An Empirical Model of Labor Supply with Social Interactions: Econometric Issues and Tax Policy Implications

Andrew Grodner
Department of Economics at East Carolina University
A423 Brewster Building, East Carolina University, Greenville, N.C. 27858
agrodner@andrewgrodner.com

Thomas J. Kniesner
Center for Policy Research and Department of Economics at Syracuse University,
426 Eggers Hall, Syracuse University, Syracuse, N.Y. 13244–1020
tkniesne@maxwell.syr.edu

November 14, 2005

Abstract

Our research econometrically tests for social interactions in labor supply. Interdependence is a response of a person's hours worked to hours worked in the reference group, which includes those sharing similar age, family structure, and location. We identify endogenous spillovers by instrumenting average hours worked in the reference group with hours worked in adjacent reference groups. Estimates of the canonical labor supply model indicate positive and non-negligible spillovers for adult men. We estimate a total wage elasticity of labor supply of 0.22, where 0.08 is the exogenous wage change effect and 0.14 is the social interactions effect. Ignoring or incorrectly considering social interactions mis-estimates the labor supply response of tax reform by ±50–60 percent.

Keywords: labor supply, social interactions, reference group, instrumental variables, social multiplier, PSID.

JEL Codes: J22, Z13

We acknowledge the financial support of the Program of Fellowships for Junior Scholars, MacArthur Research Network on Social Interactions and Economic Inequality. The authors also thank Badi Baltagi, Dan Black, Gary Engelhardt, Bo Honore, Bill Horrace, Jeff Kubik, Anil Kumar, Joseph Marschand, Christian Meghea, Jan Ondrich, and the audience members at Syracuse University, Dalhousie University, East Carolina University, the Southern Economic Association, the 2005 World Congress of the Econometric Society, and the National Tax Association for their helpful comments.

1. Introduction

Social interactions, defined as a response of individuals to the actions of people with whom they interact, may have a biological basis or stem from information gathering. Social interactions are a potentially important aspect of economic behavior because interdependencies can affect the behavioral responses of people to the expected and unexpected changes in their environment, including ones caused by public policy. We investigate the importance of social interactions in labor supply with taxes where the interdependence is a response of the individual to the hours worked by the members of a reference group.

To identify the reference group we define economic distance between individuals as a combination of personal characteristics and physical distance. The specification lets us examine whether hours worked of individuals in close economic proximity are related. We use an instrumental variables technique to control for the potential endogeneity of hours worked by the workers in the individual's reference group. Panel Study of Income Dynamics data from 1976 anchor our work to the influential studies on male labor supply by Hausman (1981) and MaCurdy et al. (1990). We find evidence of positive spillover effect in hours worked that is important for tax policy and demonstrate how the definition of the reference group plays a major role in validating the magnitude of the effect.

1.1 Key Issues

There is wide-ranging evidence that people in close proximity may have a significant effect on the individual's decisions. Researchers have identified the presence of interdependence in the decisions of giving (Andreoni and Scholz, 1998), voting (Schram and Sonnemans, 1996), consumption (Sen et al., 2001, Childers and Rao, 1992, Abu-Ismail, 1992), crime (Glaeser et al., 1996), and health (Eibner and Evans, 2005). Interdependent behavior is also present in labor economic research. There is interest in social interactions in the female labor supply (Woittiez and Kapteyn, 1998), male labor supply with taxes (Aronsson et al., 1999), retirement (Hamermesh and Slemrod, 2005), and job satisfaction (Hamermesh, 1977, 2001).

Along with evidence that there may be social interactions present, there is an equally vast literature showing that the interdependence is either economically insignificant, non-existent, or econometrically fragile. For example, studies point out the importance of the reference group choice when studying teenage behavior (Kooreman and Soetevent, 2002), cohabiting (Jacques and Chason, 1978), and workplace interaction (Baker et al., 1968). The selection of the instrumental variable used to identify social interactions can also dramatically change the results (Evans, Oates, and Schwab, 1992).

The variety of interdependence studies and their conflicting conclusions underline that identification of social interactions is econometrically complex. The first challenge the researcher need confront is the specification of the correct reference group (Durlauf, 2004). The reference group identity issue is largely ignored due to complexity, with a notable exception by Woittiez and Kapteyn (1998) and Sun (2005).

The second challenge emerges because interdependence means that people respond to the behavior of others in their reference group, so we should observe correlated behaviors in neighborhoods, regions or cities. However, there are other reasons why we may observe correlated behavior; without additional information it is impossible to distinguish endogenous effects from exogenous or contextual effects (Manski 1993).

Despite the difficulties of undertaking social interaction research, the potential benefits should outweigh the costs. Specifically, once interdependence is identified a researcher can perform a more complete welfare analysis. Policy oriented economic models can relate social interactions to health outcomes (Deaton, 2001), measurement of poverty (Pradhan, 2001), or effects of taxes (Aronsson et al., 1999; Abel 2005). Studies suggest that improving the situation of the neighborhoods, or movement of individuals between neighborhoods can greatly affect the social welfare in the local communities. A recent issue of the *Journal of Applied Econometrics* (Special Issue: Empirical Analysis of Social Interactions, 2003, September/October, Vol. 18, No. 5) also presents recent applications to peer effects in colleges, social interactions in housing demand, and interdependence in worker's productivity.

1.2 Overview of Our Research

We address the issues related to the difficulty in identifying the effect of endogenous social interactions on an individual's actions. We create a flexible measure of the economic distance that approximates the level at which individuals are interacting among each other and that reflects the varying costs of interaction (higher economic distance implies higher cost of interaction, which implies a lower level of interaction). Next we define the reference groups, each of which consists of persons who are in a close economic proximity, and compute hours worked for each person in the reference group (endogenous social interactions). We create an instrument from the mean of hours worked for persons who are in the adjacent reference group to instrument endogenous social interactions. Econometric results suggest positive and non-negligible social interactions in hours worked. The total wage elasticity of labor supply is 0.22, where about one-third (0.08) is due the exogenous wage change and about two-thirds (0.14) is due to social interaction synergies. We demonstrate how improperly accounting for social interactions can lead to mis-estimation of the labor supply effects of tax reform by ±60 percent.

2. Theory

Theories of social interactions have a long history in the economic literature. Since Becker (1974), the evolution of economic theory includes developing many forms of interactions: social norms (Lindbeck, Nyberg, and Weibull, 1999), peer-group effects (de Bartolome Charles, 1990), neighborhood effects (Durlauf, 1996), conformity effects (Bernheim, 1994), herding (Smith and Sorensen, 2000), spillovers (Roback, 1982), contagion (Rigobon, 2001), social capital (Glaeser et al., 2002; Becker and Murphy, 2000), and positional goods (Frank, 1985). In most cases the method differs according to the application, from an overlapping generations framework to a Bayesian learning model. Theoretical exercises share the common feature that the utility of the individual is somehow affected by either utility or choices made by members of the reference group, who are people with whom the individual interacts.

In our theoretical model we follow Brock and Durlauf (1995) where interactions are introduced in a baseline model with additive total utility consisting of individual utility and social utility. We assume that the economy is at the equilibrium in the sense of Durlauf (1996) with a developed neighborhood structure. In what follows we use the terms membership group, neighborhood, and community as equivalent and meaning persons who are part of the individual's reference group.

Consider a labor supply model with a negative spillover effect in hours worked:

$$V_{ig}\left(c_{ig}, h_{ig}; b_g\left(\mu_h\right)\right) = u_{ig}\left(c_{ig}, T - h_{ig}\right) - b_g\left(\mu_{hg}\right)s\left(h_{ig}\right)$$
$$\text{st.}\, c_{ig} \le h_{ig} w_g, \tag{1}$$

where $V_{ig}(\bullet)$ represents total utility of person i who belongs to the reference group g, $u(\bullet)$ represents a private utility over c (consumption) and leisure $(T-h)$, where T is total available time and h is hours worked/labor supply, and $\left[-b_g\left(\mu_h\right)s\left(h_{ig}\right)\right]$ is total social disutility of working. Unlike the canonical model here total disutility of hours worked depends on the level of $b_g(\bullet)$, which represents the importance of social disutility. For the individual i in reference group g, $b_g(\bullet)$ is increasing in average hours worked in the reference group, μ_{hg}, excluding the ith worker (thus $\mu_{hg} = \overline{h}_{(-i)g}$), with $b_g(0) = 0$, $b_g(\infty) \to \infty$, and $b_g^{'} > 0$. Total social disutility also depends on $s(\bullet)$, which is the social disutility of individual hours worked (disutility of the individual from how others judge his or her work level) with $s(0) = s_0 > 0$, $s(\infty) \to 0$, $s' < 0$, and $s'' > 0$; s_0 is autonomous social disutility, which is equal across individuals and reference groups. Finally, w_g is a wage rate in the reference group g.

