SOCIODYNAMIC DISCRETE CHOICE ON NETWORKS: THE ROLE OF UTILITY PARAMETERS AND CONNECTIVITY IN EMERGENT OUTCOMES

Elenna R. Dugundji, Universiteit van Amsterdam, e.r.dugundji@uva.nl

László Gulyás, AITIA Inc, lgulyas@aitia.ai

ABSTRACT

The reported research treats social interactions and generated feedback dynamics in the adoption of various transportation mode alternatives. We consider a model where a commuter's choice is directly influenced by the percentages of neighbors and socioeconomic peers making each choice, and which accounts for common unobserved attributes of the choice alternatives in the error structure. We explicitly address non-global interactions within different social and spatial network structures, combining econometric estimation with computational techniques from multi-agent based simulation, and present an empirical application of the model using pseudo-panel microdata collected by the Amsterdam Agency for Traffic, Transport and Infrastructure. The paper extends previous work by the authors in considering the effects of various hypothesized sociogeographic networks. We also test for the effect of the scale of the interaction, comparing municipal district clusters versus smaller 4-digit postcode clusters. We observe that the estimated utility parameters for the different sociogeographic network scenarios can generate dramatically different dynamics and thus cannot be ignored in any empirical application. However, in a hypothetical simulation experiment we find that swapping the sociogeographic networks does not significantly change the long-run outcome of the simulation, when utility parameters are held fixed. We conclude highlighting recommendations for future work.

Keywords: social influences, scale and space, mode choice behavior, travel demand modeling, multi-agent based social simulation

INTRODUCTION

There is growing awareness and interest in the influence that social factors have on transportation and land use behaviors (1, 2). We consider a model where a commuter's transportation mode choice is influenced by percentages of neighbors and socioeconomic peers making each choice. Such inter-household feedback can have very important implications for the prediction of (system-wide) results over the course of time. If such feedback exists, it can namely propel or hinder the adoption of a mode over time (3). In diverse literature this dynamically reinforcing behavior is referred to as a social multiplier, a cascade, a bandwagon effect, imitation, contagion, herd behavior, etc. (4). Our work extends the early theoretical work in three ways. First, we allow for a case where there are common unobserved attributes of the choice alternatives: we revisit a classic approach to statistical prediction in such a situation given an observed sample of decision making agents in a population, the nested logit model. Second, a key feature of our work is that we explicitly consider non-global interactions, with several different social and spatial network structures. Third, additional heterogeneity is introduced in the model through different mechanisms, such as individual-specific sociogeographic characteristics of the commuters as well as individual-specific attributes of the choice alternatives, and the availability of alternatives.

We present an exploratory application of the model to transportation mode choice using pseudo-panel microdata collected in the greater Amsterdam region. Here we combine econometric estimation with computational techniques from the field of multi-agent based simulation. This paper extends previous work by the authors (5) by further studying the role of the utility parameters and connectivity of sociogeographic networks on the emergent outcomes of the multi-agent based simulation. Finally, we conclude highlighting limitations of our present study in any extension for policy considerations on the adoption of innovation in transportation mode choice, and give our suggested recommendations for future work.

LITERATURE

Since the early theoretical work by Aoki (6), Brock and Durlauf (7) and Blume and Durlauf (8), on the long-run behavior of binary discrete choice models with social feedback, there have been a few extensions addressing the complexity of the discrete choice model kernel, the complexity of the feedback effect, and the complexity of the utility specification. A key to the early theoretical results is the assumption that the only explanatory variable in the model is the field effect. In the domain of transportation however, other explanatory variables are assumed to be significant, including individual characteristics of the decision making agents, individual-specific attributes of the alternatives as well as the availability of alternatives for individual agents. In such a case when considering explanatory variables that vary across both the alternatives and the agents, the size of the system of equations to be solved to determine long-term behavior is as large as the number of agents in the system multiplied by the number of choice alternatives minus one, without simplifying assumptions. To address this, a number of researchers have turned to simulation studies whereby the estimation results are embedded in a multi-agent based dynamic model. Dugundji and Gulyas (9) 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 (10) 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. Páez, Scott and Volz (11) 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 (5) 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 current work builds upon the latter in further exploring the relation between utility parameters and connectivity of the interaction network in determining emergent outcomes.

A distinction is hypothesized between social versus spatial interactions and between identifiable versus aggregate interactions (12, 13). We speak of "spatial" network interactions when the interdependence represents a confluence of decision makers in geographic terms. For example, decision makers may be linked based on spatial proximity of residential location, work location or some other geographical point of reference such as school, childcare, shopping, healthcare, leisure/recreation or other relevant activity location. We speak of "social" network interactions when decision makers are linked based on social circles. The decision makers need not be proximally situated in geographical terms and the interaction is not necessarily centered at a particular geographic point of reference; interaction may take place at a distance. We speak of interaction between "identifiable" decision makers when the links in the network are well-known and explicitly defined on an individual decision maker by decision maker basis. We speak of interaction between "aggregate" decision makers when interdependence is assumed to take place only at an aggregate level with links being defined, for example, more generally based on decision maker characteristics.

