across the entire MSA. Thus the local comparisons of income sorting are likely to understate, and theoretically could overstate, the true role of government in the income sorting process.

Hence, the fraction of income variation legitimately attributable to differences in governmental characteristics is likely somewhere between the 28 percent of income variation between jurisdictions and the one to two percent identified at the adjacent zip code level. The width of the gap between these two measures highlights the need for further modelling of residential choice in the presence of both local governmental differences and locational differences. Work along these lines has been undertaken by de Bartolome and Ross [5] and by Hanushek and Yilmaz [11]. A well-developed model could estimate what the R^2 distribution of income across jurisdiction boundaries would be if all of the jurisdictions were forced to have the same tax rate and same quality of public goods.[4] The difference between the simulated R^2 and the empirical value could then be attributed to differences in local government.

2 Measuring Income Sorting

A natural way to measure sorting by any characteristic within subregions (jurisdictions or neighborhoods within MSAs, "jurisdictions" hereafter when either can be meant) is to compare the average variance of the characteristic within jurisdictions to the variance at the regional level.

Kremer and Maskin [15] note that such a variance decomposition can be interpreted as the R^2 in a regression of the characteristic on a full set of dummy variables indicating individual residence in each of the jurisdictions. Indexing households by h and jurisdictions by j, and labeling income y, we have:

$$R^2 \equiv 1 - \frac{\sum_{j=1}^{J} \frac{H_j}{H} \sum_{h=1}^{H_j} \frac{(y_h - \bar{y}_j)^2}{H_j}}{\frac{\sum_{h=1}^{H} (y_h - \bar{Y})^2}{H}}, \tag{1}$$

where $\bar{y}_j$ is mean income in jurisdiction j, $\bar{Y}$ is mean income in the MSA, H_j is the number of households living in jurisdiction j, and H is the number of households in the region. The numerator of the second term on the right hand side is the population weighted average of within jurisdiction variance. The denominator is the variance at the MSA level. If jurisdictions are close to homogenous, the fraction is small, and R^2 is large (with a maximum of one). If the average squared difference between households' income within jurisdiction is equal to the average squared difference between households at the regional level then there is no sorting and R^2 is zero. Decomposing total variance, we can also interpret the R^2 measure as the ratio of the population weighted average squared deviations of jurisdiction mean

[4] Presumably the common tax rate and expenditures would affect the level of sorting.

incomes from the population mean divided by total variance. Beyond these observations, interpretation of any measure of sorting requires a model of jurisdiction choice, as in Epple and Sieg [8].

Two defects in the R^2 estimator of sorting must be addressed before taking it to the data. First, a well known problem associated with the unadjusted R^2 measure (1) is that increasing the number of regressors increases the expectation of R^2 in finite samples, even if the added regressors are orthogonal to the dependent variable (here, income). Hence in a world with no behavioral income sorting, MSAs with more jurisdictions would have greater R^2 values mechanically. Other widely used measures such as the Index of Dissimilarity and Thiel's index suffer from the same bias towards observed sorting when jurisdiction sizes are small.

The expectation of variance within a jurisdiction, when households are randomly taken from a sample of the MSA without replacement, is given by:

$$E\frac{1}{H_j-1}\sum_{h=1}^{H_j}(y_h-\bar{y}_j)^2=\frac{1}{H-1}\sum_{h=1}^{H}(y_h-\bar{Y})^2.$$

Thus, replacing H_j with H_j-1 in the numerator and H with $H-1$ in the denominator of equation (1), with random assignment of households to jurisdictions we obtain an expected R^2 of zero. With behavioral sorting, the expectation will be greater than zero.[5] In the data I consider, populations are too large for this adjustment to make a noticeable difference.

