

THE ASSOCIATION BETWEEN URBAN SPRAWL AND OBESITY: IS IT A TWO-WAY STREET?

ANDREW J. PLANTINGA

Department of Agricultural and Resource Economics
232B Ballard Extension Hall
Oregon State University
Corvallis, OR 97331-3601
plantinga@oregonstate.edu

STEPHANIE BERNELL

Department of Public Health
312 Waldo Hall
Oregon State University
Corvallis, OR 97331-6406
Stephanie.Bernell@oregonstate.edu

Draft: August 11, 2005

The authors acknowledge comments from session participants at the 2004 Association of Collegiate Schools of Planning Conference, where an earlier version of this paper was presented.

THE ASSOCIATION BETWEEN URBAN SPRAWL AND OBESITY: IS IT A TWO-WAY STREET?

ABSTRACT: In this paper we examine the relationship between obesity and the characteristics of metropolitan counties where individuals reside. Previous empirical analyses have treated urban form as exogenous to weight status. These studies find higher body mass indices (BMI) for residents of areas with development patterns associated with urban sprawl (e.g., low residential densities, poor street connectivity). The main purpose of this paper is to test whether BMI is a factor that explains the location decisions of individuals. We estimate a two-equation system with BMI and the latent variable determining the choice of a high or low sprawl county as endogenous variables. Our results support the hypothesis that BMI is a factor in determining the attributes of an individual's residential location. In particular, individuals with greater BMI have a higher probability of living in high sprawl counties. In contrast, we find limited evidence that living in a sprawling county increases BMI. Our results suggest that the association between sprawl and obesity reported in earlier studies is due largely to self-selection rather than to impacts of the urban environment on physical activity and weight.

THE ASSOCIATION BETWEEN URBAN SPRAWL AND OBESITY: IS IT A TWO-WAY STREET?

1. INTRODUCTION

This paper examines the relationship between patterns of urban land development and obesity. In the United States, obesity has risen at an epidemic rate during the past 20 years (Flegal et al., 2002; Ogden et al., 2002; Hedley et al., 2004). According to data from 1999-2002, approximately 65% of adults in the United States are overweight or obese,¹ compared to 47% between 1976 and 1980. The increases in the prevalence of overweight and obesity cut across all racial, gender and ethnic groups. Two issues have attracted much attention from researchers. First, what explains the rise in obesity rates in the U.S.? Recent economic studies (Lakdawalla and Philipson, 2002; Cutler, Glaeser, and Shapiro, 2003; Chou, Grossman, and Saffer, 2004) point to factors such as reductions in food prices and job-related exercise, increases in work hours, and the prevalence of fast-food restaurants. A second question is, what explains differences in obesity rates across groups of individuals? Previous studies have shown that obesity rates vary by factors such as income, education, race, age, sex, smoking habits, and diet (Flegal et al., 2002; Freedman et al., 2002; Wellman and Friedberg, 2002; Reidpath et al., 2002). Recently, urban planning and public health researchers have suggested that development patterns associated with urban sprawl may be related to higher rates of obesity in the U.S. population (e.g., Killingsworth and Lamming, 2001; Torres et al., 2001; Frumkin, 2002; Jackson and Kocktitzky, 2002; Frank, Engelke, and Schmid, 2003; Jackson, 2003; Killingsworth, Earp, and Moore, 2003).

There are several strands to the argument linking obesity with urban sprawl. First, poor connectivity in street networks and low-density and single-use development increase trip distances. This reduces physical activity by making walking and bicycling impractical and unsafe. Second, low-density development reduces the viability of public transportation; hence, people must commute to and from work by automobile, which can lead to traffic congestion and the diversion of time from activities such as

¹ Overweight and obesity are measured using the body mass index (BMI), computed as weight in kilograms divided by height in meters squared. Individuals with BMI between 25 and 30 are classified as overweight and individuals with BMI in excess of 30 are classified as obese.

exercising. Third, modern suburban developments do not adequately provide facilities such as parks that permit and encourage physical activity. The result is that people drive more, exercise less, and gain weight.

The most compelling evidence on the link between sprawl and obesity is found in studies by Ewing et al. (2003), Frank, Andresen, and Schmid (2004), and Lopez (2004). Ewing et al. and Lopez study residents of metropolitan counties who responded to the Behavioral Risk Factor Surveillance System (BRFSS) survey in 2000. The BRFSS collects data on the weight and height of individuals, which are used to measure BMI, in addition to demographic and other health-related variables. For each metropolitan county, Ewing et al. develop a sprawl index based on four measures of population density and two measures of street block size. The authors then regress BMI on the sprawl index for the individual's county of residence plus variables controlling for gender, age, race, education, smoking status, and diet.² Lopez constructs a metropolitan area index based on population density and estimates the effects of the sprawl index and a similar set of individual-level variables on the probability that individuals are obese. In both studies, the sprawl index is found to be significantly related to BMI. In particular, residents of counties or metropolitan areas with higher sprawl indices have systematically higher BMI and greater probability of being obese.³ Frank, Andresen, and Schmid study the likelihood of obesity in a sample of Atlanta residents, employing an individual-specific sprawl measure. Street connectivity and land-use mix are measured in a 1-kilometer buffer surrounding each respondent's residence. They find that residents of mixed-use neighborhoods tend to be less obese.

An important assumption in the urban planning/public health literature on sprawl and obesity is that urban form is exogenous to an individual's weight.⁴ In particular, it is assumed that when land use

² In all of the studies discussed here, BMI and the characteristics of an individual's location are measured at the same point in time (i.e., the contemporaneous relationship between BMI and sprawl is analyzed).

³ Kelly-Schwartz et al. (2004) examine a similar national sample of individuals (the NHANES III) and employ a metropolitan area sprawl measure, but find no correlation between sprawl and BMI.

⁴ See Figure 1 in Ewing et al. (2003), Figure 1.1 in Frank, Engelke, and Schmid (2003), and Figure 1 in Kelly-Schwartz et al. (2004).

patterns produce low levels of physical activity, individuals have a greater tendency to become obese. The urban planning/public health model considers how an individual is affected by the pattern of land use where they reside, but does not account for the individual's choice of residence. If weight affects an individual's choice of the type of neighborhood to reside in, then urban form is not exogenous to weight. The implication for earlier studies is that sprawl indices are endogenous regressors and, therefore, the coefficient estimates reported in these studies are biased and inconsistent. A second implication is that policies designed to reduce obesity through modifications of urban form—e.g., high density housing, provision of sidewalks and parks—may have limited effects if low-weight individuals sort into areas that provide opportunities for exercise. As discussed below, a number of recent government and private initiatives use urban planning approaches to promote higher levels of physical activity.

