Socio-dynamic discrete choice: Theory and application

Dugundji, E.R.

8 ECONOMETRIC CONSIDERATIONS

An econometric issue that arises in empirical estimation of models with network effects is that 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. To try to separate out effects, it is critically important to begin with an as well-specified model as possible, making use of relevant available explanatory variables as we have done in sections 7.4 and 7.7. In sections 8.1, 8.2 and 8.3 in this chapter we continue this exploration of issues in the empirical estimation of discrete choice models with field effects, by specifically testing for correlation among agents in the error structure in the particular empirical case study of mode choice to work from Chapter 7, through the use of mixed GEV family models. It is namely the intent to capture potential bias induced by endogeneity, by explicitly accounting for the supposed correlation in the error structure. Important however is that we want to avoid confounding correlation among alternatives with any potential correlation among agents. We thus use the GEV structure in the choice model kernel to capture correlation between alternatives, and use mixing to capture correlation between decision-makers. For completeness, we also compare field effect models to baseline models where limited social and spatial interdependence can be introduced as alternative-specific dummy variables. The latter modeling approach in turn can be used to move suspected endogeneity out of the non-linear choice model to a linear regression where endogeneity is easier to manage as reviewed in section 8.4. In section 8.5 we place our research into context by reviewing alternative approaches to modeling dependence between observational units from spatial and social econometrics, including important recent advances since the time of initially carrying out this research. As far as we are aware, the research presented in sections 8.1, 8.2 and 8.3 is the *first* example of the empirical estimation of a discrete choice model among multiple choice alterantives with social and/or spatial state dependence, where correlation among choice alternatives as well as correlation among decision-making agents in the error structure are also both tested.

8.1 BASELINE MODEL WITH NO INTER-AGENT DEPENDENCIES

To understand how the introduction of social and spatial network interdependencies affects the random utility model formulations, thereby prompting attention to the treatment of correlation between *decision-makers*, it is useful to first briefly review a basic model where flexible specification of correlation between discrete choice *alternatives* is addressed. The treatment of correlation between choice alternatives is namely a well-developed field that has attracted attention already many years: in the case of nested logit, dating back to Ben-Akiva (1973); in the case of binary probit, arguably dating back to Thurstone (1927). For details on treatment of correlation between alternatives, the interested reader is referred to the following excellent reference works|: Ben-Akiva and Lerman (1985), Train (2009), and Ben-Akiva and Bierlaire (1999). Although we choose the cross-nested logit model (McFadden, 1978; Small, 1987) as a starting point here owing to its fairly general closed form, the strategies presented in this section for introducing social and spatial network interdependencies could in principle be applied to any suitable choice model that sufficiently captures correlation between alternatives.

To put the methodological development into perspective, in building step-by-step from the theoretical work of Aoki (1995), Brock and Durlauf (2001a, 2002, 2006) and Blume and Durlauf (2003), first let us recall the presentation of the multinomial logit model in Chapter 2 at the outset of this dissertation. We let the utility U_{in} that a given decision-making agent n is presumed to associate with a particular elemental alternative i in his choice set C_n, be given by:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{8.1}$$

where V_{in} is the deterministic (to the modeler) or so-called "systematic" utility and ε_{in} is an error term. In the random utility maximization model, the utilities are not known to the modeler with certainty and are therefore treated as random variables (Manski 1977). The individual choice probability $P_n(i|C_n)$ that the individual decision-making entity n chooses alternative i within the choice set C_n is equal to the probability that the utility U_{in} of alternative i for individual n is greater than or equal to the utilities of all other alternatives in that individual's choice set:

$$\begin{aligned} P_{in} &\equiv P_n(i|C_n) = \Pr[U_{in} \geqslant U_{jn}, \forall j \in C_n] \\ &\Pr\left(V_{in} + \varepsilon_{in} \geqslant V_{jn} + \varepsilon_{jn}, \forall j \in C_n\right) \end{aligned} \tag{8.2}$$

By making assumptions about the joint probability distribution of the full set of disturbances ε_{in}, we can derive specific forms of the random utility model. As shown in Ben-Akiva and Lerman (1985), under the assumption of independent and identically Gumbel (type I extreme value) distributed disturbances ε_{in} with strictly positive scale parameter $\mu > 0$, the individual choice probability $P_n(i|C_n)$

that agent n chooses alternative i within nest C_n has a convenient closed form expression, given by:

$$\begin{aligned} P_n(i|C_n) &= \Pr\left(V_{in} + \varepsilon_{in} \geqslant V_{jn} + \varepsilon_{jn}, \forall j \in C_n\right) \\ &= \Pr\left[V_{in} + \varepsilon_{in} \geqslant \max_{j \in C_n}\left(V_{jn} + \varepsilon_{jn}\right)\right] \\ &= \frac{e^{\mu V_{in}}}{\sum\limits_{j \in C_n} e^{\mu V_{jn}}} \end{aligned} \tag{8.3}$$

The derivation makes convenient use of the special property that if ε_{jn} are J_n independent Gumbel distributed random variables with common scale parameter μ, then $\max \varepsilon_{jn}$ is *also* Gumbel distributed with that *same* scale parameter μ. Another special property which is important for arriving at the closed form expression, is that the difference of two Gumbel distributed random variables with a common scale parameter is *logistically* distributed. It is this latter fact from which stems the name "logit," or logistic probability unit model. When there are more than two choice alternatives, this model is called multinomial logit (MNL).

Train (2009) explains clearly the significance of the multinomial logit as well as, at the same time, its shortcoming:

> "The standard logit model exhibits independence from irrelevant alternatives (IIA), which implies proportional substitution across alternatives. ... this property can be seen either as a restriction imposed by the model or as the natural outcome of a well specified model that captures all sources of correlation over alternatives into representative utility, so that only white noise remains. Often the researcher is unable to capture all sources of correlation explicitly, so that the unobserved portions of utility are correlated and IIA does not hold. In these cases, a more general model than standard logit is needed.
>
> Generalized extreme value (GEV) models constitute a large class of models that exhibit a variety of substitution patterns. The unifying attribute of these models is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value. This distribution allows for correlations over alternatives and, as its name implies, is a generalization of the univariate extreme value distribution that is used for standard logit models."

One member of the GEV family with a fairly general closed form expression for the individual choice probability that captures correlation among alternatives is the cross-nested logit model (CNL). In accordance with the formulation in Ben-Akiva and Bierlaire

(1999), we assume a population of N decision-making entities indexed $(1, \ldots, n, \ldots, N)$. Each agent is faced with a single choice among elemental alternatives i in the composite choice set $Cn = (C_{1n}, \ldots, C_{mn}, \ldots, C_{Mn})$ of some universal choice set C of alternatives $(1, \ldots, i, \ldots, J)$. In general the composite choice set C_n will vary in size and content across agents: not all elemental alternatives i in the universal choice set may be available to all agents. The overall correlation structure of alternatives is however assumed to be the same across agents, aside from availability. The indexing $(1, \ldots, m, \ldots, M)$ represent so-called "nests" of elemental alternatives which are assumed to be correlated. The CNL model relaxes the nested logit model described earlier, such that the grouping of elemental alternatives into nests need not be mutually exclusive. A set of parameters α_{im} defined to have value between 0 to 1 inclusive, specifies the degree of "membership" of alternative i in nest m.

Let $U_{imn} = V_{in} + \varepsilon_{in} + V_{mn} + \varepsilon_{mn} + \ln \alpha_{im}$ be the utility that a given agent n is presumed to associate with a particular alternative i in choice nest C_{mn}, where V_{in} is the systematic utility associated with alternative i given choice nest C_{mn}, similarly V_{mn} is the systematic utility associated with the particular choice nest C_{mn} itself, and ε_{in} and ε_{mn} are the respective error terms independent for all $i \in C_{mn}$ and all $C_{mn} \in C_n$. Under the assumption of independent and identically Gumbel distributed disturbances ε_{in} with strictly positive scale parameter μ_m, and under the assumption of the distribution of the disturbances ε_{mn} such that the random variable $\max_{i \in C_{mn}} U_{imn}$ is Gumbel distributed with scale parameter μ, then the probability $P_n(i)$ that agent n chooses alternative i from the composite choice set C_n is simply the sum of the joint probabilities $P_n(i, C_{mn})$ across all nests:

$$P_n(i) = \sum_{m'=1}^{M} P_n(i, C_{mn}) = \sum_{m'=1}^{M} P_n(i|C_{mn}) P_n(C_{mn}) \tag{8.4}$$

The conditional probability $P_n(i|C_{mn})$ that agent n chooses alternative i given choice nest C_{mn}, the so-called "log-sum" I_{mn}, and the probability $P_n(C_{mn})$ that agent n chooses the particular choice nest C_{mn} itself, are:

$$P_n(i|C_{mn}) = \frac{\alpha_{im} e^{\mu_m V_{in}}}{\sum\limits_{j \in C_{mn}} \alpha_{jm} e^{\mu_m V_{jn}}} \tag{8.5}$$

$$I_{mn} = \frac{1}{\mu_m} \ln \sum_{j \in C_{mn}} \alpha_{jm} e^{\mu_m V_{jn}} \tag{8.6}$$

$$P_n(C_{mn}) = \frac{e^{\mu V_{mn} + \mu I_{mn}}}{\sum\limits_{m'=1}^{M} e^{\mu V_{m'n} + \mu I_{m'n}}} \tag{8.7}$$

Note that the CNL model will reduce to the nested logit model if the set of parameters α_{im} is such that the grouping of elemental alternatives into nests is mutually exclusive. This will further reduce to the multinomial logit model if the scale μ_m for the lower nests is equal to the scale μ for the upper nest. We focus on the CNL model here to make sure we have a general form that will capture correlation among alternatives. We want to avoid confounding correlation among alternatives with any potential correlation among agents.