2.1 Measurement Error

A second, and more practically important, defect in the the R^2 estimator of income sorting is that it is biased toward zero in the presence of measurement error. Suppose that reported

[5] In general, adding more jurisdictions, or equalizing the population share of jurisdictions allows for a smaller value of adjusted R^2. This only affects the expectation if behavioral sorting occurs. A finding that adding jurisdictions yields larger estimated adjusted R^2's means only that there is sorting, not necessarily that sorting behavior is more pervasive in more fragmented regions.

income is $y+v$, where v is mean zero and i.i.d. across households with variance σ_v^2. Putting aside the small denominator adjustment discussed above, our estimate of R^2 becomes

$$1 - \frac{\sum_{j=1}^{J} \frac{H_j}{H} \sum_{h=1}^{H_j} \frac{(y_h - \bar{y}_j + v_h)^2}{H_j}}{\frac{\sum_{h=1}^{H} (y_h - \bar{Y} + v_h)^2}{H}}. \tag{2}$$

In expectation, randomness of v across individuals and jurisdictions gives us:

$$\text{plim}(R_{me}^2) = \frac{\sigma_y^2 - E(y_h - \bar{y}_j)^2}{\sigma_y^2 + \sigma_v^2} = R^2 \frac{\sigma_y^2}{\sigma_y^2 + \sigma_v^2}, \tag{3}$$

Where R_{me}^2 (R^2) denotes R^2 with (without) measurement error and σ_y^2 is the variance of true income in the population. Hence as the signal to noise ratio of income approaches zero so does the estimate of income sorting, regardless of the true level.

Measurement error v can come from several sources in cross sectional survey data. First, we are typically interested in a measure of sorting by permanent wealth rather than transitory income, but annual rather than lifetime income is reported in most survey data. This would not be a problem if annual income were simply equal to a constant fraction of lifetime wealth. However, this relationship is violated both by year-specific shocks to income and by a generally upward trending age-earnings profile. Unfortunately, widely available Census data includes only aggregate counts of income within jurisdictions and does not include covariates at that geographic level of detail. A second source of error is that households may misreport their earned income in the survey year. A third problem is that income is reported in bins in the data I use.

To estimate σ_y^2 and σ_v^2 separately, we recall the formula for attenuation bias in a regression where a single right hand side variable is measured with error. If we regress some variable Z on reported income $\tilde{y}$, a noisy measure of true income y,

$$Z_h = a + b(y_h + v_h) + \epsilon_h, \tag{4}$$

then we have

$$\text{plim}(\hat{b}_{OLS}) = \frac{Cov(y, Z)}{\sigma_y^2 + \sigma_v^2}.$$

By contrast, if we find an instrument that is correlated with y, but not with v or ϵ, then the two stage least squares estimator $\hat{b_{IV}}$ has the true coefficient on income as a probability limit:

$$\text{plim}(\hat{b_{IV}}) = b = \frac{Cov(y, Z)}{\sigma_y^2}.$$

Comparing the OLS and IV estimators yields the relationship

$$\text{plim}(\frac{\hat{b_{IV}}}{\hat{b_{OLS}}}) = \frac{\sigma_y^2 + \sigma_v^2}{\sigma_y^2}. \tag{5}$$

Multiplying observed R^2 in the jurisdictional analysis by the ratio (5) undoes the bias due to measurement error, pursuant to equation (3). This ratio is equal to one plus the "noise to signal" ratio.

3 Data

I estimate the extent of income sorting using 2000 US Census (SF 3) data on the distribution of household incomes at the MSA and jurisdiction levels within 279 US MSAs. For each of these geographic entities, I observe the estimated number of households with 1999 income in each of 17 income ranges. I assume that all households deemed to have income in any income bin reported the midpoint income of the bin. For example, I consider every household in the income category of \$10,000 to \$15,000 to have an income of \$12,500. To topcoded households, I impute the jurisdiction average income for households in that category.

I follow Census Bureau definitions of metropolitan areas (MSAs) and jurisdictions. MSAs are physically continuous regions that arguably compose a single employment market.[6] I define jurisdictions as census-defined "county subdivisions." The census defines other levels of aggregation (such as places or census tracts) smaller than MSAs, but county subdivisions appear best to match the boundaries of political units that have control over schools and

[6] I define MSAs as consolidated MSAs when this is a choice.

police, to the extent that these are controlled below the county level.[7]

Computing the uncorrected R^2 estimate based on the SF 3 data for each MSA is straightforward. I address the measurement error problem in four steps. First, I regress a dependent variable, discussed below, on income and record the OLS coefficient. Second, I repeat the regression, using an instrumental variable for income in a two stage least squares regression and record the IV coefficient. Third, I use equation (5) along with the OLS and IV coefficients to estimate a signal to noise ratio. Fourth, I multiply observed R^2 by the signal to noise ratio to obtain corrected R^2 estimates.