Important avenues of transportation research in the area of identifiable interactions include coordination of individual daily activity patterns, joint participation in activities and travel, mechanisms for allocation of maintenance activities, and activity location and residential location choice behavior among household members (14-21) as well as research to understand the impact of the explicit structure of loose social networks of extended family, friends and colleagues (22-26). The topic of aggregate social interactions between individuals in different households at a market level in travel demand has only recently begun to attract attention.

SOCIOGEOGRAPHIC NETWORKS

The research reported here explores interactions between a decision maker and the aggregate actions of other decision makers proximally situated in a spatial network, and interactions between a decision maker and the aggregate actions of other decision makers associated in a socioeconomic network. We use a priori beliefs about the social and/or spatial dimension of interactions to formulate the connectivity of the network. Technically, however, interactions between identifiable decision makers may also be modeled using the approach described in this paper given the availability of suitable data, and thus methods reported here may prove to be useful in those areas as well.

In the case study to be discussed, we have rich socioeconomic data for each respondent as well as the geographic location of each respondent's residence and work location. This allows us to define aggregate interactions by grouping agents into geographic neighborhoods or into socioeconomic groups where the influence is assumed to be more likely. In the simplest case, these groups are assumed to be mutually exclusive and collectively exhaustive. That is each agent n belongs to one and only one group. The agent is assumed to be influenced by the average choice behavior of his or her group, and the influence by other groups is assumed to be negligible. At a global level, the picture is a fragmented or disconnected network of clustered groups. If we are interested in equilibrium behavior, the consequences of such an assumption are important: there is no transmission of influence across groups, and the global picture is a weighted average behavior of the separate clusters. Thus we also consider cases with overlapping groups, with agents for example connected by social group as well as by residential district, or by postcode regions of residence and work location. This leads to a giant cluster for the empirical examples under consideration, with the important implication that influence can spread throughout the entire population.

CASE STUDY

The data used in this paper originates from travel questionnaires administered by the Municipality of Amsterdam Agency for Infrastructure, Traffic and Transport, in Amsterdam and a neighboring suburb to the south of the city, Amstelveen. The data set made available by the Agency is a subset of the full modal split database, containing direct home-work trips and direct work-home trips where the purpose of the trip at the non-home location is classified as either "work" or "business." Geographical location is given in terms of the centroid of a traffic analysis zone (TAZ). There are 381 TAZ centroids in Amsterdam and 48 TAZ centroids in Amstelveen, with a total of 933 TAZs in the whole of the Netherlands. The data received includes records of trips where respondents have indicated one of the following transportation mode choices: external system public transit or internal system public transit (23.7% mode share); bicycle or moped/motorcycle (26.7% mode share); car driver or car passenger (49.6% mode share). The final sample used in the case study contains 2913 respondents. Raw variables available for use in the model are availability of public transit, car ownership, gender, income category, education level, age, in-vehicle and out-of-vehicle travel time for public transit, travel time by bicycle, and travel time and parking time for car. Availability of bicycle is generated based on a 75 minute travel time cut-off.

Definition of Interaction Variables

Now we turn to the specification of the network interdependence. We begin with a broad classification by residential district (3, 5). There are 9 districts represented in the sample, ranging in size from 223 sampled respondents to 461 sampled respondents. The mean size is 323 respondents with standard deviation 74, skewness 0.32 and kurtosis 0.19. Next using the three variables age, income and education, 13 socioeconomic groups are defined. The groups range in size from 99 sampled respondents to 385 sample respondents. The mean size is 224 respondents with standard deviation 111, skewness 0.33, and kurtosis -1.8. Finally, to be able to test the effect of spatial scale, we define a smaller neighborhood region of influence on the basis of 4-digit postcode. There are 67 postcode regions represented in the sample, ranging in size from 10 sampled respondents to 161 sampled respondents. The mean size is 43 with standard deviation 32, skewness 2.1 and kurtosis 4.4. We may hypothesize that the smaller spatial scale network interdependence defined by postcode may be more homogeneous with regard to choice behavior than that for the variables defined on the basis of district. Thus we may expect the coefficient on these variables to be relatively stronger. Network interaction variables for four scenarios are presented in Table I. Two scenarios consider clustered groups: Residential district; Social group. Two scenarios consider overlapping groups: District and Social group; Postcode and Social group.

Table I – Definition of Network Variables from Raw Data

Variable	Mean	St Dev	Min	Max
Share of agent's fellow district residents choosing :				
Public transit	0.238	0.062	0.133	0.364
Bicycle/moped/motorcycle	0.268	0.080	0.132	0.409
Car driver/passenger	0.497	0.108	0.307	0.663
Share of agent's social group peers choosing				
Public transit	0.238	0.070	0.145	0.338
Bicycle/moped/motorcycle	0.269	0.071	0.114	0.414
Car driver/passenger	0.498	0.112	0.366	0.728
Share of agent's district residents and social group peers choosing				
Public transit	0.238	0.048	0.138	0.351
Bicycle/moped/motorcycle	0.271	0.057	0.128	0.399
Car driver/passenger	0.493	0.082	0.339	0.664
Share of agent's postcode residents and social group peers choosing				
Public transit	0.237	0.061	0.126	0.387
Bicycle/moped/motorcycle	0.268	0.063	0.100	0.424
Car driver/passenger	0.498	0.097	0.341	0.741

