Transportation Research Part B 191 (2025) 103132

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Hierarchical Nearest Neighbor Gaussian Process models for discrete choice: Mode choice in New York City

Daniel F. Villarraga [*], Ricardo A. Daziano

School of Civil and Environmental Engineering, Cornell University, 220 Hollister Hall, Ithaca, NY 14853, USA

ARTICLE INFO

Keywords:
Hierarchical Gaussian process
Nearest Neighbor Gaussian Process (NNGP)
Spatially autoregressive models
Interpretable machine learning
Discrete choice
Travel mode choice
Value of Travel Time (VOTT)

ABSTRACT

Standard Discrete Choice Models (DCMs) assume that unobserved effects that influence decision-making are independently and identically distributed among individuals. When unobserved effects are spatially correlated, the independence assumption does not hold, leading to biased standard errors and potentially biased parameter estimates. This paper proposes an interpretable Hierarchical Nearest Neighbor Gaussian Process (HNNGP) model to account for spatially correlated unobservables in discrete choice analysis. Gaussian Processes (GPs) are often regarded as lacking interpretability due to their non-parametric nature. However, we demonstrate how to incorporate GPs directly into the latent utility specification to flexibly model spatially correlated unobserved effects without sacrificing structural economic interpretation. To empirically test our proposed HNNGP models, we analyze binary and multinomial mode choices for commuting to work in New York City. For the multinomial case, we formulate and estimate HNNGPs with and without independence from irrelevant alternatives (IIA). Building on the interpretability of our modeling strategy, we provide both point estimates and credible intervals for the value of travel time savings in NYC. Finally, we compare the results from all proposed specifications with those derived from a standard logit model and a probit model with spatially autocorrelated errors (SAE) to showcase how accounting for different sources of spatial correlation in discrete choice can significantly impact inference. We also show that the HNNGP models attain better out-of-sample prediction performance when compared to the logit and probit SAE models, especially in the multinomial case.

1. Introduction

When individuals face decisions that involve a finite number of alternatives, such as selecting a transportation mode for daily work commutes, they engage in a decision-making process influenced by personal preferences and the characteristics of each available option. To analyze these processes, researchers have developed Discrete Choice Models (DCMs) that assume individuals aim to maximize a latent utility function that summarizes preferences. Typically, the latent utility in DCMs includes a random component or preference shock term that accounts for unobserved effects. The simplest discrete choice models, such as the conditional logit, assume independent and identically distributed error terms. More complex models, such as the multinomial probit and random parameter logit, allow for increased flexibility in substitution patterns. It is not uncommon for researchers to model correlation and heteroskedasticity across alternatives or across observations of the same individual. However, standard discrete choice models usually ignore dependencies across individuals.

[*] Corresponding author.
E-mail addresses: dv275@cornell.edu (D.F. Villarraga), daziano@cornell.edu (R.A. Daziano).

https://doi.org/10.1016/j.trb.2024.103132
Received 30 October 2023; Received in revised form 8 November 2024; Accepted 17 November 2024

Available online 28 November 2024

Although standard DCMs assume that the unobserved effects influencing decision-making are independent and identically distributed among individuals, there are certain choice situations where this assumption may not hold true. In the case of selecting a transportation mode for daily work commutes, for example, there may be spatially correlated unobserved effects that induce correlation between individuals. These effects could include parking costs, the crime rate at public transit stations, or driving congestion for some trip directions. For instance, in cities such as NYC, parking-meter rates depend on land use, density, and parking demand (New York City Department of Transportation, 2023). Similarly, public transit crowding exhibits a spatially correlated structure (e.g. Li and Hensher 2013), and if the transit system is overcrowded at the trip origin location, individuals may be less likely to opt for public transit.

When the unobserved effects are correlated between individuals, standard DCMs may fail in several ways. Problems that arise from ignoring these dependencies include statistical inefficiency and biased standard error estimates. And more importantly, when the latent utility comprises variables that share the same correlation structure as the error term, ignoring the correlation between individuals leads to biased and inconsistent parameter estimates (Anselin et al., 2013).

