Effect of large-scale social interactions on body weight

M. Christopher Auld
Department of Economics
University of Calgary
auld@ucalgary.ca

Lisa M. Powell
Institute for Health Research and Policy
University of Illinois at Chicago
powelll@uic.edu

Abstract

The rapid rise in obesity in the United States and elsewhere in recent decades remains a puzzle. We study the individual and social determinants of body weight using the Behavioral Risk Factor Surveillance System waves 1997 through 2002, correcting self-reported height and weight using the methods of Burkhauser and Cawley (2008), and limiting attention to interactions at the county and state levels. We demonstrate that dispersion in body weight across time and space in the U.S. is not clearly excessive, and that much of this variation can be attributed to observable individual and regional characteristics. Econometric models exploiting variants of methods proposed by Glaeser et al. (2002), fixed effects, instrumental variable and split–sample instrumental variable, and other methods to address endogeneity yield mixed evidence for some moderately strong interactions on BMI, obesity, and morbid obesity, and scant evidence for interactions on underweight.

We thank Carol Bao and Jenny Zhao for excellent research assistance and gratefully acknowledge many helpful comments from seminar participants at the University of Illinois at Chicago, the University of Waterloo, the 2008 Canadian Health Economics Study Group and the 2008 American Society of Health Economists meetings. The Cooperative State Research, Education, and Extension Service (CHESS) of the USDA provided financial support. Auld also thanks the Center for Health and the Social Sciences (CHESS) and the Alberta Heritage Foundation for Medical Research (AHFMR) for support.

Preliminary draft. Comments welcome.

This version: December 2008

Effect of large-scale social interactions on body weight

1 Introduction.

A vast literature spanning disciplines investigates the obesity *epidemic*. The metaphor that obesity is an epidemic, spreading from person to person, is ubiquitous but rarely closely examined. In this paper we present econometric evidence on the causes of body weight focussing on the effect of one's neighbors' body weights on one's own body weight. Using roughly half a million observations drawn from the Behavioral Risk Surveillance System (BRFSS) from 1997 through 2002, we investigate to what extent the obesity epidemic is actually an epidemic. Specifically, our goal is to assess what can be learned about the private and social causes of body weight from standard econometric models, where 'social' for our purposes means exogenous and endogenous contextual effects at the county and state level.

In the New Social Interactions literature[1], a person's behavior depends on the behavior of those in some reference group, such as people in their neighborhood, workplace, or classroom. When such interactions are strong enough they can lead to excess variation across reference groups, the possibility of multiple equilibria, and a "social multiplier" effect in which a small change in an exogenous incentive works both directly and through an effect on the distribution of behavior in the reference group. In our work, we explore whether the rapid rise in obesity in the United States and elsewhere in response to apparently small changes in incentives such as food prices can be attributed to such a multiplier and whether variation across regions is larger than we would expect in the absence of social interaction effects on body weight.

Economists' explanations of changes in obesity rates over time have largely centered

[1]See Manski (2000) or Zanella (2004) for surveys of this literature.

on private responses to changes in incentives. Lakdawalla and Philipson (2002) argue that changes in the price of a calorie and in the opportunity cost of burning a calorie explain changes in obesity. This explanation is echoed in subsequent work by economists.[2] Research in other disciplines, such as evolutionary psychology and public health, is consilient with such explanations.[3] Social interactions complement these literatures by explaining features of the data, such as excess variance within regions or racial groups, which are difficult to explain in terms of root causes alone.

A small literature empirically examines the social determination of body weight. Costa-Font and Gil (2004) present evidence that self–image affects body weight and argue that this result suggests social influences drive obesity. Burke and Heilland (2005) present a version of the quadratic conformer model in which deviations from a desired body weight — which may depend on others' body weights — reduce utility. The authors demonstrate that a version of the model calibrated to NHANES data is able to reproduce many of the stylized facts describing changes in the distribution of body weight over time. Christakis and Fowler (2007) found evidence that one's friends cause one's body weight using the Add Health surveys, however, Cohen-Cole and Fletcher (2008) argue that this result is an artifact of model misspecification. Cohen-Cole and Fletcher (2008) report estimates of standard econometric models using data similar to Christakis and Fowler (2007) and show that there is at best weak evidence of endogenous social interactions. Summarizing, while the idea of endogenous social interactions driving weight outcomes has recently sparked attention in the literature, the evidence for such interactions remains elusive.

In this paper, we provide further evidence on the extent and nature of the effect of social influences on body weight. We study interactions at a broader level than previous empirical research. Instead of estimating the effect of one's friends weight on one's own weight, we attempt to estimate the effect of the body weight of people in the county or State in which one lives on one's own weight. This approach has benefits and drawbacks relative to examining interactions at lower levels of aggregation. The primary drawback is that it is plausible that interactions are stronger within narrowly defined social groups such as friendship networks. However, individuals may expect that their body weight will be a factor in future friendships, employment relations, and marriage

[2]Such as Chou et al. (2004), Marcelli et al. (2004), MacInnis and Rausser (2005), and Powell et al. (2007).

[3]On evolutionary reasoning and body weight see Smith (2002). See Drewnowski and Specter (2004) for a review of the relevant public health literature.

markets, suggesting that the body weights of people in the community may also have large effects on weight. These large-scale effects may be easier to reveal since they may be less confounded by the severe selection effects which generate correlations between the characteristics of friends.

Our primary source of data are the 1997 through 2002 waves of the Behavioral Risk Factor Surveillance System (BRFSS). These data are repeated cross-sections, which limits our analysis since we cannot track respondents over time. More than offsetting that downside is the enormous sample size of the BRFSS. Many of our models use sample means and other statistics calculated from the data as covariates, and the the large BRFSS samples allow us to adequately estimate those statistics, even after stratification by sex, race, and age. We use almost half a million observations from the BRFSS surveys to estimate these aggregate outcomes, and estimate models which do not require longitudinal observations entirely from these very large surveys.

We use three distinct strategies in attempting to detect the fingerprints left by social interactions. After discussing suggestive evidence for and against social interactions from descriptive statistics and variance decompositions, we implement variants suitable for repeated cross–sectional data of methods developed by Glaeser and Scheinkman (2001), Glaeser et al. (2002), and Graham and Hahn (2005), which infer the presence of social interactions from contrasts between regression coefficients at the individual and aggregate levels. Our final approach is to instrument for aggregate body weight in conventional linear–in–means quadratic conformer models, using subsets of aggregate characteristics as excluded instruments. Where appropriate, we correct for sampling–induced measurement error in regional aggregates using a split–sample instrumental variable technique. These strategies generates inconsistent estimates if strong auxiliary assumptions fail, so we interpret our estimates as providing only suggestive evidence.

Our results do not provide strong support for the hypothesis that social interactions—at this level of aggregation and over moderate time spans—are the missing piece of obesity puzzle.

1.1 Some suggestive evidence.

In this section, we remark on evidence from results from the literature which have implications for social interactions in body weight.

First, it is difficult to explain changes in obesity rates over time with changes in food prices. In Figure 1 we display obesity rates among adults constructed from NHANES data and a price index of food and beverages for roughly the last three decades. Real food prices are much higher than trend during the 1970s and then continue falling slowly for the next twenty years. This contrasts with obesity rates, which rise slowly during the 1960s and 1970s and then increase at an increasing rate for the next twenty years. Casual empiricism here suggests that social interactions may have caused the dramatic increase in obesity we observe over time.

A second observation suggesting the presence of interactions is the claim that obesity rates also exhibit high variance across countries and within other countries than the United States. For example, Vanasse et al. (2006) find that obesity rates in Canada vary from 6% to 47% across health regions. Obesity rates also vary across countries in a pattern with no obvious fundamental cause. For example, the adult female obesity rate in France is 7% compared with 15% in Portugal, 20% in the United Kingdom, and 35% in Greece.(Elmadfa and Weichselbaum, 2005).

These observations provide evidence in favor of social interactions. Relatively shallow trends in food prices are difficult to reconcile with geometrically increasing obesity rates, suggesting social interactions may be multiplying the effects of changes in fundamentals such as food prices. Consistent with that observation, obesity rates reportedly vary across regions markedly. Yet we would also observe these patterns if fundamental causes of obesity vary across time and space. We now turn to econometric models of body weight across the United States.

2 A simple model of social interactions and body weight.

In this section we briefly sketch a standard simple model of social interactions to ground the empirical work to follow.

