Spatial Cost of Living Indices and the Distribution of Public Goods

Timothy L. Hamilton[ˆ]

University of Richmond

October 14, 2015

ˆ Address for correspondence. Prof. Timothy L. Hamilton, Department of Economics, University of Richmond, VA 23173. Email: thamilt2@richmond.edu

ABSTRACT

This paper develops cost of living (COL) indices that vary across space. While conventional indices adjust for differences in prices, the cost of living defined here also reflects access to public goods. This analysis relies on the structure of a residential sorting model to estimate the COL index for each of 226 Metropolitan Statistical Areas (MSAs) in 1990 and 2000 in the United States. Empirical results show significant differences in the cost of living across the spatial landscape. This paper focuses on the Gini coefficient as a measure of inequality to demonstrate the distribution of public goods across the population.

1 Introduction

Valuation of nonmarket amenities has been extensively researched due to the role of public goods in welfare analysis and public policy. However, emphasis is generally placed on efficient levels, while considerations of the distribution of public goods across the population often receive less attention. Parallel to this issue is the fact that conventional economic analyses of distribution use only income as the measured variable. While income can be adjusted to reflect market prices, such a measure still overlooks other elements of welfare, namely public goods. This elicits the need for a broader measure of welfare in distributional analyses. To overcome these two concerns, this research seeks to construct spatially defined cost of living (COL) indices to determine a measure of welfare that reflects income, market prices, and public goods. In a further analysis of COL indices, I examine the distribution of such welfare across the population.

I employ a residential sorting model, an extension of the random utility model (RUM), in which individuals choose their residential location based on site specific amenities and labor market conditions. Optimal decisions suggest values for public amenities associated with each location. Estimation of consumer preferences then allows for calculation of the true cost of living, which reflects consumption of both market and non-market goods. Thus, COL indices are constructed from duality theory, rather than from aggregate price indices that are often inherently biased. In addition, the sorting approach allows for simple empirical identification of both observed and unobserved location amenities.

Throughout the analysis, moving costs play a considerable role. Individuals face monetary and psychological costs of obtaining various bundles of goods across the landscape. Mechanically, moving costs prevent full arbitrage of location amenities into housing and labor markets. Thus, such costs are important to properly valuing location amenities (Bayer et al. (2009)). In addition, they act to preserve significant differences in the cost of living between locations, sustaining an equilibrium that is different from that often characterized in the literature.

With no mention of public goods, income inequality has gained considerable attention in recent years and has amassed a significant presence in policy discussions and in the academic literature. Piketty and Saez (2003) examines dynamics in the income distribution over the course of the twentieth century in the United States. In particular, the authors document a growing level of income inequality over the latter half of the twentieth century, as evidenced by a higher concentration of income in the top income groups. Additional evidence of income inequality is offered by Heathcote et al. (2010), which shows increasing wage inequality from 1967-2006. Both of these studies also make note that income taxation in the United States has worked to somewhat alleviate the trend of rising inequality. Building on the notion of post-tax income as more representative of an individuals's standard of living, Armour et al. (2013) incorporates unrealized gains from asset appreciation into a more extensive measure of income, and finds that inequality has not increased to the degree suggested by the distribution of only realized income.

Given the growing literature on income inequality, the discussion may benefit from a better definition of the unequally distributed asset. Public policy may be less concerned with the distribution of income than with the distribution of a broader measure of the standard of living. Moretti (2013) addresses this issue of measuring a broader measure of the standard of living by deflating nominal wages to reflect the cost of housing. While nominal wage inequality is evident from 1980-2000, real wage inequality is less drastic due to the fact that cities with higher wage earners have also experienced a larger increase in housing prices. Thus increasing market prices counter the trend in nominal wages. Coupled with evidence which suggests that higher housing prices are the result of increases in labor demand, rather than labor supply shocks, the distribution of the standard of living appears to exhibit a decrease in inequality over the time period of the study. While Moretti (2013) illustrates changes in the distribution of the standard of living, the existence of utility-generating public goods precludes it from identifying the actual level of inequality in the standard of living.

A progression of interest in income inequality to interest in inequality of a more general measure of the standard of living serves as this paper's primary empirical motivation. Distributional issues relate to the fairness of the public provision of goods, in particular whether public amenities disproportionately benefit different income groups. This is an obvious social concern that requires empirical evidence. Development of true COL indices allow for calculation of an adjusted income measure and precise characterization of the dispersion of welfare. This paper will develop spatially explicit cost of living indices that demonstrate the degree to which access to public goods exacerbates inequality in a consistent measure of the standard of living in the United States.

2 Quality Adjusted Cost of Living Indices

A cost of living (COL) index defines the relative cost of reaching a particular level of well-being. The presence of different prices makes it more or less expensive to achieve some level of utility at different times or in different locations. A theoretically consistent means of deriving a true COL index is to take the ratio of two expenditure functions, in which the expenditure function gives the amount of income necessary to reach a base level of utility. Given accurate measures of the cost of living, it is possible to move beyond simple income measures and conduct a more complete examination of a population's well-being. Applications of COL indices may include analysis of welfare effects over time as income changes at a rate different than that of the cost of living or across locations as the cost of living may vary spatially. Conventional COL indices focus on measures that use market prices and market expenditures. However, public goods may contribute a significant amount to an individual's well-being. Of course, the non-market nature of public goods makes it difficult to observe prices and quantities of goods or even total expenditures. In the following, I briefly review traditional COL indices and discuss attempts to create COL indices and well-being measures that incorporate public goods.

Two common approximations of the true COL index are the Laspeyres index and the Paasche index. Both of these measures calculate the ratio of price times quantity across a number of goods in two regions with different prices, and assume that the same fixed basket of goods is purchased in both regions. A fixed basket of goods is necessary to make a valid comparison. Summing total expenditures is similar to calculating an expenditure function, but the requirement of a fixed basket of goods does not capture optimal behavior on the part of the consumer. The indices differ in the region that defines the fixed basket of consumption, and thus the region that defines base utility. This assumption creates the problem of no substitution of goods in the presence of different prices, and gives a biased estimate of the true cost of living. Since the reference level of utility differs, the bias moves in opposite directions between the two indices. To correct for the bias, Fisher (1922) introduced the Fisher ideal index, which takes the geometric mean of the Laspeyres and Paasche index. Konus (1939) shows that under a homogeneous quadratic utility function the Fisher index is the exact true COL. However, rather than imposing this restriction, direct estimation of the expenditure function allows for calculation of the true COL index. Kakwani and Hill (2002) estimate a variety of functional forms for an expenditure function and use parameterized functions to calculate COL indices. In an application to Thailand, the authors develop indices that reflect the spatial distribution of market prices.

Both Nordhaus (1999) and Banzhaf (2005) recognize the role of publicly provided goods, in addition to market prices, in the cost of living. Nordhaus (1999) develops an "augmented" cost of living index that deflates market expenditures to account for consumption of public goods. This index can be derived from a conventional index with a correction based on the level of public good spending relative to total consumption spending. Since public goods are implicitly purchased through higher market prices, the adjustment reflects the true market prices that consumers face. In the empirical application, public good consumption is calculated as the sum of social insurance taxes, employer provided goods, indirect business taxes, and mandated social regulations. This calculation is then used to measure the inflation rate once public good consumption is accounted for. Results show an upward bias in the traditional measure of inflation, change in the Consumer Price Index, of 19% from 1960 to 1997, an average of 0.47 percentage points per year. While the analysis focuses on national accounts rather than any notion of geographic variation in public goods, it highlights the degree to which public goods influence the cost of living.

Banzhaf (2005) provides two approaches to incorporate public goods in the cost of living. The first approach involves estimation of virtual prices of public goods. Combined with changes in public good levels, these prices reflect consumption of the public good and are used to correct price indices based only on market consumption. The second technique is based on the assumption that public goods are weak complements to some market goods (housing in this case). Similar to the approach in Nordhaus (1999), this approach estimates a new price of market consumption that reflects the true cost of achieving a reference level of utility. Public goods of interest include air quality, teacher-student ratio, and public safety in Los Angeles from 1989 to 1994. Results from the first approach show that incorporating public goods into the Bureau of Labor Statistics Los Angeles Consumer Price Index reduces the index by 0.4 and 0.3 percentage points, respectively, in the first two years where LA saw improving air quality and education. The adjusted index in the second approach suggests similar effects, with a reduction of 0.3 and 0.2 percentage points in the first two years, and a total reduction of 0.1 percentage points over the final three years. This study, although focused only a few public goods, further indicates the significant impact of including public goods in COL indices.

