A New Spatial (Social) Interaction Discrete Choice Model Accommodating for Unobserved Effects due to Endogenous Network Formation

Chandra Bhat *(Corresponding Author)*
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535; Fax: 512-475-8744
Email: bhat@mail.utexas.edu
and
King Abdulaziz University, Jeddah 21589, Saudi Arabia

ABSTRACT

This paper formulates a model that extends the traditional panel discrete choice model to include social/spatial dependencies in the form of dyadic interactions between each pair of decision-makers. In addition, the formulation accommodates spatial correlation effects as well as allows a global spatial structure to be placed on the individual-specific unobserved response sensitivity to exogenous variables. We interpret these latter two effects, sometimes referred to as spatial drift effects, as originating from endogenous group formation. To our knowledge, we are the first to suggest this endogenous group formation interpretation for spatial drift effects in the social/spatial interactions literature. The formulation is motivated in a travel mode choice context, but is applicable in a wide variety of other empirical contexts. Bhat's (2011) maximum approximate composite marginal likelihood (MACML) procedure is used for model estimation. A simulation exercise indicates that the MACML approach recovers the model parameters very well, even in the presence of high spatial dependence and endogenous group formation tendency. In addition, the simulation results demonstrate that ignoring spatial dependence and endogenous group formation when both are actually present will lead to bias in parameter estimation.

Keywords: Spatial interactions, social interactions, spatial lag, spatial drift, endogenous group formation, maximum approximate composite marginal likelihood, panel data.

1. INTRODUCTION

1.1. Background

The decision of an individual to use a particular form of travel mode, or the decision of an individual to participate in leisure and other physical activities, or the decision of a household to purchase a certain type of vehicle are all examples of activity-travel behavior choices where discrete choices of one agent may be inter-related with those of others based on spatial and/or social proximity. There is now a substantial academic literature on this topic in several fields, including marketing, economics, regional science, and transportation. This is not surprising, since spatial/social interactions can be exploited by decision-makers to achieve desired system end-states. That is, reinforcing spatial/social interactions imply that a stimulus applied to one decision agent can get magnified through the agent's interactions with other agents, so that the aggregate-level effect of a policy can be higher than the sum of individual-level effects. As a simple illustration of this point, consider a travel mode choice context, and assume that the non-motorized mode use propensity of one individual influences that of her/his residential neighbors (this may happen because, for example, the pro-bicycle attitude or health consciousness of one individual rubs off on other neighbors of the individual through social interactions). Then, a limited-funding information campaign to promote the use of non-motorized modes of transportation would do well to target individuals from different neighborhoods, rather than targeting individuals from the same neighborhood. Doing so will benefit from the "ripple wave" (or spatial multiplier) effect caused by intra-neighborhood social exchanges, so that the aggregate-level effect of the information campaign on non-motorized mode use can be substantial.

A challenge, however, when investigating the issue of social/spatial interactions, is to isolate these interactions from other "spurious" sources that may inappropriately get manifested as social/spatial interactions (in the current paper, we will refer to social/spatial interactions in the strict context of some form of dyadic interaction between individuals located in close social or spatial proximity, and accommodate it using a spatial lag effect commonly adopted in the spatial econometrics literature; this effect is also oftentimes referred to as the endogenous interaction effect in the literature, as in Elhorst, 2010 and Manski, 1993). Using Manski's terminology, the "spurious" sources of the spatial lag effect may include unobserved correlation effects and/or exogenous interaction effects (the latter are also sometimes referred to as contextual effects). The former source, as generally discussed in the literature, relates to unobserved factors that drive the decisions of decision agents located in close proximity of one another. For instance, two spatially proximate neighborhoods may both have good continuous bicycle pathways and/or seating areas along walking pathways. If these detailed bicycle/pedestrian infrastructure characteristics are not available to the analyst, and are not accounted for, the resulting unobserved correlation would get manifested as a spurious social/spatial interaction effect (because of the elevated use of non-motorized mode use in the two neighborhoods). The second spurious source, exogenous variable effects, relates to the exogenous variables of one agent directly impacting the decision-making of a neighboring agent (as opposed to the indirect impact through the social/spatial interaction effect). For instance, this may occur if the pedestrian facilities in one neighborhood affect the decision of an agent in an adjacent neighborhood to walk more.

1.2. Overview of Earlier Literature on Identification in Models of Spatial/Social Interaction

Many earlier studies have examined the case of identification with cross-sectional data in the presence of social/spatial (*i.e.*, endogenous) interactions as well as exogenous interaction effects (with or without unobserved correlation effects; see Blume *et al*. 2011 for an extensive review). Several of these studies have been undertaken in the context of a linear-in-means model (such a model relates the outcome of each individual linearly to her/his characteristics, the mean outcome or the mean expected outcome of the person's reference group, and the mean exogenous characteristics of the reference group). The studies also typically assume peer social interaction effects only within exclusive groups (that is, individuals are partitioned in groups, and individuals in a group are influenced by all members within the group and no one outside the group; in the notation of the next section, the weight matrix **W** is block-diagonal, with each block diagonal representing the interactions within a group). In these studies, the spatial/social weight used to capture the strength of the social and exogenous interaction effects between two agents is essentially $1/n_{g(i)}$ (or a minor variant of this weight), where $n_{g(i)}$ is the number of agents belonging to the group *g* of which *i* is a part. Manski (1993) uses the above specification of the interaction effects and assumes the absence of unobserved correlation effects, while Moffitt (2001) excludes the individual *i* in computing $n_{g(i)}$ and assumes that all groups have the same size. Lee (2007), on the other hand, uses the same specification as Mofitt (2001), but also allows the group sizes to vary across groups and accommodates a fixed unobserved group effect (that is, a common group unobserved effect that influences all the dependent variable outcomes in the group and may be correlated with the exogenous variables). Lee shows how the group size variation, along with the linkage of group sizes to overall sample size, allows the disentangling of social/spatial and exogenous interaction effects. Davezies *et al*. (2009) study the same formulation as Lee and show that identification is achieved even without the linking of group sizes to overall sample size, but as long as there are at least three different group sizes. Bramoulle *et al*. (2009) develop general necessary and sufficiency results for identification of social/spatial and exogenous interaction effects, and show how their results admit those of Manski, Mofitt, Lee, and Davezies as special cases. They further show that as soon as the weight matrix is not based on partitioning individuals into exclusive groups, but allows each individual to have her or his own network of reference individuals in a spatial/social network structure, this immediately is sufficient to allow identification (in an asymptotic sense) of spatial/social and exogenous interaction effects under the assumption that there are no unobserved correlation effects. They also extend their results to the case of fixed unobserved effects. However, in doing so, they go back to allocating individuals to exclusive groups and allowing group-specific unobserved effects (that generate correlation across members within the groups) while allowing a network interaction structure within the group. Using a differencing approach (similar to that used in linear panel data models), they eliminate the unobserved effects within each group (and avoid the incidental parameters problem) and then are able to disentangle the social/spatial interaction effects and the exogenous interaction effects, Lee *et al*. (2010) also study the case of multiple groups, with a spatial/social network autoregressive weight structure for the unobservables within each group to capture proximity-based preference correlations within the network as well as a group-specific unobservable effect to accommodate common environmental factors affecting individuals within each group. They eliminate the fixed effects using a specific approach they propose, and also show how the non-linearity introduced by such a network weight matrix (though now strictly within each group) facilitates identification of spatial/social interaction effects, exogenous interaction effects, and the unobserved autoregressive correlation effects within individuals within a group. It is important to note that Lee *et al*.'s approach is immediately applicable to the case of a single large network of individuals without any exclusive group assignment into smaller networks. The take-away is that, theoretically speaking, using a network structure for the social/spatial interactions and the exogenous interaction effects, as well as a (different) network structure for the unobserved correlation effects among agents, in general, will allow identification of all three effects – spatial/social interactions, exogenous interaction effects, and unobserved correlation effects.

Separate from the linear-in-means model studies discussed above, a handful of studies have examined identification considerations in discrete choice models (Brock and Durlauf, 2001, 2006, 2007; Soetevent and Kooreman, 2007; Krauth, 2006). These studies, like Manski (1993), Mofitt (2001), and Lee (2007), also consider strict group interaction effects (that is individuals are partitioned into exclusive groups, and interactions effects are confined strictly within groups). They formulate the underlying latent continuous propensity or utility of an individual for an alternative as a linear function of individual characteristics, the mean observed discrete choice share (or the mean expected discrete choice share) for the alternative in the group in which the individual belongs, and the mean exogenous characteristics of that group. These studies show that in the absence of unobserved group correlation effects, the social/spatial interaction effects can be disentangled from the exogenous interaction effects because, while the exogenous interaction effects linearly affect the utility of an alternative, the endogenous interaction effects are based off choice shares (and the transformation from the utilities to choice is non-linear). That is, the non-linearity inherent in the transformation from utility to choice probability can be exploited for identification purposes. The above holds for the case when the density function of the kernel error component defining the type of discrete choice model is pre-specified (such as an extreme value distribution), but also if the density function is left unspecified but with some additional assumptions. In the presence of unobserved group correlation effects, things become quite difficult and Brock and Durlauf (2007) identify specific conditions that may help identification in this case (such as, for example, the unimodality of the density functions of the unobserved group effects). Recently, Bhat *et al*. (2014a) discussed why identification of spatial interaction effects and exogenous interaction effects are possible in discrete choice and non-linear models using a different specification than the ones discussed above. Specifically, Bhat *et al*. use a specification in which the underlying continuous variables (alternative utilities in a discrete choice model or a propensity measure in ordinal/count models) that determine a non-continuous outcome is a linear function of exogenous variables, a weighted mean of exogenous variables (exogenous interaction effects), and a weighted mean of the underlying continuous variables of other individuals. They refer to the latter as the social/spatial interaction effect. The difference between this specification and the Brock and Durlauf-type specification is that Bhat *et al*.'s specification is essentially what is referred to as a spatial Durbin specification (see LeSage and Pace, 2011) in the spatial econometrics literature for continuous observed dependent variables, but now applied to the underlying continuous variables determining the non-continuous outcomes. The other difference between Bhat *et al*.'s specification and that of Brock and Durlauf is that Bhat *et al.* use a network weight matrix for the social/spatial as well as the exogenous interaction effects as opposed to the strict exclusive group formation specification used in Brock and Durlauf. However, Bhat *et al.*'s discussion regarding identification in their specification holds true even if the interaction effects were confined in strict groupings of individuals. Overall, there has been limited empirical work on network level interactions even in a linear context, and very little empirical work on network level interactions in a discrete choice context (much of the latter literature is from the spatial analysis field, but, until the recent introduction of a composite marginal likelihood approach by Bhat (see Bhat, 2011, Bhat *et al*., 2014a, Bhat, 2014) for accommodating network level spatial effects, it was not practical to estimate such models with moderate sized samples.

The studies above consider cross-sectional data. As soon as one has panel data, any unobserved group or network fixed effects in the case of strict group partitioning can be accommodated in a straightforward way by a differencing scheme in the linear models or using a conditional likelihood given sufficient statistics (see Chamberlain, 1984 and Bartolucii and Nigro, 2010). In this case, or the case of a single large network with panel data, it becomes easier and more stable to estimate unobserved correlation effects across individuals if we assume that the pattern of this unobserved correlation remains stable over time (effectively, we are able to control for individual-specific unobserved heterogeneity and allow network correlation effects across individuals through these unobserved heterogeneity terms).

To put the discussion above together, identification of each of the spatial/social interaction effects, exogenous variable effects, and unobserved correlation effects is theoretically feasible if a single large network is being considered in the linear model. When a discrete choice model is considered instead of a linear model, the non-linearity of such a model further facilitates theoretical identification. Finally, the presence of panel data can further help identify unobserved correlation effects in a network model if we assume that the pattern of the correlation effects do not change over time.

A caveat though to the above conclusion. While it is true that identification of the social/spatial interaction effects, exogenous interaction effects, and unobserved correlation effects can be facilitated by a network interaction structure, non-linear models, and panel data, disentangling these empirically can still be a challenge and can lead to imprecise estimation. Specifically, disentangling spatial/social interaction effects from exogenous interaction effects, particularly in the presence of unobserved correlation effects, can lead to empirical identification problems due to weak identification (see, for example, Elhorst, 2010 and Blume *et al*., 2011). That is, there is fundamental difficulty in empirically disentangling social/spatial interaction effects and exogenous interaction effects, particularly in the presence of unobserved correlation effects and regardless of the model specification. So, it is not surprising that most studies either control only for the unobserved correlation effects (with the implicit assumption of the absence of exogenous variable effects) or only for the exogenous variable effects (with the implicit assumption of the absence of unobserved correlation effects). In this study, we will control for unobserved correlation effects but assume the absence of exogenous variable effects, because of the difficulty in motivating exogenous variable effects in many transportation contexts.[1] Note also that the specification used in the current paper is based on social/spatial interaction effects and unobserved correlation effects operating at the level of underlying propensity variables dictating observed discrete outcomes. Thus, the implication of ignoring exogenous variable effects is that the propensity, for example, to use a specific mode for an individual is affected by the propensity to use that mode by those in close proximity of the individual, not directly by the income values or the built environments of those in close proximity of the individual.

[1] Our approach, in a traditional continuous dependent variable setting, leads to what is referred to as the Kelejian-Pruscha (KP) model (though, as we will discuss later, the model we propose is for an unordered-response discrete dependent variable and extends the KP model in important ways). The alternative of considering exogenous variable effects and ignoring unobserved correlation effects, in a traditional continuous dependent variable setting, is referred to as the spatial Durbin model. The reader is referred to LeSage and Pace, 2009 and Elhorst, 2010 for discussions. Both these authors suggest that, given the identification problems in introducing both these effects (along with the social/spatial interaction or spatial lag effect), there are benefits to starting with the spatial Durbin model as the most general specification. This is because ignoring spatial dependence in the errors in continuous variable models leads only to inefficiency loss, while ignoring spatial dependence in the exogenous variables leads to biased and inconsistent estimates due to the omitted variable problem. However, this discussion does not extend to discrete choice models, where the typical spatial dependence structure used for the error disturbance also adds to error heteroscedasticity, which leads to biased and inconsistent estimates of the discrete choice model. Further, as discussed in the main text, in many transportation contexts, it may be easier to motivate unobserved error correlation effects.

1.3. Unobserved Individual-Level Heterogeneity and Spatial Drift Effect

Unobserved individual heterogeneity refers to variations (due to unobserved factors) across decision agents in the preference for each alternative as well as in the responsiveness to exogenous variables. Thus, as discussed in Bhat and Sardesai (2006) and Pinjari and Bhat (2006) in a travel mode choice context, individuals may have different intrinsic preferences for specific modes due to unobserved lifestyle preferences and unobserved location factors, and may also have different sensitivities to exogenous variables such as travel times, costs, and residential built environment attributes (land-use mix, transit availability and accessibility, household density, employment density, bicycle facility density, street block density, *etc*.). For instance, a pro-bicycle attitude that does not change over time and is unobserved would lead to a higher than average utility for the bicycle alternative, while a time-conscious person would be highly sensitive to travel time during all her/his choice instances. This would then translate to an individual-specific random coefficients formulation, leading also to a stationary across-time correlation for the same individual in the case of panel or repeated mode choice data. Ignoring the presence of such unobserved heterogeneity will, in general, lead to biased and inconsistent parameters on all model parameters, including the social/spatial interaction effect. In addition, because of the spatial nature of decision agents' locations, some earlier studies have suggested that these unobserved individual-level heterogeneity effects should be correlated over decision agents based on the spatial (or social) proximity of the agents' locations, which is then referred to as spatial drift (see Bradlow *et al*., 2005 for a discussion). For example, Mittal *et al*. (2004) argues that locations in close proximity may share common climactic or lifestyle values that can lead to unobserved correlation in their sensitivity to specific marketing variables. On the other hand, almost all social interaction studies accommodate only what is referred to as a group unobserved effect in which individuals are segmented into exclusive groups with a fixed unobserved effect that commonly affects the outcomes of all individuals within the group (and also generates correlated unobserved effects among individuals within each group). Such a specification does not accommodate unobserved individual heterogeneity across individuals within the same group, and assumes no correlation effects across individuals of different groups.