A variety of direct and indirect evidence suggests that people are keenly aware of how changes in their weight are perceived by others. Averett and Korenman (1996) find that self–esteem decreases with body weight among white but not among black women, for example. Burke and Heilland (2005) present evidence that both mean and desired body weights among adult Americans have increased from 1994 to 2002, suggesting that perhaps rightward shifts in the distribution of actual weight are interdependent with similar shifts in the distribution of social norms with respect to weight. The idea is that the cost of body weight at the margin decreases as the distribution of body weight in one's reference group shifts to the right. Such preferences may arise for many reasons, including mating market phenomena in which potential partners are judged by body weight, discrimination in the labor market, or simply because body weight is subject to social norms.[4]

Models in which the behavior of people in one's reference group affects one's own behavior are developed in numerous places. Following Glaeser and Scheinkman (2001), consider a population arranged in G non-overlapping groups each of size n_{gt}. Each person i decides how many net calories to consume in period t, c_{it}. We ignore body weight dynamics and assume that body weight is contemporaneously determined by an additively separable function of c_{it} and an individual–specific component μ_i, $w_{it} = \mu_i + c_{it}$. Choosing c_{it} is, given μ_i, equivalent to choosing body weight and we adopt that notation. The costs and benefits of body weight depend on the private effects of weight such as on health, social effects such as stigma or marriage market consequences, and the opportunity cost of changes in weight in terms of other goods and services. These goals are represented by an indirect utility function which takes the form,

$$U(w_{igt},\ E[w_{igt}]\ ) = \alpha_{igt} w_{igt} - \phi E_g[w_{igt}] w_{igt} - (1/2) w_{igt}^2 \tag{1}$$

[4] Tovee et al. (1999), for example, find that BMI is the major predictor of female physical attractiveness. Cawley (2004) finds that obesity is associated with roughly 7% lower wages, although Norton and Han (2007) find no effect using genotypic variation as instruments.

where α_{igt} represents the marginal private benefit of body weight ignoring endogenous social influences, ϕ measures the strength of endogenous social interactions, $E[w_{gt}]$ is the expected mean body weight in group g at time t. Optimal responses then satisfy,

$$w^*_{igt} = \alpha_{igt} + \phi E_g[w_{igt}] \tag{2}$$

Here, θ represents the strength of endogenous social interactions: higher values of θ induce the agent to behave more like people in his reference group. Impose rational expectations by applying the expectation operator at the group–time level to find the social equilibrium,

$$E[w_{gt}] = \left[\frac{E(\alpha_{igt})}{1-\phi}\right] \tag{3}$$

Substitute the social equilibrium back into (2) to find

$$w^*_{igt} = \alpha_{igt} + \left[\frac{\phi}{1-\phi}E(\alpha_{igt})\right] \tag{4}$$

We will assume that all groups are in social equilibrium at all times, so that equation (4) characterizes the data. Further, we will adopt this framework unaltered even when w denotes a binary weight outcome, following Glaeser et al. (2002).

3 Econometric strategies.

Since estimating endogenous social interaction effects is notoriously difficult, our goal in the econometric work is to provide estimates using multiple empirical strategies.

Manski (1993) emphasizes the problems that arise in estimating endogenous interactions. In Manski's terminology, which has become standard, an **endogenous** interaction refers to simultaneous determination of an outcome of interest, an **exogenous** interaction as a direct effect of other's characteristics on the outcome of interest, and **correlated** effects as unobserved group–level influences on the outcome of interest. Generally, linear social interactions will fail to obtain asymptotic identification without restrictions on at least some of these types of interaction.

In the case of body weight these concerns play out as follows. If we observe that features of the distribution of BMI vary with region, we cannot conclude that neighbours'

BMIs are causally interdependent unless we can rule out other mechanisms which induce people who live in the same region to have similar weights. It may be that there are features of the region which commonly privately affect people, for example, fast food may be relatively inexpensive and plentiful. It may be that BMI varies with the socioeconomic composition of the region, or some other feature of regional demographics. And it may be that people with similar BMIs tend to prefer the same regions. We are able to control for some of these regional effects statistically, but not all of them.

These econometric difficulties are formidable. As Moffitt (2001) emphasizes, even under random assignment of individuals to regions, all of the parameters are not identified due to the possibility of unobserved regional influences on weight, for example, a price we do not observe which affects body weight. We do not claim to be able to surmount all of these difficulties; a "smoking gun" is unlikely to be found. Instead, we work with a variety of econometric approaches.

3.1 Selecting reference groups.

The relevant reference group a person uses in evaluating the social costs of changes in weight is ambiguous. The finest level of geographic disaggregation we have in the data is counties, and therefore we assume that people compare themselves to others within their county or their state. We report both sets of estimates for most models since we are agnostic over which geographic level is more appropriate. We allow for spillovers at the national level, under the assumption that these spillovers are constant across regions, through the inclusion of year fixed effects in most models. Such spillover effects do not affect the magnitude of the strength of our social interaction estimates.

3.2 Inferring interactions from excess variance across regions.

We begin with a simple exploratory approach suggested by Glaeser and Scheinkman (2001). Suppose that there are no social interactions or regional causes of body weights, so that body weights are distributed identically and independently across people and regions. If $V(w_{igt}) = \sigma^2$ and we observe N_g randomly selected individuals from group g in time t, then $V(\bar{W}_{gt}) = \sigma^2/N_g$. Notice that the number of people we observe at time t in group g is not the number of people who are interacting within groups—we do not

assume that social interactions only occur within BRFSS respondents.

Now suppose that there are endogenous social interactions, but continue to assume that there are no other regional determinants of weight. If $w_{igt} = \alpha_{igt} + \frac{1}{1-\phi}W_{gt}$, then in equilibrium $V(\bar{W}_g t) = \sigma^2/(N_g(1-\phi)^2)$. Since we can predict the variance across regions we would observe in the absence of interactions, and we can estimate the realized dispersion across regions, we can infer the strength of endogenous interactions. Let $\tilde{\sigma}^2$ denote the population predicted variance under no social interactions or unobserved influences on weight, and let $\ddot{\sigma}^2$ denote the probability limit of the variance we estimate from the data. Then

$$\frac{\ddot{\sigma}}{\tilde{\sigma}} = \frac{1}{1-\phi}. \tag{5}$$

Of course, any individual or regional characteristics which affect weight will tend to bias this method in the direction of finding social interactions. We reduce this bias by repeating the exercise conditional on observed individual and regional characteristics. We repeat the steps described above, except we residualize w_{igt} with respect to these characteristics, and calculate sample means of these residualized weights. Unobserved heterogeneity will, however, still tend to bias the result, and we should expect to find evidence of endogenous interactions even if none are present.

3.3 Inferring interactions from within versus between estimates.

To empirically implement equation (4), we begin by specifying:

$$\alpha_{igt} = X_{igt}\beta + P_{gt}\gamma + \lambda_g + u_{igt} \tag{6}$$

where X_{igt} are individual level covariates including age, sex, income, employment, marital status, and education, P_{gt} are regional level determinants including prices and densities of various food stores, λ_g denotes unobserved time–invariant group level determinants, and u_{igt} is an idiosyncratic error term. Then (3) and (4) imply that, in equilibrium,

$$w_{igt} = X_{igt}\beta + E[X_{igt}]\left[\frac{\phi\beta}{1-\phi}\right] + P_{gt}\left[\frac{\gamma}{1-\phi}\right] + \lambda_g^* \tag{7}$$

where $\lambda_g^* = \lambda_g(1-\phi)^{-1}$, which is a panel variant of (Manski, 1993) pure endogenous interactions model. That is, following much of the literature, we assume for the moment

that there are no contextual effects. If we take expectations at the group level and solve for equilibrium aggregate weights, we find (expressed as sample analogs),

$$\bar{W}_{gt} = \bar{X}_{gt}\left[\frac{\beta}{1-\phi}\right] + P_{gt}\left[\frac{\gamma}{1-\phi}\right] + \lambda_g^* \tag{8}$$

Notice that a change in one of the variables in X_{igt}, holding all else equal induces a change in i's weight measured by β, but if all members of i's region experience an increase in one of their X's mean weight in the region rises in equilibrium by $\frac{\beta}{1-\phi}$, by equation (8). The social multiplier is, then, $(1-\phi)^{-1}$.

Graham and Hahn (2005) consider the cross–sectional analog of this expression and suggest estimating β using the within variation and $\frac{\beta\phi}{1-\phi}$ using between variation. Replacing $E[X_{igt}]$ with its sample analog $\bar{X}_{gt}$, sample means of the covariates at the region level, introduces a measurement error problem. The estimable model is

$$w_{igt} = X_{igt}\beta + \bar{X}_{gt}\left[\frac{\phi\beta}{1-\phi}\right] + P_{gt}\gamma + \lambda_g + \nu_{igt} \tag{9}$$

where

$$\nu_{igt} = \left[\bar{X} - E(X_{gt})\right]\left[\frac{\phi\beta}{1-\phi}\right] + u_{igt} \tag{10}$$

In a cross–section, conventional fixed effects estimators cannot be applied to equation (9) because the within transformation eliminates $\bar{X}_{gt}$. Graham and Hahn (2005) suggest using the approach of Hausman and Taylor (1981) to address this difficulty, but with repeated cross-sections the usual fixed effects estimator can be deployed. However, the fixed effects estimator will be attenuated due to sampling–induced measurement error in the regional means of characteristics $\bar{X}$, a problem we correct with a split–sample instrumental variable approach described below in section 3.5.

We implement these models as follows. We first estimate equation the individual level model (7) to obtain estimates of $\hat{\beta}$. The second step is to estimate model (8), yielding estimates of $\beta(1-\phi)^{-1}$. In the final step we combine these two sets of estimates by regressing the predicted values from the aggregated model (8) on the aggregated predicted values from the individual level model, net of the predicted effects of the aggregate level means. That is, if $\hat{W}$ denotes estimates of $\bar{X}_{gt}(\frac{\beta}{1-\phi})$ from estimation of equation (8) and $\hat{W}$ denotes estimates of $X_{igt}\beta$ from estimation of equation (7), we use

the auxiliary regression

$$\hat{W} = \alpha\,[\hat{w}] + \text{noise}, \tag{11}$$

to estimate the social multiplier:

$$\frac{\hat{1}}{1-\phi} = \hat{\alpha}. \tag{12}$$

In the first two steps we use fixed effects to remove time–invariant unobserved regional effects and split–sample instrumental variable methods to correct for measurement error in the regional means. If we are able to generate consistent estimates, straightforward calculations show the slope estimate in the third stage model converges to the social multiplier.