Social disutility of individual's hours worked $s(\bullet)$ is always non-zero with the maximum value s_0 at zero hours worked.[1] Social disutility of ones hours worked seems most likely to be decreasing $(s' < 0)$ at a decreasing rate $(s'' > 0)$. The decrease of the social disutility means that as individuals work more hours they believe that others judge them less harshly. A decrease of social disutility at a decreasing rate means that as individuals work more hours the gain of appearing better in the eyes of peers is getting smaller. The worker may view certain levels of hours worked as satisfactory and care less and less about opinions of others as long as the worker reaches some accepted levels of hours worked (according to his or her personal belief system).[2]

Our final maintained hypothesis is that the importance of the social utility term, b, is increasing in the average hours worked in the individual's reference group $(b_g' < 0)$. So, when a worker sees that the environment is filled with hard-working people the worker will believe to be judged more if he or she "sticks out" more relative to the labor market performance of others. The individual may feel more negatively perceived if further down the ranking of work effort.

After setting up the Lagrangian, taking the total differential of the first-order conditions of (1), and performing comparative statics based on the properties of social interactions in labor supply just described, we obtain the result that:

$$\frac{dh}{d\mu_h} = \frac{-\overbrace{b'}^{(+)}\overbrace{s'}^{(-)}}{\underbrace{bs''(h)}_{(+)} + \underbrace{2wu_{ch} - u_{hh} - w^2 u_{cc}}_{(+)}} > 0 \qquad (2)$$

with the partial derivatives of private utility, $u_{cc} < 0$ and $u_{hh} < 0$.[3]

Equation (2) shows that an increase in average hours worked in the reference group increases the individual's hours worked. The intuition is that when the average labor supply increases, the parameter b increases, social disutility increases, and total utility decreases. To find a new maximum total utility the worker increases hours worked; although utility decreases because hours worked are a bad ($u_h < 0$), an increase in the labor supply reduces social disutility because $s' < 0$. Overall, an increase in hours worked increases total utility because the decrease in social disutility is higher than the decrease in individual utility. The model suggests that workers who are in an environment with a relatively many hard working people are induced to work more hours than when there is no social interactions effect.

We test the hypothesis that there are social interactions present in labor supply by regressing individuals' hours worked on average hours worked in their reference groups. A positive coefficient indicates the presence of a positive spillover effect in hours worked (Woittiez and Kapteyn 1999, Aronsson et al 1999). We now flesh out the econometric details of testing for social interactions in individual labor supply.

3. Econometric Model

Algebraically, $y_{ig} = \theta_0 + \theta_1 x_{ig} + \theta_2 f\left(y_{(-i)g}\right) + \theta_3 h\left(x_{(-i)g}\right) + \theta_4 u_g + \varepsilon_{ig}$, is the canonical social interactions model where y_{ig} is the behavioral outcome of interest for individual i who is a member of the reference group g, x_{ig} is a set of exogenous characteristics of an individual, $\left(\bullet\right)_{(-i)g}$ denotes other individuals in the neighborhood besides individual i, $f\left(\bullet\right)$ and $h\left(\bullet\right)$ denote some functions of outcomes or characteristics of other individuals in the neighborhood that affect individual i, u_g denote neighborhood specific effects (local amenities, institutions), ε_{ig} is white noise, and θ_k denotes coefficients.

We use here the analog canonical linear labor supply model that is

$$h = \theta + \alpha\omega + \beta\upsilon + \gamma x + \rho h_{-1} + \delta_1 \overline{h}_{(-i)g} + \delta_2 \overline{x}_{(-i)g} + \varepsilon , \qquad (3)$$

where ω is the after-tax real wage, υ is after-tax virtual income, x is a vector of individual control covariates, h_{-1} is lagged labor supplied, $\overline{h}_{(-i)g}$ is reference group average labor supplied, $\overline{x}_{(-i)g}$ is the vector of control covariate averages for the reference group, ε is the error term, and [θ, α, β, γ, ρ, δ_1, δ_2] are parameters to estimate.

3.1 Independent Variables

The net wage rate uses a marginal tax rate τ provided by the PSID, and is $(1-\tau)w$. Virtual income also uses the marginal tax rate from the PSID.[4] To control for possible endogeneity when estimating equation (3) we instrument both the after tax wage and virtual income using last year's gross wage and non-labor income (Ziliak and Kniesner 1999).

The control covariates in labor supply include number of children less than six years old, family size, an indicator if the person is more than 45 years old, the equity the family has in their house, and an indicator of a physical or nervous condition that limits the amount of work. The control covariates are standard exogenous explanatory variables in labor supply studies.

We also include hours worked in the previous year. The parameter ρ represents adjustment costs or habit formation, the degree to which an individual chooses the same hours worked from year to year. In later discussions of results we follow the habit-formation interpretation of Woittiez and Kapteyn (1998).

3.2 Social Interactions Variables

The mean for hours worked in the reference group is the sample average of hours worked for other people who are close in economic distance to the worker. In the computing the average we exclude the individual for whom we are computing a reference group mean outcome. The estimated value of the parameter δ_1 represents the effect of endogenous social interactions in hours worked.

Computing the mean of covariates takes multiple steps. We first create a proxy variable summarizing the information in the exogenous covariates. We then use factor analysis and take the first factor as a proxy variable for exogenous information. The new variable does not have a direct interpretation because it is standardized to have zero mean and unit variance, however it is highly correlated with all the exogenous variables as well as the individual's hours worked. The mean in the reference group for the created proxy variable uses the same range of the economic distance variables as used for computing mean hours worked, again excluding of the individual for whom we are computing the reference group mean. The proxy variable controls for the common characteristics of the reference group, and the estimated coefficient δ_2 indicates any presence of exogenous social interactions.

3.3 Identifying Social Interactions

The form of the equation in (3) can identify the presence of both endogenous (in the dependent variable) and exogenous (in the independent variables) social interactions. However it is well known that without additional structure none of the parameters are identified (Manski 1993, Moffitt 2001). If the reference groups are completely separable then only a randomly distributed shock that affects hours worked for some individuals and not others can help identify endogenous social interactions (Moffitt, 2001). When reference groups overlap there are a variety of empirical approaches including repeated samples (Aronsson et al., 1999), structural models (Brock and Dulauf, 2002, Kapteyn et al. 1997, Krauth forthcoming), aggregated data (Glaeser et al. 2002), within versus between variation (Graham and Hahn, 2005), or spatial econometric techniques (Kelejian and Prucha, 1998).

Alternatively, if there are workers who belong to more than one reference group, and one uses them to compute the (endogenous) mean for reference group hours worked, then hours worked by people in the adjacent reference group can be an instrument. This is much like using past values of the dependent variable in a dynamic panel data model (Arellano and Bond, 1991). We use instrumental variables techniques, where the instrument is the mean for workers in the adjacent reference groups, which are defined by a social grid with two social coordinates from factor analysis. The instrument is correlated with mean hours worked in the individual's reference group (endogenous social interactions) because individuals in the specific reference group and the adjacent reference group belong to the same economic neighborhood. The instrument should be uncorrelated with unobservables affecting individual labor supply because the particular individual does not belong to the adjacent reference group. Details of the identification strategy are explained further in the Appendix.

4. Data

We use data from the University of Michigan's Panel Study of Income Dynamics (PSID) collected in years 1975 and 1976 (PSID wave IX). One reason for using the PSID is that it is the most frequently used data to study U.S. labor supply (Blundell and MaCurdy (1999), Ziliak and Kniesner (1999)). Another reason for selecting the PSID data is that they have also been used in much influential research on how taxes affect labor supply (Hausman (1981) and MaCurdy et al. (1990)).

