

CORRECTING FOR ENDOGENEITY IN BEHAVIORAL CHOICE MODELS WITH SOCIAL INFLUENCE VARIABLES

BY JOAN L WALKER*, EMILY EHLERS*, IPSITA BANERJEE*, ELEENNA R DUGUNDJI +

* University of California at Berkeley

+ University of Amsterdam

JL Walker is the corresponding author (joanwalker@berkeley.edu)

ABSTRACT

While psychologists and behavioral economists emphasize the importance of social influences, an outstanding issue is how to capture such influences in behavioral models used to inform urban planning and policy. In this paper we focus is on operational models that do not require explicit knowledge of the individual networks of decision makers. We employ a field effect variable to capture social influences, which is calculated as the percent of population in the peer group that has chosen the specific alternative. The peer group is defined based on socioeconomic group and spatial proximity of residential location. As in behavioral economics and psychology, the concept is that one is influenced by the choices made by one's peers. However, using such a social influence variable in a behavioral model causes complications because it is likely endogenous: unobserved factors that impact the peer group also influence the decision maker, yielding correlation between the field effect variable and the error. The contribution of this paper is the use of the Berry, Levinsohn and Pakes (BLP) method to correct the endogeneity in a choice model. The two-stage BLP introduces constants for each peer group to remove the endogeneity from the choice model (where it is difficult to deal with) and insert it into a linear regression model (where endogeneity is relatively easier to deal with). We test the method using a mode choice data set from the Netherlands and readily available software, and find there is a statistically significant upward bias of the field effect parameter when endogeneity is not corrected. The procedure outlined presents a practical and tractable method for incorporating social influences in choice models.

INTRODUCTION

In an era of pervasive and new forms of social networks, there is growing interest in the influence that social factors have on transportation and land use behaviors (Dugundji, Páez and Arentze, 2008). Here we are interested in approaches to incorporate such influence in behavioral models. Important avenues of research in this domain include capturing influences from tight social networks such as among household members (see, for example, Bradley and Vovsha, 2005; Gliebe and Koppelman, 2005; Srinivasan and Bhat, 2005; Arentze and Timmermans, 2009; Timmermans and Zhang, 2009) as well as research to understand the explicit structure of loose social networks of extended family, friends and colleagues (for example, Axhausen, 2008; Carrasco et al., 2008). In this paper we focus on implicit rather than explicit social networks, meaning that we are developing models in cases in which a decision maker's actual social network are beyond the immediate family and the exact members are unknown.

Discrete choice analysis has become an industry standard in land use and transportation planning models. However as discrete choice theory is fundamentally grounded in individual choice, an outstanding challenge remains in the treatment of the interdependence of various decision-makers' choices (McFadden 2001, 2009). To capture social influences in a choice model, we employ a field effect approach first proposed as global interactions by Aoki (1995), Brock and Durlauf (2001a), Blume and Durlauf (2003) and subsequently adapted for local interactions by Dugundji and Gulyas (2003a,b). The local field effect is the share of decision makers within a defined reference group that choose a particular alternative; such a variable can be included in the utility to capture social effects with the problem of perfect correlation with an alternative specific constant. Some explorations of the dynamical behavior of such a model with application to transportation include Dugundji and Gulyas (2005), Páez and Scott (2007), Fukuda and Morichi (2007), Páez, Scott and Volz (2008), Dugundji and Gulyas (2008). Some examples of the empirical estimation of such a model with application to transportation include Goetz (2008), Goetzke and Andrade (2009), Goetze and Rave (2010).

Here we revisit the empirical estimation application of Dugundji and Walker (2005) in which we included the field effect in a cross nested logit model for mode choice. In our earlier work, the issue that such a field effect variable is endogenous was addressed by estimating computationally intensive panel effects. The objective of this paper is to develop an alternative, more tractable approach to correct for endogeneity in discrete choice models that capture social influences via a field effect variable.

The paper is organized as follows: First the relevant literature is reviewed. Second, we present the method to incorporate social influences and correct for endogeneity. Next, we present an application to mode choice where we test the method. We conclude with a summary, limitations, and directions for future research.

LITERATURE

Here we review both the social interactions literature and the endogeneity literature. In both of these cases, the literature is large, and we emphasize that which is most relevant to this work.

As many behavioral economists and psychologists have noted, choice is social; that is, an individual's decisions are not immune from the influence of others. In addition to straight rational analysis, some decisions involve a more nuanced analysis with considerations of mutual support, reciprocity, decision-making economies and efficiencies of shared information. Manski (1995) offers three hypotheses on social group interactions:

“To explain the common observation that individuals belonging to the same group tend to behave similarly, (Manski defines)... endogenous effects, wherein the propensity of an individual to behave in some way varies with the prevalence of that behavior in the group; contextual effects, wherein the propensity of an individual to behave in some way varies with the distribution of background characteristics in the group; and correlated effects, wherein individuals in the same group tend to behave similarly because they face similar institutional environments or have similar individual characteristics.”

It is often empirically difficult to distinguish the individual ramifications of each of these phenomena.

McFadden (2009) further delineated the causes of this sociality of choice. He posits that choice is influenced by information from a peer group, heuristics rooted in the behavior of others, analogies or anecdotal information garnered from associates, and constraints imposed by others. Taken together, these two scholars, along with many others, substantiate the incorporation of social influence in choice models.

Our starting point in quantitatively considering interdependence of various decision-makers' choices is a trio of papers by economists Aoki (1995), Brock and Durlauf (2001a), and Blume and Durlauf (2003). They introduce social interactions in binary discrete choice models by allowing a given agent's choice for a particular alternative to be dependent on the overall share of decision makers who choose that alternative. If the coefficient on this interaction variable is close to zero and not important (statistically insignificant) relative to other contributions to the utility, then the distribution of decision-makers' choices will not effectively change over time in relation to other decision-makers' choices. Effectively, in this hypothetical situation, choice would not be socially conditioned. However, if the coefficient on this interaction variable is strongly positive and "dominant enough" relative to other contributions to utility, there may arise a runaway situation over time as all decision makers flock to one particularly attractive choice alternative. In short, the specification captures feedback between decision makers that can potentially be reinforcing over the course of time. In diverse literature this is referred to as a social multiplier, a cascade, a bandwagon effect, imitation, contagion, herd behavior, etc. (Manski, 1995). Brock and Durlauf (2001b) give an excellent and extensive literature review. While there has been considerable attention paid to such feedback effects in the context of linear regression models, there has been considerably less emphasis in the context of discrete choice models. Some notable exceptions are discussed below.