Specification of Utility Functions

A trinary transportation mode choice model to work is estimated using the freely available, open source optimization toolkit Biogeme (27). Various piecewise linear specifications of all travel time related variables as well as age were tested against linear, quadratic and logarithmic forms of these variables. Considering various a priori hypotheses of behavior in the region and after statistical comparison of the alternative nonlinear specifications of variables against the linear versions thereof using loglikelihood ratio tests and non-nested tests (28), a baseline multinomial logit model is estimated. Estimation of three successive nested logit models first with public transit nested with bicycle, then with public transit nested with car, and finally with bicycle nested with car, show the first nesting structure to be most significant in terms of loglikelihood ratio test and in terms of the a t-test on the nest coefficient. The third nesting structure was not indicated. The nested logit model thus adds one additional parameter to the multinomial specification, namely the scale parameter for the transit-bicycle nest. The first two columns of Table II give estimation results for network interdependence defined by residential district and social group, respectively. The last two columns of Table II are treatments where social and spatial interdependence are considered jointly: agents are assumed not to distinguish between their socioeconomic peers' and their fellow district residents or neighbors when considering their choice behavior. Due to space limitations, the specifications extended to allow agents to weigh any influence from their socioeconomic peers differently from any influence from their fellow district or postcode zone neighbors are not considered here.

We conclude from t-tests on the network interaction variables, that for this particular case study and the network definitions under consideration, systematic field effects representing social and spatial network interactions between an agent and the aggregate behavior of other reference agents do indeed have explanatory power. On the basis of non-nested model specification tests in Table III, we find the fit for overlapping postcode and social group is best, as expected. The fit for broad district clusters alone is worst. Interestingly, there is no there is no statistically significant gain in fit at the 0.05 level between the scenario with social group clusters versus the scenario with overlapping district and social group. In light of the latter finding, we will find that the emergent outcomes over time when these models are embedded in a multi-agent based simulation with feedback are particularly noteworthy.

For continuity in the model development process extending the original discrete choice with interactions research by Aoki, Brock, Durlauf and Blume (6-8), a nested logit model is considered in this paper. However, it is worth mentioning that an important econometric issue arises in the empirical estimation of discrete choice models using a nested logit specification in that, while unobserved heterogeneity is accounted for across alternatives, the Gumbel error terms are still assumed to be identically and independently distributed across decision makers. It is not obvious that this is in fact a valid assumption when we are specifically considering interdependence between decision makers' choices. We might reason that if there is a systematic dependence of each decision maker's choice on an explanatory variable that captures the aggregate choices of other decision makers who are in some way related to that decision maker, as we have done, then there might be an analogous dependence in the error structure. Otherwise said, the same unobserved effects might be likely to influence the choice made by a given decision maker as well as the choices made by those in the decision maker's reference group.

The results and coefficients of such a model are likely to be biased (29). Making an analogy of inter-agent causality and correlation with the more well-understood panel data approach towards time causality and correlation (30), Dugundji and Walker (3) present and compare several modeling strategies to highlight some main hypothesized interaction effects using mixed generalized extreme value models with field and "panel" effects. Walker et al (31) re-visit this application and apply a less computationally intensive multi-stage instrumental variables approach developed by Berry (32) and Berry, Levinson and Pakes (33, 34) to control for endogeneity. Other applications addressing endogeneity in discrete choice estimation in the transportation literature are: Train and Winston (35) who also use this same multi-stage approach to correct for price endogeneity in auto ownership choice; Guevara and

Table II – Estimation Results for Nested Logit Models with Different Sociogeographic Networks