The spatial econometrics literature provides two primary models for accounting for correlation between individuals, namely: the Spatially Autoregressive Error (SAE) model and the Spatially Autoregressive Lag (SAL) model (Anselin et al., 2013). The SAE model attributes correlation between individuals to unobserved effects, such as spatially correlated unobservables (e.g, parking costs or traffic congestion). In contrast, the SAL model assumes that correlation arises from individual interactions, such as social influence or peer effects. Both models have a non-spherical variance–covariance structure, and may exhibit heteroskedasticity. In practice, the variance–covariance between individuals is defined only for the observations in the dataset used for estimation, making these models useful for inference but their use for out-of-sample inductive prediction is not straightforward. For instance, in Goulard et al. (2017), the authors study methods for out-of-sample predictions for the SAE model.

There has been some interest from the discrete choice community in accounting for correlations among individuals that arise either from social influence or from correlated unobservables. For instance, in Bhat (2015), the authors examine commute mode choice using a model that includes the spatial lag effect and spatially correlated error terms, similar to the SAL and SAE models from the spatial econometrics literature (Anselin, 2001). In their model, they account for residential self-selection by allowing random taste variations to be spatially correlated. To our knowledge, their model is one of the most comprehensive specifications that account for correlated individuals in the current literature. In Bhat et al. (2016), the authors model spatial dependence between individuals by employing a spatially auto-correlated latent construct in a generalized heterogeneous data model, and in Bhat et al. (2010), the authors propose an estimation approach for ordered response DCMs with spatial dependence structures that are isotropic and stationary. For the frequentist estimation of these models, the authors rely on composite marginal likelihood (CML) estimators that do not require simulation, which becomes impractical for moderate to large datasets (Bhat, 2011).

In this work, we propose Hierarchical Gaussian Process (HGP) models that account for correlations between individuals and generalize both the SAE and SAL models — and their extensions (Bhat, 2015). Specifically, we use the HGP model as an extension of the SAE model to study mode choice for commuting to work in New York City. The dataset we use in this study includes information from the most recent available regional household travel survey from the New York Metropolitan Transportation Council (New York Metropolitan Transportation Council, North Jersey Transportation Planning Authority, 2014), as well as estimated travel times and costs from the Google API. Our dataset includes socio-demographic variables, trip origins and destinations, approximate travel times and trip costs per travel mode, and travel mode selections for over 2400 work-related trips for the binary mode selection problem (i.e., public transit vs. private car), and more than 3200 work-related trips for the multinomial case (i.e., public transit, private car, and non-motorized).

In contrast to previous work on DCMs with correlated individuals [e.g., Bhat et al. (2016, 2010)], we use a Bayesian estimation framework for our proposed HGP models. We exploit Markov Chain Monte Carlo (MCMC) and the Nearest Neighbor Gaussian Process (NNGP) formulation (Datta et al., 2016), which extends Vecchia's original approximation (Vecchia, 1988) and significantly reduces estimation time. We provide point estimates and credible intervals for the value of time, alternative attribute and socio-demographic associated parameters, and spatial process parameters (e.g., the correlation decay rate and scale of the spatially correlated variability).

In a previous study of binary mode choice in New York City, Goetzke (2008) estimated a conditional approximation of the SAL model and found that the spatial effects importance was statistically significant. However, their analysis only accounted for unobserved spatially correlated effects at trip origins and did not include trip cost or travel time. In contrast, our study estimates and compares HGP models that generalize the SAE model, accounting for unobserved spatially correlated effects at both trip origins and destinations. We also consider a model in which autocorrelation arises from unobserved effects related to the direction and length of the trips. Furthermore, our analysis includes travel time and trip cost, providing point estimates and credible intervals for the value of time in New York City.