A related literature attempts to directly calculate a measure deemed as "quality of life" that reflects public good provision. The empirical approach uses wage and housing price differentials. The intuition is similar, as individuals trade off income and housing expenditures to gain access to public goods in a particular location. Blomquist et al. (1988) estimate wage and housing hedonic equations to attain implicit prices for public goods. Quality of life is then calculated as the sum of the value of all public goods, in which value is simply marginal price times quantity. This study assumes that all individuals have identical preferences and earning potential. Empirical estimates show significant impacts of public goods on housing prices and wages. This framework is also used for quality of life calculations in Albouy (2012), which measures differences across U.S. cities. Kahn (1995) employs a similar hedonic analysis, but only defines quality of life in terms of its rank order, rather than a cardinal measure. Lower wages and higher housing prices imply that a location has a higher quality of life as individuals arbitrage away money for public goods.

Several problems are inherent in the previous papers in determining quality of life. First, wages and income are fully arbitraged. However, in a geographic choice set that includes cities across the United States, there are likely to be significant moving costs. Therefore, the value of a location's public goods may not be fully reflected in wages and housing prices. A second, related matter is that in deriving equilibrium conditions, these studies also assume that each individual achieves the same level of utility. Exogenously determined origin locations and significant moving costs generally prevent such an outcome. By formally modeling moving costs, better estimates of the value of public goods can be obtained. Finally, the notion of quality of life in the previous studies refers to a measure of only the level of public goods. In the current analysis, the focus is on both public goods and consumption of market goods. Thus the measure of well-being developed in this paper will indicate the combined effect of wages, market prices, and public goods.

Following an expenditure function approach, Timmins (2006) uses a residential sorting model to estimate the parameters of a utility function that depends on consumption, housing, location attributes and moving costs. Preferences for such variables vary with individual characteristics. An expenditure function is then derived from the indirect utility function. The ratio of a location and individual type specific expenditure function to an area weighted average expenditure function gives the true COL index for a each of the 495 locations. When these indices are applied to the observed income distribution of Brazil, results suggest a higher degree of inequality. In what follows, I adopt this method of deriving COL indices.[1] Building upon the work of Nordhaus (1999) and Banzhaf (2005) that observed prices are an imperfect signal, this paper focuses on adjustments to real income measures based on spatial variation in the cost of living.

3 Theoretical Framework

As discussed above, COL indices can be derived directly from expenditure functions. While there exist well-behaved expenditure functions that can be directly specified and estimated given proper data, the lack of prices for public goods makes such an approach difficult. Instead, I appeal to duality theory to derive expenditure functions from indirect utility functions. The following sections outline the structural framework from which COL indices are calculated and parameterized utility functions, and thus expenditure functions, are obtained.

3.1 COL Indices

Define a conditional indirect utility function, $V = V(I, \rho, Y)$, where I denotes income, ρ is a set of prices of private goods, and Y is an exogenously determined bundle of attributes. The inclusion of Y is due to the spatial nature of COL indices in this study and its role will become clearer in the next section. An expenditure function is obtained by inverting the indirect utility function to isolate I,

$$I = E(V, \rho, Y), \tag{1}$$

which then defines the amount of income required to reach utility level V, given prices ρ and attributes Y. While expenditure functions are based on arbitrary values of V, much can be learned from making comparisons of $E(\cdot)$ at different levels of ρ and/or Y, while holding V constant. For conventional goods, cost of living (COL) indices are defined as a ratio with numerator equal to the minimum income required for purchase of a particular bundle of goods that will leave utility unchanged and denominator equal to a base income. In the context of a spatial model, COL indices make this comparison across geographic locations and include publicly provided goods as elements of the consumption bundle. Therefore, indices measure the necessary income to reach some utility level in a location relative to the necessary income to reach that same utility level in a base location. Define $E_0(\bar{V})$ as the mean necessary income to reach $\bar{V}$, where the mean is taken across the entire sample of locations. The cost of living in location n is measured relative to the average cost of living,

$$COL_n = \frac{E(\bar{V}, \rho_n, Y_n)}{E_0(\bar{V})}, \tag{2}$$

For example, a COL index of 1.2 suggests that it is 20% more expensive to reach the baseline utility level in location n, relative to the sample average. A consumer requires more income to reach $\bar{V}$ in n due to a different combination of ρ and Y. A COL index of 1 indicates that location n has an average COL, which may result from any number of combinations of ρ and Y.[2]

3.2 Model Parameters

A location sorting model is employed to estimate the necessary parameters of the indirect utility function related to individual preferences. The model discussed below is derived in detail in Hamilton and Phaneuf (2015). For the empirical application, the geographic unit of analysis is m, a metropolitan statistical area (MSA)[3]. Also note the use of k to define an individual *type*, indicating heterogeneous preferences based on observed attributes. The model uses eight different individual types based on educational attainment and the presence of children in the household[4]. Define the indirect utility function for individual i of type k in location m as

$$\ln V_{im} = \beta_I \ln I_{im} + \beta_Y \ln Y_m - MC_{im} - \beta_H \ln \rho_m^Y + \tau^k \Gamma_m^k + \zeta_m + \eta_{im} \tag{3}$$

where I_{im} denotes an individual's income in location m, Y_m is a vector of location-specific amenities and public goods in m, and ρ_m^Y is the price of housing services in m that varies across locations largely due to difference in location attributes Y. Moving costs for individual i moving to location m are indicated here by MC_{im}. Their exact definition is discussed below. Location unobservables are denoted as ζ_m, comprised of any location amenities, such as cultural or historic amenities, that may not be directly measured in Y_m.[5] Individual idiosyncratic shocks are denoted as η_{im}. These shocks include any random and unobserved individual preferences for a specific location, such as a personal connection to some particular MSA.

Location m is treated as a location with a particular level of public goods. Since an MSA can be geographically large, there still exists some variation in amenities within m. Therefore, Y_m denotes the value of MSA averages amenities. In addition to this average level, I develop the term Γ_m^k for each type-MSA combination, which is an aggregate of any variation in amenities that may exist within location m. The Γ_m^k terms are estimates from a model of residential sorting that accounts for sorting across MSAs and across smaller geographic units within MSAs. While cross-MSA sorting captures the value of the average level of amenities in an MSA, within-MSA sorting captures the value of variation in amenities within the MSA, or deviations from the average. Equation 3 thus outlines an indirect utility function that includes utility gained from the average level of location amenities in an MSA, as well as the utility gained from the choice of a smaller geographic location within each MSA. These components are derived and discussed in greater detail in Hamilton (2012) and Hamilton and Phaneuf (2015). Also, Hamilton and Phaneuf (2015) shows that equation (3) is a theoretically consistent approach to estimating differences in the value of amenities across MSAs and can be derived from a primal utility framework that satisfies the typical functional restrictions on preferences. The model is estimated as a conditional logit following Bayer et al. (2009) and Hamilton and Phaneuf (2015), resulting in the necessary parameters for computation of expenditure functions. Finally, note that the Γ_m^k parameters are indexed by k, which denotes a preference type based on attributes of the individual. This is the only portion of the model that explicitly considers preference heterogeneity. Such heterogeneity implies that expenditure functions and COL indices will be type-specific.

The necessary parameters for COL calculation along with assumptions on preference heterogeneity allow for simplification of this equation. First, I assume homogeneous preferences for location amenities and housing services. Also, analysis is not explicitly concerned with preference parameters on specific attributes. Therefore, housing prices and mean preference parameters collapse into a location-specific fixed effect. Estimation then relies on the following equation,

$$\ln V_{im}^1 = \theta_m + \beta_I \ln I_{im} - MC_{im} + \tau^k \Gamma_m^k + \eta_{im}, \tag{4}$$

$$\theta_m = \beta_Y \ln Y_m - \beta_H \ln \rho_m^Y + \zeta_m, \quad m = 1, ..., M. \tag{5}$$

Note that θ_m is simply estimated as a fixed effect, which is all that is necessary for developing expenditure functions.

