

# SOCIAL INFLUENCE AND CONSUMPTION: EVIDENCE FROM THE AUTOMOBILE PURCHASES OF NEIGHBORS

Mark Grinblatt, Matti Keloharju, and Seppo Ikäheimo\*

*Abstract*—This study analyzes the automobile purchase behavior of all residents of two Finnish provinces over several years. Using a comprehensive data set with location coordinates at the individual consumer level, it finds that the purchases of neighbors, particularly in the recent past and by those who are geographically most proximate, influence a consumer's purchases of automobiles. There is little evidence that emotional biases, like envy, account for the observed social influence on consumption.

## I. Introduction

**S**OCIAL influence is now recognized as a critical factor in the study of education outcomes, income, unemployment, relationships, and crime, but little empirical research exists on how social connections influence consumption.<sup>1</sup> Mainstream economists have long debated the role that social influence should play in the theory of consumption.<sup>2</sup> We hope to advance our understanding of this topic with a unique empirical study.

This paper investigates whether social influence exists in the consumption of a particularly important commodity—automobiles—and what might drive this influence. Using a comprehensive panel, it finds that neighbors' purchases significantly increase a consumer's decision to purchase an automobile. The influence is dominated by the very nearest neighbors, and it lasts for a remarkably short period of time.

Received for publication January 17, 2006. Revision accepted for publication July 20, 2007.

\* The Anderson School at UCLA and NBER; Helsinki School of Economics and CEPR; and Helsinki School of Economics.

We would like to thank the Finnish Vehicle Administration and the Finnish Tax Administration for providing access to the data and the Office of the Data Protection Ombudsman for recognizing the value of this project to the research community. Our appreciation also extends to Juhani Linnainmaa, Antti Lehtinen, and Juan Prajogo, who provided superb research assistance, and to Ivo Welch, who generated many insights that benefited this paper. Seminar participants at the University of California at Berkeley, Columbia University, the University of California at Los Angeles, the University of Southern California, and the Stockholm School of Economics, as well as Shlomo Benartzi, David Hirshleifer, Markku Kaustia, Juhani Linnainmaa, Avanihar Subrahmanyam, Mark Weinstein, and an anonymous referee, offered helpful comments on earlier drafts. Financial support from the Academy of Finland, the Foundation for Economic Education, and the Paulo Foundation is gratefully acknowledged.

<sup>1</sup> Liebman, Katz, and Kling (2004), Bayer, Ross, and Topa (2005), and Topa (2001) analyze social effects on income and labor. Sacerdote (2001) studies peer effects on educational outcomes. Marmaros and Sacerdote (2006) analyze how friendships form. Kling, Ludwig, and Katz (2005) and Glaeser, Sacerdote, and Scheinkman (1996) study social influence on crime. Goolsbee and Klenow (2002) is one of the few papers to study local spillovers in consumption.

<sup>2</sup> See Veblen (1898), Friedman (1957), Morgenstern (1948), and Stigler (1950) for early thoughts on the issue. Leibenstein (1950), Duesenberry (1949/1962), Becker (1974), Pollack (1976), Robson (1992), Abel (1990), Galí (1994), Bagwell and Bernheim (1996), Campbell and Cochrane (1999), Chan and Kogan (2002), and Hopkins and Kornienko (2004) develop models of interpersonal consumption effects. Bikhchandani, Hirshleifer, and Welch (1992), Bernheim (1994), and Pesendorfer (1995) study conformity and fads.

The paucity of empirical research on social influence and consumption may be due to the complexity of designing proper studies with existing data. Two econometric problems are known to hamper inferences about group and neighborhood effects.<sup>3</sup> The first is an omitted variables bias. In the absence of a perfect set of controls, one cannot validate a social influence on consumption by observing that a group of neighbors purchase similar baskets of goods. Inferences will be biased whenever individuals endogenously locate in a neighborhood because of shared commonalities. Alternatively, they may acquire these commonalities from living in the neighborhood. The universe of such commonalities is large and unobservable.

The second factor complicating inferences about social influence is a simultaneous equations bias uncovered by Manski (1993). This bias is generated by a "reflection problem"—the simultaneous feedback from the subject to the social group. The problem cannot be resolved by reduced-form modeling because the structural parameters are under-identified.

The study presented here addresses these problems with methodologies that take advantage of the exceptional quality of our data. The data set, a comprehensive panel from two major provinces in Finland, includes the exact location of all neighbors in relation to one another. It also contains an extraordinary number of common attributes for each neighbor that might account for the findings. Apart from extensive sample size and controls, the data offer major advantages over commonly used neighborhood aggregations (such as ZIP code, census tract, and block) in the literature: Daily observations on both the subject and the historical state of each subject's unique neighborhood lend themselves to a novel experimental design that controls for attributes that we do not measure.

We primarily use logit estimation to study the marginal influence of the car purchase history of the ten nearest neighbors relative to that of more distant neighbors (the "nearest neighbors approach"). In this design, the car purchases of more distant neighbors are an instrument for omitted control variables that might generate spurious inferences about near neighbor influence. The quality of our data allows us to verify the appropriateness of this instrument for addressing omitted variables bias. We find that the difference between the measurable attributes of the ten nearest neighbors and those of the more distant neighbors are unrelated to the subject's car purchases. Moreover, any

<sup>3</sup> Many prior studies of geographic or neighborhood effects have been criticized for failing to recognize these econometric issues. See, for example, Evans, Oates, and Schwab (1992), Manski (1993), and Akerlof (1997).

differences in car purchases between the ten nearest neighbors and the more distant neighbors are unrelated to any of the subject's measured attributes (except his car purchase decision). We expect that these properties extend to the subject's unmeasured attributes as well. We also check how robust this methodology is by analyzing coefficient estimates with controls for between-subject fixed effects.

The experimental design and the quality of the data also constrain the reflection problem. First, we use neighbors' lagged actions in forming the critical regressor; lagged actions are not plausibly affected by a consumer's purchases. Second, there is a lack of commutativity to neighbor nearness with our data: often, one is not among the nearest neighbors of one's nearest neighbors. These factors mitigate the impact of feedback effects on structural coefficient estimates that employ the nearest neighbors specification.

Regressions on a comprehensive sample of consumers in the most populated provinces in Finland indicate that neighbors who purchase a car, particularly those who purchased recently and are nearest in distance, increase the propensity of a consumer to purchase a car. This observed social influence controls for the consumer's income, employment status, home ownership, age, marital status, dependents, commuting costs, and sex. Instrumenting with more distant neighbors' purchases also controls for other variables (observed and unobserved) that are common to a larger community. The influence is strongest within the lowest income deciles, particularly if the influencing neighbor is of the same income decile. The effect also is more pronounced for used cars; parroting of neighbor behavior is stronger still for purchases of any particular car make (like a Honda) or any particular car model (like a Honda Accord). Such evidence argues against envy and similar emotional factors as the primary source of the social influence on consumption. Such evidence also tends to contradict alternative hypotheses based on omitted attributes and experiences shared only by very near neighbors.

Our results are organized as follows: section II describes the data, the empirical methodology, and the conditions under which our methodology addresses pertinent economic issues in the social influence literature. Section III presents results, beginning with summary statistics, and then reports results from an extensive series of regressions. Section IV discusses how one might interpret the results. Section V concludes the paper.

## II. Data and Methodological Implementation

The core data set for this study, a panel of automobile purchases in Finland, is well suited for the analysis of social influence on consumption. First, automobiles represent highly visible consumption, and thus are more likely to be associated with a behavioral social influence, like envy.<sup>4</sup> Automobile consumption also is used as an example of

publicly visible consumption in literature outside of economics.<sup>5</sup> Goods that are privately consumed, like mattresses or medicines, do not offer the same opportunity for identifying a social influence on consumption that is driven by the good's visibility. Second, for many of the subjects in our study, automobiles are a luxury rather than a necessity, and luxury goods should have more social influence than other goods.<sup>6</sup> In contrast to the United States, most of the subjects studied have access to high-quality public transportation, and the tax rate on a typical automobile (nearly 50%) and its fuel (about 70%) makes its acquisition and use costly. Finally, Finland collects data on a remarkably large number of useful control variables. As we will see, the properties of these control variables alleviate concerns that our findings are due to omitted variables, such as those associated with endogenous location effects.

We analyze variables derived from the union of two data sets: one contains data on automobile ownership and purchases; the other contains data from income tax returns.

### A. Automobile Ownership and Purchase Data

The Finnish Vehicle Administration (FVA) provided data on automobile purchases and ownership. The data set records the type of personal automobile owned by each car-owning resident on June 10, 2002, for all purchases made prior to 2002, the exact date of purchase, and registration data that allow us to infer whether the purchase was of a used or new vehicle.<sup>7</sup> (Purchases of trucks, buses, and related commercial vehicles are excluded.) The data are comprehensive for residents in the provinces of Uusimaa and East Uusimaa, which contain Greater Helsinki and the most densely populated areas in Finland.

We primarily use the data set to analyze car purchase behavior from January 1, 1999, to December 31, 2001. Car purchases prior to January 1, 1999, are used to assess the degree to which purchases over the three-year interval are influenced by the five-year history of car purchases (or nonpurchases) of each subject's neighborhood.

### B. Tax Authority Data

We develop a set of regressors from Finnish tax return data, which we linked to the FVA data set on a person-by-person basis. The tax return data set records variables as of three end-of-years 1998 through 2000 inclusive. The data are on every resident in the two provinces, both the car

<sup>5</sup> See, for example, Bearden and Etzel (1982), Solomon (1999), and Peter and Olson (2001).

<sup>6</sup> For example, Basman, Molina, and Slottje (1988) found that durables (luxury goods) had the highest marginal rate of substitution elasticities of any commodity group.

<sup>7</sup> The FVA data set contains only the most recent purchase of a car. Few purchases in the January 1, 1999, through December 31, 2001, sample period that we focus on are missing because of the typically lengthy periods over which Finnish residents tend to own the same car. In the rare instance that a person has bought more than one car in a single year, we consider only the most recent purchase in that calendar year.

<sup>4</sup> See Duesenberry (1949/1962).

owners in the June 10, 2002, FVA data set, as well as residents who do not own cars on that date. Except for address, the 1998 data are assumed to represent the data for the subjects in 1999; the 1999 data portray these variables for 2000, and so on.

We collected the following variables for each tax subject in the two provinces: income, year of birth, sex, marital status (single, married, or unmarried but cohabiting), number of dependents under eighteen years old, work-related travel costs, whether the community lived in is city, suburban, or rural,<sup>8</sup> employment status, existence of residential real estate ownership, and address. The tax return data set also reports move-in and move-out dates for each subject at a given address in a given year. Therefore, addresses for each subject are current for any given day. These addresses were converted to latitude and longitude coordinates on all subjects. The coordinates were then translated and rotated with parameters that were destroyed to maintain the anonymity of the subjects in the data sets while preserving their relative distance from one another.<sup>9</sup>

This linking of the FVA and tax data sets generates data on all residents in the provinces, both car purchasers, potential car purchasers, and their neighbors over the eight-year period, 1994–2001, with control variables over the three-year period, 1999–2001.