Variable	Residential District Clusters	Social Group Clusters	District and Social Group	Postcode and Social Group
Share of agent's network choosing each mode	1.23	1.57	1.93	1.91
	4.85	*5.90*	*5.59*	*6.27*
Alternative specific constant, defined for transit	1.02	-0.464	0.202	-0.443
	2.11	*-1.03*	*0.50*	*-1.05*
Alternative specific constant, defined for car	-0.717	-0.733	-1.11	-0.978
	-1.30	*-1.38*	*-2.14*	*-1.94*
Car ownership, defined for car	2.56	2.51	2.53	2.51
	25.2	*24.6*	*24.8*	*24.6*
Gender, defined for transit	0.288	0.269	0.242	0.249
	3.07	*3.18*	*3.12*	*3.28*
Gender, defined for car	0.260	0.327	0.276	0.310
	2.26	*2.84*	*2.46*	*2.74*
Low income, defined for bicycle	-0.211	-0.196	-0.170	-0.173
	-1.93	*-1.99*	*-1.87*	*-1.93*
Natural logarithm of age, defined for transit	-0.610	-0.155	-0.305	-0.131
	-3.31	*-1.03*	*-2.12*	*-0.97*
Age 45-59, piecewise continuously, for transit	0.0320	0.0170	0.0194	0.0146
	2.48	*1.50*	*1.80*	*1.39*
In-vehicle time, squared, for transit	-3.16e-04	-3.14e-04	-2.90e-04	-2.98e-04
	-3.85	*-3.96*	*-3.68*	*-3.85*
Out-of-vehicle time, for transit	-0.0206	-0.0186	-0.0191	-0.0185
	-3.15	*-3.16*	*-3.26*	*-3.26*
Travel time, for bicycle	-0.0442	-0.0407	-0.0375	-0.0381
	-4.33	*-4.41*	*-4.38*	*-4.70*
Natural logarithm of travel time, for car	-0.654	-0.623	-0.501	-0.545
	-2.36	*-2.39*	*-1.97*	*-2.24*
Parking time, squared, for car	-0.0131	-0.0154	-0.0136	-0.0148
	-7.89	*-9.98*	*-8.35*	*-9.53*
Scale parameter, for transit-bicycle nest	2.07	2.36	2.51	2.51
	2.05	*2.32*	*2.48*	*2.65*
Summary Statistics				
Null log likelihood (L_0)	-2977.3	-2977.3	-2977.3	-2977.3
Final log likelihood	-2060.4	-2054.4	-2055.5	-2049.4
Likelihood ratio test	1833.7	1845.7	1843.5	1855.7
Rho-squared	0.30795	0.30997	0.30960	0.31165
Adjusted rho-squared	0.30291	0.30493	0.30456	0.30662
Final gradient norm	0.0126	0.0194	0.0915	0.0101

All t-statistics (indicated in itallic below the estimated coefficient values) are against 0 except for the scale parameter for the transit-bicycle nest for which it is against 1

Table III –Tests of Model Specifications with Different Sociogeographic Networks

Model 1	Model 2	Df	z = adj.ρ21 - adj.ρ22	x = -2zL0 + df	Φ(-x1/2)	Comments
Social Group	Residential District	0	0.00202	12.04	0.00026	Don't reject model 1
District and Social Group	Residential District	0	0.00165	9.84	0.00085	Don't reject model 1
Postcode and Social Group	Residential District	0	0.00370	22.06	0.00000	Don't reject model 1
Social Group	District and Social Group	0	0.00037	2.20	0.06901	Reject model 1 at 0.05
Postcode and Social Group	Social Group	0	0.00168	10.02	0.00077	Don't reject model 1
Postcode and Social Group	District and Social Group	0	0.00205	12.22	0.00024	Don't reject model 1

The non-nested test bounds the probability of erroneously choosing the incorrect model over the true specification under the null hypothesis that model 1 with higher adjusted rho-squared is the true model
Df : difference in degrees of freedom between model 1 and model 2
Φ : standard normal cumulative distribution function

Ben-Akiva (36) who apply the "control function" approach (37-39) to correct for price endogeneity in residential housing choice; Goetzke (40) and Goetzke and Andrade (41) who account for endogeneity stemming from social network effects in a spatially autoregressive mode choice model using spatial lags as instrumental variables, Goetzke and Weinberger (42) who apply an instrumental variable probit model to test the impact of contextual and endogenous social interaction effects on auto ownership, and Goetzke and Rave (43) who derive an instrument from records with excluded trip purposes to study endogenous effects of "bicycle culture" in German cities.

For the purposes of this paper, we accept that the estimated values may be biased. Our goal here is simply to generate various plausible parameter values under different scenarios, in order to be able to characterize the long-run dynamics of the nested logit model with social feedback. While for an application for policy purposes precise parameter values would be crucially important, in this paper the focus is more abstract. We are interested in getting an idea methodologically under what conditions a runaway effect is generated and what influences this. It is very useful to understand the dynamic behavior of a simple nested logit model, before proceeding to understand the dynamic behavior of models with even more complex kernels. Such an understanding built-up step-by-step is important both theoretically and conceptually as well as for good practice in multi-agent based simulation (9).

MULTI-AGENT BASED SOCIAL SIMULATION

Using the Repast agent-based modeling platform (http://repast.sourceforge.net), we create a computational version of our nested logit models with heterogeneous agents and sociogeographic network interaction. Discrete choice estimation results describing individual heterogeneous preferences are embedded in the multi-agent based model to be able to observe the simulated evolution of choice behavior over time with sociodynamic feedback due to network effects. Example results for different random seeds are shown in Figure 1. Each run is allowed to iterate for 600,000 time steps, or roughly about 200 revisions of choices with asynchronous decision making for the sample size of 2913 agents.