The variable I_{im} indicates the need for the potential earnings of all individuals in all locations. Recall that each MSA is assumed to be a single labor market. Estimates from a regression of observed income on individual characteristics are used to predict income across locations based on location specific coefficients and an individual's characteristics. A semi-parametric correction for sorting in the labor market is proposed by Dahl (2002) and implemented in a residential sorting model in Bayer et al. (2009). This method rests on the intuition that while unobserved location attributes induce moves to labor markets, information regarding such sorting is captured by the probability of an identical type individual making the observed move.

Estimation of the model also requires ρ_m^Y, the MSA-level price of housing services. I estimate this MSA-specific price using a regression of observed home prices on attributes of the home. The price of a home can be attributed to the structural attributes of the home, such as the number of rooms and lot acreage, and the location of the home. I isolate these two impacts to find the price of a homogeneous unit of housing services in each MSA. See Appendix A for a fuller discussion of estimating potential income and the price housing services.

Finally, moving costs are characterized in a reduced form manner as a modified version of the specification in Bayer et al. (2009). I model moving costs as a discrete cost of leaving one's birth state and birth region. In addition, I allow for variation in moving costs based on the presence of children in a household, an obvious constraint to migration. For individual i in MSA m,

$$MC_{im} = \mu_{bs} D_i^{bs} + \mu_{br} D_i^{br} + \upsilon_{bs}(D_i^{bs} \times D_i^{C}) + \upsilon_{br}(D_i^{br} \times D_i^{C}), \tag{6}$$

where $D_i^{bs} = 1$ if MSA m is outside of individual i's birth state and $D_i^{br} = 1$ if MSA m is outside of individual i's birth region, with estimated coefficients μ_{bs} and μ_{br}. Moving costs should be interpreted generally here, including not simply the expenses of moving goods but also costs related to searching for a new home. In addition, individual's may tend to have preferences for an original location, implying the broadest costs associated with being away from one's birth place. While this specification may miss some of the more direct expenses of moving, such as driving to a city located 100 miles away relative to one located 50 miles away, it does account for moving costs in a broader sense. From the sample data, large percentages of individuals are observed living in birth states (52.05% in 1990 and 50.54% in 2000) and in birth regions (59.75% in 1990 and 57.15% in 2000), suggesting preferences for home locations. Also note that a very general cost of moving may be better estimated based on an individual's previous location rather than her birth location. However, the data are sparse with regard to previous locations. Additional components of equation (6) interact D_i^{bs} and D_i^{br} with a dummy variable indicating an individual with children. Thus $D_i^C = 1$ if there are children under the age of 18 living in the household of individual i.

Equation (3) defines utility for an individual in a specific location, m. Therefore, expenditure functions express the required income for an individual to attain some level of utility in m. The amount of income varies, of course, as housing prices ρ and location attributes Y vary across space. The necessary level of income also varies with preferences since an individual's ability to reach any level of utility will depend on their preferences for different market and non-market goods. This framework implies the overall cost of living of a location may be different for different individual types. For the empirical framework, I define eight unique individual types, as discussed earlier.

Coefficient estimates on income, moving costs, and the index Γ_m^k in equation (4) using maximum likelihood estimation are reported in Table A1. Parameters are highly significant and conform to expected results. Recall that estimation includes alternative specific constants as well. These alternative specific constants allow for measurement of the utility impact of a location's set of public goods. As expected, the level of public goods is positively correlated with the price of housing services. Across the eight preference types, correlation ranges from $0.22 - 0.33$ in 1990 and from $0.20 - 0.30$ in 2000. This correlation offers evidence for the capitalization of public goods into housing prices. Locations with a high price of housing tend to have highly valued public goods, indicating the importance of including public goods in addition to market prices in cost of living indices.

Given a fully parameterized model, the indirect utility function implicitly defines an expenditure function specific to each location and individual type k. The expenditure function for individual i in MSA m is easily derived from the indirect utility function in (3) by fixing utility at level $\bar{V}$ and solving for income, I_{im}, resulting in

$$I_{im} = \exp\left[\frac{1}{\beta_I}\left(\ln \bar{V} - \theta_m^k - MC_{im} - \tau^k \Gamma_m^k - \eta_{im}\right)\right], \tag{7}$$

where θ_m^k combines observed and unobserved location amenities, but is estimated as a type-location fixed-effect. This technique displays an additional benefit of the sorting model approach relative to the hedonic approach, in that location unobservables are easily controlled for. Equation (7) defines an individual-specific expenditure function that will be further discussed in subsequent sections.

3.3 Moving Costs and Cost of Living

An important advantage of sorting models relative to housing and wage hedonic functions is their ability to account for the large spatial dimension of the choice set. The present model includes MC_{im}, a variable that indicates the geographic distance between each choice alternative and the individual's starting location. Such a specification allows for a cost of moving, rather than assuming a simple choice based only on the attributes of an alternative. Bayer et al. (2009) demonstrate the implications for calculating the utility value derived from different site attributes using a hedonic framework that ignores any cost to moving. In a hedonic model, the value of site amenities is fully capitalized into housing costs and wages, and the amenity value is equal to the sum of the compensating differentials from housing and income. However, if positive moving costs exist, there is a third component contributing to the amenity value that captures the variation in the cost of moving to each location. The effect of housing and wages alone will undervalue site amenities as they ignore the additional cost of moving to a location to obtain the amenity.

In calculating adjusted income, the importance of accounting for moving costs to obtain proper values for non-market goods is obvious. However, the existence of moving costs between alternatives raises another issue in determining the cost of living. In particular, while moving costs are imperative for accurately calculating amenity values, such costs should not be included in calculating an individual's baseline cost of living. This approach implies that two identical individuals in the same location face the same cost of living index, regardless of their original locations. Still, a third identical individual may face a higher cost of living index in another location since positive moving costs prevent her from choosing the location with a lower net-of-moving-costs cost of living. Such a scenario underscores the impact that the spatial nature of public good provision has on the distribution of benefits.

Consider the case in which there are no moving costs. Then, individuals simply choose the location with the optimal bundle of income, housing prices, and amenities. If any location offers higher income, lower prices, and/or better amenities (i.e. lower cost of living), more individuals will move to that location, driving down wages and increasing prices. The equilibrium result is that all locations offer the same cost of living for identical individuals. The existence of moving costs and the previous assumption, however, preclude such a result. An alternative location may offer higher income, lower prices, and/or better amenities, but positive moving costs prevent an individual from obtaining the bundle. Therefore, the individual is worse off in her current location due to the inferior bundle of income, prices, and amenities.

As a thought experiment, assume two locations, A and B, where individual 1 begins in location A and individual 2 begins in location B. An individual chooses A if $U_{iA} > U_{iB}$ and chooses B if $U_{iA} < U_{iB}$. Assume that both locations offer the same income and price level, but B has better amenities and that both individuals have identical preferences. Further assume that equilibrium sorting results in individual 1 staying in A and individual 2 staying in B. Intuitively, it is easy to see that individual 2 is better off than individual 1 due to higher amenity value of B. Individual 1 chooses to not move to the location with a lower cost of living due to constraints arising from moving costs. From the structure of utility, individual 1's decision allows us to assign a lower bound to the cost of moving from A to B. However, to capture the idea that individual 1 is worse off, moving costs are left out of the COL calculation. While moving costs serve as a necessary factor in explaining the observed equilibrium, they are not a reasonable component in calculating the cost of living in a particular location. Therefore, the derivation of the expenditure function is modified to omit moving costs in the cost of living.