Whereas Hierarchical Gaussian Processes are not commonly used in econometrics, there are some examples of their integration in the discrete choice literature. For instance, in Sfeir et al. (2022), the authors introduced a GP latent class choice model (LCCM) where the class membership model follows a Gaussian Process, and the class-specific choice model was a standard DCM. The authors assumed that the latent utilities in the class-specific choice models were independent between individuals. Their model allowed for greater heterogeneity representation, but it came with a loss in interpretation due to the non-parametric definition of class membership using socio-demographic variables for the GP. In contrast, our model incorporates GPs directly into the utility function specification, including socio-demographic variables in a parametric way. This strategy allows us to gain insight into the influence of socio-demographics on individual choices while using GPs to model spatially correlated unobservables flexibly.

In Guo et al. (2010), the authors used GPs to model individual utilities in a non-parametric fashion, incorporating both alternative and individual similarities. Whereas this approach is highly flexible and yields good results for prediction, it is limited in terms of interpretability. As a result, their proposed model is well-suited for preference elicitation in recommendation systems but may need modification for deriving useful economic information, such as willingness to pay. In contrast, our model prioritizes straightforward inference over flexibility in the utility function specification. By sacrificing some of the flexibility in the model, we can maintain interpretability an regularity of the parameters of our models, which is essential for econometric analysis. However, we show that our models achieve better out-of-sample prediction performance than other DCMs, particularly in the multinomial case, where it surpasses the multinomial logit model's performance by around 10 percentage points

In Section 2, we provide a brief overview of standard discrete choice models. In Section 3, we discuss the SAE and SAL models. We offer a brief overview of discrete choice models (DCMs) with network effects in Section 4. In Section 5, we present the general Hierarchical Gaussian Process (HGP) model used in this work, discuss the HGP equivalents to SAE and SAL models, and describe the Nearest Neighbor Gaussian Process formulation. After that, in Section 6, we extend the HGP formulation to the multinomial case with and without independence from irrelevant alternatives (IIA). In Section 7, we present the formulation, diagnostics, and results for a binary mode choice study in NYC. Conversely, in Section 8, we apply our models to the multinomial mode choice problem. Finally, in Section 9, we provide conclusive remarks and discuss potential avenues for future research.

2. Standard discrete choice models

In the discrete choice framework employed in this work, we assume that when faced with the decision of choosing between public transit or a private car for commuting to work, an individual i will base their decision y_i on the maximization of their own utility u_i. This utility comprises a deterministic part that considers the individual's characteristics and the differences in attributes between the alternatives, as well as a random part that captures unobservables.

In a standard binary discrete choice setting, the latent utility u_i that an individual i derives from choosing a given alternative (e.g., public transit) over another (e.g., private car) is represented by Eq. (1):

$$u_i = v_i + \epsilon_i, \tag{1}$$

where ϵ_i represents the random preference shock used to account for unobserved effects, and v_i is the deterministic part of utility, as given by Eq. (2) when working with a linear specification:

$$v_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \boldsymbol{q}_i^T \boldsymbol{\gamma}, \tag{2}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are estimable vectors of parameters, $\boldsymbol{x}_i$ is a vector that represents the differences between the mode attributes (e.g., the difference in travel time), and $\boldsymbol{q}_i$ is a vector of individual characteristics (e.g., household income). In basic binary choice models, the random component of the utility ϵ_i is typically assumed to be independent and identically distributed, following either a standard normal distribution (probit model) or a standard logistic distribution (logistic regression).

In this binary setting, the probability of choosing public transit (i.e., $y_i = 1$) is given by Eq. (3):

$$P(y_i = 1) = P(\epsilon_i \leq v_i), \tag{3}$$

and the likelihood, with n individuals, for this model specification is given below (Eq. (4)):

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} P(\epsilon_i \leq v_i)^{y_i} (1 - P(\epsilon_i \leq v_i))^{1-y_i}. \tag{4}$$

For a more detailed and general description of standard discrete choice models, refer to Train (2009).

3. Spatially autoregressive models

To account for spatial correlation in unobservables within econometric settings, there are two main models that work for both continuous and limited dependent variables (Anselin et al., 2013). In this paper, we describe these approaches in the context of discrete dependent variables. The main difference between the two models is the assumption of the source of correlation for the underlying index function that is continuous (i.e., the individual utilities in the case of discrete choice). The Spatially Autoregressive Error (SAE) model assumes that correlation comes exclusively from spatially correlated errors, whereas the Spatially Autoregressive Lag (SAL) model assumes that autocorrelation arises from the interaction between individual utilities. In both cases, the resulting model has a non-spherical variance–covariance structure that can represent heteroskedasticity.