In this paper, we move away from strictly groupwise interaction effects with equal intensity of interaction among individuals within a group (equivalent to a block diagonal weight matrix with equal values of all non-diagonal entries within each block-diagonal and zero values on the diagonal) to a very general single network interaction effects model with unequal intensities of interaction among all individuals (equivalent to a general weight matrix with non-equal values of non-diagonal entries and zero values on the diagonal). In spatial econometrics, the first type of model is referred to as a local interaction model, while the second is referred to as a global interaction model (Anselin, 2003). Our reason for the use of a global interaction model is that, in spatial contexts where the grouping is generally based on geographic areas (such as census tracts or block groups or neighborhoods), the assumption that that there would be no interaction effects or no unobserved correlation effects between two individuals very close to one another in space but in different geographic areas is difficult to defend. The use of predefined spatial pockets for capturing interaction effects and correlation effects also leads to what is referred to as the modifiable areal unit problem or MAUP (see Guo and Bhat, 2008 for a detailed discussion of this issue), which is substantially reduced by our use of a global network interaction structure. In addition, we adopt a general spatial drift specification in which we allow a network-based unobserved correlation structure not simply for the overall error term, but also for individual coefficients on exogenous variables. This correlation structure is not only due to unobserved location-specific unobservables that may be correlated over space, but also motivated from the perspective of self-selection in the social interactions literature (see Mofitt, 2001 and Hartman *et al*., 2008). The resulting self-selection, if not controlled for, can manifest itself as social/spatial interactions. For example, again in the travel mode choice context, households and individuals who intrinsically prefer walking or bicycling (say due to their environmental consciousness) may locate themselves in close proximity of one another, because of unobserved factors such as good walk and bicycle path continuity of their immediate neighborhood. Similarly, households and individuals who are transit-oriented may be drawn toward locations with such built environment factors as good land-use mixing, high built-up density, and good transit availability and accessibility. The net result in the examples above is that individuals and households with similar mode use propensities and sensitivity to observed built environment attributes may be in close proximity, but this is not a result of social/spatial interactions <u>after</u> locating in a neighborhood or the causal influence of built environment attributes. That is, there is the possibility of residential self-selection of individuals based on mode-use propensities. If this residential self-selection is ignored, it can incorrectly manifest itself as a spatial lag effect based on dyadic interactions and/or built environment effects.

Interestingly, there have been studies in the transportation and other fields that accommodate residential self-selection effects due to clustering based on travel preferences, but these have not considered spatial interaction (lag) effects (see for example, Bhat and Guo, 2007, Mokhtarian and Cao, 2008, Tsai, 2009, and Bhat *et al*., 2014b). These studies run the reverse risk that true spatial lag effects can be manifested as spurious self-selection effects, completely ignoring the social multiplier effect of transportation and land-use policies. Of course, ignoring one or both of the self-selection and spatial lag effects will, in general, provide incorrect estimates of the "true" effects of neighborhood attributes on travel mode, which can potentially lead to misinformed built environment design policies.[2]

1.4. The Current Study and its Many Dimensions

The current study is set in the context of examining commute mode choice using multi-day data, though the model formulation is applicable for any other unordered-response choice situation. In doing so, we accommodate several econometric aspects relevant to spatial panel unordered discrete choice models.

Figure 1 show the different components of a comprehensive mode choice model. The elements at the bottom of the figure represent non-spatial components, while the three elements at the top of the figure represent the spatial components. The next two sections discuss these components in more detail.

1.4.1. The non-spatial components

The first non-spatial component (first box in the bottom row of Figure 1) refers to the direct effect of exogenous variables (referred to as built environment effects, but can also represent other exogenous variables in the model). The second and third non-spatial components in the bottom row correspond to the choice-occasion specific error terms in the modal utilities, with the second one capturing the cross-alternative covariance in the utilities (due to heteroscedasticity and dependence in the random utilities across alternatives) at each choice occasion of the individual. In essence, this component allows the presence of unobserved factors that engenders correlation in the utilities of modes at each choice occasion of each individual (such as, for

[2] Econometrically speaking, relative to Lee *et al*. (2010), our paper differs in that we allow a global network structure (equivalently, a single group in Lee *et al*.'s study), and also accommodate a network-based unobserved correlation structure for the overall error term as well as for individual coefficients on exogenous variables.

example, the need to travel with others simultaneously increasing or decreasing the modal utilities ascribed to the public transportation alternatives) and allows the magnitude of unobserved effects to vary across alternatives at each choice occasion of each individual (for example, ride comfort may vary substantially for public transportation based on the specific equipment used to serve a route, while ride comfort may be more uniform for the car mode). The third component entitled "cross-time fading unobserved preference" accommodates time-varying dependency effects across the utilities of the same individual at different points in time. These time-varying effects may be attributed to the effects of recent experiences and events that influence the environmental or other perceptions of individual agents. As such, these effects fade over time, with the perceptions/attitudes at a particular time being much more affected by perceptions/attitudes in the recent past than those from sometime back (for example, a sociable person may prefer public transportation modes more so than her/his peers for all choice occasions, but the person's sociable characteristic may also change over time resulting in fading correlation in public transportation utility over time). Note that these cross-time fading unobserved preference effects are being introduced in the model structure so that the formulation is generic, and is applicable to both panel as well as repeated choice data. In general, the inclusion of these time dependency effects will be more appropriate for panel data than for repeated choice data over a very short time frame. The fourth and fifth components at the bottom of Figure 1 correspond to individual-specific unobserved preference and unobserved response sensitivity, both of which are time-invariant across the many choice instances of the same individual (see Section 1.3). The presence of multi-day data enables the estimation of such individual-specific unobserved variations in preference and response sensitivity, which also generate a time-invariant dependency in the utilities of each mode across the repeated choices of the same individual.

1.4.2. The spatial components

The accommodation of the non-spatial components leads to a general aspatial panel unordered model. In this paper, we superimpose three types of spatial effects, as discussed intuitively in Sections 1.1 and 1.3, and econometrically below.

The spatial/social dependencies (originating from dyadic interactions as motivated in Section 1.1) among the choices of spatially proximate decision-makers are considered by assuming a spatial (autoregressive) lag structure (across individuals) for the utility function of each mode (labeled as the spatial lag effect in Figure 1). Such spatial lag structures have seldom been considered in the spatial econometric literature for unordered-response discrete choice models, because they require the evaluation of a multi-dimensional integral of the order of the number of decision agents multiplied by the number of alternatives less one. In the past five years or so, researchers have started to find ways to break the impasse in estimating spatial unordered models. Examples include Carrión-Flores and Irwin (2004), Smirnov (2010), Chakir and Parent (2009), and Sidharthan and Bhat (2012). Among these, the first two avoid multidimensional integration through the use of a two-step instrumental variable estimation technique after linearizing around zero interdependence, and so work well only for the case of large estimation sample sizes and weak spatial dependence. Chakir and Parent (2009) employ a Bayesian MCMC method, which requires extensive simulation, is time-consuming, is not straightforward to implement, and can create convergence assessment problems (Franzes *et al*., 2010). Sidharthan and Bhat (2012) use Bhat's (2011) maximum approximate composite marginal likelihood (MACML) approach that is based on a combination of an analytic approximation for the evaluation of the multivariate cumulative normal density (MVNCD) function and a composite marginal likelihood (CML) approach. They also show that the MACML approach is able to recover parameters accurately in simulation experiments. In this paper, we build upon the spatial lag specifications of these earlier studies by adding other spatial considerations in the model structure.

In addition to the spatial lag effect, the formulation in the paper allows the presence of a spatial structure on the individual-specific unobserved factors that influence the overall valuation (or utility) of each mode by each individual (labeled as "spatial structure on individual-specific unobserved preference" in Figure 1). We do so by employing a spatial autoregressive structure for the random effects, which captures both unobserved location-specific unobservables impacting modal outcomes that may be correlated over space, and individuals locating in space based on unobservable modal preferences (see discussion in Section 1.3). This is the traditional structure used for the spatial error model in spatial econometrics (see Pace and LeSage, 2011), though we combine this with a spatial lag structure in a Kelejian-Prucha-like specification and apply the specification to repeated choice data as opposed to cross-sectional data. As in the case of the pure spatial lag model, there have been very few applications of the spatial error model for unordered discrete choice models, again because of the dimensionality of integration involved in the resulting likelihood function. Bhat *et al.* (2010) and Sener and Bhat (2012) estimate spatial error models for unordered choice situations, but the copula approach they use is not applicable to cases where there is also a spatial lag effect nor is the copula approach easily extended to the case when spatial drift effects also are to be considered. On the other hand, the MACML method used by Sidharthan and Bhat (2012) for the spatial lag model is applicable in a straightforward way to accommodate spatial correlation effects (see Bhat *et al*., 2013 for additional details).

Finally, as discussed earlier in Section 1.3, individuals are likely to vary in their sensitivities to relevant exogenous determinants based on unobserved location-specific factors and/or self-selection driven clustering of individuals associated with unobserved lifestyle preferences. These individual-specific response variations are likely to have a spatial correlation pattern based on the residential location of individuals (this effect is labeled as "Spatial structure on individual-specific response sensitivity" in Figure 1). Such a spatially-structured response heterogeneity effect is accommodated through a spatial autoregressive structure on the random coefficients of variables (the spatial correlation effect, along with the spatially-structured response heterogeneity effect, represent the spatial drift effect, as shown in Figure 1). This model structure is more general than the Kelejian-Prucha structure (Elhorst, 2010) that includes a spatial lag structure and an additional autoregressive structure but only on the constant parameter (this is what we have labeled as "spatial structure on individual-specific unobserved preference" in Figure 1, and discussed in the previous paragraph). That is, the Kelejian-Prucha structure considers only the spatial lag and spatial correlation effects, but not the spatially-structured response heterogeneity effect. Additionally, we develop our structure for unordered discrete choice models, while the Kelejian-Prucha structure is for linear or ordered-response models.[3]

The approach we propose is also novel in its accommodation of residential self-selection through a spatial drift interpretation. In contrast, the typical approach to accommodate residential

[3] In the spatial literature, spatial drift effects have been typically incorporated using the geographically weighted regression (GWR) approach of Brunsdon *et al*. (1998) (some other approaches, such as spatial adaptive filtering and multi-level modeling, are either too *ad hoc* or too restrictive in capturing spatial drift effects, and are not often used; see Mittal *et al*., 2004 for a discussion). The GWR approach allows spatial dependence in parameters based on spatial proximity, but is not able to disentangle the spatial drift effects from the direct parameter effect. On the other hand, being able to do so is important in many cases. For example, in our mode choice context, the effects of neo-urbanist developments on mode shares, net of residential self-selection effects (as captured by spatial drift), is important in its own right. Our formulation explicitly and directly captures spatial dependence in the random coefficients and is able to isolate the mean direct effect of each exogenous variable, while also controlling for spatial lag and spatial error effects.

self-selection has been to jointly model residential location and travel choices (see Bhat and Guo, 2007 and Mokhtarian and Cao, 2008 for detailed reviews of the approaches used to accommodate self-selection effects). In these earlier approaches, it is literally infeasible to also consider spatial lag effects that may be at play in specific empirical contexts. At the same time, we have to acknowledge that, in the current study, the emphasis is on the modal choice outcome and the disentangling of spatial/social interaction effects, unobserved correlation effects and residential self-selection effects, and "true" built environment effects on mode choice. The residential location pattern itself is considered exogenous (that is, the weight matrix W used to capture social/spatial interaction patterns and spatial drift patterns in assumed exogenous).[4] In other words, the effects of changes in exogenous variables is assessed on mode choice, keeping the residential pattern fixed (even if the notion that individuals locate themselves in part based on unobserved-to-the-analyst modal preferences and tastes is accommodated; this is self-selection with respect to unobservable variables). This approach may be justified from the notion that modal choices and residential choices are made at different timescales (see also Blume *et al*., 2011 for a similar argument). Thus, an improvement in public transportation service or an improvement in bicycle infrastructure in selected corridors or areas will likely impact modal choice before potentially influencing residential choices. But in estimating the mode choice model, we accommodate for the possibility that unobserved modal choice preferences and sensitivities to modal choice determinants may themselves have affected (in part) residential choice patterns. This is achieved through the spatial drift effects. In assessing any social/spatial interaction effects and built environment effects on modal choices, these unobserved correlation effects in preferences/sensitivity across individuals based on residential choice have to be controlled for. In the longer run, the analyst can have a separate model of residential choice, followed by the mode choice model proposed here, to obtain the combined long-term effects based on residential location changes as well as modal changes.

[4] In the terminology of the social network formation literature, we accommodate self-selection effects caused by residential patterns in modeling modal choice, though we do not jointly model the coevolution of residential network formation (that is, the W matrix) and the modal outcome. That is, the network literature (see, for example, Steglich *et al*., 2010 and Christakis and Fowler, 2013) considers the network formation and behavior outcomes as closely intertwined and evolving over time, with each influencing the other in a dynamic fashion. We do not examine such a dynamic system of network and behavior coevolution in this paper.

1.5. Structure of the Paper

This paper, at once, combines spatial lag effects, spatial structure on individual-specific unobserved preference and response heterogeneity effects, and time series effects in a multinomial probit model formulation. In addition, the formulation is framed in a panel-type setting to estimate a model from multiple choice instance data from the same individual.

The rest of this paper is structured as follows. In the next section, we present the proposed model formulation. We start our model development by first incorporating the non-spatial components, followed by the spatial components. Section 3 discusses the estimation procedure and identification considerations. Section 4 then describes a simulation study undertaken to demonstrate the functionality of our formulation and its ability to recover spatial parameters. Finally, Section 5 concludes the paper and propose avenues for future research.

2. MODEL FORMULATION

In the following presentation, we will use the label "individual" to refer to a worker who has to make a mode choice decision for her/his commute. $\mathbf{IDEN}_I$ will refer to an identity matrix of size I, $\mathbf{1}_T$ will refer to a $(T \times 1)$ column vector of ones, and $\mathbf{1}_{TT}$ to a $(T \times T)$ matrix of ones. For ease in presentation, we will also assume that all alternatives are available to all individuals, and that the number of choice instances is the same across all individuals (these assumptions are innocuous and can be easily relaxed, but make the notations easier). Let the utility U_{qti} accrued by individual q (q = 1, 2, …, Q) during choice occasion t (t = 1, 2, …, T) for alternative i (i = 1, 2, …, I) be a function of a $(K \times 1)$ column vector of exogenous variables $\boldsymbol{x}_{qti}$ (excluding a constant). In the usual utility maximization decision principle, the individual chooses the alternative that provides her/him the highest utility at any choice occasion. That is, individual q selects alternative i at choice occasion t if $U_{qti} > U_{qtj} \;\forall\; j \neq i.$ As in the typical random utility maximization formulation, the individual is assumed to be aware of all factors that impact her/his utility, but some of these factors are unobserved to the analyst (that is, U_{qti} is random from the perspective of the analyst because of the presence of unobserved components).