4.1 Sample

We follow the process of the data selection in Eklöf and Sacklén (2000) who compare studies by Hausman (1981) and MaCurdy et al. (1990). Both studies estimate an almost identical linear labor supply model with income taxation. We select the observations according to the following criteria: married males 26–55 years old with positive hours worked in 1974 and 1975 (but no higher than 5096 annual hours), who are heads of households only in the cross-sectional random sub-sample; there were no changes in the family composition of the head or wife (others can change) in years 1974–1975; the head is not retired, permanently disabled, housewife, student, or other; the household resides in the United States; and the head is not self-employed or a farmer.[5] Using our exclusion criteria for the 1976 PSID we obtain 1077 observations, which is close to the Hausman sample of 1084 and the MaCurdy sample of 1018 as reported by Eklöf and Sacklén (2000).[6]

4.2 Individual Regression Variables

The wage rate comes from a direct question in the PSID, including an imputed value for workers who are not paid by the hour.[7] We also estimate a wage equation to impute hourly wages for observations with unobserved or truncated wages. In particular, we use observations that have positive and not top-coded wage rates (839 observations) to estimate a Tobit regression that uses as the dependent variable observed (un)truncated wages on a constant term, age, age squared, years of schooling, years of schooling squared, college degree, and family size. We then use the estimated wage equation to impute all wages. The procedure is similar to that in Hausman (1982), and so our mean hourly wage is \$6.17, which nearly identical to the \$6.18 reported by Hausman.

Hours worked, the dependent variable, also comes from a directly asked question in the PSID. Non-labor income is the difference between total 1975 taxable income of the husband and wife and total 1975 labor earnings of the husband. The hours worked and the non-labor income measures we use are also those of MaCurdy et al. (1990). Other independent variables include number of children less than six years old (KIDSU6), family size (FAMSIZ), an indicator variable for individuals more than 45 years old (AGE45), the amount of equity the family had in its house (HOUSEQ), and an indicator of a physical or nervous condition that limited the amount of work the respondent could do (BHLTH). Table A1 presents descriptive statistics for all regression variables.

4.2 Reference Group and Economic Distance

The definition of the individual's reference group is central in any study of interdependence (Manski 1993, 2001). Implementing the reference group concept means paying special attention to the fact that people who are in a relative proximity to each other may interact with one another because the cost of interactions among them is low. Here we use the concept of economic distance among individuals as an indicator of the potential significance and magnitude of the interdependence (Conley, 1999). We further assume that people who are close in economic distance belong to the same reference group. Economic distance is a combination of whether the workers are similar to each other demographically and in close physical proximity. We use a combination of personal and family characteristics to describe similar individuals and we use the distance between centers of counties in which individuals reside as a proxy for the relative location of the family.

To overcome the multiple difficulties of selecting from a large variety of characteristics to measure economic distance, that each characteristic variable has a difference scale, and determining the relative importance of each input variable on the economic distance, we use a statistical model of factor analysis (Woittiez and Kapteyn 1998). The factor analytic model deals naturally with characteristics having different measurement scales; the procedure standardizes individual variables then fits a linear model to find common latent variables called factors (. The intuition is that there are unobservable factors that are orthogonal to one another and that are strongly correlated with observed variables. We use the factors and call them social coordinates.

Our factor analytic model also uses all independent variables from the econometric labor supply model (3) because we assume that variables explaining labor supply can affect whether individuals interact with each other because they could be related to economic distance. We also use physical coordinates indicating the location of individuals that are the center of the county where the family resides. We use two factors to summarize demographic and physical coordinates because usually there is a much better fit with multiple factors than with only one factor, but using too many factors tends to be uninformative.[8] When we use two factors we also have a convenient feature that the computed latent variables serve as two social coordinates (SocCoord1, SocCoord2) for where individuals are located on a social interactions grid with economic distance measured by Euclidean distance between two points.

5. Estimates of Labor Supply with Social Interactions

In our study there is no clearly defined reference group so we first select persons who have interdependent choices of hours worked using social coordinates to define overlapping neighborhoods. Once the reference group is defined, we estimate the labor supply model in (3), and use instrumental variables for identification. Finally, we interpret the social interactions effects in terms of endogenous and exogenous effects.

5.1. Selecting the Reference Group

Because we do not have direct information who belongs to the reference group for a particular individual we use a statistical procedure inferring it from the location and characteristics of the group's members. We believe that our observations are representative for working married men in terms of various individual characteristics and spatial distribution.

Considering that both social coordinates are scaled similarly, we can think of the reference group as a ring of certain radius centered around a particular individual. The problem remaining is to select the radius that best proxies for the borders of the reference group. The borders selection problem is important because we use sample observations to compute the characteristics of close-by individuals (not averages from the Census, for example). Due to relatively small numbers of cases the potential effect of the measurement error may be greater. Each observation establishes possible multiple reference groups so that careful selection of borders is critical for identification here.

To find borders for the membership groups we use a result from spatial econometrics that as the reference group size expands the coefficient on endogenous social interactions tends to minus infinity (Kelejian and Prucha, 2002).[9] In our application endogenous social interactions are represented by the mean of hours worked by those in the worker's reference group, AnnHSRG_0_*R*, where *R* represents the radius of the reference group's circle. Although the group mean coefficient will become more negative as the neighborhood size increases, if there are social interactions present at a certain size of the reference group the upward bias in the endogenous variable will then overcome the statistical tendency for the coefficient to become negative (Anselin 1988).

In Table 1 we present results from a baseline labor supply regression with a non-instrumented social interactions variable, AnnHSRG_0_*R*. Estimation starts with $R = 1$, which means that the mean of hours worked uses nearby workers in the social space within the distance of 0.1 or less. It amounts to the neighborhood as having around 13 workers. As the size of the reference group increases in the social space (R increases), the number of individuals who are considered to be economically close to a particular worker increases from 44 to about 271.

As expected, the coefficient on the average hours worked for neighboring individuals is increasingly negative, going from about −0.2 to −1.5. Such a tendency will be observed for any estimator, including OLS or GMM (Kelejian and Prucha, 2002). What is crucial here is that at the size of the reference group of 0.2 (AnnHSRG_0_2) the coefficient becomes positive. Because for a simple regression the arithmetic maximum coefficient value is about −0.2, we conclude that there is an upward bias due to the endogeneity of the AnnHSRG variable. So, for the endogeneity caused by labor supply interdependence, which is the strongest for range (0,0.2), we pick 0.2 that as the radius most closely reflecting the true size of the reference group.[10] The practical consequence of our reference group selection technique is that it is about 44 persons large, which is small enough to guarantee sufficient variation between various subgroups but large enough that the computed average hours worked are meaningful and have relatively small error due to aggregation.

5.2. Social interactions and Habit Formation Effects

The focus of our research is on identifying the interdependence in hours worked using the canonical model of labor supply. We first confirm that our estimates for the uncompensated wage and income elasticities are within the range obtained in the literature. From the first column of Table 2, the baseline model's wage and income coefficients create an uncompensated wage elasticity at the means that is 0.14 and an income elasticity at the means that is –.008; both values are well within the typical range as reported in the econometric labor supply literature.

Our main results are presented in column two of Table 2, where we include both a habit formation effect (past hours worked) and social interactions effect (hours worked in the reference group). The habit formation effect interpretation is that, on average, men would work at about 60 percent of the level they have chosen last year, and the rest is due to other labor supply determining factors. The social interactions effect is one of the most important determinants of labor supply because every 10 hours increase in the reference group would increase individual's hours worked by about 6 hours. Both habit formation and social interaction effects are significant statistically and economically and are reasonable in magnitude.[11]

It is important to re-emphasize that the estimated social interactions effect, $\hat{\delta}_1$, the impact of average hours worked by persons in the worker's reference group (AnnHSRG_0_2), has the expected sign and magnitude only after the interdependence has been estimated by instrumental variables. Least Squares results in Table 1 are biased because they suggest the presence of endogenous social interactions due to the feedback effect (Durbin-Wu-Hausman test rejects at the 5 percent level).[12] Because of the difference between OLS and IV results we need emphasize the method to construct the instrument.

Details of the identification strategy are explained in the Appendix. Because we have no obvious variables to provide exogenous variation we use the structure of the data to construct the reference group work effort instrument. Taking reference groups as overlapping with boundaries as fixed, average hours worked for persons in the adjacent reference groups can be instruments. The outer boundary of the individuals for the instrument group will be exactly twice the size of the radius for each neighborhood because there may be workers who are located exactly on the boundary for both the reference group of interest and the adjacent reference group.[13] We construct hours worked for individuals in the outside ring of (0.2, 0.6], which amounts to an average of 226 observations for each instrument group; we also use the average of the proxy variable for all exogenous variables constructed using factor analysis (IndVORG_2_6). The Sargan test results in Table 2 confirm that our instruments are valid.