8.2 SOCIAL-SPATIAL INTERDEPENDENCIES IN CHOICE MODELS

An important econometric issue arises in empirical estimation of field effects in discrete choice models in that the error terms may not be identically and independently distributed across decision-makers. Namely, when we are specifically considering interdependence between individuals' choices, we might expect that if there is a systematic dependence of each decision-maker's choice on a field effect, ie. an explanatory variable p_{ig_n} that captures the aggregate choices of others who are in some way related to that decision-maker, then there might be an analogous dependence in the error structure. This can be expressed formally in the utility function through the allowance for a group-specific error term ξ_{ig_n} for alternative i for each individual's reference group g_n. In the special case that the groups are shared by multiple decision-makers (ie. each member of the reference group faces the same reference members), it is also possible to estimate a so-called "fixed" effect, an alternative specific constant for each reference group. Finally, there may be alternative-specific random taste variation on the parameter for the field effect, and dependence in the error structure can also be manifested in the portion ζ_{ig_n} of an alternative-specific deviation defining the difference between individual n's parameters and the average for the population, that is specific to each individual's reference group g_n in the taste variation on the parameter for the field effect. In this section we thus present these five strategies for introducing social and spatial network interdependencies in choice models: as a generic field effect (subsection 8.2.1), as unobserved group heterogeneity (subsection 8.2.2), as a set of alternative-specific dummy variables (subsection 8.2.3), and as a random parameter field effect, with unobserved individual heterogeneity and unobserved group heterogeneity (subsection 8.2.4).

8.2.1 *Strategy 1: Field Effect for Reference Group*

We introduce a feedback effect among agents by allowing the systematic utility V_{in} to be a linear-in-parameter β first order function of the proportion p_{in} of a given decision-maker's reference entities who have made this choice. The variable p_{in} is termed a "field variable." This model differs from the field effects model in Part II, however, in that we consider non-global interactions. We do consider an aggregated feedback, but agents see different proportions, depending on who their particular reference entities are. In fact, even agents within the same group see (slightly) different proportions of reference agents choosing each alternative because an agent's own choice is not included in these proportions. If we had data available to support treatment of network interaction effects with identifiable households, the approach could be further generalized for example to allow different weights on different ties (see section 8.5).

8.2.2 *Strategy 2: Unobserved Group Heterogeneity ("Panel" Effect)*

Next we consider correlation among agents in the disturbance of the utility. We model the correlation among agents within a group g as a panel effect. In a typical choice model application with panel data, the panel aspect captures correlation across multiple responses from the same individual. The panel implementation that we use captures correlation among different decision-makers based on the connectivity in a social and/or spatial network. By casting the research question in this way, it allows us to draw upon a wealth of literature spawned by Heckman (1981a, 1981b) on understanding the properties of choice models in the presence of both state dependence and serial correlation. Let $\xi_g = [\xi_{1g}, \ldots, \xi_{ig}, \ldots, \xi_{Jg}]^T$ be a vector of unobserved random terms where the ξ_{ig} are now additively introduced to the utility function U_{in}. The probability $P_n(i)$ that agent n chooses alternative i from the composite choice set C_n is the integral over $P_n(i|\xi_g)f(\xi_g)$ with respect to ξ_g where $f(\xi_g)$ denotes the multivariate probability density of ξ_g. It is namely our intent to capture potential bias induced by endogeneity, by explicitly accounting for the supposed correlation in the error structure. Where these error terms can be shown to be statistically insignificant, we assume that the hypothesized endogeneity has negligible effect.

Bhat (2000) proposed a related use of panel effects to study unobserved spatial heterogeneity from multiple reference groups in discrete choice models, without studying field effects. He considered a specification of the group-specific error term for alternative i of the form

$$\xi_{ig_n} \equiv \xi_{ih} + \xi_{iw} \tag{8.8}$$

where ξ_{ih} and ξ_{iw} are normally distributed random terms that capture unobserved variations across home-zones (h) and work-zones (w), respectively. Bhat's work forms the first example in travel demand literature of the use of Halton sequences for simulated estimation draws.

8.2.3 *Strategy 3: Alternative-Specific Dummy Variable for Reference Group*

For completeness, we also consider a more typical model specification with a complete set of alternative-specific dummy variables (ASDV) by reference group. This would be a case of observed group heterogeneity. In general we can expect a complete set of ASDV to perform well in terms of the loglikelihood value, in being very flexible in capturing variation across alternatives and across groups. The difficulty with adding a complete set of ASDV is the proliferation of estimated parameters if there are a large number of groups and/or large number of alternatives. In such case we might run considerable risk of overfitting our model.

8.2.4 *Strategies 4 and 5: Random Field Effect and Panel Field Effect*

Finally, we allow the field effect described in strategy #1 to have alternative-specific variance associated with it. In strategy #4 we model this variance most generally as unobserved individual heterogeneity (random parameter field effect). In strategy #5 we model this variance as unobserved group heterogeneity: the variance on the field effect is constrained to be equal for agents in the same reference group ("panel" field effect). The distinction between this last model versus strategies #1 and #2 estimated directly together is that here the unobserved group heterogeneity is on the field effect itself, instead of being estimated on its own. Strategies #4 and #5 thus gain flexibility in capturing variation across alternatives and across individuals/groups as with the complete set of ASDV in strategy #3, but without the drawback of a potential proliferation of estimated parameters with large number of groups and/or alternatives.

8.2.5 *Notation*

We formalize these strategies as follows. For notational convenience we define q_{in} to be a vector-valued function of length K_i of observable characteristics of the decision-making agent and observable attributes of the choice alternative for a given decision-making agent whereby polynomial, exponential, logarithmic, piecewise linear, and other real transformations of the agent characteristics and choice attributes may be applied as deemed appropriate by the modeler. In a classical discrete choice model, the systematic utility V_{in} is assumed to be defined by a linear-in-parameters function of the variables q_{in}:

$$V_{in} = V(\mathbf{q}_{in}) = \vartheta_i{}'\mathbf{q}_{in} = \vartheta_1 q_{in1} + \ldots + \vartheta_{K_i} q_{inK_i} \tag{8.9}$$

where $\vartheta_i = [\vartheta_1, \ldots, \vartheta_{K_i}]'$ is a vector of K_i unknown alternative-specific parameters, and the length of the vector is allowed to vary across alternatives. By defining one of the variables q_{ink} to be a dummy variable h_i for alternative i, an alternative specific constant can be included to explicitly account for any underlying bias for one alternative over another alternative. This is typically done when the number of choice alternatives J in the universal choice set is not too large, and there is no concern about implications of proliferation of parameters for model overfit in subsequent use of the model for prediction purposes.

As in chapter 7, we let p_{ig_n} represent the (hypothesized) endogenous perceptions by individual n of the choice behaviors for alternative i of others in his reference group g_n. Such perceptions are often not available to the modeler as observables from survey data; in diverse literature these are typically expressed as the expectations by individual n of his group members' (weighted) average behavior or his group members' weighted average utility. (We discuss the consequences of these different modeling approaches later in section 8.5.)

The different modeling strategies #1 to #5 may be expressed in general as follows

$$U_{in} = \gamma_{in} p_{ig_n} + \vartheta_i{}' \mathbf{q}_{in} + \xi_{ig_n} + \eta_{in} \tag{8.10}$$

where: γ_{in} is an unknown alternative-specific random parameter; ξ_{ig_n} is a group-specific error term for alternative i for each individual's reference group g_n; η_{in} is an individual-specific error term for alternative i for individual n.

In general, the parameter vector $\vartheta_i = [\vartheta_1, \ldots, \vartheta_{K_i}]'$ may also be expressed with taste variation. However we will be specifically interested in taste variation on the parameter γ_{in} for the endogenous social effect, which we can express as

$$\gamma_{in} = \beta_i + \zeta_{ig_n} + \psi_{in} \tag{8.11}$$

where:β_i is an unknown alternative-specific scalar parameter; ζ_{ig_n} is the portion of an alternative-specific deviation defining the difference between individual n's parameters and the average for the population, that is specific to each individual's reference group g_n; ψ_{in} is the portion of an alternative-specific deviation defining the difference between individual n's parameters and the average for the population, that varies across all decision-makers.

8.3 RESULTS AND DISCUSSION

All model strategies are estimated using the freely available optimization toolkit Biogeme (`http://biogeme.epfl.ch`) developed by Michel Bierlaire. 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 two nesting structure to be significant in terms of loglikelihood ratio test and in terms of a t-test on the nest coefficient. The third nesting structure was not indicated.