On the one hand, the SAE model, which involves n individuals, k mode attributes, and l socio-demographic variables, can be represented using the following matrix form for the latent utility, as shown in Eq. (5):

$$\boldsymbol{u} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\gamma} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{\epsilon}, \tag{5}$$

where, $\boldsymbol{X}$ is an $n \times k$ matrix that contains alternative attribute differences for each individual, $\boldsymbol{Q}$ is an $n \times l$ matrix that represents individual characteristics, $\boldsymbol{I}$ is an $n \times n$ identity matrix, ρ is a scalar parameter that reflects the importance of spatial effects, $\boldsymbol{W}$ is an $n \times n$ adjacency matrix that models individuals' proximity, and $\boldsymbol{\epsilon}$ is an error vector that contains n uncorrelated elements.

The SAL model, on the other hand, has a latent utility matrix form shown in Eq. (6):

$$\boldsymbol{u} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{Q}\boldsymbol{\gamma} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{\epsilon}. \tag{6}$$

If we assume that $\boldsymbol{\epsilon}$ follows a multivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, then the random utilities would follow multivariate normal distributions with non-spherical variance–covariance structure. For the SAE model, the random utility would follow:

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\gamma}, (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}(\boldsymbol{I} - \rho \boldsymbol{W})^{-\top}). \tag{7}$$

Similarly, for the SAL model:

$$\boldsymbol{u} \sim \mathcal{N}((\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{Q}\boldsymbol{\gamma}, (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}(\boldsymbol{I} - \rho \boldsymbol{W})^{-\top}). \tag{8}$$

Since the individual utilities in both models are not independent, evaluation of the likelihood function would involve computing an n-dimensional integral with no closed-form solution. Therefore, estimating the parameters in these models requires methods such as the expectation–maximization algorithm, MCMC, maximum simulated likelihood or composite marginal likelihood approaches, as described in Anselin et al. (2013).

4. State-of-the-art DCMs with network effects

As previously discussed, the model proposed by Bhat (2015) is arguably one of the most comprehensive specifications for discrete choice models accounting for correlations between observations. It works with panel data, models peer effects/social influence, and accounts for self-selection effects by allowing for correlated random preference heterogeneity. A simplified binary specification of that model in matrix form, reminiscent of a more general form of the SAE and SAL models, follows the equation presented below:

$$\boldsymbol{u} = \rho \boldsymbol{W}\boldsymbol{u} + \boldsymbol{X}(b + (\boldsymbol{I} - \rho_\beta \boldsymbol{W})^{-1}\boldsymbol{\tau}_\beta) + \boldsymbol{Q}\gamma + (\boldsymbol{I} - \rho_\epsilon \boldsymbol{W})^{-1}\epsilon \tag{9}$$

In this specification, the social part of the utility is modeled as in the SAL model, and the correlated unobserved effects are modeled as in the SAE model. However, there are a couple of terms not present in either the SAE or SAL models. These terms are the ones used to account for self-selection effects:

$$\beta = b + (\boldsymbol{I} - \rho_\beta \boldsymbol{W})^{-1}\boldsymbol{\tau}_\beta \tag{10}$$

where $\boldsymbol{\tau}_\beta$ is an uncorrelated random vector of size n, b models the expected marginal utilities with respect to either alternative attributes, and ρ_β models the importance of self-selection based on mode attributes. With $\boldsymbol{\tau}_\beta$ and ϵ normal random variables, this latent utility specification would follow a normal distribution with non-spherical variance covariance matrix as the SAE and SAL models.