Notice too the importance of including past hours in the regression. Last year's labor supply controls for the inertia when workers choose to supply labor. To some extent hours worked from 1975 also control for past social interactions effects. Most importantly, we can address the dynamic nature of utility optimization for individuals most likely consider the flow of wages to choose the profile of labor supply. The importance of using last year's annual hours as a regressor is underlined by the results in column four of Table 2; estimated social interactions effects are uneconomic, when lagged labor supply is ignored, which we interpret causing omitted variable bias.[14]

5.3. Interpreting the Social Interactions Effect

The presence of social interactions in labor supply means that individuals respond to others' hours worked by a non-negligible amount. A social interactions effect is important here because policy affecting the wages or another independent variable of a subgroup will not only affect particular individuals but also affect others in the reference group. We therefore focus on the direct versus the indirect effect of interdependence. In particular, we study the consequences of interdependence for the estimated effect of wages on labor supply, which widely used in welfare effect simulations of tax reform proposals.

Taking the mean values in equation (3) and focusing on hours worked and wages,

$$\bar{h} = \alpha\bar{\varpi} + \delta_1\bar{h} \Rightarrow \bar{h} = \alpha\frac{1}{1-\delta_1}\bar{\varpi}, \tag{4}$$

where the quantity $1/\left(1-\delta_1\right)$ is known as the global social multiplier (Glaeser et al. 2003).[15] The total effect of a wage change can be decomposed into

$$\partial\bar{h}/\partial\bar{\varpi} = \frac{\alpha}{1-\delta_1} = \alpha + \frac{\alpha\delta_1}{1-\delta_1}, \tag{5}$$

where α is the exogenous effect, and $\left(\alpha\delta_1\right)/\left(1-\delta_1\right)$ is the endogenous effect.[16] Notice that the endogenous effect depends on both the magnitude of the initial exogenous change and the social multiplier.

Multiplying equation (5) by $\bar{\varpi}/\bar{h}$ the uncompensated elasticity is

$$\eta_{hw,total} = \eta_{hw,exogenous} + \eta_{hw,endogenous}, \tag{6}$$

where $\eta_{hw,exogenous} = \alpha\bar{\varpi}/\bar{h}$ and $\eta_{hw,endogenous} = \alpha\delta_1\bar{\varpi}/\left(1-\delta_1\right)\bar{h}$. For $\delta_1 < 0.5$ the exogenous effect is always higher than the endogenous effect, but for $\delta_1 > 0.5$ the endogenous effect is higher. As we will later emphasize, the decomposition in (6) underscores how ignoring labor supply interdependencies may have serious consequences for the elasticity estimates of interest.

Using the values from column two of Table 2, the total uncompensated wage elasticity of labor supply at the means is 0.22, with an exogenous part of 0.08, and a endogenous part of 0.14.[17] In comparison, the baseline model results from column one of Table 2 are an uncompensated net wage elasticity of 0.13. When social interactions are ignored the exogenous wage effect is overestimated by about 60 percent, and the positive bias in the canonical model is because the effect of interdependence is imbedded in the single (wage) coefficient estimate. The twin findings that the wage elasticity has two unequal and sizeable parts in the social interactions model and that the wage coefficient of the traditional model suffers from sizeable omitted variable bias have important consequences for evaluating tax policy.

5.4. Implications for Tax Policy Calculations

Theoretical solutions to optimal static or dynamic taxation in the presence of social interactions externalities require values for the parameters of the utility and attendant consumption and labor supply functions (Kooreman and Schoonbeek 2004, Abel 2005). Our research clarifies the econometric subtleties of implementing labor supply models with spillover effects and then presents some of the first econometric estimates of the importance of social interactions in labor supply. The most basic results are that U.S. male labor supply data (1) reject a model ignoring social interactions against one with spillovers and (2) reject a model with spillovers treated as exogenous against one with spillovers treat as endogenous. The wage elasticity of labor supply is underestimated by about 60 percent in a model that ignores spillovers in labor supply, and the wage elasticity is overestimated by about 60 percent if one uses a social interactions model but ignores endogenous interactions as a component of the overall wage effect.

It is less obvious is how one should apply estimates that let the policy-maker apportion the total wage elasticity into segments with and without social interactions. Some back-of-the-envelope calculations are instructive. The results of the preferred model in Table 2, column 2, are that a 10 percent comprehensive tax cut would raise male labor supply by as much as 2.2 percent when social interactions are considered, but that ignoring social interactions would lead to a 40 percent under-estimate (1.3 percent). Less well established is how to use in policy calculations our decomposition of the total wage elasticity into its exogenous component (+0.08) and its endogenous social interactions component (+0.14).

To flesh out the enriched implications of a labor supply model with social interactions suppose there were a demographically based tax reform that applied only to families of a certain size or to persons of a certain marital status such that the tax rate reduction applied to 5o0 percent of the people. The effective social interactions part of the overall wage elasticity would be halved to about +0.07. In the case of a tax reform that applied to half the men the appropriate wage elasticity would be about +0.15. The correct estimate of the labor supply effect of a 10 percent tax cut would be +1.5 rather than 2.2 percent; improperly processing the interactions effects information in the econometric results could lead to approximately a 50 percent over-estimate of the effect of a demographically based tax reform. The over-estimate could be even larger in the case of a regional tax reform where a few states changed their income tax rates if one did not take account of the fraction of the reference groups' members impacted by the tax reform.

One should not use a shortcut that applies estimates from an econometric model ignoring social interactions, however. An econometric model without social interactions would under-estimate by up to 40 percent the effect of a net wage change compared to a model that includes social interactions, even if the tax change could be structured so as to have no social interactions.

The point to take away from our discussion is that not only do social interactions need to be part of the econometric framework, but also that the size of the interactions effect will depend on the way in which the tax change may limit the size of the social interactions in labor supplied.

6. Conclusion

Our research uses the canonical model of labor supply that adds possible social interactions in hours worked. We test the prediction of the model that an increase in hours worked by the members of the reference group increases hours worked for the individual (endogenous social effect). The reference group is assumed to contain persons in close economic distance to each other. Our measure of economic distance uses factor analysis, which allows mapping neighborhood variables into a two-dimensional social space. Our identification strategy builds on the likelihood that some persons belong to more than one reference group so that their hours worked may be used to instrument for endogenous labor supply of individuals in the worker's reference group.

Our regression model of labor supply social interactions uses the PSID data from

1975–1976. If social interactions in labor supply are treated as exogenous there is no estimated effect of the reference group behavior on the individual worker's behavior. When we instrument mean hours worked of the reference group we find a social interactions effect that is significant both statistically and economically. Model estimates including past hours worked also is important to identifying social interactions because the interdependence seems to operate with a time lag. In our most preferred model the estimated total wage elasticity of labor is 0.22, where about one-third is due to the exogenous wage change and two-thirds is due to social interactions effects.

The policy implications are that if one is to understand fully the labor supply and welfare effects of income taxes, which may be conditioned on demographic and location information, a model including social interactions is preferred. Equally important is proper interpretation of the social interactions model results. We demonstrate how a mis-specified model or a properly specified model that is mis-interpreted can easily lead to mis-estimates of the labor supply effects of tax reform by ±50–60 percent.

Endnotes

1. The baseline level of social disutility s_0 is exogenous, and we begin by assuming that it is constant for all individuals across all groups. The homogeneity assumption is important because if s_0 varies either across individuals due to heterogeneity or across the groups due to reference-group specific characteristics, then it is impossible to disentangle the effect of importance of social utility $b(\bullet)$ from the effect of different values for autonomous social utility s_0.
2. We could also assume $s'' < 0$ and the overall result in the paper would not change. The assumption avoids imposing additional conditions guaranteeing a unique global maximum for the total utility function, $V(\bullet)$.
3. Derivations are similar to Grodner (2003) and are omitted.
4. υ = [NLI + (τ – (TT/(TI – NLI)) × (TI – NLI))], where NLI is non-labor income, TT are total taxes, and TI is taxable income (Ziliak and Kniesner 1999).
5. Additionally two observations were dropped because the head's education was missing.
6. The difference between the number of observations used by MaCurdy et al. (1990) and our study comes from the fact that we did not exclude persons who were self-employed and farmers in 1975 but not in 1976 (individuals who changed their employment status). In the end, due to restricting the sample to individuals who also reported hours worked for year 1974 we have a final sample of 910 individuals.