### C. Data Frequency and Data Requirements

We analyze data at the yearly frequency. Because the purchase history of each neighborhood of a subject changes from day to day, we employ a methodology that alleviates concerns about the coefficient biases that seasonalities in intrayear car purchases might induce. For each of the three years studied, the actual purchase dates are used to generate a distribution of nonpurchase dates. For example, if over the entire year of 1999, there are 20 times more nonpurchasers than purchasers of a car, and if there were 200 purchases on July 12, 1999, and 50 on September 3, 1999, then we assume that there were 4,000 nonpurchases on July 12, 1999, and 1,000 nonpurchases on September 3, 1999. Doing this for every date in 1999 generates a probability distribution function of nonpurchases over 365 days that is identical to the probability distribution function for purchases. If an individual has not purchased a car in that year, his “shadow nonpurchase date” within that year is randomly assigned using this distribution. That is, the probability of someone who did not buy an automobile in all of 1999 being assigned

to July 12 as his shadow nonpurchase date is four times larger than being assigned to September 3. Nonpurchasers thus appear once in their year of nonpurchase, on their “shadow nonpurchase date,” rather than 365 times in that year. Consumers who purchased in 1999 also appear once, on their actual 1999 date of purchase. (As a consequence of this procedure, coefficients must be interpreted as annualized rates of purchase rather than daily rates of purchase.) For both a purchase and a nonpurchase observation, we capture the state of the consumer’s neighborhood, including its history, as it exists on the relevant date.

To understand how the history of purchases in a neighborhood influences purchase decisions, we compare purchases to the shadow nonpurchase decision. Since control variables (except for the history of car purchases within a particular subject’s neighborhood) do not vary day to day, this approach generates virtually the same relative coefficients as regressions using daily data (which would involve nearly a billion observations), while maintaining computational feasibility.<sup>10</sup>

For each year analyzed, we exclude residents who lack data on address or income or who resided at the same address for only a portion of the year. We also require that all subjects (whether car purchasers or not) be at least eighteen years old before the beginning of the year being analyzed. Subjects who appear to be spouses of the consumer whose action is being analyzed are excluded as neighbors.<sup>11</sup>

### D. Variable Construction and Methodology

Our analysis largely consists of regressions, with a binary action of a subject in a given year as the dependent variable. This action may be buy versus not buy a car, buy versus not buy a new car, buy versus not buy a used car, or buy versus not buy a particular make, like a Honda, or a particular model, like a Honda Accord. The right-side variables describe the history of the actions of neighbors, and attributes of the subject (including attributes of his neighborhood and point in time) whose action is the dependent variable.

Each resident appears as three observations except for those who moved in a given year. In this case, they are excluded from the year of the move. With three years of binary decisions as the dependent variable, we end up with 2,520,575 binary-decision observations.

<sup>10</sup> We have verified this by running some of our analysis with monthly data.

<sup>11</sup> Spouses are identified using the following criteria: same latitude and longitude coordinates, same move-in and move-out dates, same marital status, same number of children, opposite sex, age difference less than ten years. This exclusion avoids confounding neighborhood effects with spousal effects. For example, if one of the spouses buys a car, the other is less likely to buy a car, which would erroneously be interpreted as suggesting that a purchase by the very closest neighbor has a negative effect on the purchase behavior of the subject. While some automobiles are jointly owned, each automobile is listed as having only one primary owner.

<sup>8</sup> The classification by ZIP code is provided by *Statistics Finland*.

<sup>9</sup> The data vendor for latitude and longitude coordinates assigns exact latitude and longitude for each street intersection. The vendor, who knows the number of buildings on each side of each street, interpolates the coordinates to obtain latitude and longitude for each building. The interpolation algorithm assumes that each building between two adjacent intersections is of identical size. For example, if the distance between two adjacent intersections is 200 meters and there are eight buildings between the two intersections on a given side of the street, then each building is assumed to be 25 meters wide. All individuals living in the same building have the same latitude and longitude coordinates.

The subjects' control variables, which can change from year to year, but are relatively stable, are as follows. *Age*: The subject's age in years. We also employ the square of age to test for nonlinear effects.<sup>12</sup> *Kids*: A dummy variable that takes on the value 1 if the subject has at least one dependent that is less than eighteen years old.<sup>13</sup> *Cohabiter*: A dummy variable that takes on the value 1 if the subject individual has a live-in partner he or she is not married to. *Rural and suburban dummies*: The type of community of the individual analyzed. The zero value for both dummies is classified as a "city area." *Homeowners*: A dummy variable that is 1 if the subject has real estate or apartment wealth. *Unemployed*: A dummy variable that takes on the value of 1 if the subject collected unemployment benefits for at least one day in the prior calendar year. *Travel cost*: The subject's work-related travel costs (in thousands of euros) declared in the prior year's tax filing. *Income decile dummies*: In each year, all subjects are assigned to one of ten equal-sized deciles based on the sum of income from labor and capital. The highest income decile is the omitted dummy variable. *Year dummies*: The year of the buy versus not-buy decision. The omitted dummy is 2001.

For each subject, we rank order neighbors by their distance from the subject. The 500 closest neighbors are assigned a distance ranking from 1 to 500 with 500 being the most distant neighbor. If several individuals live exactly at the same distance, the rankings for the individuals within the distance category are assigned randomly.<sup>14</sup>

With these variables in mind, we run pooled time-series and cross-sectional regressions with each resident assigned to a single date in a given calendar year. If the subject is a purchaser, the date  $t$  is the actual purchase date in that year; if the subject is a nonpurchaser for that year, the date  $t$  is the shadow purchase date in the calendar year to which the subject is assigned by the algorithm described earlier.

Our model of the prototypical regression used in the paper can be described with the functional form:

$$\begin{aligned} & \text{Binary Decision}(\text{date } t, \text{subject } i) \\ & = f(\text{date } t \text{ car purchase history of } i\text{'s} \\ & \quad \text{neighborhood}) \\ & + g(\text{other date } t \text{ attributes of subject } i). \end{aligned} \quad (1)$$

Our primary goal is to infer the derivatives of  $f(\cdot)$  with respect to its arguments by looking at coefficients in linear and nonlinear specifications, like logit.

<sup>12</sup> For privacy protection, all persons born prior to 1910 are assumed to have been born in 1909. There are a negligible number of automobile owners within this group, for obvious reasons.

<sup>13</sup> Number of children, as a substitute variable, yields virtually identical results.

<sup>14</sup> Population density is likely to influence neighborhood relationships if we use actual distance in lieu of distance ranks. Moreover, the number of people a person is likely to know and befriend is likely to be fairly independent of population density. This argues for distance ranks as the more appropriate distinguishing characteristic of social influence.

### E. Dimensions to the Car Buying Decision

In addition to studying factors that drive the decision to buy or not buy an automobile in a given year, we also analyze the decision of which make to buy, which model to buy, and whether to purchase a new car or a used car. Each make of car, for example, Honda or Mercedes, is assigned its own code. Models are assigned dummy variables only if they can be identified as separate models and have been for sale as new cars between 1996 and 2001. Models are aggregated at the main type level. For example, Honda Accords and Honda Civics are treated as separate models, but no attempt has been made to separate the LX and EX versions of the two models.

The model year of the purchased car is not reported. Hence, we assess whether a car purchase is new or used with a decision algorithm that makes use of the registration history of an automobile. If registration followed the U.S. standard, cars whose sale date corresponded to the first date of registration would be new cars, and the remainder would be used cars. However, Finnish law differs: new vehicles sold to consumers sometimes have already been registered by the dealer. In this case, the first registration date is prior to the sale. The FVA also records an event date, which may correspond to the first date a car registered in a foreign country was brought into Finland from the foreign country. All cars with a sale date greater than six months past the earlier of the event date or the first registration date are assumed to be used cars. It is possible that some new cars sat on dealer lots for more than six months and then were sold; however, such anomalous misclassifications are likely to be rare.

For each of the most important make and model classes, we study separate regressions that analyze the decision to buy that particular make or model. Moreover, we report the average and median coefficients across each of the make (model) regressions.

### F. How Data Quality Helps Resolve Methodological Issues

The data in our study provide us with numerous advantages. For example, we have data that pinpoint the latitude and longitude of each consumer's domicile and that of his neighbors, along with the car purchase actions of each of them on every day in a multiyear period. The location precision offered by the data implies that each consumer has a unique hierarchical set of neighborhoods, ordered by distance rank, as opposed to an amorphous ZIP code or census tract. There is some, but highly imperfect, overlap in neighbor nearness—even if consumer  $j$  is one of  $i$ 's closest neighbors, the reverse need not be true. Certainly, the car purchase history of the neighborhoods of each of the ten nearest neighbors of  $i$  and  $j$  will differ. This is as much the case when the distances between such neighbors are extremely small (implying high-density buildings) as in settings where the distances that separate neighbors are great.

As we will shortly demonstrate, the nature of this data, along with the ability to assign precise dates to actions, helps address criticism of prior research on group effects.

In addition to a consumer's geographic location, precise action on every day in a multiyear sample period, as well as the five-year historical state of the actions of his 500 nearest neighbors on every day, we have data on numerous attributes for every potential consumer of an automobile (as discussed above). These attributes not only serve as controls—they also help to assess the mechanism through which interpersonal effects influence automobile consumption. Most importantly, the distribution of these control variables across near and more distant neighbors suggests that omitted variables (due to, for example, endogenous location decisions) are unlikely to account for our results.

To better understand how the quality of the data in this study advances the understanding of group effects on consumption, we now discuss how the nearest neighbors methodology addresses the two fundamental issues that plague inferences about group effects. The nearest neighbors methodology begins by studying a general specification of the influence of the behavior of the 500 nearest neighbors over the prior five years. It then excludes all neighbor behavior that appears to have negligible influence on a subject's automobile consumption. The resulting more parsimonious specification classifies a neighborhood as consisting of only the fifty nearest neighbors. This neighborhood is further subdivided into two rings: an inner ring, consisting of the ten nearest neighbors, and an outer ring, consisting of the next forty nearest neighbors.

The nearest neighbors methodology controls for unobserved variables by measuring the degree of excess influence that the car purchase history of the ten nearest neighbors have relative to the forty next nearest neighbors. The analysis of excess influence largely addresses concerns that unobserved neighborhood attributes that influence both the location and car purchase decisions might confound our inferences.

To illustrate this point by example, we linearize equation (1) and divide up the arguments of  $g$  into observed and unobserved variables. Letting  $a$ ,  $A$ ,  $x_o$ ,  $x_u$  denote the subject's action, neighborhood state vector (that is, car buying history), observed vector of subject attributes, and unobserved vector of subject attributes, respectively, equation (1) can be represented as

$$a = fA + g_o x_o + g_u x_u. \tag{2}$$

(Without loss of generality, we assume throughout the paper that  $x_u$  and  $x_o$  are uncorrelated. If they are correlated, a simple change of variables allows this result to follow: the coefficient  $g_o$  in equation (2) is increased by the coefficient from projecting  $x_u$  onto  $x_o$ , and  $x_u$  in equation (2) is replaced by the orthogonal component of the projection.) The inference problem here is that the residual in empirical specification,  $g_u x_u$ , (which incorporates the constant term) is cor-

related with  $A$ . This correlation biases the estimate of  $f$  away from 0. It can arise if the consumer's observed and unobserved attributes drive him to locate in a neighborhood with like neighbors or if the subject acquires a particular attribute from living in the neighborhood. These unobserved attributes, shared with neighbors, must also be related to car purchases to generate an inference bias.