The dependent variables, income, and the instrumental variables for the regressions that estimate measurement error are taken from the Census's 2000 IPUMS microdata file. For confidentiality purposes, this data does not report the respondent's jurisdiction, but does report Public Use Micro Areas (PUMAs), which are sub-MSA regions meant to capture housing markets. The IPUMS data reports household income, education, occupation and housing characteristics.[8] This microdata includes a bounded integer value for income, which I transform into the midpoint of the corresponding bin that would be reported in the geography-specific SF 3 data (so that a household reporting income of \$12,300 is assigned the \$12,500 midpoint of the \$10,000 to \$14,999 bin).

To estimate a range of signal to noise ratios, I regress three different dependent variables on reported household income to obtain the OLS and IV estimates in steps one and two. The dependent variables are (1) the number of rooms in the household's home, (2) the value of the household's home if they are homeowners and (3) the monthly rent paid by households if they are renters. The IV estimates are obtained by instrumenting for the transformed

[7]One might argue that within-jurisdiction school attendance zones are the relevant level of analysis. However, Tieboutian sorting involves voting, which occurs at the jurisdiction level and many decisions are undertaken at this higher level rather than the school level. Hoxby [12] considers variation in both the number of districts and the number of schools, with identification centering on districts. Some households living within an MSA do not live within any county subdivision. These residents are excluded from the empirical analysis.

[8]Not available for 2000 at the time of writing.

income variable with one of two instruments. The first instrument is the mean income for the occupation and MSA cell in which the household head works (so that all households in the Alleghany, PA PUMA headed by a real estate broker receive the same aggregated instrument). The second instrument is years of education. I perform these regressions in both logs and levels, separately in each of the 200 largest US PUMAs. The assumption required for identification is that local occupation mean income and education affect housing purchases only through household income and not through any unobserved variables included in the error term ϵ in equation (4).

This identifying assumption is not obviously tenable for either education or occupation mean income. Occupation mean income should be uncorrelated with individual shocks to 1989 income through measurement error or age-earnings profile effects to the extent that there is not age-based selection into occupations. However, macroeconomic shock are likely to survive to the occupation level and hence the instrument may not be purged of transitory variation. Hence we expect the IV estimate to be biased away from the true coefficient towards the OLS estimate. Considering education as an instrument, it is possible that education affects housing demand not just through income level, but also perhaps through a correlation with investment demand or access to capital. By using two different instruments that are plausibly biased in opposite directions and by comparing results with other estimates of the magnitude of measurement error, we can at least get a sense of the likely range for the noise to signal ratio that drives the R^2 correction.[9]

4 MSA-Level Sorting and Measurement Error Results

This section documents results from implementing the methodology described in Section 2, using the data described in Section 3.

I find considerable measurement error in household income, as indicated by comparing

[9]Using the method of moments approach proposed by Cragg [4] yields unreliable results with noise to signal ratios on the high end of the IV estimates.

OLS to IV regressions with income alone on the right hand side.[10] Table 1 shows a range of estimates nor far from those summarized in Solon [17]. Column (1) of Table 1 reports results from regressions of the form

$$x = a + by + \epsilon,$$

where x is the number of rooms, the value of the home or the monthly rent. In each case, the coefficient estimate is the population-weighted mean ratio across the 279 MSAs. Column (2) reports an IV estimate of the same regression and column (3) reports the implied noise-to-signal ratio. I perform these regressions in both levels and logs.

Examining the results of Table 1, we see different noise to signal ratios depending on the right hand side variable, and a larger ratio when income is instrumented by education rather than by occupational mean income. Based on the discussion above, these results suggest that either short term macroeconomic shocks persist when occupational mean is the instrument or that the education instrument is not valid. Using rent and price as dependent variables generates larger estimates of attenuation bias than does number of rooms. This may reflect misspecification of the rooms regression, in that rooms may not have the same price throughout a PUMA; the single price assumption would be more likely to apply to measures of hedonic bundles of housing, which rent and value plausibly represents. Nevertheless, all specifications exhibit considerable attenuation bias due to measurement error in income. Based on these results I proceed with the analysis using the modal and median noise-to-signal ratio of 1.0 in both the level and natural log of income.