In contrast to the excess variance approach, we expect this approach to yield estimates which are biased against finding endogenous interactions when we use regional fixed effects, due to the attenuation bias measurement error introduces. On the other hand, if we do not use fixed effects, we expect the estimates to be biased in favour of finding endogenous interactions, due to omitted heterogeneity bias.

3.4 Instrumental variable approaches.

In this strategy, we attempt to directly estimate the causal effect of group level outcomes on individual outcomes, addressing endogeneity using instrumental variable methods. Equations (7) and (8) imply that the equilibrium structural weight equation takes the form

$$w_{igt} = X_{igt}\beta + P_{gt}\gamma + \phi E[w_{igt}] + \lambda_g, \tag{13}$$

which may be directly estimated, after replacing $E[w]$ with its sample analog, using means of the covariates as instruments for $E[w]$. Mean weight outcomes are validly excluded from equation (13) if the assumption that there are no contextual effects holds. The idea is that the characteristics of one's neighbours can instrument neighbours' body weights if those characteristics affect own body weights but do not directly affect other's weights.

We do not assume that the entire vector of regional means are excludable. Rather, in all models we assume that regional mean income and the standard deviation of income can directly effect an individual's weight even after we hold the individual's own income

constant. The distribution of income may proxy many unobserved regional characteristics, so leaving it in the equation of interest may remove or mitigate bias from such unobservables.

Despite the *ad hoc* exclusion restrictions grounding this strategy, we believe it is of value. Particularly after sweeping out time-invariant unobserved heterogeneity with fixed effects or differencing, inferring social effects by studying the relationships between individual body weight and predictors of aggregate body weight seems a viable, if not conclusive, approach.

In contrast to the previous two approaches, we have no *a priori* expectations over the sign of the biases of our IV estimators.

3.5 IV estimation methods.

To implement this approach, we select a two–step GMM estimator which is asymptotically efficient in the presence of arbitrary skedastic patterns (we use Stata's **ivreg2** and **xtivreg2** to estimate these models). All of our models are linear probability specifications. We estimate single equation models by ordinary least squares. Covariance matrix estimates are adjusted for clustering on regions.

The dependent variables in most of our models are binary, indicators for obesity, morbid obesity and underweight, but we limit attention to linear methods for several reasons. We have roughly a half million observations, which renders some computationally demanding nonlinear methods infeasible. Further, we rely heavily on fixed effects approaches which are, again, computationally infeasible with large samples, and introduce incidental parameters difficulties. We make no attempt to use nonlinearities to identify the strength of social interactions, which often use functional form assumptions, or require knowing the number of individuals in the respondent's reference group to solve for social equilibria.

In some models we correct for measurement error in the regional means using an instrumental variable strategy suggested by Bloom et al. (2008), which is in turn an application of the split–sample instrumental variable estimator proposed by Angrist and Krueger (1995). Rather than calculating region–year means $\bar{X}_{gt}$, the sample within each year and region is randomly split into two groups 1 and 2, and the submeans $\bar{X}^1_{gt}$ and

$\bar{X}^2_{gt}$ are calculated. By virtue of random assignment to groups, the measurement error in the two submeans are independent, and therefore an instrumental variable estimator constructed using one submean as an instrument for the other is consistent by textbook arguments. We implement this procedure using the GMM estimator described above.

4 The BRFSS and other data.

Our primary data source is several waves of the Behavioral Risk Factor Surveillance System (BRFSS). Food prices were obtained from the American Chamber of Commerce Researchers Association (ACCRA) Cost of Living Index reports that contain quarterly information on prices across more than 300 US cities annually These price data are matched based on the closest city match available in the ACCRA data. Restaurant and supermarket outlet density data were obtained from business lists developed by Dun and Bradstreet (D&B).

4.1 Food prices and access measures.

A food at home food price measure is derived from thirteen general grocery food prices available in the ACCRA data constructed similarly to the food at home price measure used by Chou et al. (2004). This food at home price index was weighted using BLS expenditure weights and deflated by the CPI. A number of outlet density measures available by location (county and state) and Standard Industry Classification (SIC) codes in the D&B data are used to control for other aspects of the built environment that may affect weight outcomes. First, fast food restaurant outlet density is defined by the full set of primary 8-digit SIC codes that fall under Fast food restaurants and stands excluding coffee shops and including the two primary 8-digit SIC codes for chain and independent pizzerias. Second, a measure of full-service restaurants are defined as the number of total number of "Eating Places" minus fast food restaurants and excluding coffee shops, ice cream, soft drink and soda fountain stands, caterers, and contract food services. Third, supermarket outlet density is defined based information on food store outlets available in the D&B data set was pulled at the 6-digit SIC code level. Census population measures are used to derive per 10,000 capita measures of outlet density availability.

4.2 BRFSS, 1997–2002.

The BRFSS is an annual telephone survey of adults, including (for our sample window) all U.S. states. We limit attention to the 1997 through 2002 waves due to data limitations on the food store access measures described above. From these data, we draw information on self-reported height and weight, which we use to construct variation weight measures as described below. Other covariates which we make use of include age, sex, ethnicity, smoking status, household income, marital status, employment status, and education. Income is calculated using interval midpoints, and adjusted for inflation using the all-items CPI. We note that many of these variables may be endogenous to body weight, but our concern is controlling for population heterogeneity rather than causality at the micro level. Further information on the BRFSS sampling design can be found in Nelson et al. (2001).

For estimation, we drop observations on individuals in county–year cells with less than 100 respondents, as in such small cells we cannot credibly estimate aggregate outcomes. We also remove observations if the respondent's BMI is lower than 10 or greater than 60 in order to prevent outliers from driving results. Finally, we drop observations with missing data. We are left with an estimation sample of 450,391 observations.

Summary statistics for the estimation sample are displayed in Table 1.

4.3 Body weight outcomes.

Our main measure of body weight is body mass index (BMI), defined as weight in kilograms divided by the square of height in meters. These data are self–reported and subject to measurement error. To mitigate this concern, we implement the methods described in Burkhauser and Cawley (2008). The authors draw on data which includes both self-reported and measured height and weight and report estimates of regression models predicting measured values from observed values. We expect this procedure will have little or no effect on our models in which BMI is the outcome of interest, as the prediction equations are primarily based on age, race, and sex, covariates also in our models. However, when the outcome of interest is obesity or various other facets of the distribution of BMI, Burkhauser and Cawley (2008) show that classification of respondents (which is done unconditionally on the covariates) is sensitive to the use of

self-reported data.

We find correlations between self-reported and corrected height and weight of 99% and 98%, respectively. We use the corrected height and weight measures to estimate corrected BMI, then use corrected BMI to classify individuals as underweight (BMI$\leq$18.5), normal weight (18.5 $<$ BMI $\leq$ 30), obese (BMI $>$ 30) or morbidly obese (BMI $>$ 35). Morbid obesity is conventionally defined as a BMI of greater than 40, or BMI greater than 35 in the presence of a significant comorbidity. Our definition is then nonstandard in that we use 35 is the threshold even though we do not observe comorbidity status. We choose this threshold because we wish to study weight outcomes which may substantially affect health for a moderately large proportion of people, and few people have BMIs exceeding 40.

We study interactions on both BMI and these categorical weight outcomes because social causes of body weight may operate differently at different points in the distribution of BMI. Notice that these are not mutually exclusive categories, as someone who is morbidly obese is also necessarily obese by these definitions. There were fairly large changes in classification into underweight, obese, and morbidly obese categories across self-reported and corrected BMI measures. Based on corrected height and weight, Table 1 shows that overall mean BMI is slightly less than 26, 18% of respondents are classified as obese, 6% are morbidly obese, and 2.3% are underweight.

5 Econometric results.

5.1 Variation in weight outcomes over time.

Figure 2 displays mean weight outcomes over time. The solid lines in each panel show sample means in each year, showing mean BMI has risen and proportions obese and morbidly obese have rose while proportion underweight fell. We investigated whether these changes could be attributable to changes in demographic characteristics by residualizing with respect to the individual characteristics, prices and access measures, and county–level means as displayed in Table 1. The dashed lines in each panel show hypothetical changes in mean weight outcomes for a population with identical characteristics. Since the dashed lines in each case show larger increases in body weight (or smaller decreases

in underweight), we conclude that variation over time in weight outcomes cannot be explained by changes in observed characteristics over time, in fact, changes in characteristics obscure some of the increase in weight which would have otherwise occurred. These results lend some credibility to endogenous social interactions as an explanation.

5.2 Is there excess variance in body weight outcomes across regions?

Dramatic variation across regions in body weight outcomes which cannot be readily attributed to variation in fundamentals is evidence in favour of social interactions. Is there such excess variation across U.S. counties and States?

We first assessed the partial R^2 attributable to regional level fixed effects, and calculated heteroskedasticity–robust F-statistics against the null that the regional effects are jointly zero, varying the set of included covariates (we do not display the results to save space). When no covariates are included, the proportion of variation in BMI attributable to the State fixed effects is 0.46%. This proportion falls to 0.032% if we hold individual level demographics constant and 0.0025% if we hold constant all observed individual characteristics. These results suggest that very little of the variation in body weight across people in the United States can be attributed to regional influences, which is suggestive evidence against either endogenous social interactions or sorting on characteristics correlated with body weight. Further, roughly one third of the variation which can be attributed to regional influences is explained by basic individual level demographic characteristics (age, sex, and ethnicity). The associated F–statistics are large enough to easily reject the null that the regional effects are jointly zero, even after conditioning on the full set of individual characteristics and regional means. However, the F–statistic of 44.1 with no covariates falls to 36.3 with basic demographics and 25.5 with the full set of individual characteristics. We interpret these estimates as suggestive evidence against the hypothesis that there are economically significant regional influences on the body weight of otherwise observationally equivalent people.