Why might weight affect residential location? In section 2, we present a simple model in which an individual maximizes utility defined over weight, attributes of their residential location, and a composite good. Weight is determined by a function that depends on initial weight, locational attributes, and the composite good. Because an individual selects their location (in order to choose the corresponding attributes) and because utility depends on weight, the optimal choice of locational attributes is a function of the individual's initial weight. Thus, weight is a factor explaining the sorting of individuals into different types of locations.

A more complete theory of weight and urban land use is developed by Plantinga and Bernell (2005). These authors analyze an urban spatial model in which households maximize utility defined over housing, weight, and food subject to a fixed time budget allocated to commuting, calorie expenditure, and work. In spatial market equilibrium, the housing market adjusts to equalize utility levels across the city. Unit housing prices decline with distance to the central business district to compensate residents for commuting costs and house sizes may increase to compensate residents for locations that produce weight gain. The weight of residents and housing densities at each location are simultaneously determined. The model is used to demonstrate that residents with greater aversion to weight gain will sort into higher-density locations closer to the central business district.

In this paper, we conduct an empirical analysis to determine whether weight affects the types of locations selected by individuals. Section 3 discusses the data used in the study. Our main data source is the National Longitudinal Survey of Youth 1979 (NLSY79), which tracks a nationally representative sample of individuals who were ages 14 to 22 years in 1979. We use data from the NLSY79 for the year 2000 to measure BMI, individual attributes such as income and education, and county of residence. We use the sprawl index developed by McCann and Ewing (2003) to categorize counties as having either a high or low degree of sprawl.

Section 4 presents the empirical approach. Our econometric model is a two-equation system with BMI and the latent variable determining residence in a high or low sprawl county as endogenous variables. To investigate the dynamic relationship between sprawl and weight, the model is estimated with separate samples of individuals who moved recently and who have remained in the same county for a specified number of years. Results are presented in section 5 and a final section discusses conclusions and directions for future work.

2. THEORETICAL BACKGROUND AND EMPIRICAL APPROACH

We present a model of an individual who influences their weight through residential location and consumption of other goods. To simplify the analysis, we separate the choice of residential location from the decision about whether or not to change locations, and focus on the residential choice problem the individual faces once the decision to move has been made. We assume that weight (W) depends on initial weight, attributes of a location such as compactness and walkability, and other consumption goods:

$$(1) \quad W = f(W_0, \mathbf{N}, C)$$

where W_0 is initial weight (e.g., weight at the start of the period when consumption choices are made), $\mathbf{N} = (N_1, N_2, \dots, N_n)$ is a vector of locational attributes, and C is a composite numeraire good that may include food and exercise equipment. An individual chooses a location and the quantity of the composite good to maximize utility defined over weight, locational attributes, and other goods:

$$(2) \quad \max_{\mathbf{N}, C} U(W, \mathbf{N}, C) \quad \text{subject to} \quad \mathbf{p}\mathbf{N} + C = I, \quad W = f(W_0, \mathbf{N}, C)$$

where \mathbf{p} is a conformable vector of implicit prices for locational attributes⁵ and I is income.

Substituting for C in the utility function and denoting the multiplier on the weight constraint by λ , the Lagrangian function is written,

$$(3) \quad L = U(W, \mathbf{N}, I - \mathbf{p}\mathbf{N}) + \lambda[W - f(W_0, \mathbf{N}, I - \mathbf{p}\mathbf{N})],$$

and the corresponding first-order conditions for an interior optimum are,

$$(4) \quad U_{N_i} - p_i U_C - \lambda(f_{N_i} - p_i f_C) = 0 \quad i=1, \dots, n$$

$$(5) \quad U_W - \lambda = 0$$

$$(6) \quad W - f(W_0, \mathbf{N}, I - \mathbf{p}\mathbf{N}) = 0$$

Equations (4), (5), and (6) are combined to yield,

$$(7) \quad N_i^* = N_i(W_0, \mathbf{p}, I) \quad i=1, \dots, n$$

$$(8) \quad W^* = f(W_0, \mathbf{N}^*, I - \mathbf{p}\mathbf{N}^*)$$

where \mathbf{N}^* and W^* are the optimal levels of locational attributes and weight, respectively.

The results in (7) and (8) have important implications for the empirical analysis that follows. First, an individual's weight is shown to depend on the attributes of their neighborhood, the relationship identified in the earlier urban planning/public health literature. However, the attributes of the location are also a function of initial weight. This dependence emerges because an individual selects their location and because their utility is a function of weight. Thus, we expect a simultaneous relationship between weight and locational attributes.⁶ Second, we expect the contemporaneous relationship between weight and locational attributes to vary by individuals who have moved recently (movers) and those who have

⁵ As in the hedonic housing price literature, we assume that individuals can select any combination of residential attributes, each of which is priced implicitly. Thus, in selecting a location the individual chooses a bundle of housing characteristics.

⁶ In this formulation, locational attributes depend on initial weight rather than on realized weight (W^*). Plantinga and Bernell (2005) model the market for developed land in which locational characteristics can adjust to satisfy the demands of households. In this case, realized weight and development densities are determined simultaneously.

been in the same location for a period of time (non-movers). The effects of locational attributes on weight clearly take some time to materialize. For example, if walkability has a negative influence on weight, an individual does not lose weight the moment he or she moves to a walkable location. Rather, an individual may need to live at this location for a period of years before these effects are apparent. In general, we expect the contemporaneous effect of locational attributes on weight to be weaker for movers than non-movers. It is uncertain, *a priori*, whether the effects of weight on location choice will differ between movers and non-movers. According to (7), movers can optimally select locational attributes conditional on their current weight. Likewise, non-movers may effectively select their optimal attributes by remaining in the same location. In this case, we would expect a strong contemporaneous effect of weight on locational attributes for movers and non-movers. However, if the costs of moving are prohibitively high for some individuals, then their locational attributes may become suboptimal over time. If so, we would find a weaker sorting effect among non-movers.

In the urban planning/public health literature, the typical approach is to estimate a version of (8). Specifically, BMI is regressed on an index of locational attributes, such as a county sprawl index in Ewing et al. (2003), and individual characteristics. The individual characteristics proxy for other factors—diet, smoking, other forms of physical activity—that may influence weight. In this study, we will also account for the influence of weight in determining residential location, as shown in (7). We estimate a simultaneous equations model with contemporaneous BMI and locational attributes as the endogenous variables⁷ and individual characteristics as the exogenous variables. We estimate separate models with samples of movers and non-movers to test for different relationships between weight and sprawl for these groups and to shed light on the dynamics of this relationship—for example, do individuals have to reside in a location for a period of time before locational attributes affect their weight?

⁷ We evaluate the contemporaneous relationship between BMI and locational attributes because, as discussed below, we have data on a sprawl index for only a single year.