With no correction for measurement error, across 279 US MSAs, at the jurisdiction level I find a mean R^2 for the level of income of 0.039. The unadjusted R^2 value is 0.051 in logs. Assuming a noise to signal ratio of 1.0 for the level and log of income R^2, the mean R^2 estimates increase to .077 in levels and .102 in log income at the jurisdiction level. While the extent of sorting is generally small, it is always significantly different from zero. In all

[10] "Measurement error" is broadly defined to incorporate differences between permanent and transitory income.

Table 1: OLS and IV estimates of coefficients on income

			(1)	(2)	(3)
RHS Variable	Instrument	Functional Form	OLS	IV	Implied noise to signal
Rent	Education	levels	.003	.010	2.5
			(.0001)	(0.0142)	
Rent	Occupation Mean	levels	.003	.006	1.1
				(.0019)	
Value	Education	levels	.937	1.889	1.0
			(.333)	(.755)	
Value	Occupation Mean	levels	.937	1.456	.6
				(.476)	
Rooms	Education	levels	1.25×10^{-5}	2.48×10^{-5}	1.0
			(2.45×10^{-6})	(5.85×10^{-5})	
Rooms	Occupation Mean	levels	1.25×10^{-5}	1.96×10^{-5}	.6
				(4.71×10^{-6})	
Rent	Education	logs	.170	.421	1.5
			(.058)	(.156)	
Rent	Occupation Mean	logs	.170	.278	.7
				(.117)	
Value	Education	logs	.230	.586	1.8
			(.083)	(.240)	
Value	Occupation Mean	logs	.230	.340	.7
				(.144)	
Rooms	Education	logs	.095	.164	1.0
			(.030)	(.096)	
Rooms	Occupation Mean	logs	.095	.139	.6
				(.051)	

Notes: Each row reports a regression of the form $X = a + by + v$. The OLS column reports the OLS coefficient on income y in a regression where X is the RHS variable. The IV column reports the estimated coefficient on income when income is instrumented for by the variable listed in the Instrument column. The estimated signal to noise follows from the comparison of the two coefficient estimates. Regression coefficients are MSA population-weighted means from individual level regressions for 289 MSAs in 1989 IPUMS Census data. The standard deviation in parentheses under each estimate reflects not the standard error of the single estimated coefficient, which we expect to vary by MSA, but rather the standard error of the unweighted estimate across MSAs.

279 MSAs, an F-test rejects the null hypothesis that the jurisdictions exert no effect on household incomes.

The empirically small R^2 values imply that jurisdictions are quite heterogeneous with respect to household income, but do not necessarily imply that households do not try to sort themselves by income into jurisdictions based on differences in government. Rather, physical constraints may impede such sorting. The measured extent of income sorting may be small in part because a large share of population located in some jurisdictions implies that there must be some mixing of income. For example, a metropolitan area with just two jurisdictions cannot feature an R^2 value of one as long as there are more than two income categories with positive population. Large cities such as New York City are typically larger than the population in any single income category, and hence their associated metropolitan area must feature some income mixing.[11]

Figure 1 plots two values of R^2 for each metropolitan area (for income levels). The vertical axis measures the observed extent of sorting by jurisdiction, corrected for measurement error by multiplying the observed values by two (one plus the noise to signal ratio estimate of 1.0). The horizontal axis is income sorting that would occur if "perfect" sorting were accomplished by locating the highest income households in the smallest jurisdictions and the poorest households in the largest jurisdictions, subject to the existing distribution of numbers of households across jurisdictions. If sorting were complete, the data would lie along the 45-degree line. The figure shows that while sorting exists, so that R^2 increases with the opportunity for larger values, sorting is highly imperfect. We also see that there is a considerable range of income sorting across MSAs. However, the corrected R^2 for the level of income varies from a minimum of .0002 in Lubbock, TX to a maximum of .33 in Columbus, OH. Other highly income segregated MSAs include Philadelphia, PA (.28); St. Louis, MO (.28); Flint, MI (.27); Milwaukee, WI (.27); Cleveland, OH (.27) and New York, NY(.27). For log income, the similarly corrected range is from .0008 in Great Falls, MT to .4459 in Tallahassee, FL.