3.4 Cost of Living Calculation

Unobservable idiosyncratic shocks, η_{im}, make equation (7) individual specific. To calculate a type-location expenditure function, I simulate draws of η_{im} from a Type-I Extreme Value distribution[6] for every individual in each location. This calculation gives the average expenditure level required, for each individual type, to reach $\bar{V}$ in each location. The necessity of an average in this instance is due to the presence of heterogeneous preferences. Note that $\bar{V}$ is chosen to be the average calculated utility in the data set in 1990. I use the parameters of the indirect utility function, presented in the Appendix, along with each individual's observed income level to calculate their utility in their optimal location. I then average this utility across all individuals to get $\bar{V}$. For the index model in which expenditure functions are type-specific, average utility $\bar{V}_k$ is calculated separately for each type. The income required to reach $\bar{V}$ is averaged over the random idiosyncratic draws for each type of individual in a location to get the value of the type-location expenditure function.

Regarding simulation of the idiosyncratic term η_{im}, random draws are taken conditional on observed behavior. The optimal location choice reveals some information about the distribution of the random term, namely that draws are values such that maximum utility is achieved in the observed location,

$$V_{im} + \eta_{im} \geq V_{in} + \eta_{in} \quad \forall n \neq m, \tag{8}$$

where individual i lives in MSA m. I follow von Haefen (2003) for the conditional simulation of η. The required expenditure to reach $\bar{V}$ in each location is calculated as the average across individuals and across 1000 trials of the simulation.

After integrating over idiosyncratic shocks and removing moving costs, the expenditure function for a type k individual in MSA m becomes

$$E_m^k = \exp\left[\frac{1}{\beta_I}\left(\ln \bar{V} - \theta_m - \tau^k \Gamma_m^k\right)\right]. \tag{9}$$

A type-location COL index is defined as the ratio of the value of the expenditure function relative to that of the same individual type in a base location,

$$COL_m^k = \frac{E_m^k}{E_0^k}. \tag{10}$$

The term E_0^k denotes the mean expenditure necessary for a type k individual to reach $\bar{V}$, in which the mean is taken across MSAs. The COL index reflects whether the attributes of a location make it more or less expensive to achieve some level of utility relative to the base location. The result is an $Mx1$ vector of COL indices for each year in the price model for both moving cost specifications. For the index models that include preference heterogeneity, there exists an $Mx1$ vector of COL indices for each year and each type.

Beyond calculating such an index for each MSA, interest lies in the impact of particular amenities, such as air pollution, on the cost of living. This impact can be characterized by the correlation between a location's COL index and the utility impact of air pollution. Define PM_m as the level of PM10 (particulate matter with a diameter of 10 micrometers or less) in MSA m. Then, the relevant measure is the correlation between COL_m^k and PM_m. Since air pollution is a disamenity, a positive correlation implies that a higher cost of living is driven by poor air quality. A negative correlation implies locations with a high cost of living tend to have poor air quality, and thus the high cost of achieving some level of well-being is a result of other attributes.

3.5 Adjusted Income

The COL indices derived above are used to deflate observed income. This works to adjust income according to market prices and public goods. The resulting measure reflects an individual's income in the context of the price of housing, as in traditional COL indices, as well as the level of public amenities available at a location:

$$\mathcal{I}_{im} = \frac{I_{im}}{COL_m^k} \tag{11}$$

Adjusted income can inform us about two important features of the economy. The first is the evolution in the standard of living over time. Standard economic analysis examines standards of living by considering nominal income and prices to calculate real income. Such an approach focuses on individuals' ability to consume in private markets but does not account for their consumption of public goods. Across two periods it is possible that real income declines but the actual well-being of individuals may increase if public good consumption increases by a sufficient amount. This latter notion of well-being may be a better measure.

The second question surrounds the distribution of adjusted income, particularly as it compares to the distribution of observed real income. The distribution of observed income is of interest insomuch as it conveys information about the distribution of standard of living throughout the population. However, the general equilibrium theory that underlies residential sorting models implies that differences in income, to some degree, reflect compensation for public goods and prices in private markets. It is therefore necessary to control for differences in location amenities and housing prices to obtain what is arguably a better measure of the standard of living. This measure is estimated as the adjusted income measure explained above. An examination of the distribution of adjusted income relative to the distribution of observed income reveals the degree to which public good provision alleviates or exacerbates inequalities in the standard of living. If empirical results show a more equal distribution of adjusted income, then public goods serve to make up for variation in real income across the population. Thus a portion of the inequality in income exists as compensation for public amenities across locations, rather than differences in the overall standard of living. For example, in an economy in which all individuals received the same compensation for their production, income differences would be entirely due to compensation for spatial variation in public goods and the distribution of adjusted income would reduce to a degenerate distribution. However, if empirical results show higher inequality in the distribution of adjusted income, relative to real income, we can take this as evidence that access to public goods increases inequalities in the standard of living. In considering only real income in such a situation, the level of inequality in the standard of living would be underestimated since higher income individuals also have access to a superior set of public goods.

A final matter related to adjusted income that deserves attention is the impact of moving costs. In the absence of moving costs, public amenities should be fully capitalized into labor and housing markets, and any differences in the quality of locations will be arbitraged away. Of course in the context of the earlier discussion surrounding the removal of moving costs from expenditure functions, when costly migration is present individuals may be prevented from obtaining a lower cost of living in a distant location. This feature of the sorting process drives a wedge between expenditure functions for identical individuals so that location amenities are not entirely capitalized into labor and housing markets. While moving costs may prevent some individuals from moving to locations with a lower cost of living, moving costs will also shield particular individuals from an influx of people that could drive up housing prices and drive down wages. To test this effect, I use a comparison of adjusted income calculated with and without moving costs in the expenditure function. A more unequal distribution of adjusted income calculated without moving costs indicates that the state of the geographic landscape results in disproportionately greater moving costs for individuals that already have a lower standard of living. Such an outcome could be the result of public policy that has increased the set of public goods in already desirable locations and thus disproportionately benefits those individuals at the higher end of the standard of living distribution. Moving costs can be interpreted as transaction costs in obtaining welfare increasing consumption bundles, and thus their impact may be of concern to policy.

Given the empirical focus on the inequality of particular distributions, the primary criterion used to evaluate the equity of different distributions is the Gini coefficient, G, a measure of inequality based on the cumulative distribution of some variable. Normalize income so that $\tilde{I}_i = \frac{I_i}{I_{max}}$, which suggests an individual's relative income. Assuming some parametric distribution, a fitted cdf $F(\tilde{I}) = S$ represents the share of the population with relative income less than $\tilde{I}$. The inverse cdf, $F^{-1}(S)$, represents the level of relative income below which S percent of the population receive, known as the Lorenz curve. The Gini coefficient can be calcuated as

$$G = 1 - 2\int_0^1 F^{-1}(t)dt \tag{12}$$

If income is distributed equally among a population, $F^{-1}(S)$ is a straight line whose integral over $t \in [0, 1]$ is equal to $\frac{1}{2}$. Therefore, the Gini coefficient equals 0, indicating perfect equality. Conversely, if all income is received by a single individual and $\tilde{I}_i = 0$ for the remainder of the population, the integral of the inverse cdf approaches 0. The resulting Gini coefficient is 1, suggesting the highest degree of inequality.

I use a discrete approximation to (12) to calculate the Gini coefficient for both observed and adjusted income. Due to the arbitrary scale of utility in the model, the Gini coefficient is calculated for a sample of the entire population, as well as separately for each individual type. For the distribution of income described by the vector I,

$$G = \frac{\sum_i^n \sum_j^n |I_i - I_j|}{2n^2 \bar{I}} \tag{13}$$

where n is the number of individuals and $\bar{I}$ is mean income.

4 Data

COL indices are calculated from a parameterized residential sorting model. The sorting model requires income predictions, migration data, and observations of individuals living in MSAs. The same data are used for calculation of expenditure fucntions. These data are obtained from the Integrated Public Use Microdata Series (IPUMS), managed by the University of Minnesota. Datasets for 1990 and 2000 include a 5% sample of census long form observations, identifying individuals in their current MSAs as well as their birth and previous location. Variables also include income and individual attributes that define type identification. Based on the availability of data and consistency across years, a subset of 226 MSAs in the continental United States is chosen out of the full set of 290 defined MSAs in 1990 and 301 MSAs in 2000. This subset covers 71% $(171,413,984)$ of the U.S. population in 1990 and 70% $(190,474,896)$ in 2000. Furthermore, the sample accounts for 91% and 86% of the nation's urban population in 1990 and 2000, respectively. I restrict the sample to only household heads between the ages of 23 and 40 in an attempt to capture mobile individuals that are active participants in the housing and labor markets. Of course this also implies that results and policy implications are conditional on the subset of the population from which the data are drawn. After random sampling, the final sample includes $37,165$ individuals for the year 2000 and $39,058$ individuals for the year 1990, spanning the 226 MSAs. Appendix A further discusses the data used to construct variables and estimate necessary parameters.

