Association Internationale des Economistes du Sport
AIES IASE
International Association of Sports Economists

Working Paper Series, Paper No. 06-13

Economic Determinants of Participation in Physical Activity and Sport

Brad R. Humphreys† and Jane E. Ruseski††

July 2006

Abstract

This paper examines the economic determinants of participation in physical activity by developing and analyzing a consumer choice model of participation and by testing the predictions of this model using data drawn from the Behavioral Risk Factor Surveillance Survey (BRFSS). Both emphasize that individuals face two distinct decisions: (1) should I participate in sport?; and (2) how much time should I spend participating in sport? The evidence highlights the importance of selectivity. The economic factors that affect these two decisions work in opposite directions; factors that increase the likelihood of participation generally decrease the amount of time spent participating.

JEL Classification Codes: I20, I12, I18, L83

Keywords: physical activity, time allocation, health production

* Previous drafts of this paper were presented at the IASE conference “New perspectives in Sports Economics” in Bochum, Germany, May 2006 and the Western Economics Association Annual meetings, in San Diego, CA, July 2006. We thank Chad Meyerhofer, Dennis Coates, Paul Downward, and Joe Riordan for their helpful comments on earlier drafts of this paper and Chad Zanocco for his excellent research assistance.

---

†Department of Recreation, Sport and Tourism, University of Illinois, 104 Huff Hall, 1206 South Fourth Street, Champaign, IL 61820, (217) 333-4410 (phone), (217) 244-1935 (fax), brh@uiuc.edu

††Department of Kinesiology and Community Health, University of Illinois,124B Huff Hall MC588, 1206 South Fourth Street, Champaign, IL 61820 USA; Phone: 217-333-4242; Email:jruseski@uiuc.edu

Introduction

There is a growing perception among policymakers and public health researchers that individual's decisions about participating in physical activity, including sport, has an important economic component. Relatively little attention has been paid to this topic by economists. In this paper, we develop an economic model of the decision to participate in physical activity and test the predictions of this model using a nationally representative data set containing detailed information on participation in sport and other physical activities.

Much of this interest in the economic determinants of physical activity stems from the growing literature on the economic causes and consequences of obesity. Poor nutrition and physical inactivity are discretionary activities that can have a major impact on chronic diseases such as obesity (Cawley, 2004). Many plausible explanations for the rise in obesity have been advanced and a variety of policy interventions have been proposed to reduce the rate of obesity. However, the prevalence of meeting nutrition and and physical activity guidelines is low in the United States (Hill et al., 2004). Despite the important policy and public health aspects of participation in physical activity, little economic research has focused on the topic. There are a few notable exceptions to this. One is the recent research by Cawley, et al. (2005) on physical education in the United States. Another is research on participation in physical activity from a leisure demand perspective; Davies (2002) is a recent example of research in this area. Still another is statistical analysis by Farrell and Shields (2002) on the economic and demographic determinants of sporting participation in England. A final exception is the tangentially related research on the economic returns to participating in intercollegiate athletics in the United States spawned by interest in Title IX.

A possible explanation for the failure in meeting physical activity guidelines is a poor understanding of the influence of economic factors on participation in physical activity and sport. Economics is useful for furthering our understanding because it provides a framework for studying how people allocate their time to competing activities and what economic, environmental and demographic factors affect the decision to be physically active. Once the decision to participate is made, the next decisions involve what activity, how often, how intense and how long. This paper begins to bridge this gap in the literature by combining and adapting components of the SLOTH developed by Cawley (2004) and commonly used recreation and leisure demand models to investigate the economic determinants of participation in physical activity. The SLOTH framework is based on Becker's (1964) model of labor and leisure choice. This model assumes that individuals derive satisfaction from the consumption of "basic commodities" like meals. These basic commodities are produced by households using time and market commodities according to a production function. The consumption of basic goods has labor market implications because any time used to produce and enjoy basic commodities represents time spent not working. Participation in physical activity and sport is a basic commodity in this context.

The model developed in this paper incorporates the idea that individuals make two separate but related decisions: (1) should I participate in sport?; and (2) how much time should I spend participating in sport? The model generates predictions about the relationship between physical activity, economic factors like income, and the opportunity cost of time. The model developed here generates a number of predictions that provide new insight into the economic determinants of participation in physical activity and sport. For example, the model predicts that the effect of changes in income on the decision to participate in physical activity and the effect on the amount of time spent in physical activity may have opposite signs. Participation in physical activity rises with income, but time spent in physical activity declines with income. These predictions are empirically testable, and we find strong support for these predictions through the analysis of a large, nationally representative data set containing a wealth of information about individual's participation in physical activity.

A Model of Participation in Physical Activity

Our model of participation in physical activity is an extension of the SLOTH framework of time allocation. (Cawley 2004) The framework is based on Becker's (1964) model of labor and leisure choice. The basic idea is that people are involved in the production of their own health. People combine time with market goods to produce health. The SLOTH framework assumes that individuals choose how to allocate their time to maximize their utility subject to budget, time and biological constraints where SLOTH is an acronym for the activities to which individuals allocate their time. Specifically, S represents time spent sleeping; L represents time spent at leisure, O represents time spent at paid work; T represents time spent in transportation; and H represents time spent in home production, or unpaid work. Participation in physical activity and sport is captured in L, as is time spent in sedentary leisure activities such as watching television or playing computer games.

We combine the key temporal elements of the SLOTH framework with a recreation demand model (McConnell, 1992) to analyze both time allocation decisions and decisions about the purchase of good and services related to active and passive leisure. The key behavioral decisions behind our model are the separate but related decisions to participate in physical activity and how long to participate per episode of physical activity. We extend the SLOTH framework by recognizing that some of the activities in the SLOTH framework require time and goods and services purchased in the marketplace. Suppose individuals choose how to allocate their time to various activities according to the following utility function:

$$\max U(S, L_F(T_F, G), L_U(T_U, V), O + T, C(T_C, X)) \tag{1}$$

where S is time spent sleeping; L_F is a function describing leisure time devoted to physical activity; L_U is a function describing leisure time devoted to sedentary activity; $O + T$ is total time spent at work and getting to and from work; and C is a function describing time spent in household production or unpaid work. The functions L_F, L_U and C have time arguments (T_F, T_U and T_C) and market goods and services arguments (G, V and X). These functions recognize that an individual's decision to participate in physical activity requires not only time but also goods and services purchased in the marketplace. Before introducing the time and budget constraints, we adapt components of the recreation demand model to reformulate and simplify the utility function to emphasize the physical activity participation and duration decisions. Equation (1) can be restated more compactly

$$\max U(a, t, z) \tag{2}$$

where a represents the individual's decision to participate in physical activity; t is the amount of time spent per episode of physical activity; and z represents the individual's decision to engage in the other activities in the SLOTH framework. z then is composite of S, O, T, C and L_U; a is equivalent to L_F; and t is equivalent to T_F in the modified SLOTH framework.

Individuals choose how to best allocate their time and what bundle of goods and services to purchase subject to time and budget constraints. The budget constraint is

$$Y = F_a + c_a at + c_z z \tag{3}$$

where F_a is the fixed cost of engaging in physical activity; c_a is the variable cost associated with engaging in physical activity; and c_z is the cost all other goods and services. The fixed costs of physical activity are one-time costs or flat recurring costs that individuals incur to participate in physical activity but do not depend on how many times the individuals participates. An example of a fixed cost is the monthly membership dues at a health club. An individual pays this flat, fixed amount regardless of how many times he uses the gym during the month. Variable costs of physical activity are costs that do depend on the amount of time or the number of times the individual engages in physical activity. Examples of variable costs are equipment maintenance costs, coaches fees and personal trainer fees.