2.1. Incorporating Built Environment Effects, Cross Alternative Choice Occasion-Specific Covariance, and Cross-Time Fading Unobserved Preference (the first three non-spatial components in Figure 1)

Consider the following utility structure:

$$U_{qti} = \tilde{a}_i + \boldsymbol{b}'\boldsymbol{x}_{\boldsymbol{qti}} + \tilde{\varepsilon}_{qti} \tag{1}$$

where $\tilde{a}_i$ represents the generic preference across all individuals for alternative i, $\boldsymbol{b}$ is a $(K \times 1)$ vector of coefficients, and $\tilde{\varepsilon}_{qti}$ is a normally distributed choice instance-specific error term uncorrelated with $\boldsymbol{x}_{\boldsymbol{qti}}$ and also uncorrelated across individuals. The vector $\boldsymbol{b}$ captures the built environment effects (along with the effects of other exogenous variables contained in $\boldsymbol{x}_{\boldsymbol{qti}}$). Next, to allow a general cross alternative choice occasion-specific covariance structure on the $\tilde{\varepsilon}_{qti}$ terms as well as a cross-time fading unobserved preference component, consider a first order autoregressive (AR1) temporal dependency process:

$$\tilde{\varepsilon}_{qti} = \rho\tilde{\varepsilon}_{q,t-1,i} + \tilde{\eta}_{qti}, \tag{2}$$

where ρ $(0 < \rho < 1)$ is a temporal autoregressive parameter. The error term $\tilde{\eta}_{\boldsymbol{qti}}$ is temporally uncorrelated, but can be correlated across modes due to unobserved factors at time t that simultaneously increase or decrease the utility of specific combinations of modes: $\tilde{\boldsymbol{\eta}}_{\boldsymbol{qt}} = (\tilde{\eta}_{qt1}, \tilde{\eta}_{qt2}, \ldots, \tilde{\eta}_{qtI})' \sim MVN_I(\mathbf{0}_I, \tilde{\boldsymbol{\Psi}})$, where $MVN_I(0, \tilde{\boldsymbol{\Psi}})$ refers to a multivariate normal distribution of dimension I with mean $\mathbf{0}_I$ and covariance $\tilde{\boldsymbol{\Psi}}$. The covariance matrix $\tilde{\boldsymbol{\Psi}}$ also allows the variance of the utilities to vary across alternatives at each choice occasion (see Section 1.4.1). Note also that, in Equation (2), we allow the preference for each alternative to be correlated across choice occasions of the individual, though there is a fading correlation based on the duration between choice occasions (see Section 1.4.1). If $\rho = 0$, the implication is that there is no cross-time fading unobserved preference component across the choice instances of the same individual, and the model in Equation (1) collapses to a simple cross-sectional MNP model. On the other hand, if $\rho \neq 0$ but $\tilde{\boldsymbol{\Psi}} = \mathbf{IDEN}_I$, the model in Equation (1) is a panel MNP model but with an independent and identically distributed (IID) error structure at each choice instance of the individual.

2.2. Incorporating Time-Invariant Individual-Specific Unobserved Preference and Response Sensitivity (the final two non-spatial components in Figure 1)

Individuals are likely to be characterized by some unobserved personality traits that are time-invariant (at least over the span of the data collection period of a panel) and that impact the utilities they associate with alternatives at each choice instance as well as their responsiveness to relevant exogenous variables (see discussion in Sections 1.3 and 1.4.1). This may be captured as follows:

$$U_{qti} = \tilde{a}_i + \breve{\alpha}_{qi} + (\boldsymbol{b} + \tilde{\boldsymbol{\beta}}_{\boldsymbol{q}})'\boldsymbol{x}_{\boldsymbol{qti}} + \rho\tilde{\varepsilon}_{q,t-1,i} + \tilde{\eta}_{qti} = \tilde{\alpha}_{qi} + \boldsymbol{\beta}_q'\boldsymbol{x}_{\boldsymbol{qti}} + \rho\tilde{\varepsilon}_{q,t-1,i} + \tilde{\eta}_{qti} \tag{3}$$

where $\tilde{\alpha}_{qi} = \tilde{a}_i + \breve{\alpha}_{qi}$ with $\tilde{a}_i$ representing the average (across individuals) inherent preference for alternative i and $\breve{\alpha}_{qi}$ representing the deviation of individual *q's* unobserved preference for alternative i from that of the average, and $\boldsymbol{\beta}_{\boldsymbol{q}} = \boldsymbol{b} + \tilde{\boldsymbol{\beta}}_{\boldsymbol{q}}$ with $\boldsymbol{b}$ representing the average (across individuals) response sensitivity vector and $\tilde{\boldsymbol{\beta}}_{\boldsymbol{q}}$ representing the deviation vector of individual *q's* responsiveness from that of the average. To complete the specification, assume a multivariate normal distribution for $\breve{\boldsymbol{\alpha}}_q = (\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, ..., \breve{\alpha}_{qI}) \sim MVN_I(\mathbf{0}_I, \vec{\boldsymbol{\Lambda}})$ and for $\tilde{\boldsymbol{\beta}}_{\boldsymbol{q}} \sim MVN_K(\mathbf{0}_K, \vec{\boldsymbol{\Omega}})$.

2.3. Incorporating Social/Spatial Interaction (first spatial component in Figure 1)

As discussed in Sections 1.1 and 1.4.2, we incorporate social/spatial dependencies through a spatial lag specification for the utilities as follows:

$$U_{qti} = \delta \sum_{q'} w_{qq'} U_{q'ti} + \tilde{\alpha}_{qi} + \boldsymbol{\beta}_q'\boldsymbol{x}_{\boldsymbol{qti}} + \rho\tilde{\varepsilon}_{q,t-1,i} + \tilde{\eta}_{qti} \tag{4}$$

where δ $(0 < \delta < 1)$ is the spatial lag autoregressive parameter, $w_{qq'}$ is the usual distance-based spatial weight corresponding to individuals q and q' ($w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$ for all individuals q). The weight $w_{qq'}$ is assumed, as is typical in spatial econometrics, to converge to zero as the social/spatial distance between individuals q and q' tends to infinity.

2.4. Incorporating Spatial Drift (or Self-Selection) Effects (the final two spatial components in Figure 1)

The spatial drift effects (discussed in Sections 1.3 and 1.4.2) are captured by allowing a structured dependency pattern in both the intrinsic mode-use preferences and the sensitivity to mode specific exogenous variables (*i.e.*, in the $\tilde{\alpha}_{qi}$ terms for each alternative *i* and in the $\boldsymbol{\beta}_{\boldsymbol{q}}$ vector of Equation (4)). To do so, we use a spatial autoregressive structure for $\breve{\alpha}_{qi}$: $\breve{\alpha}_{qi} = \theta \sum_{q'} w_{qq'} \breve{\alpha}_{qi} + \tilde{\tau}_{qi}$, where θ $(0 < \theta < 1)$ is a residential self-selection parameter that captures the intensity of residential clustering based on generic mode preferences. $\tilde{\tau}_{qi}$ is an unobserved individual-specific error term for mode *i* that is independent across individuals, but can be correlated across modes so that $\tilde{\boldsymbol{\tau}}_q = (\tilde{\tau}_{q1}, \tilde{\tau}_{q2}, ..., \tilde{\tau}_{qI})' \sim MVN_I(\mathbf{0}_I, \tilde{\boldsymbol{\Lambda}})$. Similar to the spatial structure on the unobserved preferences, we assume a spatial autoregressive structure on the unobserved sensitivities for each variable *k*: $\beta_{\boldsymbol{q}k} = b_k + \tilde{\beta}_{qk}$, where b_k is the mean effect of the *k*th variable in the $\boldsymbol{x}_{qti}$ vector, and $\tilde{\beta}_{qk} = \lambda_k \sum_{q'} w_{qq'} \tilde{\beta}_{q'k} + \tilde{\gamma}_{qk}$. λ_k $(0 < \lambda_k < 1)$ is the residential self-selection parameter that captures the intensity of residential clustering based on the unobserved sensitivities for the attribute represented by the *k*th variable in the $\boldsymbol{x}_{qti}$ vector, and $\tilde{\gamma}_{qk}$ is an individual-specific normally distributed term capturing unobserved sensitivity to the *k*th variable in the $\boldsymbol{x}_{qti}$ vector. Define $\tilde{\boldsymbol{\gamma}}_q = (\tilde{\gamma}_{q1}, \tilde{\gamma}_{q2}, ..., \tilde{\gamma}_{qK})$, and assume that $\tilde{\boldsymbol{\gamma}}_q$ is a realization of a multivariate normal distribution with a mean vector of zeros and covariance matrix $\tilde{\boldsymbol{\Omega}}$: $\tilde{\boldsymbol{\gamma}}_q \sim MVN_K(\mathbf{0}_K, \tilde{\boldsymbol{\Omega}})$.

2.5. Matrix Form of Model

We now put together Equation (4) with the spatial drift effects discussed in the previous section, and write the resulting equation in matrix form. To do so, define $\tilde{\mathbf{A}}$ as an $(I \times 1)$ vector whose elements are $\tilde{\mathbf{A}} = (\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_I)'$ and $\mathbf{W}$ as the $(Q \times Q)$ spatial weight matrix with elements $w_{qq'}$ Further, also define $\breve{\boldsymbol{\alpha}}_q$ as an $(I \times 1)$ vector $\breve{\boldsymbol{\alpha}}_q = (\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, ..., \breve{\alpha}_{qI})'$, $\breve{\boldsymbol{\alpha}}$ as a $(QTI \times 1)$ vector $\breve{\boldsymbol{\alpha}} = [(\mathbf{1}_T \otimes \breve{\boldsymbol{\alpha}}_1)', (\mathbf{1}_T \otimes \breve{\boldsymbol{\alpha}}_2)', ..., (\mathbf{1}_T \otimes \breve{\boldsymbol{\alpha}}_Q)']'$, and $\tilde{\boldsymbol{\tau}}$ as a $(QTI \times 1)$ vector

$\tilde{\boldsymbol{\tau}} = \left[(\mathbf{1}_T \otimes \tilde{\boldsymbol{\tau}}_1)', (\mathbf{1}_T \otimes \tilde{\boldsymbol{\tau}}_2)', ..., (\mathbf{1}_T \otimes \tilde{\boldsymbol{\tau}}_Q)'\right]'$. Then $\breve{\boldsymbol{\alpha}} = \left[\theta \mathbf{W} \otimes (\mathbf{1}_{TT} \otimes \mathbf{IDEN}_I)\right]\breve{\boldsymbol{\alpha}} + \tilde{\boldsymbol{\tau}}$. Letting $\mathbf{G}$ be the $(QTI \times QTI)$ matrix given by $\mathbf{G} = \left[\mathbf{IDEN}_{QTI} - \theta \mathbf{W} \otimes \left(\mathbf{1}_{TT} \otimes \mathbf{IDEN}_I\right)\right]^{-1}$, we can write $\breve{\boldsymbol{\alpha}} = \mathbf{G}\,\tilde{\boldsymbol{\tau}}$. Next, let $\tilde{\boldsymbol{\gamma}} = \left(\tilde{\boldsymbol{\gamma}}_1, \tilde{\boldsymbol{\gamma}}_2, ..., \tilde{\boldsymbol{\gamma}}_Q\right)'$ $(QK \times 1)$ vector, $\boldsymbol{b} = \left(b_1, b_2, ..., b_K\right)'$ $(K \times 1)$ vector, $\tilde{\boldsymbol{\beta}}_{\boldsymbol{q}} = \left(\tilde{\beta}_{q1}, \tilde{\beta}_{q2}, ..., \tilde{\beta}_{qk}\right)'$ $(K \times 1)$ vector, $\tilde{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\beta}}_2', ..., \tilde{\boldsymbol{\beta}}_Q'\right)'$ $(QK \times 1)$ vector, and let $\boldsymbol{\lambda}$ be a $(K \times K)$ diagonal matrix with elements λ_k. Then, $\tilde{\boldsymbol{\beta}} = \left[\mathbf{W} \otimes \boldsymbol{\lambda}\right]\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\gamma}}$. Let $\mathbf{D}$ be the $(QK \times QK)$ matrix given by $\mathbf{D} = \left(\mathbf{IDEN}_{QK} - \left[\mathbf{W} \otimes \boldsymbol{\lambda}\mathbf{IDEN}_K\right]\right)^{-1}$. Then, we can write $\tilde{\boldsymbol{\beta}} = \mathbf{D}\tilde{\boldsymbol{\gamma}}$.

We will now write Equation (4) in matrix form, using previously defined notation as well as several additional matrices: $\mathbf{x}_{qt} = (\boldsymbol{x}_{qt1}, \boldsymbol{x}_{qt2}, ..., \boldsymbol{x}_{qtI})'$ $(I \times K)$ matrix, $\mathbf{x}_q = (\mathbf{x}_{q1}', \mathbf{x}_{q2}', ..., \mathbf{x}_{qT}')'$ $(TI \times K)$ matrix, $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_Q')'$ $(QTI \times K)$ matrix, $\mathbf{U}_{qt} = (U_{qt1}, U_{qt2}, ..., U_{qtI})'$ $(I \times 1)$ vector, $\mathbf{U}_q = (\mathbf{U}_{q1}', \mathbf{U}_{q2}', ...\mathbf{U}_{qT}')'$ $(TI \times 1)$ vector, $\mathbf{U} = (\mathbf{U}_1', \mathbf{U}_2', ...\mathbf{U}_Q')'$ $(QTI \times 1)$ vector, $\tilde{\boldsymbol{\eta}}_{qt} = (\tilde{\eta}_{qt1}, \tilde{\eta}_{qt2}, ..., \tilde{\eta}_{qTI})'$ $(I \times 1)$ vector, $\tilde{\boldsymbol{\eta}}_q = (\tilde{\boldsymbol{\eta}}_{q1}', \tilde{\boldsymbol{\eta}}_{q2}', ...\tilde{\boldsymbol{\eta}}_{qT}')'$ $(TI \times 1)$ vector, $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\eta}}_1', \tilde{\boldsymbol{\eta}}_2', ...\tilde{\boldsymbol{\eta}}_Q')'$ $(QTI \times 1)$ vector. Further, define the following matrices:

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} (T \times T \text{ matrix}),\ \tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 & 0 & 0 & 0. & \cdots & 0 \\ 0 & \mathbf{x}_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{x}_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \mathbf{x}_Q \end{bmatrix} (QTI \times QK \text{ matrix}),$$