Since the early theoretical work on binary discrete choice models with social interactions, there have been a few extensions addressing both the complexity of the discrete choice model kernel as well as the complexity of the field effect and the utility specification. For example, Brock and Durlauf (2002) have extended their results on the behavior of binary logit models to multinomial logit models. Dugundji (2003) makes Brock and Durlauf's multinomial results precise for trinary multinomial choice and extends the analytical results for the case of nested logit with global interactions. Also, while the behavior over time derived in early work assumed each decision maker to be influenced by all other decision makers (global interactions), Dugundji and Gulyas (2003a,b) introduce a more general case where each decision maker is influenced by only a subset of decision makers (non-global interactions). This is important since a field effect defined globally would be perfectly correlated with a set of alternative specific constants in a discrete choice model. They present a framework for studying local field effects in discrete choice models, making a distinction between social network interactions versus spatial network interactions, identifiable versus aggregate interactions, proximal versus tangential versus global interactions, and exogenous versus endogenous interactions. They derive theoretical results for equilibrium behavior of binary choice on the abstract class of random networks and small world networks.

A key to the theoretical results is the assumption that the only explanatory variable in the model is the field effect. While such a specification may be plausible for a fad, it is much less intuitive for transportation mode choice where other explanatory variables would be assumed to be significant, including both attributes of the alternatives, as well as characteristics of the decision-making agents. Dugundji and Gulyas (2005) thus present results using simulated data for a binary logit model with non-global interactions and other explanatory variables included in the utility in an application to intercity

travel behavior for a parameter sweep of network density across a series of networks in the abstract class of random networks. Páez and Scott (2007) present results using simulated data for a binary logit application to telecommute behavior for a range of networks of different sizes defined by a similarity on a two-dimensional matrix of personal characteristics. Fukuda and Morichi (2007) present results for a binary logit application to illicit bicycle parking in Tokyo using empirical data where the field effect is defined by reference group on the basis of the railway station at which the user most frequently parks. Páez, Scott and Volz (2008) present results using simulated data for a multinomial logit application to residential location choice for a parameter sweep of degree distribution and clustering parameter across a series of networks in the abstract class of poisson networks. Dugundji and Gulyas (2008) present results for the behavior over time of a nested logit model in an application to transportation mode choice in Amsterdam, using empirical data and an empirical treatment of which decision makers influence each other defined on the basis of socioeconomic group (income, education, age) and spatial proximity of residential location.

The focus in this paper is on the potential endogeneity of field effect variables in the estimation of the behavioral models. Conceptually, the issue is well captured by Manski (1993) in his presentation of the “reflection” problem inherent in studying social effects, which is “the difficulty in determining whether the average behavior of some group influences the behavior of the individuals that comprise the group.” Endogeneity is a classic econometric issue for which there exists a large literature. The problem arises when an explanatory variable is correlated with the unobserved factors, and such a situation leads to biased and inconsistent parameter estimates. Endogeneity surfaces due to errors in variables, simultaneous determination, and omitted attributes, among other causes. It is an issue in both linear models (for example, regression) as well as non-linear models (for example, discrete choice). Dealing with endogeneity in regression is a fairly well understood process and estimation methods such as instrumental-variables estimation are described in most econometric textbooks (see, for example, Pindyck and Rubinfeld, 1998, or Green, 2007). Indeed, the solution we employ to correct the endogeneity of our social influence model is to move the endogeneity from the non-linear model to a linear model so that such techniques can be used.

Dealing with endogeneity in discrete choice models is less well studied. There are a few particularly useful summary papers that deal with the issue. The first is a workshop paper by Louviere et al. (2006) that presents work on recent progress on endogeneity in choice models. The second is a chapter on endogeneity from Train (2009), which provides an excellent review of methods that have been developed to estimate choice models with endogenous explanatory variables. The primary methods he covers are BLP (Berry, Levinsohn and Pakes, 1995 and 2004) and the control function approach (Hausman, 1978; Heckman, 1978) and its full maximum likelihood extension (as in Villas-Boas and Winer, 1999). The reader is referred to Train (2009) for a description of these methods, as well as additional citations containing both theoretical developments as well as applications. Guevara (2010) is also an excellent resource. The aspect he covers particularly well (Chapter 3) is why dealing with endogeneity is more complicated in discrete choice than in regression. The basic issue is that the corrections often lead to changes in the error term of the utility function, which leads to a change in scale of the model. For example, the standard 2-stage instrumental variable approach used in regression produces consistent estimates of all of the parameters of the utility function *with the exception* of the scale of the model (Newey, 1985) and this causes complications for forecasting. As forecasting is important for transportation, we seek, if possible, a method that also provides consistent estimates of the scale. It is worth noting that there has been little published regarding endogeneity in the transportation literature. The

only applications of which we are aware beyond earlier work by the authors (Dugundji and Walker, 2005) are the original BLP (1995, 2004) papers cited above on the auto market, Guevara and Ben-Akiva (2006) who applied the control function approach to correct for price endogeneity in residential housing choice, Train and Winston (2007) who used BLP to correct for price endogeneity in auto ownership choice, Goetzke (2008) and Goetzke and Andrade (2009) who apply a conditional spatially auto-regressive binary choice model with moving mode share averages to study endogenous effects in public transit and walking respectively, and Goetzke and Rave (2010) who derive an instrument from records with excluded trip purposes to study endogenous effects of “bicycle culture” in German cities. Goetzke’s work in particular is notable due to the attention paid in the last three mentioned articles on specific policy implications of endogenous social network effects in the empirical cases studied.

The approach we employ in this paper is the BLP method, which applies when endogeneity occurs at the level of a market segment. This is well-suited to the issue of capturing social influences in our case where the individual network is not explicit, because the aggregate peer groups can be thought of in terms of market segments. BLP has an advantage over other approaches for dealing with endogeneity in discrete choice, because the correction does not modify the scale of the choice model. Therefore forecasting with the resulting BLP-corrected model is more straightforward. The procedure will be described more fully when we discuss the method. Note that extension of this work for the case of individual-specific and not market-specific networks could be done by employing the other primary method of correcting for endogeneity in choice models, which is the control function approach (see Train, 2009; and Guevara, 2010).

To summarize, we aim to capture social influences in discrete choice models using a field effect variable. However, this variable is likely endogenous exhibiting what Manski (1993) terms the *reflection* problem. Therefore, we aim to correct for endogeneity using the BLP procedure proposed by Berry, Levinsohn, and Pakes (1995, 2004). This endogeneity correction is the contribution of this paper.

METHOD

In this section we first introduce the use of a field effect variable to capture social influences in a choice model. This is an approach developed previously and described in the background section above. Next, we describe the issue of endogeneity that arises and propose a correction. The approach we take follows a suggestion from McFadden (2009).

In incorporating social influences into choice models, ideally, one could pinpoint the extent of a decision-maker’s social network and gauge the effect of this network on the decision-maker’s choice calculus. Impalpable, social networks vary across individuals; and furthermore, the degree to which social networks influence a given decision maker also varies. Because it is difficult, and potentially invasive, to determine the precise extent and supremacy of one’s social network, an aggregate “field effect” is introduced to replicate this influence. This field effect collects people with similar attributes into a peer group and captures social interactions by allowing each person’s choice to depend on the overall choice probabilities of the other people in the decision-maker’s peer group. More explicitly, the field effect is defined as the percent of people in the peer group exhibiting a given behavior.