The time constraint is

$$T^* = at + \theta z \tag{4}$$

where T^* is the time available for consumption activities such as physical activity and θ is time spent consuming z. Assume that T^*, t and θ are measured in the same units such as hours. Let T be the total time available for work and all other activities. Hence, $T^* = T - h$ where h is time spent working. If individuals can choose the amount of hours they work, then wage earnings w can be expressed in terms of total time available and time spent not working

$$wh = w(T - at - \theta z). \tag{5}$$

Equation (5) captures the notion that any time spent in physical activity and other activities is time not available for work and reduces earnings. Thus, the wage is the opportunity cost of engaging in activities other than work. The full budget (or income) constraint includes the opportunity cost of time

$$y_0 + w(T - T^*) = F_a + p_a at + p_z z \tag{6}$$

where $p_a = c_a + w$ and $p_z = c_z + \theta w$ are the full costs of participating in physical activity and other activities.

Comparative Static Analysis

Consumers choose a, t and z to maximize utility subject to the full income constraint. The lagrangian for this problem is

$$V = U(a, t, z) - \lambda(F_a + p_a \cdot a \cdot t + p_z z - y) \tag{7}$$

The first order conditions characterizing the optimal choices of a, t and z are found by partially differentiating V with respect to the choice variables and the lagrange multiplier

$$\begin{aligned}
\frac{\partial V}{\partial a} &= \frac{\partial U}{\partial a} - \lambda p_a t &&= 0 \\
\frac{\partial V}{\partial t} &= \frac{\partial U}{\partial t} - \lambda p_a t &&= 0 \\
\frac{\partial V}{\partial z} &= \frac{\partial U}{\partial z} - \lambda p_z &&= 0 \\
\frac{\partial V}{\partial \lambda} &= -(F_a + p_a \cdot a \cdot t + p_z z - y) &&= 0.
\end{aligned}$$

We conduct a comparative static analysis of the consumer's choice problem. We analyze the effects of changes in income and the opportunity cost of time on the decisions to participate in physical activity and the amount of time spent participating in physical activity. In the comparative static analysis we treat the decision to participate in physical activity as a continuous, but discrete count variable rather than a dichotomous variable restricted to take on the values of zero or one. This approach is consistent with the time dimension of the participation in physical activity data used in our empirical analysis, the month prior to the survey. Each episode of physical activity requires a separate participation decision, so the participation decision is made repeatedly over time. As a result, observed episodes of physical activity are not limited to zero or one over the relevant time period.

We first derive comparative static expressions for the the effect of a change in income (dy) on both the participation decision a and the optimal amount of time spent in physical activity t. The comparative static expression for the effect of change in income (dy) on a holding dt, dp_a, dp_z and dF_a constant and setting $dt/dy = 0$ is

$$\frac{\partial a}{\partial y} = \frac{U_{az}p_z - p_a t U_{zz}}{p_z(-U_{aa}p_z + U_{za}p_a t) - p_a t(-U_{az}p_z + U_{zz}p_a t)} \tag{8}$$

The detailed derivation of the comparative static results are contained in the technical appendix. Convexity of the indifference curves requires the denominator of equation (8) to be positive and the sign depends on the sign of the numerator. $U_{zz} < 0$ by assumption, so the sign of the numerator depends on the first term, which contains the term U_{az} that cannot be signed *a priori*. If $U_{az} > 0$, then $\frac{\partial a}{\partial y} > 0$.

Our intuition is that the cross-partial derivative, $U_{az} > 0$, should be positive. This cross partial describes the relationship between the marginal utility from participating in physical activity and the marginal utility from other activities like meals or watching television. Participating in physical activity may lead to increased enjoyment of other non-active leisure activities. For example, if an individual decides to go to the gym and work out, the marginal utility from a meal in a restaurant later in the evening could be greater than the marginal utility received by a non-participant.

Next, we evaluate the comparative static derivative dt/dy to examine the effect of changes in income on the optimal amount of time spent in physical activity by holding da, dp_a, dp_z and dF_a constant and setting $da/dy = 0$. We hold da constant because the decision about the amount of time an individual participates in physical activity is only relevant if the individual chooses to participate. The comparative static expression is

$$\frac{\partial t}{\partial y} = \frac{U_{tz}p_z - p_a a U_{zz}}{p_z(-U_{tt}p_z + U_{zt}p_a a) - p_a a(-U_{tz}p_z + U_{zz}p_a a)} \tag{9}$$

Again, convexity of the indifference curves requires the denominator of equation (9) to be positive and the sign depends on the sign of the numerator. $U_{zz} < 0$ by assumption, so the sign of the numerator depends on the first term in the numerator, which contains the term U_{tz} that cannot be signed *a priori*. If $U_{tz} < 0$, then $\frac{\partial t}{\partial y} < 0$. This cross-partial is difficult to sign and we have no intuition about what this sign might be. However, we empirically estimate the effect of changes in income on the amount of time spent in physical activity below. This estimate will shed light on the sign of this term, and the nature of the relationship between utility derived from time spent in physical activity and utility derived from consuming other goods and spending time in other leisure activities.

The opportunity cost of time affects the decision to participate in physical activity and the amount of time devoted to physical activity. Recall, $p_a = c_a + w$ and $p_z = c_z + \theta w$. The opportunity cost of time is the wage rate w. Expanding the lagrangian to explicitly show the full cost of time spent in physical activity and all other activities is

$$V = U(a, t, z) - \lambda(F_a + (c_a + w) \cdot a \cdot t + (c_z + \theta w)z - y). \tag{10}$$

The individual's choices are the same in this expanded model; namely, to choose a, t and z to maximize utility. Differentiating the lagrangian with respect to a, t, z and λ yields the system of first order conditions

$$\begin{aligned}
\frac{\partial V}{\partial a} &= \frac{\partial U}{\partial a} - \lambda(c_a t + wt) &&= 0 \\
\frac{\partial V}{\partial t} &= \frac{\partial U}{\partial t} - \lambda(c_a a + wa) &&= 0 \\
\frac{\partial V}{\partial z} &= \frac{\partial U}{\partial z} - \lambda(c_z + \theta w) &&= 0 \\
\frac{\partial V}{\partial \lambda} &= -(F_a + (c_a + w) \cdot a \cdot t + (c_z + \theta w)z - y) &&= 0.
\end{aligned}$$

We examine the effect of a change in the opportunity cost of time, (dw), on the participation decision, a and the amount of time spent in physical activity, t. Since the participation decision and the time decision are sequential, we solve for the comparative static result, da/dw by holding dt constant and setting $\frac{dt}{dy} = 0$. The comparative static expression is

$$\frac{\partial a}{\partial w} = \frac{-(ta + \theta z)}{|J_{FI_p}|} \cdot (U_{zz} t(c_a + w) - U_{az}(c_z + \theta w) - \frac{\lambda t}{|J_{FI_p}|} \cdot ((-c_z - \theta w)^2 - \theta(-c_z - w)(-c_a - w)) \quad (11)$$

where

$$|J_{FI_p}| = (c_z + \theta w)(-U_{aa}(c_z + \theta w)) + U_{za}(t(c_a + w)) - (c_a + w)t(-U_{az}(c_z + \theta w) + U_{zz} t(c_a + w)).$$

The first term in Equation (11) resembles the familiar term from standard consumer theory for the income effect of a change in the price of a market good. In the standard consumer theory model, an increase in the price of a good effectively decreases the consumer's real income, and the income effect is greater as the importance of the good in the consumer's budget increases. The income effect of a change in the opportunity cost of time has the opposite effect on the consumer's real income than the income effect of a change in the price of some market good. This occurs because an increase in the opportunity cost of time effectively means a higher wage and an increase in real income. If physical activity is a normal good, then we would expect the income effect of an increase in the opportunity cost of time to be positive. Recall that the numerator of the first term on the right hand side of Equation (11), $ta + \theta z = T^*$, is the weighting factor on the income effect. This term is the amount of time available for all activities other than work. The total amount of time spent in all non-work activities, rather than only the time in physical activity, determines the effect of a change in the opportunity cost of time on participation. This occurs because the opportunity cost of time is the same for physical activity and all other activities and individuals are constrained by the total amount of time available.

The second term is the substitution effect of a change in the opportunity cost of time. The substitution effect is negative which means that the likelihood of participating in physical activity decreases as the opportunity cost of time increases. The sign of $\partial w/\partial a$ depends on both the sign of the income and substitution effect. We cannot sign the entire expression for $\partial w/\partial a$ because the substitution and income effect move in opposite directions. Our empirical estimates will shed some light on which effect dominates this comparative static result.