7. See Ziliak and Kniesner (1999) for elaboration.

8. To elaborate,, the first factor loads primarily on demographic information and explains about 75 percent of the total variation in the variables and the second (rotated) factor loads primarily on location measures and explains about 15 percent of the information.

9. The intuition behind the result is that as the size of the group used to produce the average grows it approaches a similar value for everyone and increasingly collinear with the regression constant term.

10. Results from a Moran I test (Anselin 2001, p. 323) confirm the presence of social social interaction in hours worked and that the radius we adopt to define the reference group is reasonable.

11. In the linear form for the labor supply the coefficient on the hours worked for the reference group needs to be less than 1. Otherwise, one hour worked increase in the mean hours worked for the reference group would induce workers to increase their labor supply by more than one hour, which in turn would increase the hours worked for other workers in the individual's reference group even further. The labor market equilibrium would be unstable and a small positive shock to hours worked for any individual in the reference group would cause an exploding domino effect resulting in all workers choosing maximum feasible hours in the limit.

12. A feedback effect means that individuals working more hours at the same increase the level of hours in their reference group, which in turn affects them again

directly. The circular process, just like in the macroeconomic multiplier, continues until the effect fades out.

13. The result stems from symmetric boundaries around each member. We thank Dan Black for that observation.

14. Even though the instrument has the desired effect on the social interactions effect by increasing the coefficient of ANNHSRG_0_2, the correlation between past hours and hours in the reference group cannot be ignored.

15. The so-called global social multiplier represents the full effect of social interactions computed on the highest level of aggregation.

16. It is called a partial effect in Aronsson et al. (1999).

17. The elasticity calculations we discuss do not involve the coefficient of lagged labor supply; they treat the lagged dependent variable as simply a control for heterogeneity and are best viewed as short-run elasticities. Long-run steady-state elasticities that use the coefficient of lagged labor supply as an estimated partial adjustment effect would be about 2.5 times the ones we discuss.

15. Variables y_{3g_2} and x_{3g_2} appear in the reduced form equation because y_{3g_2} and x_{3g_2} are correlated with $y_{2g_1g_2}$, which is missing.

Table 1. Selection of the reference group using simple regression

	(1) AnnualHours	(2) AnnualHours	(3) AnnualHours	(4) AnnualHours	(5) AnnualHours	(6) AnnualHours
AfterTaxWage	**52.5361**	**68.4734**	**71.6107**	**71.8010**	**68.8683**	**68.9533**
	(36.3663)	(35.9633)*	(35.4610)**	(35.8352)**	(35.6676)*	(35.4563)*
VirtualInc	−**0.0034**	−**0.0031**	−**0.0040**	−**0.0047**	−**0.0051**	−**0.0055**
	(0.0059)	(0.0058)	(0.0058)	(0.0057)	(0.0057)	(0.0057)
KIDSU6	34.7968	8.9791	19.0675	29.5884	41.8152	48.5633
	(30.6237)	(29.3458)	(29.8259)	(29.7923)	(29.7989)	(29.9947)
FAMSIZ	−19.9075	−23.4939	−20.2682	−18.8505	−11.9343	−5.9835
	(14.7150)	(13.9574)*	(14.3347)	(14.0465)	(14.1303)	(14.4647)
AGE45	−24.1466	−23.3877	−25.6232	−27.5392	−30.9817	−34.0938
	(6.6709)***	(6.4811)***	(6.4667)***	(6.6709)***	(6.7804)***	(7.0095)***
HOUSEQ	0.0010	0.0007	0.0013	0.0016	0.0018	0.0019
	(0.0014)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
BHLTH	−141.4119	−122.3102	−135.2098	−157.1920	−181.7707	−185.7980
	(88.9913)	(82.6035)	(80.4038)*	(81.5759)*	(81.0654)**	(80.2034)**
AnnHSRG_0_1	−0.1978					
	(0.0867)**					
AnnHSRG_0_2		0.0626				
		(0.1432)				
AnnHSRG_0_3			−0.2042			
			(0.2411)			
AnnHSRG_0_4				−0.4982		
				(0.3231)		
AnnHSRG_0_5					−0.9685	
					(0.3566)***	
AnnHSRG_0_6						−1.5021
						(0.4709)***
Constant	2,540.7951	1,915.0347	2,472.8496	3,115.6458	4,144.2077	5,303.9724
	(259.6370)***	(361.2642)***	(557.6828)***	(707.8054)***	(783.2533)***	(1,007.0001)***
Observations	879	910	918	922	922	922
Average obs in ref group	13.33	44.53	89.19	142.56	204.13	271.21
Instruments	WageRate75, NLIncome75	WageRate75, NLIncome75	WageRate75, NLIncome75	WageRate75, NLIncome75	WageRate75, NLIncome75	WageRate75, NLIncome75

Standard errors in parentheses
Endogenous variables in bold
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 2. Regressions with Social Interactions, Habit Formation, and Various Sets of Instruments

Dependent Var: Annual Hours Worked	**(1) Baseline**	**(2) Full**	**(3) Only habit formation**	**(4) Only social interactions**
AfterTaxWage	**66.6982**	**38.5373**	**30.5734**	**81.6429**
	(35.5604)*	(28.6798)	(28.1246)	(37.3766)**
VirtualInc	−**0.0031**	**0.0000**	**0.0011**	−**0.0055**
	(0.0058)	(0.0047)	(0.0045)	(0.0061)
IndVRG_0_2	−318.8201	−317.4740	−284.0008	−385.0609
	(381.9788)	(307.9343)	(302.0874)	(401.1535)
KIDSU6	−135.0654	−155.5360	−116.3623	−214.5065
	(178.1115)	(144.4119)	(140.8569)	(188.1673)
FAMSIZ	−61.2033	−57.2676	−44.1304	−87.9325
	(48.6610)	(39.5923)	(38.4820)	(51.5978)*
AGE45	10.0278	24.6100	17.7155	23.8719
	(40.9405)	(33.0979)	(32.3824)	(43.1129)
HOUSEQ	0.0032	0.0025	0.0026	0.0030
	(0.0031)	(0.0025)	(0.0025)	(0.0033)
BHLTH	−28.8600	119.5895	60.9646	90.2927
	(144.5166)	(119.0663)	(114.3946)	(155.0286)
AnnualHours75		0.5950	0.5940	
		(0.0270)***	(0.0265)***	
AnnHSRG_0_2		**0.6379**		**1.3128**
		(0.2689)**		(0.3532)***
Constant	2,173.2953	−484.0063	903.5016	−671.7148
	(202.1274)***	(607.4100)	(169.7451)***	(792.5331)
Observations	910	910	910	910
Sargan test		0.212		0.081
P-value		0.64502		0.77653
Instruments	WageRate75	WageRate75	WageRate75	WageRate75
	NLIncome75	NLIncome75	NLIncome75	NLIncome75
		AnnHSORG_2_6		AnnHSORG_2_6
		IndVORG_2_6		IndVORG_2_6

Standard errors in parentheses
Endogenous variables in bold
* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix. Econometric Issues of Identification

To make correct inferences the labor supply model in (3) must deal with special identification issues involving social interactions. Econometric problems involving interdependence of individuals were first discussed in Manski (1993); more recent discussions include Manski (2000, 2002), Moffitt (2001), Brock and Durlauf (2001), and Durlauf (2004). For the purpose of illustrating econometric concerns, again consider the general regression model:

$$y_{ig} = \theta_0 + \theta_1 x_{ig} + \theta_2 f\left(y_{(-i)g}\right) + \theta_3 h\left(x_{(-i)g}\right) + \theta_4 u_g + \varepsilon_{ig},$$

Where here y_{ig} is the hours worked for individual i who is a member of the reference group g, x_{ig} is a vector of exogenous characteristics of the individual, $(\bullet)_{(-i)g}$ denotes all other individuals in the neighborhood besides the individual i, $f(\bullet)$ and $h(\bullet)$ denote some functions of outcomes or characteristics of other individuals in the neighborhood that affect individual i, u_g denotes neighborhood specific effects, ε_{ig} is a white-noise disturbance, and θ_k denotes coefficients.

Simultaneity problem (*endogenous social interactions*, *parameter* θ_2): comes from the correlation of actions made by individuals belonging to the same reference group, meaning that the choices of persons are interdependent and individuals' error terms in the same reference group should have a nonzero covariance.

Endogenous group membership problem (*exogenous social interaction, parameter* θ_3): comes from the possibility that people may select neighborhoods (reference groups) on the basis of their individual characteristics. In our model of labor supply people who work more hours may choose communities/membership groups that consist of people who also work more or share the same characteristics.