[11]Eliminating the largest city in an MSA has a negligible effect on estimated sorting for larger MSAs.

The corrections for measurement error and for feasible sorting reveal that, on average, 9.1 percent of feasible sorting is attained. That is, the average corrected R^2 divided by the maximized R^2 on the vertical axis of Figure 1 is equal to 0.091.

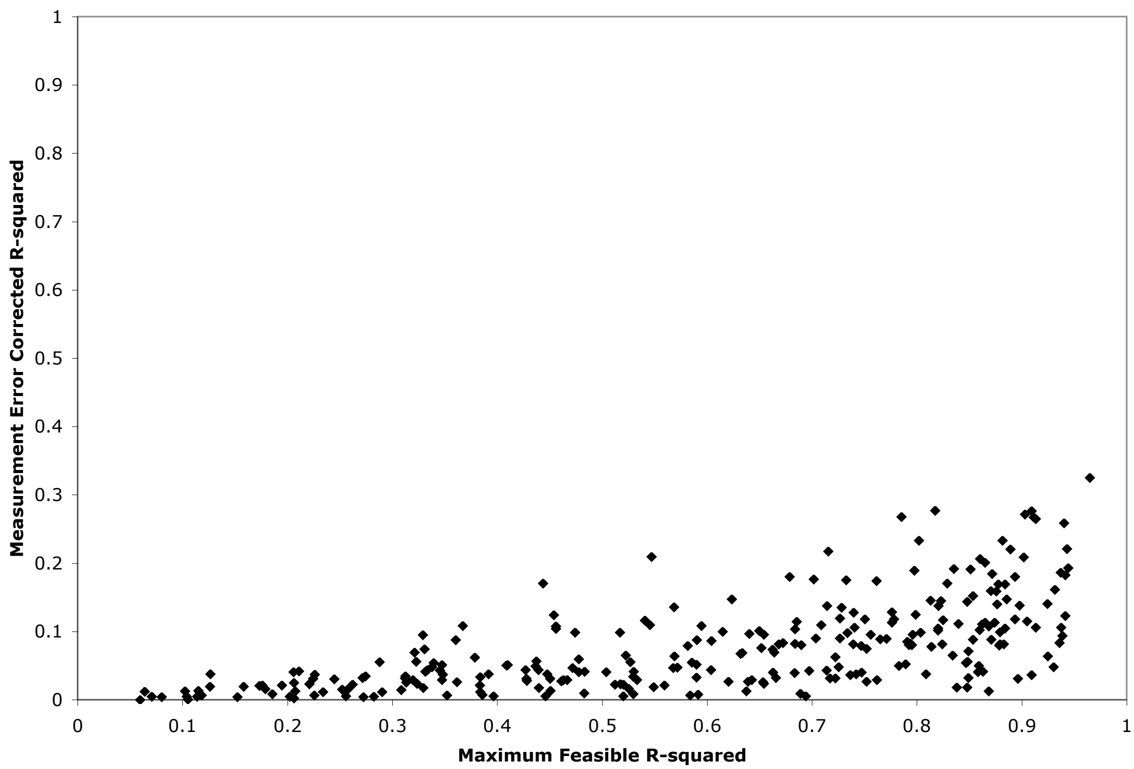
5 Decomposing Income Sorting

A correction of the R^2 estimate of the extent of sorting is necessitated by measurement error. However, the corrected income sorting measure likely overstates the role of differences across local governments in the distribution of household income across jurisdictions within MSAs. The likely overstatement arises because local governments are differentiated both by governmental characteristics and by locational amenity which may have nothing to do with Tieboutian sorting. By examining the extent of income sorting while controlling for locational differences, we may be able to estimate the role of local government in the sorting process. At least, we should get an idea of the range of how much income sorting can be explained by differences in local governments.