Next we examine (dt/dw) by holding da constant and setting $\frac{da}{dw} = 0$. The comparative static result is

$$\frac{\partial t}{\partial w} = \frac{-(ta + \theta z)}{|J_{FI_d}|} \cdot (U_{zz} a(c_a + w) - U_{tz}(c_z + \theta w) - \frac{\lambda a}{|J_{FI_d}|} \cdot ((-c_z - \theta w)^2 - \theta((-c_z - \theta w)(-c_a - w)) \quad (12)$$

where

$$|J_{FI_d}| = (c_z + \theta w)(-U_{tt}(c_z + \theta w)) + U_{tz}(a(c_a + w)) - (c_a + w)a(-U_{tz}(c_z + \theta w) + U_{zz}a(c_a + w)).$$

The interpretation of this comparative static result is the same as the interpretation of Equation (11). The first term is the income effect and the second term is the substitution effect. Again, this term cannot be signed *a priori* because of the opposing signs of the income and substitution effects. We empirically estimate the effect of changes in the opportunity cost of time on the amount of time spent in physical activity below.

In summary, the model developed in this section describes consumer's decisions about participating in physical activities, time spent participating in physical activities, time spent in non-active leisure activities, the purchase of other goods, services, and the time spent consuming these other goods and services. We conduct a comparative statics analysis to examine the effect of changes in income and the opportunity cost of time on participation and time spent in physical activity. To our knowledge, no previous research has developed and solved a formal consumer choice model of physical activity participation and time decisions.

Because of the lack of research in this area, a formal empirical test of some of the predictions of this model is an important step in research into the economic determinants of physical activity. In the following section we describe a large, nationally representative data set that contains a rich amount of data on participation in physical activity and other economic and demographic factors. We then use these data to test the economic predictions that emerge from the consumer choice model and estimate the effect of social and behavioral factors on physical activity.

Data Description and Sample Statistics

Since little previous research has focused on the economic determinants of participation in sport and physical activity, we empirically test the predictions of our consumer choice model in order to assess their validity. We use data from the Behavioral Risk Factor Surveillance System (BRFSS). The survey is conducted annually by telephone to a random representative sample of the population over the age of 18 in each U.S. state by the Center for Disease Control and Prevention in conjunction with U.S. states. The survey collects uniform state-specific data on preventative health factors, behavioral risk factors, and other economic and demographic characteristics and includes a rotating selection of modules one of which is on exercise and physical activity.

The BRFSS physical activity data is a rich source of information on participation in physical activities in the United States and has been used in some previous economic research. For example, Chou, et al. (2002a,2002b) used this data set to examine the link between obesity and physical activity. The survey asks about both frequency and duration of participation, which provides a relatively complete picture of self reported physical activities. The survey also asks questions about demographic factors like age, gender, race, ethnicity, and marital status, and questions about economic factors like income and labor market participation. This makes the BRFSS data an ideal setting for examining the economic determinants of physical activity. The physical activity module is not included in every year. We use data from the 2000 BRFSS survey, which included a module about physical activity and exercise.

184,450 persons were surveyed in the 2000 BRFSS survey. The 2000 survey included residents of Puerto Rico, and the exercise module was not administered to residents of Illinois that year. After excluding these observations, and some observations for individuals with a reported age under 18, a sample of 175,246 individuals remained. Table 1 shows some basic summary statistics for this sample of 175,246 individuals.

Table 1: Summary Statistics

Variable	Mean	Standard Deviation
Age	46.88	17.21
Female	0.59	0.49
Black	0.08	0.27
Hispanic	0.07	0.25
Married	0.54	0.50
Number of Children	0.75	1.17
Employed	0.64	0.48
Retired	0.18	0.38
Income	$46,524	34,003
High School Dropout	0.12	0.32
High School Graduate	0.32	0.47
Some College	0.28	0.45
College Graduate	0.29	0.45
Health "Excellent"	0.22	0.42
Health "Very Good"	0.33	0.47
Health "Good"	0.29	0.47
Physically Active	0.72	0.45
Average Times per Week Participating	2.36	2.49
Average Minutes Per Week Participating	197	330

The average age of an individual in the sample was just under 47 years. 59% of the individuals sampled were female. In terms of minority representation, the sample was 8% black and 7% Hispanic. These categories are mutually exclusive in terms of race and ethnicity in the BRFSS survey methodology, which divides the sample into four categories ("white non-Hispanic," "black non-Hispanic," "black," and "other"). Over half of the respondents were married, and the average number of children present in the household was 0.75. The average number of children present in households that have children is 1.96 and the average number of children in households with a married couple and at least one child is 2.01.

64% of those surveyed were employed. Those who were not in the labor force were identified as short and long term unemployed persons, students, homemakers, people unable to work because of disabilities, and retired persons. 19% of the respondents were retired, 12% dropped out of high school, 32% were high school graduates, 28% had attended one or more years of college without graduating, and 29% were college graduates.

The BRFSS survey asks respondents about income from all sources. This is somewhat limiting because it potentially includes income from sources like pensions, capital, and government assistance programs in addition to income earned from work. Time allocation decisions depend heavily on the opportunity cost of time, which is related to the hourly wage. To the extent that the income variable reported in the BRFSS survey includes unearned income, this variable will be a poor proxy for the hourly wage. The BRFSS survey reports income in ranges. The ranges in the survey are less than $10,000, between $10,000 and $15,000, between $15,000 and $20,000, between $20,000 and $25,000, between $25,000 and $35,000, between $35,000 and $50,000, between $50,000 and $75,000, and greater than $75,000. Following Ruhm (2005), the level of income for each individual is coded as the midpoint of the range reported, or 150% of the unbounded top range. Only 150,648 people responded to the income question in the 2000 BRFSS survey. This sub-sample forms the basis for the empirical work in the following sections. From Table 1, the average level of income in the sample was $46,524.

Physical Activity Measures

The 2000 BRFSS survey contained a module of questions on physical activity. These questions were asked to the entire sample except residents of Illinois. The basic physical activity question in the BRFSS survey is

> *During the past month, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?*

We initially define participation in physical activity using this survey question. From Table 1, 72% of the sample answered yes to this question. This is a relatively high participation rate for physical activity, but the question is not qualified by any statement about the duration of the activity, and many people may answer yes even when they spend very little time participating in physical activity. Fortunately, BRFSS contains more detailed questions about physical activity and exercise than a simple participation question. The survey also elicits detailed information about the type of activity, the frequency of participation, and the duration of participation in physical activity. The first question that elicits additional detail about the type of physical activity undertaken is

> *What type of physical activity or exercise did you spend the most time doing during the past month?*

The responses to this question were a long list of activities from running to raking the lawn. Table 2 contains the entire list of physical activities reported and their frequency in the sample. The responses on Table 2 show a wide variation in reported physical activities in the sample. The activities also differ in a number of important ways. They require different amounts of travel, equipment, time and effort on the part of the participant. Some, like walking or running, can be done alone while others, like soccer or volleyball, require additional participants; still others can be done alone or in groups. Because of the heterogeneity of the activities, future research will consider grouping the activities by common characteristics like equipment required, level of exertion, calories burned and other identifiable factors.

Walking is by far the most common physical activity. Just over 50% of the sample, 65,620 individuals, reported walking as their primary or secondary type of physical activity over the past month. No other type of physical activity was even close. The high reported frequency of walking as the primary form of physical activity in the sample probably reflects the relatively low cost of this activity. Unlike many of the other activities listed on Table 2, walking does not require any specialized equipment or facilities. It can be done in almost any setting under almost any conditions.

The survey asks about a first and a second activity participated in over the last month. 31% of those surveyed reported participating in two activities over the previous month. In this paper, we simply add up time spent in the primary and secondary physical activity to get a total measure of time spent in physical activity per week.