It is evident that restricting this specification allows us to recover either the SAE or SAL models, as well as the standard discrete choice model presented earlier. To illustrate this, setting $\rho = 0$ and $\rho_\beta = 0$ would yield the SAE model, and further setting $\rho_\epsilon = 0$ would yield the standard choice model discussed at the beginning of this paper. In this model, if matrix $\boldsymbol{W}$ is populated with inverse distances between observations, the underlying spatial processes are assumed to be isotropic and stationary. Additionally, similar to the SAE and SAL models, the only spatial process parameters are the importance parameters ρ associated with the peer effects, unobserved spatial effects, or residential self-selection. Estimation for these models in the frequentist setting can be done using CML-based approaches as proposed by Bhat (2011).

5. Hierarchical Gaussian process models

As mentioned above, both the SAE and SAL models assume that the latent utility vector follows a multivariate normal distribution, where the variance–covariance depends on the individuals' proximity, represented by the adjacency matrix $\boldsymbol{W}$, and the importance of spatial effects, denoted by ρ. It is worth noting that the variance–covariance is defined for n observations, which are typically the ones used for estimation in practice. Additionally, with this model structure, the underlying spatial process is assumed to be isotropic and stationary with a single estimable parameter ρ. Relaxing the isotropic and stationary assumption might not be trivial and would involve modifying the spatial-proximity function used to populate $\boldsymbol{W}$. Depending on the way that $\boldsymbol{W}$ is constructed, predictions for unseen locations might not be straightforward to compute.

Multivariate normal distributions are limited to n dimensions, whereas Gaussian processes (GPs) can describe infinite-dimensional random variables. Formally, a GP is a stochastic process from which any collection of random variables follows a multivariate normal distribution (MacKay and Mac Kay, 2003). Unlike multivariate normal distributions, which require a mean vector and a variance–covariance matrix, GPs are characterized by mean and variance–covariance functions – that can even model non-isotropic and non-stationary spatial processes – over a continuous domain. This fact makes it possible to think of GPs as distributions over functions in a continuous domain Wilson and Izmailov (2020), such as those defined by location coordinates.

In this paper, we propose the use of hierarchical Gaussian processes (HGPs) to account for spatially correlated unobserved effects in binary choice settings. The term "hierarchical" refers to the fact that we model latent random utilities u_i using a GP, and then use a link function to describe their relationship with the observed decisions y_i. Our general model expresses the latent utility u_i for each individual i as follows:

$$u_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{q}_i^T\boldsymbol{\gamma} + Z(\boldsymbol{s}_i). \tag{11}$$

Here, $\boldsymbol{x_i}$ denotes a vector of alternative attribute differences (e.g., between public transit and private car), $\boldsymbol{q_i}$ represents a vector of socio-demographic characteristics for individual i, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are parameter vectors, $\boldsymbol{s_i}$ is a location or direction vector (e.g., the origin coordinates of the trip), and $Z \sim GP(0, k(\boldsymbol{s_i}, \boldsymbol{s_j}|\boldsymbol{\theta}))$ is a GP that models unobserved effects. The covariance function $k(\boldsymbol{s_i}, \boldsymbol{s_j}|\boldsymbol{\theta})$ (i.e. kernel) with parameters $\boldsymbol{\theta}$ gives the covariance between any two individual pairs (i, j). For instance, in the empirical study in this paper, we use the exponential kernel with a nugget parameter, defined below:

$$k(s_i, s_j|\sigma^2, \phi, \tau^2) = \sigma^2 \exp\left(-\phi\|\boldsymbol{s_i} - \boldsymbol{s_j}\|\right) + \tau^2, \tag{12}$$

with scale parameter for the spatially correlated effects σ^2, correlation decay rate ϕ, and nugget τ^2. The nugget parameter τ^2 is used to model the non-correlated unobserved effects, and its value is set to one to normalize the scale of the utilities. Other potential choices for the kernel include the square exponential kernel, often referred to as the radial basis function kernel in the machine learning literature.

The observed mode choices y_i are distributed Bernoulli $y_i \sim \text{Ber}(\Lambda(u_i))$, and the conditional probability of choosing the alternative of interest (e.g., transit over private car) for the ith individual is then given by:

$$P(y_i = \text{Transit}|\boldsymbol{q_i}, \boldsymbol{x_i}, Z) = \Lambda(u_i), \tag{13}$$

where $\Lambda(\cdot)$ is a link function that maps the real line to $(0, 1)$ (e.g. logistic).