Thus a cross-nested structure with public transit nested with both bicycle and car is a logical choice. This cross-nesting structure was tested against a general structure where all 3 alternatives could belong to each of 2 nests, and showed no significant loss of fit. Furthermore, the nest coefficient on the public transit and car nest and the nest coefficient on the public transit and bicycle nest are constrained to be equal. Finally constraining the two parameters $\alpha_{pt,m}$ defining the degree of "membership" of the public transit alternative in each of the two nests m to be equal also showed no significant loss of fit, and improvement in the t-test on $\alpha_{pt,m}$. The baseline cross-nested model thus adds two additional parameters, namely a generic nest coefficient and a generic membership in the nests.

Example estimation results for the case of network interdependence defined by socioeconomic group are given in Table 8.1.

All model strategies are shown, with the exception of the complete set of ASDV that is omitted out of space limitations for its 39 estimated parameters. The loglikelihood, number of estimated parameters and adjusted rho-squared are given in Table 8.2 for all model strategies, and various reference group treatments. The summary results from Table 8.1 re-appear in the first column of Table 8.2. The second and third columns of Table 8.2 give summary results for network interdependence defined by residential district and residential postcode, respectively. The fourth and the fifth columns 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. In the remaining sixth through ninth columns of Table 8.2, the specification is extended to allow agents to weigh any influence from their socioeconomic peers differently from any influence from their fellow district residents and furthermore differently from any influence from their more immediate neighbors.

All estimations shown in Table 8.2 that involve mixing (model strategies #2, #1+2, #4, #5) have 500 Halton draws. Two parallel sets of estimations carried out first with 250 Halton draws and then afterwards with an alternative optimization algorithm, all yield coefficients with magnitudes within 2 standard deviations of the reported estimations. All sigmas for the random parameters and "panel" effects were started at 1.0 to be sure that any move towards zero is indeed a real effect and not an artifact of simulation error.

Specification tests are given in Table 8.3 and Table 8.4. We conclude from loglikelihood ratio tests on the field effect model versus the baseline model, that for this particular case study and the network definitions under consideration, systematic field effects representing social

Table 8.1: Example estimation results for various model strategies with network interdependence defined by socioeconomic peer group.

MODEL STRATEGY:			#1	#2	#1+#2	#4	#5
	Baseline	Baseline	Field Effect	Panel Effect	Field+ Panel	Random Field	Panel Field
ESTIMATED PARAMETERS	MNL	CNL	CNL	CNLP	CNLP	MNCL	MNCLP
Share of agent's socioeconomic			1.39e+0		1.12e+0	2.10e+0	1.06e+0
peer group choosing each mode			5.20		3.44	4.21	3.29
Alternative specific constant for transit	9.33e-1	6.30e-1	-8.49e-1	4.40e-1	-8.88e-1	-9.20e-1	-8.43e-1
	1.18	1.10	-1.39	0.68	-1.24	-1.34	-1.18
Alternative specific constant for car	9.61e-1	3.62e-1	-2.59e-2	2.95e-1	1.27e-2	-5.16e-1	4.23e-2
	2.08	0.90	-0.07	0.73	0.03	-0.88	0.11
Car ownership, defined for car	2.60e+0	2.19e+0	2.09e+0	2.17e+0	2.12e+0	2.92e+0	2.11e+0
	25.30	17.10	16.70	17.04	16.67	6.03	16.70
Gender, defined for transit	5.61e-1	4.05e-1	4.12e-1	4.45e-1	4.23e-1	4.42e-1	4.25e-1
	4.67	4.21	4.41	4.53	4.43	4.37	4.46
Gender, defined for car	4.02e-1	3.17e-1	3.93e-1	3.62e-1	3.91e-1	4.48e-1	3.91e-1
	3.31	2.99	3.69	3.35	3.63	3.28	3.64
Low income, defined for bicycle	-4.77e-1	-3.32e-1	-3.39e-1	-2.95e-1	-3.02e-1	-3.25e-1	-3.02e-1
	-2.88	-2.46	-2.57	-2.17	-2.26	-2.34	-2.26

ESTIMATED PARAMETERS, CONT'D	MNL	CNL	#1	#2	#1+2	#4	#5
Natural log of age, defined for transit	-9.94e-1	-7.37e-1	-2.85e-1	-6.53e-1	-2.71e-1	-2.70e-1	-2.73e-1
	-4.42	-4.32	-1.60	-3.44	-1.32	-1.33	-1.35
Age 45 to 59, piecewise, for transit	4.68E-02	3.47e-2	2.60e-2	2.10e-2	1.89e-2	2.94e-2	1.73e-2
	2.44	2.47	1.92	1.36	1.27	1.88	1.17
In-vehicle time for transit, squared	-4.08e-4	-3.18e-4	-3.09e-4	-3.03e-4	-3.03e-4	-3.89e-4	-3.02e-4
	-4.59	-4.34	-4.34	-4.20	-4.25	-3.84	-4.24
Out-of-vehicle time for transit	-2.20e-2	-1.65e-2	-1.71e-2	-1.77e-2	-1.77e-2	-2.25e-2	-1.76e-2
	-2.53	-2.48	-2.65	-2.68	-2.72	-2.71	-2.72
Travel time for bicycle	-8.30e-2	-6.30e-2	-6.08e-2	-6.26e-2	-6.17e-2	-6.39e-2	-6.14e-2
	-15.50	-10.30	-10.20	-10.34	-10.27	-8.02	-10.26
Natural log of travel time for car	-1.48e+0	-9.93e-1	-9.43e-1	-9.50e-1	-9.40e-1	-9.63e-1	-9.32e-1
	-7.56	-5.25	-5.12	-5.14	-5.11	-3.95	-5.09
Parking time for car, squared	-1.37e-2	-1.20e-2	-1.18e-2	-1.16e-2	-1.16e-2	-1.69e-2	-1.15e-2
	-9.21	-9.15	-9.05	-8.92	-8.92	-5.04	-8.91

ESTIMATED PARAMETERS, CONT'D	MNL	CNL	#1	#2	#1+2	#4	#5
Scale for transit-bicycle nest		1.53e+0	1.60e+0	1.56e+0	1.59e+0	1.65e+0	1.60e+0
(t-statistic against 1)		3.10	3.35	3.25	3.33	2.38	3.37
Membership parameter for transit		6.36e-1	6.48e-1	6.41e-1	6.47e-1	6.57e-1	6.48e-1
		19.80	20.60	20.38	20.60	14.37	20.75
(t-statistic against 1)		-11.30	-11.20	-11.38	-11.24	-7.50	-11.25
Variance of random coeff for transit				1.96e-3	-1.23e-3	4.50e-2	-4.01e-3
				0.02	-0.02	0.03	-0.02
Variance of random coeff for bicycle				2.75e-1	1.80e-1	-3.56e-2	6.63e-1
				3.75	2.78	-0.03	2.55
Variance of random coeff for car				-3.37e-3	1.64e-3	2.74e+0	-1.03e-3
				-0.02	0.03	3.06	-0.01
Draws				500	500	500	500
Final log-likelihood	-2073.16	-2066.03	-2052.15	-2054.22	-2048.69	-2047.91	-2048.75

All t-statistics (indicated in itallic below the estimated coefficient value) are against 0, except where noted against 1. Compare #1 with the nested logit model with field effect in the second column of Table 6.29.

and spatial network interactions between an agent and the aggregate behavior of other reference agents do indeed have explanatory power. Also, on the basis of non-nested tests, the fit is better than models with "panel" effects representing unobserved heterogeneity between reference groups and no systematic field effects. Finally, on the basis of loglikelihood ratio tests, there is no statistically significant gain in fit at the 0.05 level by adding the "panel" effects, when the systematic field effects are already included in the model.

For the case of network interdependence defined by residential district, a loglikelihood ratio test on the model with a complete set of alternative specific dummy variables (ASDV) for reference group versus the field effect model shows no significant additional explanatory power of the ASDV. With network interdependence defined by socioeconomic group, a loglikelihood ratio test shows that the complete set of ASDV does indeed have explanatory power over the field effect model, but it comes at the expense of 23 additional variables.

Loglikelihood ratio tests on the "random field" effect versus the field effect, shows that there is indeed significant unobserved individual heterogeneity at the 0.05 level on the field variables with all network definition treatments except residential district. For residential district this may not be so surprising given our finding that the complete set of ASDV also showed no statistically significant improvement over the field effect, but gives now further information: not only is there no additional explanatory power due to group heterogeneity, there is also no additional explanatory power due to unobserved individual heterogeneity.

Loglikelihood ratio tests on the "panel field" effect versus the field effect, show that there is no significant unobserved group heterogeneity at the 0.05 level on the field variables. At the 0.1 level however, for the case of network interdependence defined by socioeconomic group there is significant unobserved group heterogeneity. This is in-line with our finding that the complete set of ASDV for socioeconomic group showed statistically significant improvement over the field effect model for socioeconomic group. Similarly on the basis of non-nested tests on the "random field" effect versus the "panel field" effect, only for network interdependence defined by socioeconomic group can we not safely bound the probability of incorrectly selecting the wrong model at less than 0.05, under the hypothesis that the model with higher adjusted rho-squared (random field effect) is the correct one.