The nearest neighbors methodology addresses this issue by identifying separate  $A$ s for two subneighborhoods, with subscripts 1 and 2 denoting the inner- and outer-ring neighborhoods, respectively. In this case, consumption behavior takes the form

$$a = f_1 A_1 + f_2 A_2 + g_o x_o + g_u x_u, \text{ with} \tag{3}$$

$$A_1 = A_2 + e, \tag{4}$$

where  $e$  is uncorrelated with  $x_o$  and  $x_u$ . The latter assumption will be justified shortly. Provided this assumption holds, then the regression

$$a = f_1(A_1 - A_2) + g_o x_o + [g_u x_u + (f_1 + f_2)A_2], \tag{5}$$

with the residual being the term in brackets, conservatively estimates  $f_1$ .<sup>15</sup> The nearest neighbors methodology uses equation (5)'s specification.

How does one prove that the error term in equation (4),  $e$ , is uncorrelated with  $x_u$ , a vector that we cannot measure? The next section demonstrates that all of the numerous observable characteristics,  $x_o$ , are uncorrelated with  $e$ , measured as the difference in the prior ten-day automobile purchase rate between the inner- and outer-ring neighborhoods. It is therefore quite reasonable to conclude that this lack of correlation extends to those characteristics that we cannot measure, the  $x_u$ 's. As an added measure of assurance, we also demonstrate that the car purchase action of the subject is uncorrelated with any difference in the average observable characteristics between the inner- and outer-ring neighborhoods. Finally, as we outline later, the results on used cars, and on the role of social influence for the purchase of similar makes and models of cars, are inconsistent with any plausible explanation for our results based on an omitted variable bias.

To understand why the data also address the reflection problem, we develop a simple example of feedback between a subject and a neighborhood (with no additional control variables to complicate the exposition). Let  $a$  denote the action of the subject and  $A$  (measured in similar units to  $a$

<sup>15</sup> The conservative estimation stems from  $A_2$  being negatively correlated with  $A_1 - A_2$ , which biases a positive estimate of  $f_1$  towards 0. The negative correlation is not only an empirical fact, but a theoretical conclusion from the highly plausible assumptions that  $A_2$  is positively correlated with  $A_1$ , and is drawn from a probability distribution with the same mean and no larger variance than  $A_1$ , as is the case here. If  $A_2$  is included as a regressor, the residual is correlated with  $A_2$ . This can bias estimates of both neighborhood coefficients, depending on the structure of the joint dependence of  $A_1$ ,  $A_2$ , and  $x_u$ .

for expositional simplicity) denote the action of the neighborhood. A system of equations describing how each influences the other is below:

$$a = fA + u, \quad (6)$$

$$A = ha + v. \quad (7)$$

This is a simpler representation of equation (5), used to illustrate a criticism of social influence research in the literature: due to the reflection back to the neighborhood from the subject,  $u$  and  $A$  are correlated, biasing estimates of  $f$ , and there is no reduced form to this system to possibly identify  $f$ .

How serious is the estimation bias induced by reflection? Equations (6) and (7) are equivalent to a pair of equations with one of the equations

$$a = f \left[ \frac{1}{1 - fh} v + \frac{h}{1 - fh} u \right] + u, \quad (8)$$

and where the second equation is the statement that the term in brackets in equation (8) equals  $A$ . A bias arises from two possible sources. The first is a possible correlation between  $u$  and  $v$ , the omitted variable bias from shared unobserved common attributes. The second is a reflection bias associated with the magnitude of  $h$  because the  $u$  inside the bracket is (perfectly) correlated with the residual  $u$  to the right of the bracket. (One can ignore the negligible bias induced by  $fh$  possibly being nonzero, as this is clearly a second-order effect.) If  $u$  and  $v$  are independent, equation (8) implies that the simultaneous equations bias is negligible if  $h/(1 - fh)$  is close to 0.

The nearest neighbors approach in equation (5) measures  $A$  as  $A_1 - A_2$ , the difference between the rate of inner- and outer-ring purchase rates. Thus,  $v$  is largely sampling error, arising from differences in unobserved attributes across two highly similar rings of neighbors. We contend, based on empirical evidence, that such a  $v$  is largely independent of  $u$ . Thus, if there is negligible reflection from the subject to the neighborhood, the bias from estimating the structural equation directly is minimal. This is largely the case here.  $h$  is close to 0 for two reasons: First, neighborhood actions are mostly in the past. Second, the nearest neighbors methodology measures  $A$  as the difference between the ten-day actions of the inner- and outer-ring neighbors. The feedback to the difference in inner- and outer-ring purchases—already minimized by past purchases dominating its makeup—are likely to be far smaller than the effect of the difference on the subject. This is because of a lack of commutativity in the nearest neighbors relationship, as discussed earlier.

A few empirical figures from an analysis of our data set helps to quantify this lack of commutativity. A consumer has only a 41% chance of being one of the ten nearest neighbors of one of his ten nearest neighbors and about a 10% chance of being one of any random ten of the outer-ring neighbors

of one's ten nearest neighbors.<sup>16</sup> The 31% difference in these two percentages suggests that the nearest neighbors approach generates at least a threefold reduction ( $1/0.31$ ) in any bias from the reflection problem.<sup>17</sup> The reduction is more than fivefold for regression estimation using data points from urban areas only.

We reiterate that these calculations do not consider the substantial reduction in the reflection problem due to past purchases being the driving component of the neighborhood purchase variable. It would be difficult to credibly claim that a consumer's purchases influenced his neighbors' purchases three days, five days, or ten days in the past. Even in the absence of the lack of commutativity in the nearest neighbors' relationship, the daily frequency at which we collect sales data allows estimation that minimizes any bias induced by feedback from the consumer to his neighbors.

The nearest neighbors methodology also controls for biases associated with "between-subjects" fixed effects if differences between inner- and outer-ring purchases are not related to these fixed effects. However, it has an advantage over the traditional panel approach of adding dummy variables for each subject when a stable neighborhood characteristic interacts with a sudden temporal event. For example, if gullibility is an unmeasured attribute of a neighborhood, a newspaper advertisement could make neighborhood residents simultaneously purchase a car. This would be spuriously interpreted as social influence, even with traditional fixed-effects estimation. The problem is that adding dummy variables alone does not control for the full neighborhood attribute—a gullible set of neighbors who have just read an advertisement. Another difficulty is that subject dummies in logit regressions eliminate most of the data observations (about 80%). This is because the logit likelihood function for such estimation is unaffected by consumers who never purchase a car.

We believe that the nearest neighbors methodology effectively controls for fixed effects without explicit controls for each subject. To verify this, we study coefficients from the nearest neighbors specification with additional explicit controls for fixed effects. Using both logit and OLS, we estimate neighbor influence using fixed-effect dummies for each subject. All of our specifications and estimation methods arrive at the same conclusions about near neighbor

<sup>16</sup> Where the influence is strongest, for the nearest neighbor, the consumer has only a 16% chance of being a nearest neighbor of one's nearest neighbor (and drops to 9% for urban areas).

<sup>17</sup> The calculation assumes that equation (7)'s reverse influence,  $h$ , is identical for each of one's inner-ring neighbors. In general, if  $x$  is the probability of being an inner-ring neighbor of one's inner-ring neighbors and  $4y$  is the probability of being an outer-ring neighbor, the reflection problem in equation (8) is reduced from  $h$  to  $(x - y)h$ . For example, if  $x$  and  $y$  are identical, there is no reflection problem. Here, one is four times as likely to be an outer-ring as an inner-ring neighbor, as the relative group sizes would imply if nearness is not commutative. In this case, the subtraction of the outer-ring purchase rate from the inner-ring purchase rate eliminates the bias from  $h$ .

influence. Hence, additional explicit controls for fixed effects have no impact on our findings.

### III. Results

Table 1 presents summary statistics on the data. Panel A presents the number of residents who purchased or did not purchase an automobile, both new and used, in each of the three years of the study. Panel B breaks the purchases down by month of the year.<sup>18</sup> As can be seen from this panel, car purchases are relatively rare events with a pronounced seasonality. The warmer weather months and early fall generate more car buying.

Panel C of table 1 presents car buying propensities based on several control variables. As can be seen from panel C, car buying propensities are smaller for those who are unattached to a significant other or who lack children; renters; females; and urban dwellers. The propensities increase in income, which is our proxy for socioeconomic status.

#### A. Marginal Effects of Control Variables

Our analysis begins with a logit regression. The dependent dummy variable equals 1 if the subject purchases a car in a particular year. The regressors are a set of control variables and 135 variables representing the state of the car purchase history of the subject's neighborhood. The first column of numbers in table 2 presents logit regression coefficients for the regression's control variables (described in the last section) estimated over all subjects. The coefficients for all income decile dummies, except the ninth, are negative and monotonic in the deciles. The lower the income, the lower is the likelihood of purchasing a car, other things equal. Despite statistical significance, arising from the large sample size, the marginal effect of income rank on car buying propensity is about the same for the eighth, ninth, and tenth income deciles. Older people also have a larger propensity to purchase a car, but very old people, as indicated by the age-squared coefficient, have less of a propensity to buy a car than middle-aged people. Males, subjects with children, those who are married or cohabiting with an unmarried partner, homeowners, and those with high travel costs also are more likely to purchase cars. Those collecting unemployment benefits are more likely to

purchase cars, perhaps because they have lost access to a company car or other transportation provided by their employer. The coefficients on the urban and suburban dummies indicate that subjects with greater distances to travel are more likely to purchase cars.

The remaining columns in table 2 present coefficients for the same control variables, where the logit regressions are estimated separately for city, suburban, and rural communities. The control variables have much the same impact as they did in the overall logit regression except that the effect of being single (as opposed to married or cohabiting) no longer has a negative effect on car buying propensity in suburban and rural areas. This may have something to do with the impact of public transportation in cities with young professionals who are single and prefer not have a car. No similar transportation alternative may be available in suburban and rural areas.

Finally, the spread of the income coefficients is larger in cities than in suburban areas and it is larger in suburban areas than in rural areas. This is consistent with the argument that a purchase of a car has the least necessity attached to it in cities (where public transportation tends to function well) and the least luxury attached to it in rural areas (where public transportation is less available).

#### B. The Influence of Neighbors on the Automobile Purchase Decision

Figure 1 illustrates the impact of neighbors' automobile purchases from the previously discussed regression, by graphing the coefficients on the regression's remaining 135 variables. These variables are the number of cars purchased by neighbors at a certain distance rank interval and within a certain time interval.