$\mathbf{C} = \left[\mathbf{IDEN}_{QTI} - \mathbf{IDEN}_Q \otimes \left(\rho\mathbf{R} \otimes \mathbf{IDEN}_I\right)\right]^{-1} = \mathbf{IDEN}_Q \otimes \left(\left[\mathbf{IDEN}_{TI} - \left(\rho\mathbf{R} \otimes \mathbf{IDEN}_I\right)\right]^{-1}\right)$ $(QTI \times QTI)$ matrix, $\mathbf{S} = \left[\mathbf{IDEN}_{QTI} - \{(\delta\mathbf{W} \otimes \mathbf{IDEN}_T) \otimes \mathbf{IDEN}_I\}\right]^{-1}$ $(QTI \times QTI)$ matrix, $\mathrm{Var}(\tilde{\boldsymbol{\tau}}) = \mathbf{IDEN}_Q \otimes \left(\mathbf{1}_{TT} \otimes \tilde{\boldsymbol{\Lambda}}\right)$, $\mathrm{Var}(\tilde{\boldsymbol{\gamma}}) = \mathbf{IDEN}_Q \otimes \tilde{\boldsymbol{\Omega}}$, $\boldsymbol{\Lambda} = \mathbf{G}\,\mathrm{Var}(\tilde{\boldsymbol{\tau}})\mathbf{G}'$, $\boldsymbol{\Psi} = \mathbf{IDEN}_{QT} \otimes \tilde{\boldsymbol{\Psi}}$ $(QTI \times QTI$ matrix$)$ and $\boldsymbol{\Omega} = \tilde{\mathbf{x}}\mathbf{D}\,\mathrm{Var}\left(\tilde{\boldsymbol{\gamma}}\right)\left(\mathbf{D}\tilde{\mathbf{x}}\right)'$. Then Equation (4) in matrix notation takes the following form:

$$\mathbf{U} = \mathbf{S}\left[\left(\mathbf{1}_{\mathbf{QT}} \otimes \tilde{\mathbf{A}}\right) + \mathbf{x}\boldsymbol{b} + \mathbf{G}\tilde{\boldsymbol{\tau}} + \tilde{\mathbf{x}}\mathbf{D}\tilde{\boldsymbol{\gamma}} + \mathbf{C}\tilde{\boldsymbol{\eta}}\right] \sim MVN_{I \times T \times Q}(\tilde{\mathbf{B}}, \tilde{\boldsymbol{\Xi}}), \tag{5}$$

where $\tilde{\mathbf{B}} = \mathbf{S}\left[(\mathbf{1}_{\mathbf{QT}} \otimes \tilde{\mathbf{\Lambda}}) + \mathbf{x}\boldsymbol{b}\right]$ and $\tilde{\mathbf{\Xi}} = \mathbf{S}\left[\mathbf{\Lambda} + \mathbf{\Omega} + \mathbf{C\Psi C'}\right]\mathbf{S'}$. In the equation above, $\boldsymbol{b}$ represents the effects of built environment and other observed modal/individual attributes (the first non-spatial component), $\mathbf{\Psi}$ captures the cross-alternative choice occasion specific covariance (the second non-spatial component), $\mathbf{C}$ captures the cross-time fading unobserved preference effect (the third non-spatial component), $\mathbf{G}\tilde{\boldsymbol{\tau}}$ captures the combination of unobserved preference heterogeneity (the fourth non-spatial component) and the structured spatial drift dependency pattern in the intrinsic mode-use preferences (the second spatial component), $\tilde{\mathbf{x}}\mathbf{D}\tilde{\boldsymbol{\gamma}}$ captures the combination of unobserved response heterogeneity (the fifth non-spatial component) and the structured spatial drift dependency in the sensitivity to mode specific exogenous variables (the third spatial components), and the matrix $\mathbf{S}$ appears because of the spatial/social interaction effect (the first spatial component).

The model proposed in this paper is very general, and nests several other spatial models in the literature as special cases. Table 1 illustrates these restricted versions of our general model. By using nested statistical tests, these restricted versions can be tested against our general model.

2.6. Effects of Exogenous Variables

The end-objective of models of the type discussed in the previous section is to examine the impact of exogenous variables on the discrete outcome of interest. In the setting that motivated the current model formulation, once estimated, the parameters of the model can be used to forecast the effect on mode choices of changing demographics or built environment variables embedded in the $\boldsymbol{x}$ vector. It can also be used by policy makers in different ways to examine a change in a variable for one individual on the mode choice probabilities of that individual (direct effect), and on the mode choice probabilities of other individuals (indirect effect). But, to summarize these effects, it is typical to compute "average" effects, as we discuss later. Further, the effects themselves can be computed in several ways. Here, we propose doing so in a way that is generalizable to any explanatory variable (whether it is a continuous explanatory variable or not) and to any magnitude of change in the explanatory variable (the procedure suggested in LeSage and Pace, 2009 and LeSage *et al*. (2011), on the other hand, is specific to continuous explanatory variables and to an infinitesimal change in an explanatory variable).

In many panel models of the type investigated here, the panel setting arises from a repeated choice situation (such as commute mode choice over different days of the week, or mode choice in a series of stated preference questions). This can be used to provide stable estimates of unobserved individual-specific unobserved heterogeneity effects, spatial drift effects, and social/spatial interaction effects. Once estimated, the choice occasion index t can be discarded (with ρ=0), and the utility function of individual q at any choice occasion may be written as:

$$U_{qi} = \delta \sum_{q'} w_{qq'} U_{q'i} + \widetilde{\alpha}_{qi} + \boldsymbol{\beta}_q' \boldsymbol{x}_{\boldsymbol{qi}} + \widetilde{\eta}_{qi};\ \widetilde{\alpha}_{qi} = \widetilde{a}_i + \breve{\alpha}_{qi},\ \boldsymbol{\beta}_q = (\boldsymbol{b} + \widetilde{\boldsymbol{\beta}}_{\boldsymbol{q}}), \tag{6}$$

with $\breve{\alpha}_{qi} = \theta \sum_{q'} w_{qq'} \breve{\alpha}_{qi} + \widetilde{\tau}_{qi}$, and $\widetilde{\beta}_{qk} = \lambda_k \sum_{q'} w_{qq'} \widetilde{\beta}_{q'k} + \widetilde{\gamma}_{qk}$. All notations are preserved from Section 2.4. Next define the following (for ease in presentation, we maintain the same notations as in Section 2.4 for the re-defined vectors and matrices):

$\mathbf{U}_q = (U'_{q1}, U'_{q2}, ..., U'_{qI})'$ and $\widetilde{\boldsymbol{\eta}}_q = (\widetilde{\eta}'_{qi}, \widetilde{\eta}'_{q2}, ..., \widetilde{\eta}'_{qI})'$ ($I \times 1$ vectors),

$\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2, ..., \mathbf{U}'_Q)'$ and $\widetilde{\boldsymbol{\eta}} = (\widetilde{\boldsymbol{\eta}}'_1, \widetilde{\boldsymbol{\eta}}'_2, ..., \widetilde{\boldsymbol{\eta}}'_Q)'$ ($QI \times 1$ vectors),

$\breve{\boldsymbol{\alpha}}_q = (\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, ..., \breve{\alpha}_{qI})'$ ($\mathrm{I} \times 1$ vector), $\breve{\boldsymbol{\alpha}} = \left[(\breve{\alpha}_1)', (\breve{\alpha}_2)', ..., (\breve{\alpha}_Q)'\right]'$ ($QI \times 1$ vector),

$\widetilde{\boldsymbol{\tau}} = \left[\widetilde{\boldsymbol{\tau}}_1', \widetilde{\boldsymbol{\tau}}_2', ..., \widetilde{\boldsymbol{\tau}}_Q'\right]'$ ($QI \times 1$ vector), $\widetilde{\boldsymbol{\gamma}} = \left(\widetilde{\boldsymbol{\gamma}}_1, \widetilde{\boldsymbol{\gamma}}_2, ..., \widetilde{\boldsymbol{\gamma}}_Q\right)'$ $(QK \times 1)$ vector

$$\mathbf{x}_q = (\mathbf{x}_{q1}, \mathbf{x}_{q2}, ..., \mathbf{x}_{qI})' \ (I \times K \text{ matrix}),\ \mathbf{x} = (\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_Q)' \ (QI \times K \text{ matrix}), \tag{7}$$

$$\widetilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 & 0 & 0 & 0. \cdots 0 \\ 0 & \mathbf{x}_2 & 0 & 0 \cdots 0 \\ 0 & 0 & \mathbf{x}_3 & 0 \cdots 0 \\ \vdots & \vdots & \vdots & \vdots \ldots \vdots \\ 0 & 0 & 0 & 0 \ldots \mathbf{x}_Q \end{bmatrix} (QI \times QK \text{ matrix}),$$

$\mathbf{S} = \left[\boldsymbol{IDEN_{QI}} - (\delta \boldsymbol{W} \otimes \boldsymbol{IDEN_I})\right]^{-1}$ ($QI \times QI$ matrix),

$\mathbf{G} = \left[\mathbf{IDEN}_{QI} - \theta\, \mathbf{W} \otimes \mathbf{IDEN}_I\right]^{-1}$ ($QI \times QI$ matrix),

$\mathbf{D} = \left(\mathbf{IDEN}_{QK} - \left[\mathbf{W} \otimes \boldsymbol{\lambda} \mathbf{IDEN}_K\right]\right)^{-1}$ ($QK \times QK$ matrix),

$\boldsymbol{\Psi} = \mathbf{IDEN}_Q \otimes \widetilde{\boldsymbol{\Psi}}$ ($QI \times QI$ matrix), $\boldsymbol{\Omega} = \widetilde{\mathbf{x}}\mathbf{D}\,[\mathbf{IDEN}_Q \otimes \widetilde{\boldsymbol{\Omega}}]\left(\mathbf{D}\widetilde{\mathbf{x}}\right)'$ ($QI \times QI$ matrix), and $\boldsymbol{\Lambda} = \mathbf{G}[\mathbf{IDEN}_Q \otimes \widetilde{\boldsymbol{\Lambda}}]\mathbf{G}'$ ($QI \times QI$ matrix).

Then, using other notations as in Section 2.5 and Section 3, we may write the following counterpart of Equation (5):

$$\mathbf{U} \;= \mathbf{S}\Big[\big(\mathbf{1}_Q \otimes \widetilde{\mathbf{A}}\big) + \mathbf{x}\boldsymbol{b} + \mathbf{G}\widetilde{\boldsymbol{\tau}} + \widetilde{\mathbf{x}}\mathbf{D}\widetilde{\boldsymbol{\gamma}} + \widetilde{\boldsymbol{\eta}}\Big] \sim MVN_{I \times Q}(\widetilde{\mathbf{B}}, \widetilde{\boldsymbol{\Xi}}), \tag{8}$$

where $\widetilde{\mathbf{B}} = \mathbf{S}\big[(\mathbf{1}_{\mathbf{Q}} \otimes \widetilde{\mathbf{A}}) + \mathbf{x}\boldsymbol{b}\big]$ and $\widetilde{\boldsymbol{\Xi}} = \mathbf{S}\,[\boldsymbol{\Lambda} + \boldsymbol{\Omega} + \boldsymbol{\Psi}]\mathbf{S'}$.

The above $QI \times 1$-vector $\mathbf{U}$ can be simulated many times (say 5000 times) using the estimated values of the parameters. Next, the utilities across alternatives for each individual q for each of the draws are compared, and the alternative with the highest utility is designated as the "chosen" alternative for each of the draws for each individual. The predicted share of each alternative across the draws for each individual is an estimate of the probability of choice of the mode for the individual. The aggregate share (across individuals) of each mode is then readily obtained by averaging the individual-level probabilities of each mode.

The procedure to compute the effect of variables is then as follows. For continuous variables, one may increase the exogenous variable by some percentage value to estimate the impact. For discrete binary variables, one can set the value to zero for all individuals and then change to the value of one for all individuals to estimate the impact. Specifically, consider the case of increasing a continuous exogenous variable by 10%. To estimate this impact, first compute the aggregate modal shares as described in the previous paragraph for the base case. Then, increase the exogenous variable for the first individual by 10% (while keeping all other values fixed), and compute the predicted modal probabilities for the individual. Subsequently, the percentage change (from the base case) in the modal probabilities may be computed, and designated as the direct effect corresponding to the first individual. Similarly, the percentage change (from the base case) in the predicted modal probabilities for this first individual because of a 10% increase in the exogenous variable of all other individuals (but not individual 1) is obtained, and designated as the indirect effect corresponding to a change in the exogenous variable for all other individuals. Finally, the overall percentage change (from the base case) in the predicted modal shares of the first individual because of a 10% increase in the tax rate of all individuals (including individual 1) is also obtained, and labeled as the total effect for the first individual. This process can be repeated in turn for each of the individuals in the sample. Next, the overall measure of direct, indirect, and total percentage effects may be obtained as the average across the individual-specific direct, indirect, and total percentage changes, respectively. Note that the total percentage effect will not be equal to the sum of the direct and indirect effects because we work with percentage changes.

3. ESTIMATION APPROACH

To develop the likelihood function for our model, let individual q's observed choice of mode at choice occasion t be m_{qt}. Next, define $\mathbf{M}$ as an $[(I-1)\times T\times Q]\times[I\times T\times Q]$ block diagonal matrix, with each block diagonal having $(I-1)$ rows and I columns corresponding to the t^{th} choice occasion for individual q. Let this $(I-1)\times I$ matrix correspond to an $(I-1)$ identity matrix with an extra column of -1's added as the $m_{qt}{}^{\text{th}}$ column. For instance, consider the case of Q = 2, T = 2, and I = 4. Let individual 1 be observed to choose mode 2 at choice occasion 1 and mode 1 at choice occasion 2, and let individual 2 be observed to choose mode 3 in choice occasion 1 and mode 4 in choice occasion 2. Then $\mathbf{M}$ takes the form below.

$$\mathbf{M}=\left[\begin{array}{cccc|cccc|cccc|cccc}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\hline
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0\\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{array}\right] \tag{9}$$

Let $\mathbf{B}=\mathbf{M}\tilde{\mathbf{B}}$ and $\mathbf{\Xi}=\mathbf{M}\tilde{\mathbf{\Xi}}\mathbf{M}'$, and let ϖ be the collection of parameters to be estimated: $\varpi=\left[\boldsymbol{b}';\text{Vech}(\tilde{\mathbf{\Omega}});\tilde{\mathbf{A}}',\text{Vech}(\tilde{\mathbf{\Lambda}}),\text{Vech}(\tilde{\mathbf{\Psi}}),\delta,\rho,\lambda_1,\lambda_2,...\lambda_K,\theta\right]'$, where Vech($\tilde{\mathbf{\Omega}}$) represents the row vector of upper triangle elements of $\tilde{\mathbf{\Omega}}$. Then, the likelihood of the observed sample may be written succinctly as:

$$L_{ML}(\varpi)=\text{Prob}[\mathbf{MU}<0]=F_{QT(I-1)}(-\mathbf{B},\mathbf{\Xi})=\Phi_{QT(I-1)}(\boldsymbol{\omega}_{\mathbf{\Xi}}^{-1}(-\mathbf{B}\ ),\mathbf{\Xi}^*), \tag{10}$$

where $F_{QT(I-1)}$ is the multivariate cumulative normal distribution of $Q \times T \times (I-1)$ dimensions, $\Phi_{QT(I-1)}$ is the standard multivariate cumulative distribution of $Q \times T \times (I-1)$ dimensions, $\boldsymbol{\omega}_{\boldsymbol{\Xi}}$ is the diagonal matrix of standard deviations of $\boldsymbol{\Xi}$ (that is, $\boldsymbol{\omega}_{\boldsymbol{\Xi}}$ is a diagonal matrix with entries that represent the square root of the diagonal entries of $\boldsymbol{\Xi}$), and $\boldsymbol{\Xi}^* = \boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Xi}\boldsymbol{\omega}_{\boldsymbol{\Xi}}^{-1}$.

Despite advances in simulation techniques and computational power, the evaluation of a very high dimensional integral as in Equation (7) is literally infeasible using traditional frequentist and Bayesian simulation techniques. In a recent paper, Bhat (2011) proposed a maximum approximate composite marginal likelihood (MACML) inference approach in such cases.