An interesting finding with respect to this particular case study is that the specifications allowing agents to weigh both any influence from their socioeconomic peers as well as any influence from their spatial network, and furthermore allowing the possibility to weight these influence differently, all outperform models where each agent belongs to one and only one group. Also interesting is that the speci-

Table 8.2: Loglikelihood, number of estimated parameters and adjusted rho-squared for various reference-group treatments

MODEL STRATEGY	SOC	DTR	PC4	DSD	PSD	SOC+DTR	DTR+PC4	SOC+PC4	SOC+DTR+PC4
#1: Field effect	-2052.15	-2054.88	-2009.55	-2051.9	-2047.61	-2042.92	-2005.1	-1999.58	-1994.65
	16	16	16	16	16	17	17	17	18
	0.305	0.304	0.320	0.305	0.307	0.308	0.321	0.323	0.324
#2: Panel effect	-2054.22	-2063.61	-2061.22	-2058.55	-2059.17	-2058.55	-2061.22	-2059.17	-2059.17
	18	18	18	18	18	18	18	18	18
	0.304	0.301	0.302	0.303	0.302	0.303	0.302	0.302	0.302
#1+#2: Field and panel	-2048.69	-2054.87	-2009.55	-2048.54	-2043.86	-2041.27	-2005.1	-1999.34	-1994.49
	19	19	19	19	19	20	20	20	21
	0.306	0.303	0.319	0.306	0.307	0.308	0.320	0.322	0.323
#4: Random field effect	-2047.91	-2052.01	-2004.79	-2047.08	-2042.51	-2033.31	-2002.55	-1989.9	-1984.37
	19	19	19	19	19	23	23	23	27
	0.306	0.304	0.320	0.306	0.308	0.309	0.320	0.324	0.324
#5: Panel field effect	-2048.75	-2054.87	-2009.55	-2048.06	-2044.42	-2040.95	-2005.1	-1999.1	-1994.02
	19	19	19	19	19	23	23	23	27
	0.305	0.303	0.319	0.306	0.307	0.307	0.319	0.321	0.321

Table 8.3: Likelihood ratio specification tests. The likelihood ratio rest rejects the null hypothesis that restrictions are true if -2[$L^R - L^U$] is greater than χ^2_{DF}, where L^R is the final log-likelihood of the restricted model, L^U is the final log-likelihood of the unrestricted model, and DF is the difference in degrees of freedom between the models.

MODEL(S)	DF	-2[$L^R - L^U$]	$\chi^2_{DF}(0.05)$	P-VALUE
Field effect (#1: unrestricted) vs. Baseline (restricted)				
SOC	1	27.76	3.84	0.000
DTR	1	22.3	3.84	0.000
PC4	1	112.96	3.84	0.000
DSD	1	28.26	3.84	0.000
PSD	1	36.84	3.84	0.000
SOC+DTR	2	46.22	5.99	0.000
DTR+PC4	2	121.86	5.99	0.000
SOC+PC4	2	132.9	5.99	0.000
SOC+DTR+PC4	3	142.76	7.81	0.000
Panel effect (#2: unrestricted) vs. Baseline (restricted)				
SOC	3	23.62	7.81	0.000
DTR	3	4.84	7.81	0.184
PC4	3	9.62	7.81	0.022
DSD	3	14.96	7.81	0.002
PSD	3	13.72	7.81	0.003
SOC+DTR	3	14.96	7.81	0.002
DTR+PC4	3	9.62	7.81	0.022
SOC+PC4	3	13.72	7.81	0.003
SOC+DTR+PC4	3	13.72	7.81	0.003
Field and panel effect (#1+#2: unrestricted) vs. Field effect (#1: restricted)				
SOC	3	6.92	7.81	0.074
DTR	3	0.02	7.81	0.999
PC4	3	0	7.81	1.000
DSD	3	6.72	7.81	0.081
PSD	3	7.50	7.81	0.058
SOC+DTR	3	3.3	7.81	0.348
DTR+PC4	3	0	7.81	1.000
SOC+PC4	3	0.48	7.81	0.923
SOC+DTR+PC4	3	0.32	7.81	0.956

MODEL(S)	DF	$-2[L^R - L^U]$	$\chi^2_{DF}(0.05)$	P-VALUE
Random field effect (#4: unrestricted) vs. Field effect (#1: restricted)				
SOC	3	8.48	7.81	0.037
DTR	3	5.74	7.81	0.125
PC4	3	9.52	7.81	0.023
DSD	3	9.64	7.81	0.022
PSD	3	10.20	7.81	0.017
SOC+DTR	6	19.22	12.59	0.004
DTR+PC4	6	5.1	12.59	0.531
SOC+PC4	6	19.36	12.59	0.004
SOC+DTR+PC4	9	20.56	16.92	0.015
Panel field effect (#5: unrestricted) vs. Field effect (#1: restricted)				
SOC	3	6.8	7.81	0.079
DTR	3	0.02	7.81	0.999
PC4	3	0	7.81	1.000
DSD	3	7.68	7.81	0.053
PSD	3	6.38	7.81	0.095
SOC+DTR	6	3.94	12.59	0.685
DTR+PC4	6	0	12.59	1.000
SOC+PC4	6	0.96	12.59	0.987
SOC+DTR+PC4	9	1.26	16.92	0.999
Field effect model (#1) with various network interdependence				
S+D VS S	1	18.46	3.84	0.000
S+D VS D	1	23.92	3.84	0.000
D+P VS D	1	99.56	3.84	0.000
D+P VS P	1	8.9	3.84	0.003
S+P VS S	1	105.14	3.84	0.000
S+P VS P	1	19.94	3.84	0.000
S+D VS DSD	1	17.96	3.84	0.000
S+P VS PSD	1	96.06	3.84	0.000
S+D+P VS D+P	1	20.9	3.84	0.000
S+D+P VS S+P	1	9.86	3.84	0.002

SOC (S): Socioeconomic peer group clusters
DTR (D): Residential district clusters
PC4 (P): Residential 4-digit postcode clusters
DSD: Overlapping residential district and socioeconomic group
PSD: Overlapping residential postcode and socioeconomic group

Table 8.4: Non-nested specification tests. The non-nested test bounds the probability of erroneously choosing the incorrect model over the true specification under the null hypothesis that the model with higher adjusted rho-squared ($\bar{\rho}^2$) is the true model.

MODEL(S)	DF	$r =$ diff.$\bar{\rho}^2$	$s =$ $2rL_0+$DF	$\Phi(-s^{1/2})$	COMMENTS
Field effect (#1: higher $\bar{\rho}^2$) vs. Panel effect (#2)					
SOC	-2	0.00137	6.14	0.00061	Don't reject
DTR	-2	0.00360	19.46	0.00001	Don't reject
PC4	-2	0.01803	105.34	0.00000	Don't reject
DSD	-2	0.00291	15.30	0.00005	Don't reject
PSD	-2	0.00455	25.12	0.00000	Don't reject
SOC+DTR	-1	0.00559	32.26	0.00000	Don't reject
DTR+PC4	-1	0.01919	113.24	0.00000	Don't reject
SOC+PC4	-1	0.02035	120.18	0.00000	Don't reject
SOC+DTR+PC4	0	0.02167	129.04	0.00000	Don't reject
Random field effect (#4: higher $\bar{\rho}^2$) vs. Panel field effect (#5)					
SOC	0	0.00028	1.68	0.09746	Reject at 0.05
DTR	0	0.00096	5.72	0.00839	Don't reject
PC4	0	0.00160	9.52	0.00102	Don't reject
DSD	0	0.00033	1.96	0.08076	Reject at 0.05
PSD	0	0.00064	3.82	0.02532	Don't reject
SOC+DTR	0	0.00257	15.28	0.00005	Don't reject
DTR+PC4	0	0.00086	5.10	0.01196	Don't reject
SOC+PC4	0	0.00309	18.40	0.00001	Don't reject
SOC+DTR+PC4	0	0.00324	19.30	0.00001	Don't reject
Field effect model (#1) with various network interdependence					
DSD VS SOC	0	0.00008	0.50	0.23975	Reject at 0.2
DSD VS DTR	0	0.00100	5.96	0.00732	Don't reject
PSD VS SOC	0	0.00152	9.08	0.00129	Don't reject
PC4 VS PSD	0	0.01278	76.12	0.00000	Don't reject

SOC: Socioeconomic peer group clusters

DTR: Residential district clusters

PC4: Residential 4-digit postcode clusters

DSD: Overlapping residential district and socioeconomic group

PSD: Overlapping residential postcode and socioeconomic group fications allowing agents to weigh any influence from their fellow district residents differently from any influence from their more immediate neighbors also perform better than a uniform spatial network.

8.4 ENDOGENEITY

Our main aim in sections 8.1, 8.2 and 8.3 has been to explore the empirical estimation of discrete choice models with social and/or spatial feedback effects, including various forms of unobserved (group) heterogeneity. We specifically test for correlation among agents in the error structure, as well as in the systematic utility. Nonetheless an open question may still be the issue of potentially biased estimates due to endogeneity in the field effect, even despite our efforts to explicitly model unobserved heterogeneity in strategies #2, #1+2, #4 and #5.

An alternative strategy for introducing social and spatial network dependencies in choice models is a multi-stage approach, estimating first a complete set of ASDV for reference group as in strategy #3, and then second, writing each ASDV as the sum of a field effect plus an error term so as to be able to apply instrumental variables in a linear regression. Note that applying contextual variables in the utility in the place of the feedback effect is a slightly different research question than the feedback effect itself as endogeneity is removed. In this section we will review this approach as well as another estimation method that entails finding instruments which are correlated with the suspected endogenous variable but uncorrelated with the error term (Train 2009).