Correlated unobservables problem, parameter θ_4): comes from the likelihood that there are some reference group-specific characteristics that vary across reference groups. The most natural way to deal with this problem is to include reference group-specific fixed effects in a panel data model.

To clarify our identification strategy we draw on the elegant discussion of the linear interdependence model presented by Moffitt (2001). For so-called expositional purposes we assume that there are no correlated unobservables in the model (although it does not change the main result), there are two reference groups in the population indexed by g_1 and g_2, and that each group consists of only two individuals (the entire population consists of three individuals). The difference from the discussion in Moffitt (2001) is that there is one individual, 2, who is a member of both reference groups 1 and 2. Thus, assuming that individuals respond to the mean of outcomes for other individuals in their reference group we have the following structural form of the model:

Group g_1

$$\text{individual 1 belongs only to group } g_1: \quad y_{1g_1} = \theta_0 + \theta_1 x_{1g_1} + \theta_2 y_{2g_1g_2} + \theta_3 x_{2g_1g_2} + \varepsilon_{1g_1}$$

$$\text{individual 2 belongs to both groups } g_1 \text{ and } g_2: \quad y_{2g_1g_2} = \theta_0 + \theta_1 x_{2g_1g_2} + \theta_2 \frac{y_{1g_1} + y_{3g_2}}{2} + \theta_3 \frac{x_{1g_1} + x_{3g_2}}{2} + \varepsilon_{2g_1g_2},$$

Group g_2

$$\text{individual 2 belongs to both groups } g_1 \text{ and } g_2: \quad y_{2g_1g_2} = \theta_0 + \theta_1 x_{2g_1g_2} + \theta_2 \frac{y_{1g_1} + y_{3g_2}}{2} + \theta_3 \frac{x_{1g_1} + x_{3g_2}}{2} + \varepsilon_{2g_1g_2}$$

$$\text{individual 3 belongs only to group } g_2: \quad y_{3g_2} = \theta_0 + \theta_1 x_{3g_2} + \theta_2 y_{2g_1g_2} + \theta_3 x_{2g_1g_2} + \varepsilon_{3g_2},$$

We also assume that the correlation of the error terms within the reference groups is non-zero but correlation of the error terms across the reference groups (for individuals who do not belong to another reference group) is zero. Specifically, $\operatorname{cov}\left(\varepsilon_{1g_1}, \varepsilon_{2g_1g_2}\right) \neq 0$, $\operatorname{cov}\left(\varepsilon_{2g_1g_2}, \varepsilon_{3g_2}\right) \neq 0$, but $\operatorname{cov}\left(\varepsilon_{1g_1}, \varepsilon_{3g_2}\right) = 0$. The last is a critical identifying assumption validating the instrument we propose in the following discussion.

The reduced form for individual 1 in group 1 is the following (where the same can be done for individual 3 in group 2):

$$\begin{aligned} y_{1g_1} = &\underbrace{\frac{2\left(\theta_0+\theta_0\theta_2\right)}{2-\theta_2^2}}_{\alpha} + \underbrace{\frac{2\theta_1+\theta_2\theta_3}{2-\theta_2^2}}_{\beta} x_{1g_1} + \underbrace{\frac{\left(\theta_1\theta_2+\theta_3\right)}{2-\theta_2^2}}_{\gamma} x_{2g_1g_2} + \underbrace{\frac{\left(\theta_2\varepsilon_{2g_1g_2}+\varepsilon_{1g_1}\right)}{2-\theta_2^2}}_{\nu_{1g_1}} \\ &+ \frac{\theta_2^2}{2-\theta_2^2} y_{3g_2} + \frac{\theta_2\theta_3}{2-\theta_2^2} x_{3g_2} \end{aligned} \tag{A1-1}$$

Notice that the above reduced form is not for the entire system but only for the reference group 1. Due to separability of reference groups, an endogenous variable y_{3g_2} (in the full simultaneous system of equations) can be treated as exogenous in the reference group 1, together with x_{3g_2}, because both are neither correlated with the error term ν_{1g_1} nor correlated with independent variables x_{1g_1} and $x_{2g_1g_2}$.[18] They can be omitted from the estimation of the reduced form (A1-1) (they can be viewed like an independent measurement error in the dependent variable and we can still obtain consistent estimates for [α, β, γ].

Without the loss of generality we can, for the moment, focus on the reduced form:

$$y_{1g_1} = \alpha + \beta x_{1g_1} + \gamma x_{2g_1g_2} + \nu_{1g_1}. \tag{A1-2}$$

Moffitt (2001) discusses the reduced form equation in (A1-2); he points out that the system is not identified because we have four parameters to recover ($\theta_0, \theta_1, \theta_2, \theta_3$), but can obtain only three estimates (α, β, γ). Both Manski and Moffitt show that the estimate of γ indicates whether there is any type of interactions because $\theta_2 = \theta_3 = 0$ implies $\gamma = 0$. Therefore, if one specifies the correct model the reduced form, (A1-2), can give information on the presence social effects.

To identify θ_2 and θ_3 separately we need to use the assumption that there are overlapping reference groups in the population and that for some individuals the reduced form equation looks like (A1-1). Then variables y_{3g_2} and x_{3g_2} work as exclusion restrictions for the individual 1 relative to the equation on individual 2. Variables y_{3g_2} and x_{3g_2} can both be additional variables in the reduced form equation or they can both be instruments for the endogenous variables $y_{2g_1g_2}$ in the structural equation for individual 1. We use an instrumental variables technique to estimate the structural equation for individual 1 to obtain a direct estimate for θ_2. A strategy similar to the one described above has been used by Case and Katz (1992), who instrument for the endogenous effect using the average levels of adjacent neighbors' characteristics that are supposedly exogenous. Similarly, Evans et al. (1992) instrument school composition with city-wide variables for the unemployment rate.

Our identification strategy is illustrated schematically in Graph 1. We present the hypothetical two-dimensional social coordinate space with two reference groups: g_1 and g_2. Suppose now that individual $y^0_{g_1}$ belongs to the reference group g_1 and is affected by the outcomes of the members of the reference group, represented by the observations labeled as $y^1_{g_1}$ and $y^2_{g_1g_2}$ (empty and gray-filled circles). If we use the mean of all $y^1_{g_1}$ and $y^2_{g_1g_2}$ observations (referred further as $\overline{y}^{(-0)}_{g_1}$) as an independent variable in the regression

(5) to identify endogenous social interaction in $y^0_{g_1}$ the coefficient will be biased due to the fact that observations $y^1_{g_1}$ and $y^2_{g_1g_2}$ are also affected by the outcome $y^0_{g_1}$, which causes endogeneity in the $\overline{y}^{(-0)}_{g_1}$. However, if there are observations in the reference group g_1 that also belong to the neighboring reference group g_2, then part of $\overline{y}^{(-0)}_{g_1}$ attributed to the outcomes $y^2_{g_1g_2}$ can be instrumented by the outcomes of the members of the reference group g_2, denoted by $y^3_{g_2}$. We can use instrumental variables (IV) estimation because $y^3_{g_2}$ are correlated with all $y^2_{g_1g_2}$ observations because they belong to the same reference group, but $y^3_{g_2}$ are not correlated with the error term associated with either $y^0_{g_1}$ or $y^1_{g_1}$ observations because they do not belong to the same reference group. Observations $y^3_{g_2}$ are transitorily correlated with the outcomes $y^0_{g_1}$ and $y^1_{g_1}$ only through the deterministic part of observations $y^2_{g_1g_2}$.

Notice that in practice if we instrument observations $y^2_{g_1g_2}$ with outcomes $y^3_{g_2}$ there may still be observations $y^1_{g_1}$ that are not instrumented and thus will make a part of the $\overline{y}^{(-0)}_{g_1}$ endogenous, which is the case presented in Graph 1. Instead of using just one reference group we can imagine using a full set of observations in the adjacent reference groups that form the ring around the particular reference group (represented by the dotted circle on the graph).

Graph 1. Demonstration of the identification strategy for the endogenous social interactions.