3.1. The MACML Estimation Technique

The MACML approach combines a composite marginal likelihood (CML) estimation approach with an approximation method to evaluate the multivariate standard normal cumulative distribution (MVNCD) function. The composite likelihood approach replaces the likelihood function with a surrogate likelihood function of substantially lower dimensionality, which is then subsequently evaluated using an *analytic approximation* method rather than simulation techniques. This combination of the CML with the specific analytic approximation for the MVNCD function is effective because it involves only univariate and bivariate cumulative normal distribution function evaluations, regardless of the spatial and/or temporal complexity of the model structure. In simulation studies, the approach has been able to recover parameters and their covariance matrix estimates quite accurately and precisely because of the smooth nature of the first and second derivatives of the approximated analytic log-likelihood function (unlike the non-smooth first and second derivatives that arise in simulation approaches). The MVNCD approximation method is based on decomposition into a product of conditional probabilities, and the subsequent approximation of the conditional probabilities using a linear regression model structure (see Bhat, 2011). Note that this approximation method has nothing to do with linearization around the spatial correlation parameter, as in Klier and McMillen (2008) and Smirnov (2008). Rather, it simply is an approximation to evaluate the MVNCD function in any context.

The MACML approach, similar to the parent CML approach (see Varin *et al.*, 2011 and Bhat, 2014 for recent reviews of the CML approach), maximizes a surrogate likelihood function that compounds much easier-to-compute, lower-dimensional, marginal likelihoods (see Lindsay *et al.*, 2011, Bhat, 2011, and Yi *et al.*, 2011). The CML approach, which belongs to the more general class of composite likelihood function approaches (see Lindsay, 1988), may be explained in a simple manner as follows. Instead of developing the likelihood function for the entire set of Q observations, as in Equation (7), one may compound (multiply) pairwise probabilities of observation q having chosen alternative i at time period t and j at time t' and observation q' having chosen alternative i at time period t and j at time t', and so on. The CML estimator (in this instance, the pairwise CML estimator) is then the one that maximizes the compounded probability of all pairwise events.[5] The properties of the CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004, Yi *et al.*, 2011). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191, Xu and Reid, 2011), the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood; for a formal proof, see Xu and Reid, 2011 and Bhat, 2014). The CML function may be written as

$$L_{CML}(\varpi) = \prod_{q=1}^{Q} \prod_{q'=q}^{Q} \prod_{t=1}^{T} \prod_{t'=t}^{T} L_{qq'tt'} \text{ with } q \neq q' \text{ when } t = t' \tag{11}$$

where $L_{qq'tt'} = \text{Prob}(Q_{qt} = m_{qt}, Q_{q't'} = m_{q't'}) = \Phi_{2(I-1)}(\boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'tt'}}^{-1}(-\ddot{\mathbf{B}}_{qq'tt'}), \ddot{\boldsymbol{\Xi}}_{qq'tt'}^{*})$, Q_{qt} is an index for the alternative chosen by individual q at time t, $\ddot{\mathbf{B}}_{qq'tt'} = \boldsymbol{\Delta}_{qq'tt'}\mathbf{B}$, $\ddot{\boldsymbol{\Xi}}_{qq'tt'} = \boldsymbol{\Delta}_{qq'tt'}\boldsymbol{\Xi}\boldsymbol{\Delta}'_{qq'tt'}$, $\ddot{\boldsymbol{\Xi}}_{qq'tt'}^{*} = \boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'tt'}}^{-1} \ddot{\boldsymbol{\Xi}}_{qq'tt'} \boldsymbol{\omega}_{\ddot{\boldsymbol{\Xi}}_{qq'tt'}}^{-1}$, and $\boldsymbol{\Delta}_{qq'tt'}$ is a $2(I-1) \times QT(I-1)$-selection matrix with an identity matrix of size $(I-1)$ occupying the first $(I-1)$ rows and the $\left[(q-1)\times(I-1)\times T + (t-1)\times(I-1)+1\right]^{th}$ through $\left[(q-1)\times(I-1)\times T + t\times(I-1)\right]^{th}$ columns, and another identity matrix of size $(I-1)$ occupying the last $(I-1)$ rows and the

[5] This is the same estimator as the one used by Wang *et al*. (2013) in a spatial binary probit context and that they label as the partial maximum likelihood estimator (PMLE). However, papers using the pairwise CML for spatial econometric contexts and for even more general discrete choice contexts than the spatial binary probit were published by Bhat and colleagues earlier, as in Bhat and Sidharthan (2012), Sener and Bhat (2012), Castro *et al*. (2012), Bhat (2011), and Bhat *et al*. (2010). Some of these combine the CML method with the MVNCD approximation, which then becomes the MACML approach of Bhat (2011).

$[(q'-1)\times(I-1)\times T+(t'-1)\times(I-1)+1]^{th}$ through $[(q'-1)\times(I-1)\times T+t'\times(I-1)]^{th}$ columns. The matrix $\boldsymbol{\Delta}_{qq'tt'}$ has zero elements everywhere else.

The CML function above requires the computation of the multivariate normal cumulative distribution (MVNCD) function that is of dimension $2\times(I-1)$ integrals (instead of $Q\times T\times(K-1)$ in the full maximum likelihood case). Such integrals may be computed easily using the MVNCD approximation method embedded in the MACML method. The CML estimator of ϖ (obtained by maximizing the logarithm of Equation (8) with respect to ϖ) is consistent and asymptotically normally distributed with asymptotic mean ϖ and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix (see Bhat, 2014 for a detailed discussion).[6] To write the covariance matrix, let $\widetilde{W}=[QT(QT-1)]/2$ be the total number of pairings used in the CML function of Equation (11), and let the MVNCD function in $L_{qq'tt'}$ of Equation (8) be evaluated using the MACML method, Then, the covariance matrix of the MACML estimator is $\frac{[\hat{\boldsymbol{G}}]^{-1}}{\widetilde{W}}=\frac{[\hat{\boldsymbol{H}}^{-1}]\,[\hat{\boldsymbol{J}}]\,[\hat{\boldsymbol{H}}^{-1}]'}{\widetilde{W}}$, where

$$\hat{\boldsymbol{H}}=-\frac{1}{\widetilde{W}}\left[\sum_{q=1}^{Q}\sum_{q'=q}^{Q}\sum_{t=1}^{T}\sum_{t'=t}^{T}\frac{\partial^2\log L_{qq'tt'}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]_{\hat{\theta}_{MACML}} \quad q'\neq q \text{ when } t=t', \tag{12}$$

and the **J** matrix maybe empirically estimated as:

[6] The CML estimator loses some asymptotic efficiency from a theoretical perspective relative to a full likelihood estimator, because information embedded in the higher dimension components of the full information estimator are ignored by the CML estimator. However, as presented in Bhat (2014), many studies have found that the efficiency loss of the CML estimator (relative to the maximum likelihood (ML) estimator) is negligible to small in applications on finite samples. Besides, in spatial models, a maximum simulated likelihood (MSL) approach is needed for estimation because of the high dimensionality of integration. When simulation methods are used, there is also a loss in asymptotic efficiency in the maximum simulated likelihood (MSL) estimator relative to a full likelihood estimator (McFadden and Train, 2000). Consequently, it is difficult to state from a theoretical standpoint whether the CML estimator efficiency will be higher or lower than the MSL estimator efficiency. Bhat (2014) presents many studies that empirically compare the CML and MSL finite sample efficiency results in models where it is practical to implement the MSL, and concludes that "….any reduction in the efficiency of the CML approach relative to the MSL approach is in the range of non-existent to small". In addition, while the MSL method encountered convergence problems even for relatively simple aspatial models, they noted that the CML approach exhibited no such problems, and had the benefit of substantially faster computational times.

$$\hat{\mathbf{J}} = \frac{1}{Q}\left[\sum_{q=1}^{Q}\left[\frac{1}{\widetilde{W}}\left(\left[\sum_{l_{\tilde{q}}=1}^{Q}\sum_{l'=l_{\tilde{q}}}^{Q}\sum_{t=1}^{T}\sum_{t'=t}^{T} \frac{\partial \log L_{l_{\tilde{q}}l'tt'}}{\partial \boldsymbol{\theta}}\right]\left[\sum_{l_{\tilde{q}}=1}^{Q}\sum_{l'=l_{\tilde{q}}}^{Q}\sum_{t=1}^{T}\sum_{t'=t}^{T} \frac{\partial \log L_{l_{\tilde{q}}l'tt'}}{\partial \boldsymbol{\theta}'}\right]\right)\right]\right]_{\hat{\theta}_{MACML}},$$

$$q' \neq q \text{ when } t = t'$$

The number of pairings in the MACML procedure in Equation (11) is $[QT(QT-1)]/2$. Strategies to reduce the pairings are discussed in detail in Bhat (2014), though we consider all pairings in the simulation study of the current paper.

3.2. Identification and Positive Definiteness Considerations

As usual, appropriate scale and level normalization must be imposed on $\widetilde{\mathbf{A}}$, $\widetilde{\mathbf{\Lambda}}$ and $\widetilde{\mathbf{\Psi}}$ for identifiability. Specifically, only utility differentials matter in discrete choice models. However, as discussed in Bhat (2011), the MACML inference approach, like the traditional GHK simulator, takes the difference in utilities against the chosen alternative during estimation. Thus, consider that the individual q chooses alternative m_{qt} at choice instance t. This implies that values of $\alpha_{qim_{qt}} = \widetilde{\alpha}_{qi} - \widetilde{\alpha}_{qm_{qt}}$ $(i \neq m_{qt})$, and the covariance matrices $\mathbf{\Lambda}_{m_{qt}}$ and $\mathbf{\Psi}_{m_{qt}}$ are needed for individual q at choice occasion t. However, though different random effects differentials and different covariance matrices are used for different individuals and different choice occasions, all of these must originate in the same values of the undifferenced error term vector $\widetilde{\mathbf{A}}$ and covariance matrices $\widetilde{\mathbf{\Lambda}}$ and $\widetilde{\mathbf{\Psi}}$. To achieve this consistency, we normalize $\widetilde{\alpha}_{q1} = 0 \;\forall\; q$. This implies that $\widetilde{a}_1 = 0$. Also, we develop $\mathbf{\Lambda}$ from $\mathbf{\Lambda}_1$ by adding an additional row on top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros. Similarly, we construct $\mathbf{\Psi}$ from $\mathbf{\Psi}_1$ by adding a row on top and a column to the left. This first row and the first column of the matrix $\widetilde{\mathbf{\Psi}}$ are also filled with zero values. However, an additional normalization needs to be imposed on $\widetilde{\mathbf{\Psi}}$ because the scale is also not identified. For this, we normalize the element of $\widetilde{\mathbf{\Psi}}$ in the second row and second column to the value of one. Note that all these normalizations do not place any restrictions, and a fully general specification is the result. But they are needed for econometric identification, because, in MNP models, only the covariance matrix of the utility differences (from a base utility) are estimable and the scale of one of the utility differences has to be fixed.

A final issue regarding estimation. The analyst needs to ensure the positive definiteness of the three covariance matrices $\tilde{\mathbf{\Omega}}, \tilde{\mathbf{\Lambda}},$ and $\tilde{\mathbf{\Psi}}$. Also, the spatial parameters $\delta, \rho, \lambda_1, \lambda_2, ...\lambda_K, \theta$ need to be restricted between 0 and 1. In our estimation, the positive-definiteness of each of the three covariance matrices $\tilde{\mathbf{\Omega}}, \tilde{\mathbf{\Lambda}},$ and $\tilde{\mathbf{\Psi}}$ is guaranteed by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of these matrices, and maximizing with respect to these elements of the Cholesky factor. To ensure the constraints on the spatial parameters, we parameterize all of the spatial elements appropriately. For example, we parameterize $\delta = 1/[1+\exp(\tilde{\delta})]$, estimate $\tilde{\delta}$, and then obtain the estimate of δ.

4. SIMULATION STUDY

In this section, we undertake a simulation experiment with two objectives in mind. The first objective is to examine the ability of the MACML inference approach to recover the parameters from finite samples in our proposed model of travel mode choice. The second is to examine the effects of ignoring spatial lag effects and residential self-selection (*i.e.*, spatial drift effects) when both are actually present. The use of a simulated exercise is valuable because the true parameters underlying the data generating process (DGP) are set by the analyst, and the analyst can evaluate the behavior of the MACML estimator for different levels of the spatial lag and spatial drift effects.

4.1. Experimental Design

A four-alternative choice situation ($I = 4$) with five choice occasions ($T = 5$) is considered for the simulation exercise. A total of $Q = 200$ individuals are assumed. The choice instance specific $\tilde{\varepsilon}_{qti}$ error terms can have a first order autoregressive temporal dependency process: $\tilde{\varepsilon}_{qti} = \rho\tilde{\varepsilon}_{q,t-1,i} + \tilde{\eta}_{qti}$. However, for repeated choice situations, we do not expect temporal variations in unobserved factors influencing an individual's utility for a mode, and so we set $\rho = 0$ in the simulation. The covariance matrix $\tilde{\mathbf{\Psi}}$ for the error term vector $\tilde{\boldsymbol{\eta}}_{qt} \left[= \left(\tilde{\eta}_{qt1}, \tilde{\eta}_{qt2}, ..., \tilde{\eta}_{qtI} \right)' \right]$ is specified to be diagonal and fixed with variances of 0.5 along the diagonal. Such a matrix is a restrictive case of the more general $\tilde{\mathbf{\Psi}}$ covariance matrix discussed in Section 2.1. Note that such a structure simplifies the simulation, since the elements of $\tilde{\boldsymbol{\Psi}}$ are not estimated. The reason for such a restriction on the $\tilde{\boldsymbol{\Psi}}$ matrix in the simulation design is to direct attention on the issues of spatial effects, which is the focus of the current paper.

Three independent variables are used in the utility equation and each of them is generated from a standard univariate normal distribution (these are the elements of the $\boldsymbol{x}_{qti}$ vector). We consider the coefficient on the first variable to be fixed, but allow spatial drift (residential self-selection) effects in the next two elements of the coefficient vector. In the notation of the previous section, $\tilde{\beta}_{q1} = 0$. For the remaining two coefficients, we use the autoregressive specification as follows: $\tilde{\beta}_{qk} = \lambda_k \sum_{q'} w_{qq'} \tilde{\beta}_{q'k} + \tilde{\gamma}_{qk}$ (k=2,3). The spatial drift (residential self-selection) parameter is fixed across both the random coefficients in the experiments; $\lambda_2 = \lambda_3 = \lambda.$ The covariance matrix $\tilde{\boldsymbol{\Omega}}$ for the two random coefficients is specified as follows:

$$\tilde{\boldsymbol{\Omega}} = \begin{bmatrix} 0.81 & 0.54 \\ 0.54 & 1.00 \end{bmatrix} = \mathbf{L}_{\tilde{\Omega}} \mathbf{L}'_{\tilde{\Omega}} = \begin{bmatrix} 0.90 & 0.00 \\ 0.60 & 0.80 \end{bmatrix} \begin{bmatrix} 0.90 & 0.60 \\ 0.00 & 0.80 \end{bmatrix} \tag{13}$$

The Cholesky decomposition of $\tilde{\boldsymbol{\Omega}}$ guarantees the positive definiteness of $\tilde{\boldsymbol{\Omega}}$. In the estimations, the likelihood function is reparameterized in terms of the lower Cholesky factor $\mathbf{L}_{\tilde{\Omega}}$, and the three associated Cholesky parameters $l_{\tilde{\Omega}1} = 0.9$, $l_{\tilde{\Omega}2} = 0.6$, and $l_{\tilde{\Omega}3} = 0.8$ are estimated. Collectively, these three parameters, stacked vertically into a column vector, will be referred to as $\mathbf{l}_{\tilde{\Omega}}$. The mean effects vector for the coefficients; $\boldsymbol{b}\left[= (b_1, b_2, b_3)'\right]$; is set to $\boldsymbol{b}$ = (0.5, 0.8, 1)'. In the simulation experiments, we do not consider separate random effects, as these are but one type of random coefficients on dummy variables defined in the $\boldsymbol{x}_{qti}$ vector. That is, in the notation of the previous section, $\tilde{\alpha}_{qi} = \tilde{a}_i + \breve{\alpha}_{qi} = 0 \;\forall\, q, i.$

To examine the potential impacts of different levels of spatial lag dependence and spatial drift (*i.e.*, residential self-selection dependence) on the ability of the MACML approach to recover model parameters, we consider two values of the spatial lag autoregressive coefficient δ corresponding to low dependence (δ = 0.25) and high dependence (δ = 0.75), as well as two values of the spatial drift (residential self-selection) autoregressive coefficient λ corresponding to low dependence (λ = 0.25) and high dependence (λ = 0.75). Thus, in total, there are four possible combinations of the spatial lag and drift autoregressive coefficients considered in the simulations. In the simulations, the 200 individuals are located on a rectangular grid with the longer side containing 50 locations spaced 1 unit apart and the shorter side containing 4 locations spaced 1 unit apart. The spatial weight matrix **W** (of size 200×200) is created using the inverse of the distance on the coordinate plane between observational units.