The state of Massachusetts grants considerable power over local public goods such as schools, fire and police to jurisdictions,[12] making Massachusetts a popular subject for analysis of jurisdiction choice. If differences across local governments should drive income sorting anywhere, it should do so in Massachusetts. As we might expect, and supportive of the idea that jurisdictions drive sorting, the Boston MSA is more income sorted than average. For the 212 jurisdictions within both the state of Massachusetts and the Boston CMSA, I estimate an R^2 in the level of income as described above of .14, uncorrected for measurement error, which corresponds to a corrected value of approximately .28, assuming a noise-to-signal ratio of 1.0. This makes Boston the ninth most sorted of US MSAs. For log income the unadjusted figure is .10, which would be adjusted to .20 for measurement error, ranking 32nd among MSAs. The uncorrected variance decomposition results are consistent with those of Epple and Sieg [8], who find an R^2 of .11 using Boston MSA data from the 1980 census. However,

[12] Subject to some limitations on the level of and changes to property taxes.

Figure 1: Income Sorting: Observed corrected R^2 values and maximized values with "perfect sorting"



Notes: Each observation is drawn from a different US MSA. Observed corrected R^2 is the weighted average ratio of income variance within jurisdictions to MSA level income variance, modified to correct for measurement error. The correction is to multiply the observed R^2 by one plus a noise to signal ratio of one (based on estimates presented in Table 1). Maximum feasible R^2 is estimated by assigning census households into maximally homogenous jurisdictions, as described in the text.

without further analysis we cannot determine to what extent this income sorting is driven by differences in tax and expenditure characteristics across jurisdictions as opposed to different locational characteristics, such as different housing stock and access to regional amenity that are not directly related to local government policy.

To disentangle these possibilities, I estimate R^2 sorting measures for a large number of subregions within the Boston MSA. Each subregion is composed of a pair of adjacent neighborhoods, and the R^2 measures the fraction of the variance of income for the population of the subregion that can be explained by living on one side or the other of the boundary between the neighborhoods. Each neighborhood lies strictly within the boundaries of a single jurisdiction. The neighborhoods I use for these purposes are Census-defined "5 digit Zip Code Tabulation Areas," (ZCTAs) These roughly capture carrier routes for the US postal service. Some jurisdictions have multiple ZCTAs and others have only one. Thus it is possible for two neighboring ZCTAs to be in the same jurisdiction or to be in different jurisdictions.

By comparing the extent of income sorting in "regions" composed of two adjacent neighborhoods when the neighborhood pairs are (a) in the same jurisdiction and (b) in neighboring jurisdictions, we gain insight into the role of local government in the sorting process. Under the assumption that a jurisdiction boundary involves no less separation of amenity and housing conditions than a boundary drawn for postal delivery convenience within a jurisdiction, and the admittedly unrealistic assumption that tax and expenditure policies are randomly distributed across locations within an MSA, the difference between sorting measures (a) and (b) can be interpreted as an upper bound on the sorting directly attributable to governmental differences. A trivial further restriction is that postal service does not affect residential choice. Realistically, underlying demand for amenity, lot size and high quality public goods are likely to be correlated. Recognizing this, a more modest use of the R^2 arising from comparisons of types (a) and (b) is to obtain a sort of lower bound on the role of government in the sorting process.[13]

[13]Theoretically, if for some reason local public goods were negatively correlated with provision in neighboring jurisdictions, this exercise could overstate the role of governmental differences. This seems implausible.

For each 5-digit Zip Code Tabulation Area (ZCTA) that lies in both the Boston MSA and the state of Massachusetts, I identify the closest neighboring ZCTA, based on centroid-to-centroid Euclidean distance. I then estimate the R^2 value for the fraction of log income variance in the joint population of these neighboring ZCTAs that is explained by being in one or the other ZCTA. Next, I compare the R^2 values for ZCTA pairs that lie in different jurisdictions to the R^2 values for ZCTA pairs that lie on opposite sides of a jurisdiction boundary. Because all ZCTAs lie within a single jurisdiction, those are the only kinds of ZCTA boundaries. The nearest neighbor relationship need not be symmetric; when it is symmetric, I discard one of the two repeated observations.[14]

Turning to estimation of the effect of being in adjacent "neighborhoods" but different jurisdictions, I find that jurisdiction boundaries do, in fact, matter. Of 352 ZCTA pair R^2 estimates for log income, 192 are within jurisdiction comparisons, and they have a mean of .016. The remaining 160 variance decompositions are across jurisdiction boundaries, and have a mean of .030. The difference of 1.4 percent (2.8 percent corrected for measurement error) is significantly different from zero, as reported in the first column of Table 2.