The likelihood for the model with n observations takes the form described below:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}) = \int\int \cdots \int f(u_1, u_2, \ldots, u_n) \prod_{i=1}^{n} \Lambda(u_i)^{y_i}(1 - \Lambda(u_i))^{1-y_i} du_1 du_2 \ldots du_n. \tag{14}$$

Here, $f(u_1, u_2, \ldots, u_n)$ is the probability density function of a multivariate normal distribution. Unfortunately, the n-dimensional integral in the likelihood (as presented in Eq. (14)) does not have a closed-form solution. Therefore, estimation once again requires using methods such as the EM algorithm or MCMC. In this paper, we implement MCMC with the Nearest Neighbor Gaussian Process formulation, which is discussed later.

It is important to note that, since the mode attribute differences $\boldsymbol{x_i}$ and socio-demographics $\boldsymbol{q_i}$ are included parametrically, the gain in flexibility from modeling unobservables with a GP in the utility specification does not compromise the model's structural interpretability. On the contrary, the kernel parameters $\boldsymbol{\theta}$ provide insights into the correlated unobservables.

5.1. SAE and SAL models HGP equivalence

As GPs generalize multivariate normal distributions to infinite-dimensional spaces, the HGP model presented in this section generalizes the SAE model discussed previously. To see how this works, let us define $\boldsymbol{\Omega}$ as $(\boldsymbol{I} - \rho\boldsymbol{W})^{-1}(\boldsymbol{I} - \rho\boldsymbol{W})^{-T}$ so that the random utility in the SAE model follows a multivariate normal distribution, given by:

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{X\beta} + \boldsymbol{Q\gamma}, \boldsymbol{\Omega}). \tag{15}$$

This n-dimensional random vector can be generated from a GP that is defined as:

$$u_i \sim GP(\boldsymbol{x_i}^\top \boldsymbol{\beta} + \boldsymbol{q_i}^\top \boldsymbol{\gamma}, k(\boldsymbol{s_i}, \boldsymbol{s_j}|\boldsymbol{\theta})). \tag{16}$$

Here, $k(\boldsymbol{s_i}, \boldsymbol{s_j}|\boldsymbol{\theta})$ is some kernel function that reproduces $\boldsymbol{\Omega}$ for a fixed collection of n observations. It is straightforward to see that this GP is equivalent to the one proposed in the previous section, where $u_i = \boldsymbol{x_i}^\top \boldsymbol{\beta} + \boldsymbol{q_i}^\top \boldsymbol{\gamma} + GP(\boldsymbol{0}, k(\boldsymbol{s_i}, \boldsymbol{s_j}|\boldsymbol{\theta}))$.

The HGP equivalent for the SAL model is less evident. Recall that the SAL model assumes that the correlation between individuals comes from the utility interaction rather than the exclusive correlation in the error term. The random utility in the SAL model is given by:

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\Omega^{1/2} X\beta} + \boldsymbol{\Omega^{1/2} Q\gamma}, \boldsymbol{\Omega}), \tag{17}$$

where $\boldsymbol{\Omega^{1/2}}$ is $(\boldsymbol{I} - \rho\boldsymbol{W})^{-1}$. Therefore, a generating GP for the SAL model could be defined as:

$$u_i \sim GP([\boldsymbol{X\beta}]^\top \boldsymbol{G_i^{1/2}} + [\boldsymbol{Q\gamma}]^\top \boldsymbol{G_i^{1/2}}, k(\boldsymbol{s_i}, \boldsymbol{s_j})), \tag{18}$$

where $\boldsymbol{G_i^{1/2}}$ is the i−th row of the matrix square root for the gram matrix $\boldsymbol{G} = \boldsymbol{\Omega}$ reproduced by $k(\boldsymbol{s_i}, \boldsymbol{s_j}|\theta)$. In this case, the mean for the GP is a function of the attributes and individual characteristics associated with the whole collection of individuals. In other words, the expected value for the individuals' utilities depends on their own information and the one from other observations.