Suppose for simplicity that there are no random taste variations on the parameter γ_{in} for the field effect :

$$\gamma_{in} \equiv \beta_i \colon \zeta_{ig_n} = 0, \ \psi_{in} = 0 \tag{8.12}$$

We can then write the general expression (8.10)for the utility as

$$U_{in} = \beta_i p_{ig_n} + \vartheta_i{}' \mathbf{q}_{in} + \xi_{ig_n} + \eta_{in} \tag{8.13}$$

where the systematic utility is given by:

$$V_{in} = \beta_i p_{ig_n} + \vartheta_i{}' \mathbf{q}_{in} \tag{8.14}$$

and the error term is composed of two parts:

$$\varepsilon_{in} = \xi_{ig_n} + \eta_{in} \tag{8.15}$$

In the two estimation methods discussed in this section, we assume by construction that the individual-specific error term η_{in} for alternative i for individual n is uncorrelated with all the variables p_{ig_n} and q_{in} in the systematic utility. Furthermore we assume the variables

q_{in} are truly exogenous, meaning that they are uncorrelated with both the group-specific error term ξ_{ig_n} for alternative i for each individual's reference group g_n, as well as the individual-specific error term η_{in} for alternative i for individual n. If a social field effect or network effect p_{ig_n} is correlated with the reference group-specific error term ξ_{ig_n}, standard estimation methods will fail to retrieve consistent estimators of model parameters.

8.4.1 *The BLP Approach*

The BLP approach, developed in a series of papers by Berry (1994) and Berry, Levinsohn, and Pakes (1995, 2004), requires the endogenous variable to be shared (i.e. equal value) between multiple decision makers in the same "market."

$$p_{ig_n} \equiv p_{ig} \tag{8.16}$$

The error is then also presumed to be shared (i.e. equal value) between multiple decision makers in the same "market."

$$\xi_{ig_n} \equiv \xi_{ig} \tag{8.17}$$

This approach may be suitable in social influence models when the suspected endogeneity occurs via a field effect. In this case each reference group is interpreted as a "market."

The key to the BLP approach is to isolate the terms in (8.13) for the social field effect p_{ig} and the reference group-specific error ξ_{ig}, that are suspected to be correlated, and to replace these with a set of market-specific, alternative-specific constants κ_{ig} as follows:

$$\begin{aligned} U_{in} &= \beta_i p_{ig} + \vartheta_i{}' \mathbf{q}_{in} + \xi_{ig} + \eta_{in} \\ &= [\beta_i p_{ig} + \xi_{ig}] + \vartheta_i{}' \mathbf{q}_{in} + \eta_{in} \\ &= \kappa_{ig} + \vartheta_i{}' \mathbf{q}_{in} + \eta_{in} \end{aligned} \tag{8.18}$$

where

$$\kappa_{ig} \equiv \beta_i p_{ig} + \xi_{ig} \tag{8.19}$$

The issue with the suspected correlation between p_{ig} and ξ_{ig} is thus effectively moved out of the choice model to a linear model (8.19) where it is more straightforward to handle via two-stage instrumental variables approach. The field effect is expressed in terms of a linear-in-parameters function of a vector of exogenous variables z_{ig} :

$$p_{ig} = \theta_i{}' \mathbf{z}_{ig} + \nu_{ig} \tag{8.20}$$

where: $\theta_i = [\theta_i, \dots, \theta_i]'$ is a vector of $K*_i$ unknown alternative-specific parameters, whereby the length of the vector is allowed to vary across alternatives; and ν_{ig} is a group-specific error term for alternative i for each reference group g.

The fitted values of the field effect are then used to estimate (8.19):

$$\kappa_{ig} = \beta_i \hat{p}_{ig} + \omega_{ig} \tag{8.21}$$

where

$$\hat{p}_{ig} = \hat{\theta}_i' \mathbf{z}_{ig} \tag{8.22}$$

Since by design, the fitted values of the field effect are uncorrelated with the error ω_{ig}, ordinary least squares regression (8.21) will result in a consistent estimate of the parameter β_i. The least square residuals of (8.21) and the fitted values of the field effect are then inserted back into the choice model as follows:

$$\begin{aligned} U_{in} &= \kappa_{ig} + \vartheta_i{}' \mathbf{q}_{in} + \eta_{in} \\ &= [\beta_i \hat{p}_{ig} + \hat{\omega}_{ig}] + \vartheta_i{}' \mathbf{q}_{in} + \eta_{in} \\ &= \beta_i \hat{p}_{ig} + \vartheta_i{}' \mathbf{q}_{in} + \hat{\omega}_{ig} + \eta_{in} \end{aligned} \tag{8.23}$$

where

$$\hat{\omega}_{ig} = \kappa_{ig} - \hat{\kappa}_{ig} = \kappa_{ig} - \hat{\beta}_i \hat{p}_{ig} \tag{8.24}$$

This three-stage approach is thus summarized as follows:

1. Estimate a choice model with constants corresponding to each alternative and market

2. The field effect variable is regressed on exogenous instruments and the fitted values of the field effect are calculated for each market.

3. Regress the set of constants from step 1 on the fitted values of the field effect from step 2 in a linear model and calculate the residuals. The choice model is then corrected by replacing the constants with the entire right hand side of this linear model, including residuals.

Walker et al. (2011) use this approach to deal with endogeneity in the field effect term in a multinomial logit model of mode choice in the Netherlands.

8.4.2 *The Control Function Approach*

The control function approach, proposed by Hausman (1978), Heckman (1978) and Heckman and Robb (1985), is an approach for dealing with endogeneity at the individual-level. Rivers and Vuong (1988) provide an application to binary probit models with fixed coefficients.

Guevara (2010) provides an application to multinomial choice models with fixed coefficients. Petrin and Train (2010) provide an application to multinomial choice models with random coefficients. This approach is in principle applicable in social influence models both for field effects and conditional autoregressive network effects, however the only application to social influence models in the transportation literature that the authors are aware of is for the case of field effects which we review below.

In the case of field effects, we again make the assumptions (8.16) and (8.17) in the expression (8.13) and we also require that it is possible to express the field effect as in (8.20) in terms of a linear-in-parameters function of a vector of exogenous variables z_{ig} plus an error term ν_{ig}:

$$\begin{aligned} U_{in} &= \beta_i p_{ig} + \vartheta_i{}' \mathbf{q}_{in} + \xi_{ig} + \eta_{in} \\ p_{ig} &= \theta_i{}' \mathbf{z}_{ig} + \nu_{ig} \end{aligned} \tag{8.25}$$

As above, our concern is that the social field effect p_{ig} might be correlated with the reference group-specific error term ξ_{ig}. Explicitly in the system (8.25), the concern is the possible correlation between the error terms ξ_{ig} and ν_{ig}.

The key to the control function approach is to decompose the reference group-specific error term ξ_{ig} into its conditional expectation given ν_{ig}, and the deviations σ_{ig} around this expectation. The conditional expectation $E(\xi_{ig}|\nu_{ig})$ is called the "control function". Importantly, by design, the deviations σ_{ig} are uncorrelated with ν_{ig} and thus also uncorrelated with social field effect p_{ig}. The utility function in the system (8.25) can then be written:

$$U_{in} = \beta_i p_{ig} + \vartheta_i{}' \mathbf{q}_{in} + E(\xi_{ig}|\nu_{ig}) + \sigma_{ig} + \eta_{in} \tag{8.26}$$

where

$$\xi_{ig} \equiv E(\xi_{ig}|\nu_{ig}) + \sigma_{ig} \tag{8.27}$$

The issue with the suspected correlation between p_{ig} and ξ_{ig} is thus effectively removed from the choice model through the explicit inclusion of the control function $E(\xi_{ig}|\nu_{ig})$ as an explanatory variable in the utility function. The remaining error $\sigma_{ig} + \eta_{in}$ in the utility function is no longer correlated with any explanatory variables. A change of scale is however expected between the utility functions in (8.25) and (8.26) since ξ_{ig} and σ_{ig} have different variance.

Since by design, the exogenous variables z_{ig} are uncorrelated with the error ν_{ig}, ordinary least squares regression of the second equation in the system (8.25) will result in a consistent estimate of the vector of parameters θ_i. The least square residuals are then inserted back into the control function in the choice model as follows:

$$U_{in} = \beta_i p_{ig} + \vartheta_i{}' \mathbf{q}_{in} + E(\xi_{ig}|\hat{\nu}_{ig}) + \sigma_{ig} + \eta_{in} \tag{8.28}$$

where

$$\hat{\nu}_{ig} = p_{ig} - \hat{p}_{ig} = p_{ig} - \hat{\theta}_i' \mathbf{z}_{ig} \tag{8.29}$$

This two-stage approach is thus summarized as follows:

1. The field effect variable is regressed on exogenous variables and residuals are calculated.

2. The residuals from step 1 are then used in an auxiliary variable (the control function) in the estimation of the choice model. The auxiliary variable can be used to test whether the field effect variable is exogenous.