Thus, to capture social influences we incorporate the field effect variable into random utility models. Starting with a basic utility equation U for a given alternative i for a given decision maker n

$$U_{in} = V(x_{in}, s_n; \beta) + \varepsilon_{in}, \quad (1)$$

where $V(x_{in}, s_n; \beta)$ is the systematic portion of the utility (consisting of attributes of the alternatives x_{in} , characteristics of the decision maker s_n , and taste parameters β) and ε_{in} is the unobservable portion of the utility. Now incorporating the field effect, we have

$$U_{in} = V(x_{in}, s_n; \beta) + \gamma F_{in} + \varepsilon_{in} \quad (2)$$

where F_{in} is the proportion of those people in the peer group of decision maker n who chose alternative i and γ is a parameter that captures the magnitude of the influence.

This is a choice model with social influences, which can be introduced in any random utility model and can be estimated. However, equation 2 as written may suffer endogeneity, leading to inconsistent estimates of the parameters (both γ and β).

ENDOGENEITY

The endogeneity issue arises because similar, unobservable environment and preferences impact both the decision maker being modeled as well as the behavior of the people in the decision-maker's peer group. This is to be expected if a decision-maker's choice depends on the behavior of other people related to her—as assumed by the theory of sociality. Then, it is also likely that the same observed and unobserved attributes influence both the individual and everyone in her peer group. These unobserved effects influence the field effect F_{in} and are captured in the random term ε_{in} , and therefore F_{in} and ε_{in} may be correlated and F_{in} is suspected to be an endogenous variable. We expect the field effect variable and the error to be positively correlated, and therefore an estimate of γ from equation 2 will be biased upward.

For example, a decision maker may be influenced by those who reside in the same neighborhood. However, in a transport and land-use context, people who live proximally to the decision maker will face the same attributes of alternative modes as the decision maker. The attributes that make transit or driving attractive or unattractive to a particular commuter may similarly affect his or her neighbors. Or, if the decision maker lives along an attractive bike path, then her neighbor also lives along or near the bike path, which makes bicycling arguably easier or more fun for both neighbors. Inevitably in this scenario, both the observed factors (i.e. travel time) and unobserved factors (i.e. aesthetics of the bike route or perceived safety of bus ride) will be similar for the decision maker and those who live nearby.

While the influence of land-use and spatial factors is critical in the mode choice context, social influence extends beyond spatiality. Decision makers may be influenced by those in their same social class. However, presumably there are unobserved factors that affect people in the same social class. For instance, some classes may have a stigma surrounding transit use—a la “only poor people take transit.” Alternatively, if bicycling is considered hip in certain social circles, this would represent a shared unobserved preference.

It is the shared unobserved effects, whether taste heterogeneity or omitted attributes, for persons in the same peer group that lead to the endogeneity of the field effect. That is, the field effect variable, capturing

these unobserved influences of spatial proximity or social class, is correlated with the error in the decision-maker's utility. Using the field effect variable without correcting for endogeneity, assuming it exists, is a fatal flaw as it produces inconsistent estimates of the parameters.

BLP CORRECTION

The approach we employ here to correct for endogeneity is that proposed by Berry, Levinsohn and Pakes (1995 and 2004). Their approach is relevant when the endogeneity can be considered at a market level, applying similarly to decision makers within a given market. In our case, the markets are peer groups, where the groups are defined based on spatial proximity and social class along the lines of the examples described above. Each spatial and social group then, for the sake of the BLP procedure, can be considered a market.

The BLP procedure involves decomposing the error into two parts: the endogenous-causing part and the random portion. The utility equation shown above then becomes

$$U_{in} = V(x_{in}, s_n; \beta) + \gamma F_{in} + \ddot{\epsilon}_{in} + \dot{\epsilon}_{in}, \quad (3)$$

where $\ddot{\epsilon}_{in}$ is correlated with F_{in} , and $\dot{\epsilon}_{in}$ is uncorrelated with F_{in} (and x_{in}, s_n).

One key to the BLP procedure is to isolate the endogenous-causing components, that is, F_{in} and $\ddot{\epsilon}_{in}$. The terms are thus rearranged as follows:

$$U_{in} = [\gamma F_{in} + \ddot{\epsilon}_{in}] + V(x_{in}, s_n; \beta) + \dot{\epsilon}_{in}. \quad (4)$$

The first term $[\gamma F_{in} + \ddot{\epsilon}_{in}]$ represents the observable and unobservable components of utility relevant to the peer group. The second term $V(x_{in}, s_n; \beta)$ represents the remainder of the systematic utility of the individual (that is, the portion not related to the peer group). The error term $\dot{\epsilon}_{in}$, by construction, is orthogonal to all explanatory variables in the model, including F_{in} and $\ddot{\epsilon}_{in}$.

The second key to BLP is related to the setup of the procedure, which is the assumption that the endogeneity occurs at a market level. In our case, each peer group is a market and we add a market delineator m to denote the peer group to which a decision maker belongs. Modifying the utility equation accordingly leads to:

$$U_{in_m} = [\gamma F_{im} + \ddot{\epsilon}_{im}] + V(x_{in_m}, s_{n_m}; \beta) + \dot{\epsilon}_{in_m}. \quad (5)$$

Thus, the first term $[\gamma F_{im} + \ddot{\epsilon}_{im}]$ represents the unobservable and observable components of utility relevant to the individuals' peer group m . It represents the average, or common, utility of a given choice in a given group. This term varies across peer groups but does not vary across individuals in the same group. The second term $V(x_{in_m}, s_{n_m}; \beta)$ represents the systematic portion of utility that varies across decision makers. The error term $\dot{\epsilon}_{in_m}$ is orthogonal to all explanatory variables and varies across decision makers.

The trick in the BLP procedure is to now replace the peer group effect with market specific constants α_{im} for each alternative i and each peer group m , such that the new utility equation is:

$$U_{in_m} = \alpha_{im} + V(x_{in_m}, s_{n_m}; \beta) + \dot{\varepsilon}_{in_m} , \quad (6)$$

where $\alpha_{im} = [\gamma F_{im} + \ddot{\varepsilon}_{im}]$.

These constants capture the average effects of the peer group. There is no endogeneity issue in the choice model as written this way, and therefore the parameters α_{im} and β are estimated via usual choice modeling procedures. (A very large number of markets may require that the constants be estimated via the “contraction” approach described in BLP, although our application did not require this.) Note, though, that we are interested in the social effect as represented by the parameter γ , which is not estimated via the choice model.