The existence proof and mathematical form for the kernels $k(\boldsymbol{s_i}, \boldsymbol{s_j}|\boldsymbol{\theta})$ that reproduce $\boldsymbol{\Omega}$ in the SAE and SAL models depend on how $\boldsymbol{W}$ is defined. These proofs are beyond the scope of this paper.

5.2. Nearest Neighbor Gaussian Process

As previously discussed, the likelihood for the model presented in Eq. (14) involves computing an n-dimensional integral that lacks a closed-form solution. We employ MCMC with the Nearest Neighbor Gaussian Process (NNGP) formulation to estimate our

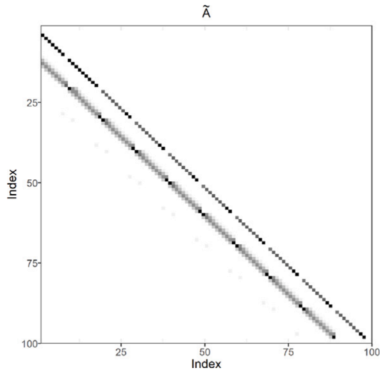

Fig. 1. Graphical representation for matrix $\boldsymbol{A}$ in a hypothetical NNGP.

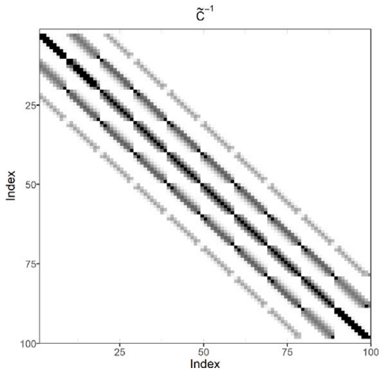

Fig. 2. Graphical representation for $\tilde{\boldsymbol{\Omega}}^{-1}$ in a hypothetical NNGP.

models. In each iteration of the MCMC algorithm, the original HGP model (as described earlier) requires computing the Cholesky decomposition of an $n \times n$ variance–covariance matrix. However, this computation carries a high computational cost of $\mathcal{O}(n^3)$, which can become prohibitive when handling datasets containing thousands of observations. For most empirical applications, this computational cost would render model estimation impractical.

To address this computational cost challenge, one could use Vecchia's likelihood approximation (Vecchia, 1988). This sparse approximation reduces the computational cost to $\mathcal{O}(nm^3)$, where m is usually much smaller than n. In this paper, we adopt the NNGP formulation (Datta et al., 2016), which extends the original Vecchia's sparse approximation to a well-defined sparsity-inducing spatial process.

To understand how the NNGP formulation reduces the computational burden, let us first consider the posterior distribution for the HGP model presented earlier:

$$p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta} | \boldsymbol{X}, \boldsymbol{Q}) \sim p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}) \times \mathcal{N}(\boldsymbol{u} | \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\gamma}, \boldsymbol{\Omega}) \times \mathrm{Ber}(\boldsymbol{y} | \Lambda(\boldsymbol{u})). \tag{19}$$

Here, $p(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta})$ is the prior distribution for the utility and kernel parameters, whereas $\boldsymbol{\Omega}$ is the variance–covariance matrix generated by $k(\boldsymbol{s}_i, \boldsymbol{s}_j | \theta)$. Evaluating $\mathcal{N}(\boldsymbol{u} | \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\gamma}, \boldsymbol{\Omega})$ in this posterior requires the inverse of $\boldsymbol{\Omega}$. We can define the random variables $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$ to obtain this inverse:

$$\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}), \quad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}),$$

where the elements in $\boldsymbol{D}$ are the conditional variances for $\boldsymbol{\omega}$, denoted as $d_{jj} = \mathrm{Var}(\omega_j | \omega_i : i < j)$. With these definitions, we can represent the random variable $\boldsymbol{\omega}$ using the system of equations described below (Eq. (20)).

$$\omega_1 = \eta_1, \quad \omega_i = a_{i1}\omega_1 + a_{i2}\omega_2 + \cdots + a_{i,i-1}\omega_{i-1} + \eta_i. \tag{20}$$