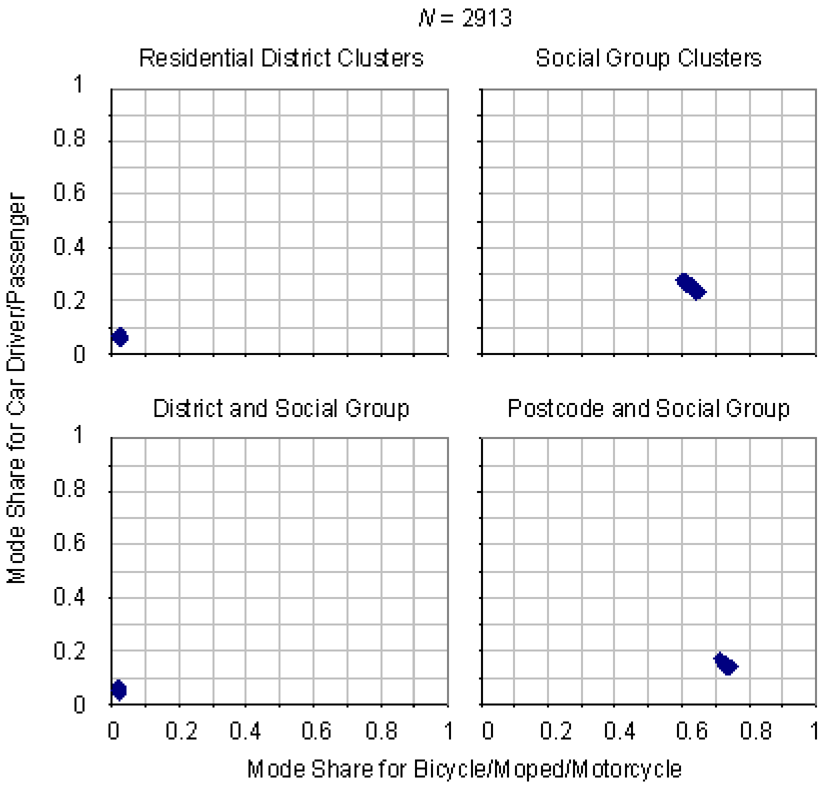

Figure 1 – Observed long-run mode shares for multi-agent simulation of nested logit models with social feedback on different sociogeographic networks

There are several immediately striking features of the long-run results. First, we notice that in all scenarios, the long-run mode shares in Figure 1 moved significantly away from the initial overall modal split (23.7% public transit share; 26.7% bicycle or moped/motorcycle share; 49.6% auto driver or auto passenger share). Second, we notice that the long-run results are also fairly stable: there is little variation in the long-run results for a given scenario. This is true for both scenarios with overlapping groups where influence has the possibility, in principle, to spread through the entire sample. Since it is also true for both scenarios with disconnected clusters where there is no possibility for transmission of influence across groups, this implies that the modal split within the clusters was effectively the same across clusters for a given scenario. Third, we notice in all cases the auto share strongly decreased. This is especially remarkable since the auto mode had initially a share about twice as large as either of the other modes. We see that the feedback effect was thus indeed significant in dynamically hindering the auto mode in the long-run in all scenarios in a well-defined manner.

What is curious is that the feedback effect one hand dynamically propels the transit mode for the case of network interaction by residential district clusters, and by overlapping residential district and social group, and on the other hand dynamically propels the bicycle mode for the case of network interaction by social group clusters, and by overlapping postcode and social group. This is a dramatic difference, emphasizing how important it would be in an application for policy purposes to know in the case of clusters whether influence actually works through neighbors or though socioeconomic peers, or in the case of overlapping groups what the regional scale is of neighborhood influence. We can in any case conclude resoundingly: if a feedback effect can be assumed, the precise details of the connectivity sociographic networks matter!

It is important to recognize however that there are two stages in our process where the sociogeographic network enters. First, the network enters in the econometric estimation in determining the value of the estimated coefficients. Second, the network enters in the multi-agent based simulation in determining the course of the spread of influence when the feedback is strong enough. We may wonder then what is the driving factor of the results: is it simply the strength of the feedback effect relative to the other components of the utility? or is it the connectivity of the network during the transmission process? or both? For example, if a feedback effect can be assumed, in a campaign to promote a particular mode or new service, we would want to know whether to focus efforts on the way the mode is promoted to make the adoption most convincing, or whether to focus for example, on seeding opinion-makers to try to influence the connectivity of the sociogeographic network. Concretely, if say, Twitter were used to get the word out to market a mode, is it the art of sending an enticing enough tweet to generate many re-tweets? or is it the number of followers that receive the tweet and the shape of the network? or both?

To gain some insight to the answer with regard to this particular case study, we run a hypothetical simulation experiment with sociogeographic networks swapped, while holding the utility parameters fixed. Example results for different random seeds are shown in Figure 2. We find that only in the case of the social group parameters did the connectivity of the network seem to have some slight effect on the outcome of the multi-agent simulation. In our particular case study, we conclude that the strength of the feedback effect relative to the other components of the utility is the dominant factor in generating the long-run results. That is, in our particular case study, the connectivity appears not to be very relevant at the transmission stage. This said, it is important to note that the networks studied here are fairly dense by definition, due to the nature of the aggregate interaction assumed within groups. Earlier work by the authors (5) on a simple binary choice model with social interactions on abstract classes of networks over a sweep of network density from a classical case of independent agents on one hand to a fully-connected network on the other hand, holding utility parameters constant, indicated that sparse networks were more sensitive in the outcomes of transmission.