BRFSS also contains detailed information about how frequently individuals in the survey participated in physical activities in the past month, and how much time the individuals spent in each activity on average. These data provide enough detail to construct an estimate of the number of times per week and minutes per week that each individual in the survey spent participating in some

Table 2: Distribution of Physical Activities

Activity	Frequency	Percent
Walking	65620	50.20
Gardening	9213	7.05
Running	8172	6.25
Weightlifting	5437	4.16
Other	4581	3.50
Aerobics class	4357	3.33
Bicycling for pleasure	4032	3.08
Golf	3948	3.02
Home exercise	3250	2.49
Basketball	2424	1.85
Health club exercise	2319	1.77
Swimming laps	2148	1.64
Jogging	1719	1.32
Calisthenics	1608	1.23
Bicycling machine exercise	1272	0.97
Hiking cross country	1080	0.83
Tennis	947	0.72
Softball	791	0.61
Dancing-aerobics/Ballet	784	0.60
Mowing lawn	585	0.45
Bowling	544	0.42
Soccer	505	0.39
Judo/Karate	479	0.37
Snow skiing	465	0.36
Volleyball	476	0.36
Horseback riding	438	0.34
Hunting large game - deer, elk	436	0.33
Skating - ice or roller	426	0.33
Fishing from riverbank or boat	312	0.24
Racqetball	299	0.23
Stair climbing	301	0.23
Boxing	178	0.14
Surfing	159	0.12
Snow shoveling by hand	136	0.10
Carpentry	113	0.09
Waterskiing	115	0.09
Boating (canoeing, rowing, sailing)	100	0.08
Raking lawn	110	0.08
Touch football	104	0.08
Canoeing, rowing in competition	84	0.06
Rowing machine exercise	74	0.06
Mountain climbing	65	0.05
Badminton	35	0.03
Painting/Papering house	43	0.03
Snow shoeing	37	0.03
Backpacking	27	0.02
Rope skipping	26	0.02
Scuba diving	32	0.02
Sledding, tobogganing	26	0.02
Table tennis	21	0.02
Handball	14	0.01
Squash	13	0.01
Stream fishing in waders	12	0.01
Snorkeling	10	0.01
Snow blowing	9	0.01
Paddleball	4	0.00

physical activity. In estimating the duration of participation, we included the reported amount of time spent in both the primary and secondary physical activity.

The last two lines of Table 1 show the average frequency of participation and time spent participating in the sample. The average person in the sample participated in physical activity one and a third times per week. The average amount of time per week spent participating in physical activity was 197 minutes, about three and one quarter hours. A forty hour work week contains 2400 minutes.

Note that the survey asks about physical activities undertaken in the past month. The survey is administered throughout the year. Note that the sample does not constitute a panel. Each individual is contacted only once, but because of the large sample size, the survey takes a year to complete. Many physical activities take place outdoors, leading to some seasonal variation in participation. Table 3 shows the rate of participation in physical activities in the sample by month. The average participation rate of 72% across all months in the sample clearly masks considerable seasonal variation in participation. This seasonal variation is probably due to changes in climate. Runners and joggers may only exercise during warm months in northern states. We account for this seasonal variation in the participation rate in the empirical work described below and its effect on time spent in physical activity.

Table 3: Participation by Month of Survey

Month	% Participating
January	0.67
February	0.67
March	0.71
April	0.72
May	0.76
June	0.77
July	0.78
August	0.76
September	0.75
October	0.73
November	0.71
December	0.67

Empirical Analysis of Participation in Physical Activity

The decision to participate in physical activity conceptually resembles the familiar labor supply decision from labor economics. In this context, individuals have an expected benefit from participating in physical activity and face a shadow price of their leisure time that depends on the hourly wage and other factors. In labor supply models, utility maximizing individuals compare their reservation wage to the wage they can earn in the labor market and participate in the labor market (supply a positive number of hours of work) if the market wage is greater than or equal to their reservation wage. Heckman (1974, 1976) developed several widely used estimators that can be applied to these situations. Here, if the expected benefit of participating in physical activity exceeds the shadow price of an individual's time, then that individual will participate in physical activity.

Let A_i be the amount of time that individual i spends in some physical activity, X_i be a vector of variables, including characteristics of individual i that might explain the time that individual i spends in physical activity and β be a vector of unobservable parameters. The data set used here contains both individuals who participate in physical activity ($i = 1, 2, \ldots N_1$ where $A_i > 0$) and individuals that do not participate in physical activity ($i = N_1 + 1, N_1 + 2, \ldots, N$, where $A_i = 0$). The set of individuals who participate in physical activity will be referred to as S_1 and the set of individuals who do not participate will be referred to as S_2.

Given the available data, it would be possible to estimate an equation explaining observed time spent in physical activity

$$A_i = \beta X_i + e_i \tag{13}$$

where e_i is an unobservable mean zero, constant variance random variable capturing factors other than X_i that affect individual is decision to participate in physical activity. However, if the time spent in physical activity by non-participants is set to zero and the parameters of equation (13) estimated using the Ordinary Least Squares (OLS) estimator, the parameter estimates will be inconsistent because the model incorrectly assumes that equation (13) can be applied to all individuals in the sample. This is the well-known selectivity problem in econometrics. Heckman (1974, 1976) developed a two-step procedure to deal with selectivity of this type. The Heckman selectivity correction is based on a reduced form approach to individual's participation decision.

Note that if all individuals in the sample participated in physical activity, then the expected value of the time spent in physical activity would be

$$E[A_i] = \beta X_i$$

but when $A_i = 0$ for some individuals the expected value of the time spent in physical activity is

$$E[A_i] = Prob(A_i > 0) \cdot E[A_i | A_i > 0] + Prob(A_i \leq 0) \cdot 0.$$

Applying Heckman's approach implies that individuals make two choices related to physical activity: a choice to participate in physical activity (the participation decision) and a choice about how much time to spend in physical activity conditional on the decision to participate (the time decision). To implement Heckman's approach, partition X_i into two sets of variables (X_{i1}, X_{i2}) where X_{i1} affects the participation decision and X_{i2} affects the time decision.

Given this partitioning of X_i, the time decision can be expressed

$$A_i = \beta_1 X_{i1} + u_i \quad \text{if} \quad A_i = \beta_1 X_{i1} + u_i > 0$$

and otherwise

$$A_i = 0.$$

The participation decision is modeled as a function of observable factors (X_{i2}), an unobservable mean zero, constant variance random error term (ν_i), and some unobservable factor w_i* that captures the benefit that the individual gets from participating in physical activity. Let σ_ν be the variance of ν. If $w_i^* > 0$ then the individual participates in physical activity but if $w_i* \leq 0$ then the individual does not participate. Formally

$$\begin{aligned} A_i > 0 \quad &\text{if} \quad w_i^* > 0 \\ A_i = 0 \quad &\text{if} \quad w_i^* \leq 0 \end{aligned}$$

and w_i^* is determined by

$$w_i^* = \beta_2 X_{i2} + \nu_i.$$

The covariance between ν_i and u_i is $\sigma_{u\nu}$. This can be shown to equal $\sigma_\nu \sigma_u \rho$ where ρ the coefficient of correlation between the two error terms.

This model for the determination of w_i^* implies a selection rule based on the sign of this unobservable variable

$$\begin{array}{lll} A_i > 0 & \text{if} & \nu_i > -\beta_2 X_{i2} \\ A_i = 0 & \text{if} & \nu_i \leq -\beta_2 X_{i2}. \end{array}$$

This selection rule simply indicates that an individual compares the benefit of participating in physical activity, reflected by the realization of ν_i to the cost of participating in the activity, represented by $\beta_2 X_{i2}$. If the benefit exceeds the cost, then w_i^* is positive and the individual participates. If the cost exceeds the benefit, then w_i^* is negative and the individual does not participate. Based on this selection rule, the expected value of A_i is

$$E[A_i] = \Phi\left(\frac{\beta_2 X_{i2}}{\sigma_\nu}\right) \cdot E[A_i | \nu_i > -\beta_2 X_{i2}] + \left[1 - \Phi\left(\frac{\beta_2 X_{i2}}{\sigma_\nu}\right)\right]$$

where $\Phi(\cdot)$ is the standard normal distribution function. Also note that

$$\begin{aligned} E[A_i | \nu_i > -\beta_2 X_{i2}] &= \beta_1 X_{i1} + E[u_i | \nu_i > -\beta_2 X_{i2}] \\ &= \beta_1 X_{i1} + \frac{\sigma_{u\nu}}{\sigma_\nu} \\ &= \beta_1 X_{i1} + \rho \sigma_u h\left(\frac{\beta_2 X_{i2}}{\sigma_\nu}\right). \end{aligned}$$

The Heckman two step procedure is a sequential procedure:

1. Estimate β_2/σ_ν using the probit estimator. This involves maximizing the likelihood function

$$\ell = \prod_{i \in S_2} \left[1 - \Phi\left(\frac{\beta_2 X_{i2}}{\sigma_\nu}\right)\right] \cdot \prod_{i \in S_1} \Phi\left(\frac{\beta_2 X_{i2}}{\sigma_\nu}\right) \tag{14}$$

   with respect to β_2/σ_ν.

