Lifestyles, Demographics, Dietary Behavior, and Obesity: A Switching Regression Analysis

Steven T. Yen, Zhuo Chen, and David B. Eastwood

Objectives. To investigate the effects of lifestyles, demographics, and dietary behavior on overweight and obesity.

Data Source. Continuing Survey of Food Intakes by Individuals 1994–1996, U.S. Department of Agriculture.

Study Design. We developed a three-regime switching regression model to examine the effects of lifestyle, dietary behavior, and sociodemographic factors on body mass index (BMI) by weight category and accommodating endogeneity of exercise and food intake to avoid simultaneous equation bias. Marginal effects are calculated to assess the impacts of explanatory variables on the probabilities of weight categories and BMI levels.

Principal Findings. Weight categories and exercise are found to be endogenous. Lifestyle, dietary behavior, social status, and other sociodemographic factors affect BMI differently across weight categories. Education, employment, and income have strong impacts on the likelihood of overweight and obesity. Exercise reduces the probabilities of being overweight and obese and the level of BMI among overweight individuals.

Conclusion. Health education programs can be targeted at individuals susceptible to overweight and obesity. Social status variables, along with genetic and geographic factors, such as region, urbanization, age, and race, can be used to pinpoint these individuals.

Key Words. Lifestyle, obesity, ordinal probit, overweight, switching regression

Obesity has become a major public health issue in developing and developed countries alike. Overweight and obese individuals are at increased risks for a long list of physical ailments, including hypertension (high blood pressure [BP]), hypercholesterolemia (high blood cholesterol), diabetes, coronary heart disease, stroke, cancer, poor reproductive health, and psychological problems such as depression and eating disorders (Stunkard and Wadden 1993, p. 224; National Institutes of Health [NIH] 1998; Jensen et al. 2004; Costa-Font and Gil 2005). Obesity accounts for approximately 365,000 deaths in the United States per year, second only to tobacco use (Mokdad et al. 2005). The total

direct and indirect costs attributable to overweight and obesity amounted to US\$117 billion in 2000 (U.S. Department of Health and Human Services [USDHHS] 2001). As the then Surgeon General David Satcher noted, these problems, if left unabated, may soon cause as many preventable diseases and deaths as tobacco use (USDHHS 2001).

Ogden et al. (2002) noted that the prevalence of overweight children in the United States had continued to increase, especially among Mexican American and non-Hispanic black adolescents. Kuchler and Variyam (2003) and Mokdad et al. (2003) confirmed the deleterious effects of an overweight U.S. population. Given the increased prevalence of obesity and mounting evidence of its deleterious health effects, the Surgeon General called upon citizens to work together "in finding solutions to this public health crisis" (USDHHS 2001, p. xiv).

The increasing prevalence of overweight and obese people has created an urgent need to identify the causes, which can provide a framework to develop programs and policies to constrain and ultimately reduce the incidence. Lifestyle, health knowledge, social policies, and neighborhood characteristics have been hypothesized as the underlying factors of the obesity epidemic within and outside of the United States. Lifestyle variables have long been identified as contributing factors (e.g., Huffman and Rizov 2007). Kuchler and Variyam (2003) found nonsmokers more likely to become obese than smokers. Chen, Yen, and Eastwood (2007) examined the relationship between smoking and bodyweight and found that smoking may not cause sustained weight loss. Chou, Grossman, and Saffer (2004) found that the number of fast-food and full-service restaurants, food consumed at home, and prices of cigarettes and alcohol were related to obesity. Lin, Huang, and French (2004) examined the relationship among eating behavior, dietary intake, physical activity, attitude toward diet and health, sociodemographic variables, and body mass index (BMI) among women and children in the 1994–1996 Continuing Survey of Food Intakes by Individuals (CSFII). Significant correlations between women's BMI and age, race, dietary patterns, TV watching, and smoking were found.

Address correspondence to Steven T. Yen, Ph.D., Department of Agricultural Economics, The University of Tennessee, 302 Morgan Hall, Knoxville, TN 37996-4518; e-mail: syen@utk.edu. Zhuo Chen, Ph.D., is with The Chicago Center of Excellence in Health Promotion Economics (formerly), The University of Chicago, Chicago, IL. David B. Eastwood, Ph.D., is with the Department of Agricultural Economics, The University of Tennessee, Knoxville, TN.

Kan and Tsai (2004), using a sample from Taiwan, found a relationship between individuals' knowledge concerning the health risks of obesity and their tendencies to be obese. Nayga (2000) found positive effects of health knowledge on obesity among U.S. adults after controlling for education using the Diet and Health Knowledge Survey component of the 1994 CSFII. TV watching, video games, and computer uses have been blamed for childhood obesity in Switzerland (Stettler, Signer, and Suter 2004) and in the United States (Vandewater, Shim, and Caplovitz 2004).

Recent analyses of cross-sectional data indicate that women receiving food stamps are more likely to be overweight and obese and also weigh more than nonparticipating eligibles (Fox and Cole 2004; Chen, Yen, and Eastwood 2005). Using data from National Health and Nutrition Examination Surveys, Ver Ploeg et al. (2007, p. 10), however, found that the effect of FSP has vanished because "the BMI of the rest of the population has caught up to the BMI levels of food stamps recipients."

Declining food prices associated with long-run technological change have been found to contribute to the rising proportions of overweight and obese people in high-income countries (Philipson and Posner 1999; Lakdawalla and Philipson 2002; Cutler, Glaeser, and Shapiro 2003). Chen and Meltzer (2008) concluded that relative income had contributed to obesity in rural China.

Existing studies have often focused on one weight category (overweight). However, the effects of sociodemographic and lifestyle variables on body weight may differ by weight category and such differentiated effects help identify target populations for public policy intervention. Audrain, Klesges, and Klesges (1995), for instance, found that smoking increased resting energy expenditure in both normal-weight and obese smokers, but the metabolic effect was larger and lasted longer among normal-weight smokers. The theoretical biology literature confirms that unobserved energy expenditure, used to maintain basic metabolism, increases with bodyweight (Christiansen, Garby, and Sørensen 2005), which leads to our hypothesis that the relationship between obesity and its underlying factors may differ across weight categories.

This study addresses the differentiated effects of sociodemographic and lifestyle variables on BMI by weight category. We first examine the effects of sociodemographic and lifestyle variables on body weight by estimating ordinary least-squares (OLS) regressions for weight-segmented subsamples and the pooled sample. These regression estimates allow testing of the hypothesis that the relationship between obesity and the underlying factors is constant across weight categories. On rejection of this "equal relationship hypothesis," we tested the next hypothesis that determination of weight categories is

exogenous. A comparison among the segmented sample results motivates the use of a more sophisticated statistical model, namely the endogenous switching regression model (SRM), which permits a test of the exogeneity hypothesis.¹

In the next section, we present a theoretical framework that motivates our empirical specification and describe the dataset and variables used. We then compare OLS estimates of the BMI regressions by weight categories. A multiregime endogenous SRM is then developed. Results for the SRM are presented next, which is followed by concluding remarks.