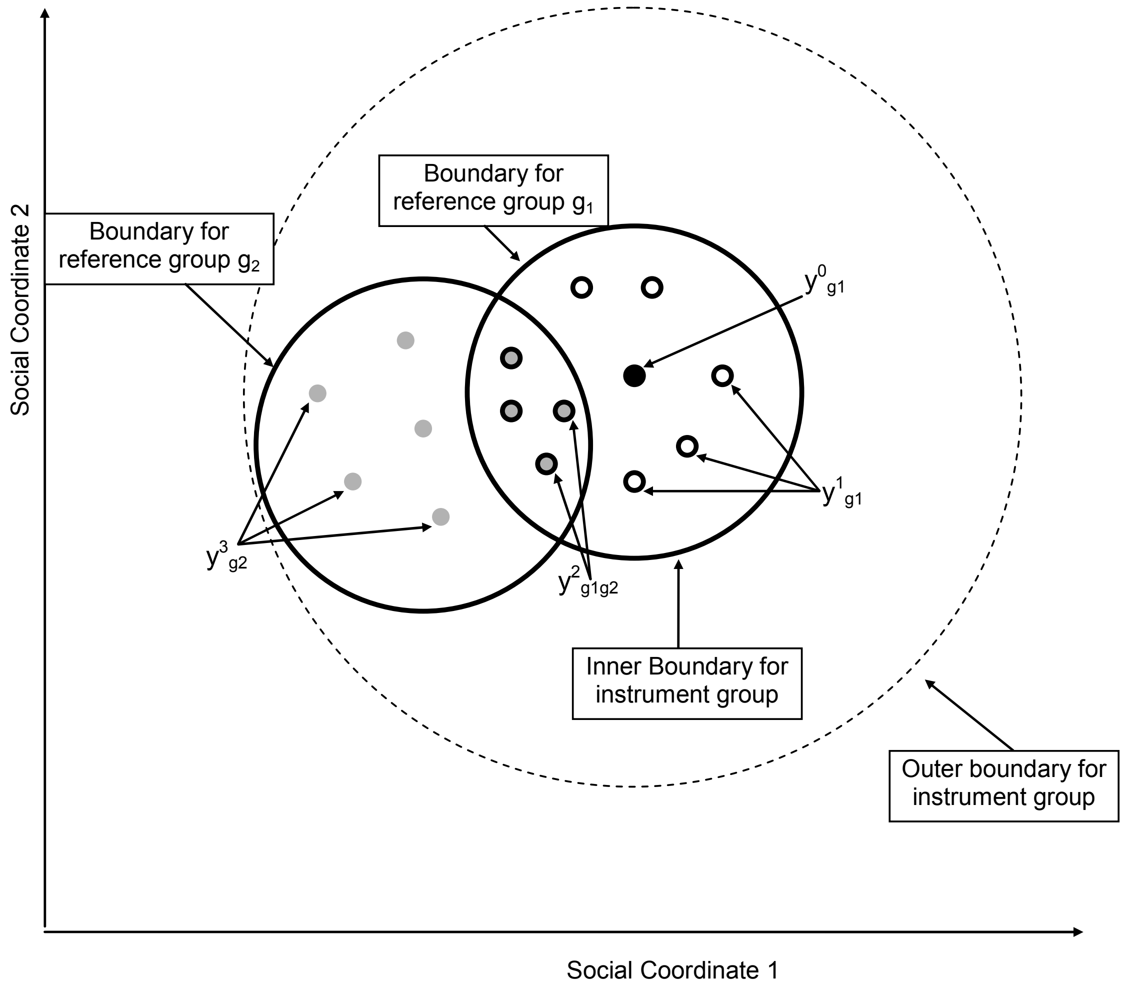

Table A1. Descriptive Statistics

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
AnnualHours	910	2236.864000	536.701100	288.000000	4917.000000
AnnualHour75	910	2247.385000	540.086500	320.000000	4500.000000
AfterTaxWage	910	4.692693	1.198573	0.542700	7.488000
WageRate	910	6.272303	1.794132	0.670000	9.900000
WageRate75	910	5.479915	0.636453	3.655642	6.837162
VirtualInc	910	5138.557000	4364.210000	−965.000000	45593.000000
NLIncome	910	3710.268000	4700.172000	−7900.000000	57640.000000
NLIncome75	910	3298.155000	3984.506000	−10000.000000	26000.000000
AnnHSRG_0_2	910	2210.108000	134.322300	1180.000000	2950.667000
AnnHSORG_2_6	910	2214.600000	53.725650	2009.458000	2477.579000
IndVORG_2_6	910	301892.000000	5444115.000000	−1.534880	1.369547
IndVRG_0_2	910	352303.000000	0.035230	−1.760330	1.598270
KIDSU6	910	0.445055	0.696331	0.000000	3.000000
FAMSIZ	910	3.873626	3.873626	2.000000	9.000000
AGE45	910	1.748352	3.108485	0.000000	11.000000
HOUSEQ	910	18511.900000	16930.990000	−5000.000000	120000.000000
BHLTH	910	.051648	.221438	0.000000	1.000000

References

Abel, Andrew B. 2005. "Optimal Taxation When Consumers Have Endogenous Benchmark Levels of Consumption." *Review of Economic Studies* 72(1) (January): 21–42.

Abu-Ismail, F. F. 1992. "Reference Group Influences in the Uncertain Decision Making Process: Empirical Findings on Buying of Personal Investment Products." *Middle East Business and Economic Review* 4(1) (January): 41–48.

Andreoni, James, and John Karl Scholz. 1998. "An Econometric Analysis of Charitable Giving with Interdependent Preferences." *Economic Inquiry* 36(3): 410–428.

Anselin, Luc. 2001. "Spatial Econometrics." In *A Companion to Theoretical Econometrics*. Badi H. Baltagi (ed.), Malden, MA, Blackwell Publishers, Inc., pp. 310–330.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*. 58(2) (April): 277–297.

Aronsson, Thomas, Sören Blomquist, and Hans Sacklén. 1999. "Identifying Intertemporal Behaviour in An Empirical Model of Labour Supply." *Journal of Applied Econometrics* 14(6): 607–626.

Baker, Frank, Gregory M. O'Brien, and Alan Sheldon. 1968. "Reference Groups of Occupational Therapists: Role Orientation in A Changing Mental Hospital." *American Journal of Occupational Therapy* 22(3): 197–202.

Bartolome de, Charles A. M. 1990. "Equilibrium and Inefficiency in a Community Model with Peer Group Effects." *Journal of Political Economy* 98(1): 110–133.

Becker, Gary S. 1974. "A Theory of Social Interactions." *Journal of Political Economy* 82(6) (November-December): 1063–1093.

Becker, Gary S., and Kevin M. Murphy. 2000. *Social Economics: Market Behavior in a Social Environment*, Cambridge, Massachusetts, Harvard University Press.

Benabou, Roland. 1996. "Equity and Efficiency in Human Capital Investment: the Local Connection." *Review of Economic Studies* 63(2) (April): 237–264.

Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy* 102(5) (October): 841–877.

Blundell, Richard, and Thomas MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics*. Volume 3A, Ashenfelter, Orley; Card, David (eds.), Amsterdam, Elsevier Science, North-Holland, pp. 1559–1695.

Brock, William A., and Steven N. Durlauf. 2001. "Interactions-based Models, (SSRI)." In *Handbook of Econometrics,* Volume 5*,* James J. Heckman and Edward Leamer (eds.), Amsterdam, Elsevier Science, North-Holland, pp. 3297–3380.

Brock, William A., and Steven N. Durlauf. 2002. "A Multinomial-Choice Model of Neighborhood Effects." *American Economic Review* 92(2) (May): 298–303.

Card, David. 1994. "Intertemporal Labour Supply: An Assessment (Chapter 2)." In *Advances in Econometrics Sixth World Congress: Vol. II,* Christopher A. Sims (ed.), New York, Cambridge University Press: pp. 49–78.

Case, Anne C., and Lawrence F. Katz. 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." National Bureau of Economic Research (NBER) Working Paper, No. 3705, (May).

Childers, Terry L., and Akshay R. Rao. 1992. "The Influence of Familial and Peer-Based

Reference Groups on Consumer Decisions." *Journal of Consumer Research* 19(2) (September): 198–211.

Clark, Robert E. 1972. *Reference Group Theory and Delinquency*, New York, Behavioral Publications.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92(1) (September): 1–45.

Crouter, Ann C., and Beth Manke. 1997. "Development of A Typology of Dual-Earner Families: A Window Into Differences Between and Within Families in Relationships, Roles, and Activities." *Journal of Family Psychology* 11(1) (March): 62–75.

Dabelko, David D. 1976. "Reference Group Theory, Social Comparison Theory, and the Study of Politics." *Journal of Social Psychology* 99(2) (August): 283–288.