We then assessed observed variation in regional means relative to the variation we would observe through sampling noise if there are no regional effects. Figure 2 shows kernel density estimates of the mean BMI of BRFSS respondents aggregated at the state

level (left panel) and county level (right panel)[5]. In each figure, the solid density is the observed density of BMI residualized with respect to observed individual and regional characteristics. The dashed density is the distribution of regional means for a hypothetical population which is identically and independently distributed across the U.S. with the same individual variance and regional sample sizes as the dataset. The distributions would overlap (up to sampling noise) if observed characteristics explained all of the excess regional variation in body weight across the U.S. At both the state and county levels, the hypothetical means display less variation than means calculated from the data, but the distributions are strikingly similar. That is, observed variation is only modestly greater than we would expect to observe if there were no unobserved determinants of BMI, including endogenous social interactions, which vary systematically across regions. This simple analysis is strongly suggestive against powerful endogenous social interactions.

More formally, Table 2 displays estimates of the social multiplier calculated using equation (5). Notice first that for all weight outcomes, and at either the county or state levels, individual level variation in weight outcomes dwarfs variation across regions. The estimates in Panel A do not correct for observed influences on weight and therefore are likely to be heavily biased in favor of finding social interactions. At the county level, the implied social multipliers for BMI, obesity, and morbid obesity are about 3, 2.5, and 2, respectively. At the State level these estimates rise to about 4.5, 4, and 3. These estimates are roughly half the analogous estimates for smoking reported by Cutler and Glaeser (2007), but they are still large. In contrast, at both the county and State levels the estimated multiplier on underweight is roughly unity. Even when we do not correct for observed influences on weight, we do not find even suggestive evidence that underweight is determined by endogenous social interactions.

In the bottom panel, notice that correcting for observed influences on weight markedly reduces the implied social multipliers. At the county level, BMI, obesity, and morbid obesity are estimated to be subject to multipliers of about 1.3, 1.2, and 1.1. At the state level, these effects rise to about 2, 1.6, and 1.5. Again, underweight does not appear to be interacted upon. The lack of excess variation across regions after controlling for observable characteristics can also be noted in Table 4, which displays estimates of weight models including regional fixed effects. All else equal, the proportion of variance attributable to the region effects varies from 0.12% in the underweight and obesity mod-

[5]Kernel density estimates use the Epanechnikov kernel and bandwidth chosen by cross-validation.

els to 0.28% in the BMI model. In these models, the null that all of the region effects are zero can be safely rejected, but that is largely an artifact of the large sample size: the region effects have no *substantive* effect on weight outcomes.

From this exercise we conclude that underweight certainly does not display excess variation across regions. We find substantial multipliers for mean BMI and the obesity outcomes when we do not correct for observed influences on weight, but much of that excess variation is attributable to our fairly limited set of individual and regional covariates. If we compare regions comprised of statistically identical individuals, we find modest evidence for social multipliers between about 1.2 and 2. Mean BMI consistently displays more unexplained variation across regions than either obese or morbidly obese status.

5.3 Comparing aggregate outcomes to predicted outcomes from individual regressions.

Table 3 displays estimates of body weight social multipliers estimated using the methods described in Section 3.3. In the top panel, we consider interactions at the county level. When we use OLS without fixed effects in either the individual or aggregate level models, the implied social multiplier on BMI is 1.032, trivially different from unity. Similarly small multipliers are recovered for obesity and morbid obesity, and the estimate on underweight is less than one at 0.79, that is, as the proportion of individuals around a given individual who are underweight increases, the probability the given individual is underweight falls. The bias of these estimates is unclear: the estimates are biased up because unobserved regional influences on weight will appear to be social multipliers, and biased down because of attenuation resulting from measurement error in the regional means.

In the next row, we correct for measurement error using split–sample instrumental variable estimation. The multipliers on BMI, obesity, and morbid obesity rise moderately since the attenuation bias has been removed, but are still quite modest at 1.14, 1.27, and 1.34. In the third row, we also remove the upward bias resulting from unobserved regional influences on weight, and the resulting estimates are near or below unity.

Turning to the State level estimates in the bottom panel of Table 3, we find similar

results to those we found at the county level, although the upward bias resulting from unobserved State level influences appears to be lower in magnitude than at the county level. For example, the estimated BMI social multiplier is 1.05 when we use OLS, 1.06 when we correct for attenuation bias using split–sample instrumental variables, and 1.075 when we both correct for measurement error and unobserved state level influences on weight. We find comparable results for obesity and morbid obesity, and again find multipliers less than unity for underweight.

These results suggest that the body weight social multiplier is probably small, or that there is no no-trivial multiplier.

5.4 IV estimates of the effect of aggregate outcomes on individual body weights.

Finally, we consider estimating the structural body weight equation (13) directly using means of the covariates as instruments for the regional weight outcomes. Before presenting structural estimates, in Table 4 we first present estimates of reduced forms of these models, corrected for measurement error using the split–sample technique described in section 3.5. To save space, we only report estimates using State level aggregates; estimates from the county level models are qualitatively similar. Equation (9) suggests that if there exist strong endogenous social interactions, we should find that means of covariates are strongly associated with individual outcomes after holding individual–level covariates fixed. For example, if we fix a respondent's education and exogenously move him to a region in which mean education levels are higher, we would, under this strategy, infer that endogenous social interactions are driving body weight if the respondent's body weight fell. Notice that if the higher education region differs from the original region in ways which affect body weight and which we cannot statistically hold constant, our strategy will fail.

Table 4 shows that many of the individual level characteristics are highly correlated with weight outcomes. Smokers weigh less, and are less likely to be obese or morbidly obese, and are more likely to be underweight, than observationally equivalent non–smokers. Employed people are more likely to be normal weight than either under or over weight, and similarly probability of either under or over weight falls with income at a declining rate. The (undisplayed) estimates on the age and ethnicity controls

are precisely estimated and usually of the expected signs. Food prices and restaurant access are associated with weight outcomes in expected patterns: higher food prices are associated with lower weights, higher fast food outlet density predicts higher weight, higher probability of obesity and morbid obesity, and lower probability of underweight, whereas non–fast food restaurant density predicts the opposite. These associations are of similar magnitude and usually retain statistical significance when regional fixed effects are used. Unexpectedly, and in contrast to results we report in Auld and Powell (2008) for adolescents, we do not find that higher supermarket density is associated with lower probability of either overweight or underweight.

Confident that the model recovers estimates on individual level covariates similar to those in the literature, we now examine the estimates on the state level means. When we do not include regional fixed effects, we find that several regional demographic characteristics are moderately highly associated with weight outcomes. Ethnic composition statistically significantly affects each outcome, for example, holding the respondent's own ethnicity and other characteristics constant, increasing the proportion of black people in the region by ten percentage points decreases BMI by 0.13 units (t=5.43), decreases probability of obesity by 3.5 percentage points (t=1.96), decreases probability of morbid obesity by 2.1 percentage points (t=1.90), and increases probability of underweight by 0.27 units (t=3.63). More smokers are associated with higher BMI and probabilities of obesity and morbid obesity, regions with more men have lower BMI and less obesity and morbid obesity, and regions with more unmarried people are associated with lower BMI, obesity, and morbid obesity. However, none of these associations survive when fixed effects are used, suggesting that unobserved heterogeneity across regions is driving the cross-sectional results.

Various literatures have explored the idea that income inequality has a direct impact on health, a contextual effect which operates even after holding individual income fixed.[6] The estimates presented in Table 4 do not support the idea that body weight responds to such a contextual effect. Mean income never has an economically nor statistically significant effect on any body weight outcome across specifications. The dispersion of income is associated with higher BMI and obesity rates—a $10,000 increase in the standard deviation of income is associated with an increase in BMI of 0.8 units (t=6.14) and a 6.8 percentage point increase in obesity rates (t=7.2)—but these effects disappear when fixed effects are used, suggesting they are attributable to unobserved regional

[6]See Subramanian and Kawachi (2004) for a review of this topic.

influences. The rate of employment (as calculated from BRFSS respondents) has no effect on any outcome across specifications except the underweight model without fixed effects. Note that Table A-1 in the Appendix shows that dispersion in income appears to have a contextual effect on weight outcomes when we do not correct for measurement error in the regional means, so that we might incorrectly conclude such an effect exists if other contextual effects are estimated with attenuation. These results suggest that labour market outcomes do not have substantial contextual effects, even though they are are highly associated with weight outcomes at the individual level.

We note here that, unlike in Bloom et al. (2008), our results are sensitive to correction for measurement error in the regional means. In an Appendix, Table A-1 displays the same models as in Table 4, estimated using OLS and OLS with fixed effects rather than Table 4's instrumental variable method. The parameter estimates in Table A-1 tend to be substantially smaller in magnitude than those in Table 4, suggesting measurement error substantially attenuates the OLS and OLS-FE models.