METHOD

Theoretical Framework

The empirical specification is based on the consumer utility maximization theory. Conditional on a set of sociodemographic and lifestyle variables S_1 , an individual derives utility from body weight (W), physical activity or exercise (E), and levels of food (F) and other goods (C) consumed, with corresponding prices P_E , P_F , and P_C . Body weight is a function of food consumed and exercise, conditional on another set of sociodemographic and lifestyle variables S_2 . The utility function

$$U = U(W(F, E; S_2), E, F, C; S_1) \quad (1)$$

is maximized subject to an income (I) constraint limiting the amount of money spent:

$$P_F F + P_E E + P_C C = I \quad (2)$$

This setup is similar to that of Philipson and Posner (1999) and Schroeter, Lusk, and Tyner (2008). Solving the constrained utility maximization problem in (1) and (2) yields the optimal levels of food (F^*), exercise (E^*), other consumer goods (C^*), as well as body weight (W^*). The optimal weight can therefore be expressed as

$$W^* = W^*(F^*(P_F, P_E, P_C, I; S_1, S_2), E^*(P_F, P_E, P_C, I; S_1, S_2); S_1) \quad (3)$$

W^* is a function of F^* and E^* and, importantly, these two variables are endogenous (i.e., functions of prices, income, and sociodemographic and lifestyle variables). Note that while prices are not available in a single cross section, some of the price variations are likely explained by the sociodemographic variables S_1 and S_2 , especially regional and urbanization variables. Multiple food items, used in the empirical specification below, can

be accommodated by allowing F to be a vector. We treat the determination (endogenization) of F^* and E^* differently because inclusion of many food items in a simultaneous-equations framework would require too many variables and would be further complicated by censoring (observed zero values) in these foods items. We take a simpler approach of predicting the food quantities with a censored (Tobit) regression model (Amemiya 1985). These predicted quantities are used as regressors in the weight category and level equations discussed below. This procedure to circumvent the endogeneity issue was used by Bryant (1988) and Yen (1993). The other endogenous variable, exercise, is incorporated by including a probit equation in the SRM and this variable is used as an explanatory variable in both the weight category and the level equations.

Data and Sample

Data are drawn from the CSFII, which collected information on food intakes, sociodemographic, behavioral, and anthropometric measurements for a multistage stratified area probability sample of noninstitutionalized individuals in the United States (USDA-ARS 2000). Our sample includes 9,209 adults aged ≥ 18 .

BMI, the first endogenous variable, is defined as self-reported weight (in kilograms) divided by the square of self-reported height (in meters). Based on criteria established by the U.S. Federal and international organizations, the sample is stratified into three groups by BMI level: underweight or normal weight in one category (below 24.9), overweight (between 25.0 and 29.9), and obese (30.0 and above) (NIH 1998; World Health Organization 1998; USDHHS 2000).² These groups correspond to the segmented sample and endogenous SRM estimates presented below. For the whole sample, 45 percent have normal or below normal weights (mean = 22.09, SD = 2.03), 36 percent are overweight (mean = 27.17, SD = 1.44), and 19 percent are obese (mean = 34.41, SD = 4.64). Variation in BMI is the largest within the obese group.

The second endogenous variable, exercise, is a binary indicator of whether the individual exercises regularly. Three health-related dummy variables are used to explain exercise—whether the individual had been diagnosed with cancer, had had any BP or heart problems (BP/heart), and whether self-reported health status was fair or better (healthy). The remaining endogenous variables are intakes of nine food products. As suggested above, to circumvent the endogeneity of these variables, a censored (Tobit) regression model (Amemiya 1985) is estimated to account for censoring in each food

product consumed, and the predicted food intake variables are used in the weight category and BMI equations.³ Exogenous explanatory variables are education, age, income, hours watching TV or playing video games per day (TV–videos), and dummy variables indicating urbanization (city, suburban), census region (northeast, midwest, south), race (white, black, Asian), ethnicity (Hispanic), home ownership (home owner), occupation (professional, manager, farmer, clerical), and whether the individual was smoking cigarettes (smoker) or was on a vegetarian diet (vegetarian). Definitions of all variables and sample statistics are presented in Table 1.

OLS Estimates by Weight Categories

Our first hypothesis was that sociodemographic and lifestyle variables have the same effects on BMI across weight categories. Potentially differentiated effects of explanatory variables on BMI were investigated with separate OLS regressions of the BMI equation for three weight categories and for the pooled sample. Results, presented in Table 2, suggest notable differences among coefficients of these regressions. For instance, age contributes to BMI according to the under/normal weight and pooled samples, but this variable does not affect BMI for the overweight and obese categories. TV and video games increase BMI within the overweight, obese, and pooled samples but not in the under/normal weight sample. Statistical significance of explanatory variables is much sparser among the overweight and obese categories than among the under/normal weight category. These differentiated effects of explanatory variables on BMI across weight categories are likely to be disguised in a pooled sample regression and highlight the importance of estimating the BMI equation by weight categories. Based on the log likelihood values of the segmented and pooled sample regressions, the hypothesis that all regression coefficients are equal across weight categories is rejected according to a likelihood ratio (LR) test ($LR = 16,683.64$, $df = 64$, $p\text{-value} < .0001$).

A Multiregime SRM with a Binary Endogenous Regressor

The inference is the BMI regression should be estimated for segmented samples and not for the pooled sample. We then consider the potential endogeneity of sample stratification. If selection into each weight category is endogenous or otherwise nonrandom, segmented- and pooled-sample estimates are subject to sample selection biases and loss of statistical efficiency. Nonrandom determination of weight categories can be accommodated and

Table 1: Definitions and Sample Statistics of Explanatory Variables (Sample Size = 9,209)

<i>Variable</i>	<i>Definition</i>	<i>Mean</i>	<i>SD</i>
Endogenous variables			
BMI category	BMI category: 1 = under/normal; 2 = overweight; 3 = obese	1.74	0.76
Exercise	Exercised regularly (2–4 times a week)	0.49	0.50
BMI	Body mass index (kg/m ²)	26.24	5.24
Continuous explanatory variables			
Education	Highest grade completed (years)	12.65	3.09
Age	Age in years	48.86	17.96
TV–videos	Average hours of TV watching or video games per day	2.88	2.19
Income	Per-capita annual income (thousand USD)	15.52	12.70
Other continuous variables: predicted food intake (g/day, 2-day average)			
Grains	Grains	298.73	40.57
Vegetables	Vegetables	223.47	27.34
Fruits	Fruit products	168.06	54.83
Milk and dairy	Milk and other dairy products	233.11	27.94
Meat	Meat products	219.64	32.16
Fats and oils	Fats and oils	17.14	3.45
Sugar and sweets	Sugar and sweets	26.82	3.06
Alc. bev.	Alcoholic beverages	135.11	71.26
Nonalc. bev.	Nonalcoholic beverages	959.64	119.00
Dummy variables (yes = 1; no = 0)			
City	Resided in central city	0.30	
Suburban	Resided in a suburban area	0.45	
Rural	Resided in a rural area (reference)	0.25	
Northeast	Resided in the northeast	0.18	
Midwest	Resided in the midwest	0.24	
South	Resided in the south	0.37	
West	Resided in the west (reference)	0.21	
White	Race is White	0.82	
Black	Race is Black	0.11	
Asian	Race is Asian	0.02	
Other race	Race is other (reference)	0.05	
Hispanic	Of the Hispanic origin	0.04	
Home owner	Was a home owner	0.71	
Professional	Employed as a professional	0.15	
Manager	Employed in a managerial role	0.09	
Farmer	Employed in the farming sector	0.01	
Clerical	Employed in a clerical role	0.09	
Service	Employed in service and other categories (reference)	0.66	
Male	Gender is male	0.51	
Smoker	Currently smoking cigarettes	0.25	
Vegetarian	On a vegetarian diet	0.03	
Healthy	Self-reported health is fair or better	0.83	
Cancer	Had been diagnosed of cancer	0.06	
BP/heart	Had been diagnosed of blood pressure or heart problem	0.29	

Note: Intake for each food item was predicted with a censored regression model. See text for details.

tested by an endogenous SRM. SRMs date back to Roy (1951), who was concerned with an individual's decision between earning income as a fisher or hunter and have been used extensively in economics. Vijverberg (1993) reviews their applications in labor economics, which estimate earning differentials by union/nonunion status, public/private sector, occupational status, migrant/stayer distinction, formal/informal sector, and level of education and in housing demand by renter/owner status and household credit by demand/supply constraint. Important applications in food and health include investigation of shopping frequencies and food intake decisions (Wilde and Ranney 2000), effect of food label use on nutrient intakes (Kim, Nayga, and Capps 2000), comparison of body weights between individuals in the United States and Canada (Auld and Powell 2006), and use of preventive care among the immigrant population (Pylypchuk and Hudson 2008). All existing SRM applications feature regression functions for two sample regimes, governed by a binary probit equation (Amemiya 1985, pp. 399–400). This framework is generalized in this paper to one with multiple regression functions for multiple sample regimes (weight categories). Another generalization we consider is endogenization of one of the regressors—exercise, which is binary.