5 Variation in the Cost of Living Indices

There are several factors that contribute to variation across locations in COL indices. Consider a simplified, though unrealistic, scenario in which moving costs and idiosyncratic preference differences do not exist. In such a case, COL indices reflect only income differences for identical occupations in different locations. A location with a high cost of living must offer indivdiuals higher income. Otherwise, it is optimal to move to a location with a lower cost of living. This follows directly from the theory of compensating differentials in the labor market. The aggregate impact of amenities and market prices, that which is captured by COL indices, should be fully arbitraged away in the labor market to reach a spatial equilibrium.

In terms of adjusted income, a location's cost of living compensates for cross-location wage differences. Therefore, for a population of identical individuals, adjusted income will be equal across the entire population. For a population with different earning power, COL indices will still compensate lower wage earners within a particular "skill group"; i.e. compensation is for geographic differences in wages, rather than for wage differences due to factors such as education or occupation. This causes the distribution of adjusted income to move closer to an equal distribution. In such a case, the distribution of adjusted income reflects the distribution of earning power.

In the presence of costly migration, however, the arbitrage mechanism discussed above does not operate to its fullest extent. The value of amenities and prices is not fully arbitraged through the labor market since there is a discrete cost of obtaining any given bundle. Even if an alternative location offers a lower cost of living and identical income, an individual may not take advantage of the opportunity to move if the utility gain is less than the cost of moving. The impact of moving costs generates further differences in COL indices across space, and thus in adjusted income. The final impact is due to a combination of the spatial distribution of individual types (variation in individuals' earning power and cost of moving) relative to the spatial distribution of location amenities. The effect of costly migration on the distribution of adjusted income will be further discussed with the empirical results.

6 Empirical Results

6.1 COL Indices

In interpreting the results of COL indices, I concentrate exclusively on the spatial distribution of public goods. The empirical framework is based on location decisions in a specific time period, rather than choices across time, revealing relative values for location-specific amenities across cities. Therefore, while it is tempting to study changes in the cost of living from 1990 to 2000, such an analysis goes beyond the identifying assumptions of the model. To be able to make comparisons over time, the individual's utility function must be identical in both years. Such an assumption implies that any unobservable scale of an individual's utility is constant in both years. While this paper does restrict utility to the degree that preference parameters on market goods and location amenities are identical across time, this paper does not make this further restriction, as it is differenced out in the cross-location empirical specification.

Tables 1 and 2 report summary statistics for COL indices in 1990 and 2000. Since indices are normalized with average cost of living as the base, rather than a particular location, all means are equal to 1 and are omitted. The impact of public goods is most evident by ranking the MSAs by COL looking at the MSAs with the highest and lowest COL indices. Since the model implies type-specific COL indices, we have to look at 8 different sets of MSA rank orderings, one for each preference type. However, in each year there are a few MSAs that show up consistently in the top five highest and lowest COL indices, respectively, across all types. First, consider the MSAs with the lowest COL indices. One can interpret low indices as indicating a desirable locations based on the overall cost of living. Thus a location with a very strong set of public goods may in fact have a high cost of living, and be deemed undesirable, if the price of housing is extremely high. Alternatively, some locations may be desirable on the grounds on a low price of housing. Across types, we see Danville, VA, Joplin, MO, Los Angeles, CA, and Phoenix, AZ in 1990 and Dothan, AL, Fargo-Moorehad, ND/MN, Fort Smith, AR/OK, and Laredo, TX in 2000. Some of these results are intuitive, as MSAs with very low market prices have a low cost of living. However, cities such as Los Angeles, CA and Phoenix, AZ with high housing prices still have a low COL index due to the set of location amenities they offer. Similar results hold for cities with high COL indices. For 1990, we see cities such as Ann Arbor, MI, Danbury, CT, Santa Cruz, CA, and Youngstown-Warren, OH/PA with very high costs of living. For 2000, Ann Arbor, MI, Danbury, CT, San Jose, CA, San Francisco, CA, and Stamford, CT all consistently have high costs of living. Again there is a mix of MSAs with high market prices and low market prices, indicating the importance of location amenities in the COL calculation. More formally, the correlation coefficient for housing prices and COL indices ranges across types between only 0.25 and 0.41 for 1990 and between 0.30 and 0.45 for 2000.

Also, recall that costly migration is omitted from calculation of COL indices for reasons discussed earlier. While index calculations that include moving costs are not reported here, I offer a brief discussion of the differences in the results of the two methods. First, consider the range of estimates. The 10% and 90% quantiles for COL indices across individual types for 1990 range from .75 to .82 and 1.21 to 1.30. For 2000, these ranges are .72 to .78 and 1.19 to 1.28. Compare this to results derived when moving costs are included. For 1990, the 10% and 90% quantiles are .80 to .85 and 1.20 to 1.23 respectively, and .75 to .81 and 1.16 to 1.20 for 2000. In general, the range of COL indices across MSAs becomes slightly smaller when moving costs are included in the index. This is not surprising given the mechanism of residential sorting. In removing moving costs from the COL calculation, I am removing a component of disutility. Thus, individuals in low-COL MSAs are currently facing higher moving costs, presumably incurred to reach such locations. Similarly, individuals in high-COL MSAs are currently facing relatively lower moving costs, having not incurred the same magnitude of costs to reach a low-COL location. Second, variation across types is smaller once moving costs are included in the COL calculation. The mean (averaged across MSAs) cross-type standard deviation in the COL index is .041 for 1990 and .040 for 2000. Compare this to .027 for 1990 and .029 for 2000 when moving costs enter the calculation. This suggests that the degree to which variation in the landscape offers alternative residential locations for individuals may differ based on an individual's education and family structure. Therefore different proportions of different types may be observed to have incurred moving costs. This is an indirect result of different magnitudes of residential sorting among the different types of individuals, due in part to variation in original locations of individuals.

Finally, I run a linear regression using COL indices as the dependent variable and the logged price of housing services ρ^Y and the logged level of particulate matter concentrations PM_{10}[7] respectively, as independent variables. Though a very simplified approach, it is useful for seeing the role of one aspect of environmental quality in the cost of living. Regressions are estimated separately for each individual type and contain a year fixed effect. While the magnitude of the parameters does not offer an interesting interpretation, all coefficient estimates are positive and significant. Estimates are shown in Table 3. Both housing prices and concentrations of PM_{10} impact a location's cost of living. Though it is obvious in the former case, these results suggest the importance of considering differences in air pollution across cities, even in an overly simplistic regression. In general, access to public goods and environmental quality are meaningful factors in determining an individual's cost of living, in additional to market prices.

6.2 Adjusted Income

Given COL indices for each type, MSA, and year, it is now possible to calculate adjusted income. Recall that adjusted income is nominal income normalized to reflect the location-specific cost of achieving a particular level of well-being. For example, if the COL index is 1.5 and an individual earns \$90,000, adjusted income is calculated as \$90,000/1.5 = \$60,000. Intuitively, the individual's income in the high cost of living location is the equivalent of \$60,000 in the base location. Consistent comparisons of well-being are based on such an approach, which accounts for both market prices and local amenities.

The level of variation in COL indices points to the imperfect nature of income as a proxy for the standard of living. Furthermore, controlling for differences in market prices only partially corrects for this. For the remainder of the analysis, I will focus on the distribution of three different sets of income data: 1. nominal income deflated for national inflation, referred to as real income, 2. nominal income deflated for MSA-specific housing prices, referred to as price-deflated income and 3. nominal income adjusted for both MSA-specific housing prices and public goods, referred to as adjusted income.