In Table 5 we report estimates of structural models. Each model includes individual and aggregate level covariates as in Table 4 and full sets of year and region effects. The estimated effects of county and state level mean BMI on individual BMI are neither economically nor statistically significant. In contrast, each of the binary weight outcomes obtains statistical significance at at least one of the county or state levels. Focusing on the county level estimates, moving an individual from a region in which there are zero obese people to one in which everyone is obese would, all else equal, increase the individual's own probability of obesity by 33 percentage points (t=3.01). The analogous estimates for morbid obesity and for underweight are 40 percentage points (t=3.81) and 52 percentage points (t=3.44). Put another way, these estimates suggest that the social multipliers on obesity, morbid obesity, and underweight are about 1.5, 1.7, and 2.1.

However, we emphasize that these estimates must, even more so perhaps than those previously discussed, be interpreted with caution. Diagnostic tests against weak instruments and invalid overidentifying restrictions are displayed at the bottom of Table 5. The p–values on the J–statistic based tests of the overidentifying restrictions tell us to fail to reject the null that the exclusion restrictions are valid. However, we remain concerned that our instruments are not valid. We have little theoretical reason to suppose that demographic shifts such as changes in ethnic or educational composition around an individual have no direct effect on the individual's weight, and Dahlberg et al. (2002)

present Monte Carlo evidence that similar overidentifying restriction tests in dynamic panel models have low power in the presence of measurement error. We are not estimating dynamic models such as Dahlberg et al. (2002) study, but it seems plausible that concerns over power carry from their context to ours.

The Kleiberger–Paap F–statistic measures the ability of the excluded instruments to explain the endogenous aggregate weight outcomes (see Kleibergen (2007) for details). Across outcomes and aggregation levels, the instruments explain only modest variation in the regional weight outcomes. The F–statistics for the BMI, obesity, and morbid obesity models range from four to almost seven, while those on underweight are only slightly over two at either level of aggregation. These values are large enough to reject the hypothesis that the instruments are irrelevant at conventional significance levels, however, simulation evidence presented by Stock and Yogo (2002) suggests that they are small enough that bias and size distortions are substantial. Since the bias is in the same direction as the OLS bias, and OLS is biased in favour of finding endogenous social interactions, we should conclude that these estimates are biased up. In light of the model's poor ability to explain underweight and the previously reported results suggesting underweight is least subject to interactions, we are particularly skeptical of the estimates on underweight.

6 Conclusions.

Endogenous social interactions in body weight—a causal effect of other's body weights on one's own weight—could help explain the obesity pandemic in the Western world. This paper presents evidence on the strength of such interactions at the county and State levels in the United States from 1997 to 2002. We find variation in sociodemographic characteristics explains much of the variation in body weight across regions and that individual level predictors of weight outcomes within regions are similar to predictors across regions. This is evidence against the hypothesis that there are very substantial endogenous interactions on weight. Similarly, we infer small multipliers when we generate contrasts between individual and aggregate level weight models. After holding individual characteristics constant, we do not find that regional educational, ethnic, income, employment, or family structure measures are statistically significantly associated with individual weight outcomes when regional fixed effects are used. Finally, using re-

gional characteristics as instruments for regional weight outcomes, we find no significant multiplier on BMI but multipliers between 1.5 and 2 on underweight, obese, and morbid obesity. However, our instruments may not be valid. Finally, we find little evidence that regional macroeconomic conditions, including mean income, dispersion in income, and the employment rate, have effects on body weight after holding individual–level economic outcomes fixed. Taken together, we cautiously interpret these results as evidence against very strong social interactions on obesity, for example, the evidence suggests that obesity is not subject to social norms as strong as those driving smoking decisions.

Our analysis faces difficult limitations. We search only for contemporaneous relationships; interactions which slowly causes changes in weight over many years will not be found by our methods. Further, we cannot assess national level endogenous social interactions. We found that changes in body weight over the six year interval we study cannot be attributed to changes in observed individual level characteristics or to changes in food prices and food store access, suggesting some such unobserved process is causing changes in body weight at the national level. Further, even though we use large samples from the BRFSS cross–sections, the aggregate level outcomes we calculate are subject to sampling–induced measurement error, which will attenuate our estimates of the strength of social interactions in some models. Finally, with repeated cross-sections, we cannot control for sorting across regions on outcomes related to body weight.

References

Angrist, J. and A. Krueger (1995). Split–sample insturmental variable estimates of the return to schooling. *Journal of Business and Economic Statistics 13*, 225–235.

Auld, M. and L. Powell (2008). Economics of food energy density and adolescent body weight. Technical report. (forthcoming).

Averett, S. and S. Korenman (1996). The economic reality of the beauty myth. *Journal of Human Resources 31*(2), 304–330.

Bloom, D., D. Canning, I. Günther, and S. Linnemayr (2008). Social interactions and fertility in developing countries. Working paper, Harvard University.

Burke, M. and F. Heilland (2005). Social dynamics of obesity. CSED working paper 40, Brookings Institution.

Burkhauser, R. and J. Cawley (2008). Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics 27*, 519–529.

Cawley, J. (2004). The impact of obesity on wages. *Journal of Human Resources 39*(2), 451–74.

Chou, S.-Y., M. Grossman, and H. Saffer (2004). An economic analysis of adult obesity: Results from the behavioral risk factor surveillance system. *Journal of Health Economics 23*, 565–587.

Christakis, N. and J. Fowler (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine 357*(4), 370–379.

Cohen-Cole, E. and J. M. Fletcher (2008). Is obesity contagious?: social networks vs. environmental factors in the obesity epidemic. Quantitative Analysis Unit Working Paper QAU08-2, Federal Reserve Bank of Boston.

Costa-Font, J. and J. Gil (2004, November). Social interactions and the contemporaneous determinants of individuals' weight. *Applied Economics 36*(20), 2253–2263.

Cutler, D. M. and E. L. Glaeser (2007, October). Social interactions and smoking. NBER Working Papers 13477, National Bureau of Economic Research, Inc.

Dahlberg, M., E. Johansson, and P. Tovmo (2002, July). Power properties of the sargan test in the presence of measurement errors in dynamic panels. Working Paper Series 2002:13, Uppsala University, Department of Economics.

Drewnowski, A. and S. Specter (2004). Poverty and obesity: the role of energy density and energy costs. *American Journal of Clinical Nutrition 79*, 6–16.

Elmadfa, I. and E. Weichselbaum (2005). *European Nutrition and Health Report 2004*. Karger.

Flegal, K., M. Carroll, C. Ogden, and C. Johnson (2002). Prevalence and trends in obesity among U.S. adults 1999–2002. *Journal of the American Medical Association 288*, 1723–1727.

Glaeser, E. and J. Scheinkman (2001). Measuring social interactions. In S. Durlauf and H. P. Young (Eds.), *Social Dynamics*, pp. 83–131. Brookings.

Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (2002). The social multiplier. Harvard Institute of Economic Research Working Papers 1968, Harvard - Institute of Economic Research.

Graham, B. S. and J. Hahn (2005, July). Identification and estimation of the linear-in-means model of social interactions. *Economics Letters 88*(1), 1–6.

Hausman, J. and W. Taylor (1981). Panel data and unobservable individual effects. *Econometrica 49*(6), 1377–1398.

Kleibergen, F. (2007, July). Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics 127*(1), 181–216.

Lakdawalla, D. and T. Philipson (2002). The growth of obesity and technological change: A theoretical and empirical examination. Working paper 8946, National Bureau of Economic Research.

MacInnis, B. and R. Rausser (2005). Childhood obesity: Is there something amiss with food processing? Working paper, University of California at Riverside.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*, 531–542.

Manski, C. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives 14*(3), 115–136.

Marcelli, E., D. Cutler, S. Subramanian, and I. Kawachi (2004). Neighbourhood and metropolitan-level determinants of adult overweight and obesity in the United States. Working paper, Harvard University.

Moffitt, R. (2001). Policy interventions, low–level equilibria, and social interactions. In S. Durlauf and H. P. Young (Eds.), *Social Dynamics*, pp. 45–82. Brookings.

Nelson, D., D. Holtzman, J. Bolen, C. Stanwyck, and K. Mack (2001). Reliability and validity of measures from the behavioral risk factor surveillance system (BRFSS). *Social and Preventive Medicine 46(S)*(1), S03–S42.

Norton, E. and E. Han (2007). Genetic information, obesity, and labor market outcomes. Working paper, University of North Carolina at Chapel Hill.

Powell, L., M. Auld, F. J. Chaloupka, P. M. O'Malley, and L. Johnston (2007). Access to fast food and food prices: Relationship with fruit and vegetable consumption and overweight status among adolescents. *Advances in Health Economics and Health Services Research 17*, 23–48.

Smith, T. (2002). Obesity and nature's thumbprint: How modern waistlines can inform economic theory. Working paper, University of California, Santa Barbara.

Stock, J. H. and M. Yogo (2002, October). Testing for weak instruments in linear iv regression. Working Paper 284, National Bureau of Economic Research.

Subramanian, S. and I. Kawachi (2004). Income inequality and health: What have we learned so far? *Epidemiological Reviews 26*, 78–91.

Tovee, M. J., D. S. Maisey, J. L. Emery, and P. L. Cornelissen (1999). Visual cues to female physical attractiveness. *Proceedings: Biological Sciences 266*(1415), 211–218.

Vanasse, A., M. Demers, A. Hemiari, and J. Courteau (2006). Obesity in canada: Where and how many? *International Journal of Obesity 30*, 677–84.

Zanella, G. (2004). Social interactions and economic behavior. Working paper, University of Wisconsin–Madison.