Let z_1 , z_2 , and x be vectors of exogenous variables with conformable parameter vectors α_1 , α_2 , and β_j . The switching equation is specified as ordinal probit for the weight category variable (d):

$$d = j \text{ if } \xi_{j-1} < z_1' \alpha_1 + \gamma h + u_1 \leq \xi_j, \quad j = 1, 2, 3 \quad (4)$$

where γ is a scalar parameter and, for identification, the threshold parameters ξ 's are standardized such that $\xi_0 = -\infty$, $\xi_1 = 0$, $\xi_3 = \infty$, and ξ_2 is estimable (McKelvey and Zavoina 1975). Exercise (h) is modeled with a binary probit model

$$h = 1 \text{ if } z_2' \alpha_2 + u_2 > 0; \quad h = 0 \text{ otherwise} \quad (5)$$

which appears as an endogenous regressor (with scalar coefficients γ and δ_j) in the regime switching equation (4) and each of the category regression equations:

$$y = x' \beta_j + \delta_j h + \sigma_j v_j \text{ if } d = j, \quad j = 1, 2, 3 \quad (6)$$

where σ_j is the scale parameter. Note that vectors z and x contain lifestyle and sociodemographic variables as well as Tobit-predicted food intake variables discussed above. The error terms (u_1 , u_2 , v_1 , v_2 , v_3) are assumed to be distributed as multivariate normal with zero means, unitary variances, and covari-

Table 2: BMI Regressions: OLS Estimates by Weight Categories and for Pooled Sample (Dependent Variable = BMI)

<i>Variable</i>	<i>Under/Normal</i>	<i>Overweight</i>	<i>Obese</i>	<i>Pooled Sample</i>
Constant	− 131.491*** (36.178)	51.282* (30.947)	− 132.843 (128.057)	− 456.487*** (62.980)
Education	0.036 (0.023)	− 0.010 (0.019)	− 0.002 (0.078)	− 0.015 (0.039)
Age	0.089*** (0.023)	− 0.028 (0.021)	0.117 (0.088)	0.318*** (0.042)
TV-videos	− 0.003 (0.015)	0.026** (0.013)	0.149*** (0.047)	0.165*** (0.025)
Income	0.019 (0.099)	− 0.093 (0.080)	0.537 (0.389)	0.262 (0.173)
City	− 0.128 (0.086)	− 0.058 (0.073)	0.003 (0.313)	− 0.610*** (0.151)
Suburban	0.100 (0.079)	0.017 (0.065)	− 0.415 (0.272)	− 0.354*** (0.136)
Northeast	0.124 (0.099)	− 0.022 (0.085)	− 0.722* (0.377)	0.100 (0.176)
Midwest	0.194** (0.093)	0.024 (0.080)	− 0.280 (0.350)	0.436*** (0.165)
South	0.185** (0.085)	0.090 (0.074)	− 0.401 (0.335)	0.235 (0.153)
White	0.614 (0.405)	0.267 (0.346)	0.757 (1.476)	5.434*** (0.704)
Black	0.769* (0.421)	0.477 (0.355)	0.923 (1.504)	7.172*** (0.723)
Asian	0.064 (0.426)	− 0.201 (0.416)	− 2.744 (2.305)	2.656*** (0.780)
Hispanic	0.344* (0.195)	0.180 (0.147)	− 0.607 (0.684)	− 0.399 (0.324)
Home owner	− 0.098 (0.078)	0.001 (0.065)	− 0.451* (0.268)	− 0.182 (0.135)
Professional	− 1.148*** (0.277)	0.024 (0.231)	− 0.594 (0.980)	− 2.654*** (0.478)
Manager	− 1.020*** (0.278)	0.130 (0.230)	− 0.340 (0.966)	− 2.133*** (0.477)
Farmer	0.390 (0.322)	0.190 (0.184)	− 0.545 (0.907)	0.743* (0.456)
Clerical	0.049 (0.109)	0.012 (0.102)	− 0.396 (0.413)	− 0.075 (0.199)
Male	0.854*** (0.074)	− 0.075 (0.063)	− 1.291*** (0.278)	0.670*** (0.130)
Smoker	− 0.419*** (0.071)	− 0.082 (0.063)	− 0.889*** (0.268)	− 1.086*** (0.127)

continued

Table 2. *Continued*

<i>Variable</i>	<i>Under/Normal</i>	<i>Overweight</i>	<i>Obese</i>	<i>Pooled Sample</i>
Vegetarian	− 0.846*** (0.163)	− 0.375** (0.183)	− 1.909** (0.932)	− 2.304*** (0.332)
Grains	8.986*** (2.975)	− 4.150* (2.478)	16.204 (10.435)	27.067*** (5.135)
Vegetables	6.790** (3.435)	− 0.075 (2.882)	− 8.514 (11.771)	3.408 (5.898)
Fruits	2.152** (0.924)	1.109 (0.825)	− 0.072 (3.438)	13.128*** (1.635)
Milk and dairy	− 1.759 (1.127)	− 1.685* (0.952)	12.986*** (3.944)	0.159 (1.954)
Meat	− 10.716*** (3.442)	4.399 (2.921)	− 19.368 (12.084)	− 29.977*** (5.981)
Fats and oils	− 3.741*** (1.254)	0.138 (1.088)	− 6.222 (4.571)	− 18.518*** (2.199)
Sugar and sweets	− 7.040*** (1.539)	0.694 (1.349)	− 4.953 (5.505)	− 21.121*** (2.703)
Alc. bev.	− 1.396*** (0.398)	0.187 (0.320)	− 1.801 (1.305)	− 4.056*** (0.667)
Nonalc. bev.	22.862*** (5.082)	− 3.254 (4.432)	27.835 (18.447)	76.432*** (8.944)
Exercise	0.249*** (0.063)	− 0.095* (0.052)	− 0.196 (0.231)	− 0.426*** (0.110)
R^2	0.090	0.024	0.084	0.081
F	13.18	2.56	5.06	25.93
(df ₁ , df ₂)	(31, 4145)	(31, 3254)	(31, 1714)	(31, 9209)
Log likelihood	− 8687.913	− 5822.630	− 5079.361	− 27931.72

Notes: Standard errors in parentheses. Asterisks indicate levels of significance:

***1%, **5%, *10%.

p -value < .0001 for all tests for overall significance.