The final step of the BLP procedure is to estimate via linear regression the market-specific constants as explained by the field effect variable, or:

$$\alpha_{im} = \gamma F_{im} + \ddot{\varepsilon}_{im} . \quad (7)$$

While the endogeneity issue remains (F_{im} is correlated with $\ddot{\varepsilon}_{im}$), it is more straightforward to correct for endogeneity in the linear model. For this we use a two-stage instrumental variables approach. In the first stage, instrumental variables I_{im} (correlated with the field effect variable F_{im} and uncorrelated with the error $\ddot{\varepsilon}_{im}$) are used to explain the field effect variable F_{im} as follows:

$$F_{im} = \theta_i + \theta_F I_{im} + v_{im} , \quad (8)$$

where v_{im} is a random error (orthogonal to I_{im}) and θ_i and θ_F are estimated parameters. We will discuss the selection of instruments when we present the empirical application.

In the second stage, the market-specific constants are regressed on the fitted value of the field effect from the first stage ($\hat{F}_{im} = \hat{\theta}_i + \hat{\theta}_F I_{im}$) as follows:

$$\alpha_{im} = \gamma_i + \gamma_F \hat{F}_{im} + \ddot{\varepsilon}_{im} , \quad (9)$$

As \hat{F}_{im} is orthogonal to $\ddot{\varepsilon}_{im}$, this regression results in a consistent estimate of γ_F , which captures the effect of the field effect variable on the utility. This can then be inserted back into the choice model (replacing the market-specific constants with equation 9) so that the choice model captures the effect of the peer group. Note that the entire right-hand side, including the fitted value of the error, is included in the final choice model.

In summary, the BLP process removes the endogeneity from the choice model via the use of market-specific constants. The endogeneity is then dealt with in a linear regression setting (with instrumental variables) to obtain consistent estimates of the social influence effect. This consistent estimate of the field effect parameter is then reintroduced to the choice model to obtain a choice model that captures social influences. Further, in contrast to the other methods of correcting for endogeneity, the scale of the choice model is not modified with the procedure.

There are two important design issues required to implement this procedure. The first is the definition of the peer groups (m as introduced in equation 5). The second is the selection of the instrumental variables (I_{im} in equation 7). We will describe how we handle these issues in our application, described next.

APPLICATION

To test such a correction, we re-visit the application in Dugundji and Walker (2005), which incorporated a field effect variable into a model of mode choice to work. In the 2005 paper, an attempt was made to correct for the endogeneity of the field effect variable via a correlated error structure. While the a priori hypothesis was that endogeneity existed, evidence of such was not uncovered in the 2005 paper. Further, the correlated error structure approach was intensive computationally requiring use of a high performance computing cluster to explore the range of model specifications and different network connectivity definitions studied in the paper within a reasonable timeframe. The objective here is to develop the same type of model, but employ the BLP procedure to correct for endogeneity, which is computationally more tractable.

The data used originate from travel questionnaires administered by the Municipality of Amsterdam Agency for Infrastructure, Traffic, and Transport between 1992 and 1997 in the Netherlands' capital city, Amsterdam, and a neighboring suburb to the south, Amstelveen. Due to limitations in the available data, a trip-based home-to-work model is considered. The estimation dataset consists of 2913 respondents.

The a priori hypothesis regarding sociality is that one's mode choice to work is dependent on the behavior, beliefs and preferences of people in the decision-maker's peer group. Therefore, we include a field effect variable in the mode choice model to capture this effect.

The mode choice model is specified as a choice among auto, transit, and bike. The explanatory variables include level of service variables (time and cost by mode), socio-demographic information (sex, income, education, and age), and the field effect variable to capture sociality. The key issue in including the field effect variable is the definition of the peer group; this is described next.

DEFINITION OF THE PEER GROUP

The question is how to define a decision-maker's network of influence without having direct knowledge of his or her individual ties. Based on the hypothesis that people are most influenced by both those within spatial proximity and those who are in similar socio-economic groups. We use observable factors on these dimensions of space and socio-economics to define the peer groups, and we refer to these as *spatial* and *social* reference groups.

We assign *spatial* reference groups to everyone in our sample according to his or her residential postal code (a zonal definition used in Amsterdam). These postal codes range in size from less than a square kilometer (in the center of Amsterdam) to 15 square kilometers (in Amstelveen). The average postal code area is 3.31 km². These two municipalities in the Netherlands are relatively dense, with a population density of 4,451 people per km² (Netherlands Central Bureau for Statistics, 1999). However, there is tremendous variability within our sample. The smallest postal code has only 193 inhabitants, while the largest has 21,130 inhabitants (Netherlands Central Bureau for Statistics, 1999). The average postal code size is 8,252 inhabitants. Our a priori assumption is that these postal code boundaries delineate spatial peers and that spatially proximate individuals are more similar, exerting a stronger influence than individuals who live farther away. There are 67 postal codes and, therefore, 67 spatial reference groups.

We also define a *social reference group* analog based on income, under the assumption that people are influenced by others in his or her own income class regardless of physical location. Here the idea is that

one's friends and colleagues are likely of a similar income class. The people in our sample reported their income category; there were six categories ranging from 0 to above 60,000 Dutch guilders per year. As such, we define six social reference groups.

Thus, the reference group of each decision maker is defined by the postal code in which they live and their level of income. There are two primary issues with our approach to specifying the reference groups. The first is the issue of sharp boundaries, in which we assume that decision makers are only influenced by those within either their social reference group or spatial reference group, and those individuals outside these groups are not influential. Our definition draws on the first law of geography (Tobler, 1970), which is that everything is related to everything else, but nearer things are more related than farther things. That is, while everyone is influenced by everyone else, we make the assumption that the strongest influence is by those who are nearest either spatially or socially and the influence outside of these boundaries is marginal. Further, we delineate community boundaries using the postal code definition used in Amsterdam and social boundaries using the income classification of the Amsterdam survey. This leads us to the second issue, which is the classic modifiable area unit problem or MAUP (see Openshaw, 1984). We are limited by our dataset in terms of how we can define both our spatial boundaries and our class boundaries. We investigated this issue as best we could with the existing dataset. On the spatial side, we did have an additional zonal classification that consisted of larger zones (and fewer in number). However, these zones proved to be too large to reflect any significant social influence in the models. On the social side, we explored different classifications of the social reference group (incorporating age and education), however such definitions did not improve the specification over the income classification used for the results presented in this paper. Our explorations were of course limited by the definition of demographic categories collected in the survey. There are other spheres of influence which could be explored beyond our analysis, including spatial influence at the work location, other demographic influences such as household structure (e.g., children or not), and alternatives to the sharp boundary approach (e.g., use of a spatial weighting matrix as in Goetzke, 2008).

The result of these two issues (sharp boundaries and MAUP) are that the definitions of our reference group boundaries become integrally important to the specification, and the definitions we make regarding zones and class essentially become a part of the model. What we aim to do in this paper is take a stab at defining these groups (flaws and all) and test the approach of the use of the field effect and the BLP correction. What we will show in the estimation results is that this approach appears to work quite well. The technique will then only be improved as the definitions of these reference groups improve and the underlying assumptions made stronger.