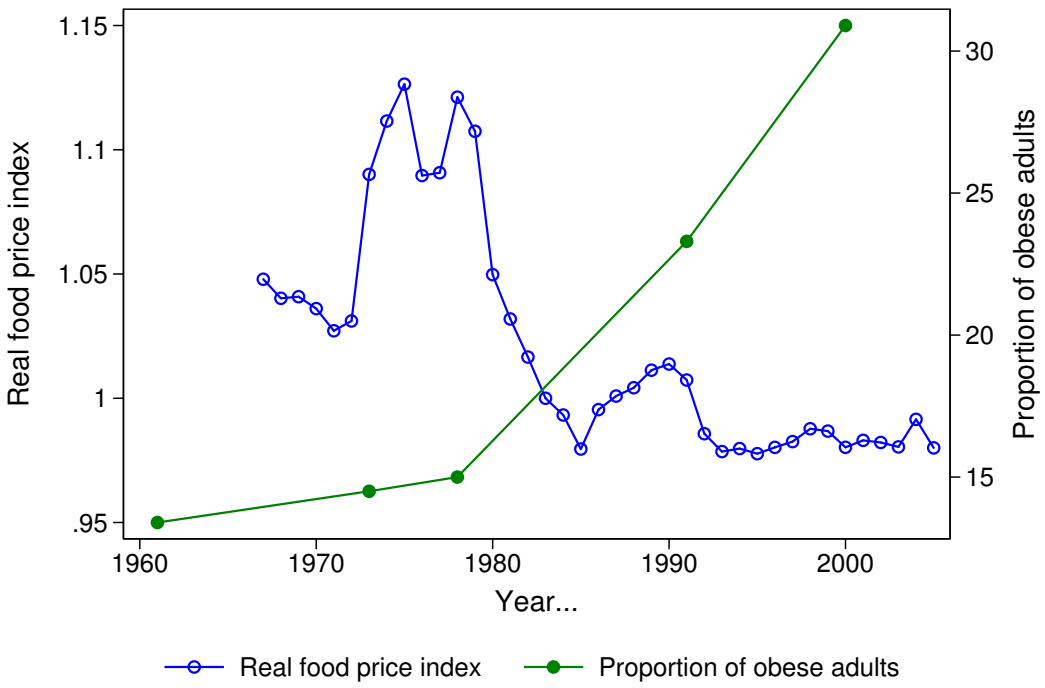

Figure 1: Adult obesity and the price of food. Obesity rates for all adults aged 20 through 74 calculated by Flegal et al. (2002) from NHANES data. Food price is the price index of foods and beverages deflated by the all–items CPI. Source: Bureau of Labor Statistics.

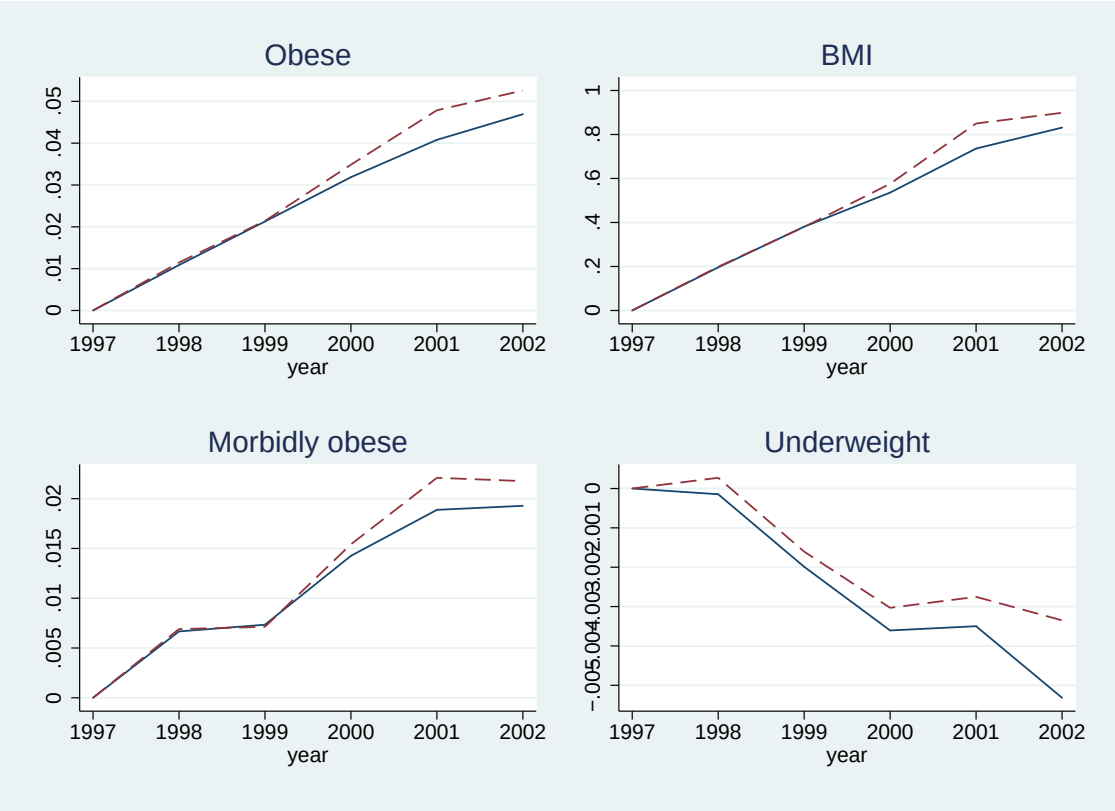

Figure 2: **Changes in weight outcomes over time.** In each panel, the solid line shows unadjusted changes in the mean of the outcome relative to 1997 and the dashed line shows changes in regression–adjusted weight outcomes holding individual and county–level mean characteristics constant.

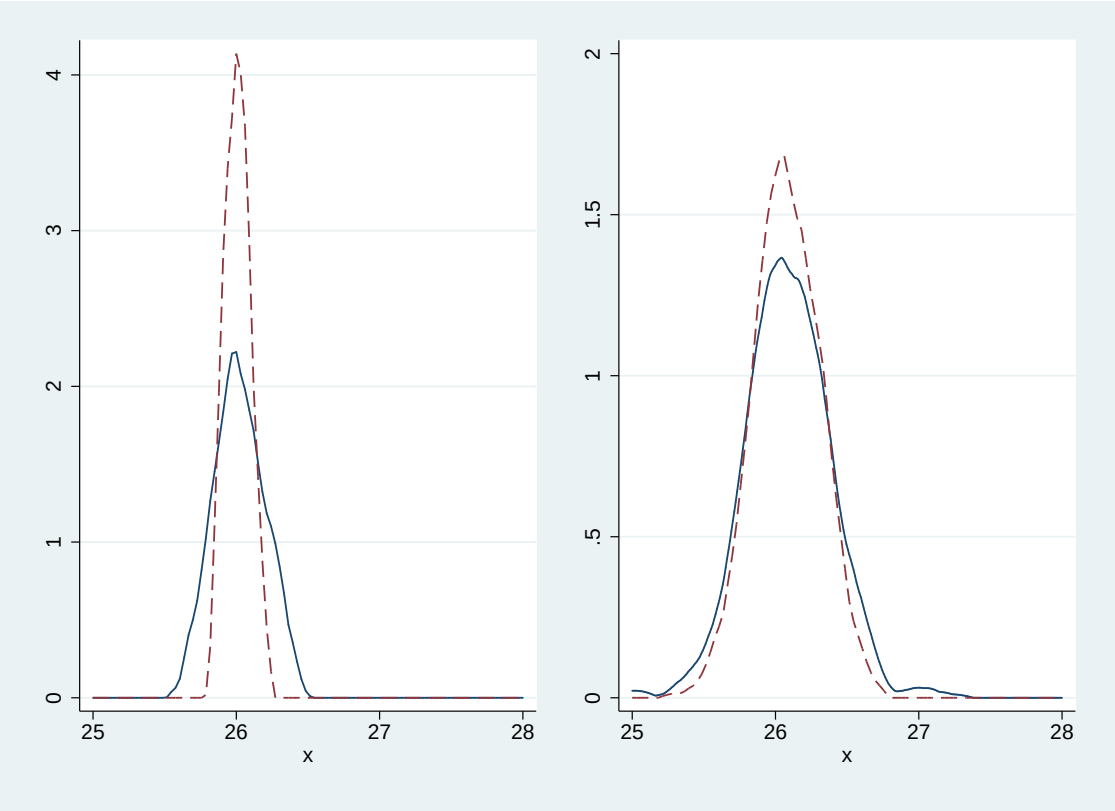

Figure 3: Observed and hypothetical distributions of regional means. Kernel density estimates. County means on left, state means on right. In each figure, the dashed line is the density of sample means from a simulated dataset in which, by construction, there are no unobserved regional influences on BMI or endogenous social interactions. The solid density is the estimated distribution from the data, residualized with respect to observable individual and regional characteristics.

variable	mean	standard deviation
Individual level variables.		
BMI	25.977	5.275
obese	0.18	0.384
morbidly obese	0.062	0.241
underweight	0.023	0.151
age	45.563	16.498
age^2	2348.121	1666.316
black	0.085	0.28
hispanic	0.069	0.253
other race	0.041	0.198
male	0.434	0.496
smoker	0.228	0.419
high school	0.283	0.45
some college	0.287	0.452
BA +	0.347	0.476
never married	0.188	0.391
dev / wid /sep	0.286	0.452
employed	0.682	0.466
income ($10,000s)	4.298	2.249
income^2	23.533	20.199
County level aggregates. *Means unless stated otherwise*		
BMI	25.928	0.658
age	46.408	2.331
black	0.084	0.108
hispanic	0.071	0.086
other race	0.041	0.03
smoker	0.222	0.039
male	0.424	0.033
high school	0.289	0.06
some college	0.282	0.043
BA+	0.336	0.09
never married	0.189	0.058
div / wid / sep	0.291	0.041
employed	0.652	0.056
income	4.292	0.587
standard dev. of income	2.168	0.131