Figure 1 graphs these 135 coefficients for each of the four logit regressions. If each neighbor car purchase on a given day has the same influence, no matter how distant the neighbor or how far in the past, and influences are linearly additive, then the 135 coefficients would be identical. Obviously, figure 1 suggests that they are not. The coefficients for the nearest neighbors and the most recent purchases by those neighbors, graphed closest to the origin, are substantially larger than those found elsewhere in the graph. There is a sharp peak in each of the graphs, corresponding to the nearest neighbor on the same day, which has a coefficient of 1.3.<sup>19</sup> Each of the surfaces in the four graphs decline as the neighbors become more distant and their purchases occur further back in time.

Neighbor purchases from more than ten days ago have little influence. In figure 1, panel A (all observations), the average of all of the coefficients from days  $-11$  to  $-30$  is below 0.01; for the more distant past, the averages are even smaller. All but two of the coefficients associated with purchase behavior more than ten days in the past are below

<sup>18</sup> The seasonalities (by month and year) in the fraction of new versus used cars are partly due to a truncation effect. A new car owned on June 10, 2002, tends to have been owned for a longer period of time than a used car. Since the more distant years and early calendar months in our sample tend to be furthest from June 10, 2002, we see the new car fraction largest in the early calendar months and distant years. For the same reason, the trend toward more car purchases over time is a biased representation of what actually took place. Cars bought in 1999 and sold in 2001 appear only as 2001 purchases in our sample. However, cars tend to be held for a fairly long period of time in Finland, so the increased frequency of purchases may partly be due to Finnish economic growth, which peaked in 2000. This truncation does not affect our conclusions about social influence, which are robust when analyzed with monthly data or run separately for each calendar year.

<sup>19</sup> As suggested earlier, we have been careful about excluding spouses.

TABLE 1.—DESCRIPTIVE STATISTICS OF AUTOMOBILE PURCHASES AND NONPURCHASES

Panel A. Number of Purchases and Nonpurchases by Year				
	1999	2000	2001	Totals
New car purchases	19,922	24,066	19,993	63,981
Used car purchases	34,100	49,367	63,725	147,192
Purchases, totals	54,022	73,433	83,718	211,173
Nonpurchases	774,467	773,942	760,993	2,309,402
Purchases and nonpurchases, totals	828,489	847,375	844,711	2,520,575

Panel B. Number of Purchases and Nonpurchases by Month				
Month	Purchases	Nonpurchases	Totals	Fraction of new
1	15,280	168,861	184,141	0.394
2	13,696	150,493	164,189	0.333
3	17,363	191,357	208,720	0.329
4	17,816	197,846	215,662	0.334
5	20,402	223,330	243,732	0.337
6	18,999	208,854	227,853	0.316
7	18,984	208,076	227,060	0.280
8	19,752	213,846	233,598	0.281
9	19,052	208,150	227,202	0.279
10	19,541	210,715	230,256	0.270
11	17,098	184,738	201,836	0.257
12	13,190	143,136	156,326	0.227
Totals	211,173	2,309,402	2,520,575	0.303

Panel C. Propensity to Purchase by Year				
	Propensity to Purchase by Year			Totals
	1999	2000	2001	
Females	0.038	0.051	0.056	0.048
Males	0.096	0.128	0.148	0.124
18–24	0.036	0.059	0.085	0.060
25–29	0.064	0.095	0.128	0.096
30–34	0.078	0.109	0.136	0.107
35–39	0.084	0.111	0.132	0.109
40–44	0.084	0.109	0.128	0.107
45–49	0.082	0.106	0.119	0.102
50–54	0.080	0.104	0.111	0.098
55–59	0.075	0.095	0.100	0.091
60–64	0.062	0.077	0.078	0.073
65–69	0.049	0.058	0.058	0.055
70–	0.021	0.025	0.025	0.024
Single	0.048	0.068	0.083	0.067
Cohabiter	0.086	0.120	0.147	0.118
Married	0.081	0.104	0.113	0.099
No kids	0.055	0.074	0.086	0.072
Kids	0.090	0.119	0.136	0.115
Lowest income	0.026	0.036	0.046	0.036
2	0.029	0.046	0.061	0.046
3	0.028	0.043	0.056	0.042
4	0.033	0.046	0.058	0.046
5	0.050	0.066	0.080	0.066
6	0.060	0.081	0.098	0.080
7	0.071	0.098	0.114	0.095
8	0.091	0.119	0.135	0.115
9	0.109	0.139	0.149	0.132
Highest income	0.120	0.149	0.151	0.140
Nonhomeowner	0.045	0.066	0.087	0.066
Homeowner	0.081	0.103	0.108	0.098
Employed	0.068	0.089	0.100	0.084
Unemployed	0.045	0.070	0.094	0.081
Urban	0.055	0.073	0.083	0.070
Suburban	0.081	0.106	0.119	0.102
Rural	0.090	0.122	0.144	0.119
Positive travel costs	0.108	0.140	0.160	0.136
Zero travel costs	0.053	0.072	0.082	0.069
Whole sample	0.065	0.087	0.099	0.084

For each of the three years 1999 to 2001, panel A reports the total number of car purchases and nonpurchases in two Finnish provinces. Automobile purchases are classified into two main categories, new cars and used cars. A car is assumed new if its sale occurs no more than six months after the first registration day. Individuals who did not purchase a car in a given year are recorded as nonpurchasers. Panel B reports the monthly distribution of purchases and nonpurchases. In a given year, the number of nonpurchases for a particular month has been computed by assuming that the distribution of nonpurchase dates is the same as the distribution of purchase dates. The fraction of new automobile purchases indicates the proportion of new car purchases to all purchases. Panel C reports the propensity to purchase in each of the three years based on classifications using the following control variables: gender, age, marital status (single, cohabiter, or married), dependents under eighteen years (yes/no), total income rank deciles (based on labor plus capital income), homeownership status, employment status, the type of community in which the subject is living (urban, suburban, or rural), and travel costs declared in taxation (yes/no).



TABLE 2.—BASELINE LOGIT REGRESSIONS OF NEIGHBOR INFLUENCE BY TYPE OF COMMUNITY

	Coefficients				t-values			
	All	City	Suburban	Rural	All	City	Suburban	Rural
Independent variables								
Constant	-3.468	-3.560	-3.342	-2.857	-124.90	-98.33	-54.13	-38.94
Male	0.884	0.977	0.836	0.668	170.63	139.40	82.26	55.72
Age	0.029	0.028	0.033	0.030	26.94	20.27	14.88	11.45
Age squared	-0.001	-0.001	-0.001	-0.001	-45.14	-34.59	-22.65	-19.49
Kids	0.022	0.026	0.017	0.042	3.41	2.94	1.37	2.86
Married	0.084	0.146	-0.015	-0.057	13.30	17.63	-1.14	-3.74
Cohabits	0.109	0.162	0.018	-0.038	8.75	9.40	0.75	-1.37
Rural	0.203				25.09			
Suburban	0.096				15.00			
Homeowner	0.168	0.176	0.180	0.147	28.36	22.72	14.70	10.39
Unemployed	0.119	0.128	0.089	0.118	11.40	9.27	4.15	4.87
Travel cost	0.049	0.012	0.091	0.070	18.21	2.46	18.76	15.36
Individual's social class								
Lowest	-1.227	-1.347	-1.147	-0.990	-71.30	-57.64	-33.91	-24.69
2	-0.887	-1.019	-0.763	-0.642	-66.69	-57.37	-28.56	-20.24
3	-0.703	-0.824	-0.565	-0.471	-55.57	-48.58	-22.05	-15.86
4	-0.564	-0.700	-0.412	-0.295	-47.52	-43.89	-17.47	-10.51
5	-0.327	-0.428	-0.206	-0.141	-31.52	-30.80	-10.18	-5.59
6	-0.205	-0.285	-0.106	-0.077	-21.47	-22.22	-5.71	-3.27
7	-0.121	-0.183	-0.056	0.006	-13.52	-15.27	-3.25	0.27
8	-0.014	-0.054	0.030	0.081	-1.65	-4.81	1.89	3.78
9	0.040	0.017	0.068	0.105	4.95	1.61	4.53	4.97
Year 1999	-0.138	-0.144	-0.157	-0.208	-14.84	-12.78	-7.27	-7.65
Year 2000	-0.015	-0.020	-0.025	-0.039	-2.13	-2.21	-1.64	-2.04
Pseudo R <sup>2</sup>	0.081	0.083	0.065	0.067				
N	2,520,575	1,636,620	552,648	331,307				

Table 2 reports coefficients and test statistics for control variables for four logit regressions with a three-year panel. The dependent variable in all regressions is a dummy variable indicating whether an individual purchased a car in a given year. Nonpurchasers for a year are assigned to days within a year based on the daily distribution of purchases within that year. The control variables include male dummy, the subject's age in years, the square of age, a dummy variable that is 1 if the subject has at least one dependent, marital status dummy (1 = married), a cohabits dummy (1 = have a live-in partner), rural and suburban dummies depending on the type of community in which the subject lived, homeownership dummy (if the subject had real estate or apartment wealth the previous year), unemployment dummy (if the subject collected unemployment benefits during the prior year), travel costs (the subject's work-related travel costs in thousands of euros during the prior year), income decile dummies based on the sum of labor and capital income, and year dummies for years 1999 and 2000. The 135 time-distance variables included in the regression are reported in figure 1. Each time-distance variable is computed as the number of cars purchased by the neighbors at that distance rank and time interval.

0.05 and only one of the two, for the very nearest neighbor, exceeds 0.1; most coefficients are far smaller.

Beyond the ten nearest neighbors, there is only modest influence and virtually all of it is from purchases within the ten prior days.<sup>20</sup> Only two of the coefficients from these more distant neighbors in the very recent past exceed 0.05, and most are far smaller. There is very little influence among the most distant neighbors. One cannot find a coefficient above 0.05 for neighbors that are not among the fifty nearest.

Panels B–D of figure 1 indicate that neighborhood influence varies inversely with population density: rural areas exhibit the greatest neighborhood influence while cities exhibit the least. These differences are largely driven by the influence of the nearest neighbor purchasing a car 0–4 days prior to the date of the car buying decision. On day 0, for example, the coefficient on the same day, the nearest neighbor dummy coefficient is more than twice as large for rural areas as it is for cities.

Because the coefficients in these figures are from logit regressions, and the probability (and odds ratio) of buying

an automobile on a given day is close to 0, it is easy to translate the logit coefficients in the figures (and our subsequent tables) into marginal probability effects: the exponential of the logit coefficient will approximately be the proportional scaling of the automobile purchase probability for one additional neighbor purchase (within that neighbor category). For example, assume the annualized probability of purchasing an automobile is 0.08. A logit coefficient of 0.14 implies that the 0.08 probability increases to 0.092 ( $\exp(0.14) \times 0.08$ ) for one additional neighbor purchase in the relevant neighbor category. The increase from 0.08 to 0.092 on an annualized basis, or alternatively, from 0.08/365 to 0.092/365 in the equivalent daily probability, represents a 15% increase ( $\exp(0.14) - 1$ ) in the propensity of buying an automobile. By contrast, a logit coefficient of 0.4 translates to a 49% increase in the same propensity.