The usefulness of the control function approach depends on the ability to express the distribution of ξ_{ig} conditional on ν_{ig}. For example, if it can be assumed that ξ_{ig} and ν_{ig} are jointly normal with zero mean and constant covariance matrix, then the expectation of ξ_{ig} conditional on ν_{ig} is simply $\rho_i \nu_{ig}$ where ρ_i reflects the covariance. The utility function (8.26) can then be written:

$$U_{in} = \beta_i p_{ig} + \vartheta_i{}' \mathbf{q}_{in} + \rho_i \nu_{ig} + \sigma_{ig} + \eta_{in} \tag{8.30}$$

where

$$E(\xi_{ig} | \nu_{ig}) = \rho_i \nu_{ig} \tag{8.31}$$

The deviations σ_{ig} in this case are then normal with constant variance. If the error term η_{in} is also normal then the choice probability is mixed probit. If the error term η_{in} is extreme value then the choice probability is mixed logit. If the deviations σ_{ig} can be neglected then the choice probability does not involve mixing. The coefficient ρ_i can be used to test whether the field effect p_{ig} is exogenous, under the null hypothesis $\rho_i = 0$. The two-stage estimation approach leads to consistent estimates of the parameters up to a scale.

Goetzke and Weinberger (2012) use the control function approach in their study of the influence of endogenous and contextual social effects on automobile ownership in New York City.

8.4.3 *Instrument Selection*

Guevara (2010) presents an excellent overview of the state-of-the-art in testing the validity of instruments in discrete choice models. Sargan (1958) and later Basman (1960) noted that validation of instruments may be possible in linear regression models if there are more instruments than endogenous variables, using the residuals of the instrumental variable regression. Lee (1992) noted that the objective function of a two-stage minimum chi-squared estimator proposed by Amemiya (1978) and further studied by Amemiya (1979, 1983), Lee

(1981) and Newey (1987) can be similarly used to construct a test for over-identifying restrictions for the probit model when an endogenous variable can be written in terms of instrumental variables. Guevara and Ben-Akiva (2008) propose an extension of the Sargan test specifically for logit models using the generalized residuals concept for non-linear models developed by Cox and Snell (1968) and applied by McFadden (1987) for a test of omitted attributes, extending a result by Engle (1984) for binary logit. Guevara (2010) also proposes an additional "direct" test for the validity of instruments in discrete choice models, based on an alternative implementation of McFadden's (1987) test for omitted attributes used by Train, Ben-Akiva and Atherton (1989). Whereas the Amemiya-Lee-Newey test and Guevara and Ben-Akiva's (2008) test require the calculation of auxiliary variables and the estimation of several auxiliary regressions, Guevara's "direct" test can be carried out using a simple log-likelihood ratio test with a single re-estimation of the original choice model, again provided that there are more available instruments than endogenous variables.

8.5 EXTENSIONS

As far as we are aware, the research presented in sections 8.1, 8.2 and 8.3 is the *first* example of the empirical estimation of a discrete choice model among multiple choice alterantives with social and/or spatial state dependence, where correlation among choice alternatives as well as correlation among decision-making agents in the error structure are also both tested. That this was achievable with readily available software was facilitated by two features: (1) the implementation in Biogeme (Bierlaire 2003) of mixed GEV panel models; and (2) the specific formulation of the interdependence between decision-makers in clusters or overlapping groups, whereby the state dependence in the systematic utility can be modeled as a field effect, and the serial correlation in the error structure can be modeled as a "panel" effect (albeit correlation between choice observation by different decision-makers, rather than a classical correlation between the choice observations by a single decision-maker over time). As noted in subsection 8.2.2, the implementation of a "panel" effect in the context of spatial correlation of errors between different decision-makers in the context of home and work location was done earlier by Bhat (2000), but without modeling the state dependence in the systematic utility.

There is however a long tradition of research in spatial and social econometrics in studying correlation between observational units, although as far as we are aware before the time of our initial work, this research had focused primarily on linear regression with continuous dependent variables, not discrete dependent variables. The spatial models that do deal with discrete dependent variables are binary choice, typically using binary probit-based models, *not* choice between multiple discrete response alternatives whereby correlation between choice alternatives would be a concern in addition to correlation between decision-makers. To place our work in context, we very briefly review the framework of this earlier spatial and social econometric research in subsections 8.5.1, 8.5.2 and 8.5.3, respectively on the conditional autoregressive lag model, the simultaneous autoregressive lag model and the simultaneous autoregressive error model.

One reason for the lack of attention until recently to discrete choice models between multiple choice alternatives in the presence of correlation between decision-makers, has to the with the fact that in a general multivariate probit model formulation with fully flexible specification of errors, the estimation requires the highly computationally demanding solution of a multi-dimensional integral that is as large as the number N of decision-makers in the sample times the number of choice alternatives minus one $(J-1)$, without simplifying assumptions such as we make with the specific formulation of the interdependence between decision-makers in clusters or overlapping groups. In many applications without such simplifying assumptions, the solution of such an N x $(J-1)$ integral by simulated estimation is prohibitive. In our work presented in sections 8.1, 8.2 and 8.3, through the formulation of the inter-agent interdependence in clusters or overlapping groups and the use of mixed GEV panel models implemented in Biogeme, a closed form GEV structure is used to capture correlation between alternatives, and the (panel) mixing is used to capture correlation between decision-makers.

Since the time of our initial work, there have been two landmark research contributions by Bhat (2009, 2011) that facilitate accomodating flexible correlation between decision-makers in circumstances in discrete choice models between multiple choice alternatives. In chronological order, this is first the copula approach to specifying correlation between decision-makers reviewed in subsection 8.5.4, and then subsequently the maximum approximate composite likelihood estimation approach for multinomial probit-based unordered response models which accomodates both state dependence in the systematic utility (see subsection 8.5.2) as well as flexible correlation in the error structure (see subsection 8.5.3).

8.5.1 *Conditional Autoregressive Lag*

Spatial and social econometric models classically use a weighting matrix $w_{inn'}$ to represent social or spatial distance. The weighting matrix describes the degree of influence between observational units. In the consideration of discrete response the weighting matrix may in general be different for different choice alternatives i. A simple measure of *absolute* influence is given by:

$$w_{inn'} = \begin{cases} 1 \text{ if } n' \in g_n \\ 0 \text{ otherwise} \end{cases} \tag{8.32}$$

For example, Sidharthan et al (2011) test the effect of county comembership. A measure of *relative* influence, normalized by the number N_{g_n} of persons in the reference group of agent n, is given by:

$$w_{inn'} = \begin{cases} 1/N_{g_n} \text{ if } n' \in g_n \\ 0 \text{ otherwise} \end{cases} \tag{8.33}$$

Páez and Scott (2007) and Páez, Scott and Volz (2008) give a useful summary of these ideas; Akerlof (1997) and Leenders (2002) provides further discussion. Various other flexible formulations are also possible. For example, Sidharthan et al. (2011) estimate an *inverse distance* weighting matrix, where people living closer to an individual exert more influence than people living farther away. They also test this against other formulations of the weight matrix such as based on *inverse of exponentiated distance*, as well as specifications of income and age similarity created using demographic "distance" measures.

State dependence between decision-makers can be expressed in terms of the weight matrix as follows:

$$p_{ig_n} \equiv p_{in} = \sum_{n'=1}^{N} w_{inn'} y_{in'} = [w_{in1} \cdots w_{inN}] \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iN} \end{bmatrix} = \mathbf{w}_{in} \mathbf{y}_i \tag{8.34}$$

where $y_{in'}$ equals one if individual n' chose alternative i and zero otherwise. Note that field effect variables studied in this dissertation can similarly be expressed in general in this way, with suitable definition of an N x N weight matrix.

As is the case with the field effect formulation, we must be careful as the social lag term $y_{in'}$ is likely to suffer endogeneity errors due to omitted variable bias and simultaneity. A useful property of this model is that it can be estimated using standard software as long as the weighing matrix has no parameters (i.e. the weights are fully known before estimation) and appropriate measures are taken for handling endogeneity in $y_{in'}$.

Goetzke (2008) analyzes transit mode choice using a spatial autogressive formulation of the weighting matrix but assumes that $y_{in'}$ is exogenous to simplify model estimation. Adjemian et al. (2010) use a similar model to predict auto ownership by class with a series of binary logit models. In simulation studies, Páez and Scott (2007) and Páez, Scott and Volz (2008) modify equation (8.34) by having the utility an individual gains from choosing an alternative depend on the past choices of peers:

$$\begin{aligned} p_{ig_n t} \equiv p_{in,t} &= \sum_{n'=1}^{N} w_{inn'} y_{in',t-1} \\ &= [w_{in1} \cdots w_{inN}] \begin{bmatrix} y_{i1,t-1} \\ \vdots \\ y_{iN,t-1} \end{bmatrix} = \mathbf{w}_{in} \mathbf{y}_{i,t-1} \end{aligned} \tag{8.35}$$

8.5.2 *Simultaneous Autoregressive Lag*

Since conditional autoregressive models have simultaneity issues, an alternative approach is to model the decision process as a system of simultaneous equations. Behaviourally, the formulation is different from the conditional autoregressive and field effect formulations as the individual is assumed to be affected by perceptions of the *preferences* of others, modeled by the latent utilities $U_{in'}$, rather than the perceptions of the others' actual *decisions* $y_{in'}$.

$$p_{ig_n} \equiv p_{in} = \sum_{n'=1}^{N} w_{inn'} U_{in'} = [w_{in1} \cdots w_{inN}] \begin{bmatrix} U_{i1} \\ \vdots \\ U_{iN} \end{bmatrix} = \mathbf{w}_{in} \mathbf{U}_i \tag{8.36}$$