The simulation experiments entail assuming underlying "true" values for the above parameters and generating data sets for estimation. Specifically, using the pre-specified parameters, we develop the mean vector **B** and variance matrix $\mathbf{\Delta}$ of the utility vector **U** in Equation (5) for each of the four combinations of δ and λ just discussed. A $(QTI \times 1)$ vector of the utility vector **U** is drawn from the multivariate normal distribution with mean **B** and covariance structure $\mathbf{\Delta}$. Then, for each individual and choice occasion, the alternative with the highest utility is designated as the chosen alternative. This variable constitutes the discrete dependent variable. For each of the four combinations, the data generation process just discussed is undertaken 20 times with different realizations of the utility vector **U** from the values of **B** and $\mathbf{\Delta}$ to assemble a total of 80 datasets. The MACML estimator is then applied to each data set to estimate eight parameters: three mean coefficients on the exogenous variables (corresponding to the b_1, b_2, and b_3 coefficients), the three standard deviation elements of the lower triangular Cholesky decomposition of covariance matrix $\tilde{\mathbf{\Omega}}$, the spatial lag parameter δ, and the residential self-selection parameter λ.[7] To be more specific, the MACML estimator is applied to each dataset 10 times with different permutations for the ordering of the conditional probabilities in the MVNCD computation to obtain the approximation error, computed as the standard deviation of estimated parameters among the 10 different estimates on the same data set (note that, within a dataset and for each of the ten runs, the permutation used varies across individuals, but is the same across iterations for a given individual). In future studies, one can use more than 20 datasets for each combination, and more than 10 estimations for each dataset. Here, we confined ourselves to 20 datasets and 10 permutations per data set, because a total of 800 estimations (20

[7] In the MACML approach, a single random permutation is generated for each choice instance (the random permutation varies across choice instances, but is the same across iterations for a given choice instance) to decompose the multivariate normal cumulative distribution (MVNCD) function into a product sequence of marginal and conditional probabilities (see Section 2.1 of Bhat, 2011). We also tested higher number of permutations, but noticed little difference in the estimation results, and hence settled with the single permutation per individual.

datasets per combination × 4 combinations × 10 permutations per dataset and each combination) have to be undertaken even with only 20 datasets.

4.2. Performance Evaluation

The performance of the MACML inference approach in estimating the parameters of the model and their standard errors is evaluated as follows:

(1) Estimate the MACML parameters for each data set and for each of 10 independent sets of permutations. Estimate the standard errors (s.e.) using the Godambe (sandwich) estimator.

(2) For each data set *s*, compute the mean estimate for each model parameter across the 10 random permutations used. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the **absolute percentage (finite sample) bias** (APB) of the estimator as:

$$APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$$

(3) Compute the standard deviation of the MED values across datasets, and label this as the **finite sample standard deviation or FSSD** (essentially, this is the empirical standard deviation).

(4) For each data set *s*, compute the mean s.e. for each model parameter across the 10 draws. Call this MSED, and then take the mean of the MSED values across the 20 data sets and label this as **the asymptotic standard error or ASE** (essentially this is the standard error of the distribution of the estimator as the sample size gets large).

(5) Next, to evaluate the accuracy of the asymptotic standard error formula as computed using the MACML inference approach for the finite sample size used, compute the absolute percentage bias of the asymptotic standard error (APBASE) for each parameter relative to the corresponding finite sample standard deviation.

$$APBASE = \left| \frac{\text{ASE - FSSD}}{\text{FSSD}} \right| \times 100$$

(6) Compute the standard deviation of the parameter values around the MED parameter value for each data set, and take the mean of this standard deviation value across the data sets; label this as the **approximation error** (**APERR**). Compute the APERR as a percentage of FSSD.

4.3. Additional Restrictive Model Comparisons with the Proposed Model

In addition to the above exercise to investigate the ability of the MACML model to recover parameters, we also examine the potential problems that could arise from ignoring spatial lag effects and spatial drift (residential self-selection) effects. To do so, we estimate two additional models on the 20 data sets generated for each combination of spatial and temporal dependence levels. The first model ignores the spatial lag autocorrelation coefficient δ (that is, assumes δ = 0), while the second model assumes away the spatial-drift based residential self-selection autocorrelation coefficient λ (that is, assumes λ = 0). For ease in presentation, we will refer to the first model as the "No spatial lag" (NSL) model, and the second as the "No spatial drift" (NSD) model. We compare these two restrictive formulations with our general spatial (GS) model based on the mean APB measure across all parameters and the adjusted composite log-likelihood ratio test (ADCLRT) value (see Pace *et al*., 2011 and Bhat, 2011). For the comparisons, we use a single replication per data set (the replication is the same one for the generalized spatial (GS) model and all the restrictive models; that is, we use a single permutation per individual that varies across individuals but is held fixed across the GS and other models). We do so rather than run 10 replications for each of the GS and the more restrictive models because, as we will present in the next section, the approximation error in the parameters is negligible for any given data set. The ADCLRT statistic needs to be computed for each data set separately, and compared with the chi-squared table value with the appropriate degrees of freedom. Here we identify the number of times (corresponding to the 20 model runs, one run for each of the 20 data sets) that the ADCLRT value rejects the NSL and NSD models in favor of our proposed GS model. This is essentially an exploration of the power or the probability that we reject the null hypothesis (no spatial drift or no spatial lag) when the null hypothesis is false.

4.4. Simulation Results

4.4.1. Recoverability of parameters in the low spatial lag case

Tables 2a and 2b present the results for the low spatial lag case, with the first table representing the case of low spatial drift and the second table representing the case of high spatial drift. As indicated earlier, there are four alternatives $(I=4)$ in each set of alternatives, leading to a six $[=(I-1)*2]$ dimensional integral in the CML function.

The parameter estimation results in Table 2a indicate that the MACML method does very well in recovering the parameters for the low spatial dependency and low spatial drift case, as can be observed by comparing the mean estimates of the parameters (see the first sub-column under the main column entitled "parameter estimates") with the true values (see the second column). The absolute percentage bias (APB) is no more than 2.7% for any parameter (see the column entitled "Absolute Percentage Bias") with an overall mean value of 1.377% across all parameters, as indicated at the bottom of the table (see the row labeled "overall mean value across parameters"). However, the APB values are somewhat higher for the low spatial dependency and high spatial drift case in Table 2b, with a maximum APB value of 9.2% and an overall mean value of 5.712%. This is to be expected because a high value of the spatial drift parameter λ leads to higher interdependence (due to unobserved factors) among individuals, which leads to a more non-linear surface of the CML function over which to optimize (relative to the case of a low spatial drift parameter). Not surprisingly, the highest degradation between the low spatial drift and high spatial drift cases is in the recovery ability of the spatial drift parameter itself. In the low spatial drift case, the mean APB for the spatial drift parameters λ_1 and λ_2 is a mere 1.08% relative to 9.48% in the high spatial drift case. The APB values for the elements of the Cholesky parameter vector $\mathbf{l}_{\tilde{\Omega}}$ and the spatial lag autocorrelation coefficient δ are also quite a bit higher in the high spatial drift case relative to the low spatial drift case, because these elements enter the CML function in a rather non-linear fashion through sub-matrices of the covariance matrix $\mathbf{M\Delta M'}$. On the other hand, there is no substantial difference in recovery ability for the elements of the $\boldsymbol{b}$ vector between the low and high spatial drift cases, because these elements enter the CML function more linearly through the sub-vector of the mean vector $-\mathbf{MB}$. Overall, it is important to note that, even in the high spatial drift case, the MACML still does a remarkable job of recovering parameters.

The results in Table 2a also indicate good empirical efficiency of the MACML estimator for the low spatial lag-low spatial drift case, with the finite sample standard deviation (FSSD) ranging from less than one-seventieth of the mean estimated value (for λ_2) to about a sixth of the mean estimated value (for $l_{\tilde{\Omega}3}$). Across all the parameters, the FSSD is 12.33 % of the mean parameter estimate. The MACML retains good empirical efficiency in the low spatial lag-high spatial drift case of Table 2b too, though the increase in the finite sample standard error (FSSD) values from the low spatial lag-low spatial drift case is clearly discernible. Across all parameters, the FSSD is 22.25 % of the mean parameter estimate for this second case, which is almost double of the first case. Importantly, between Tables 2a and 2b, the mean of the FSSD values across the elements of the ***b*** vector (as a % of the corresponding mean parameter estimates) jumps from 10.89% to 27.81%, and the mean of the FSSD values across the λ_2 and λ_3 parameters increases from 4.50% to 20.14%.

The finite sample standard errors and the asymptotic standard errors obtained using the Godambe matrix in the MACML method are close, with the APBASE values ranging between 0.46-20.47 for all of the parameters in both the cases. The average APBASE across all the parameters is less than 10% for both the cases, indicating that the asymptotic formula is performing well in estimating the finite sample standard error.

Finally, the last columns of Tables 2a and 2b present the approximation error (APERR) for each of the parameters as a percentage of the FSSD.[8] While there is no hard metric for whether or not an APERR value is reasonable, APERR values of less than 25% of the FSSD may be considered reasonable (for instance, in Sandor and Train, 2004, the highest simulation deviation is about 24% of the sampling standard deviation in the best simulation method). The APERR values (as a percentage of the FSSD) are presented in the last column of the tables. The values range from 1.21%-5.25% for all parameters except the λ_2 and λ_3 parameters. However, the APERR values (as a percentage of the FSSD) increase quite substantially to the order of 40%

[8] As pointed out by McFadden (1989) and Sandor and Train (2004), approximation methods of any kind to evaluate the maximum of an analytically-intractable function will tend to show more variance in the convergent values (in repeated applications of the approximation with different sets of simulation draws in a simulated setting or different permutations of the conditional probability sequence in our MACML estimation setting) as the function gets flatter near the maximum. This is because errors introduced by simulation or other approximations can move the maximum considerably when the function is flat near the maximum. Of course, large sampling variances of the parameters embedded in a function means a large sampling variance of the function being maximized; that is, the larger the sampling variance of parameters, the higher in general will be the approximation error. Thus, we examine the extent of approximation error as a percentage of the finite sample standard deviation (or FSSD).

for λ_2 and λ_3 in Table 2a. However, this is because of the very low FSSD values for these two parameters in Table 2a. The APERR values themselves are still very small for both these parameters. Essentially, while the ratio of APERR to FSSD is of the same order for other parameters, this does not seem to hold for the λ_2 and λ_3 parameters, an issue that needs further investigation in future studies. The APERR values (as a percentage of the FSSD) are lower and of the order of 10% in the case with high spatial drift, though these values remain the highest compared to other values in Table 2b. Across all parameters, the APERR (as a percentage of FSSD) is of the order of 10% for the low spatial drift case and 5% for the high spatial drift case, suggesting that, overall, even a single permutation (per observation) of the MACML estimator provides adequate precision, in the sense that the convergent values are about the same for a given data set regardless of the permutation used for the decomposition of the multivariate probability expression within the MACML approach.

4.4.2. Recoverability of parameters in the high spatial lag case

Tables 2c and 2d present the results for the high spatial lag case, with the first table focusing on the low spatial drift case and the second table focusing on the high spatial drift case.

The results in Table 2c show again that the MACML method does very well in recovering the parameters for the low spatial drift case, with the order of the APB being about the same as in Table 2a. However, the APB values are once again somewhat higher for the high spatial drift case in Table 2d compared to the low spatial drift case in Table 2c. This mirrors the difference between the low and high spatial drift cases for the low spatial lag case. Also, as in the low spatial lag case, the highest degradation between the low spatial drift and high spatial drift cases for the high spatial lag case is again in the recovery ability of the spatial drift parameter itself. In the low spatial drift case, the mean APB for the spatial drift parameters λ_2 and λ_3 is a mere 0.56% relative to 8.25% in the high spatial drift case. The APB values for the elements of the Cholesky parameter vector $\mathbf{l}_{\tilde{\Omega}}$ are also quite a bit higher in the high spatial drift case (mean APB value of 3.81%) relative to the low spatial drift case (mean APB of 0.96%). Overall, however, it is important to note that, even in the high spatial lag-high spatial drift case, the MACML still does a remarkable job of recovering parameters, with an overall mean value of 4.216%.

As in the low spatial lag effect-low spatial drift case (Table 2a), the standard error estimates of the parameters for the high spatial lag effect-low spatial drift (Table 2c) case also indicate good empirical efficiency of the MACML estimator. Across all parameters, the finite sample standard error (FSSD) is about 15.20 % of the mean parameter estimate. However, the finite sample standard error (FSSD) is higher for the high spatial lag effect-high spatial drift case (Table 2d) at about 22.71 % of the mean parameter estimate. Once again, between Tables 2c and 2d, the mean of the FSSD values across the elements of the ***b*** vector (as a percentage of the corresponding mean parameter estimates) jumps from 16.44% to 26.43%, and the mean of the FSSD values across the λ_1 and λ_2 parameters increases from 6.34% to 23.21%. All of these FSSD increases between the low spatial drift and high spatial drift cases are of the same order in this high spatial lag case as in the low spatial lag case.

The finite sample standard deviations and the asymptotic standard errors obtained using the Godambe matrix in the MACML method are close for both the low spatial drift case (Table 2c) and the high spatial drift case (Table 2d), with the APBASE values ranging between 0.73%-30.07% for all of the parameters in both the cases, indicating that the asymptotic formula is performing very well in estimating the finite sample standard deviation (note also that the values of the APBASE of about 30% correspond to the smallest FSSD values; the absolute difference between the FSSD and ASE is small even in these cases of high APBASE). As in the low spatial dependency case, the average APBASE across all the parameters is of the order of 10% in both the cases.

Finally, the approximation errors (APERR) for each of the parameters in Tables 2c and 2d are negligible in magnitude, with the APERR (as a percentage of the FSSD or the ASE), averaged across all the parameters, being of the order of 5% of the sampling error for the first case and of the order of 1.8% of the sampling error for the second case. This is again evidence that just a single permutation (per observation) of the MACML estimator provides adequate precision even for the higher spatial lag dependency cases.[9]

---

[9] A peculiar observation related to the approximation error (as a percentage of the FSSD) is that it declines quite considerably for the λ_2 and λ_3 parameters as one moves from low spatial lag to high spatial lag and from low spatial drift to high spatial drift. Why this is so is left for future exploration.