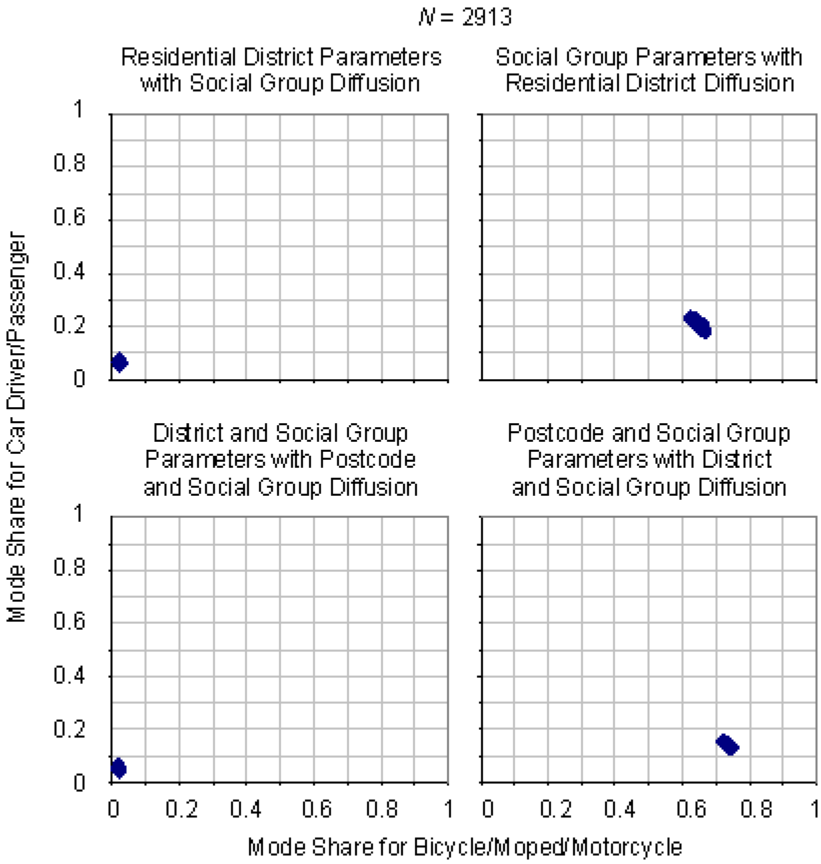

Figure 2 – Observed long-run mode shares for hypothetical experiment with multi-agent simulation of nested logit models where estimated utility parameters and sociogeographic networks are swapped

CONCLUSIONS AND RECOMMENDATIONS

We have extended previous work on discrete choice with social interactions in important ways. We consider a model where an agent's choice is directly influenced by the percentages of the agent's neighbors and socioeconomic peers making each choice, under four different scenarios. Two scenarios depict influence within a disconnected network of clustered groups. Two scenarios depict influence within overlapping social and spatial groups. Given the availability of appropriate data, our approach is principle directly extendable to the identifiable agent case. We observe that the estimated utility parameters for different hypothetical sociogeographic network scenarios can generate dramatically different dynamics. This finding underscores the need for more empirical research to understand actual sociogeographic influence networks (44-49).

A challenging direction of on-going work by the authors addresses evolving networks in coupling with the evolving behavioral dynamics (50). A motivation for this direction of work is to be able to account for residential mobility, occupational mobility and other life cycle changes in social-spatial networks impacting transportation mode choice (51). An important distinction can be namely understood in the land use transportation planning problem domain between network interactions impacting choices, such as transport mode choice, which do not endogenously affect the decision maker's reference position in the network (eg. whether an agent chooses to travel by car versus rail for an intercity trip will not affect the fact of who the agent's neighbors are), as opposed to network interactions impacting "sorting" type choices, such as residential location choice, which do indeed endogenously affect the decision maker's reference position in a spatial network and potentially also within a social network (eg. in moving to a new neighborhood an agent per definition acquires new neighbors). The econometric aspects of estimating utility parameters for a residential location choice model with social and spatial network interactions are non-trivial and data intensive (52).

Putting these desirable data and modeling features together for policy purposes, a challenging set of statistical questions arises for the econometric estimation with regard to the sampling frame for data collection. Various extensions of the maximum likelihood procedure for discrete choice models have been developed for estimation with general stratified samples (eg. exogenous samples, where sampling strata are segmented by the decision maker characteristics and/or attributes of alternatives, and choice-based samples, where each choice alternative corresponds to a separate stratum), enriched samples (eg. pooling of exogenously stratified samples with one or more choice-based samples), and double or multi-stage samples (eg. carrying first a small survey and then using information obtained to design a second survey). See for example, Manksi and Lerman (53), Manski and McFadden (54), Cosslett (55) and Daganzo (56, 57) for early work. An intriguing direction for further research, when collecting data on data on social networks using a technique such as snowball sampling (45), is what modifications may be necessary in the estimation procedure for the utility parameters of (complex) discrete choice models capable of capturing endogenous effects, contextual effects and correlated effects, and what the formal properties of estimates are under such a sampling scheme where the selection of decision makers is inherently interdependent by design and the choice behavior, characteristics, choice attributes and links are followed over time.