Once the reference groups are defined, the field effect variables are calculated for each individual, which is the percentage of people in each group who choose to drive to work, the percentage who choose to take transit, and the percentage who choose to bike. Each decision maker has both a spatially and a socially-defined reference group; therefore, each decision maker has two sets of field effect variables (one spatial, one social). In our application, aggregate data were not available so we used the sample data (2913 respondents) to calculate these modal shares by reference group. These field effects capture the average attractiveness of each mode to work within a given group. By entering the field effect into the utilities, we allow each person's choice to depend on the overall choice probabilities of the other people in the decision-maker's reference group.

SPECIFICATION OF THE INSTRUMENTAL VARIABLES

As described above, the BLP procedure removes the endogeneity from the choice model by specifying group specific constants that subsume the correlated field effect and error term into a constant for each reference group (the α_{im} defined above in equation 6). In a second stage, these market-specific constants are then regressed on the field effect. This is useful because we can then employ the instrumental variables regression estimation techniques to correct for endogeneity. The issue in the application is what to use for instruments.

The definition of an instrument is a variable that is correlated with the endogenous variable and uncorrelated with the error. In a spatial context, there are a set of natural instruments, which are the values of the endogenous variable in the spatially adjacent zones. These values are assumed to be correlated with the problem variable and uncorrelated with the error. The first part of this assumption is straightforward in our case in that the mode share in a zone is likely correlated with the mode share in the adjacent zones. This is due to the spatial continuity of both the transportation network and social structure. The second part of the assumption, that the instrument is uncorrelated with the error, is always trickier to defend. In our case we need to rely on our zonal definitions as being true to our concept of reference groups, where the boundaries are meaningful. For our instrument, we assume that the predominant social influences are coming from individuals within the decision-makers' zone and those from outside of the zone are marginal. While this can never be perfect, the postal code definitions in the Netherlands are generally defined such that there is homogeneity within a zone in terms of land uses and built environment, and heterogeneity across zones as is typical for urban planning. Further, we draw again on the first law of geography, that everything is related to everything else but near things are more similar to distant things. So while the assumption of a sharp boundary may not be reasonable, it is reasonable to assume a decay effect with distance. With these assumptions in mind, we choose to use the average field effect of the adjacent postal codes as the instrument for the endogenous field effect term.

For the social reference group field effect, we use a social parallel to the spatial instrument. That is, we use the values of the endogenous variable in the "socially" adjacent zones. For the case of income classes, we use the value of the field effect for the next lowest income class as the instrument. The idea here is that while the strongest influence is likely from those in one's own income group, there may be some influence from a higher income group due to status seeking behavior and therefore this higher income group would not serve as a good instrument. As the lowest income class does not have a lower income group, we used the field effect from the third income group as the instrument (to avoid the next highest group, which is assumed to have greater influence). We did explore different instruments (for example using both the income group above and that below), and the specification that makes use of only the lower income group as the instrument was both intuitively the most attractive and statistically the strongest.

One issue with our specification arises because each person has both a spatial reference group and an income reference group, and we treat these groups as being independent. This leads to an inconsistency with our instruments, because for the spatial instrument we assume that one is not influenced by people in the surrounding zones. However, we do assume that one is influenced by people of the same income class, and there will be people in the same income class in the surrounding zones. A simple solution would be to define each individual's reference groups as either the union or intersection of their spatial and social reference groups. However, we didn't have the luxury of this definition, because we didn't

have enough data to populate these more detailed reference groups. Even if we did have enough data for our particular application, such an approach would not be scalable to further detail of the reference group (e.g., based on education and household structure as well) and we wish to have a scalable method. Therefore, we have decided to proceed with this inconsistency. Our assumption is that while there may be some overlap, it is negligible relative to the overall effect. With the instruments for the income-based field effects, this is likely a reasonable assumption as there are a large number of postal codes and so the overlap would be minimal. The instruments for the zonal-based field effects are more suspect as we have relatively few income groups and so there may be non-negligible overlap. This issue would partially be alleviated by a richer specification of the social influence groups such that the overlap would be less significant. It is worthwhile to note that we worked with an alternative specification to overcome this problem by regressing the sum of the field effects together rather than each individually. In this case, the instruments were specific to the social-spatial combination and the overlap was removed. The results of this alternative approach were similar to the results presented in this paper, which provides some evidence that the inconsistency cited above is minimal. We retained the specification described in this paper because it is more scalable and because the issue of overlap will diminish as the definition of the reference groups become more refined.

There is another classical set of instruments that are often used, which are the other exogenous explanatory variables in the model. We explored the use of these, but in this case they were not effective for several reasons. First, the exogenous variables in the choice model are person specific (for example, level of service of transport modes to work), and therefore cannot be used in the aggregate, market-specific regressions. We tried using aggregate level of service variables such as various zonal accessibility measures, but these did not prove significant. One issue is that our access to aggregate transportation data was limited. Another issue is that transportation level of service variables were already controlled for in the choice model at the individual level, and therefore were not significant as instruments at the market level.

Now that the method has been outlined (field effect and BLP), the application described (mode choice to work in the Netherlands), the definition of the reference group made (by residential postal code and income group), and the instruments presented (use of groups that are spatially or socially ‘near’), we are ready to present the estimation results.

ESTIMATION RESULTS

The results are presented in Tables 1-3. First we will explain the layout of these tables and then will discuss the results. Detailed estimation results from different models are presented in Table 1 (discrete choice) and Table 2 (linear regression). The models include:

- A base choice model without the field effect variable (Table 1, model 1) – equation 1
- An uncorrected choice model with the field effect variable used directly (Table 1, model 2) – equation 2
- BLP corrected models, which include three parts:
 - A choice model with market-specific constants (Table 1, model 3) – equation 6
 - First stage of instrumental-variables regression (Table 2, model 2a) – equation 8
 - Second stage of instrumental-variables regression (Table 2, model 2b) – equation 9

- Additionally, the uncorrected regression estimation is presented (Table 2, model 1) – equation 7 – to compare with both the uncorrected choice models as well as the corrected parameters from the instrumental variables estimation.

Finally, as our emphasis is on the parameter of the field effect variable, we summarize in Table 3 the field effect estimated parameters from all model runs. The choice models were estimated using the open-source software BIOGEME 1.8 (Bierlaire, M. 2003).

Now we will discuss the results, starting with the choice models in Table 1. The first column is the base choice model (auto versus transit versus bike) with no social influences (equation 1). The base model estimates the effect on the mode choice decision of level of service attributes, socio-demographics, and alternative-specific constants. The signs of the level of service variables are as expected and, with the exception of parking cost, are significant. The socio-demographic variables include gender, age, education and income. There are no surprising signs and, for the most part, these parameters are significant.