It is obvious that any positive level of public good consumption generates some benefits and increases an individual's welfare. However, the distribution of these benefits across the population may be of greater interest. As discussed earlier, this question is approached using the Gini coefficient for the adjusted income distribution relative to that of the real income distribution. Table 4 shows Gini coefficients in 1990 and 2000 for the distributions of income, price-deflated income, and adjusted income. Calculations are done for the population as a whole as well as for each individual type. Table 4 reports results for the distribution of all individuals in a sample of $920,208$ individuals for 1990 and $782,323$ individuals for 2000.

For both years, a higher Gini coefficient for adjusted income, relative to real income, denotes a more unequal distribution. Before further discussion of this point, however, note that there is a decrease in inequality when price-deflated income is compared to real income. Market prices, represented here as housing prices, tend to have an equalizing force on the distribution of well-being. Low-income individuals face lower housing prices and thus experience an increase in their relative living standards.

Significant increases in the Gini coefficient when both location amenities and prices are introduced indicate a higher degree of inequality in living standards across individuals when both income and public goods are considered. Much of this effect can be attributed to public goods, as is evident from a larger gap between adjusted income and real income, relative to adjusted income and price-deflated income. Higher income individuals, therefore, tend to live in locations that offer a better set of location amenities and market prices. Note that an increase in the Gini coefficient for adjusted income relative to price-deflated income goes beyond the idea that individuals with higher income simply consume more public goods. This result also accounts for housing price variation that reflects different levels of public goods. In other words, the adjustment is made for the aggregate impact of the benefit of public goods net of the cost of housing price adjustments for such goods. The larger Gini coefficient on adjusted income demonstrates more inequality that is driven by unequal consumption of high-quality locations, in additional to the observed in inequality in the income distribution.

An increase in the Gini coefficient holds for all individual types, as is evident in Table 5, though the adjustments for public goods are less drastic since a portion of the inequality in adjusted income is due to cross-type differences. For each level of education, the level of inequality in adjusted income is considerably higher for those individuals with children. Results also demonstrate that inequality in adjusted income is greater among individuals with higher levels of education. However, it is important to make clear that at least some portion of these differences can be explained by differences in the distribution of income before factoring in the value of public goods. These results should therefore be taken as additional evidence of the impact of including public goods in inequality measures, rather than evidence of differences in inequality among different subgroups of the population.

To isolate the impact of public amenities, I focus on the distributions of price-deflated income and adjusted income. The Lorenz curves displayed in Figure 1 plot the underlying distributions used to calculate the Gini coefficients, and are plotted along with the forty-five degree line that signifies a perfectly equal distribution. Recall that higher inequality is denoted by curves that are bowed further from this forty-five degree line. The graph in Panel A shows changes in the real income distribution from 1990 to 2000 to demonstrate some magnitude. Panels B and C both indicate a higher degree of inequality in adjusted income, as discussed above. While these changes look small graphically, they are fairly substantial in the context of changes in real income.

As an additional means of giving context to differences in the distribution of adjusted income relative to price-deflated income, consider scenarios that would alter the real income distribution so that it reaches a level of inequality identical to that of the adjusted income distribution. While there are an infinite number of ways to redistribute income to achieve any particular Gini coefficient, I describe here an empirical exercise that demonstrates the difference in inequality across the real income and adjusted income distributions. Consider redistributing income from the bottom half of the distribution to the top half. For 1990, the adjusted income Gini coefficient is replicated in the real income distribution by reducing all incomes below the median by $1,180, and evenly distributing that money to all individuals above the median income level. For 2000, the required individual reduction in income is $941. Rather than focusing on the bottom half versus the top half of the income distribution, consider the bottom 20% and the top 20%. In this case, redistributing $1,890 from the bottom results in an income distribution with a Gini coefficient identical to that of the adjusted income distribution. The corresponding amount for the 2000 distribution is $1,499. While differences in the distributions of real and adjusted incomes may appear small when looking at the Gini coefficient or the Lorenz curve, these differences become quite substantial when put in the more intuitive context of a monetary redistribution. Note that this redistribution exercise illustrates the magnitude of differences in values of Gini coefficient only to provide context for comparing the distribution of adjusted income distribution to that of price-deflated income, but not to consider any particular redistribution policy. Any redistributions in the economy would certainly have larger general equilibrium impacts.

Recall that the above calculations are based on expenditure functions that do not include moving costs in an individual's expenditure function for reasons discussed earlier. In particular, the omission of moving costs in COL calculation makes for a better interpretation of the standard of living. However, a comparison of results obtained when omitting versus including moving costs highlights whether moving costs disproportionately impact different segments of the population. Gini coefficients for both sets of results are shown in Table 6. The top row and the first row for each type are simply repeats of numbers reported in Table 4. The second rows, respectively, show Gini coefficients when moving costs are included in expenditure functions used to calculate COL indices. In all cases, there is a slight decrease in the level of inequality when moving costs are included. Such a result corresponds to the earlier discussion of COL indices and the inclusion of moving costs. The decrease in inequality suggests that individuals with lower adjusted incomes are currently facing lower costs of migrating away from their birthplace. Put another way, individuals with higher standards of living, relative to the rest of the population, disproportionately have incurred high moving costs to reach betters locations. This fits well with the intuition of residential sorting.

Finally, an interesting empirical conclusion is derived from the change in the Gini coefficient from 1990 to 2000. When measured using adjusted income, there has been a considerable increase in inequality from 1990 to 2000. However, that increase largely mimics the increase in inequality of price-deflated income. While it is evident that access to public goods exacerbates the apparent level of inequality, the severity with which this happens has not increased over the study period. Rather, the change in inequality over time has been driven by the real income distribution, with an inherent degree of inequality driven by public goods. Still, these results, combined with the more general results concerning inequalities, suggest the importance of distributional considerations in public good provision.

7 Conclusion

This study has developed location-specific cost of living indices that account not only for variation in market prices, as do conventional COL indices, but also access to public goods. A model of residential location sorting reveals implicit values for public goods and accounts for costly migration in obtaining such goods. Given that public amenities offer recreation, consumption, and health benefits it is reasonable to include them in any measure of welfare. Furthermore, wage and price differences across space reflect compensation for non-market goods. Therefore it becomes important to look beyond income and market prices to gain a complete understanding of the standard of living.

The previously discussed analyses in Nordhaus (1999) and Banzhaf (2005) found that conventional estimates of inflation were biased upwards, implying a downward bias in real income, as public goods offer benefits that increase the standard of living. Building on the notion of biased real income, this paper further examines the role that public goods play in standard of living estimates by considering the spatial variation in public goods. Therefore the level bias in real income, though not explicitly measured, is likely to involve considerable variation. Empirical results show significant differences in the cost of living across MSAs in the United States. We see locations that have high prices and are thus often considered high cost of living cities that have a low cost of living due to their bundle of public goods. The opposite holds for some cities with low market prices.

COL indices allow for calculation of adjusted income, an arguably better measure of well-being that is easily comparable across locations with varying prices and public goods. Of primary importance is an empirical analysis of distributional issues. While there is some level of inequality inherent in the income distribution, results of this study suggest that those inequalities are exacerbated when consumption of non-market goods is considered. Such a result goes beyond the obvious conclusion that wealthy individuals implicitly purchase more public goods. Rather, empirical findings point to the result that wealthy individuals enjoy a lower cost of living, so that a particular level of well-being is more easily attained. Thus we see a higher degree of inequality across the population when considering a measure of standard-of-living defined in a broader sense than income. These conclusions are even more significant when one considers the result predicted by a theory of compensating wages. Rather than a more equally defined distribution of adjusted income due to higher wages existing as compensation for lower levels of location attributes, empirical results show that the inclusion of public goods in individuals' welfare demonstrates a population with greater inequality.