3. DATA DESCRIPTION

The National Longitudinal Survey of Youth 1979 (NLSY79) is the primary data source for this analysis. The NLSY79 is conducted by the U.S. Department of Labor, currently on a biennial basis. Respondents are a nationally representative sample of 12,686 men and women who ranged in age from 14 to 22 years when first surveyed in 1979. We use data from the NLSY79 for the year 2000, when the respondents were from age 35 to 43 years. The variables include weight, height, race, sex, smoking status, income, number of children, marital status, age, and educational attainment. The weight data in the NLSY79 is self-reported. Because heavier individuals are believed to systematically underreport their weight, we use the weight correction procedure developed by Cawley (1999) to adjust the data.⁸ Height data were last collected in 1985 when respondents were ages 20 to 28. We assume that heights did not change over the intervening years. The adjusted weight and height data are used to compute the BMI for each respondent (denoted ADJBMI).

The NLSY79 uses 26 race categories, which we combined into four exhaustive categories: black, hispanic, asian, and white (non-black, non-hispanic, and non-asian). Dummy variables (DBLACK, DHISPANIC, DASIAN, AND DWHITE) are created for each of these categories. DSEX, DSMOKE, and DMARRIED are dummy variables that takes the value 1 if, respectively, the respondent is male, smokes daily, and is married, and 0 otherwise. INCOME is defined as net family income in 2000. The data on number of children are used to create dummy variables for no children (DCHILD0), one child (DCHILD1), two children (DCHILD2), and three or more children (DCHILD3). AGE is defined as the age of the respondent in 2000 and AGESQ is the square of this variable. Educational status is measured by dummy variables for attainment less than high school (EDUC1), high school (EDUC2), and some college or more (EDUC3). The NLSY79 database reports the state that each respondent resides in. Using

⁸ The adjustments were developed through a comparison of self-reported and measured weights of individuals and vary according to sex and race (white and non-white). We thank Darius Lakdawalla for providing us with the adjustment factors.

standard regional designations, we construct dummy variables for the Western (WEST), North Central (NCENTRAL), Northeast (NEAST), Southeast (SEAST), and South Central (SCENTRAL) regions.

We also applied to the Department of Labor to obtain confidential data on the county of residence.⁹ These data allow us to match the residence of respondents in 2000 to a sprawl index discussed below. As well, we use data on the county of residence in earlier years to distinguish between recent movers and non-movers. A respondent is defined as a mover if there was a change in their county of residence during a specified period prior to 2000. Non-movers are those respondents whose county of residence did not change over this time. We vary to length of the period (from 2 to 10 years in increments of 2 years) to produce different subsamples of movers and non-movers.

Urban form was measured using the county sprawl index developed by McCann and Ewing (2003). The authors use principal components analysis to combine four county measures of population density (persons per square mile, % of population living at densities below 1,500 persons per square mile, % of population living at densities above 12,500 persons per square mile, county population divided by square miles of urban land) and two county measures of street block size (average block size, % of blocks 1/100 of a square mile or less in size). Most of the data used to construct the sprawl index is from the 2000 Census of Population, the same year as the individual-level data extracted from the NLSY79. The sprawl index is developed for 448 metropolitan counties. The scores range from 63 for the county with the most sprawl (Geauga) to 352 for the county with the least (New York City). Unfortunately, the sprawl index is available only for a single year, which limits our ability to fully exploit the longitudinal structure of the NLSY79. We denote the index by SPRAWL. For reasons discussed in the next section, we make use of a binary variable that differentiates between low and high sprawl counties: DENSE takes the value 1 if $SPRAWL \geq 120$ and 0 if $SPRAWL < 120$.

⁹ Ewing et al. (2003) use data from the BRFSS survey, which provides many more observations than the NLSY79. However, obtaining and using BRFSS data on county of residence is exceedingly complicated, which is one reason we have opted to use the NLSY79 data.

Our final sample includes 3,867 respondents to the NLSY79 who reside in counties for which the sprawl index was measured¹⁰ and for whom we have complete observations of the variables defined above. Summary statistics for our sample are reported in Table 1. The mean adjusted BMI is 28.1 and approximately 30% of the sample are obese (BMI greater than 30). The respondents are predominately white (52%), followed by black (29%), hispanic (19%), and a negligible share of asians. One-half of the respondents are male, 23% smoke, and 55% are married. The average age of respondents is 39 years and the mean family income is approximately \$60,000. Similar shares of the sample have no children (18%) or one child (21%) and similar shares have two children (33%) or three or more children (28%). A large majority of the sample has a high school diploma (41%) or at least some college (50%). Only 9% of the sample has educational attainment less than high school. The largest shares of the sample reside in the North Central and Western regions (each 23%), followed by the Northeast (22%), Southeast (16%), and South Central regions (7%). Approximately 11% of the sample moved at least once in the previous two years (i.e., between 1998 and 2000). Using the alternative definitions of movers, we find that 19%, 27%, 36%, and 40% of the sample moved in the past 4, 6, 8, and 10 years, respectively.¹¹

Figure 1 depicts the distribution for the sprawl index. Specifically, it shows the number of counties represented in the sample with a sprawl index value in a given range. The distribution is skewed slightly to the right with a rough bell shape for values below 130 and some high-valued outliers. The mean of the distribution is 100 and the median is 97. To define the variable DENSE, we initially select 120 as the dividing line between low and high sprawl counties. By this definition, approximately 10% of the 448 metropolitan counties are low sprawl and 26% of the respondents in our sample (the mean of DENSE in Table 1) live in low sprawl counties. Our definition of low sprawl counties reflects the

¹⁰ McCann and Ewing (2003) report the sprawl index for 446 counties and for two county groups (Adams-Denver in Colorado and Arlington-Alexandria-Fairfax in Virginia). We separate these county groups into single counties and assume that the sprawl index in each is the same as in the respective county group.

¹¹ The size of the overall sample (i.e., movers and non-movers) falls to 3,678 when we employ the 2-year definition of movers and to 3,363 for the 10-year definition. This is due to missing observations for the county of residence in some years.

conventional wisdom that the large majority of U.S. metropolitan areas are characterized by sprawl. However, given that a cut-off of 120 is somewhat arbitrary, we also consider alternative definitions of DENSE using 115 and 125 as the lower bound for low sprawl counties. With the 115 (125) cut-off, 14% (8%) of the counties are low sprawl and 35% (23%) of the sample lives in a low sprawl county.