Substituting (8.36) into (8.10) we have

$$U_{in} = \gamma_{in} \mathbf{w}_{in} \mathbf{U}_i + \vartheta_i{}' \mathbf{q}_{in} + \xi_{ig_n} + \eta_{in} \tag{8.37}$$

Suppose for simplicity of exposition that there are no random taste variations in (8.11) on the parameter γ_{in} for the state dependence between decision-makers:

$$\gamma_{in} \equiv \beta_i \colon \zeta_{ig_n} = 0,\ \psi_{in} = 0 \tag{8.38}$$

and there is no *additional* network-specific error term for alternative i for each individual's reference group g_n in the utility function (8.37)

$$\xi_{ig_n} = 0 \tag{8.39}$$

If we let W_i, Q_i and η_i each respectively be the N vertically stacked terms w_{in}, q_{in} and η_{in}, then expressing (8.37) in vector form we have a system of N simultaneous equations:

$$\mathbf{U}_i = \beta_i \mathbf{W}_i \mathbf{U}_i + \vartheta_i{}' \mathbf{Q}_i + \eta_i \tag{8.40}$$

In the spatial econometrics literature, (8.40) is referred to as the spatially autoregressive lagged dependent variable model (SAL). Rearranging terms, we have:

$$\begin{aligned}\mathbf{U}_i - \beta_i \mathbf{W}_i \mathbf{U}_i &= \vartheta_i{}' \mathbf{Q}_i + \eta_i \\ (I - \beta_i \mathbf{W}_i)\mathbf{U}_i &= \vartheta_i{}' \mathbf{Q}_i + \eta_i \\ \mathbf{U}_i &= (I - \beta_i \mathbf{W}_i)^{-1}(\vartheta_i{}' \mathbf{Q}_i) + (I - \beta_i \mathbf{W}_i)^{-1}\eta_i\end{aligned} \tag{8.41}$$

where I is the N x N identity matrix. The term $(I - \beta_i W_i) - 1$ is referred to as the "*Leontif inverse*" (Anselin 2002). Equivalently, making the definitions:

$$\begin{aligned}\mathbf{V}_i &\equiv (I - \beta_i \mathbf{W}_i)^{-1}(\vartheta_i{}' \mathbf{Q}_i) \\ \varepsilon_i &\equiv (I - \beta_i \mathbf{W}_i)^{-1}\eta_i\end{aligned} \tag{8.42}$$

for the vectors $V_i = [V_{i1}, \ldots, V_{iN}]'$, and $\varepsilon_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iN}]'$ we can write the system of N simultaneous equations simply as:

$$\mathbf{U}_i = \mathbf{V}_i + \varepsilon_i \tag{8.43}$$

In contrast with (8.1), it is important to emphasize in (8.43) with the definitions (8.42) that the utility U_{in} for a single individual n is thus related to *both* the systematic utility and the error of all the other individuals in the system through the N x N weight matrix W_i in the Leontif inverse. Due to the general correlation of the error for a given individual with the error of other individuals, special techniques are necessary for estimation. In the case that there is flexible correlation between multiple alternatives with unordered response, estimation of the simultaneous autoregressive lag model will in general involve a multidimensional integral of degree given by the number N of individuals in the sample times the number of choice alternatives in the universal choice set minus one $(J - 1)$.

Fleming (2004) surveys various approaches for estimation of binary probit simultaneous autoregressive lag models, including the Expectation Maximization (EM) algorithm due to McMillen (1992), and Bayesian estimation with a Monte Carlo Markov Chain procedure involving Gibbs sampling with a Metropolis-Hastings algorithm due to LeSage (2000) and Bolduc et al (1997). Extending Vijverberg (1997), Beron et al. (2003) and Beron and Vijverberg (2004) propose a recursive importance sampling (RIS) technique that is a more general form of the well-known GHK simulator (Geweke 1989, Hajivassiliou 1993, Keane 1993, 1994) to directly evaluate the likelihood function involving maximum simulated likelihood. Fleming (2004) also proposes a weighted non-linear version of the two stage least squares (or instrumental variables) estimator described by Kelejian and Prucha (1998) for the continuous-dependent variable case.

In a landmark paper, Bhat (2011) proposes the maximum approximate composite marginal likelihood (MACML) estimation approach for multinomial probit models, including formulations for panel models, as well as social/spatial independencies. A key component of Bhat's approach is the analytic approximation for the evaluation of the multivariate standard normal cumulative distribution (MVNCD) function by decomposition into a product of conditional probabilities, as proposed by Solow (1990) based on Switzer (1977), and refined by Joe (1995). The second key component of Bhat's approach is composite marginal likelihood (CML) inference, i.e. maximizing a "surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods" (Lindsay 1988, Cox and Reid 2004, Zhao and Joe 2005; Varin 2008, Varin et al. 2011). By combining MVNCD analytic approximation with CML inference, Bhat's resulting "simulation-free" estimation involves only univariate and bivariate cumulative normal distribution function evaluations, permitting estimation of model structures otherwise infeasible. In a companion paper, Bhat and Sidharthan (2011) demonstrate in a simulation evaluation that the MACML estimation method for mixed multinomial probit models is substantially faster, more stable and the absolute percentage bias are smaller than maximum simulated likelihood (MSL) estimation. Bhat and Sidharthan (2012) present an extension of the MACML approach to accommodate non-normal mixing in cross-sectional and panel multinomial probit using the multivariate skew-normal distribution function. In a series of applications, Sidharthan et al. (2011) apply the MACML approach to children's school travel mode choice accounting for effects of spatial and social interaction, Sidharthan and Bhat (2012) incorporating spatial dynamics and temporal dependency in land use change models, and Paleti et al. (2013) model household vehicle type choice accommodating spatial dependence effects.

Whalen et al (2012) estimate an ordinal probit model with a spatially lagged dependent variable in an application exploring the influence of membership in tertiary street communities (T-communities), on sense of community in a university town. They apply a composite marginal likehood (CML) approach whereby a pairwise likelihood function is formed by the product of the likelihood contributions of all pairs (or couplets) of individuals, without the necessity for an MVNCD approximation.

8.5.3 *Simultaneous Autoregressive Error*

Analogous to the formulation in subsection 8.5.2 of a simultaneous autoregressive lag, correlations in the error terms can also be expressed in terms of the weighting matrix $w_{inn'}$ representing social or spatial distance. Suppose again for simplicity of exposition that there are no random taste variations in (8.11) on the parameter γ_{in} for the state dependence between decision-makers:

$$\gamma_{in} \equiv \beta_i \colon \zeta_{ig_n} = 0, \ \psi_{in} = 0 \tag{8.44}$$

The error in the general utility functions (8.1) and (8.10) can then be expressed as:

$$\varepsilon_{in} = \xi_{ig_n} + \eta_{in} \tag{8.45}$$

We allow for spatially correlated errors by making the definition

$$\xi_{ig_n} \equiv \lambda_i \sum_{n'=1}^{N} w_{inn'}\varepsilon_{in'} = \lambda_i [w_{in1} \cdots w_{inN}] \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iN} \end{bmatrix} = \lambda_i \mathbf{w}_{in}\varepsilon_i \tag{8.46}$$

Substituting (8.46) into (8.45) we have

$$\varepsilon_{in} = \lambda_i \mathbf{w}_{in}\varepsilon_i + \eta_{in} \tag{8.47}$$

If we let W_i and η_i each respectively be the N vertically stacked terms w_{in} and η_{in}, then expressing (8.47) in vector form we have a system of N simultaneous equations:

$$\varepsilon_i = \lambda_i \mathbf{W}_i \varepsilon_i + \eta_i \tag{8.48}$$

Re-arranging terms, we have:

$$\begin{aligned} \varepsilon_i - \lambda_i \mathbf{W}_i \varepsilon_i &= \eta_i \\ (I - \lambda_i \mathbf{W}_i)\varepsilon_i &= \eta_i \\ \varepsilon_i &= (I - \lambda_i \mathbf{W}_i)^{-1}\eta_i \end{aligned} \tag{8.49}$$

where I is the N x N identity matrix. Now expressing the utility function (8.10) in vector form under the conditions (8.44), we have a system of N simultaneous equations:

$$\mathbf{U}_i = \beta_i \mathbf{p}_i + \vartheta_i{}'\mathbf{Q}_i + (I - \lambda_i \mathbf{W}_i)^{-1}\eta_i \tag{8.50}$$

where p_i and Q_i are respectively the N vertically stacked terms p_{ig_n} and q_{in}. In the spatial econometrics literature, (8.50) is referred to as the spatially autoregressive error model (SAE). Equivalently, making the definitions:

$$\mathbf{V}_i \equiv \beta_i \mathbf{p}_i + \vartheta_i{}'\mathbf{Q}_i \tag{8.51}$$

for the vectors $V_i = [V_{i1}, \ldots, V_{iN}]'$, we can write the system of N simultaneous equations as:

$$\mathbf{U}_i = \mathbf{V}_i + (I - \lambda_i \mathbf{W}_i)^{-1}\eta_i = \mathbf{V}_i + \varepsilon_i \tag{8.52}$$

Again, in contrast with (8.1), it is important to emphasize in (8.52) that the utility U_{in} for a single individual n is thus related to the error of all the other individuals in the system through the N x N weight matrix W. As is the case with the simultaneous autoregressive lag model, when there is flexible correlation between multiple alternatives with unordered response, estimation will in general involve a multidimensional integral of degree given by the number N of individuals in the sample times the number of choice alternatives in the universal choice set minus one $(J-1)$.