It should again be noted that the empirical analysis, including parameter estimates and COL calculations, focused on only the subset of the U.S. urban population with household heads aged 23-40. Thus I exclude several important demographic groups, including young adults, middle-aged individuals, retirees, and those living in rural areas. Also note that the unit of analysis in this study is an individual. Thus the underlying model assumes individual preferences and location decisions, rather than a household-level decision. The extent to which such an assumption influences parameter estimates determines the impact it may have on calculated COL indices. This caveat is also important as it relates to the general results of the distribution of public goods, as a household-level analysis may be preferable for discussions of standard of living. While the empirical strategy presents a framework for analyzing the distribution of a broad welfare measure and illustrates the role of public goods in such a measure, findings should be interpreted with restrictions in mind and one should be cautious regarding inferences about the general population. A final caveat is that this study uses income to determine the contribution of market goods in an individual's standard of living. The alternative approach is to use wealth, including real and financial assets, which may offer a better characterization of an individual's access to market consumption.

In the context of a growing literature examining economic inequality, results from this paper indicate the importance of accounting for public goods when analyzing the distribution of a broad measure of the standard of living. While the existing literature documents income inequality that is partially mitigated by progressive income taxes and wealth in other assets, the inclusion of public goods in an individuals's consumption bundle appears to increase the level of inequality in the United States. Though Moretti (2013) presents evidence that such inequality decreased from 1990 to 2000, as discussed earlier, the level of inequality in the standard living during this time period was considerably higher than that of conventional income measures.

Appendices

A Model Parameters

A.1 Model Construction

The following section discusses estimation of necessary components of the residential sorting model in equation (3). Recall that the indirect utility function is of the form

Potential earnings, denoted by I_{im}, indicates the level of income that would be available to an individual i in MSA m. Estimation of the sorting model requires potential earnings of all individuals in all locations. Recall that each MSA is assumed to be a single labor market. Estimates from a regression of observed income on individual characteristics are used to predict income across locations based on location specific coefficients and an individual's characteristics. A semi-parametric correction for sorting in the labor market is proposed by Dahl (2002) and implemented in a residential sorting model in Bayer et al. (2009). This method rests on the intuition that while unobserved location attributes induce moves to labor markets, information regarding such sorting is captured by the probability of an identical type individual making the observed move.

An income regression is estimated in which the dependent variable is the log of weekly income, as reported on the U.S. Census long form. The sample includes $874,809$ total observations for 1990 and $749,618$ total observations for 2000, with a range of MSA populations of $465 - 45,921$ (mean of $3,976$) and $390 - 34,476$ (mean of $3,637$) for each year, respectively. Regressors include gender, marital status, race/ethnicity, age, part-time employment, citizenship, education, and industry, as well as type-specific migration probabilities to account for nonrandom sorting. Observations include employed household heads not in the military and not disabled. A separate wage hedonic is run for each MSA using A-14, controlling for effects related to labor demand and allowing the income effect of individual traits to vary across locations. The equation for estimation is

$$\ln I_{im} = \alpha_I X_i^{Inc} + \alpha_{P1} P_{k,R1:R2} + \alpha_{P2} \left(P_{k,R1:R2}\right)^2 \tag{A-14}$$

Individual attributes are included in X_i^{Inc}, while $P_{k,R1:R2}$ is the observed probability that an individual of type k migrated from region $R1$ to region $R2$. Data constraints limit the probabilities to migration between regions, rather than between MSAs.

Income is predicted for each individual in each MSA for 1990 and 2000. Note that the migration probabilities act only as controls for consistent estimates, and so are not included when predicting income. The weekly wage prediction is multiplied by the individual's number of weeks worked to obtain the final income prediction, $\hat{I}_{im}$.

Results for the hedonic wage regression include 226 sets of coefficients representing each MSA/labor market, so only summary statistics are discussed. Coefficient means all have expected signs, with positive wage premiums for individuals who are male, married, white, older, educated, U.S. citizens, full time workers, and are in management positions. In some cases, the minimum coefficient estimate may be negative for a variable that has a positive (and expected positive) mean coefficient estimate. Similarly, the maximum coefficient estimate may be positive for a variable that has an negative (and expected negative) mean coefficient estimate. The fact that such results show up for some MSAs is likely a function of supply and demand idiosyncracies in local labor markets. The last two rows report coefficients on type specific migration probabilities and migration probabilities squared. There is very little interpretation that follows these parameters, but their significance suggests that the approach is valid. These regression coefficients are then used to predict annual incomes for each individual in each MSA.

The empirical model then requires the price of a homogeneous unit of housing services, rather than the price of a house. From an individual's budget constraint, the market price for house i in tract j of MSA m can be expressed as

$$P_{im} = \rho_m^Y H_i \exp(\nu_{im}), \tag{A-15}$$

where ν_{im} is a house-specific idiosyncratic shock.

Define the index of housing services as a function that maps the vector of discrete structural variables of a property h_i into a continuous index defined by $H_i = \exp(\phi h_i)$. With this equation (A-15) can be rewritten as

$$P_{im} = \rho_m^Y \exp(\phi h_i) \exp(\nu_{im}). \tag{A-16}$$

Empirically, I am interested in an estimate of $\ln \rho_m^Y$ for the sorting model. Taking logs,

$$\ln P_{im} = \ln \rho_m^Y + \phi h_i + \nu_{im}, \tag{A-17}$$

and the parameter of interest, $\ln \rho_m^Y$, is estimated as a location m fixed effect.

The price that serves as the dependent variable in (A-17) is the annual cost of housing. For units that are rented, this value is simply the annual rent plus utilities and fees. For owned units, however, an annual rent must be imputed from the housing value. Rents are calculated in a manner similar to Albouy (2012) and Blomquist et al. (1988), following methods described in Poterba (1992). In equilibrium, the ratio of the rental value to the house price is the user cost of owner-occupied housing. This ratio is equal to the sum of the nominal interest rate, property tax rate, risk premium, maintenance costs and depreciation, less the inflation rate. Since property taxes are observed and treated as an additional annual cost after calculating annual rent, the property tax rate is omitted from calculation of the user cost of owner-occupied housing. Following Poterba (1992), I assign maintenance costs and depreciation each to be 2% and the risk premium on home ownership at 4%. The nominal interest rate is the average commitment rate on new fixed mortgages and the inflation rate is calculated as a five year average of the CPI inflation rate. The respective values for the interest rate and inflation are 10.13 and 4.12 for 1990 and 8.05 and 2.54 for 2000. The resulting rent to value ratio is 11.48 for 1990 and 11.50 for 2000. Given a rent to value ratio of 11.5, a $100,000 house takes on an annual housing cost of $8,695, equivalent to an apartment with monthly rent of $725. Final calculations for the cost of housing are found by adding annual property taxes and utility payments to estimates of the annual cost of housing.

This hedonic is estimated on a sample of $262,735$ households for 1990 and $233,095$ households for 2000 across the 226 MSAs, obtained from the IPUMS dataset. Note that the set of households used for the MSA hedonic estimation is not the same as the set of individuals used for the macro sorting model. Both data sets are subsets of the same set of household observations, but different selection criteria lead to different sets of observations.

Equation (A-17) estimates an MSA-specific price for each year. The hedonic regression gives highly significant and expected results. Price increases for homes that have larger living quarters, are more recently built, and on larger plots of land. In addition, single family detached homes and apartment units are more expensive than attached single family homes.

References

David Albouy. Are big cities bad places to live? estimating quality of life across metropolitan areas. *Working Paper*, 2012.

Philip Armour, Richard V. Burkhauser, and Jeff Larrimore. Deconstructing income and income measures: A crosswalk from market income to comprehensive income. *American Economic Review*, 103(3):173–177, 2013.

H. Spencer Banzhaf. Green price indices. *Journal of Environmental Economics and Management*, 49:262–280, 2005.

Patrick Bayer, Nathaniel Keohane, and Christopher Timmins. Migration and hedonic valuation: The case of air quality. *Journal of Environmental Economics and Management*, 58(1):1–14, 2009.

Glenn C. Blomquist, Marc C. Berger, and John P. Hoehn. New estimates of qaulity of life in urban areas. *The American Economic Review*, 78(1):89–107, March 1988.

Gordon B. Dahl. Mobility and the return to education: Testing a roy model with multiple markets. *Econometrica*, 70(6):2367–2420, 2002.

Irving Fisher. *The Making of Index Numbers*. Houghton Mifflin Company, 1922.