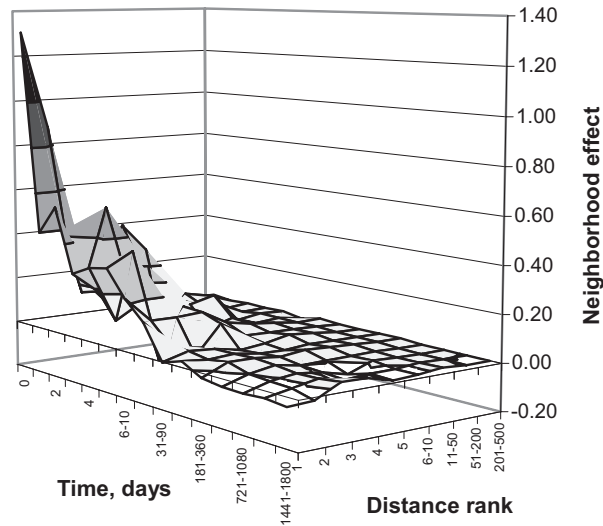
### C. A Parsimonious Representation of the Neighborhood Effect

Figure 1 suggests that there is an effect from the broader community that, while fairly small, does not decay as distance increases beyond the tenth nearest neighbor or more than ten days in the past. The existence of this “outer ring” effect could be due to an omitted variable bias that exaggerates the neighborhood effect even for the ten nearest

<sup>20</sup> The rapid decay in social influence as a function of distance is consistent with the results in Marmaros and Sacerdote (2006). They find that sharing the same freshman year dorm (but not the hallway) doubles the number of emails sent to a fellow student. When they share the same floor, the number of e-mails expands by a factor of 4.6.

FIGURE 1.—THE JOINT EFFECT OF TIME AND DISTANCE RANK ON NEIGHBOR INFLUENCE

A. Whole Sample



B. Urban Communities

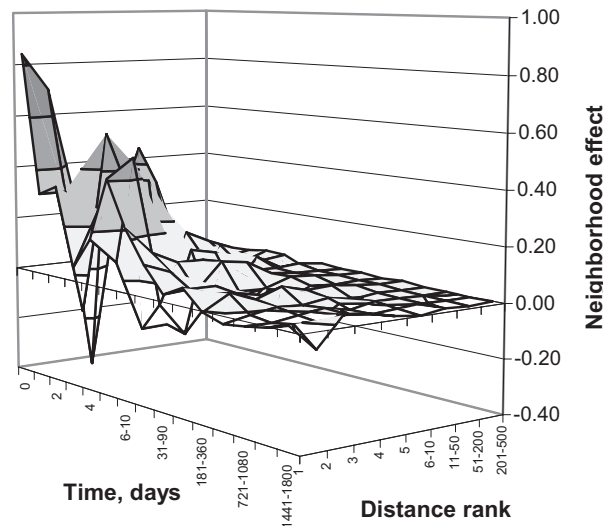


Figure 1 plots 135 time-distance variable coefficients for the logit regressions of neighbor influence described in table 2. The dependent variable in all regressions is a dummy variable indicating whether an individual purchased a car in a given year. Nonpurchasers for a year are assigned to days within a year based on the daily distribution of purchases within that year. Each time-distance variable is computed as the number of cars purchased by the neighbors at that distance rank and time interval. There are nine distance rank intervals and fifteen time intervals. Distance intervals denoted by numbers 1 through 5 represent the number of purchases of each of the five nearest neighbors (usually zero or one), whereas intervals 6–10, 11–50, 51–200, and 201–500 represent the collective number of purchases of 5–300 neighbors, depending on the interval. Time intervals  $t_1$ – $t_2$  refer to the number of purchases by a particular group of neighbors between  $t_1$  calendar days ago and  $t_2$  calendar days ago. A single number means that  $t_1$  equals  $t_2$ . Panel A plots the coefficients for the whole sample, panel B for individuals living in urban communities, panel C for individuals living in suburban communities, and panel D for individuals living in rural communities. The coefficients for the control variables are reported in table 2.

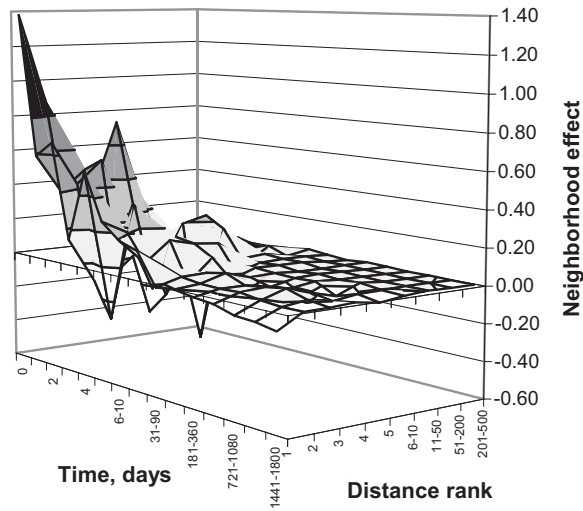
neighbors. For example, we have no data that might indicate whether a particular community has excellent or poor public transportation. Cross-sectional variation across communities in this unobservable dimension could generate a spurious neighborhood effect. We want to understand how automobile purchase decisions by neighbors arising from their specific idiosyncratic preferences alter purchase propensities among very near neighbors. To isolate this effect, even in the presence of unobservable control variables, we create the variable

*Neighborhood effect*: the number of cars purchased by the ten nearest neighbors in the last ten days less the expected number of purchases among the ten nearest neighbors,

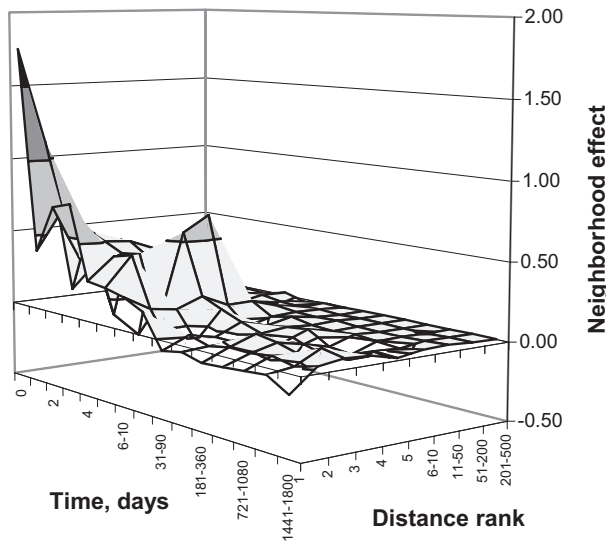
where the “expected number of purchases among the ten nearest neighbors” is 1/4 the number of purchases over the prior ten days among neighbors ranked 11th through 50th in nearness. The latter, represented as  $A_2$  in equations (3)–(5), can be viewed as a base neighborhood purchase rate.

FIGURE 1.—(CONTINUED)

C. Suburban Communities



D. Rural Communities



Subtracting this outer-ring purchase rate controls for omitted common attributes shared by residents of both the inner and outer ring that influence purchases of residents within the combined inner- and outer-ring neighborhood, as well as purchases of the subject. The neighborhood effect variable constructed from this subtraction, which is represented by  $A_1 - A_2$  in equation (5), is negatively correlated with the outer-ring purchase rate. We now present evidence that it is uncorrelated with the subject's attributes and that the difference in inner- and outer-ring attributes is unrelated to the subject's propensity to purchase an automobile.

D. Empirical Evidence on the Impact of Omitted Variables on Estimation

An omitted variable, such as an endogenous location effect, could only be the spurious source of our subsequent

results if (a) differences in this variable between the inner- and outer-ring neighbors are correlated with the subject's car purchase decision and (b) the variable lies outside the subspace generated by all of our measured control variables. These two conditions seem implausible, not just because of the number of control variables we employ, but because of the similarities in distance from the subject shared by the inner and outer rings. The average distance of the subject to an inner-ring neighbor is 26 meters, while the average distance to an outer-ring neighbor is 93 meters. The inner- and outer-ring neighbors truly share the same neighborhood with the subject.

We also verify this empirically. Table 3 presents data on the relation between the observable control characteristics,  $x_o$ , and the difference in the inner- and outer-ring neighborhood variable. The first column presents the correlation

TABLE 3.—DIFFERENCES BETWEEN THE INNER- AND OUTER-RING NEIGHBORS AND OMITTED VARIABLES BIAS

Independent variables	Correlation with Inner-Ring Purchases – Outer-Ring Purchases/4	Inner-Ring Average/Outer-Ring Average – 1		
		Subject Classification		
		All	Purchases	Nonpurchases
Male	0.03%	0.19%	–0.02%	0.21%
Age	–0.16%	–0.22%	–0.59%	–0.19%
Kids	0.15%	0.28%	0.87%	0.22%
Married	0.04%	–0.37%	–0.13%	–0.39%
Cohabiter	0.16%	–0.65%	1.21%	–0.84%
Single	–0.07%	0.42%	0.04%	0.45%
Homeowner	0.00%	0.10%	0.35%	0.08%
Unemployed	0.04%	–0.80%	–0.72%	–0.81%
Travel cost	–0.03%	0.38%	0.97%	0.31%
Social class (1–10)	0.11%	0.07%	0.24%	0.06%
Auto purchases, conditional on # days before individual's purchase				
0–10		1.33%	14.25%	–0.07%
11–30		0.59%	2.18%	0.42%
31–90		0.40%	0.59%	0.38%
91–1,800		0.32%	0.21%	0.33%

Table 3 analyzes the degree to which differences in the purchases (attributes) of inner- and outer-ring neighbors are related to the attributes (purchases) of the subject. Inner-ring neighbors are the ten nearest while outer-ring neighbors are those ranked 11–50 in nearness. The first column reports the correlation between the attribute of the subject and the difference between the inner-ring neighbor purchases and one-quarter of the outer-ring neighbor purchases. The remaining columns report the percentage deviation of a ratio of averages of various variables from 1. The numerator in the ratio is the average of the variable for the inner-ring neighbors. The denominator is the average of the variable for the outer-ring neighbors. The percentage deviations are computed for all subjects, for all subjects who purchase a car, and for all subjects who do not purchase a car.

between each of ten control characteristics of the subject and the neighborhood difference variable. It does not appear as if any of the ten observed control variables are correlated with the neighborhood difference variable. The remaining three columns present data on the similarity of fourteen characteristics between the ten nearest neighbors of the subject, and between neighbors who are among the fifty nearest (but not within the ten nearest) of the subject. Across all subjects, the observed attributes of inner-ring neighbors are highly similar to those of the outer-ring neighbors. The two rightmost columns suggest that the similarity in the attributes of the inner- and outer-ring neighbors also applies when we divide the subjects into car purchasers and non-purchasers, with one exception: purchasers of cars have inner-ring neighbors with approximately 14% more car purchases in the last ten days per neighbor than outer-ring neighbors. In short, differences in each of the observable control variables between the inner- and outer-ring neighborhoods are unrelated to a subject's propensity to purchase a car. Only purchases by near neighbors influence this propensity.<sup>21</sup>

#### E. An Analysis of Neighbor Influence with the Nearest Neighbors Approach

The first column of table 4 describes the logit regression results using the more parsimonious one-variable representation of the neighborhood effect in lieu of the more complex 135 neighborhood variables. The control variables

have approximately the same coefficients as those reported in table 2. The coefficient on neighborhood effect, 0.112, is highly significant, with a *t*-statistic of 9.71. In other words, the logged odds ratio increases by 0.112 if, in the same ten-day time frame, your ten nearest neighbors purchased one additional car relative to your more distant neighbors. A logit coefficient of 0.112 means that the probability of a car purchase is scaled up by a factor of about 12% (multiplied by about 1.12) for each additional near neighbor purchase in the last ten days. Given that the daily probability of buying a car is close to 0, one still achieves a negligible probability of a car purchase on a given day, no matter how many neighbors have purchased cars in the last ten days. However, as a percentage of that low probability of a car purchase on a given day, the increase is quite substantial.<sup>22</sup>

To help understand what factors might drive the influence of neighbors, we analyze how the subject's income decile and his decile relative to his neighbors affects the neighbor influence coefficient. (For example, envy is unlikely to drive the consumption of a subject whose income level precludes purchases of luxuries.) Figure 2 plots the neighbor influence coefficient for the regression in table 4, panel A, run separately for each income decile. Those in the lowest income decile are most influenced by neighbor purchases. Here, the lowest income decile's logit coefficient of 0.25 corresponds to a 1.29 times larger probability of buying a car on a given day for each recent purchase by one of the ten nearest neighbors.