One might be concerned that within jurisdiction zip code boundaries tend to be in large jurisdictions, so that the differences in geography represented by zip code boundaries are not comparable to the differences in geography represented by jurisdiction boundaries. The second column of Table 2 estimates the effect of zip code pairs being in different jurisdictions conditional on some characteristics of the ZCTAs. The column (2) regressions are of the form:

$$R_{ij}^2 = a + b * DIFJUR_{ij} + \gamma_1 * DISTANCE_{ij} + \gamma_2 VAR_{ij} + MEAN_{ij} + u_{ij}, \qquad (6)$$

where $DIFJUR_{ij}$ indicates whether ZCTAs i and j are in different jurisdictions, $DISTANCE$ is Euclidean distance (in degrees divided by 1 million), VAR_{ij} is the variance of income for the subregion formed by ZCTAs i and j, and $MEAN_{ij}$ is mean income in the subregion.

[14] In the comparisons below, standard errors are clustered on the ZCTA in the pair with the lower identification number. The fact that each ZCTA is likely to be paired with other ZCTAs with both higher and lower identification numbers implies that standard errors are likely understated.

$DIFJUR$ should be associated with differentiation in government policy, since only pairs of ZCTAs within the same jurisdiction share the same government (although there may be differences in average government service quality between neighborhoods). It is the effect on sorting of these policy differences that we wish to estimate through the coefficient b. The important question for identification is whether $DIFJUR$ is also correlated with the error term u through unobserved differences in conditions across jurisdiction boundaries that do not exist across within-jurisdiction ZCTA boundaries and are not directly attributable to government policy (or through differences that exist across random boundaries but not across jurisdictional boundaries). Such differences might include differences in housing quality (to the extent that these differences are not driven by zoning), topography, or access to regional amenities. These unobserved differences seem no less likely to occur across jurisdiction boundaries than postal route boundaries, so the coefficient b is plausibly an upper bound on the reduced form effect of local governments boundaries on sorting.[15]

Table 2 describes the results of the regression (6). Including covariates significantly reduces the estimated effect of the added sorting associated with being in a different jurisdiction in the case of log income. However, the estimated effect remains significant at just under one percent in log income, uncorrected for measurement error. In the presence of controls and correcting for measurement error, the results are virtually identical for the level of income. Jurisdiction boundaries explain an additional 1.5 percent of income variation on top of that

[15] There is some reason to be concerned that ZCTA boundaries might be associated with a large degree of amenity differentiation. The ZCTA boundaries are drawn to follow Census-defined block groups, so a number of block groups lie within each ZCTA. The Census defines block groups so that their populations are relatively homogenous. Hence any division of collections of block groups is likely to reflect a relatively significant break in amenity (this problem shrinks when the number of component block groups is large). It is hard to know whether such boundaries are likely to represent greater breaks in amenity than older jurisdiction boundaries. When neighboring zip codes are replaced by more randomly differentiated within-jurisdiction populations, the difference between the R^2 estimates of types (a) and (b) grows by approximately one percent. The results are unchanged if multiple R^2s are calculated based on multiple neighbors for each ZCTA.

explained by within-jurisdiction ZCTA boundaries.

6 Conclusions

Jurisdictions are segregated by income relative to metropolitan areas. Correcting for measurement error in income increases the estimated extent of sorting, but sorting remains far from complete. In the Boston MSA, where jurisdictions have considerable authority, they explain approximately 28 percent of variation in household lifetime income. Observed sorting at the jurisdiction level may be generated by differences in tax and spending policies and the related quality of public goods such as schools, by differences in extra-governmental amenity and housing quality or, most likely, by a combination of these factors. The empirical results suggest that housing quality and extra-governmental amenity play large roles in the sorting process. Jurisdictional differences account for only approximately two percent of the variation in household income when the population is drawn from adjacent neighborhoods which presumably share many extra-governmental characteristics. Given these results, future development of the theory of local political economy should be embedded in a setting in which households' location choice is affected by partly exogenous, spatially correlated variation in housing quality.