Timothy Hamilton. *A Multi-Level Residential Sorting Model with an Application to Cost of Living Indices*. PhD thesis, North Carolina State University, 2012.

Timothy L. Hamilton and Daniel J. Phaneuf. An integrated model of regional and local residential sorting with application to air quality. *Journal of Environmental Economics and Management (forthcoming)*, 2015.

Jonathan Heathcote, Fabrizio Perri, and Giovani L. Violante. Unequal we stand: An empirical analysis of economic inequality in the united states, 1967-2006. *Review of Economic Dynamics*, 13(1):15–51, January 2010.

Matthew E. Kahn. A revealed preference approach to ranking city quality of life. *Journal of Urban Economics*, 38:221–235, 1995.

Nanak Kakwani and Robert J. Hill. Economic theory of spatial cost of living indices with application to thailand. *Journal of Public Economics*, 86:71–97, 2002.

H. Allen Klaiber and Daniel J. Phaneuf. Valuing open space in a residential sorting model of the twin cities. *Journal of Environmental Economics and Management*, 60(2):57–77, September 2010.

A. A. Konus. The problem of the true index of the cost of living. *Econometrica*, 7(1):10–29, 1939.

Nicolai V. Kuminoff, V. Kerry Smith, and Christopher Timmins. The new economics of equilibrium sorting and policy evaluation using housing markets. *Journal of Economic Literature*, 51(4): 1007–1062, 2013.

Enrico Moretti. Real wage inequality. *American Economic Journal: Applied Economics*, 5(1): 65–103, 2013.

William D. Nordhaus. Beyond the cpi: An augmented cost-of-living index. *Journal of Business and Economic Statistics*, 17(2), April 1999.

Thomas Piketty and Emmanuel Saez. Income inequality in the united states, 1913-1998. *Quarterly Journal of Economics*, 118:1–39, February 2003.

James M. Poterba. Taxation and housing: Old questions, new answers. *The American Economic Review*, 82(2):237:242, May 1992.

Christopher Timmins. Estimating spatial differences in the brazilian cost of living with household location choices. *Journal of Development Economics*, 80:59–83, 2006.

Constant Tra, Anna Lukemeyer, and Helen Neil. Evaluating the welfare effects of school quality improvements: A residential sorting approach. *Journal of Regional Science (forthcoming)*, 2012.

Roger H. von Haefen. Incorporating observed choice into the construction of welfare measures from random utility models. *Journal of Environmental Economics and Management*, 45(2):145–165, March 2003.

Table 1: Summary Statistics for COL Indices: 1990

	Type I	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8
10% Percentile	0.804	0.788	0.755	0.749	0.818	0.785	0.760	0.757
Median	0.991	0.966	0.976	0.977	0.991	0.975	0.969	0.961
90% Percentile	1.215	1.267	1.277	1.292	1.208	1.268	1.270	1.305

Table 2: Summary Statistics for COL Indices: 2000

	Type I	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	Type 8
10% Percentile	0.779	0.758	0.740	0.739	0.773	0.772	0.747	0.718
Median	0.994	0.976	0.983	0.981	0.986	0.979	0.972	0.973
90% Percentile	1.199	1.256	1.253	1.244	1.190	1.246	1.266	1.285

Table 3: Regression of COL Index on Prices, PM_{10}

	ln Prices		ln PM_{10}	
	coef.	s.e.	coef.	s.e.
No H.S. Degree, No Children	.2669	.0602	.0043	.0008
H.S. Degree, No Children	.3169	.0621	.0050	.0009
College Degree, No Children 3	.2818	.0667	.0052	.0009
Graduate/Prof. Degree, No Children	.2495	.0675	.0055	.0010
No H.S. Degree, w/ Children	.2470	.0497	.0045	.0008
H.S. Degree, w/ Children	.3041	.0706	.0050	.0009
College Degree, w/ Children	.2685	.0722	.0052	.0009
Graduate/Prof. Degree, w/ Children	.2239	.0716	.0057	.0010

Table 4: Gini Coefficient

	1990	2000
Income	.3055	.3502
Price Deflated	.3048	.3469
Adjusted	.3247	.3647

Table 5: Type-Specific Gini Coefficients

	1990		2000	
	Price-Deflated	Adjusted	Price-Deflated	Adjusted
No H.S. Degree, No Children	.2545	.2685	.2895	.3026
H.S. Degree, No Children	.2554	.2647	.2757	.2878
College Degree, No Children	.2767	.3018	.3283	.3441
Graduate/Prof. Degree, No Children	.3260	.3574	.3683	.3881
No H.S. Degree, w/ Children	.2623	.2740	.2950	.3056
H.S. Degree, w/ Children	.2666	.2741	.2913	.3021
College Degree, w/ Children	.2990	.3154	.3633	.3752
Graduate/Prof. Degree, w/ Children	.3737	.3917	.4321	.4448

Table 6: Gini Coefficients: Impact of Moving Costs in Adjusted Income

		1990	2000
General	w/out MC	.3247	.3647
	w MC	.3188	.3595
No H.S. Degree, No Children	w/out MC	.2685	.3026
	w MC	.2648	.3009
H.S. Degree, No Children	w/out MC	.2647	.2878
	w MC	.2593	.2826
College Degree, No Children	w/out MC	.3018	.3441
	w MC	.2970	.3410
Graduate/Prof. Degree, No Children	w/out MC	.3574	.3881
	w MC	.3558	.3865
No H.S. Degree, w/ Children	w/out MC	.2740	.3056
	w MC	.2707	.3043
H.S. Degree, w/ Children	w/out MC	.2741	.3021
	w MC	.2685	.2963
College Degree, w/ Children	w/out MC	.3154	.3752
	w MC	.3090	.3702
Graduate/Prof., w/ Children	w/out MC	.3917	.4448
	w MC	.3882	.4428

Table A1: Macro Sorting Parameters

		Coef.	t-stat	τ
Income	β_I	1.673	63.330	
MC_{State}	μ_{bs}	2.670	134.273	
MC_{Region}	μ_{br}	1.173	58.132	
$\mathrm{MC}_{State}\times$ Children	ν_{bs}	0.230	8.512	
$\mathrm{MC}_{Region}\times$ Children	ν_{br}	0.145	5.200	
MSA Index: Γ^1	τ^1	0.718	7.305	0.672
MSA Index: Γ^2	τ^2	0.005	0.202	0.501
MSA Index: Γ^3	τ^3	1.137	23.874	0.757
MSA Index: Γ^4	τ^4	1.955	15.302	0.876
MSA Index: Γ^5	τ^5	0.724	11.524	0.673
MSA Index: Γ^6	τ^6	-0.191	-9.403	0.452
MSA Index: Γ^7	τ^7	0.293	7.357	0.573
MSA Index: Γ^8	τ^8	0.771	11.614	0.684

Observations: 1990 = 39058, 2000 = 37165

All parameters are significant at the 1% level

Figure Titles

Figure 1: Lorenz Curves for Price-Deflated and Adjusted Income

Notes

[1]Though not explicitly concerned with COL measure, a number of other papers use similar models to examine consumption of public goods. For example, see Klaiber and Phaneuf (2010) or Tra et al. (2012) for applications and Kuminoff et al. (2013) for an examination of such models.

[2]It is equally valid to measure COL relative to a particular location, using a specific location's necessary expenditures as the denominator in equation (2). In that case, a location's COL is interpreted relative to the specific base location and any two locations are directly comparable as they have identical denominators.

[3]An MSA is a U.S. Census designated geographic entity that is equivalent to the metropolitan area around a major city. In 2000, there were 301 MSAs in the continental U.S. comprising roughly 81% of the U.S. population.

[4]Individual type k is defined as a unique combination of having or not having children and one of four levels of educational attainment: no high school diploma, high school diploma, college degree, graduate/professional degree.

[5]While this is an important part of estimation in residential sorting models in general, it becomes trivial in what follows since this analysis is not concerned with identifying the impacts of specific amenities in Y_m

[6]This distribution is used based on its function as the assumed distribution for the idiosyncratic shock in the logit estimation of the underlying sorting model

[7]PM_{10} is used in this application due to its prevalence in the literature valuing air quality.