<sup>21</sup> The *p*-values for the standard test statistics for table 3 greatly overstate significance because of the overlap in the neighborhood constituencies. We defer to our other analyses for statistical significance. What is clear from this table is that automobile purchase differences across the inner and outer rings are an order of magnitude larger for purchasers versus nonpurchasers than the control variables.

<sup>22</sup> The summary statistics here are also impressive. Of the 211,173 purchases, there were 7,662 purchases by the ten nearest neighbors in the prior ten days, which is 1,983 more purchases than would be expected by chance, assuming that purchasers and nonpurchasers are otherwise identical and there are no endogenous location effects. We control for the latter in the analysis found in the body of the paper.

TABLE 4.—A PARSIMONIOUS REPRESENTATION OF SOCIAL INFLUENCE AND EFFECTS OF RELATIVE INCOME

	Model 1	Model 2	Model 3	Model 4
Independent variables				
Neighborhood effect conditional on				
All observations	0.112 9.71			
Neighbor's income decile lower than individual's income decile		0.083 4.55	0.083 4.55	
Neighbor's income decile the same as individual's income decile		0.146 5.07		
Neighbor's income decile greater than individual's income decile		0.115 6.54		
Neighbor's income decile greater than or equal to individual's income decile			0.123 8.22	
Neighbor's income decile minus individual's income decile equals				
-3				0.103
-2				2.27
-1				0.050
0				1.27
1				0.009
2				0.25
3				0.146
4				5.07
5				0.087
6				2.58
7				0.136
8				3.58
9				0.108
Other				2.49
				0.140
				6.57

Table 4 reports results from parsimonious neighborhood effect regressions, analogous to those in table 2 and figure 1. Instead of the battery of 135 time-distance variables in figure 1, the neighbor effect is the number of automobiles purchased by the ten nearest neighbors in the last ten days less one quarter the number of purchases by the neighbors ranked 11th through 50th in nearness in the last ten days. This parsimonious regression specification includes the same control variables as in table 2, but the coefficients on the control variables are omitted for brevity. Table 4 reports coefficients and *t*-statistics (below the coefficient) for four variations of the logit regressions. The dependent variable in all regressions is a dummy variable indicating whether an individual purchased a car in a given year or not. In model 1, the neighbor effect is the number of automobiles purchased by the ten nearest neighbors in the last ten days less one-quarter the number of purchases among the neighbors ranked 11th through 50th in nearness in the last ten days. In models 2 thru 4, neighbor purchases are computed in an analogous manner, but are divided into two or more subcategories depending on the income decile of the neighbors in relation to that of the subject. The income deciles of a subject and her neighbor are based on their total income (labor plus capital income). Income class 1 refers to the lowest total income decile of all individuals in the sample.

The remaining columns in table 4 quantify the effect of relative income on consumption in more detail. The second and third columns show that the car purchase behavior of neighbors in the same income decile has the greatest influence, while the least influence is among neighbors in lower-

income deciles than the prospective automobile purchaser. In economic terms, model 2 implies that each recent purchase by a near neighbor whose income is below yours multiplies your daily propensity of buying a car by 1.09, while the corresponding effect from a neighbor in the same

FIGURE 2.—THE EFFECT OF INCOME DECILE ON NEIGHBOR INFLUENCE

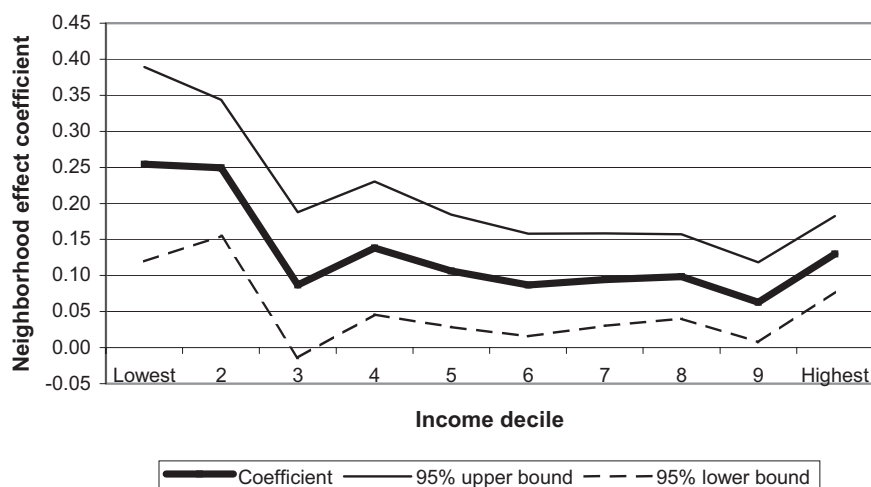


Figure 2 plots the neighbor effect coefficients and their 95% upper and lower bounds for each income decile. The results are obtained from ten logit regressions where each regression is restricted to only those individuals belonging to the income decile. The dependent variable in all regressions is a dummy variable indicating whether an individual purchased a car in a given year. The neighbor effect is the number of automobiles purchased by the ten nearest neighbors in the last ten days less one-quarter the number of purchases among the neighbors ranked 11th through 50th in nearness in the last ten days. A subject's income decile is based on the sum of labor and capital income. The control variables are the same as in table 2.

TABLE 5.—USED VERSUS NEW CARS: NEIGHBOR INFLUENCE REGRESSIONS

	Buy Any Car vs. Not	Buy New Car vs. Not	Buy Used Car vs. Not
Neighborhood effect conditional on			
Neighbor bought new car	0.082	0.084	0.072
	3.80	2.33	2.83
Neighbor bought used car	0.124	0.012	0.159
	9.10	0.48	10.20

Table 5 reports coefficients and *t*-statistics (below the coefficient) for three logit regressions. In the first column, the dependent variable is a dummy variable indicating whether an individual purchased any car in a given year. In the second column, the dependent variable is a dummy variable indicating whether an individual purchased a new car in a given year. A car is assumed new if its sale occurs no more than six months after the first registration day. In the third column, the dependent variable is a dummy variable indicating whether an individual purchased a used car in a given year. The new (used) car neighbor effect is the number of new (used) automobiles purchased by the ten nearest neighbors in the last ten days less one-quarter the number of new (used) car purchases among the neighbors ranked 11th through 50th in nearness in the last ten days. The control variables are the same as in table 2, but their coefficients are omitted for brevity.

income decile is 1.16. The fourth column, model 4, elaborates on this by studying dummy coefficients on relative income deciles, and offers similar conclusions to model 3. Model 4 also shows that the influence of neighbors in the three higher income deciles is about twice as large as the influence in the three lower deciles.

#### F. Purchases of Used versus New Cars and of Particular Makes and Models

The first coefficient column in table 5 indicates that a neighbor's used car purchase affects the probability of a purchase more than a new car purchase. The used car coefficient is about 50% larger than the new car coefficient. Table 5 also analyzes new car purchases and used car purchases separately. In the second column, the dependent (dummy) variable is 1 only if the subject makes a new car purchase. In the third column, it is 1 only if the subject makes a used car purchase. Clearly, used car purchases by neighbors influence used car purchases to a greater extent than new car purchases by neighbors influence used car purchases. For example, each recent near neighbor used car purchase increases the logged odds ratio of purchase 0.159, an increase in the purchase propensity of more than 17%. Similarly, new car purchases by neighbors influence new car purchases more than new car purchases influence used car purchases.

Table 6 panel A analyzes the logit regression of tables 2 and 3 separately for each of the fifteen most popular makes of automobiles. Panel A focuses on two influence variables rather than one. "Same make" is the number of purchases of the make listed in the row among the ten nearest neighbors within the last ten days (adjusted for the expected number of purchases of that make, in a manner analogous to the adjustment employed for the influence variable used previously in the paper). The "other makes" variable is the number of purchases of a make other than that listed in the row among ten nearest neighbors within the last ten days (adjusted for the expected number of purchases of the other makes). Clearly, a purchase by a neighbor tends to generate a purchase of the same make. The average coefficient for "same make" is more than five times the size of the influence coefficient for "other makes." For about half the

makes, there is no significant influence on the purchase probability arising from a neighbor's purchase of a different make.

The difference is even stronger for the average coefficient of the same model when we look at the ten most popular models. The variables for same make and model and same make different model are computed analogously to the influence variables studied in panel A. As table 6, panel B reports, the median "same make and model" influence coefficients are almost twice as large as the "same make different model" influence coefficients and almost ten times larger than the "other make" influence coefficient. Indeed, as panel B indicates, neighbors' purchases of different makes significantly influence purchases of only three of the ten most popular models.

One cannot easily explain the stronger findings for used cars and for similar makes and models with alternatives beyond those presented here. For our results to be due to something other than a direct influence by a neighbor's purchase, there would have to be an unmeasured phenomenon at a particular point in time that influences the car purchases of nearest neighbors but not slightly more distant neighbors. Even implausible stories like trees falling on a pair of nearest neighbors' automobiles cannot explain why the neighbors would go out and buy a similar make. Also, because the similar make tends to be a used car, and dealerships (which are the primary source for used cars) generally carry many different makes and models of used cars, there is no alternative story tied to shared trips to the same dealership that can easily explain the stronger same make and same model results. If some of the results are due to shared trips, the intent of the trip is to buy a similar make or model of car. This would mean there is a social influence in the consumption of automobiles. We will shortly turn to an analysis of what drives the social influence.