Alc. bev., alcoholic beverages; BMI, body mass index; nonalc. bev., nonalcoholic beverages; OLS, ordinary least squares.

ance matrix $R = [\rho_{ij}]$. The model, consisting of equations (4)–(6), is estimated by the method of maximum likelihood, with the sample likelihood function programmed in the Gauss programming language. Derivation of the sample likelihood function is available in an appendix online. Note that, apart from the additional probit equation (5) for exercise, the equations (4) and (6) represents a generalization of an important trichotomous-regime model suggested by Poirier (1978) in which the two (upper and lower) extreme outcomes of y are treated as “spikes” (versus regression functions), and the errors are uncorrelated between the switching and regression equations. As in the two-regime SRMs (Amemiya 1985), correlations among error terms (v_1 , v_2 , v_3) of

the regression equations are not estimable because the sample regimes are mutually exclusive. When error correlations between the dietary equation and the rest of the system are set to zeros, the model reduces to one with an exogenous regressor (henceforth, “exogenous” model), in which case the probit equation (5) can be estimated separately from the SRM equations (4) and (6). Further, restricting all (remaining) estimable error correlations to zeros results in an independent model that can be estimated by an ordered probit for d and a binary probit for h using the full sample and OLS by category for y , as has been done in the segmented OLS estimation (Table 2). We test both nested specifications against the full model with LR tests. Note that the model equations (4) and (6) reduce to the conventional two-regime SRM used extensively in the aforementioned literature when the number of categories in (4) is 2.

As in two-regime SRM and sample selection models, the effects of exogenous variables can be explored further by calculating marginal effects. Define a dichotomous indicator $k = 2h - 1$ such that $k = -1$ if $h = 0$ and $k = 1$ if $h = 1$. Then, the probability of being in weight category j conditional on exercise status $h = 1$ or 0 is

$$\Pr(d = j|h) = \left\{ \Phi_2(\xi_j - z'_1\alpha_1 - \eta h, kz'_2\alpha_2; -\kappa\rho_{21}) - \Phi_2(\xi_{j-1} - z'_1\alpha_1 - \eta h, kz'_2\alpha_2; -\kappa\rho_{21}) \right\} / \Phi(kz'_2\alpha_2), \quad j = 1, 2, 3 \quad (7)$$

where ρ_{21} is the correlation between error terms u_1 and u_2 with pdf $g(u_1, u_2; \rho_{21})$, and Φ and Φ_2 are the univariate and bivariate standard normal cumulative distribution functions, respectively. Define the bivariate standard normal probability as

$$G(a_1, b_1, a_2, b_2; \rho_{21}) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} g(u_1, u_2; \rho_{21}) du_1 du_2 \quad (8)$$

with integration limits

$$a_1 = \xi_{j-1} - z'_1\alpha_1 - \eta h, b_1 = \xi_j - z'_1\alpha_1 - \eta h, a_2 = -\infty, b_2 = -z'_2\alpha_2 \quad (9)$$

Then, the category means of y conditional on no exercise ($h = 0$) are

$$E(y|d = j, h = 0) = x'\beta_j + \sigma \left\{ \sum_{i=1}^2 \rho_{i,j+2} \gamma_i \phi(a_i) - \sum_{i=1}^2 \rho_{i,j+2} \lambda_i \phi(b_i) \right\}, \quad j = 1, 2, 3 \quad (10)$$

where

$$\gamma_1 = \{G(a_1, b_1, a_2, b_2; \rho_{21})\}^{-1} \left\{ \Phi \left(\frac{b_2 - \rho_{21} a_1}{(1 - \rho_{21}^2)^{1/2}} \right) - \Phi \left(\frac{a_2 - \rho_{21} a_1}{(1 - \rho_{21}^2)^{1/2}} \right) \right\} \quad (11)$$

$$\gamma_2 = \{G(a_1, b_1, a_2, b_2; \rho_{21})\}^{-1} \left\{ \Phi \left(\frac{b_1 - \rho_{21} a_2}{(1 - \rho_{21}^2)^{1/2}} \right) - \Phi \left(\frac{a_1 - \rho_{21} a_2}{(1 - \rho_{21}^2)^{1/2}} \right) \right\} \quad (12)$$

$$\lambda_1 = \{G(a_1, b_1, a_2, b_2; \rho_{21})\}^{-1} \left\{ \Phi \left(\frac{b_2 - \rho_{21} b_1}{(1 - \rho_{21}^2)^{1/2}} \right) - \Phi \left(\frac{a_2 - \rho_{21} b_1}{(1 - \rho_{21}^2)^{1/2}} \right) \right\} \quad (13)$$

$$\lambda_2 = \{G(a_1, b_1, a_2, b_2; \rho_{21})\}^{-1} \left\{ \Phi \left(\frac{b_1 - \rho_{21} b_2}{(1 - \rho_{21}^2)^{1/2}} \right) - \Phi \left(\frac{a_1 - \rho_{21} b_2}{(1 - \rho_{21}^2)^{1/2}} \right) \right\} \quad (14)$$

The category conditional mean with exercise, that is, $E(y|d = j, h = 1)$, is also given by (10), with an additional term $\delta_j h$ on the right-hand side and with integration limits changed slightly to $a_2 = -z_2' \alpha_2$ and $b_2 = \infty$ in (9). The conditional mean (10) is an extension of that in Tunali (1986) and Fische, Trost, and Lurie (1981) with dual single truncations to one with dual double truncations in the random error terms, and detailed derivations are available in S. T. Yen and J. Rosinski (unpublished data). Marginal effects of explanatory variables on the probabilities of weight categories and on BMI are calculated by differentiating (or differencing, in the case of discrete variables) the above expressions for the probabilities and conditional means. For statistical inferences, standard errors of marginal effects are calculated by the δ -method (Rao 1973, p. 388).

To evaluate the effects of exercise on BMI levels, we also calculate the average treatment effects (ATEs), which are often used in similar sample selection models (Heckman, Tobias, and Vytlačil 2001) in the program evaluation literature. Using the conditional probability (7) and conditional mean (10), the effects of exercise on the probabilities and levels of BMI for each individual of being in the j th category are, respectively,

$$TE_p = \Pr(d = j|h = 1) - \Pr(d = j|h = 0), j = 1, 2, 3 \quad (15)$$

$$TE_y = E(y|d = j, h = 1) - E(y|d = j, h = 0), j = 1, 2, 3 \quad (16)$$

ATEs can be calculated as the sample means of the individual effects (TE_p and TE_y) across the sample.

RESULTS

SRM Regression Estimates

Besides variables included in the other equations, healthy, cancer, and BP/heart are used solely to explain exercise, the justification being that these health belief and condition variables may affect exercise decisions more directly than they do body weight and that endogenization of exercise nevertheless allows these variables to affect body weight indirectly by way of the exercise equation. Another specification issue relates to the choice of variables for the switching equation and regression equations. Unlike two-step estimation, in which exclusion conditions are needed for identification, in maximum-likelihood estimation, the nonlinear identification criteria (i.e., linear independence of the first derivatives of the log likelihood) are met in the current system through the functional form and distributional assumptions (Buchel and van Ham 2003; Pylypchuk and Hudson 2008). To avoid overburdening the functional form for identification, however, the system is estimated with one additional variable in the switching equation. Specifically, the squared term of age is used in the switching equation and is found significant at the 1 percent level, which suggests a nonlinear effect of age on the prevalence of overweight and obesity.

As suggested above, significant error correlations between the exercise equation and the other equations would indicate endogeneity of the former, whereas significance of the error correlations between the switching equation and the regression equations would suggest endogenous regime switching (i.e., determination of weight categories). Based on the likelihood values of the full and restricted models, LR tests suggest rejection of both the independent model (LR = 108.94, df = 7, p -value < .001) and the exogenous model (LR = 64.25, df = 4, p -value < .001).

Table 3 presents parameter estimates for the full model. The LR tests for endogeneity discussed above are confirmed by significance of the error correlations: between the switching equation and the exercise equation and between these equations and the BMI equation for the under/normal weight category.