2. Use the estimate of β_2/σ_ν to estimate $h(\cdot)$ and add this variable to the time equation. The expanded time equation

$$A_i = \beta_1 X_{i1} + \frac{\sigma_{u\nu}}{\sigma_\nu} h\left(\frac{\beta_2 X_{i2}}{\sigma_\nu}\right) + u_i \tag{15}$$

   is estimated using OLS.

The OLS estimator generates unbiased and consistent estimates of the parameters of equation (15) because of the correction for selectivity captured by equation (14). For the Heckman procedure to work, X_{i2} must contain some variables not in X_{i1}, and these variables must identify the participation decision.

Results and Discussion

Again, the process of participating in physical activity clearly includes selectivity. Individuals are deciding to either participate in a physical activity or not participate, and after making the participation decision they determine how long to participate. We use the two step Heckman procedure to account for selectivity in the behavior of the survey participants. The first stage of the Heckman procedure analyzes the participation decision. This involves estimating a model with a discrete dependent variable that is equal to one if the individual participates in some physical activity and is equal to zero if the individual does not participate. The vector of explanatory variables must contain some variables that do not enter the time equation in order to identify the participation decision. Ideally, these variables should explain some of the observed participation behavior.

The parameters of the first stage regression model are estimated using the probit estimator. This is a maximum likelihood estimator that determines the values of β_2/σ_ν, the ratio of the parameter estimates to the standard error of the residual, that minimizes the likelihood function defined by equation (14). We include the individual's age, the number of children in the indivdual's household, and a series of indicator variables for gender, race, ethnicity, education, reported health, the month the individual was surveyed, and the state where the individual lives as explanatory variables in the first stage regression.

Table 4 shows the results of the estimation of the first stage selectivity model using maximum likelihood. In general, the parameter estimates from models with discrete dependent variables cannot be directly interpreted like the parameters of other regression models. However, Table 4 reports the probability derivative

$$b_i^* = \left. \frac{\partial \Phi(\beta_2 X_{i2})}{\partial X_{i2}} \right|_{x=\overline{x}} = \phi(\overline{X}_{i2} b_2) b_i$$

which can be interpreted as the change in the probability that the dependent variable is equal to one for a one unit change in the explanatory variable.

In general, the explanatory variables are significant and of the predicted sign. Each additional year of age reduces the probability that an individual participates in some physical activity by three tenths of one percentage point, or a reduction of two percent per decade. Each additional child in the household reduces the probability that an individual participates in some physical activity by just more than one percentage point. Married people are one per cent less likely to participate in physical activity.

Higher income is associated with a higher probability of participating in some physical activity, but the size of the effect is relatively small. Each additional ten thousand dollars in income increases the probability that an individual participates in some physical activity by one percent. This effect probably reflects the cost of equipment and facilities associated with participation in many of the activities. Higher income allows individuals to afford gym memberships, tennis rackets, home exercise equipment and other goods and services needed to participate in these physical activities. This also verifies the predication that $\partial a/\partial y > 0$ that emerges from the model developed above. Recall that the explanation for this is that the cross partial derivative of the utility function with respect to a and z, participation in physical activity and time spent (and market supplied goods and services purchased) in non-active leisure activities.

Employed persons are about one percent less likely to participate in physical activity than the reference category, which includes long and short term unemployed persons and persons, persons unable to work because of disability, students, and homemakers. The difference would probably be larger if disabled and discouraged workers were excluded. If employed people have a higher

Table 4: Participation Equation Estimation Results

Variable	Probability Derivative	P-value
Age	-0.003	0.000
Married	-0.013	0.000
Number of Children	-0.011	0.000
Income (thousands)	0.001	0.000
Employed	-0.008	0.013
Retired	0.076	0.000
High School Graduate	0.054	0.000
Some College	0.107	0.000
College Graduate	0.156	0.000
Female	-0.021	0.000
Black	-0.043	0.000
Hispanic	-0.059	0.000
Reported Health Excellent	0.163	0.000
Reported Health Very Good	0.141	0.000
Reported Health Good	0.081	0.000
January	-0.108	0.000
February	-0.121	0.000
March	-0.067	0.000
April	-0.047	0.010
May	-0.014	0.018
July	-0.007	0.260
August	-0.009	0.125
September	-0.028	0.001
October	-0.047	0.000
November	-0.076	0.000
December	-0.116	0.000
Observations	150,648	
Pseudo R^2	0.091	

opportunity cost of time than the unemployed, home makers, and students, then the sign of this parameter suggests that the unsigned cross partial derivative from the model, U_{tz} is negative.

The dummy variables for educational attainment exhibit an interesting pattern. The probability of participation increases with the level of education. The omitted category is people who did not complete high school. High school graduates are less likely to participate in physical activities than high school dropouts, those who attended some college are more likely to participate than high school graduates, and college graduates are more likely to participate than those with some college. This pattern could be due to occupational sorting of individuals with different levels of education. For example, if high school dropouts tend to work part time or seasonally, they might have more free time to participate in physical activities than high school graduates, if high school graduates tend to work full time. Alternatively, Grossman's (1972) model of health production predicts that individuals with more education are more efficient in producing health. If participation in physical activity leads to improved health, then the estimated parameters on the education variables could be part of the mechanism through which individuals produce better health, supporting the predictions of Grossman's model.

After controlling for differences in income and the presence of children, females are two percent less likely to participate in some physical activity than males. This probably reflects greater responsibility in childcare and home production activities. Blacks and Hispanics are less likely to participate in some physical activity than whites. Many of these differences may simply reflect that employed minorities have poor access to the goods and services needed to participate in physical than whites.

The reported health of individuals has a significant impact on participation in physical activity. The reference group is individuals who reported their general health as "fair" or poor. It is important to remember that this estimated is conditional on the effect of age on participation in some physical activity. Individuals in poor health would be expected to walk less, or perform any physical activity. Also, since the participation decision is estimated using maximum likelihood, any correlation between reported health and the true health of the individual, which will be captured by the equation error term in the participation decision model, will not lead to econometric problems.

The month indicator variables are also highly significant and follow a logical pattern. Recall that the survey asks about participation in a physical activity in the previous month, and individuals are surveyed throughout the calendar year. Many physical activities take place outside, and a fair number may have strong seasonal participation components, like gardening. The pattern on the month dummy variables shows that individuals in the sample were more likely to report participation in a physical activity in the summer and less likely to report participation in the winter. Again, note that the data do not constitute a panel. This seasonal variation in participation is is across different individuals in the sample.

The model included a set of dummy variables indicating the state in which individuals lived. These state effects might capture the quantity and distance of recreation areas and sports facilities, climate, and variation in travel costs due to transportation infrastructure and traffic congestion. The estimates on these variables are not reported, but many of them were significant and the signs were both positive and negative. These state indicator variables may capture a variety of underlying factors related to climate, geography, population density, patterns of urbanization, transportation networks, and the supply of goods and services related to physical activity like playing fields, sporting goods stores, and health clubs across states. The significance of the state of the state indicators suggests that future research should focus on collecting additional state-specific data to quantify these factors, in order to better identify specific factors that affect participation in physical activity. Overall, the model explains just under six percent of the observed variation in participation.