Due to the general correlation of the error for a given individual with the error of other individuals, special techniques are again necessary for estimation. All of the full spatial information estimators reviewed in Fleming (2004) for binary probit spatial autoregressive lag models mentioned in subsection 8.5.2 can also be used for binary probit spatial autogressive error models: the EM algorithm, RIS simulator and the Gibbs sampler. Fleming (2004) also proposes a weighted non-linear feasible generalized least squares estimator based on the work of Kelejian and Prucha (1999). In the special case that errors are heteroscedastic and the off-diagonal terms of the variance-covariance matrix are zero, a two-step Generalized Method of Moments (GMM) procedure due to Pinkse and Slade (1998) can be used. For multinomial probit spatial autogressive error models, Bhat's (2011) MACML estimation approach discussed in subsection 8.5.2 can be used.

8.5.4 *Copula Approach*

Bhat and Sener (2009) propose the novel use of a copula (Sklar 1959, 1973) to accommodate spatial error correlation across individuals in discrete choice models, thus avoiding the need for the specification of a weight matrix as in the case of the autoregressive models. They explain:

> "A copula is a device or function that generates a stochastic dependence relationship among random variables with pre-specified (parametric) marginal distributions. In essence, the copula approach separates the marginal distributions from the dependence structure, so that the dependence structure is entirely unaffected by the marginal distributions assumed. This provides substantial flexibility in correlating random variables, which may not even have the same marginal distributions."

Bhat and Sener (2009) consider the Farlie-Gumbel-Morgenstern (FGM) family of copulas which is suited to incorporate spatial correlation and has a simple analytic form allowing for either positive or negative dependence. By combining a binary logit distribution for each individual error term with the FGM copula to generate the dependence between error terms, the result is a spatial logit model structure with a closed-form solution for joint choice probabilities. In an application to teenagers' weekday physical activity participantion, they show their spatially correlated heteroscedastic binary logit model leads to a better model fit than a standard binary logit model which does not take the error correlations between individuals into account.

Bhat, Sener and Eluru (2010) and Sener, Eluru and Bhat (2010) extend the copula approach to ordered choice models with applications respectively to spatial correlation on one hand and social correlation in clusters (family units) on the other.

For the spatial correlation case, Bhat, Sener and Eluru (2010) compare FGM and Gaussian copulas paired with ordinal logit and ordinal probit models. By using the composite marginal likelihood (CML) inference approach to estimation discussed earlier in subsection 8.5.2, they again avoid the need for simulation techniques. In an application to the daily episode frequency of teenagers' weekday recreational activity participation, they find that a Logistic-Gaussian model has the best fit for both physically active and inactive recreation, highlighting the value of separating the univariate marginal distribution of a given individual from the multivariate dependence structure between individuals.

For the case of social correlation in family units, Sener, Eluru and Bhat (2010) consider Clayton, Gumbel, Frank, and Joe variants of Archimedean copulas, since they provide closed form multivariate cumulative distribution functions when there are identical dependencies between pairs of individual within a cluster. The number of computations increases rapidly with the number of individuals within a cluster, but they find this is not problematic with clusters sizes of 6 or less individuals due to the closed form structures. In an application to weekend physical activity participation levels within family units comparing the four Archimedean copulas paired with ordinal logit and ordinal probit models, they find that a Logistic-Clayton model specification provides the best data fit, revealing a clear asymmetry in the dependence relationship of the physical activity propensities of individuals in the same family.

Sener and Bhat (2012) extend the copula approach to unordered multinomial choice, comparing FGM and Gaussian copulas and a new Generalized Gumbel (GG) copula proposed by Bhat (2009). They explain while the former copulas are "radially symmetric and assume the property of asymptotic independence", the GG copula allows "asymmetric and extreme tail dependence, ie. the dependence is higher in the right tail than in the left tail". They apply CML inference to avoid the need for simulation. In an application to teenagers' social and recreational activity participation, they find that the MNL-GG model has the best fit, implying that teenagers in close residential proximity tend to have more correlation of higher activity levels, than correlation of lower activity levels.

8.6 CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

In this chapter we re-visit the framework for conceptualizing the interdependence of decision-makers' choices, making a distinction between social versus spatial network interdependencies and between identifiable versus aggregate agent interdependencies, presented earlier at the outset of Chapter 7. We discuss five strategies for introducing social and spatial network interdependencies into choice models, focusing on feedback effects and on correlated effects. Due to the nature of available data, we are unable to consider identifiable agent interaction in the case study. Instead we consider aggregate agent interdependencies and apply the model strategies for nine variations on three network treatments, one of these defined by socioeconomic group and two defined spatially based on residential location. According to likelihood ratio and non-nested specification tests, the best performing model strategy for this particular case study is what we term a "random field effect" model, namely one with unobserved individual heterogeneity on the group mean state dependence.

By casting the research question on social and spatial network interdependencies in terms of an analogy with the treatment of typical panel data, we gain two important benefits. First it allows us to build on some well-understood properties of typical panel data models. Second, but not least, it allows us to apply freely available software for the estimation of the models with social-spatial interdependence. In so doing, we hope to stimulate researchers and practitioners to adopt these techniques due to the relatively lower entry barrier for researchers and practitioners than could be the case if dedicated own code would need to be written or if expensive software would need to be purchased.

The modeling strategies presented in this chapter can be useful for researchers and practitioners who have a priori reason to believe there is a feedback effect ("true" state dependence) in their work, for example, on theoretical or qualitative grounds. Even without such a priori knowledge, the so-called field variable applied throughout this dissertation can still be a very powerful and practical means of capturing variation in a data set, especially in avoiding proliferation of estimated parameters as compared to estimating a complete set of fixed effects, although strict caution must then be exercised in the interpretation of predictions over time. Access to temporal panel data would allow a possibility to remove ambiguity and be more definitive about results. It is ultimately our hope that researchers and practitioners will start to pay more attention to the importance and consequences of agent interdependence. In particular it would be an important impulse to science and informed policy decision-making to see more temporal panel data collected with the aim towards understanding the potential inherent dynamic associated with feedback effects.

Our goal in this dissertation has been a systematic, step-by-step exploration of discrete choice with social and spatial interactions, where the conditions in the seminal work of Aoki (1995), Brock and Durlauf (2001a, 2002, 2006) and Blume and Durlauf (2003), assuming homogeneous decision-makers, global interactions and laws of large of numbers, are relaxed one at time. We have started our journey with binary choice where correlation between alternatives is not relevent, and then proceeded to multinomial choice where it can indeed be relevant. The research presented in Part II allows for unobserved preference heterogeneity between choice alternatives by studying the nested logit model. Next, by drawing on the computational possibilities permitted through social simulation of multi-agent systems (MAS), Part III relaxes the assumption of global interactions and considers instead local interactions within several hypothesized social and spatial network structures. Additional heterogeneity is thus hereby induced by the influence on a given decision-maker's choice by the particular network connections he or she has and the particular perceived percentages, for example, of the agent's neighbors or socio-economic peers making each choice. Discrete choice estimation results controlling these heterogeneous individual preferences are embedded in a multi-agent based simulation model in order to observe the evolution of choice behavior over time with socio-dynamic feedback due to the network effects. The MAS approach also gives us an additional advantage in the possibility to test size effects, and thus relax the assumption of large numbers, as well as test the effect of different initial conditions. Finally an extra benefit is gained via the MAS approach in that we are not confined to study only the equilibrium behavior, and have the possibility here to observe the time-varying trajectories of the choice behavior. Averaged over time, the emergent behavior can yield a quite different picture than the theoretical results predicted in the seminal work. To bring the work of this dissertation full circle, it would be desirable at this point to *return* to the theoretical analysis in Part II and develop mean field benchmark results for the cross-nested logit model. After doing this, it would be desirable to *return* to the agent-based modeling analysis in Part III and explore simulation results when additional unobserved heterogeneity is explicitly added to the utility and to taste variations on the estimated parameter for the field effect. It is to be expected that the additional symmetry-breaking with the cross-nested logit model will yield even more equilibrium regimes and additional new complex patterns of bifurcation behavior than that already presented for the nested logit model in Chapter 5. Furthermore it is hypothesized that the additional fluctuations when adding a panel effect, a random field effect or a panel field effect as described in strategies #2, #4 and #5 in this chapter, may yield interesting results as an extension of the simulation results in Chapter 7. However as this dissertation is already exceptionally lengthy as it is, both of these exercises are left for future research.

We maintain in any case that a systematic approach to incrementally relaxing conditions is most appropriate for deeper understanding of such a non-linear complex system. It would be difficult to appreciate bifurcations in a cross-nested model, without having understood bifurcations in a multinomial logit and nested logit model first. Likewise, it would be difficult to appreciate the changes due to adding further unobserved and induced heterogeneous fluctuations, without having understood models without these fluctuations. Having the in-depth and thorough understanding developed systematically in this dissertation forms a solid foundation and jumping off point for further work, which we hope will lead to and inspire much continued research into the future – theoretical, computational, and econometric.