The exogenous and independent models, both rejected by LR tests in favor of the full model, produce notably different regression coefficients from the full model. For instance, whereas the independent model (Table 2) suggests statistical insignificance of age in the overweight equation, the full model suggests significance (at the 5 percent significance level) of age in that equation. Differences are also found in other variables such as suburban, northeast, and vegetarian.

According to parameter estimates of the full model (Table 3), age has a significant and nonlinear effect on the regime switching equation. Other significant variables in the switching equation are urbanization (city and suburban), region (midwest), race (Asian), professional category (clerical), gender (male), smoker, vegetarian, and exercise. Interestingly, none of the food intake variables is a determinant of weight categories. Over one half of the variables are significant (at the 10 percent level or lower) in the regression equation for under/normal weight, whereas significance is more scant in the regression equations for the other weight categories. Because the effects of exercise and other explanatory variables on the probabilities of weight categories and BMI level have important policy implications, we investigate the roles of these variables further by calculating ATEs and marginal effects.

Effects of Exercise

According to the ATEs presented in Table 4, exercise ameliorates overweight and obesity. Specifically, regular exercise increases the probability of being in the under/normal weight category by .04 and decreases the probabilities of overweight and obesity slightly (by .02 and .03, respectively). In terms of level, regular exercise increases BMI by 0.19 on average among individuals in the under/normal weight group while decreasing the BMI by 0.09 among individuals who are overweight.

Marginal Effects of Variables

Table 5 presents the marginal effects of significant explanatory variables on the probabilities of BMI categories, conditional on exercise status and calculated at the means of all variables by exercise category. We present only the results for significant variables. The SRM produces notably different marginal effects of variables, and the rest of the analysis is based on this model. The marginal effects for many of these variables are notably different, in both signs and magnitudes, between individuals who exercise regularly and those who do not. Such differences are most often seen for the probabilities of the overweight category. For instance, whereas TV–videos increase the probability of being overweight (at the 1 percent level of significance) conditional on exercise, this variable has a negative effect on overweight conditional on no exercise (at the 10 percent level of significance). Differentiated effects on the probability of being overweight by exercise status are also seen for other variables, such as education, age, suburban, northeast, midwest, home owner, professional, clerical, healthy, and BP/heart. Echoing findings by Stettler,

Table 3: Multiregime Switching Regression Estimates of BMI with Endogenous Exercise

Variable	Switching: Ordered Probit	Endogenous: Exercise	Regression Equations of BMI		
			Under/Normal	Overweight	Obese
Constant	5.567 (13.626)	− 0.058 (0.112)	− 69.691* (37.509)	76.350** (33.975)	− 35.810 (153.357)
Education	− 0.003 (0.008)	0.021*** (0.006)	0.019 (0.024)	− 0.005 (0.019)	0.024 (0.084)
Age	0.062*** (0.009)	− 0.010*** (0.001)	0.057** (0.025)	− 0.044** (0.022)	0.054 (0.105)
Age ² /1,000	− 0.740*** (0.045)				
TV-videos	0.006 (0.006)	− 0.055*** (0.007)	0.016 (0.017)	0.013 (0.015)	0.093** (0.045)
Income	0.056 (0.037)	0.099*** (0.021)	− 0.072 (0.107)	− 0.100 (0.086)	0.542 (0.454)
City	− 0.151*** (0.034)	− 0.146*** (0.039)	0.028 (0.095)	− 0.036 (0.080)	0.096 (0.354)
Suburban	− 0.128*** (0.030)	− 0.149*** (0.035)	0.235*** (0.088)	0.024 (0.071)	− 0.383 (0.312)
Northeast	− 0.008 (0.040)	− 0.127*** (0.046)	0.180* (0.108)	− 0.046 (0.088)	− 0.819** (0.423)
Midwest	0.090** (0.037)	− 0.034 (0.042)	0.167* (0.100)	− 0.008 (0.084)	− 0.397 (0.372)
South	0.019 (0.035)	− 0.046 (0.039)	0.198** (0.091)	0.071 (0.076)	− 0.483 (0.343)
White	− 0.075 (0.156)	− 0.107 (0.073)	0.054 (0.441)	− 0.032 (0.366)	− 0.108 (1.633)
Black	0.173 (0.161)	− 0.149* (0.084)	0.051 (0.473)	0.072 (0.393)	− 0.297 (1.769)
Asian	− 0.835*** (0.182)	− 0.404*** (0.113)	0.006 (0.451)	− 0.296 (0.442)	− 2.645 (4.572)
Hispanic	− 0.065 (0.071)	− 0.094 (0.083)	0.478** (0.209)	0.204 (0.148)	− 0.415 (0.647)
Home owner	− 0.036 (0.030)	0.031 (0.033)	− 0.116 (0.082)	0.008 (0.066)	− 0.392 (0.296)
Professional	0.057 (0.104)	− 0.201*** (0.044)	− 0.641** (0.290)	0.139 (0.255)	− 0.210 (1.158)
Manager	0.116 (0.104)	− 0.219*** (0.051)	− 0.568** (0.292)	0.210 (0.250)	− 0.072 (1.090)
Farmer	0.094 (0.104)	− 0.005 (0.119)	0.251 (0.321)	0.134 (0.191)	− 0.747 (1.545)
Clerical	− 0.101** (0.045)	− 0.200*** (0.050)	0.181 (0.124)	0.000 (0.108)	− 0.433 (0.435)
Male	0.310*** (0.030)	0.377*** (0.028)	0.518*** (0.103)	− 0.100 (0.092)	− 1.383*** (0.395)
Smoker	− 0.311***	− 0.114***	− 0.252***	− 0.027	− 0.680*

continued

Table 3. *Continued*

Variable	Switching: Ordered Probit	Endogenous: Exercise	Regression Equations of BMI		
			Under/Normal	Overweight	Obese
	(0.028)	(0.032)	(0.080)	(0.075)	(0.364)
Vegetarian	− 0.229***	0.385***	− 0.903***	− 0.221	− 1.372
	(0.081)	(0.085)	(0.179)	(0.208)	(1.854)
Grains	− 1.327		5.099*	− 5.432**	11.243
	(1.086)		(2.954)	(2.644)	(11.737)
Vegetables	− 0.790		5.193	− 0.561	− 9.756
	(1.160)		(3.326)	(2.950)	(12.175)
Fruits	0.296		0.827	0.429	− 2.315
	(0.344)		(0.980)	(0.871)	(4.074)
Milk and dairy	0.458		− 1.405	− 1.552	12.347***
	(0.390)		(1.143)	(0.962)	(4.087)
Meat	1.518		− 6.330*	5.827*	− 13.167
	(1.256)		(3.412)	(3.085)	(13.923)
Fats and oils	0.066		− 1.413	1.121	− 2.542
	(0.482)		(1.347)	(1.187)	(5.879)
Sugar and sweets	0.288		− 4.228***	1.753	− 1.217
	(0.588)		(1.619)	(1.461)	(6.314)
Alc. bev.	− 0.212		− 0.922**	0.423	− 0.943
	(0.134)		(0.388)	(0.346)	(1.451)
Nonalc. bev.	− 0.941		13.085**	− 7.084	13.058
	(1.981)		(5.434)	(4.904)	(22.823)
Exercise	− 1.156***		2.121***	− 0.239	− 1.002
	(0.082)		(0.391)	(0.358)	(1.257)
Healthy		0.397***			
		(0.037)			
Cancer		− 0.082			
		(0.052)			
BP/heart		− 0.163***			
		(0.030)			
μ	0.926***				
	(0.031)				
σ			2.201***	1.455***	4.608***
			(0.101)	(0.036)	(0.162)
Error correlations					
Exercise	0.657***				
	(0.055)				
Under/normal	− 0.500***	− 0.508***			
	(0.076)	(0.090)			
Overweight	− 0.148	0.086			
	(0.118)	(0.156)			
Obese	− 0.213	0.185			
	(0.111)	(0.149)			
Log likelihood	− 34502.638				

Notes: Variables income and all food quantities consumed (from grains to nonalc. bev.) are log-arithmically transformed. Asymptotic standard errors in parentheses. Asterisks indicate levels of significance:

***1%, **5%, *10%.

Alc. bev., alcoholic beverages; BMI, body mass index; BP, blood pressure; nonalc. bev., nonalcoholic beverages.