Table 1: Summary statistics. *Continues on next page...*

...Table 1 continued.

variable	mean	standard deviation
State level aggregates. *Means unless stated otherwise.*		
BMI	26.016	0.489
age	47.176	1.578
black	0.068	0.069
hispanic	0.068	0.072
other	0.039	0.022
smoker	0.226	0.03
male	0.423	0.025
high school	0.311	0.042
some college	0.276	0.036
BA+	0.305	0.058
never married	0.172	0.033
div / wid /sep	0.292	0.027
employed	0.637	0.044
income	4.145	0.457
standard dev. of income	2.187	0.121
Prices and food store access.		
price of food at home	1.135	0.114
density of non-fast food restaurants	12.132	3.198
density of fast food "restaurants"	2.774	0.753
density of supermarkets	0.541	0.249
cigarette tax rate	47.608	29.054

Table 1: Summary statistics. Demographic variables from the Behavioral Risk Factor Surveillance System, 1997–2002. Food prices and access from ACCRA and D&B, as described in the text. Outlet densities are per 10,000 capita. N=450,391.

A. Unadjusted weight outcomes.

Group	standard deviation	Average Group size	BMI	Obese	Morbidly obese	under–weight
Individual		1	5.276	0.384	0.241	0.151
County	predicted	494	0.237	0.173	0.011	0.007
	actual		0.661	0.042	0.021	0.007
	ratio		2.788	2.425	1.916	0.973
State	predicted	3,470	0.089	0.006	0.004	0.002
	actual		0.398	0.025	0.013	0.002
	ratio		4.450	3.925	3.210	1.068

B. Residualized with respect to observed characteristics.

Group	standard deviation	Average Group size	BMI	Obese	Morbidly obese	under–weight
Individual		1	4.996	0.369	0.235	0.149
County	predicted	494	0.225	0.165	0.010	0.007
	actual		0.294	0.203	0.012	0.007
	ratio		1.311	1.223	1.193	0.902
State	predicted	3,470	0.085	0.006	0.004	0.002
	actual		0.169	0.010	0.006	0.002
	ratio		1.992	1.623	1.481	0.926

Table 2: Inferring social interactions from excess variation across regions. Table shows standard deviations predicted under the assumption that body weights are i.i.d. contrasted with actual variances calculated from the BRFSS. High ratios of actual to predicted variances suggest social interactions are driving outcomes. Weight outcomes in the bottom panel are residuals from regressions on quadratics in age and in income, ethnicity dummies, education, marital status, employment status, food prices and food store access measures, cigarette tax rates, and regional means of the individual characteristics. N=450,391.

A. County level estimates.

	BMI	Obese	morbidly obese	under–weight
		Without fixed effects.		
OLS	1.032	1.088	1.066	0.794
	(85.69)	(98.25)	(99.30)	(70.16)
IV	1.140	1.277	1.343	0.644
	(33.53)	(30.05)	(25.78)	(25.06)
		With fixed effects.		
IV-FE	1.002	1.062	0.994	0.668
	(90.85)	(100.57)	(75.27)	(22.78)

B. State level estimates.

	BMI	Obese	Morbidly obese	under–weight
		Without fixed effects.		
OLS	1.052	1.158	1.142	0.608
	(30.24)	(31.92)	(36.36)	(10.78)
IV	1.059	1.165	1.147	0.591
	(29.41)	(31.12)	(33.79)	(10.40)
		With fixed effects.		
IV-FE	1.075	1.050	0.8685	0.734
	(31.06)	(23.69)	(18.72)	(12.67)

Table 3: Estimates of body weight social multipliers. Each cell represents a separate regression; each estimate is of a social multiplier. Estimates were obtained by regressing individual weight outcomes on individual and aggregate level covariates, then collapsing the data at the region–year level and regressing between effects (aggregate) predicted values on predicted outcomes using the individual level models. IV models correct for measurement error using a split–sample method. t–ratios in parentheses are based on standard errors are from final stage OLS regression and do not reflect sampling error in the first two stages. Covariates in all models are as described in the notes to Table 2. N=450,391.

	BMI		obese		morbidly obese		underweight	
state effects?	no	yes	no	yes	no	yes	no	yes
Selected individual–level covariates:								
smoker	-1.1683	-1.1697	-0.0589	-0.0590	-0.0283	-0.0284	0.0177	0.0177
	(-65.30)	(-65.30)	(-46.18)	(-46.21)	(-34.98)	(-35.04)	(28.16)	(28.09)
employed	-0.0367	-0.0352	-0.0076	-0.0075	-0.0095	-0.0094	-0.0058	-0.0058
	(-1.83)	(-1.76)	(-5.30)	(-5.22)	(-10.04)	(-9.95)	(-9.08)	(-9.09)
income	-0.1958	-0.1949	-0.0184	-0.0184	-0.0160	-0.0160	-0.0070	-0.0070
	(-11.06)	(-11.00)	(-14.68)	(-14.64)	(-18.71)	(-18.70)	(-12.60)	(-12.63)
income^2	-0.0001	-0.0003	0.0005	0.0005	0.0009	0.0009	0.0007	0.0007
	(-0.06)	(-0.14)	(3.73)	(3.64)	(10.16)	(10.11)	(11.46)	(11.45)
State level means:								
Age	0.0024	0.0043	-0.0002	-0.0005	-0.0008	0.0017	-0.0005	0.0010
	(0.19)	(0.06)	(-0.22)	(-0.11)	(-1.31)	(0.52)	(-1.26)	(0.45)
Black	-1.3201	0.4244	-0.0352	-0.0122	-0.0218	-0.0389	0.0272	0.0022
	(-5.43)	(0.25)	(-1.96)	(-0.10)	(-1.90)	(-0.48)	(3.63)	(0.04)
Hispanic	-1.2132	1.2885	-0.0537	-0.0171	-0.0201	0.1367	0.0224	0.0582
	(-7.17)	(0.77)	(-4.29)	(-0.14)	(-2.54)	(1.74)	(4.24)	(1.13)
Other ethn.	3.1166	2.3061	0.1581	-0.0391	0.1185	0.0320	-0.0572	-0.0562
	(4.95)	(1.36)	(3.41)	(-0.32)	(4.00)	(0.41)	(-3.00)	(-1.08)
Smoker	2.9106	4.3973	0.2263	0.4419	0.0727	0.2450	0.0031	0.0768
	(3.95)	(1.43)	(4.16)	(1.96)	(2.09)	(1.70)	(0.14)	(0.82)
Male	-4.9331	-2.2422	-0.2483	-0.1314	-0.1678	-0.0664	0.0193	0.0369
	(-5.53)	(-1.19)	(-3.78)	(-0.95)	(-4.04)	(-0.76)	(0.70)	(0.63)
High School	0.4875	-0.1804	0.0972	-0.0988	0.0724	0.0833	0.0479	0.1213
	(0.89)	(-0.08)	(2.40)	(-0.59)	(2.80)	(0.78)	(2.89)	(1.71)
Some college	0.1919	1.0143	0.0654	-0.0559	0.0405	0.1530	0.0376	0.1402
	(0.36)	(0.35)	(1.67)	(-0.26)	(1.62)	(1.13)	(2.33)	(1.58)
BA+	-1.8117	1.1673	0.0277	0.1390	0.0011	0.1397	0.0957	0.1755
	(-2.46)	(0.46)	(0.51)	(0.75)	(0.03)	(1.19)	(4.26)	(2.25)
never married	-0.4483	-1.1124	-0.1409	-0.0954	-0.0491	0.0932	-0.0366	0.0025
	(-0.83)	(-0.53)	(-3.54)	(-0.62)	(-1.94)	(0.95)	(-2.22)	(0.04)
div / wid / sep	-3.0576	-6.5431	-0.2124	-0.2793	-0.0697	-0.2879	0.0092	-0.0138
	(-4.00)	(-1.97)	(-3.79)	(-1.15)	(-1.95)	(-1.85)	(0.39)	(-0.14)
employed	0.7549	-0.8290	-0.0179	-0.1317	-0.0035	-0.0780	-0.0508	-0.0595
	(1.71)	(-0.51)	(-0.55)	(-1.09)	(-0.17)	(-1.02)	(-3.71)	(-1.17)
income	0.0609	-0.3638	-0.0024	-0.0195	0.0044	-0.0125	-0.0015	0.0030
	(1.07)	(-1.50)	(-0.57)	(-1.10)	(1.67)	(-1.11)	(-0.87)	(0.40)
std. dev. inc	0.7982	0.2375	0.0687	0.0167	0.0158	-0.0102	-0.0065	-0.0059
	(6.14)	(0.67)	(7.20)	(0.64)	(2.62)	(-0.61)	(-1.56)	(-0.53)
price of food	-0.8499	-0.7620	-0.0567	-0.0519	-0.0286	-0.0227	0.0024	0.0045
	(-9.52)	(-6.55)	(-8.61)	(-6.03)	(-6.75)	(-4.08)	(0.92)	(1.30)
full service rests.	-0.0542	-0.0523	-0.0028	-0.0026	-0.0011	-0.0010	0.0002	0.0002
	(-20.28)	(-18.50)	(-14.57)	(-12.85)	(-9.03)	(-7.57)	(2.51)	(2.36)
fast food rests.	0.0546	0.0548	0.0024	0.0023	0.0017	0.0018	-0.0003	-0.0004
	(4.10)	(3.80)	(2.45)	(2.11)	(2.69)	(2.64)	(-0.86)	(-0.95)
supermarkets	0.0544	-0.0108	0.0027	-0.0001	0.0025	-0.0004	0.0004	0.0028
	(1.55)	(-0.26)	(1.06)	(-0.03)	(1.54)	(-0.20)	(0.36)	(2.28)
cig. tax	0.0029	0.0002	0.0002	-0.0000	0.0001	0.0001	-0.0000	-0.0000
	(6.24)	(0.16)	(5.49)	(-0.02)	(3.65)	(1.27)	(-2.36)	(-0.28)
R^2	0.1011	0.0976	0.0742	0.0713	0.0457	0.0438	0.0121	0.0120
σ_λ								
σ_u		5.0043		0.3689		0.2356		0.1519
ρ								