Deaton, Angus. 2001. "Inequalities in Income and Inequalities in Health." In *The Causes and Consequences of Increasing Inequality*, Finis Welch, (ed.). Chicago, University of Chicago Press, pp. 285–313.

Dornstein, Miriam. 1988. "Wage Reference Groups and Their Determinants: A Study of Blue-Collar and White-Collar Employees in Israel." *Journal of Occupational Psychology* 61(3) (September): 221–235.

Durlauf, Steven N. 1996. "A Theory of Persistent Income Inequality." *Journal of Economic Growth* 1(1) (March): 75–93.

Durlauf, Steven N. 2004. "Neighborhood Effects." In *Handbook of Urban Economics Vol. 4*, JV Henderson and J-F Thisse (eds.). Elsevier Science, North-Holland: pp. 2173–2242.

Eibner, Christine E., and William N. Evans. 2005. "Relative Deprivation, Poor Health Habits and Mortality." *Journal of Human Resources* 40(3) (Summer): 591–620.

Eklöf, Matias, and Hans Sacklén. 2000. "The Hausman-MaCurdy Controversy: Why Do the Results Differ across Studies?" *Journal of Human Resources* 35(1) (Winter): 204–220.

Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. "Measuring Peer Group Effect: A Study of Teenage Behavior." *Journal of Political Economy* 100(5) (October): 966–991.

Frank Robert H. 1985. "The Demand for Unobservable and Other Nonpositional Goods." *The American Economic Review* 75(1) (March): 101–116.

Glaeser, Edward L., David Laibson, and Bruce Sacerdote. 2002. "An Economic Approach to Social Capital." *Economic Journal* 112(483) (November): F437–F458.

Glaeser, Edward L, Bruce Sacerdote, and Jose A. Scheinkman. 1996. "Crime and Social Interactions." *Quarterly Journal of Economics*. 111(2) (May): 507–548.

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman. 2003. "The Social Multiplier." *Journal of the European Economic Association*. 1(2-3) (April-May): 345–353.

Graham, Bryan S., and Jinyong Hahn. 2005. "Identification and Estimation of the Linear-in-Means Model of Social Interactions." *Economics Letters* 88(1) (July) 1–6.

Grodner, Andrew. 2003. "A Partial Equilibrium Analysis in the Markets with Social Interactions." Unpublished manuscript. Syracuse University.

Hamermesh, Daniel S. 1977. "Economic Aspects of Job Satisfaction." In *Essays in Labor Market and Population Analysis,* Ashenfelter Orley C. and Oates Wallace E. (ed.).

New York, Wiley, pp. 53–72.

Hamermesh, Daniel S. 2001. "The Changing Distribution of Job Satisfaction." *Journal of Human Resources* 36(1) (Winter): 1–30.

Hamermesh, Daniel S. and Joel Slemrod. 2005. "The Economics of Workaholism: We Should Not Have Worked on This Paper." National Bureau of Economic Research (NBER) Working Paper, No. 11566, (August).

Hausman, Jerry A. 1981. "Labor Supply." *In How Taxes Affect Economic Behavior*, Henry J. Aaron and Joseph A. Pechman (ed.), Washington, D.C.: The Brookings Institution, pp. 27–83.

Jacques, Jeffrey M., and Karen J Chason. 1978. "Cohabitation: A Test of Reference Group Theory among Black and White College Students." *Journal of Comparative Family Studies* 9(2) (Summer): 147–165.

Johnson, Richard A., and Wichern Dean W. 1992. *Applied Multivariate Statistical Analysis*, Englewood Cliffs, New Jersey, Prentice Hall.

Kapteyn, Arie, Sara van de Geer, Huib van de Stadt, and Tom Wansbeek. 1997. "Interdependent Preferences: An Econometric Analysis." *Journal of Applied Econometrics*. 12(6) (November-December): 665–686.

Kelejian, Harry H., and Ingmar R. Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *Journal of Real Estate Finance and Economics* 17(1) (July): 99–121.

Kelejian, Harry H., and Ingmar R. Prucha. 2002. "2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights." *Regional Science and Urban*

Economics 32(6) (November): 691–707.

Kelley, Harold H. 1968. “Two Functions of Reference Groups.” In *Readings in Reference Group Theory and Research*, Herbert H. Hyman and Eleanor Singer (eds.), New York: Free Press; Collier-Macmillan Limited: London, pp. 77–83.

Kimmel, Jean, and Thomas J. Kniesner. 1998. “New Evidence on Labor Supply: Employment versus Hours Elasticities By and Marital Status.” *Journal of Monetary Economics* 42(2) (July): 289–301.

Kooreman, Peter and Lambert Schoonbeek. 2004. “Characterizing Pareto Improvements in an Interdependent Demand System.” *Journal of Public Economic Theory* 6(3) (August): 427–443.

Kooreman, Peter, and Adriaan Soetevent. 2002. “A Discrete Choice Model with Social Interactions; an Analysis of High School Teen Behavior.” (February): University of Groningen.

Krauth, Brian. forthcoming. “Simulation-Based Estimation of Peer Effects.” *Journal of Econometrics*.

Lindbeck, Assar, Sten Nyberg, and Jorgen W. Weibull. 1999. “Social Norms and Economic Incentives in the Welfare State.” *Quarterly Journal of Economics* 64(1) (February): 1–35.

MaCurdy, Thomas, David Green, Harry J. Paarsch. 1990. “Assessing Empirical Approaches for Analyzing Taxes and Labor Supply.” *Journal of Human Resources* 25(3) (Summer): 415–490.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects.” *Review of Economic Studies* 60(204) (July):.531–542.

Manski, Charles F. 2000. “Economic Analysis of Social Interactions.” *Journal of Economic Perspectives* 14(3) (Summer): 115–136.

Moffitt, Robert A. 2001. “Policy Interventions, Low-Level Equilibria and Social Interactions.” In *Social Dynamics,* S. Durlauf and H.P. Young (eds.), Washington, D.C.: MIT and Brookings Institution Press, pp. 45–82.

Menno Pradhan, Asep Suryahadi, Sudarno Sumarto, and Lant Pritchett. 2001. “Eating Like Which Joneses? An Iterative Solution to the Choice of a Poverty Line Reference Group.” *Review of Income and Wealth*. 47(4) (December): 473–487.

Rettig, Salomon, Frank N. Jacobson, and Benjamin Pasamanick. 1959. “Subjective Status and the Nature of the Reference Group.” *Journal of Social Psychology* 50: 233–240.

Rigobon, Roberto. 2001. “Contagion: How to Measure It?” National Bureau of Economic Research (NBER) Working Paper, No. 8118. (February): pp. 1–78.

Roback, Jennifer. 1982. “Wages, Rents, and the Quality of Life.” *Journal of Political Economy* 90(6) (December): 1257–1278.

Schram, Arthur, and Joep Sonnemans. 1996. “Why People Vote: Experimental Evidence.” *Journal of Economic Psychology*. 17(4) (August): 417–442.

Schaffner, Julie Anderson. 1995. “Attached Farm Labor, Limited Horizons and Servility.” *Journal of Development Economics* 47(2) (August): 241–270.

Sen, Sankar, Zeynep Gurhan-Canli, and Vicki. Morwitz. 2001. “Withholding Consumption: A Social Dilemma Perspective on Consumer Boycotts.” *Journal of Consumer Research* 28(3) (December): 399–417.

Sherif, Muzafer, and Carolyn W. Sherif. 1969. *Social Psychology*, New York, Evanston, and London, Harper & Row, Publishers.

Stark, Oded, and Edward J. Taylor. 1991. "Migration Incentives, Migration Types: The Role of Relative Deprivation." *Economic Journal*. 101(408) (September): 1163–1178.

Sun, Yixiao. 2005. "Estimation and Inference in Panel Structure Models." Department of Economics, University of California, San Diego, Working Paper (August).

Triest, Robert K. 1990. "The Effect of Income Taxation on Labor Supply in the United States." *Journal of Human Resources* 25(3) (Summer): 491–516.

Woittiez, Isolde, and Arie Kapteyn. 1998. "Social Interactions and Habit Formation in a Model of Female Labour Supply." *Journal of Public Economics* 70(2) (November): 185–205.

Ziliak James P., and Thomas J. Kniesner. 1999. "Estimating Life Cycle Labor Supply Tax Effects." *Journal of Political Economy* 107(2) (April): 326–359.