ACKNOWLEDGMENTS

This work was supported in part by the Netherlands Organisation for Scientific Research, for a generous grant for the AMADEUS program directed by Harry Timmermans and coordinated by Theo Arentze. The authors would also like to gratefully acknowledge discussion with Frank le Clercq, Loek Kapoen, Joan Walker, George Kampis, József Váncza and András Márkus at various stages in the development of the research. The authors claim full responsibility for any errors. Special thanks are also due to Guus Brohm, Yvon Weening, Nelly Kalfs and Hans Niepoth at the Agency for Infrastructure, Traffic and Transport of the Municipality of Amsterdam for making data available for the empirical application, and to Willem Vermin, Bert van Corler, Lilit Axner, John Donners, and the High Performance Computing support team at SARA Computing and Networking Services, Science Park Amsterdam.

REFERENCES

1. Dugundji, E. R., A. Páez and T. A. Arentze (2008). Social networks, choices, mobility, and travel. Environment and Planning B: Planning and Design, 35: 6, 956-960.
2. Dugundji, E. R., A. Páez, T. A. Arentze, J. L. Walker, J. A. Carrasco, F. Marchal, and H. Nakanishi (2011). Transportation and social interactions. Transportation Research Part A, 45: 4, 239-247.
3. Dugundji, E. R. and J. L. Walker (2005). Discrete Choice with Social and Spatial Network Interdependencies: An Empirical Example Using Mixed GEV Models with Field and Panel Effects. Transportation Research Record: Journal of the Transportation Research Board, 1921, 70-78.
4. Manksi, C. F. (1995). Identification Problems in the Social Sciences. Harvard University Press, Cambridge, MA.
5. Dugundji, E. R. and L. Gulyás (2008). Socio-Dynamic Discrete Choice on Networks: Impacts of Agent Heterogeneity on Emergent Equilibrium Outcomes. Environment and Planning B: Planning and Design, 35: 6, 1028-1054.
6. Aoki, M. (1995). Economic fluctuations with interactive agents: Dynamic and stochastic externalities. Japanese Economic Review, 46: 2, 148-165.
7. Brock, W. A. and S. N. Durlauf (2001). Discrete Choice with Social Interactions. Review of Economic Studies, 68, 235-260.
8. Blume, L. and S. N. Durlauf (2003). Equilibrium Concepts for Social Interaction Models. International Game Theory Review, 5: 3, 193-209.
9. Dugundji, E. R. and L. Gulyás (2005). Socio-dynamic discrete choice in an empirical example on intercity travel demand: An agent-based approach, analytical benchmark and some issues in estimation. Presented at Workshop on Modeling Urban Social Dynamics, University of Surrey, Guildford, UK.
10. 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.
11. 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: Planning and Design, 35: 6, 1055-1069.
12. Dugundji, E. R. and L. Gulyás (2003). Empirical Estimation and Multi-Agent Based Simulation of a Discrete Choice Model with Network Interaction Effects. Presented at 8th International Conference on Computers in Urban Planning and Urban Management, Sendai, Japan.
13. Dugundji, E. R. and L. Gulyás (2003). An Exploration of the Role of Global versus Local and Social versus Spatial Networks in Transportation Mode Choice Behavior in The Netherlands. Presented at AGENT 2003: Challenges in Social Simulation, Argonne National Laboratory, Chicago, IL.
14. Gliebe, J. P. and F. S. Koppelman (2002). A model of joint activity participation between household members. Transportation, 29, 49-72.
15. Gliebe, J. P. and F. S. Koppelman (2005). Modeling household activity-travel interactions as parallel constrained choices. Transportation, 32, 449-471.

16. Scott, D. M. and P. S. Kanaroglou (2002). An activity-episode generation model that captures interactions between household heads: development and empirical analysis. Transportation Research Part B, 36, 875-896.
17. Bradley, M. and P. Vovsha (2005). A model for joint choice of daily activity pattern types of household members. Transportation, 32, 545-571.
18. 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.
19. Timmermans, H. J. P. and J. Zhang (2009). Modeling household activity travel behavior. Transportation Research Part B, 43, 187-190.
20. Arentze, T. A. and H. J. P. Timmermans. A need-based model of multi-day, multi-person activity generation. Transportation Research Part B, 43, 2009, 251-265.
21. Kang, H. and D. M. Scott (2010). Exploring day-to-day variability in time use for household members. Transportation Research Part A, 44: 8, 609-619.
22. Carrasco, J. A. and E. J. Miller (2006). Exploring the propensity to perform social activities: A social network approach. Transportation, 33: 5, 463-480.
23. 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: Planning and Design, 35: 6, 961-980.
24. Axhausen, K. W. (2008). Social networks, mobility biographies, and travel: Survey challenges. Environment and Planning B: Planning and Design, 35: 6, 981-996.
25. Schwanen, T. (2008). Managing uncertain arrival times through sociomaterial associations. Environment and Planning B: Planning and Design, 35: 6, 997-1011.
26. Lovejoy, K. and S. Handy (2011). Social networks as a source of private-vehicle transportation: The practice of getting rides and borrowing vehicles among Mexican immigrants in California. Transportation Research Part A, 45: 4, 248-257.
27. Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. Presented at 3rd Swiss Transportation Research Conference, Ascona, Switzerland.
28. Ben-Akiva, M. and S. R. Lerman (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA, chapter 7.
29. Train, K. E. (2009). Discrete Choice Methods with Simulation, 2nd edition. Cambridge University Press, New York, chapter 13.
30. Heckman, J. J. (1981). The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process. In: Structural Analysis of Discrete Data with Econometric Applications (C. F. Manski and D. McFadden, eds.). MIT Press, Cambridge, MA.
31. Walker, J. L., E. Ehlers, I. Banerjee, and E. R. Dugundji (2011). Correcting for endogeneity in behavioral choice models with social influence variables. Transportation Research Part A, 45: 4, , 362-374.
32. Berry, S. (1994). Estimating discrete choice models of product differentiation. RAND Journal of Economics, 25, 242-262.
33. Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. Econometrica, 63, 841-889.