Table 5: Time Equation Estimation Results

Variable	Parameter Estimate	P-value
Age	0.908	0.000
Married	-18	0.000
Income (thousands)	-0.270	0.000
Employed	31	0.000
Retired	-0.178	0.947
High School Graduate	-37	0.000
Some College	-67	0.000
College Graduate	-104	0.000
Female	-75	0.000
Black	15	0.003
Hispanic	26	0.000
Constant	519	0.000
λ	-324	0.000
Wald χ^2	8539	0.000

In the second stage, the time equation, equation (15) is estimated using OLS. Table 5 contains the parameter estimates and P-values from estimating equation (15) with OLS. The vector of month indicator variables was also included in the time equation, but the estimated parameters on these variables are not reported. Most were significant, and indicated that time spent in physical activity was higher in the summer and lower in other months. This regression model explains variation in the amount of time individuals spend participating in physical activity. The second term on the right hand side of equation (15) is calculated from the first stage regression, and is sometimes referred to as the "inverse Mills ratio" in the literature. The parameter on this variable indicates the correlation between the error term on the selectivity equation and the time equation, and including this term ensures that the other parameter estimates do not reflect selectivity bias.

The identifying variables in the first stage regression are the number of children present in the household, the three variables indicating reported health status, and the vector of dummy variables indicating state of residence. The parameter estimates for the time equation were not sensitive to including the reported health status variables and the state of residence indicators in the second stage.

The estimates of the time equation reveal several interesting features. First, after correcting for selectivity, the sign on the age parameter is positive so individuals who choose to participate in physical activity tend to increase the time spent as age increases, other things constant. The increase in time spent in physical activity is about one minute per year. Married people spend about 18 minutes less per week participating in physical activities than single people.

The parameters on the race, gender and ethnicity show interesting patterns. Although blacks and Hispanics are less likely to participate in physical activity, those who do choose to participate spend more time in physical activity than whites. One implication of this result is that interventions aimed at increasing the participation of these groups might be very effective, in that the individuals induced to switch from non-participation to participation would spend a significant amount of time engaging in physical activity.

Females spend less time in physical activities than males. This difference could be due to occupational choices, if females tend to sort themselves into occupations that require more hours of work, or offer less job flexibility than males. Examples of such occupations include nursing, primary and secondary education, and secretarial work. If could also reflect differences in the underlying types of physical activities preferred by males and females.

Education has a negative effect on the amount of time individuals spend participating in physical activity, and the decrease in time spent in physical activity increases with the level of education. This could reflect occupational sorting, if more education leads individuals to work in positions that provide less flexibility in working hours, or require the individual to work more than 40 hours per week. Note that the effect of increasing education on the decision to participate in physical activity is opposite to the effect on increasing education on time spent participating in physical activity. Employment has a positive effect on the amount of time spent in physical activity. Employed people spend 32 more minutes per week in physical activity than the reference group, which includes homemakers, students, disabled people, and the unemployed.

We interpret the educational and employment variables as proxies for the opportunity cost of time. In general, the effect of changes in the opportunity cost of time on time spent in physical activity has two possible effects. Higher opportunity cost of time is positively related to higher hourly earnings, so as hourly income rises, the opportunity cost of any non-work related activity increases, and individuals will spend less time participating in these activities; this is a substitution effect. But participation in physical activity, and other leisure and recreation activities, are normal goods, and people demand more of all normal goods as increases in the hourly wage raise income; this is the income effect. The substitution effect tends to reduce time spent participating in physical activity and the income effect tends to increase time spent participating in physical activity. The results on Table 5 indicate that the substitution effect dominates in these data, to the extent that individuals with higher levels of education tend to have higher opportunity cost of time. It is also interesting that the effect of increasing education on time spent participating in physical activity is opposite to the effect of increasing education on the participation decision. The positive sign on the employment variable suggests that the income effect dominates on this margin. Employed people have a higher opportunity cost of time than homemakers, students, and the unemployed. However, this higher opportunity cost is in part due to higher hourly wages, which also affects time spent in physical activity through the income effect.

The effect of increases in income on time spent in physical activity is negative. Individuals with higher income spend less time participating in physical activity. This verifies the predication that $\partial t/\partial y < 0$ that emerges from the model developed above. It is also opposite in sign from the effect of increases in income on the decision to participate in physical activity. The opposing signs on income in the participation equation and the time equation makes the selectivity correction used here very important to understanding the overall effect of changes in income on participation in physical activity. In particular, ignoring the effect of changes in income on the participation decision would clearly lead to inappropriate inferences about the overall effect of income on physical activity.

λ is the inverse of the Mills ratio. From equation (15), the coefficient on this variable shows the correlation between the equation error for the selectivity equation and the equation error on the time spend equation. Since the estimated parameter on λ is negative, there is negative correlation between these two variables. Unobservable factors which lead to an increased probability of participation reduce time spent participating in physical activities. The Wald χ^2 statistic is a test of the validity of the selectivity correction. The null hypothesis is no selectivity effects. This null is not accepted, suggesting that the selectivity correction is important in this context.

Conclusions and Suggestions for Future Research

This research examines participation and time spent in physical activity by developing a consumer's choice model containing these elements and by empirically testing some predications generated by this model using a large nationally representative data set. A number of interesting conclusions emerge from the analysis. The predictions of the model are supported by the empirical results. Increasing income has a positive effect on participation in physical activity and a negative effect on time spent in physical activity. Changes in the opportunity cost of time, as captured by employment status and educational attainment, have a mixed effect on both participation and time spent. This is to be expected given the offsetting income and substitution effects in this setting. The empirical support for this model suggests that this approach may be useful in further research on the economic determinants of physical activity.

The model furthers our understanding of individual's decisions about participation and time spent in physical activity by embedding these factors in a standard consumer choice model and expanding the budget constraint to include a limit on the total amount of time available for all activities and by modeling the full cost of participating in physical activities. This approach expands on the $SLOTH$ framework developed by Cawley (2004) and generates specific, empirically testable predications about the effects of changes in income and the opportunity cost of time on the decisions to participate and spend time in physical activity.

The results have important implications for those designing policy interventions aimed at increasing participation in physical activity. The presence of children reduces the probability of participation, so successful policy interventions should be linked to daycare in some way. Participation appears to decline with age, but time spent increases with age. These results suggest that programs aimed at increasing participation in older populations might be particularly effective, as the time spent in physical activity in this population would be relatively large. Both participation and time spent in physical activity display seasonal variation; both decline during cold weather months and increase during warm weather months. These results suggest that policy interventions aimed at increasing physical activity should take into account this seasonal variation. Finally, the model and empirical results suggest that the opportunity cost of time plays a key role in both the participation and time decision. Any policy interventions that ignore this dimension of the decision to participate in physical activity may not be very effective.

The empirical results underscore the importance of selectivity in understanding the economic determinants of physical activity. Individuals make two related choices, a participation decision and a time decision. The sign of the parameter on the inverse Mills ratio, and the clear rejection of the null in the Wald test clearly show the importance of correcting for selectivity in this setting. Because the effect of the selectivity is strongly negative – factors that increase the likelihood of participation tend to reduce time spent – correcting for the effects of selectivity are crucial to a complete understanding of the economic determinants of physical activity. Ignoring the effects of selectivity will clearly lead to incorrect inference in empirical analysis, and might also lead to ineffective policy interventions, if they are designed based on results that do not account for the effects of selectivity. Our results suggest this might be the case for policy interventions targeted by gender, race and ethnicity.

The geographic indicator variables are significant in the participation decision but we cannot learn much about what underlying factors contribute to the observed variation in the participation across states with these data. Supply side factors, climatic factors, and differences in commuting time, transportation networks and amount of urban sprawl across states may affect the participation decision. We plan to collect additional state-specific data in future research to learn more about the specific factors that explain variation in participation in future research.

While the model provides new insight into economic determinants of participation and time spent in physical activity, it also has considerable room for improvement. One clear extension of the model is to include physical activity as an input to the production of health. This extension should allow us to examine the economic links between physical activity and obesity, and also explicitly link physical activity to the consumption of health goods and services. Grossman's (1972) model of health production provides one possible way to expand this model.