Table 4: Effects of Exercise on Probabilities and Levels of BMI

<i>Category</i>	<i>Probability</i>	<i>BMI</i>
Under/normal	0.044*** (0.009)	0.193*** (0.070)
Overweight	− 0.018*** (0.003)	− 0.086* (0.053)
Obese	− 0.026*** (0.007)	− 0.030 (0.253)

Notes: Asymptotic standard errors in parentheses. Asterisks indicate levels of significance:

***1%, *10%.

BMI, body mass index.

Signer, and Suter (2004) and Vandewater, Shim, and Caplovitz (2004) that TV, video games, and computer use contribute to childhood obesity, we find that TV–videos decrease the probability of having under/normal weight and increase the probability of being obese, both conditional and unconditional on exercise. The magnitudes of marginal effects can be very different between exercisers and nonexercisers too. For instance, vegetarians are 9.47 percent less likely to be obese than nonvegetarians conditional on exercise but 13.96 percent less likely to be obese conditional on no exercise.

There is more agreement qualitatively, in terms of signs and statistical significance, between the two sets of marginal effects on the probabilities for the under/normal weight and obese subgroups—conditional on no exercise and on exercise. Conditional or unconditional on regular exercise, residing in the city or a suburban area, having a clerical job, being healthy, being Asian, a home owner, a smoker, and a vegetarian all increase the probability of under/normal weights and decrease the probability of being in the obese subgroup, whereas TV–videos, residing in the midwest, having a managerial job, and being male have the opposite effects—a negative effect on the probability of having an under/normal weight and a positive effect on the probability of obesity. Interestingly, men are more likely to be overweight and obese than women, both conditional on exercise and no exercise.

Table 6 presents the marginal effects of significant explanatory variables on BMI level by exercise status. These marginal effects are also calculated at the sample means of all explanatory variables by exercise status. The model produces notably different effects of some variables on BMI from those suggested by the independent model. For instance, whereas the independent model suggests age does not affect BMI among individuals who are over-

Table 5: Marginal Effects of Significant Explanatory Variables on Conditional Probabilities of BMI Categories

Variable	Conditional on No Exercise			Conditional on Exercise		
	Under/Normal	Overweight	Obese	Under/Normal	Overweight	Obese
Continuous explanatory variables						
Education	0.44 (0.34)	0.05 (0.10)	− 0.49 (0.31)	0.54 (0.35)	− 0.20* (0.11)	− 0.34 (0.25)
Age	0.43 (0.39)	− 0.03* (0.02)	− 0.40 (0.38)	0.06 (0.42)	0.01 (0.10)	− 0.07 (0.32)
TV-videos	− 1.06*** (0.20)	− 0.13* (0.08)	1.19*** (0.20)	− 1.33*** (0.22)	0.49*** (0.10)	0.83*** (0.18)
Binary explanatory variables						
City	3.89*** (1.21)	− 0.38** (0.18)	− 3.51*** (1.22)	4.08*** (1.36)	− 0.77** (0.40)	− 3.31*** (1.00)
Suburban	2.90*** (1.06)	− 0.35** (0.15)	− 2.55** (1.10)	2.96*** (1.20)	− 0.43 (0.33)	− 2.53*** (0.90)
Northeast	− 1.62 (1.46)	− 0.25 (0.19)	1.87 (1.39)	− 2.13 (1.63)	0.92* (0.52)	1.21 (1.14)
Midwest	− 4.16*** (1.35)	− 0.04 (0.23)	4.20*** (1.32)	− 4.73*** (1.50)	1.38*** (0.47)	3.35*** (1.09)
Black	− 8.46 (6.27)	− 1.72 (0.58)	10.18 (6.47)	− 10.18 (7.00)	1.86 (1.69)	8.32 (5.41)
Asian	30.42*** (7.32)	− 11.01*** (2.31)	− 19.42*** (6.15)	30.30*** (7.69)	− 14.40*** (3.19)	− 15.90*** (4.94)
Home owner	1.91* (1.07)	0.07 (0.13)	− 1.98* (1.08)	2.23* (1.21)	− 0.68** (0.34)	− 1.56* (0.90)
Professional	− 5.13 (3.86)	− 0.68 (0.48)	5.80 (4.23)	− 6.46 (4.51)	2.10** (0.98)	4.36 (3.58)
Manager	− 7.50** (3.67)	− 1.12 (0.78)	8.62** (4.32)	− 9.38** (4.37)	2.60 (0.75)	6.78* (3.74)
Clerical	1.15 (1.63)	− 0.59*** (0.22)	− 0.56 (1.50)	0.75 (1.81)	0.18 (0.63)	− 0.93 (1.21)
Male	− 6.84*** (1.06)	1.03*** (0.27)	5.81*** (1.08)	− 6.80*** (1.21)	1.16*** (0.45)	5.64*** (0.87)
Smoker	11.25*** (1.08)	− 1.32*** (0.43)	− 9.93*** (0.96)	12.13*** (1.17)	− 3.89*** (0.56)	− 8.25*** (0.77)
Vegetarian	17.31*** (3.47)	− 3.35** (1.58)	− 13.96*** (2.10)	17.18*** (3.23)	− 7.71*** (1.71)	− 9.47*** (1.65)
Healthy	5.45*** (0.58)	1.28*** (0.46)	− 6.73*** (0.93)	7.99*** (1.08)	− 2.60*** (0.51)	− 5.39*** (0.69)
BP/heart	− 2.40*** (0.47)	− 0.43** (0.19)	2.82*** (0.60)	− 3.19*** (0.67)	1.21*** (0.30)	1.98*** (0.40)

Notes: All probabilities are multiplied by 100. Asymptotic standard errors in parentheses. Asterisks indicate levels of significance:

***1%, **5%, *10%.

BMI, body mass index; BP, blood pressure.

Table 6: Marginal Effects of Significant Explanatory Variables on Conditional BMI Levels