34. S. Berry, 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.
35. 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.
36. Guevara, C. A. and M. Ben-Akiva (2006). Endogeneity in residential location choice models. Transportation Reseach Record: Journal of the Transportation Research Board, 1977, 60-66.
37. Hausman, J. (1978). Specification tests in econometrics. Econometrica, 46: 6, 1251-1272.
38. Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. Econometrica, 46, 931-959.
39. Heckman, J. and R. Robb (1985). Alternative methods for evaluating the impacts of interventions: An overview. Journal of Econometrics, 30, 239-267.
40. Goetzke, F. (2008). Network effects in public transit use: evidence from a spatially autoregressive mode choice model. Urban Studies, 45, 407-417.
41. Goetzke, F. and P. M. Andrade (2010). Walkability as a summary measure in a spatially autoregressive mode choice model: An instrumental variable approach. In: Progress in Spatial analysis: Methods and Applications (A. Páez, R. N. Buliung, J. Le Gallo and S. Dall'erba, eds.), Advances in Spatial Science. Springer-Verlag, Berlin, Heidelberg.
42. Goetzke, F. and R. Weinberger (2012). Separating contextual from endogenous effects in automobile ownership models. Environment and Planning A, 44, 1032-1046.
43. Goetzke, F. and T. Rave (2011). Bicycle use in Germany: Explaining differences between municipalities with social network effects. Urban Studies, 48: 2, , 427-437.
44. Axsen, J., and K. S. Kurani (2012). Interpersonal influence within car buyers' social networks: Applying five perspectives to plug-in hybrid vehicle drivers. Environment and Planning A, 44, 1047-1065.
45. Kowald, M. and K. W. Axhausen (2012). Focusing on connected personal leisure networks: Selected results from a snowball sample. Environment and Planning A, 44, 1085-1100.
46. Arentze, T. A., P. van den Berg, and H. J. P. Timmermans (2012). Modeling social networks in geographic space: Approach and empirical application. Environment and Planning A, 44, 1085-1100.
47. Dam, I., D. M. Scott, A. Páez, and R. Wilton (2012). Investigating the effects of social influence on the choice to telework. Environment and Planning A, 44, 1016-1031.
48. Carrasco, J. A. (2012). Social capital, social networks, and travel: An empirical illustration. Environment and Planning A, 44, 1066-1084.
49. Hackney, J. and M. Kowald (2011). Exponential random graph models of the Zurich snowball survey. Presented at FUTURENET Workshop: Social Network Analysis in Transport, Manchester, UK.
50. Gulyás, L. and E. R. Dugundji (2006). Emergent opinion dynamics on endogenous networks. In: Interaction and Emergent Phenomena in Societies of Agents (G. P.

Trajkovski and S. G. Collins, eds.), Technical Report FS-06-05. AAAI Press, Menlo Park, CA.

51. Dugundji, E. R., L. L. Kapoen, F. le Clercq, T. A. Arentze, H. J. P. Timmermans, and K. J. Veldhuisen (2001). The Long-term Effects of Multi-modal Transportation Networks: The Residential Choice Behavior of Households. Presented at 9th World Conference on Transportation Research, Seoul, Korea.
52. Dugundji, E. R. (2006). Residential choice and the geography of family networks: Some considerations. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C.
53. Manski, C. F. and S. R. Lerman (1977). The estimation of choice probabilities from choice-based samples. Econometrica, 45, 1977-1988.
54. Manski, C. F. and D. McFadden (1981). Alternative estimators and sample designs for discrete choice analysis. In: Structural Analysis of Discrete Data with Econometric Applications (C. F. Manski and D. McFadden, eds.). MIT Press, Cambridge, MA.
55. Cosslett, S. (1981). Efficient estimation of discrete choice models. In: Structural Analysis of Discrete Data with Econometric Applications (C. F. Manski and D. McFadden, eds.). MIT Press, Cambridge, MA.
56. Daganzo, C. F. (1980). Optimal sampling strategies for statistical models with discrete dependent variables. Transportation Science, 14: 4, 324-345.
57. Daganzo, C. F. (1982). Errata. Transportation Science, 16: 1, 107.