4.4.3. Qualitative comparison of proposed method relative to earlier methods

As indicated in the first section, many methods have been suggested to estimate spatial discrete choice models (see Bhat and Sener, 2009 and Bille, 2013 for reviews). However, a problem with these approaches is that they are, in general, not practically feasible for moderate-to-large samples of the size usually used in most transportation, urban science, and geographic micro-level data contexts. But we provide a qualitative comparison of the proposed method to other spatial discrete choice estimation methods based on a study conducted by Calabrese and Elkink (2014) (CE) for the spatial case of a cross-sectional binary probit model with a simple spatial lag specification. We should caution that the comparisons below are intended simply to provide a picture of the performance of the methods. They are not strictly valid because of different spatial specifications (we include spatial lag, spatial drift effects and unobserved heterogeneity, while CE's study is for a spatial lag specification), binary versus multinomial specifications, different spatial weight matrices, cross-sectional versus panel analysis, different numbers of explanatory variables in the simulation models, and different sample sizes.

Our method does well in recovering parameters (including the spatial autoregressive coefficient) regardless of the value of the autoregressive coefficient. On the other hand, McMillen's (1992) Expectation Maximization (EM) method and Klier and McMillen's (2008) linearized Generalized Method of Moments (LGMM) method are well known not to do very well at high values of the spatial coefficient. In Calabrese and Elkink's (2014) simulation study, for a true value of $\delta = 0.8$ with a sample size of 1500, the authors obtained a mean APB of over 30% for δ for the EM method (over 70% for the LGMM method), relative to less than 1% APB in our more comprehensive spatial multinomial choice panel case with a true value of $\delta = 0.75$ and a sample size of 1000 (200 individuals x 5 choice occasions per individual). The FSSD values for δ are of the same order between the two methods, at about 1.6% of the true value for the EM method and 3.0% for our method. The FSSD for the LGMM method, though, is higher at about 15% of the true value. Additionally, both the EM and LGMM methods do not provide an estimate of the precision of δ, as estimated from a single sample (that is, they do not provide an asymptotic standard error estimate of δ), and the LGMM method does not guarantee that δ will be in the interval [-1,1]. The recursive importance sampling (RIS) technique of Beron and Vijverberg (2004) (which is a full-information maximum simulated likelihood technique), and the markov chain Monte Carlo (MCMC) Bayesian method of LeSage (2000) as implemented by

Thomas (2007) perform relatively well in CE's study for all values of the spatial autoregressive coefficient. For the low range of δ (δ=0.1 in CE's study and δ=0.25 in our study), the RIS has an APB for δ of about 4% compared to 19% for the MCMC and 0-5% in our study depending on the intensity of the spatial drift effect (see the APB values of δ in Tables 2a and 2b). The FSSD values, as a percentage of the true value, are in the 20-30% range for the RIS and MCMC methods for this low spatial autoregressive case compared to the 13-20% range (based on low or high spatial drift intensities) for our method. For the high range of δ (δ=0.8 in CE's study and δ=0.75 in our study), the RIS has an APB for δ of about 2% compared to 10% for the MCMC method and less than 1% for our method (regardless of spatial drift intensity). The FSSD values, as a percentage of the true value, are in the 2-4% range for the RIS and MCMC methods for this high spatial autoregressive case compared to 3% in our method.

In terms of the coefficients on the explanatory variables, CE's study indicates an APB in the 1-2% range for the EM, RIS, and MCMC methods, and 7% for the LGMM method, in the low spatial autoregressive coefficient case. This is as compared to an APB in the range of 0-3% in our approach. The FSSD are in the range of 8-16% of the true value for all methods. However, in the high spatial autoregressive case, the other methods have rather high APBs, ranging from 35% for the MCMC to 50% for the RIS to over 80% for the EM/GLMM methods. This is in contrast to an APB of 0-6% in our approach. The FSSD values are, in general, quite high and in the range of about 25% for the RIS and our method in this high spatial autoregressive case. Interestingly, the FSSD are smaller for the MCMC (FSSD of about 7%) and the EM/GLMM methods (FSSD of about 1%).

Overall, in terms of recovering the spatial autoregressive coefficient and the coefficients on the explanatory variables, our method does at least as well (and generally better than) the RIS and MCMC methods (both of which are known to have a downward sample bias), though these latter methods are unmanageable for the kinds of specifications and sample sizes on which we have implemented the MACML method in this paper. The appeal of our approach relative to the EM and GLMM methods should be quite obvious.

4.4.4. Comparison between the generalized spatial model and more restrictive models

In this section, we compare the performance of the generalized spatial (GS) model formulation with the more restrictive formulations, when the data generated actually conforms to the GS (see

Section 4.3). This comparison provides a sense of the biases that may accrue because of using a restrictive specification. Table 3 provides the results. As may be observed, two sets of mean APB values (across model parameters) are computed for the GS model, one for comparison with the "no spatial lag" (NSL) model and the second for comparison with the no spatial drift (NSD) model. For comparison with the NSL model, the mean APB value for the GS model is computed without considering the APB value for the δ parameter, because the δ parameter is fixed at zero in the NSL model. For comparison with the NSD model, the mean APB value for the GS model is computed without considering the APB values for the λ_1 and λ_2 parameters (both of which are fixed to zero in the NSD model).

The results indicate that the mean APB values are higher for the NSL and NSD models than for the GS model. Not surprisingly, the NSL model performs better in the two low spatial lag cases than in the two high spatial lag cases, since ignoring spatial lag dependence when such dependence is low should be of less consequence than ignoring such dependence when high. However, even in the two low spatial lag cases, the NSL model may be rejected compared to the "correct" GS specification based on the adjusted composite likelihood ratio test (ADCLRT) statistic (note that the GS specification rejects the simpler NSL specification for each of the twenty data sets generated). The NSL model performs very poorly for the two high spatial lag cases, with very poor ability to recover parameters. Similar results hold when comparing the NSD model (which ignores spatial drift) with the GS model, though the deterioration in the NSD model is not as substantial when moving from the low spatial drift case to the high spatial drift case (relative to the deterioration in the NSL model when moving from the low spatial lag to the corresponding high spatial lag case). Also, the NSD model performs better (in terms of mean APB) as compared to the NSL model for all four combinations of spatial lag and spatial drift parameters. But the NSL model too is rejected all twenty times in favor of the GS model for all four combinations, based on the ADCLRT test. An interesting suggestion from the simulation results above is also that ignoring the spatial lag effects (when present) is of much more serious consequence than ignoring spatial drift (*i.e.*, residential self-selection) effects (when present). Further theoretical and empirical exploration of this finding is left for future work.

Overall, the simulation results show that, irrespective of the magnitude of the spatial lag and spatial drift effects, the MACML estimator recovers the parameters of the proposed GS model very well. The MACML estimator also seems to be quite efficient, based on the low asymptotic standard error estimates of the parameters compared to the mean estimates of the parameters.

5. CONCLUSIONS

In many choice contexts, the discrete choice of one agent may be inter-related with those of others based on spatial and/or social proximity. The examination of such interactions is an area of active research in many fields, because a deeper understanding of the nature of such dependencies can be exploited by decision-makers to achieve desired system end-states in an efficient and cost-effective manner. A challenge, however, when investigating the issue of social/spatial interactions is to isolate these interactions from other "spurious" sources that may inappropriately get manifested as social/spatial interactions. These "spurious" sources can include correlation in unobserved factors of proximally located individuals and endogenous group formation. Thus, in a travel mode choice context, individuals and households with similar mode-use propensities may be drawn toward neighborhoods with specific observed built-environment attributes. That is, there is the possibility of residential self-selection of individuals based on mode-use propensities. This self-selection is introduced in our formulation in the form of a spatial (drift) structure on individual-specific random effects and sensitivities to observed exogenous factors. In this way, the formulation, at once, considers a whole suite of non-spatial and spatial considerations in a single unified panel multinomial probit model.

The paper proposes a maximum approximate composite marginal likelihood (MACML) inference approach to estimate the resulting discrete choice model. A simulation exercise is undertaken to evaluate the ability of the MACML approach to recover model parameters as well as to assess the empirical efficiency of the MACML estimator. The simulation results show that the MACML model recovers the parameters of the proposed model remarkably well. The MACML estimator is also quite efficient in the overall, and the asymptotic formula (based on the inverse of the Godambe information matrix) performs well in estimating the finite sample standard errors. In addition, the results clearly highlight the bias in estimates if spatial lag and/or spatial drift effects are ignored when both are actually present. Thus, in a travel mode choice context, the effects of built environment variables, and the intensity of social/spatial interactions and residential self-selection effects, can be mis-stated if the analyst ignores one or more of the many non-spatial and spatial considerations that may be at work. The result can be misinformed policies directed toward effecting changes in travel mode choice behavior.

Future efforts should focus on empirical applications of the proposed formulation, now that the feasibility of the approach has been demonstrated through a systematic simulation study. Indeed, it is hoped that our formulation and estimation approach will open the door for the extensive exploration of multiple aspatial and spatial effects impacting choice decisions.

ACKNOWLEDGEMENTS

This research was partially supported by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning (D-STOP) Tier 1 University Transportation Center. The author would also like to acknowledge support from a Humboldt Research Award from the Alexander von Humboldt Foundation, Germany. The author is grateful to Lisa Macias for her help in formatting this document, and to Chrissy Bernardo, Rajesh Paleti, and Subodh Dubey for GAUSS coding help and providing the results from the simulation runs. Four anonymous reviewers provided useful comments on an earlier version of this paper.

References

Anselin L (2003) Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review* 26: 153-166.

Bartolucci F, Nigro V (2010) A dynamic model for binary panel data with unobserved heterogeneity admitting a consistent conditional estimator. *Econometrica* 78(2): 719-733.

Beron KJ, Vijverberg WPM (2004) Probit in a spatial context: a Monte Carlo analysis. In: Anselin L, Florax RJGM, Rey SJ (eds) *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, Berlin.

Bhat CR (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7): 923-939.

Bhat, CR (2014) The composite marginal likelihood (CML) inference approach with applications to discrete and mixed dependent variable models. *Foundations and Trends in Econometrics* 7(1): 1-117.

Bhat CR, Guo JY (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B* 41(5): 506-526.

Bhat CR, Sardesai R (2006) The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B* 40(9): 709-730.

Bhat CR, Sener IN, Eluru N (2010) A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B* 44(8-9): 903-921.

Bhat CR, Elhorst JP, LeSage JP (2013) On spatial effects, interpretation, and estimation in econometric models. Paper based on a workshop held at the 9th Invitational Choice Symposium, Noordwijk, The Netherlands.

Bhat CR, Paleti R, Singh, P (2014a) A spatial multivariate count model for firm location decisions. *Journal of Regional Science* 54(3): 462-502.

Bhat CR, Astroza C, Sidharthan R, Jobair Bin Alam M, Khushefati WH (2014b) A joint count-continuous model of travel behavior with selection based on a multinomial probit residential density choice model. *Transportation Research Part B* 68: 31-51.

Bille AG (2013) Computational issues in estimation of the spatial probit model: A comparison of various estimators. *The Review of Regional Studies* 43: 131-154.

Blume LE, Brock WA, Durlauf SN, Ioannides YM (2011) Identification of social interactions. In: Behhabib J, Jackson, MO, Bisin A (eds) *Handbook of Social Economics Volume 1B*, Chapter 18: 853-964, North-Holland, San Diego, CA.

Bradlow ET, Bronnenberg B, Russell GJ, Arora N, Bell DR, Duvvuri SD, Hofstede FT, Sismeiro C, Thomadsen R, Yang S (2005) Spatial models in marketing. *Marketing Letters* 16(3): 267-278.

Bramoullé Y, Djebbari H, Fortin B (2009) Identification of peer effects through social networks. *Journal of Econometrics* 150: 41-55.

Brock W, Durlauf S (2001) Discrete choice with social interactions. *Review of Economic Studies* 68: 235-60.

Brock W, Durlauf S (2006) Multinomial choice with social interactions. In Blume L, Durlauf S (eds) *The Economy as an Evolving Complex System*, Vol. 3, Oxford University Press, New York.

Brock W, Durlauf S (2007) Identification of binary choice models with social interactions. *Journal of Econometrics* 140: 52-75.

Brunsdon C, Fotheringham S, Charlton M (1998) Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)* 47(3): 431-443.

Calabrese R, Elkink JA (2014) Estimators of binary spatial autoregressive models: A Monte Carlo study, *Journal of Regional Science* 54(4): 664-687.

Carrión-Flores C, Irwin EG (2004) Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics* 86(4): 889-904.

Castro, M, Paleti, R, and Bhat, CR (2012) A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. *Transportation Research Part B*, 46(1): 253-272.

Chakir R, Parent O (2009) Determinants of land use changes: A spatial multinomial probit approach. *Papers in Regional Science* 88(2): 327-344.

Chamberlain, G (2010) Binary response models for panel data: identification and information, *Econometrica* 78(1), 159-168.

Cox DR, Reid N (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3): 729-737.

Davezies L, d´ Haultfoeuille X, Foughre D (2009) Identification of peer effects using group size variation. *Econometrics Journal*: *3*97-413.

Christakis NA, Fowler JH (2013) Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine* 32(4): 556–577.

Elhorst JP (2010) Applied spatial econometrics: raising the bar. *Spatial Economic Analysis* 5(1): 10-28.

Franzese RJ, Hays JC, Schaffer LM (2010) Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. APSA 2010 Annual Meeting Paper. Available at: http://ssrn.com/abstract=1643867.

Godambe VP (1960) An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4): 1208–1211.

Guo JY, Bhat CR (2007) Operationalizing the concept of neighborhood: application to residential location choice analysis. *Journal of Transport Geography* 15(1): 31-45

Hartman WR, Manchanda P, Nair H, Bothner M, Dodds P, Godes D, Hosanagar K, Tucker C (2008) Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Letters* 19(3-4): 287-304.

Klier T, McMillen DP (2008) Clustering of auto supplier plants in the U.S.: GMM spatial logit for large samples. *ASA Journal of Business & Economic Statistics* 26(4): 460-471.

Krauth B (2006) Social interactions in small groups. *Canadian Journal of Economics* 39: 414-433.

Lee L-F (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140: 333-374.

Lee L-F, Liu X, Lin X (2010) Specification and estimation of social interaction models with network structure: contextual factors, correlation and fixed effects. *Econometrics Journal* 13: 145-176.

LeSage JP (2000) Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32(1): 19-35.

LeSage JP, Pace RK (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton.

LeSage JP, Pace RK, Lam, N Campanella, R and Liu, X (2011) New Orleans business recovery in the aftermath of Hurricane Katrina, *Journal of the Royal Statistical Society: Series A*, 174(4): 1007-1027.

Lindsay BG (1988) Composite likelihood methods. *Contemporary Mathematics* 80: 221-239.

Lindsay BG, Yi GY, Sun J (2011) Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* 21(1): 71-105.

Manski CF (1993) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3): 531-542.

McFadden D, Train K (2000) Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15(5): 447-470.

McMillen DP (1992) Probit with spatial autocorrelation. *Journal of Regional Science* 32: 335-348.

Mittal V, Kamakura WA, Govind R (2004) Geographic patterns in customer service and satisfaction; an empirical investigation. *Journal of Marketing* 68: 48-62.

Moffitt R (2001) Policy interventions, low-level equilibria, and social interactions. In Durlauf S, Young HP (eds) *Social Dynamics*, MIT Press, Cambridge.

Mokhtarian PL, Cao X (2008) Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B* 42(3): 204-228.

Molenberghs G, Verbeke G (2005) *Models for Discrete Longitudinal Data*. Springer Series in Statistics. Springer Science + Business Media, Inc., New York.