4. EMPIRICAL MODELS AND ESTIMATION PROCEDURES

We use data most comparable to that used by Ewing et al. (2003). Thus, we begin by demonstrating that we obtain results similar to theirs. We specify our BMI equation as:

$$(9) \quad \begin{aligned} \text{ADJBMI}_i = & \beta_0 + \beta_1 \text{DBLACK}_i + \beta_2 \text{DHISPANIC}_i + \beta_3 \text{DOTHER}_i + \beta_4 \text{DSEX}_i + \beta_5 \text{DSMOKE}_i \\ & + \beta_6 \text{AGE}_i + \beta_7 \text{AGESQ}_i + \beta_8 \text{DEDUC1}_i + \beta_9 \text{DEDUC2}_i + \beta_{10} \text{WEST}_i \\ & + \beta_{11} \text{NCENTRAL}_i + \beta_{12} \text{NEAST}_i + \beta_{13} \text{SEAST}_i + \beta_{14} \text{SPRAWL}_i + \varepsilon_i \end{aligned}$$

where i ($i=1, \dots, N$) indexes individuals, β_0 is the intercept term, β_j ($j=1, \dots, 14$) are coefficients on variables defined above, and ε_i is a spherical disturbance term. The intercept captures the effects of the omitted dummy variable categories (asian, female, non-smoker, some college or more, South Central region). There are a number of differences between (9) and the Ewing et al. model. As discussed above, our BMI measure is adjusted for reporting bias, though this adjustment does not appear to greatly affect BMI values. Ewing et al. use four categorical variables for age (30 to 44, 45 to 64, etc.). Because our sample is of individuals who are all between 35 and 43 years, we enter age and the square of age as continuous variables. Ewing et al. include one additional category for educational attainment (some college) and a variable for fruit and vegetable consumption. The NLSY79 does not provide information on diet and so this variable is excluded from our model.

Ewing et al. employ a two-stage hierarchical modeling approach to estimate their model. In the first stage, they estimate separate models for residents of each county. For each model, BMI is regressed on individual characteristics and an intercept term which measures county-specific effects. In the second stage, they regress the intercept estimates on the sprawl index. We lack sufficient observations to

estimate the first-stage county models. Instead, we include regional dummy variables in (9) and estimate the equation with least squares. The dummy variables may control for regional differences in factors such as climate that could influence an individual's propensity to exercise. The key difference between the Ewing et al. approach and ours is that they include a county-specific random effect, which should increase the efficiency of their estimates relative to our model with regional fixed effects. Because (9) does not include all of the variables listed in Table 1, fewer observations are omitted because of missing values. Thus, we estimate (9) with N=4,700 observations rather than the 3,867 observations mentioned above. We weight our data for sampling intensity using the sampling weight variable in the NLSY79. Finally, White's (1980) estimate of the covariance matrix is used to produce heteroskedastic-consistent standard errors for (9) and all linear equations estimated below.

The next model allows for the possibility that residential choice, and thus the sprawl index, is determined simultaneously with BMI.¹² The general specification of our model is:

$$(10) \quad \begin{aligned} \text{ADJBMI}_i &= \gamma_1 \text{SPRAWL}_i + \mathbf{X}_{1i} \boldsymbol{\beta}_1 + \varepsilon_{1i} \\ \text{SPRAWL}_i &= \gamma_2 \text{ADJBMI}_i + \mathbf{X}_{2i} \boldsymbol{\beta}_2 + \varepsilon_{2i} \end{aligned}$$

where i ($i=1, \dots, N$) indexes individuals, γ_1 and γ_2 are parameters on the endogenous variables (ADJBMI_i and SPRAWL_i), \mathbf{X}_{1i} and \mathbf{X}_{2i} are vectors of exogenous variables, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are conformable parameter vectors, and ε_{1i} and ε_{2i} are error terms. One approach would be to enter ADJBMI_i and SPRAWL_i as continuous variables and estimate (10) with two-stage least squares. However, because the sprawl index uniquely identifies counties, the variables in \mathbf{X}_{2i} would need to explain the choice of individuals to locate in particular counties. As demonstrated in the literature on the economics of migration (e.g., Hunt and Meuller, 2004), such decisions depend on anticipated returns to an individual's skills in distinct labor markets, location-specific amenities, distances from origin to alternative destination counties, and so

¹² See Boarnet and Crane (2001) for a related empirical study of travel behavior in which residential location is an endogenous regressor.

forth. Modeling county-level location decisions of individuals would be a substantial undertaking and is beyond the scope of this study.¹³

Rather than model particular location decisions, we assume that individuals make a higher-level choice of the type of county to reside in. We define two types of counties, those with low and high degrees of sprawl. We model the choice of county type, which is assumed to depend on individual attributes such as income, education, marital status, and so on. We do not explicitly model the lower-level choice of a particular county conditional on county type.¹⁴ Using the notation in (10), our structural model is specified:

$$(11) \quad \begin{aligned} \text{ADJBMI}_i &= \gamma_1 y_i^* + \mathbf{X}_{1i} \boldsymbol{\beta}_1 + \varepsilon_{1i} \\ y_i^* &= \gamma_2 \text{ADJBMI}_i + \mathbf{X}_{2i} \boldsymbol{\beta}_2 + \varepsilon_{2i} \end{aligned}$$

where y_i^* is a latent variable determining the choice of county type: $\text{DENSE}_i = 1$ if $y_i^* \geq 0$ and $\text{DENSE}_i = 0$ if $y_i^* < 0$. As in the discrete-choice literature (e.g., Train 2003), y_i^* can be interpreted as the net gain to an individual from choosing a low sprawl county. We might prefer to think of BMI as depending on the choice of county type, DENSE_i , rather the latent variable y_i^* . However, specifying (11) in terms of the discrete variable DENSE_i would mean that ε_{1i} and ε_{2i} have continuous and discrete distributions, respectively, complicating the estimation procedure. Because (11) is formulated in terms of the continuous latent variable, we can assume that ε_{1i} and ε_{2i} are independently and identically distributed (iid) bivariate normal random variables. Equation (11) can then be estimated using the simultaneous equation probit estimator discussed by Amemiya (1978).

¹³ Even in a state-of-the-art national migration study like Hunt and Meuller (2004), a higher-level of aggregation (states) is used to categorize origins and destinations.

¹⁴ As discussed above, this decision depends on factors such as anticipated labor market outcomes and amenities. By modeling the county type decision only in terms of individual attributes, we are assuming the labor market and amenity factors do not differ across the county types.

The exogenous variables in \mathbf{X}_{1i} and \mathbf{X}_{2i} measure individual characteristics that have direct effects on BMI and county type choice, respectively. \mathbf{X}_{1i} contains many of the same variables as in (9): the race, sex, smoking, age, and education variables, and the regional dummies. These variables proxy for the many factors that can influence an individual's weight, including physical activities other than those dependent on location (e.g., gym membership). We also include the family income variable (INCOME) in \mathbf{X}_{1i} . Income is determined in part by educational attainment, so the education variables measure additional effects on BMI beyond those transmitted through the income variable. \mathbf{X}_{2i} includes the race, sex, age, education, and income variables. We hypothesize that blacks and hispanics are more likely to choose denser counties because these groups have traditionally resided in inner cities. We expect younger people to be choose denser counties with greater probability because of a desire to live in the metropolitan downtown. In contrast, females may be less likely to live in such areas because of concerns about crime. Highly educated people may choose dense, downtown areas because of interest in cultural amenities. The effects of income on residential choice are less clear. High income may explain the decision to live in vibrant metropolitan cores such as New York City, but also the decision to live in lower density suburbs if living space is a normal good. We include the marriage (DMARRIED) and children (DCHILD1, DCHILD2, DCHILD3) variables in \mathbf{X}_{2i} . We expect married individuals with children to prefer less dense, suburban counties.