In model 2 of Table 1 the field effect variables have been added directly to the utility (as in equation 2), where we introduced the field effects additively (that is, $U_{in_m} = \gamma^{social} F_{im}^{social} + \gamma^{spatial} F_{im}^{spatial} + \dots$). The parameter is generic across both the spatial field effect (defined by postal code) and the social field effect (defined by income group), that is $\gamma^{social} = \gamma^{spatial}$. Spatial- and social-specific parameters did not significantly improve model fit. The income dummies are removed from the specification because of the correlation with the socially-defined field effect terms. The field effect parameter (2.49) is statistically significant (t-stat of 12.3) and its addition significantly improves the fit of the model. The positive sign on the field effects is as expected: the probability of choosing a mode increases with the share of one's peer group that selects that mode. This suggests that people are influenced by both the mode choices of their neighbors and also of the decisions of other people in the decision-maker's social class. The magnitude, sign and significance of the other variables in the model change only marginally once the field effect is added. The only exceptions are the parameters on parking time and parking cost, which retain the correct negative sign however parking time has a decreased effect and parking cost has an increased effect. In both the base model and the base model with field effects, a minute spent bicycling is still about twice as onerous as a minute spent traveling to work via auto. One minute spent traveling by transit is the least onerous of all three modes, probably due to the convenience and comfort of transit in the Netherlands. As expected, out-of-vehicle travel time (e.g. waiting, access and egress) is more onerous than time spent traveling. The socioeconomic parameters in the model—age, education, income and gender—similarly enter both models.

While the choice model with the field effects appears attractive affirming our hypothesis regarding the interdependence of behavior among people, these field effects are suspected to be endogenous. If endogenous, the parameter estimates of the field effect variable (as well as all other parameters) are inconsistent. We expect in this case that there is a positive correlation between the field effect and the error, and therefore we expect that the parameter estimates for the field effect as reported in model 2 of Table 1 will be biased upward, overemphasizing the actual effect.

To correct for the endogeneity, we apply the BLP procedure. The first step is to estimate market-specific constants in the choice model (equation 6), and the estimation results of this choice model are in model 3,

the final column in Table 1. With 67 postal codes and six income groups, 144 constants are estimated for these markets: 6 income auto constants, 6 income transit constants, 66 postal code auto constants (one market is constrained for identification), and 66 postal code transit constants (again, one market is constrained). The ranges of the constants are reported to provide a summary of the values. The log-likelihood, of course, increases with the large number of constants, but the adjusted rho-square decreases over both models 1 and 2. Again the shifts from the other parameters in the model are marginal.

The second step of BLP is to regress these market specific constants on the field effect variable (equation 7). The estimation results are shown in Table 2. As in the choice model, we define a generic parameter on the field effect, because we found that spatial- and social-specific effects were not a significant improvement. Note that, without a correction, the regression also has the issue that the field effect is endogenous. However, for comparison purposes, we do report the uncorrected regression as model 1 in Table 2. The first thing to note is that these uncorrected field effect parameters from the regression (2.93, t-stat of 15.2) are similar in magnitude and significance to the uncorrected choice model shown in model 2 of Table 1 (2.49, t-stat of 12.3), which is as expected.

To correct for endogeneity, we use an instrumental variables estimation as described by equations 8 (first stage) and equation 9 (second stage). The instruments are as described above in the section titled ‘specification of the instrumental variables’. The results are shown in Table 2, where model 2a is the first stage to obtain the fitted values of the field effect, which are exogenous by construction. While we constrain the field effect parameter to be generic across social and spatial field effects, we allow for different intercept terms for both the social and spatial effects for both modes of auto and bus. The instruments appear to work reasonably well in that the instrument is significant with a correct sign. The fitted values from stage 1 of the regression are used to replace the endogenous field effect in the second stage. The second stage regression is reported as model 2b and here we finally have a parameter estimate for the field effect for which the endogeneity has been corrected and therefore is consistently estimated. Further, the correction performs as we would expect: the fit of model 2b (corrected) is lower than model 1 (uncorrected), the parameter on the field effect decreases (from 2.93 to 1.48) and its significance also decreases (t-stat drops from 15.2 to 2.7). Importantly, note that even after the endogeneity is corrected, the field effect parameter is still significant.

As the parameter on the field effect is the focus of this work, we summarize in Table 3 the findings discussed above. Namely, while the field effect parameters in the uncorrected choice model and uncorrected regression are of similar magnitude and significance, the parameter in the corrected regression exhibits a decrease in both magnitude and significance. Further, while the significance drops substantially, the field effect remains a significant factor in the model. To test whether the bias in the uncorrected models are significant, we apply a Hausman specification test (Hausman, 1978) on the null hypothesis that the field effect is *not* endogenous. The Hausman test is applicable when comparing two estimators with the following properties: one estimator is consistent and efficient under the null hypothesis but inconsistent under the alternative hypothesis (this is our uncorrected model) and the other estimator is consistent (not efficient) under both the null and the alternative hypothesis (this is our BLP corrected model). We apply the test to the parameter on the field effect. The Hausman test statistic is then: $(\hat{\gamma}_{uncorrected} - \hat{\gamma}_{corrected})(StdErr(\hat{\gamma}_{corrected})^2 - StdErr(\hat{\gamma}_{uncorrected})^2)^{-1}(\hat{\gamma}_{uncorrected} - \hat{\gamma}_{corrected})$, which has a chi-square distribution with 1 degree of freedom, and its square root has a normal distribution. We report t-statistics from this test in Table 3. We compare both the uncorrected choice model and the

uncorrected regression model to the BLP corrected model. In both cases we reject the null hypothesis that the field effect variable is not endogenous.

CONCLUSION

In this paper we present a tractable and statistically-valid (in that endogeneity is corrected) approach for incorporating social influences in a discrete choice model. We applied the approach to an application of mode choice to work, where the assumption is that an individual's mode choice to work does not occur in isolation but is influenced by the choices made by one's peers, friends, family and neighbors. While we do not observe each of these interactions nor the precise social network, we define a field effect variable to capture spatial and social interdependence. The field effect is defined as the proportion of people within one's peer group who choose a particular mode to work. We then apply the Berry Levinsohn and Pakes (BLP) procedure to correct correlation between the field effect and the unobserved error term. We find that after we correct for endogeneity, the spatially- and socially-defined reference group is a significant factor in mode choice behavior. An important lesson is that we demonstrated a significant upward bias (verified with a Hausman specification test) when a field effect variable is directly inserted in the choice model; the uncorrected parameter on the field effect was over 60% higher than the corrected parameter. While this is a first step that appears promising, there are further issues to explore, including the way in which the reference groups are defined, the issue of MAUP, and the validity of the instruments.