Table 2: Regressions of Income R^2 for Neighboring Zip Code Tabulation Areas on a Variable Indicating Different Jurisdictions, Boston MSA

	(1)	(2)	(3)	(4)
DIFJUR	.004/ .008	.008** / .016	.014** / .028	.009** / .018
	(0.002)	(0.0032)	(.002)	(.003)
$DISTANCE/10^{13}$		2.04**		-2.87**
		(.548)		(.805)
$MEAN$		$5.82/10^7$**		.008
		$(2.05/10^7)$		(.007)
VAR		$2.24/10^{12}$		.062**
		$(3.31/10^{12})$		(.016)
Constant	.002**	.045**	.030**	-.107
	(.002)	(.0022)	(.009)	(.075)
Income Measured in	Level	Level	ln	ln
Regression R^2	.01	.06	.04	.14
Observations	358	358	358	358

Notes: ** Denotes significance at 5 percent, * at 1 percent. Dependent variable is one minus the population-weighted average ratio of within ZCTA income variance to total variance in a region composed of two adjacent ZCTAs. DIFJUR indicates whether the two ZCTA are in different jurisdictions. The first coefficient on DIFJUR is the coefficient uncorrected for measurement error, the second is corrected for measurement error assuming a .5 noise to signal ratio. All other coefficients and all standard errors do not correct for measurement error. $DISTANCE$ is the Euclidean distance (in millionths of degrees) between the centroids of each ZCTA.

References

[1] P. Bayer, R. McMillan, and K. Rueben. An equilibrium model of sorting in an urban housing market: A study of the causes and consequences of residential segregation. mimeo, Yale University, 2002.

[2] R. Benabou, Workings of a city: Location, education, and production, *Quarterly Journal of Economics*, 108(3):619–652, August 1993.

[3] S. Black, Do better schools matter? parental valuation of elementary education, *Quarterly Journal of Economics*, pages 577–599, May 1999.

[4] J. G. Cragg, Using higher moments to estimate the simple errors-in-variables model, *Rand Journal of Economics*, 28(0):S71–S91, 1997.

[5] C. de Bartolome and S. L. Ross, Equilibria with local governments and commuting: income sorting vs income mixing, *Journal of Urban Economics*, 54(1):1–20, 2003.

[6] F. Enchenique and R. Fryer. On the measurement of segregation. Working Paper 11258, NBER, 2005.

[7] D. Epple, R. Filimon, and T. Romer, Equilibrium among local jurisdictions: Toward an integrated treatment of voting and residential choice, *Journal of Public Economics*, 24: 281–308, 1984.

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

[9] R. Fernandez and R. Rogerson, Income distribution, communities, and the quality of public education, *Quarterly Journal of Economics*, 11(1):135–164, February 1996.

[10] E. Glaeser and D. Cutler, Are ghettos good or bad?, *Quarterly Journal of Economics*, 112:827–872, 1997.

[11] E. Hanushek and K. Yilmaz. Can money buy a better education? working paper, Hoover Institution, 2005.

[12] C. M. Hoxby. Does competition among public schools benefit students and taxpayers? Technical report, NBER, December 1994. W4979.

[13] Y. Ioannides and T. Seslen. Neighborhood income and wealth distributions. Unpublished, Tufts University, 1999.

[14] M. Kremer, How much does sorting increase inequality?, *The Quarterly Journal of Economics*, pages 115–139, February 1997.

[15] M. Kremer and E. Maskin. Wage inequality and segregation by skill. 1996.

[16] J. Rothstein. What do parents choose? manuscript, UC Berkeley, 2003.

[17] G. Solon, Biases in the estimation of intergenerational earnings correlations, *Review of Economics and Statistics*, 71:172–174, 1989.

[18] C. M. Tiebout, A pure theory of local expenditures, *Journal of Political Economy*, 64 (4):442–444, October 1956.

[19] F. Westhoff, Existence of equilibria in economies with a local public good, *Journal of Economic Theory*, 14:84–112, 1977.

[20] W. J. Wilson. *The Truly Disadvantaged: The Inner City, the Underclass and Public Policy*. The University of Chicago Press, 1987.