We estimate (11) using the two-stage procedure described by Maddala (1983, p. 244-5). Solving (11) for the reduced-form equations and expressing them in vector notation, we have:

$$(12) \quad \mathbf{ADJBMI} = \mathbf{X}\boldsymbol{\Pi}_1 + \boldsymbol{\nu}_1$$

$$(13) \quad \mathbf{y}^* = \mathbf{X}\boldsymbol{\Pi}_2 + \boldsymbol{\nu}_2$$

where \mathbf{ADJBMI} and \mathbf{y}^* are vectors of the endogenous variables, \mathbf{X} is a matrix of the distinct exogenous variables in \mathbf{X}_{1i} and \mathbf{X}_{2i} , $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ are parameter vectors, and $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ are error vectors with variances σ_1^2 and σ_2^2 , respectively, and covariance σ_{12} . The first stage involves estimating the

parameters of (12) by least squares and those of (13) by probit maximum likelihood. Probit estimation requires a normalization of the error distribution to identify the model. Standard practice is to normalize the variance of the error term to 1, yielding in our case:

$$(14) \quad \mathbf{y}^{**} = \frac{\mathbf{y}^*}{\sigma_2} = \mathbf{X} \frac{\boldsymbol{\Pi}_2}{\sigma_2} + \frac{\boldsymbol{\nu}_2}{\sigma_2} = \mathbf{X}\boldsymbol{\Pi}_2^* + \boldsymbol{\nu}_2^*$$

Substitution of \mathbf{y}_i^{**} in (11) yields the structural model, in vector notation:

$$(15) \quad \mathbf{ADJBMI} = \gamma_1 \sigma_2 \mathbf{y}^{**} + \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$$

$$(16) \quad \mathbf{y}^{**} = \frac{\gamma_2}{\sigma_2} \mathbf{ADJBMI} + \mathbf{X}_2 \frac{\boldsymbol{\beta}_2}{\sigma_2} + \frac{\boldsymbol{\varepsilon}_2}{\sigma_2}$$

First-stage estimation of (12) and (14) yields $\hat{\boldsymbol{\Pi}}_1$ and $\hat{\boldsymbol{\Pi}}_2^*$. These estimates are used to form predicted values of the endogenous variables, namely $\mathbf{X}_1 \hat{\boldsymbol{\Pi}}_1$ and $\mathbf{X}_2 \hat{\boldsymbol{\Pi}}_2^*$, which are substituted, respectively, for \mathbf{ADJBMI} in (15) and \mathbf{y}^{**} in (16). Consistent estimates of the parameters $\gamma_1 \sigma_2$, $\boldsymbol{\beta}_1$, γ_2 / σ_2 , and $\boldsymbol{\beta}_2 / \sigma_2$ are then produced by applying least squares and probit maximum likelihood, respectively, to (15) and (16).

Maddala (1983, p. 245) provides the covariance matrices for the two-stage estimates of $(\gamma_1 \sigma_2, \boldsymbol{\beta}_1)$ and $(\gamma_2 / \sigma_2, \boldsymbol{\beta}_2 / \sigma_2)$. The first covariance matrix is a function of $c = \sigma_1^2 - 2\gamma_1 \sigma_{12}$. Amemiya (1978, p. 1200) explains how to derive consistent estimates of these terms. The second matrix depends on σ_{12} / σ_2 , which we estimate from the residuals of (12) and (14). The structural parameters are identified only in the case of $\boldsymbol{\beta}_1$. Amemiya (1978) discusses a method for estimating the structural parameters of the model from the estimates of the reduced-form coefficients. We use the Amemiya procedure to estimate γ_1 , the unidentified structural parameter in (15). The structural parameters in the probit equation (γ_2 and $\boldsymbol{\beta}_2$) must be estimated with only four degrees of freedom, and this produces imprecise estimates. Therefore, we report the estimates of $(\gamma_2 / \sigma_2, \boldsymbol{\beta}_2 / \sigma_2)$. As in standard probit estimation, the

normalization on σ_2 is required to identify the model and affects only the interpretation of the parameters. Namely, the effects of the observed variables are interpreted relative to the standard deviation of the unobserved error term.

In all cases, we estimate the model in (15) and (16) with data on respondents for the year 2000. To investigate dynamic processes that may affect the relationship between BMI and county type (see section 2), we estimate separate models with samples of movers and non-movers. Our basic model is produced using the 4-year definition of movers and the 120 lower bound for low sprawl counties. We also conduct sensitivity analysis by estimating models with alternative definitions of movers and alternative definitions of low and high sprawl counties.

5. RESULTS

Estimation of equation (9) produces results similar to those reported in Ewing et al. (2003) (Table 2). The dummy variables for male (DSEX), black (DBLACK), hispanic (DHISPANIC), less than high school (DEDUC1), and high school (DEDUC2) are positive and significantly different from zero at the 5% significance level. These characteristics increase BMI relative to an asian female with some college or more. The dummy variable for white (DOTHER) is positive but not significantly different from zero. AGE has a positive and significant effect on BMI. The negative effect of AGESQ indicates that the incremental effect of age diminishes as age increases. Smoking (DSMOKE) and the sprawl index (SPRAWL) have negative and significant effects on BMI. Recall that higher values of the sprawl index correspond to counties with less sprawl. Thus, this result suggests that living in a less sprawling county reduces BMI. The coefficient on the county sprawl index is -0.0062 with a standard error of 0.0026. The coefficient estimated by Ewing et al. (-0.0034) falls within the 95% confidence interval for this estimate. The regional dummies are not significantly different from zero, indicating no significant differences relative to the omitted category (the South Central region).

The estimation results for the model in (15) and (16) are presented for movers (Table 3) and non-movers (Table 4) using the 4-year definition to define movers.¹⁵ The results support the prediction of our theoretical model that BMI is a factor in determining the type of county in which individuals locate. For movers and non-movers, BMIHAT has a negative and significant effect on the probability that an individual locates in a dense county. This result suggests that individuals with high BMI sort into sprawling counties. That BMI influences the location of non-movers suggests that these individuals effectively sort into county types by remaining in the same location. Further, we find that the latent variable \mathbf{y}^{**} does not have a significant effect on BMI. Individuals with a higher probability of living in a dense county do not have systematically higher or lower BMI. The dependence of BMI on location is strongly rejected for movers (the t -statistic on \mathbf{y}^{**} is -0.81). This supports our conjecture (see section 2) that individuals would need to live in the same place for a number of years before the effects of location on BMI are apparent. For non-movers, the t -statistic is -1.57. We reject the hypothesis of a non-zero effect at the 5% level, but would fail to reject the hypothesis at the 12% level. Thus, we find some weak evidence that location affects BMI for non-movers.