Table 1: Choice Model Estimation Results

Mode choice to Work (Logit) Alternatives: Auto, Transit, Bike		model 1			model 2			model 3		
		Base Model			Choice Models with Field Effect Variable (Endogenous)			First stage of BLP: Choice Models with Market-specific Constants *		
Parameter (& relevant utility)		Estimate	StdErr	t-test	Estimate	StdErr	t-test	Estimate	StdErr	t-test
Alt-specific Constants	Constant (auto)	0.233	0.189	1.2	-1.78	0.20	-8.8	72 constants, range: -2.5 to 3.7		
	Constant (transit)	-2.33	0.22	-10.8	-1.99	0.22	-9.0	72 constants, range: -3.8 to 2.8		
Level of service	Total travel time, minutes (bike)	-0.0740	0.0048	-15.4	-0.0708	0.0048	-14.9	-0.0754	0.0050	-15.1
	Total travel time, minutes (auto)	-0.0407	0.0079	-5.1	-0.0348	0.0080	-4.3	-0.0396	0.0086	-4.6
	In-vehicle travel time, minutes (transit)	-0.0199	0.0060	-3.3	-0.0149	0.0059	-2.5	-0.0180	0.0064	-2.8
	Out-of-vehicle-travel time, minutes (transit)	-0.0234	0.0082	-2.9	-0.0311	0.0085	-3.7	-0.0290	0.0091	-3.2
	Parking time, minutes (auto)	-0.166	0.022	-7.7	-0.0668	0.0230	-2.9	-0.118	0.027	-4.4
	Parking cost, dutch guilders (auto)	-0.0512	0.0363	-1.4	-0.154	0.036	-4.2	-0.112	0.040	-2.8
Socio demographics	Female dummy (auto)	-0.0115	0.1045	-0.1	0.0193	0.1072	0.2	0.0447	0.1146	0.4
	Female dummy (transit)	0.565	0.115	4.9	0.572	0.117	4.9	0.659	0.127	5.2
	young age dummy, 12-29 years old (auto)	-0.436	0.151	-2.9	-0.344	0.150	-2.3	-0.370	0.159	-2.3
	middle age dummy, 30-44 years old (auto)	-0.387	0.135	-2.9	-0.351	0.139	-2.5	-0.345	0.143	-2.4
	young age dummy, 12-29 years old (transit)	0.283	0.159	1.8	0.259	0.161	1.6	0.364	0.174	2.1
	middle age dummy, 30-44 years old (transit)	-0.129	0.152	-0.9	-0.165	0.154	-1.1	-0.0997	0.1608	-0.6
	Low education dummy, < high school (auto)	0.861	0.168	5.1	0.728	0.165	4.4	0.858	0.182	4.7
	Middle education dummy, high school (auto)	0.611	0.121	5.1	0.497	0.121	4.1	0.602	0.131	4.6
	Low education dummy, < high school (transit)	0.783	0.174	4.5	0.684	0.175	3.9	0.850	0.199	4.3
	Middle education dummy, high school (transit)	0.639	0.124	5.1	0.582	0.127	4.6	0.677	0.142	4.8
	Low income dummy (auto)	-0.989	0.150	-6.6						
Middle income dummy (auto)	-0.707	0.112	-6.3							
Sociality	Field Effect (auto, transit, bike)				2.49	0.20	12.3			
Number of observations		2888			2888			2888		
Number of parameters		20			19			160		
Log-likelihood at convergence		-2406			-2326			-2288		
Adjusted rho-square		0.230			0.256			0.223		

Table 2: Second stage of BLP: Regressing the market-specific constants

Model 1: Uncorrected				Model 2: Corrected				
Uncorrected Regression (OLS)	Dependent variable: Market specific constants from choice model				Dependent variable: Field effects			
	<i>- Model 1 -</i>				<i>- Model 2a -</i>			
		Estimate	StdErr	t Stat		Estimate	StdErr	t Stat
	Intercept Spatial (transit)	1.21	0.06	21.5	Intercept Spatial (transit)	-0.00897	0.02055	-0.4
	Intercept Spatial (auto)	0.250	0.070	3.6	Intercept Spatial (auto)	0.0582	0.0281	2.1
	Intercept Social (transit)	-3.45	0.19	-18.5	Intercept Social (transit)	-0.00704	0.06815	-0.1
	Intercept Social (auto)	-2.09	0.19	-11.2	Intercept Social (auto)	0.0846	0.0679	1.2
	Field effect	2.93	0.19	15.2	Field effect instrument	0.756	0.094	8.0
	Number of observations	144			Number of observations	144		
	Adjusted R Square	0.897			Adjusted R Square	0.540		
				Dependent variable: Market specific constants from choice model				
				<i>- Model 2b -</i>				
					Estimate	StdErr	t Stat	
				Intercept Spatial (transit)	1.18	0.09	13.1	
				Intercept Spatial (auto)	0.559	0.146	3.8	
				Intercept Social (transit)	-3.53	0.30	-11.8	
				Intercept Social (auto)	-1.96	0.30	-6.5	
				Fitted field effect	1.48	0.54	2.7	
				Number of observations	144			
				Adjusted R Square	0.751			

Notes

Calculation of the field effect

Because the market-specific constants are "auto relative to bike" and "transit relative to bike" (bike is the base alternative in the choice model), the field effects ("F") enter these regressions as "F_Auto - F_Bike" and "F_Transit - F_Bike"

The 146 observations are:

- 67 postal codes X 2 alternative-specific constants (auto and transit)
- + 6 social classifications X 2 alternative-specific constants (auto and transit)

Table 3: Summary of the parameter estimates for the field effect from various models

Estimator	Estimate	StdErr	t-test	Hausman test	
				t-test	conclusion
Choice - uncorrected	2.49	0.20	12.3	2.0	--> reject null of no endogeneity (i.e., 2.49 is significantly higher than 1.48)
Regression (OLS) - uncorrected	2.93	0.19	15.2	2.9	--> reject null of no endogeneity (i.e., 2.93 is significantly higher than 1.48)
Regression (2 stage IV) - CORRECTED	1.48	0.54	2.7		