Variable	Conditional on No Exercise			Conditional on Exercise		
	Under/Normal	Overweight	Obese	Under/Normal	Overweight	Obese
Continuous explanatory variables						
Age	0.044* (0.025)	-0.046* (0.027)	0.046 (0.112)	0.048 (0.090)	-0.044** (0.022)	0.053 (0.107)
TV-videos	0.008 (0.014)	0.027 (0.032)	0.162*** (0.043)	-0.013 (0.014)	0.026** (0.013)	0.120*** (0.044)
Grains	0.014 (0.010)	-0.020 (0.021)	0.032 (0.042)	0.014 (0.074)	-0.019** (0.008)	0.032 (0.039)
Milk and dairy	-0.005 (0.007)	-0.006 (0.005)	0.056*** (0.018)	-0.005 (0.040)	-0.006 (0.011)	0.054*** (0.018)
Meat	-0.081 (0.080)	0.029** (0.015)	-0.051 (0.067)	-0.025 (0.081)	0.028* (0.015)	-0.052 (0.064)
Sugar and sweets	-0.150*** (0.060)	0.069 (0.055)	-0.030 (0.239)	-0.152*** (0.060)	0.069 (0.055)	-0.035 (0.240)
Alc. bev.	-0.009 (0.007)	0.003 (0.002)	-0.010 (0.013)	-0.007*** (0.002)	0.002 (0.003)	-0.008 (0.010)
Nonalc. bev.	0.013 (0.009)	-0.008 (0.006)	0.012 (0.023)	0.013** (0.006)	-0.008 (0.005)	0.012 (0.025)
Binary explanatory variables						
City	-0.116 (0.085)	-0.054 (0.074)	0.051 (0.334)	-0.141* (0.085)	-0.056 (0.074)	-0.001 (0.333)
Northeast	0.146 (0.101)	-0.021 (0.086)	-0.686* (0.422)	0.100 (0.101)	-0.023 (0.086)	-0.777* (0.421)
Midwest	0.225** (0.095)	0.030 (0.082)	-0.234 (0.364)	0.196** (0.095)	0.027 (0.082)	-0.294 (0.363)
South	0.201** (0.086)	0.088 (0.076)	-0.403 (0.342)	0.181** (0.085)	0.086 (0.076)	-0.447 (0.341)
Hispanic	0.410** (0.196)	0.203 (0.145)	-0.400 (0.638)	0.386** (0.193)	0.202 (0.145)	-0.444 (0.638)
Home owner	-0.136* (0.078)	-0.011 (0.065)	-0.476* (0.292)	-0.117 (0.078)	-0.010 (0.065)	-0.441 (0.293)
Professional	-0.641** (0.285)	0.200 (0.253)	0.088 (1.164)	-0.730*** (0.283)	0.198 (0.253)	-0.070 (1.172)
Manager	-0.525* (0.288)	0.293 (0.250)	0.319 (1.113)	-0.635** (0.285)	0.293 (0.250)	0.139 (1.121)
Male	0.830*** (0.077)	-0.079 (0.063)	-1.378*** (0.290)	0.912*** (0.078)	-0.075 (0.063)	-1.217*** (0.294)
Smoker	-0.501*** (0.072)	-0.106* (0.064)	-0.989*** (0.303)	-0.487*** (0.072)	-0.104* (0.064)	-0.954*** (0.303)
Vegetarian	-0.938*** (0.158)	-0.400** (0.196)	-2.195 (1.840)	-0.808*** (0.158)	-0.350* (0.198)	-1.696 (1.845)
Healthy	0.077*** (0.031)	-0.080* (0.049)	-0.425*** (0.139)	0.243*** (0.078)	-0.088* (0.053)	-0.191*** (0.069)

continued

Table 6. *Continued*

<i>Variable</i>	<i>Conditional on No Exercise</i>			<i>Conditional on Exercise</i>		
	<i>Under/Normal</i>	<i>Overweight</i>	<i>Obese</i>	<i>Under/Normal</i>	<i>Overweight</i>	<i>Obese</i>
BP/heart	− 0.035** (0.015)	0.035 (0.022)	0.181*** (0.067)	− 0.098*** (0.034)	0.033 (0.021)	0.067** (0.028)

Notes: Asymptotic standard errors in parentheses. Asterisks indicate levels of significance: ***1%, **5%, *10%.
Alc. bev., alcoholic beverages; BMI, body mass index; BP, blood pressure; nonalc. bev., nonalcoholic beverages.

weight (Table 2), the variable decreases BMI among individuals who are overweight according to the SRM. Differences are also found in other variables, such as TV–videos. Similar to their effects on probabilities of weight categories, the effects of most variables on BMI level are also similar with or without regular exercise. Exceptions are found for a few variables. For instance, conditional on exercise, grain intake decreases BMI among the overweight while alcoholic beverages and residing in the city decrease BMI among individuals with under/normal weight; these variables do not affect BMI conditional on no exercise. Home ownership decreases BMI among individuals who have under/normal weights and individuals who are obese, conditional on no exercise; this variable does not affect BMI conditional on exercise. Age is related to an increase in BMI level within the under/normal subgroup but decreases BMI within the overweight subgroup. Interestingly, TV–videos increase BMI among those overweight and obese. Some of the food intake variables affect BMI level. Specifically, grains have a negative effect and meat intake has a positive effect on BMI among individuals who are overweight, whereas milk and dairy products have a positive effect on BMI among the obese. Sugar/sweets, alcoholic beverages, and being vegetarian all decrease BMI while nonalcoholic beverages increase BMI among individuals with under/normal weight. There is some evidence of regional differences, with under/normal weight individuals from the midwest and south being heavier than their counterparts from other regions. Ethnicity is a factor, with Hispanic having a positive effect on BMI among individuals with under/normal weight. Occupations also matter. Specifically, having a professional or managerial job decreases BMI among individuals with under/normal weight. Gender differences also exist, with men in the under/normal weight category having higher BMI and men in the obese category having lower BMI than their female counterparts. Smokers consistently have lower BMI across all

weight categories conditional on exercise and no exercise. Finally, conditional on exercise and no exercise, being healthy has a positive effect on BMI among individuals with under/normal weight but a negative effect on BMI among individuals who are overweight.

CONCLUSIONS AND DISCUSSIONS

Motivated by results of preliminary OLS analyses, we develop a multiregime SRM to investigate the differentiated roles of lifestyle, dietary, and sociodemographic variables on BMI by weight categories. We also accommodate endogeneity of an important lifestyle/behavioral variable—exercise. Weight categories and exercise are found to be endogenous. The multiregime SRM is particularly appealing because the differentiated effects of variables on BMI have important implications, and these differentiated effects would have been disguised with the use of more traditional analytical frameworks. Results suggest that sociodemographic characteristics, lifestyle, and dietary behavior affect the probabilities of individuals being in weight categories and the level of BMI. Also, these effects differ by weight category. Therefore, it is important to differentiate these effects by weight category when designing cost-effective intervention programs. They can be used to pinpoint those who are more susceptible to being overweight and obese and to target individuals for health intervention and nutritional education.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Funding was provided by Cooperative Agreements No. 58-5000-7-0123 and No. 58-4000-7-0029 with the Economic Research Service of the U.S. Department of Agriculture. The views expressed herein are those of the authors and do not necessarily reflect the views or policies of the USDA. There is no conflict of interest to disclose.

Disclosures: None.

Disclaimers: None.

NOTES

1. Other approaches include quantile regression. Although quantile regression appears to be a viable alternative to investigate the difference in the effects of

explanatory variables on bodyweight, we are more interested in examining such difference across weight categories. Quantiles are relative positions, but it has been well-established that health risks differ across categories with absolute thresholds.

2. Merger of the under and normal weight categories was due to the small proportion of the former in the sample.
3. Regressors in the Tobit equations include age, household size, household income (as a percentage of Federal poverty level), and dummy variables indicating education (primary school, college, and postgraduate), race, food stamp recipient status, being on a special diet, a white-collar worker, and a household meal planner.