For movers and non-movers, we find that BMI is positively related to being male and older and negatively related to income and smoking (Tables 3 and 4). For non-movers, the race and education variables are also significant determinants of BMI. Specifically, BMI is higher for blacks and hispanics and for individuals with lower educational attainment. Qualitatively, the results for non-movers are similar to those found for the single-equation BMI model (Table 2), the key difference being that the effect of location weakens once the endogeneity of the sprawl variable is accounted for. The results for the DENSE equations show that, for movers, age has a positive but diminishing effect on the decision to locate in a dense county. None of the other exogenous variables are significantly different from zero in the structural model for movers (the coefficient on the sex variable is positive and significant at the 10% level). The estimates of the reduced-form parameters (not reported) indicate that being married has a

¹⁵ Likelihood ratio tests reject the null hypothesis that the models are the same for the two samples.

significant and negative effect on the probability of living in a dense county.¹⁶ Among non-movers, low educational attainment increases the probability that individuals will reside in a dense county and having three or more children reduces the probability. The race variables for black and hispanic are positive and significant at the 10% level. In the reduced-form model for non-movers, the marriage variable is negative and significantly different from zero.

We test the sensitivity of our results to assumptions about the definitions of movers and low sprawl counties (Table 5). Movers are alternatively defined as individuals who have moved at least once in the past 2, 6, 8, and 10 years. Low sprawl counties are alternatively defined as counties with a sprawl index exceeding 115 and 125. The coefficients on the endogenous variables (\mathbf{y}^{**} and BMIHAT) are reported for all combinations of mover and sprawl definitions (Table 5). The results show that, for movers, a higher probability of living in a dense county has an insignificant effect on BMI (\mathbf{y}^{**} , the latent variable in the BMI equation, is not significantly different from zero in all cases). In most cases, the same result is found for non-movers. The latent variable \mathbf{y}^{**} has a significant effect on BMI when the 2-year definition of movers is used, but an insignificant effect for all other definitions. For movers and non-movers, the effects of BMIHAT on the location in a dense county are significantly different from zero in 23 out of 30 cases (26 out of 30 cases if a 10% confidence level is used). Overall, we find that the results do not change appreciably when we vary the definitions of movers and low sprawl counties.

6. CONCLUSIONS

In this study, we examine the link between urban sprawl and obesity. Previous studies in the urban planning and public health literature have found that residents of sprawling areas tend to have

¹⁶ The parameters in the structural model measure direct effects on the corresponding left-hand side endogenous variable. This effect feeds back through the system as the induced change in the endogenous variable influences the other endogenous variable, and so on. The total effect of an exogenous variable on an endogenous variable accounts for all such feedbacks between the endogenous variables, and is found by solving the structural model. Estimates of the reduced-form parameters in (12) and (14) are consistent estimates of the total effects of each exogenous variable.

higher BMI. In these studies, the attributes of an individual's residence are treated as an exogenous determinant of their weight. We consider the alternative possibility that BMI is a factor in determining the type of location chosen by an individual. A two-equation system with BMI and the latent variable determining the choice of a high or low sprawl county as endogenous variables is estimated. We find evidence that individuals with higher BMI tend to select locations with a high degree of sprawl. This result was found for movers and non-movers and was robust to alternative definitions of movers and low sprawl counties. In contrast, we find much weaker evidence that residence in a sprawling area results in higher BMI. In some cases, the latent variable determining location choice had a significant (or weakly significant) effect on BMI for non-movers; for movers, the effect was insignificant in all cases. Our interpretation of these results is that the association between sprawl and obesity reported in earlier studies is due largely to sorting rather than to impacts of the urban environment on physical activity and weight.

Our findings have important policy implications. The authors of earlier studies (Ewing et al., 2003; Frank, Andresen, and Schmid, 2004), while stressing that the association between sprawl and obesity does not imply causality, use their results to argue that obesity and other health problems can be reduced through modifications of the urban environment.¹⁷ Urban planning approaches to increase physical activity have been proposed or are underway. For example, the Atlanta Regional Commission has proposed to invest \$1.1 billion in new bicycle and pedestrian facilities and \$500 million to extend a livable community initiative (Atlanta Regional Commission, 2004). The Robert Wood Johnson Foundation's Active Living by Design (ALbD) program establishes and evaluates innovative approaches to increasing physical activity through community design, public policies and communications strategies. A number of federal agencies (National Institutes of Health, Centers for Disease Control and Prevention,

¹⁷ Ewing et al. write, "... there is moderate support for the assertion that urban form can have significant (positive or negative) influences on health and health-related behaviors. If this assertion is true, health practitioners can improve public health by advocating for more compact development patterns." Frank, Andresen, and Schmid write, "This (public health) goal may be achieved through a variety of policy options that include ... longer-term changes in the built environment, such as increased mixed use, density, and street connectivity that make walking an attractive and viable option." Later on, they state that, "The change in land use mix from 0.15 to 0.30, although a substantial increase, was not outside policy control in certain areas of the Atlanta region."

Department of Transportation, Environmental Protection Agency, National Park Service, National Forest Service) have programs that use design approaches to promote physical activity. Our results suggest that such policies may have limited effects. We find that individuals with low BMI tend to locate in counties with less sprawl. This suggests that efforts to curb sprawl, and thereby make communities more exercise-friendly, may simply attract those individuals who are predisposed to physical activity.

Obesity and residential choice are complicated phenomena, and so more research is needed to identify the relationship between them. In the current study, we observe the county sprawl index at only one point in time, which restricts us to modeling obesity status and residential choice as contemporaneous outcomes. We attempt to shed light on the dynamic relationship between sprawl and obesity by separately considering movers and non-movers. However, more progress could be made if repeated observations of the sprawl index were available. For example, we could test whether *changes* in residential location (e.g., a move from a low to high sprawl county) result in different patterns of weight change. If not, we could attribute the association between sprawl and obesity to self-selection with greater confidence. The sprawl index used in this study is measured at the county level. While this is the best information available at this time, further investigation into sub-county variation in urban development patterns is warranted. Frank, Andreson, and Schmid (2004) use measures of urban form in the area surrounding each individual's residence. This approach seems preferable but would be difficult to implement on a national scale.