#### G. Robustness: Fixed-Effects Estimation

This subsection analyzes specifications that supplement our earlier analysis with specific controls for between-subject fixed effects. Table 7 reports that the OLS-estimated neighborhood effect coefficient, 0.0092, actually increases slightly when controlling for neighborhood fixed effects, to

TABLE 6.—EFFECTS OF THE SIMILARITY OF MAKE AND MODEL ON NEIGHBOR INFLUENCE

Panel A. Effects of the Similarity of Make Only on Neighbor Influence					
	Coefficients		t-values		
	Same Make	Other Makes	Same Make	Other Makes	
Make					
Opel	0.379	0.145	7.07		4.70
Toyota	0.516	0.131	10.31		4.28
Ford	0.410	0.106	6.30		2.99
Nissan	0.479	0.068	7.54		1.89
Volkswagen	0.232	0.012	3.02		0.29
Volvo	0.374	0.101	4.26		2.43
Peugeot	0.308	0.077	3.11		1.72
Renault	0.570	0.094	5.70		2.00
Mazda	0.456	0.081	3.75		1.56
Fiat	0.391	0.090	2.92		1.68
Citroen	0.473	0.031	3.03		0.50
Mercedes-Benz	0.532	-0.013	3.54		-0.21
Honda	0.807	0.097	4.80		1.48
Saab	1.078	0.217	6.94		3.47
Mitsubishi	-0.289	0.014	-0.84		0.18
Average	0.448	0.083			
Median	0.456	0.090			

Panel B. Effects of the Similarity of Make and Model on Neighbor Influence						
	Coefficients			t-values		
	Same Model	Different Model	Other Makes	Same Model	Different Model	Other Makes
Make and model						
Toyota Corolla	0.677	0.429	0.159	7.95	4.51	3.77
Opel Astra	0.330	0.088	0.042	2.54	0.77	0.79
Volkswagen Golf	0.350	0.149	-0.009	2.14	0.98	-0.15
Opel Vectra	0.818	0.406	0.223	4.81	3.29	3.60
Nissan Primera	0.858	0.216	-0.023	5.33	1.49	-0.33
Ford Escort	0.703	0.505	0.221	3.24	3.47	3.24
Nissan Almera	0.681	0.513	0.075	3.34	3.69	1.05
Mazda 323	0.743	0.620	0.112	3.44	2.82	1.52
Toyota Avensis	0.716	0.432	0.070	3.25	3.22	0.93
Mazda 626	0.328	0.074	0.048	1.19	0.27	0.61
Average	0.620	0.343	0.092			
Median	0.692	0.418	0.072			

Panel A reports coefficients and t-statistics for fifteen logit regressions. The dependent variable is a dummy variable indicating whether an individual purchased a car representing the given make in a given year. The same make (other makes) neighbor variable is the number of automobiles representing the same (a different) make purchased by the ten nearest neighbors in the last ten days less one-quarter the number of same (different) make purchases among the neighbors ranked 11th through 50th in nearness in the last ten days. Panel B shows the results for ten logit regressions. The dependent variable is a dummy variable indicating whether an individual purchased a car representing the given model in a given year. The same model (same make, other models) neighbor effect variable is the number of automobiles representing the same model (different models, same make) purchased by the ten nearest neighbors in the last ten days less one-quarter the number of same model (different models, same make) purchases among the neighbors ranked 11th through 50th in nearness in the last ten days. The other makes neighborhood effect is computed as in panel A. The control variables in both panels are the same as in table 2, but their coefficients are omitted for brevity.

0.0096. This magnitude implies that if there is one additional automobile purchase of the inner-ring neighbors, ceteris paribus, the annualized propensity to purchase rises by about 0.96%. Thus, if the unconditional annualized

probability of purchasing an automobile is 8%, it increases to 8.96% as a consequence of that one additional purchase. The approximately 12% increase in the annualized car purchase rate is very similar to the implied increase from logit estimation that was discussed earlier.

The final column in table 7 reports results from the logit regression of the earlier specification, but with fixed effects. The coefficient on the neighborhood variable, 0.098, is similar to the 0.112 coefficient obtained earlier without explicit controls for between subject fixed effects. In contrast to OLS addition of fixed-effects controls, which negligibly increases the neighborhood difference coefficient, here we see a decrease in the coefficient. However, since the logit fixed-effects estimation is based on only about 20% of the comparison regression's observations, the minor differences in coefficient estimates are not at all surprising. In short, the similarity in these coefficients suggests that the

TABLE 7.—FIXED-EFFECTS REGRESSIONS

	OLS		Logit Fixed Effects
	Standard	Fixed Effects	
Independent variables			
Neighborhood effect	0.0092	0.0096	0.0981
	10.01	7.81	6.99
Adjusted R <sup>2</sup>	4.23%	0.03%	0.03%
N	2,520,575	2,520,575	546,606

Table 7 reports the coefficients and t-values (below the coefficient) for OLS and logit regressions with and without fixed-effects dummies. The specification in the "standard" column is identical to that in table 4, model 1, but is estimated with OLS. It reports results that do not explicitly control for fixed effects. The "fixed effects" columns control for between-subject fixed effects by adding dummy variables for each subject. The same control variables as those in table 4 are used in all of the regressions below. The specification in the "logit fixed effects" column is identical to that in the OLS fixed effects column, except that logit estimation is used in lieu of OLS estimation.

nearest neighbors methodology does an adequate job of controlling for between-subject fixed effects, even when the fixed effects are not specifically controlled for.

#### IV. Interpreting the Results

There appears to be overwhelming evidence that the consumption of automobiles is influenced by neighbors' actions. This section analyzes the forces that lie behind the social influence we have found in consumption. The results can be summarized as follows.

- Neighbor nearness is critical to the social influence.
- The influence is fairly short-lived.
- The influence is stronger in rural areas.
- If neighborhood influence is a "keeping up with the Joneses" phenomenon, the Joneses that are most important for influencing a consumer are those of the same income decile and to a lesser extent, those in greater income deciles.
- Neighbor influence is strongest among the lowest income classes.
- Neighbor influence is stronger for used car purchase decisions than for new car purchase decisions.
- New car purchases by near neighbors influence used car purchases, but they have a stronger influence on the new car purchase decision. Even stronger is the effect of the used car purchases of neighbors on the used car purchase decision.
- The strongest influence is found for same makes and models of cars, which accounts for much of the neighbor influence effect.

One naturally gravitates to an emotional explanation for the influence of neighbors' purchases on consumption, like envy. Such an explanation permeates popular thinking and is a central motive for advertising. Economic research has long recognized that emotions driven by external consumption may alter preferences. Duesenberry (1949/1962) wrote, "What kind of reaction is produced by looking at a friend's new car? The result is likely to be a feeling of dissatisfaction with one's own . . . car. (The dissatisfaction) . . . will lead to an increase in expenditure." Veblen's *The Theory of the Leisure Class* (1899/1931) postulated that upper classes would try to distinguish themselves from the lower classes by consuming luxury goods. The lower classes would try to emulate this behavior. And today, there is a long list of models associated with a "keeping up with the Joneses" effect.

On balance, envy does not seem to fit with the neighbor influence facts we have documented here. Behavioral models of social influence based on envy would almost certainly argue that new car purchases by neighbors would have a greater influence on purchase behavior. The prominence of near neighbors for this influence is unsurprising if envy is the source of consumption emulation. However, it is diffi-

cult to explain how quickly the social influence of those nearest neighbors decays. Envy is a more persistent emotion. The Mercedes in your neighbor's driveway does not go away after a few days, a few months, or even a few years. If envy of it were driving you to consume, there is no reason to believe that influence would largely disappear after days.

There also is no a priori reason to think that envy would lead those in the lowest income deciles to be the most susceptible to emotional urges in their consumption of a luxury good, like an automobile. The loss to intrinsic well-being from a consumption "mistake" due to envy is far greater for someone with less income. Under fully rational consumption, the basket of goods of a lower-income consumer has a larger proportion of necessities. For this consumer, indulgence in luxury goods due to envy is more likely to crowd out necessities like shelter or food. If a consumption "mistake," like a Mercedes-Benz purchase arising only from envy of a neighbor is a greater sacrifice for those in the lowest income classes, it is less likely to happen.<sup>23</sup>

If envy cannot explain the results observed in the data, could other emotional explanations account for them? We considered whether snobbery or other forms of status signaling could explain our results. However, the stronger pattern of neighbor influence in rural areas, for used cars, and for same makes and models is not consistent with "one-upping the Joneses." Rural areas, which exhibit the greatest social influence, are less prone to status signaling, perhaps because there is no point to signaling status via consumption when your neighbors already know that status.<sup>24</sup> Used cars are not as much a sign of status as new cars, yet show the greatest social influence. Finally, if status signaling drives the result, why is consumption emulation so much stronger for same makes and models? When the neighbor buys a Buick, the status seeker should purchase a Cadillac, not another Buick.

On the other hand, we do not know what triggers the initial purchase by neighbors. Lacking data with which to analyze those initial purchases makes it difficult to rule out status signaling as a potential motive for those purchases. In this regard, status signaling also is consistent with the result that purchases by higher-income-class neighbors have a stronger influence than the purchases of neighbors in lower-income classes.<sup>25</sup>

If there is one emotional explanation for our findings, it is most likely to be conformity. The strongest evidence for this is that near neighbors' purchases of used cars, new cars, and

<sup>23</sup> Duesenberry (1949/1962, last paragraph) suggested that social considerations are most important in the upper-income groups.

<sup>24</sup> See Veblen (1899/1931, pp. 88–89). There is another explanation that accounts for some but not all of this result. In urban settings, many buildings contain more than ten residents. In these cases, we cannot pinpoint the ten nearest neighbors. (In cases where we can, the urban and rural influence coefficients are far closer, but the rural coefficient is still larger.)

<sup>25</sup> This prediction is in Veblen (1899/1931, p. 103).



particular makes and models increase the propensity to buy used cars, new cars, and those same makes and models, respectively. The greater mimicry of neighbors within one's own income decile also is consistent with conformity as the source of social influence. On the other hand, it is difficult to reconcile conformity as a fast-decaying emotion, essentially gone within ten days of the neighbor's purchase.

If there is no emotional motivation for consumption emulation, is it possible that other seemingly plausible explanations account for our findings? For example, could neighbor influence simply reflect that the purchase of a car triggers a sale of the purchaser's previously owned car to a neighbor or vice versa? This explanation does not seem to work either: the ages of cars purchased by inner-ring neighbors (ranked 1–10) in the prior ten days are significantly closer to one another ( $t$ -value = 2.46) than the ages of cars purchased by the more distant outer-ring neighbors (ranked 11–50). This is the opposite of what an abundance of neighbor-to-neighbor inner-ring sales would predict.

Hypotheses based on advertising, dealer sales, or any omitted variable common to a larger neighborhood cannot account for our findings either. Such hypotheses, which dispute the existence of any social influence, have a difficult time explaining why the ten nearest neighbors have so much more influence than the next forty nearest neighbors.