Pace L, Salvan A, Sartori N (2011) Adjusting composite likelihood ratio statistics. *Statistica Sinica* 21(1): 129-148.

Pace RK, LeSage JP (2011) Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. Available at SSRN: http://ssrn.com/abstract=1966039

Pinjari AR, Bhat CR (2006) On the nonlinearity of response to level of service variables in travel mode choice models. *Transportation Research Record* 1977: 67-74

Sándor Z, Train K (2004) Quasi-random simulation of discrete choice models. *Transportation Research B* 38: 313-327.

Sener IN, Bhat CR (2012) Flexible spatial dependence structures for unordered multinomial choice models: Formulation and application to teenagers' activity participation. *Transportation* 39(3): 657-683.

Sidharthan R, Bhat CR (2012) Incorporating spatial dynamics and temporal dependency in land use change models. *Geographical Analysis* 44(4): 321-349.

Smirnov OA (2010) Spatial econometrics approach to integration of behavioral biases in travel demand analysis. *Transportation Research Record* 2157: 1-10.

Steglich C, Snijders TAB, Pearson M (2010) Dynamic networks and behavior: separating selection from influence. *Sociological Methodology* 40(1): 329-393.

Soetevent A, Kooreman P (2007) A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics* 22: 599 624.

Thomas TS (2007) Impact of economic policy options on deforestation in Madagascar, Companion Paper for the Policy Research Report on Forests, Environment, and Livelihoods, The World Bank, Washington DC. Accessed October 11, 2014: http://siteresources.worldbank.org/INTKNOWLEDGEFORCHANGE/Resources/491519-1199818447826/a_primer_for_bayesian_spatial_probits_with_an_application_to_deforestation_in_madagascar.pdf.

Tsai Y (2009) Impacts of self-selection and transit proximity on commute mode choice: Evidence from Taipei Rapid Transit System. *Annals of Regional Science* 43(4): 1073-94.

Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Statistica Sinica* 21(1): 5-42.

Xu X, Reid N (2011) On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference* 141(9): 3047-3054.

Yi GY, Zeng L, Cook RJ (2011) A robust pairwise likelihood method for incomplete longitudinal binary data arising in clusters. *The Canadian Journal of Statistics* 39(1): 34-51.

Zhao Y, Joe H (2005) Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3): 335-356.

LIST OF FIGURES

LIST OF TABLES

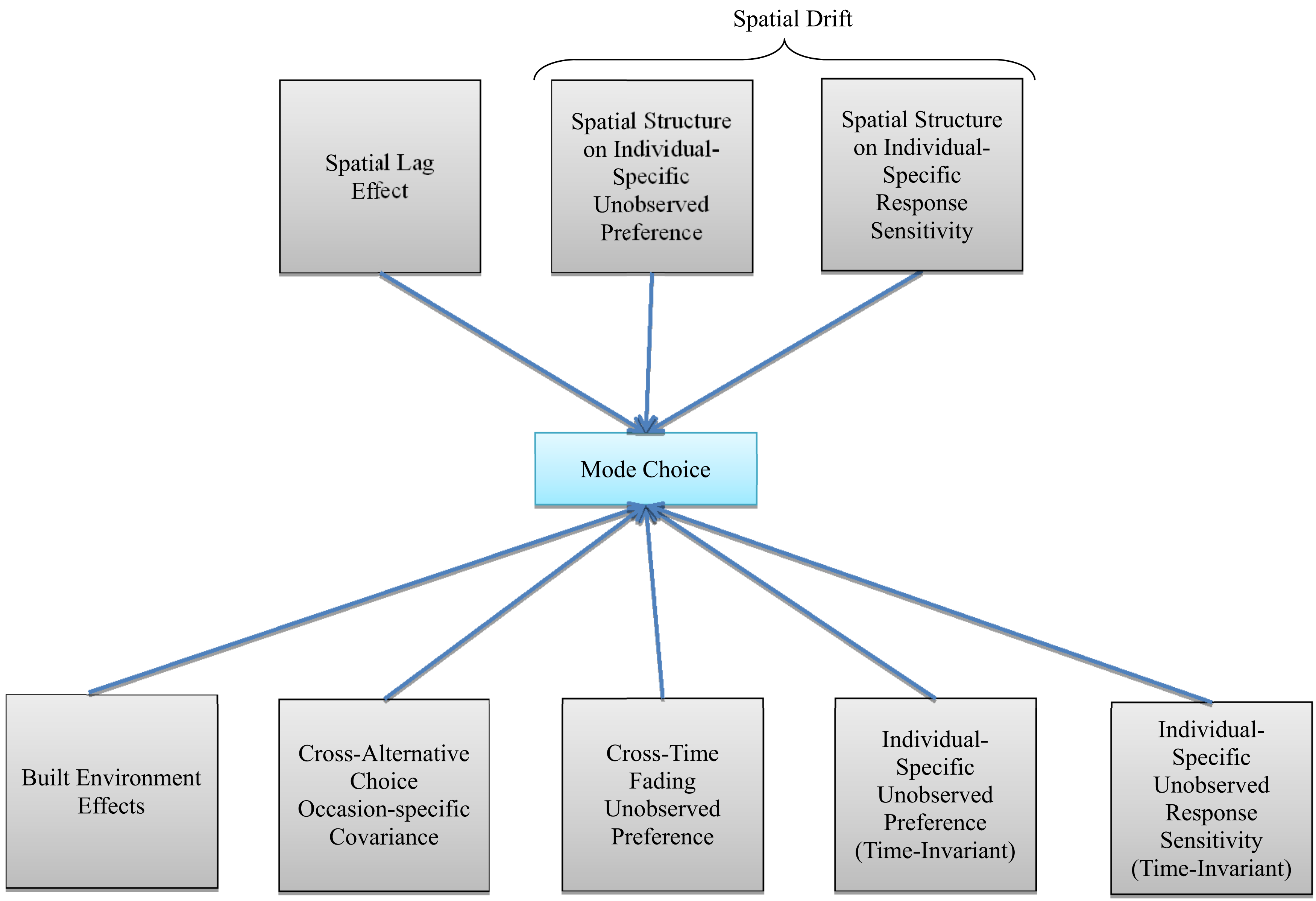

Figure 1: Components of a generic mode choice model

Table 1: Restricted versions of proposed model structure

Restriction	Conclusion if the restriction is true
$\delta = 0$	No spatial lag (spatial drift due to residential self-selection, and time-varying and time-invariant panel effects, still present)
$\theta = 0$ and $\lambda_k = 0 \; \forall k$	No spatial drift (residential self-selection) effects
$\rho = 0$	No time-varying unobserved effects
$\tilde{\mathbf{\Lambda}} = \mathbf{0}_{IxI}$	No intrinsic preference effects, and cannot identify spatial drift (residential self-selection) effects based on intrinsic preferences (*i.e.* θ parameter)
$\tilde{\mathbf{\Omega}} = \mathbf{0}_{KxK}$	No heterogeneity in response to exogenous variables, and cannot identify spatial drift (self-selection) effects based on exogenous variables (*i.e.*, the $\boldsymbol{\lambda}$ matrix parameters)
All of the above	Simple cross-sectional MNP

Table 2a: Simulation results for the four-alternative case with 20 datasets for low spatial dependency and low spatial drift (based on a total of 20×10 runs/dataset=200 runs)

Parameter	True Value	Parameter Estimates			Standard Error-related Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Dev (FSSD)	Asymptotic St. Err. (ASE)	APB of the asymt. St. Err. (APBASE)	Approx. error as a % of FSSD (APERR)
b_1	0.50	0.502	0.002	0.304	0.0433	0.0431	0.46	3.00
b_2	0.80	0.791	0.009	1.074	0.1005	0.0948	5.67	1.69
b_3	1.00	0.977	0.023	2.288	0.1107	0.127	14.72	1.90
$l_{\tilde{\mathbf{\Omega}}1}$	0.90	0.876	0.024	2.696	0.1217	0.1332	9.45	2.88
$l_{\tilde{\mathbf{\Omega}}2}$	0.60	0.614	0.014	2.297	0.1082	0.1186	9.61	2.96
$l_{\tilde{\mathbf{\Omega}}3}$	0.80	0.794	0.006	0.787	0.1436	0.1142	20.47	2.50
δ	0.25	0.252	0.002	0.789	0.0497	0.0455	8.45	1.21
λ_2	0.25	0.251	0.001	0.403	0.0035	0.0036	2.85	42.86
λ_3	0.25	0.246	0.004	1.753	0.0187	0.0163	12.84	39.57
Overall mean value across parameters			0.009	1.377	0.0778	0.0774	9.39	10.95

[a] The mean composite log-likelihood value for the low spatial dependency and low spatial drift model at converged parameter is -19710.55

Table 2b: Simulation results for the four-alternative case with 20 datasets for low spatial dependency and high spatial drift (based on a total of 20×10 runs/dataset=200 runs)

Parameter	True Value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Dev (FSSD)	Asymptotic St. Err. (ASE)	Abs.Bias of the asymt. St. Err. (APBASE)	Approx. error as a % of FSSD (APERR
b_1	0.50	0.495	0.005	1.038	0.0494	0.0397	19.64	3.64
b_2	0.80	0.791	0.009	1.131	0.2736	0.3075	12.39	1.24
b_3	1.00	0.965	0.035	3.478	0.3749	0.3523	6.03	1.04
$l_{\tilde{\Omega}1}$	0.90	0.817	0.083	9.203	0.1477	0.1709	15.71	4.20
$l_{\tilde{\Omega}2}$	0.60	0.645	0.045	7.526	0.1545	0.1590	2.91	3.75
$l_{\tilde{\Omega}3}$	0.80	0.753	0.047	5.852	0.1332	0.1261	5.33	5.25
δ	0.25	0.261	0.011	4.223	0.0329	0.0300	8.88	3.95
λ_2	0.75	0.679	0.071	9.453	0.1556	0.1626	4.31	11.44
λ_3	0.75	0.679	0.071	9.504	0.1465	0.1509	3.00	8.46
Overall mean value across parameters			0.042	5.712	0.1631	0.1665	8.69	4.77

[a] The mean composite log-likelihood value for the low spatial dependency and low spatial drift model at converged parameter is -19769.08

Table 2c: Simulation results for the four-alternative case with 20 datasets for high spatial dependency and low spatial drift (based on a total of 20×10 runs/dataset=200 runs)

Parameter	True Value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Dev (FSSD)	Asymptotic St. Err. (ASE)	Abs.Bias of the asymt. St. Err. (APBASE)	Approx. error as a % of FSSD (APERR
b_1	0.50	0.502	0.002	0.378	0.1077	0.1174	9.00	3.06
b_2	0.80	0.806	0.006	0.796	0.1173	0.1367	16.54	3.84
b_3	1.00	0.980	0.020	2.024	0.1304	0.1323	1.46	5.52
$l_{\tilde{\Omega}1}$	0.90	0.898	0.002	0.183	0.2234	0.2442	9.31	3.36
$l_{\tilde{\Omega}2}$	0.60	0.592	0.008	1.401	0.1388	0.1236	10.95	4.90
$l_{\tilde{\Omega}3}$	0.80	0.790	0.010	1.299	0.1851	0.1705	7.88	4.75
δ	0.75	0.746	0.004	0.562	0.0229	0.0274	19.65	3.49
λ_2	0.25	0.248	0.002	0.729	0.0182	0.0201	10.44	10.44
λ_3	0.25	0.249	0.001	0.394	0.0133	0.0173	30.07	6.01
Overall mean value across parameters			0.006	0.863	0.1063	0.1099	12.81	5.04

[a] The mean composite log-likelihood value for the low spatial dependency and low spatial drift model at converged parameter is -11532.19

Table 2d: Simulation results for the four-alternative case with 20 datasets for high spatial dependency and high spatial drift (based on a total of 20×10 runs/dataset=200 runs)

Parameter	True Value	Parameter Estimates			Standard Error Estimates			
		Mean Est.	Abs. Bias	Absolute Percentage Bias (APB)	Finite Sample St. Dev (FSSD)	Asymptotic St. Err. (ASE)	Abs.Bias of the asymt. St. Err. (APBASE)	Approx. error as a % of FSSD (APERR
b_1	0.50	0.514	0.014	2.732	0.1155	0.1122	2.86	1.47
b_2	0.80	0.843	0.043	5.370	0.2605	0.2396	8.02	1.27
b_3	1.00	1.014	0.014	1.391	0.2627	0.2924	11.31	1.37
$l_{\tilde{\Omega}1}$	0.90	0.925	0.025	2.824	0.2341	0.2506	7.05	1.75
$l_{\tilde{\Omega}2}$	0.60	0.628	0.028	4.705	0.1649	0.1637	0.73	2.00
$l_{\tilde{\Omega}3}$	0.80	0.769	0.031	3.887	0.1870	0.2294	14.77	2.46
δ	0.75	0.754	0.004	0.540	0.0210	0.0148	29.52	0.95
λ_2	0.75	0.680	0.070	9.333	0.1721	0.1748	1.57	3.37
λ_3	0.75	0.696	0.054	7.165	0.1470	0.1317	10.41	0.88
Overall mean value across parameters			0.031	4.216	0.1739	0.1788	9.58	1.72

[a] The mean composite log-likelihood value for the low spatial dependency and low spatial drift model at converged parameter is -11693.61

Table 3: Effects of ignoring spatial effects when present

Evaluation Metric	$\delta = 0.25, \lambda_1 = \lambda_2 = 0.25$ **Low spatial lag- Low spatial drift**			$\delta = 0.25, \lambda_1 = \lambda_2 = 0.75$ **Low spatial lag-High spatial drift**			$\delta = 0.75, \lambda_1 = \lambda_2 = 0.25$ **High spatial lag- Low spatial drift**			$\delta = 0.75, \lambda_1 = \lambda_2 = 0.75$ **High spatial lag- High spatial drift**		
	GS Model	**NSL Model**	**NSD Model**	**GS Model**	**NSL Model**	**NSD Model**	**GS Model**	**NSL Model**	**NSD Model**	**GS Model**	**NSL Model**	**NSD Model**
Mean APB												
For comparison of GS model with the no-spatial lag (NSL) model	1.23	2.88	-	5.50	7.89	-	0.80	41.93	-	6.24	44.21	-
For comparison of GS model with the no-spatial drift (NSD) model	1.50	-	3.03	4.58	-	6.66	0.99	-	2.36	3.05	-	3.72
Mean composite log-likelihood value at convergence	-19,711	-20,047	-19,946	-19,768	-19,997	-19,914	-11,535	-21,590	-11,735	-11,694	-21,915	-11,929
Number of times the adjusted composite likelihood ratio test (ADCLRT) statistic favors the ORSH model	-	All twenty times when compared with $\chi_1^2 = 6.64$ value (mean ADCLRT statistic is 166.49)	All twenty times when compared with $\chi_2^2 = 9.21$ value (mean ADCLRT statistic is 135.63)	-	All twenty times when compared with $\chi_1^2 = 6.64$ value (mean ADCLRT statistic is 127.03)	All twenty times when compared with $\chi_2^2 = 9.21$ value (mean ADCLRT statistic is 109.96)	-	All twenty times when compared with $\chi_1^2 = 6.64$ value (mean ADCLRT statistic is 267.76)	All twenty times when compared with $\chi_2^2 = 9.21$ value (mean ADCLRT statistic is 118.10)	-	All twenty times when compared with $\chi_1^2 = 6.64$ value (mean ADCLRT statistic is 287.62)	All twenty times when compared with $\chi_2^2 = 9.21$ value (mean ADCLRT statistic is 166.80)