REFERENCES

- Amemiya, T. 1978. The Estimation of a Simultaneous Equation Generalized Probit Model. *Econometrica* 46(5):1193-1205.
- Atlanta Regional Commission. 2004. 2004 Annual Report. The Shape of Things to Come. Atlanta, GA.
- Boarnet, M.G., and R. Crane. 2001. *Travel by Design: The Influence of Urban Form on Travel*. Spatial Information Systems Series. Oxford and New York: Oxford University Press.
- Cawley, J. 1999. Rational Addiction, the Consumption of Calories, and Body Weight. PhD Dissertation. University of Chicago, Chicago, IL.
- Chou, Shin-Yi, Michael Grossman, and Henry Saffer. 2004. "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System," *Journal of Health Economics*, 23, 565-587.
- Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro. 2003. "Why Have Americans Become More Obese?," *Journal of Economic Perspectives*, 17, 93-118.
- Estrella, A. 1998. "A New Measure of Fit for Equations with Dichotomous Dependent Variables." *Journal of Business Economics and Statistics* (April):198-205.
- Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., and S. Raudenbush. 2003. "Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity," *American Journal of Health Promotion*, Sept./Oct., 47-57.
- Flegal, K.M., Carroll, M.D., Ogden, C.L., and C.L. Johnson. 2002. Prevalence and Trends in Obesity Among U.S. Adults. *Journal of the American Medical Association* 288(14):1723-7.
- Frank, L.D., Andresen, M.A., and T.L. Schmid. 2004. Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars. *American Journal of Preventive Medicine* 27(2):87-96.
- Frank, L.D., Engelke, P.O., and T.L. Schmid. 2003. *Health and Community Design: The Impact of the Built Environment on Physical Activity*. Washington, DC: Island Press.
- Freedman, D.S., Khan, L.K., Serdula, M.K., Galuska, D.A., and W.H. Dietz. 2002. Trends and Correlates of Class 3 Obesity in the United States From 1990 Through 2000. *Journal of the American Medical Association* 288:1758-1761.
- Frumkin, H. 2002. "Urban Sprawl and Public Health," *Public Health Reports*, 117, 201-218.
- Hedley, A.A., Ogden, C.L., Johnson, C.L., Carroll, M.D., Curtin, L.R., and K.M. Flegal. 2004. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. *Journal of the American Medical Association* 291:2847-50.
- Hunt, G.L., and R.E. Mueller. 2004. Canadian Immigration to the U.S., 1985-1990: Estimates from a Roy Selection Model of Differences in Returns to Skill. *Review of Economics and Statistics* 86(4):988-1007.

- Jackson, R.J. 2003. "The Impact of the Built Environment on Health: An Emerging Field," *American Journal of Public Health*, 93, 1382-1383.
- Jackson, R.J., and C. Kochtitzky. 2002. "Creating a Healthy Environment: The Impact of the Built Environment on Public Health." Centers for Disease Control and Prevention.
- Kelly-Schwartz, A.C., Stockard, J., Doyle, S., and M. Schlossberg. 2004. Is Sprawl Unhealthy? A Multilevel Analysis of the Relationship of Metropolitan Sprawl to the Health of Individuals. *Journal of Planning Education and Research* 24:184-96.
- Killingsworth, R., Earp, J., and R. Moore. 2003. "Supporting Health Through Design: Challenges and Opportunities," *American Journal of Health Promotion*, 18, 1-3.
- Killingsworth, R.E., and J. Lamming. 2001. "Development and Public Health: Could our Development Patterns be Affecting our Personal Health?," *Urban Land*, July, 12-17.
- Lakdawalla, Darius and Tomas Philipson. 2002. "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination." National Bureau of Economic Research Working Paper 8946.
- Lopez, R. 2004. Urban Sprawl and Risk for Being Overweight or Obese. *American Journal of Public Health* 94(9):1574-9.
- Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press.
- McCann, B.A., and R. Ewing. 2003. Measuring the Health Effects of Sprawl: A National Analysis of Physical Activity, Obesity, and Chronic Disease. Smart Growth America, Surface Transportation Policy Project, September.
- Ogden, C.L., Flegal, K.M., Carroll, M.D., and C.L. Johnson. 2002. Prevalence and trends in overweight among US children and adolescents, 1999-2000. *Journal of the American Medical Association* 288:1728-32.
- Plantinga, A.J., and S. Bernell. 2004. A Spatial Economic Analysis of Urban Land Use and Obesity. *Journal of Regional Science* 45(3):473-492.
- Reidpath, D.D., Burns, C., Garrard, J., Mahoney, M., and M. Townsend. 2002. An Ecological Study of the Relationship Between Social and Environmental Determinants of Obesity. *Health and Place* 8(2):141-5.
- Torres, G.W., Pittman, M., Hollander, M., Kraft, M.K, and E. Henry. 2001. *Active Living Through Community Design*. White Paper Prepared for the Robert Wood Johnson Foundation, February.
- Train, K.E. 2003. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.
- Wellman, N.S. and B. Friedberg. 2002. Causes and Consequences of Adult Obesity: Health, Social and Economic Impacts in the United States. *Asia Pacific Journal of Clinical Nutrition* 11 Suppl 8: S705-9.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48(May):817-38.

TABLE 1. Summary Statistics

Variable	Description	Summary Statistics
ADJBMI	Adjusted body mass index	Mean = 28.1 Percent > 30 = 30.1 Minimum = 10.6 Maximum = 57.3
DBLACK	Dummy variable for black	Mean = 0.29
DHISPANIC	Dummy variable for hispanic	Mean = 0.19
DASIAN	Dummy variable for asian	Mean = 0.003
DOTHER	Dummy variable for white	Mean = 0.52
DSEX	Dummy variable for sex (male = 1)	Mean = 0.50
DSMOKE	Dummy variable for smoking (daily = 1)	Mean = 0.23
DMARRIED	Dummy variable for married (yes = 1)	Mean = 0.55
INCOME	Family income (\$1000 per year)	Mean = \$61 Minimum = \$0 Maximum = \$333
DCHILD0	Dummy variable for no children	Mean = 0.21
DCHILD1	Dummy variable for 1 child	Mean = 0.18
DCHILD2	Dummy variable for 2 children	Mean = 0.33
DCHILD3	Dummy variable for 3 or more children	Mean = 0.28
AGE	Age of respondent	Mean = 39 years
DEDUC1	Educational attainment less than high school	Mean = 0.09
DEDUC2	Educational attainment is high school	Mean = 0.41
DEDUC3	Educational attainment is some college or more	Mean = 0.50
WEST	Dummy variable for Western region	Mean = 0.22
NCENTRAL	Dummy variable for North Central region	Mean = 0.23
NEAST	Dummy variable for Northeast region	Mean = 0.22
SEAST	Dummy variable for Southeast region	Mean = 0.16
SCENTRAL	Dummy variable for South Central region	Mean = 0.07
DENSE	Dummy variable equal to 1 for residence in a county with SPRAWL \geq 120	Mean = 0.26

TABLE 2. Estimation Results for Single-Equation BMI Model (Equation 9)