Finally, the decision to participate in physical activity needs to be explicitly linked to economic outcomes like employment and earnings. Previous research by Long and Caudill (1991), Barron, et al. (2000), and Eide and Ronan (2001) show a clear link between participation in physical activity and labor market outcomes and lieftime earnings. This suggests an important link between participation in physical activity and human capital and labor productivity. Much of the previous literature focused on participation in team sports in secondary schools and college. The importance of age in explaining observed participation and time spent in the broad measures of physical activity examined here suggest that a closer look at the relationship between this type of activity and labor market outcomes warrants additional attention.

References

Barron, J., B. Ewing and G. Waddell, (2000). "The Effects of High School Athletic Participation on Education and Labor Market Outcomes," *Review of Economics and Statistics,* 82(3), pp. 409-421.

Becker, G. (1964). "A Theory of the Allocation of Time," *The Economic Journal*, 75(299), pp. 493-513.

Cawley, J. (2004). "An Economic Framework for Understanding Physical Activity and Eating Behaviors," *American Journal of Preventive Medicine*, 27(3s), pp. 117-125.

Cawley, J., C. Meyerhofer and D. Newhouse (2005). "The Impact of State Physical Education Requirements on Youth Physical Activity and Overweight," NBER Working paper 11411.

Chou, S., M. Grossman and H. Saffer (2002a). "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System," NBER Working paper 9247.

Chou, S., M. Grossman and H. Saffer (2002b). "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System," *Journal of Health Economics* 23(3), pp. 565-587.

Davies, L. (2002). "Consumers' Expenditure on Sport in the UK: Increased Spending or Underestimation?" *Managing Leisure*, 7(1), pp. 83-102.

Eide, E. and N. Ronan, (2001). "Is Participation in High School Athletics an Investment or a Consumption Good? Evidence from High School and Beyond," *Economics of Education Review*, 20(5), pp. 431-442.

Farrell, L. and Shields, M. (2002). "Investigating the Economic and Demographic Determinants of Sporting Participation in England," *Journal of the Royal Statistical Society*, 165(Part 2), pp. 335-348.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), pp. 223-255.

Heckman, J. J. (1974). "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42(4), pp. 679-694.

Heckman, J. J. (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5(4), pp. 475-492.

Hill, J., J. Sallis, and J. Peters (2004). "Economic Analysis of Eating and Physical Activity: A Next Step for Research and Policy Change," *American Journal of Preventive Medicine*, 27(3s), pp. 111-116.

Long, J. and S. Caudill, (1991). "The Impact of Participation in Intercollegiate Athletics on Income and Graduation," *Review of Economics and Statistics*, 73(3), pp. 525-532.

McConnell, K. (1992). "On-Site Time in the Demand for Recreation," *American Journal of Agricultural Economics*, 74(4), pp. 918-925.

Ruhm, C. (2005). "Healthy Living in Hard Times," *Journal of Health Economics*, 24(2), pp. 341-363.

Technical Appendix

Consumers choose a, t and z to maximize utility subject to the full income constraint. The lagrangian for this problem is

$$V = U(a, t, z) - \lambda(F_a + p_a \cdot a \cdot t + p_z z - y) \tag{16}$$

The first order conditions characterizing the optimal choices of a, t and z are found by partially differentiating V with respect to the choice variables and the lagrange multiplier

$$\begin{aligned}
\frac{\partial V}{\partial a} &= \frac{\partial U}{\partial a} - \lambda p_a t &= 0 \\
\frac{\partial V}{\partial t} &= \frac{\partial U}{\partial t} - \lambda p_a t &= 0 \\
\frac{\partial V}{\partial z} &= \frac{\partial U}{\partial z} - \lambda p_z &= 0 \\
\frac{\partial V}{\partial \lambda} &= -(F_a + p_a \cdot a \cdot t + p_z z - y) &= 0.
\end{aligned}$$

Totally differentiate each first order conditions and set this total differential equal to zero to form the system of equations needed for a comparative static analysis of the consumer's choice problem

$$\begin{aligned}
U_{aa}da + U_{at}dt + U_{az}dz - \lambda t dp_a - \lambda p_a dt - p_a t d\lambda &\equiv 0 \\
U_{ta}da + U_{tt}dt + U_{tz}dz - \lambda a dp_a - \lambda da p_a - p_a a d\lambda &\equiv 0 \\
U_{za}da + U_{zt}dt + U_{zz}dz - \lambda dp_z - p_z d\lambda &\equiv 0 \\
-dF_a - p_a t da - p_a a dt - a t dp_a - z dp_z - p_z dz + dy &\equiv 0.
\end{aligned}$$

The system of total differential equations can be expressed compactly in matrix form

$$\begin{bmatrix} U_{aa} & U_{at} - \lambda p_a & U_{az} & -p_a t \\ U_{at} - \lambda p_a & U_{tt} & U_{tz} & -p_a a \\ U_{za} & U_{zt} & U_{zz} & -p_z \\ -p_a t & -p_a a & -p_z & 0 \end{bmatrix} \begin{bmatrix} da \\ dt \\ dz \\ d\lambda \end{bmatrix} = \begin{bmatrix} \lambda dp_a t \\ \lambda dp_a a \\ \lambda dp_z \\ a t dp_a + z dp_z + dF_a - dy \end{bmatrix} \tag{17}$$

and this coefficient matrix is the familiar Jacobian, $|J|$, from standard consumer theory. The Jacobian forms the basis of a comparative static analysis of the effect of changes in exogenous variables on the choice variables. We analyze the effects of changes in income and changes in the opportunity cost of time on the decision to participate in physical activity and on the decision about how much time to spend participating in physical activity.

We first derive comparative static expressions for the the effect of a change in income (dy) on both the participation decision a and the optimal amount of time spent in physical activity t. To find the effect of change in income (dy) on a and t holding dp_a, dp_z and dF_a, divide the system of total differential equations through by dy

$$\begin{bmatrix} U_{aa} & U_{at} - \lambda p_a & U_{az} & -p_a t \\ U_{at} - \lambda p_a & U_{tt} & U_{tz} & -p_a a \\ U_{za} & U_{zt} & U_{zz} & -p_z \\ -p_a t & -p_a a & -p_z & 0 \end{bmatrix} \begin{bmatrix} \frac{da}{dy} \\ \frac{dt}{dy} \\ \frac{dz}{dy} \\ \frac{d\lambda}{dy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}.$$

Since the participation and time decisions are separate decisions, we first solve for the comparative static derivative da/dy by holding dt constant and settng $dt/dy = 0$. In matrix form, the restricted model becomes

$$\begin{bmatrix} U_{aa} & U_{az} & -p_a t \\ U_{za} & U_{zz} & -p_z \\ -p_a t & -p_z & 0 \end{bmatrix} \begin{bmatrix} \frac{da}{dy} \\ \frac{dz}{dy} \\ \frac{d\lambda}{dy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

This matrix is the Jacobian matrix for the participation decision and is denoted $|J_p|$. The comparative static derivative in matrix form is

$$\frac{\partial a}{\partial y} = \frac{|J_a|}{|J_p|} = \frac{\begin{vmatrix} 0 & U_{az} & -p_a t \\ 0 & U_{zz} & -p_z \\ -1 & -p_z & 0 \end{vmatrix}}{\begin{vmatrix} U_{aa} & U_{az} & -p_a t \\ U_{za} & U_{zz} & -p_z \\ -p_a t & -p_z & 0 \end{vmatrix}}.$$

Finding the determinants $|J_a|$ and $|J_p|$ and substituting yields

$$\frac{\partial a}{\partial y} = \frac{|J_a|}{|J_p|} = \frac{U_{az} p_z - p_a t U_{zz}}{p_z(-U_{aa} p_z + U_{za} p_a t) - p_a t(-U_{az} p_z + U_{zz} p_a t)} \tag{18}$$

Next, we solve for the comparative static derivative dt/dy to examine the effect of changes in income on the optimal amount of time spent in physical activity by holding da constant and setting $da/dy = 0$. We hold da constant because the decision about the amount of time an individual participates in physical activity is only relevant if the individual chooses to participate. The restricted model in matrix form becomes

$$\begin{bmatrix} U_{tt} & U_{tz} & -p_a a \\ U_{zt} & U_{zz} & -pz \\ -p_a a & -p_z & 0 \end{bmatrix} \begin{bmatrix} \frac{dt}{dy} \\ \frac{dt}{dy} \\ \frac{d\lambda}{dy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