REFERENCES

- Aoki, M. (1995) Economic fluctuations with interactive agents: Dynamic and stochastic externalities. *Japanese Economic Review*, 46(2), 148–165.
- Arentze, T.A. and H.J.P. Timmermans (2009), A need-based model of multi-day, multi-person activity generation. *Transportation Research B*, 43, 251-265.
- Axhausen, K.W. (2008) Social networks, mobility biographies, and travel: Survey challenges. *Environment and Planning B*, 35(6), 981-996.
- Berry, S., J. Levinsohn and A. Pakes (1995) Automobile prices in market equilibrium. *Econometrica*, 63, 841–889.
- Berry, S., J. Levinsohn and A. Pakes (2004). Differentiated products demand systems from a combination of micro and macro data: The new vehicle market. *Journal of Political Economy*, 112(1), 68-105.
- Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- Blume, L. and S. Durlauf (2003), Equilibrium concepts for social interaction models. *International Game Theory Review*, 5 (3), 193-209.
- Bradley M. and P. Vovsha (2005) A model for joint choice of daily activity pattern types of household members. *Transportation*, 32, 545-571.
- Brock, W. and S. Durlauf (2001a) Discrete choice with social interactions. *Review of Economic Studies*, 68, 235-260.
- Brock, W. and S. Durlauf (2001b) Interaction-based models. In: J. Heckman, and E. Leamer, (eds.), *Handbook of Econometrics*, vol. 5. North-Holland.
- Brock, W. and S. Durlauf (2002) A multinomial choice model of neighborhood effects. *American Economic Review*, 92(2), 298-303.
- Carrasco, J.A., B. Hogan, B. Wellman and E.J. Miller (2008) Collecting social network data to study social activity-travel behavior: An egocentric approach. *Environment and Planning B*, 35(6), 961-980.
- Dugundji E.R. (2003) Discrete choice on networks: nested logit. Presented at the *Graduate Workshop on Computational Economics*, Santa Fe Institute, New Mexico, USA. (<http://zia.hss.cmu.edu/econ/projects03.html>)
- Dugundji, E.R. and L. Gulyas (2003a) Empirical estimation and multi-agent based simulation of a discrete choice model with network interaction effects.” In *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management*, Sendai, Japan.
- Dugundji, E.R. and L. Gulyas (2003b) An exploration of the role of global versus local and social versus spatial networks in transportation mode choice behavior in the Netherlands.” In *Proceedings of AGENT 2003: Challenges in Social Simulation*, Argonne National Laboratory, Chicago, USA. (<http://www.dis.anl.gov/pubs/51409.pdf>)
- Dugundji, E.R. and L. Gulyas (2005) Socio-dynamic discrete choice in an empirical example on intercity travel demand: An agent-based approach, analytical benchmark and some issues in estimation. In *Proceedings of MUSD 2005: Workshop on Modeling Urban Social Dynamics*, University of Surrey, Guildford, UK.

- Dugundji, E.R. and L. Gulyas (2008) Socio-dynamic discrete choice on networks in space: Impacts of agent heterogeneity on emergent outcomes. *Environment & Planning B*, 35(6), 1028-1054.
- Dugundji, E.R., A. Páez and T.A. Arentze (2008) Social networks, choices, mobility, and travel. *Environment & Planning B*, 35(6), 956- 960.
- Dugundji, E.R. and J.L. Walker (2005) Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record*, 1921, 70-78.
- Fukuda, D. and S. Morichi (2007) Incorporating aggregate behavior in an individual's discrete choice: An application to analyzing illegal bicycle parking behavior. *Transportation Research Part A*, 41, 313-325.
- Gliebe, J.P. and F.S. Koppelman (2005) Modeling household activity-travel interactions as parallel constrained choices, *Transportation*, 32, 449-471.
- Goetzke, F. (2008) Network effects in public transit use: evidence from a spatially autoregressive mode choice model. *Urban Studies*, 45, 407-417.
- Goetzke F. and P.M. Andrade (2009) Walkability as a summary measure in a spatially autoregressive mode choice model: An instrumental variable approach. *Advances in Spatial Sciences*, Springer.
- Goetzke F. and T. Rave (2010, forthcoming) Bicycle use in Germany: Explaining differences between municipalities with social network effects. *Urban Studies*.
- Guevara, C.A. (2010, forthcoming) Endogeneity and sampling of alternatives in spatial choice models. PhD Dissertation, MIT, USA.
- Guevara, C. and M. Ben-Akiva (2006) Endogeneity in residential location choice models. *Transportation Research Record*, 1977, 60–66.
- Green, W.H. (2007) *Econometric Analysis*, 6th Edition, Prentice Hall.
- Hausman, J. (1978) Specification tests in econometrics. *Econometrica*, 46(6), 1251–1272.
- Heckman, J. (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46, 931–959.
- Manski, C.F. (1995) *Identification Problems in the Social Sciences*. Harvard University Press.
- Manski, C.F. (1993) Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies*, 60(3), 531-542.
- McFadden, D. (2001) Economic choices. *American Economic Review*, 91(3), 351-378.
- McFadden, D. (2009) Sociality, rationality, and the ecology of choices. Working paper, UC Berkeley.
- Newey, W.K. (1985) Semiparametric estimation of limited dependent variable models with endogenous explanatory variables. *Annales de L'insee*, 59/60, 219-237.
- Netherlands Central Bureau for Statistics (1999). *Wijk en buurt register*.
- Louviere, J., K. Train, M. Ben-Akiva, C. Bhat, D. Brownstone, T.A. Cameron, R.T. Carson, J.R. Deshazo, D. Fiebig, W. Greene, and D. Hensher (2006) Recent progress on endogeneity in choice modeling. *Marketing Letters*, 16(3-4), 255-265.

- Openshaw, S. (1984) *Concepts and Techniques in Modern Geography: Number 38 - The Modifiable Areal Unit Problem*. Geo Books, Norwick.
- Páez, A. and D.M. Scott (2007) Social influence on travel behavior: A simulation example of the decision to telecommute. *Environment and Planning A*, 39, 647-665.
- Páez, A., D.M. Scott and E. Volz (2008) A discrete-choice approach to modeling social influence on individual decision making. *Environment and Planning B*, 35(6), 1055-1069.
- Pindyck R.S. and D.L. Rubinfeld (1998) *Econometric Models and Economic Forecasts*, 4th Edition, Irwin McGraw-hill.
- Srinivasan, S. and C.R. Bhat (2005) Modeling household interactions in daily in-home and out-of-home maintenance activity participation. *Transportation*, 32, 523-544.
- Timmermans, H.J.P. and J. Zhang (2009), Modeling household activity travel behavior, *Transportation Research B*, 43, 187-190.
- Tobler W. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.
- Train, K.E. (2009) *Discrete Choice Methods with Simulation*, 2nd edition. Cambridge University Press.
- Train, K.E. and C. Winston. (2007) Vehicle choice behavior and the declining market share of US automakers. *International Economic Review*, 48(4), 1469-1496.
- Villa-Boas M and R. Winer (1999) Endogeneity in brand choice models. *Management Science*, 45, 1324-1338.

ACKNOWLEDGMENTS

We would like to acknowledge useful discussions with Daniel McFadden regarding sociality in discrete choice models and the potential of BLP for correcting endogeneity. Kenneth Train and Angelo Guevara have been very helpful regarding the issues and methods related to endogeneity in discrete choice models. Of course any issues that arise are the fault of the authors alone. Our special thanks also go to Nelly Kalfs and Hans Niepoth of the Department of Traffic Research of the Municipality of Amsterdam Agency for Infrastructure, Traffic, and Transport for data assistance.