REFERENCES

- Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- Audrain, J. E., R. C. Klesges, and L. M. Klesges. 1995. "Relationship between Obesity and the Metabolic Effects of Smoking in Women." *Health Psychology* 14 (2): 116–23.
- Auld, M. C., and L. M. Powell. 2006. "The Economics of Obesity: Research and Policy Implications from a Canada–U.S. Comparison." In *Health Services Restructuring in Canada: New Evidence and New Directions*, edited by C. M. Beach, R. P. Chaykowski, S. Shortt, F. St-Hilaire, and A. Sweetman, pp. 305–32. Kingston, Ontario: School of Policy Studies Queen's University.
- Bryant, W. K. 1988. "Durables and Wives Employment Yet Again." *Journal of Consumer Research* 15 (1): 37–47.
- Buchel, F., and M. van Ham. 2003. "Overeducation, Regional Labor Markets, and Spatial Flexibility." *Journal of Urban Economics* 53 (3): 482–93.
- Chen, Z., and D. O. Meltzer. 2008. "Beefing up with the Chans: Evidence for the Effects of Relative Income and Income Inequality on Health from the China Health and Nutrition Survey." *Social Science & Medicine* 66 (11): 2206–17.
- Chen, Z., S. T. Yen, and D. B. Eastwood. 2005. "Effects of Food Stamp Participation on Body Weight and Obesity." *American Journal of Agricultural Economics* 87 (5): 1167–73.
- . 2007. "Does Smoking Have a Causal Effect on Weight Reduction?" *Journal of Family and Economic Issues* 28 (1): 49–67.
- Chou, S. Y., M. Grossman, and H. Saffer. 2004. "An Economic Analysis of Adult Obesity: Results from the Behavior Risk Factor Surveillance System." *Journal of Health Economics* 23 (3): 565–87.
- Christiansen, E., L. Garby, and T. I. A. Sørensen. 2005. "Quantitative Analysis of the Energy Requirements for Development of Obesity." *Journal of Theoretical Biology* 234 (1): 99–106.
- Costa-Font, J., and J. Gil. 2005. "Obesity and the Incidence of Chronic Diseases in Spain: A Seemingly Unrelated Probit Approach." *Economics and Human Biology* 3 (2): 188–214.
- Cutler, D. M., E. L. Glaeser, and J. M. Shapiro. 2003. "Why Have Americans Become More Obese?" *Journal of Economic Perspectives* 17 (3): 93–118.

- Fishe, R. P. H., R. P. Trost, and P. M. Lurie. 1981. "Labor Force Earnings and College Choice of Young Women: An Examination of Selectivity Bias and Comparative Advantage." *Economics of Education Review* 1 (2): 169–91.
- Fox, M. K., and N. Cole. 2004. *Nutrition and Health Characteristics of Low-Income Populations: Vol. 1, Food Stamp Program Participants and Nonparticipants, E-FAN-04-014-1*. Washington, DC: United States Department of Agriculture, Economic Research Service.
- Heckman, J., J. L. Tobias, and E. Vytlacil. 2001. "Four Parameters of Interest in the Evaluation of Social Programs." *Southern Economic Journal* 68 (2): 210–23.
- Huffman, S. K., and M. Rizov. 2007. "Determinants of Obesity in Transition Economies: The Case of Russia." *Economics and Human Biology* 5 (3): 379–91.
- Jensen, T. K., A. M. Andersson, N. Jorgensen, A. G. Andersen, E. Carlsen, J. O. Petersen, and N. E. Skakkebaek. 2004. "Body Mass Index in Relation to Semen Quality and Reproductive Hormones among 1,558 Danish Men." *Fertility and Sterility* 82 (4): 863–70.
- Kan, K., and W. Tsai. 2004. "Obesity and Risk Knowledge." *Journal of Health Economics* 23 (5): 907–34.
- Kim, S. Y., R. M. Nayga Jr., and O. Capps Jr. 2000. "The Effect of Food Label Use on Nutrient Intakes: An Endogenous Switching Regression Analysis." *Journal of Agricultural and Resource Economics* 25 (1): 215–31.
- Kuchler, F., and J. N. Variyam. 2003. "Mistakes Were Made: Misperception as a Barrier to Reducing Overweight." *International Journal of Obesity* 27 (7): 856–61.
- Lakdawalla, D., and T. J. Philipson. 2002. *The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination*. NBER Working Paper # 8946.
- Lin, B. H., C. L. Huang, and S. A. French. 2004. "Factors Associated with Women's and Children's Body Mass Indices by Income Status." *International Journal of Obesity* 28 (4): 536–42.
- McKelvey, R. D., and W. Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4: 103–20.
- Mokdad, A. H., E. S. Ford, B. A. Bowman, W. H. Dietz, F. Vinicor, V. S. Bales, and J. S. Marks. 2003. "Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors, 2001." *Journal of the American Medical Association* 289 (1): 76–9.
- Mokdad, A. H., J. S. Marks, D. F. Stroup, and J. L. Gerberding. 2005. "Correction: Actual Causes of Death in the United States, 2000." *Journal of the American Medical Association* 293 (3): 293–4.
- National Institutes of Health (NIH). 1998. "Clinical Guidelines on the Identification, Evaluation and Treatment of Overweight and Obesity in Adults: The Evidence Report." *Obesity Research* 6 (suppl 2): 51S–209S.
- Nayga, R. M. 2000. "Schooling, Health Knowledge and Obesity." *Applied Economics* 32 (7): 815–22.
- Ogden, C. L., K. M. Flegal, M. D. Carroll, and C. L. Johnson. 2002. "Prevalence and Trends in Overweight among US Children and Adolescents, 1999–2000." *Journal of the American Medical Association* 288 (14): 1728–32.
- Philipson, T. J., and R. A. Posner. 1999. *The Long-Run Growth in Obesity as a Function of Technological Change*. NBER Working Paper # 7423.

- Poirier, D. J. 1978. "The Use of the Box-Cox Transformation in Limited Dependent Variable Models." *Journal of American Statistical Association* 73 (362): 284-7.
- Pylypchuk, Y., and J. Hudson. 2008. "Immigrants and the Use of Preventive Care in the United States." *Health Economics*, published online. doi: 10.1002/hec.1401.
- Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*. 2d Edition. New York: John Wiley & Sons.
- Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135-46.
- Schroeter, C., J. Lusk, and W. Tyner. 2008. "Determining the Impact of Food Price and Income Changes on Body Weight." *Journal of Health Economics* 27 (1): 45-68.
- Stettler, N., T. M. Signer, and P. M. Suter. 2004. "Electronic Games and Environmental Factors Associated with Childhood Obesity in Switzerland." *Obesity* 12 (6): 896-903.
- Stunkard, A. J., and T. A. Wadden (Eds.). 1993. *Obesity: Theory and Therapy*. 2d Edition. New York: Raven Press.
- Tunali, I. 1986. "A General Structure for Models of Double-Selection and an Application to a Joint Migration-Earnings Process with Remigration." *Research in Labor Economics* 8 (B): 235-84.
- U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS). 2000. *Continuing Survey of Food Intakes by Individuals 1994-96, CD-ROM*. Washington, DC: U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS).
- U.S. Department of Health and Human Services (USDHHS). 2000. *Healthy People 2010: Volume I and II*. 2d Edition. Washington, DC: U.S. Government Printing Office.
- . 2001. *The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity*. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General.
- Vandewater, E. A., M. Shim, and A. G. Caplovitz. 2004. "Linking Obesity and Activity Level with Children's Television and Video Game Use." *Journal of Adolescence* 27 (1): 71-85.
- Ver Ploeg, M., L. Mancino, B. Lin, and C. Wang. 2007. "The Vanishing Weight Gap: Trends in Obesity among Adult Food Stamp Participants (US) (1976-2002)." *Economics and Human Biology* 5 (1): 20-36.
- Vijverberg, W. P. M. 1993. "Measuring the Unidentified Parameter of the Extended Roy Model of Selectivity." *Journal of Econometrics* 57 (1-3): 69-89.
- Wilde, E. P., and C. K. Ranney. 2000. "The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework." *American Journal of Agricultural Economics* 82 (1): 200-13.
- World Health Organization. 1998. *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation on Obesity*. Geneva, Switzerland: World Health Organization.
- Yen, S. T. 1993. "Working Wives and Food Away from Home: The Box-Cox Double Hurdle Model." *American Journal of Agricultural Economics* 75 (4): 884-95.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Development of the Likelihood Function.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.