This is the Jacobian matrix for the time decision and is denoted $|J_d|$. The comparative static result is

$$\frac{\partial t}{\partial y} = \frac{|J_t|}{|J_d|} = \frac{\begin{vmatrix} 0 & U_{tz} & -p_a a \\ 0 & U_{zz} & -pz \\ -1 & -p_z & 0 \end{vmatrix}}{\begin{vmatrix} U_{tt} & U_{tz} & -p_a a \\ U_{zt} & U_{zz} & -p_z \\ -p_a a & -p_z & 0 \end{vmatrix}}.$$

Solving for the determinants $|J_t|$ and $|J_d|$ and substituting yields

$$\frac{\partial t}{\partial y} = \frac{|J_t|}{|J_d|} = \frac{U_{tz}p_z - p_a a U_{zz}}{p_z(-U_{tt}p_z + U_{zt}p_a a) - p_a a(-U_{tz}p_z + U_{zz}p_a a)} \tag{19}$$

The opportunity cost of time affects the decision to participate in physical activity and the amount of time devoted to physical activity. Recall, $p_a = c_a + w$ and $p_z = c_z + \theta w$. The opportunity cost of time is the wage rate w. Expanding the lagrangian to explicitly show the full cost of time spent in physical activity and all other activities is

$$V = U(a,t,z) - \lambda(F_a + (c_a + w) \cdot a \cdot t + (c_z + \theta w)z - y). \tag{20}$$

The system of total differential equations expressed compactly in matrix form is

$$\begin{bmatrix} U_{aa} & U_{at} - \lambda(c_a + w) & U_{az} & -(c_a t + wt) \\ U_{ta} - \lambda(c_a + w) & U_{tt} & U_{tz} & -(c_a a + wa) \\ U_{za} & U_{zt} & U_{zz} & -(c_z + \theta w) \\ -t(c_a + w) & -a(c_a + w) & -(c_z + \theta w) & 0 \end{bmatrix} \begin{bmatrix} da \\ dt \\ dz \\ d\lambda \end{bmatrix} = \begin{bmatrix} \lambda(dc_a t + dwt) \\ \lambda(dc_a a + dwa) \\ \lambda(dc_z + dw\theta + d\theta w) \\ I \end{bmatrix}$$

where

$$I = atdc_a + (at + \theta z)dw + zdc_z + wzd\theta + dF_a - dy.$$

In this expression the coefficient matrix is the Jacobian $|J|$ for the system of equations based on the expanded or full income constraint. This matrix is denoted $|J_{FI}|$.

We examine the effect of a change in the opportunity cost of time, (dw), on the participation decision, a and the amount of time spent in physical activity, t. Divide the system of total differential equations through by dw, holding dc_a, dc_z, dF_a, and $d\theta$ constant. The system in matrix form becomes

$$\begin{bmatrix} U_{aa} & U_{at} - \lambda(c_a + w) & U_{az} & -(c_a t + wt) \\ U_{ta} - \lambda(c_a + w) & U_{tt} & U_{tz} & -(c_a a + wa) \\ U_{za} & U_{zt} & U_{zz} & -(c_z + \theta w) \\ -t(c_a + w) & -a(c_a + w) & -(c_z + \theta w) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial a}{\partial w} \\ \frac{\partial t}{\partial w} \\ \frac{\partial z}{\partial w} \\ \frac{\partial \lambda}{\partial w} \end{bmatrix} = \begin{bmatrix} \lambda t \\ \lambda a \\ \lambda \theta \\ ta + \theta z \end{bmatrix}$$

Since the participation decision and the time decision are sequential, we solve for the comparative static result, da/dw by holding dt constant and setting $\frac{dt}{dy} = 0$. The model in matrix form becomes

$$\begin{bmatrix} U_{aa} & U_{az} & -(c_a t + wt) \\ U_{za} & U_{zz} & -(c_z + \theta w) \\ -t(c_a + w) & -(c_z + \theta w) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial a}{\partial w} \\ \frac{\partial z}{\partial w} \\ \frac{\partial \lambda}{\partial w} \end{bmatrix} = \begin{bmatrix} \lambda t \\ \lambda \theta \\ ta + \theta z \end{bmatrix}$$

The coefficient matrix is the Jacobian matrix for the participation decision in the model with expanded income constraint and is denoted $|J_{FI_p}|$. The comparative static derivative is

$$\frac{\partial a}{\partial w} = \frac{|J_{FI_a}|}{|J_{FI_p}|} = \frac{\begin{vmatrix} \lambda t & U_{az} & -(c_a t + wt) \\ \lambda \theta & U_{zz} & -(c_z + \theta w) \\ ta + \theta z & -(c_z + \theta w) & 0 \end{vmatrix}}{\begin{vmatrix} U_{aa} & U_{az} & -(c_a t + wt) \\ U_{za} & U_{zz} & -(c_z + \theta w) \\ -t(c_a + w) & -(c_z + \theta w) & 0 \end{vmatrix}}$$

Finding the determinants of $|J_{FI_a}|$ and $|J_{FI_p}|$, substituting and rearranging yields

$$\frac{\partial a}{\partial w} = \frac{-(ta+\theta z)}{|J_{FI_p}|} \cdot (U_{zz}t(c_a+w) - U_{az}(c_z+\theta w) - \frac{\lambda t}{|J_{FI_p}|} \cdot ((-c_z-\theta w)^2 - \theta(-c_z-w)(-c_a-w)) \quad (21)$$

where

$$|J_{FI_p}| = (c_z+\theta w)(-U_{aa}(c_z+\theta w)) + U_{za}(t(c_a+w)) - (c_a+w)t(-U_{az}(c_z+\theta w) + U_{zz}t(c_a+w)).$$

Next we examine (dt/dw) by holding da constant and setting $\frac{da}{dw} = 0$. The matrix equation becomes

$$\begin{bmatrix} U_{tt} & U_{tz} & -a(c_a+w) \\ U_{zt} & U_{zz} & -(c_z+\theta w) \\ -a(c_a+w) & -(c_z+\theta w) & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial t}{\partial w} \\ \frac{\partial z}{\partial w} \\ \frac{\partial \lambda}{\partial w} \end{bmatrix} = \begin{bmatrix} \lambda a \\ \lambda\theta \\ ta+\theta z \end{bmatrix}$$

The coefficient matrix is the Jacobian for the time decision in the model with the expanded income constraint and is denoted $|J_{FI_d}|$. The comparative static derivative is

$$\frac{\partial t}{\partial w} = \frac{|J_{FI_t}|}{|J_{FI_d}|} = \frac{\begin{vmatrix} \lambda a & U_{tz} & -a(c_a+w) \\ \lambda\theta & U_{zz} & -(c_z+\theta w) \\ ta+\theta z & -(c_z+\theta w) & 0 \end{vmatrix}}{\begin{vmatrix} U_{tt} & U_{tz} & -a(c_a+w) \\ U_{zt} & U_{zz} & -(c_z+\theta w) \\ -a(c_a+w) & -(c_z+\theta w) & 0 \end{vmatrix}}$$

Finding the determinant of $|J_{FI_t}|$ and $|J_{FI_d}|$, substituting and rearranging yields

$$\frac{\partial t}{\partial w} = \frac{-(ta+\theta z)}{|J_{FI_d}|} \cdot (U_{zz}a(c_a+w) - U_{tz}(c_z+\theta w) - \frac{\lambda a}{|J_{FI_d}|} \cdot ((-c_z-\theta w)^2 - \theta((-c_z-\theta w)(-c_a-w)) \quad (22)$$

where

$$|J_{FI_d}| = (c_z+\theta w)(-U_{tt}(c_z+\theta w)) + U_{tz}(a(c_a+w)) - (c_a+w)a(-U_{tz}(c_z+\theta w) + U_{zz}a(c_a+w)).$$