Variable	Estimate	t-statistic
C	-59.845*	-2.1515
DSEX	1.3501*	7.3892
AGE	4.3825*	3.0727
AGESQ	-0.0558*	-3.0580
DOTHER	0.6929	0.7248
DBLACK	3.1565*	3.2702
DHISPANIC	2.1199*	2.1587
DEDUC1	1.4276*	3.5544
DEDUC2	0.8105*	4.1156
DSMOKE	-1.0141*	-4.5181
WEST	-0.2069	-0.6437
NCENTRAL	0.0224	0.0738
NEAST	0.1468	0.4541
SEAST	-0.3964	-1.2529
SPRAWL	-0.0062*	-2.3604

Dependent variable = ADJBMI

Number of observations = 4700

Standard error of regression = 5.344

Adjusted R-square (transformed data) = 0.05

Note: * indicates the coefficient estimate is significantly different from zero at the 5% level

TABLE 3. Estimation Results for Simultaneous Equation Model: Movers (4-year Definition)

Variable	BMI Equation (15)		DENSE Equation (16)	
	Estimate	Asym. <i>t</i> -stat	Estimate	Asym. <i>t</i> -stat
<i>Endogenous</i>				
BMIHAT			-0.328*	-2.438
DENSE (y^{**})	-0.789	-0.811		
<i>Exogenous</i>				
Intercept	-130.9	-1.654	-72.47*	-2.243
DSEX	1.441*	3.479	0.479	1.827
AGE	8.056*	2.001	4.126*	2.363
AGESQ	-0.102*	-1.971	-0.053*	-2.356
DOTHER	-1.188	-0.489	-0.240	-0.223
DBLACK	0.154	0.062	0.259	0.239
DHISPANIC	-0.064	-0.024	0.733	0.671
DEDUC1	0.068	0.079	-0.152	-0.418
DEDUC2	-0.178	-0.372	-0.263	-1.263
INCOME	-0.012*	-3.116	-0.003	-1.305
DMARRIED			-0.231	-1.103
DCHILD1			0.383	1.328
DCHILD2			-0.014	-0.054
DCHILD3			0.059	0.218
DSMOKE	-1.136*	-2.148		
WEST	0.237	0.180		
NCENTRAL	0.900	1.049		
NEAST	0.301	0.187		
SEAST	-0.246	-0.250		
Dependent variable = ADJBMI			Dependent variable = DENSE	
Number of observations = 703			Number of observations = 703	
Standard error of regression = 5.251			Scaled R-square (Estrella 1998)	
Adjusted R-square (transformed data) = 0.03			= 0.09	

Note: * indicates the coefficient estimate is significantly different from zero at the 5% level

TABLE 4. Estimation Results for Simultaneous Equation Model: Non-Movers (4-year Definition)

Variable	BMI Equation (15)		DENSE Equation (16)	
	Estimate	Asym. <i>t</i> -stat	Estimate	Asym. <i>t</i> -stat
<i>Endogenous</i>				
BMIHAT			-0.144*	-2.275
DENSE (\mathbf{y}^{**})	-1.182	-1.573		
<i>Exogenous</i>				
Intercept	-45.38	-1.328	-4.775	-0.483
DSEX	0.502*	2.248	0.069	0.997
AGE	3.499*	1.996	0.398	0.745
AGESQ	-0.044*	-1.985	-0.005	-0.735
DOTHER	1.762	1.123	-0.010	-0.023
DBLACK	5.136*	3.244	0.910	1.718
DHISPANIC	3.914*	2.452	0.906	1.850
DEDUC1	1.570*	3.560	0.432*	3.331
DEDUC2	0.634*	2.319	-0.013	-0.159
INCOME	-0.008*	-3.891	0.000	0.207
DMARRIED			-0.145	-1.663
DCHILD1			-0.172	-1.660
DCHILD2			-0.117	-1.233
DCHILD3			-0.229*	-2.360
DSMOKE	-1.270*	-4.601		
WEST	1.119	0.972		
NCENTRAL	0.994	1.007		
NEAST	1.638	1.231		
SEAST	0.597	0.726		
Dependent variable = ADJBMI			Dependent variable = DENSE	
Number of observations = 2904			Number of observations = 2904	
Standard error of regression = 5.670			Scaled R-square (Estrella 1998)	
Adjusted R-square (transformed data) = 0.05			= 0.07	

Note: * indicates the coefficient estimate is significantly different from zero at the 5% level

TABLE 5. Effects of the Endogenous Variables for Alternative Definitions of Movers and Non-Movers and Low and High Sprawl Counties

Definition	DENSE (y^{**})		BMIHAT	
	Estimate	Asym. t -stat	Estimate	Asym. t -stat
Low sprawl counties defined by SPRAWL \geq 120				
<i>Movers</i>				
2-year	0.349	0.441	-0.386*	-2.323
4-year	-0.789	-0.811	-0.328*	-2.438
6-year	-0.259	-0.284	-0.494*	-2.713
8-year	-0.421	-0.556	-0.306*	-2.472
10-year	-0.427	-0.516	-0.258	-1.949
<i>Non-movers</i>				
2-year	-1.635*	-2.160	-0.090	-1.575
4-year	-1.182	-1.573	-0.144*	-2.275
6-year	-1.290	-1.567	-0.125*	-2.101
8-year	-1.065	-1.156	-0.233*	-3.027
10-year	-1.352	-1.309	-0.250*	-2.956
Low sprawl counties defined by SPRAWL \geq 115				
<i>Movers</i>				
2-year	0.233	0.293	-0.245*	-1.973
4-year	-1.446	-1.180	-0.210*	-2.086
6-year	-0.502	-0.479	-0.281*	-2.384
8-year	-0.711	-0.820	-0.105	-1.334
10-year	-0.666	-0.829	-0.027	-0.308
<i>Non-movers</i>				
2-year	-1.439*	-2.018	-0.153*	-2.545
4-year	-0.977	-1.451	-0.201*	-2.974
6-year	-0.995	-1.343	-0.193*	-2.957
8-year	-0.659	-0.836	-0.329*	-3.593
10-year	-0.915	-1.022	-0.366*	-3.501
Low sprawl counties defined by SPRAWL \geq 125				
<i>Movers</i>				
2-year	0.296	0.396	-0.351*	-2.152
4-year	-0.786	-0.950	-0.379*	-2.464
6-year	-0.221	-0.264	-0.506*	-2.601
8-year	-0.286	-0.348	-0.322*	-2.411
10-year	-0.325	-0.370	-0.225	-1.687
<i>Non-movers</i>				
2-year	-1.692*	-2.115	-0.108	-1.746
4-year	-1.395	-1.623	-0.128*	-1.975
6-year	-1.480	-1.638	-0.088	-1.474
8-year	-1.259	-1.292	-0.162*	-2.301
10-year	-1.478	-1.413	-0.197*	-2.479

Note: * indicates the coefficient estimate is significantly different from zero at the 5% level

FIGURE 1. Distribution of Sprawl Index