There seems to be more credence to information as the source of social influence. If information drives the influence coefficient, information about a specific make or model's qualities or pricing would probably be more important than information that it is a good time to buy a car. Learning that a particular make of car accelerates very nicely, that the seats are comfortable, or that research done by the neighbor suggests it gets great fuel mileage or doesn't tend to require frequent repairs is more likely to be useful to a prospective consumer. Similarly, it is more useful to learn that a particular dealer who specializes in Volvo wagons is likely to give favorable financing terms because of his current inventory situation. Such information may not be as readily available from public sources. Even if it is information of a type that is advertised in the classified section of a newspaper, friends may further publicize the information by calling attention to it with a purchase or by announcing it in conversation.

Advertising, reviews, and warranties all serve to mitigate the asymmetric information problem in new car purchases, or serve as an additional set of factors that influence purchases. They operate to a lesser degree in the used car market if at all. Hence, the information story would also predict that neighbor influence might be stronger for used car purchases. This is a category of good where quality concerns or pricing ambiguity may be particularly important.

The information story also is consistent with the demographic pattern of neighbor influence. Consumers in the lower-income deciles are more prone to purchase used cars. Lower-income consumers and those in rural areas also may

find it relatively more difficult to access sources of information that substitute for neighbor "word-of-mouth."

What is still difficult to reconcile with all of these hypotheses is that purchases by very near neighbors on the same day or in the very recent past drive much of the neighbor influence phenomenon. It is plausible that neighbors exchange information about the attributes of automobiles or dealer pricing and this information sharing induces similar purchases among neighbors. For the same-day purchases, it is likely that neighbors who have shared information are shopping together. If the information story is behind the neighbor influence coefficient pattern, the value of the neighbor's information from the purchase (or prepurchase research) should decline with time. For one, new models of the neighbor's car and substitutes for it are being introduced all the time. Public information about these automobiles, via consumer and government testing units, also may dilute the value of the neighbor's information over time. Nevertheless, it is difficult to accept that if information drives the neighbor influence, it should decline to something so negligible within ten days.

The rapid decline of the neighbor influence coefficient may simply reflect that there is predisposition on the part of the subject to buy a car that is difficult to detect. Putsis and Srinivasan (1994) document that three months is the median amount of time between a car's date of purchase and the date the buyer first contemplated its purchase. In some instances (thousands for a large sample like ours), it is the neighbor purchase that tips the scale in favor of the purchase. If this is the proper explanation, it is perhaps more likely to be information about pricing than about automobile quality that drives the influence.<sup>26</sup> For example, a consumer may travel to the dealer and find that the dealer is more flexible about pricing or financing than expected. He might also discover that the dealer has a large inventory to dispose of or an inventory containing exactly the same hard-to-get color that the neighbor is looking for. Communicating this relatively private time-sensitive information to close friends may lead to an abnormally large number of purchases within a concentrated social circle.

A recent and fascinating paper by Kapteyn et al. (2007) is consistent with our findings.<sup>27</sup> This paper studies the Street Prize of the Dutch Post Code Lottery, which showers about 9,000 euros in (after-tax) cash on all residents of a post district who bought a lottery ticket. In addition, there is one superior prize within the post district, which awards a BMW or similar-sized cash prize at the winner's option. The paper finds that there is no increase in BMW purchases in the winning post-code and that proximity to a BMW winner does not increase consumption. If price-related information is driving our short-lived neighborhood effect, dropping a

<sup>26</sup> Putsis and Srinivasan (1994) document that the actual sticker price, trade-in value, and manufacturer's discount are among the most significant determinants of the expeditiousness of a purchase.

<sup>27</sup> We are grateful to an anonymous referee for pointing this out to us.

free BMW on a neighborhood would not raise BMW consumption among the neighbors of the winner.

## V. Conclusion

It has generally been difficult to address the issue of social influence in the consumption function. Akerlof (1997, p. 1007) perhaps best sums up what numerous authors have observed—that statistically significant neighborhood effects are difficult to interpret when there are neighborhood fixed effects. Data, even at the ZIP code level, cannot eliminate such fixed effects. Among other issues, omitted variables that draw consumers to a particular neighborhood may underlie similarities in consumption among neighbors. Our unique data set on Finnish automobile consumption, consumer location, and consumer attributes has allowed us to develop a methodology to test for neighborhood effects in consumption.

Contrary to Friedman's (1957) contention that such influence does not exist, we are confident that this study has documented a highly significant social influence in Finnish automobile consumption. One's nearest neighbors' purchases appear to influence purchases, particularly of the same make and model, and of used cars, and within a short time frame. It is difficult to argue that emotional biases lie behind the observed social influence when that influence is intensified to such a degree by these factors and with this pattern.

Information transmission of some sort is better at explaining why consumers are observed to keep up with the Joneses, but puzzles remain. In particular, the fact that the neighbors exerting influence are particularly close suggests that there may be geographic barriers to learning that are worth investigating. Moreover, the rapidity with which the social influence decays could be related to the dissemination of information. However, it is surprising, nonetheless, given our priors about reaction times to information that might be relevant to a large purchase.

This paper provides many answers but also raises questions that beg for additional research. For example, the speed of the decay of the social influence coefficient has led us to conjecture that it is information about prices rather than information about intrinsic utility from the good that is responsible for our findings. However, we have not been able to establish this with any other evidence from our data. Kapteyn et al. (2007), while consistent with this conjecture, have issues with sample size and data limitations that still allow reasonable persons to dispute whether our conjecture is correct.

Our study lacked data on other sources of networking that might lead to a social influence. To truly test the robustness of our findings, it would be useful to research whether similar phenomena are observed among socially connected subjects in workplace and educational settings. Finally, there is little reason to believe that the consumption functions of Finland's residents would be materially different

from those of the residents of other nations. However, studies in other settings where the speed of information transmission differs from that in Finland may help assess whether the conjectures we have put forth here are truly valid.

## REFERENCES

- Abel, Andrew B., "Asset Prices under Habit Formation and Catching up with the Joneses," *American Economic Review* 80 (1990), 38–42.
- Akerlof, George, "Social Distance and Social Decision," *Econometrica* 65 (1997), 1005–1027.
- Bagwell, Laurie Simon, and Douglas B. Bernheim, "Veblen Effects in a Theory of Conspicuous Consumption," *American Economic Review* 86 (1996), 349–373.
- Basmann, Robert L., David J. Molina, and Daniel J. Slottie, "A Note on Measuring Veblen's Theory of Conspicuous Consumption," this REVIEW 70 (1988), 531–535.
- Bayer, Patrick, Stephen L. Ross, and Giorgio Topa, "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," Yale University working paper no. 927 (2005).
- Bearden, William, and Michael J. Etzel, "Reference Group Influence on Product and Brand Purchase Decisions," *Journal of Consumer Research* 9 (1982), 183–194.
- Becker, Gary S., "A Theory of Social Interactions," *Journal of Political Economy* 82 (1974), 1063–1093.
- Bernheim, B. Douglas, "A Theory of Conformity," *Journal of Political Economy* 102 (1994), 841–877.
- Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Changes as Informational Cascades," *Journal of Political Economy* 100 (1992), 992–1026.
- Campbell, John, and John Cochrane, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107 (1999), 205–251.
- Chan, Yeung Lewis, and Leonid Kogan, "Catching up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices," *Journal of Political Economy* 110 (2002), 1255–1285.
- Duesenberry, James S., *Income, Saving, and the Theory of Consumer Behavior* (Cambridge: Harvard University Press, 1962, fourth printing). Originally published 1949.
- Evans, William, Wallace Oates, and Robert Schwab, "Measuring Peer Group Effects: A Study of Teenage Behavior," *Journal of Political Economy* 100 (1992), 966–991.
- Fishman, Arthur, "Search Technology, Staggered Price-Setting, and Price Dispersion," *American Economic Review* 82 (1992), 287–298.
- Friedman, Milton, *A Theory of the Consumption Function* (Princeton: Princeton University Press, 1957).
- Gali, Jordi, "Keeping Up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices," *Journal of Money, Credit and Banking* 26 (1994), 1–8.
- Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman, "Crime and Social Interactions," *Quarterly Journal of Economics* 111 (1996), 507–548.
- Goolsbee, Austan, and Peter J. Klenow, "Evidence on Learning and Network Externalities in the Diffusion of Home Computers," *Journal of Law and Economics* 45 (2002), 317–343.
- Hopkins, Ed, and Tatiana Kornienko, "Running to Keep in the Same Place: Consumer Choice as a Game of Status," *American Economic Review* 94 (2004), 1085–1107.
- Kapteyn, Arie, Peter Kooreman, Peter Kuhn, and Adriaan Soeteven, "Measuring Social Interactions from the Dutch Post Code Lottery," University of California at Santa Barbara working paper (April 2007).
- Kling, Jeffrey R., Jens Ludwig, and Lawrence F. Katz, "Neighborhood Effect on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment," *Quarterly Journal of Economics* 120 (2005), 87–130.
- Leibenstein, Harvey, "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand," *Quarterly Journal of Economics* 64 (1950), 183–207.

- Liebman, Jeffrey B., Lawrence F. Katz, and Jeffrey R. Kling, "Beyond Treatment Effects: Estimating the Relationship between Neighborhood Poverty and Individual Outcomes in the MTO Experiment," John F. Kennedy School of Government working paper no. 36 (2004).
- Manski, Charles, "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60 (1993), 531–542.
- Marmaros, David, and Bruce Sacerdote, "How Friendships Form?" *Quarterly Journal of Economics* 121 (2006), 79–119.
- Morgenstern, Oskar, "Demand Theory Reconsidered," *Quarterly Journal of Economics* 62 (1948), 165–201.
- Pesendorfer, Wolfgang, "Design Innovation and Fashion Cycles," *American Economic Review* 85 (1995), 771–792.
- Peter, J. Paul, and Jerry C. Olson, *Consumer Behavior and Marketing Strategy*, 6th ed. (New York: McGraw-Hill/Irwin, 2001).
- Pollack, Robert A., "Independent Preferences," *American Economic Review* 66 (1976), 309–320.
- Putsis, William P., Jr., and Srinivasan Narasimhan, "Buying or Just Browsing? The Duration of Purchase Deliberation," *Journal of Marketing Research* 31 (1994), 393–402.
- Robson, Arthur J., "Status, the Distribution of Wealth, Private and Social Attitudes to Risk," *Econometrica* 60 (1992), 837–857.
- Sacerdote, Bruce, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics* 116 (2001), 681–704.
- Solomon, Michael R., *Consumer Behavior: Buying, Having, and Being*, 4th ed. (Upper Saddle River, NJ: Prentice Hall, 1999).
- Stigler, George, "The Development of Utility Theory. II," *Journal of Political Economy* 58 (1950), 373–396.
- Topa, Giorgio, "Social Interactions, Local Spillovers, and Unemployment," *Review of Economic Studies* 68 (2001), 261–295.
- Veblen, Thorstein, "Why Is Economics Not an Evolutionary Science?" *Quarterly Journal of Economics* 12 (1898), 373–397.
- , *The Theory of the Leisure Class. An Economic Study of Institutions* (New York: Random House, 1931, tenth printing). Originally published 1899.

Copyright of *Review of Economics & Statistics* is the property of MIT Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.