Table 4: Reduced form social interaction models. Split–sample IV estimates to correct for sampling error in regional means. Numbers in parentheses are t–ratios. All models also include a constant, year effects, and (at the individual level) a quadratic in age, ethnicity dummies, a sex dummy, and educational status indicators. σ_λ is the estimated standard deviation of the state effects, and σ_u is the estimated standard deviation of the individual level disturbance. ρ is the proportion of error variance attributable to the region effects. Standard errors are corrected for clustering N = 450,391.

	BMI		obese		morbid		underweight	
Regional:	county	state	county	state	county	state	county	state
BMI	0.0621	0.1157						
	(0.47)	(0.89)						
obesity			0.3281	0.0575				
			(3.01)	(0.35)				
morbid obesity					0.4001	0.2919		
					(3.81)	(1.96)		
underweight							0.5195	0.6762
							(3.44)	(3.77)
R^2	0.0931	0.0972	0.0692	0.0709	0.0434	0.0435	0.0137	0.0121
Hansen's J	4.5409	8.9360	14.6679	8.9870	9.8758	14.5393	11.7789	6.6259
J p–value	0.9197	0.5382	0.1446	0.5333	0.4515	0.1498	0.3001	0.7602
Kleibergen–Paap F	4.9439	6.7944	4.6363	4.2672	3.9185	4.1385	2.4121	2.3823

Table 5: GMM Estimates of body weight models. Selected parameter estimates from models in which aggregate mean BMI appears as an endogenous regressor. Excluded instruments are aggregate mean education, employment, marital status, and ethnicity outcomes. Hansen's J refers to the heteroskedasticity–robust J statistic against the null that the overidentifying restrictions are valid. The Kleiber–Paap rk statistic is a measure of instrument strength, see the text. All models include full sets of region and year fixed effects and individual–level covariates as described in Table 1.

APPENDIX

Selected individual–level covariates:								
smoker	-1.1402	-1.1415	-0.0580	-0.0581	-0.0272	-0.0272	0.0172	0.0172
	(-28.87)	(-28.83)	(-25.92)	(-25.89)	(-20.66)	(-20.58)	(22.63)	(22.60)
employed	-0.0232	-0.0224	-0.0074	-0.0074	-0.0090	-0.0089	-0.0061	-0.0061
	(-0.83)	(-0.79)	(-3.98)	(-3.92)	(-7.05)	(-6.96)	(-8.36)	(-8.38)
income	-0.1881	-0.1881	-0.0174	-0.0175	-0.0156	-0.0156	-0.0069	-0.0070
	(-6.57)	(-6.57)	(-9.09)	(-9.07)	(-12.21)	(-12.23)	(-10.89)	(-10.89)
income^2	-0.0007	-0.0007	0.0004	0.0004	0.0009	0.0009	0.0007	0.0007
	(-0.25)	(-0.25)	(2.10)	(2.10)	(6.85)	(6.88)	(9.64)	(9.67)
State level means:								
age	0.0323	-0.0033	0.0015	0.0009	0.0002	-0.0001	-0.0005	-0.0000
	(1.45)	(-0.15)	(1.21)	(0.48)	(0.32)	(-0.06)	(-1.26)	(-0.02)
black	-0.8811	-0.2868	-0.0263	-0.0287	-0.0111	-0.0430	0.0187	-0.0039
	(-2.32)	(-0.27)	(-1.21)	(-0.44)	(-0.82)	(-0.97)	(2.72)	(-0.13)
hispanic	-1.4732	0.0333	-0.0692	-0.0492	-0.0247	0.0376	0.0228	-0.0190
	(-3.54)	(0.04)	(-2.78)	(-0.74)	(-1.68)	(0.98)	(3.31)	(-0.66)
other ethn.	1.6265	2.3311	0.1011	0.0527	0.0741	0.0454	-0.0486	-0.0485
	(1.74)	(1.73)	(1.80)	(0.60)	(2.16)	(1.02)	(-1.65)	(-1.57)
smoker	2.3514	0.7292	0.1586	0.1135	0.0607	0.0665	0.0006	0.0326
	(2.66)	(0.50)	(3.21)	(1.28)	(1.96)	(1.32)	(0.03)	(1.10)
male	-2.4467	-0.3173	-0.1613	-0.0403	-0.0982	-0.0255	0.0057	-0.0062
	(-2.00)	(-0.33)	(-2.24)	(-0.56)	(-2.31)	(-0.61)	(0.31)	(-0.28)
high school	-0.3232	-1.9618	0.0528	-0.0849	0.0564	0.0386	0.0332	0.0448
	(-0.26)	(-1.16)	(0.79)	(-0.89)	(1.52)	(0.78)	(1.75)	(1.30)
some college	-0.0557	-1.3263	0.0482	-0.0896	0.0439	0.0424	0.0229	0.0269
	(-0.05)	(-0.83)	(0.73)	(-0.93)	(1.22)	(0.76)	(1.14)	(0.81)
BA+	-2.3030	-1.7082	-0.0348	-0.0105	-0.0110	0.0409	0.0663	0.0670
	(-2.29)	(-1.33)	(-0.58)	(-0.14)	(-0.35)	(0.95)	(3.09)	(1.97)
never married	-0.1956	0.3512	-0.0892	0.0515	-0.0289	0.0704	-0.0223	0.0151
	(-0.24)	(0.33)	(-1.62)	(0.70)	(-0.99)	(2.21)	(-1.45)	(0.37)
div / wid /sep	-2.1583	-0.7298	-0.1518	-0.0607	-0.0507	-0.0459	-0.0009	-0.0274
	(-1.78)	(-0.68)	(-2.08)	(-0.78)	(-1.16)	(-0.82)	(-0.05)	(-0.85)
employed	1.0597	0.2330	0.0130	-0.0314	0.0051	-0.0445	-0.0416	-0.0493
	(1.48)	(0.30)	(0.29)	(-0.54)	(0.22)	(-1.12)	(-3.15)	(-1.80)
income	0.0087	-0.1058	-0.0024	-0.0062	0.0024	-0.0019	-0.0003	0.0061
	(0.08)	(-0.75)	(-0.43)	(-0.61)	(0.70)	(-0.25)	(-0.15)	(1.44)
std. dev. income	1.0296	0.7585	0.0787	0.0560	0.0238	0.0169	-0.0074	-0.0074
	(4.07)	(2.96)	(5.88)	(3.98)	(3.36)	(1.99)	(-2.03)	(-1.41)
price food	-0.7989	-0.7368	-0.0557	-0.0468	-0.0289	-0.0195	0.0011	0.0045
	(-3.46)	(-3.99)	(-4.70)	(-5.74)	(-4.20)	(-3.67)	(0.29)	(1.28)
non-ff rests.	-0.0571	-0.0557	-0.0030	-0.0028	-0.0012	-0.0011	0.0003	0.0002
	(-9.59)	(-11.80)	(-8.15)	(-9.08)	(-5.74)	(-6.23)	(2.54)	(2.50)
fast food rests.	0.0576	0.0573	0.0024	0.0023	0.0017	0.0017	-0.0007	-0.0008
	(2.47)	(2.74)	(1.82)	(1.73)	(1.81)	(2.19)	(-1.56)	(-1.99)
smarket density	0.0066	-0.0786	0.0012	-0.0039	0.0021	-0.0011	0.0006	0.0025
	(0.08)	(-0.80)	(0.24)	(-0.62)	(0.80)	(-0.32)	(0.39)	(1.96)
cig. tax	0.0035	-0.0000	0.0002	0.0000	0.0001	0.0000	-0.0000	-0.0000
	(3.01)	(-0.00)	(2.78)	(0.62)	(2.88)	(1.60)	(-2.37)	(-0.38)
R^2	0.1013	0.0980	0.0741	0.0714	0.0452	0.0436	0.0119	0.0118
σ^2_λ		0.2634		0.0125		0.0095		0.0053
σ^2_u		4.9871		0.3672		0.2341		0.1520
ρ		0.0028		0.0012		0.0016		0.0012

Table A-1: OLS and OLS-FE models of body weight outcomes. Numbers in parentheses are t–ratios. All models also include a constant, year effects, and (at the individual level) a quadratic in age, ethnicity dummies, a sex dummy, and educational status indicators. σ_λ is the estimated standard deviation of the state effects, and σ_u is the estimated standard deviation of the individual level disturbance. ρ is the proportion of error variance attributable to the region effects. Standard errors are corrected for clustering N = 450, 391.