In matrix notation, that system is represented in Eq. (21).

$$\boldsymbol{\omega} = \boldsymbol{A}\boldsymbol{\omega} + \boldsymbol{\eta}, \quad \boldsymbol{A} \text{ strictly lower triangular.} \tag{21}$$

With that representation, it is straightforward to see that $\boldsymbol{\Omega}$ is given by $\boldsymbol{\Omega} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{D}(\boldsymbol{I} - \boldsymbol{A})^{-\top}$, so its inverse can be computed using Eq. (22).

$$\boldsymbol{\Omega}^{-1} = (\boldsymbol{I} - \boldsymbol{A})^{\top}\boldsymbol{D}^{-1}(\boldsymbol{I} - \boldsymbol{A}). \tag{22}$$

Therefore, computing the inverse for $\boldsymbol{\Omega}$ is equivalent to solving for $\boldsymbol{A}$ and $\boldsymbol{D}$ in the system of equations described in (20), which takes $\mathcal{O}(n^3)$ flops. To reduce the computational burden of this inverse computation, the elements in the strictly lower triangular matrix $\boldsymbol{A}$ are conditioned to be equal to zero for all a_{ij} in which j does not belong to the nearest neighbors $N(i)$ of i. Formally, this condition is summarized in (23).

$$a_{ij} = 0 \quad \forall (i,j) \in \{(i,j) : [j > i] \vee [j \notin N(i)]\} \tag{23}$$

For a matrix $\boldsymbol{A}$ with the structure described in (23), the inverse for $\boldsymbol{\Omega}$ can be computed in $\mathcal{O}(nm^3)$ flops, where m is the maximum number of neighbors in $N(i)$. An additional gain from this formulation is that the resulting inverse $\tilde{\boldsymbol{\Omega}}^{-1}$ is sparse, which allows for efficient computations of quadratic forms (Finley et al., 2019).

In Fig. 1, we show a graphical representation of matrix $\boldsymbol{A}$ for a hypothetical NNGP with a hundred observations, and in Fig. 2 the sparse inverse of the variance–covariance matrix $\tilde{\boldsymbol{\Omega}}^{-1}$ associated with that process.

For more details on the NNGP formulation and the algorithmic implementations to find $\tilde{\boldsymbol{\Omega}}^{-1}$, please refer to Datta et al. (2016) and Finley et al. (2019), and to Zhang (2018) for a Stan implementation of non-hierarchical NNGPs.

6. HGP model extension to multiple alternatives

When considering multiple unordered alternatives, such as multiple transportation modes, extending our models is straightforward if independence from irrelevant alternatives (IIA) is assumed (which is a reasonable assumption in some choice situations, including a fully specified model, and for cases with non-labeled alternatives). For cases where IIA might not be a reasonable assumption, other substitution patterns can be easily modeled by including an error term correlated across alternatives for each individual with any desired structure. In this section, we describe both cases.

6.1. Multinomial HGPs with independence from irrelevant alternatives (IIA)

Assuming independence from irrelevant alternatives would lead to an HGP model with independent Gaussian Processes for each alternative j. The latent utility for individual i and alternative j is given by Eq. (24), presented below:

$$u_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{q}_i^T \boldsymbol{\gamma}_j + Z_j(\boldsymbol{s}_i). \tag{24}$$

where $\boldsymbol{x}_{ij}$ is a vector of alternative attributes observed by individual i for alternative j, $\boldsymbol{\gamma}_j$ is a vector of estimable parameters associated with socio-demographics $\boldsymbol{q}_i$ and alternative j, and Z_j is the Gaussian Process associated with alternative j. This Gaussian process Z_j can be assumed to be identical across alternatives, they could have different scales and decay rates, or they could even have different kernel forms to provide isotropic and anisotropic latent utilities depending on the alternative. To simplify the estimation of our models, we assume that Z_j is identical across alternatives.

6.2. Multinomial HGPs without IIA

Considering flexible substitution patterns while also accounting for individual correlations is not a trivial problem. Following the formulation presented in Riihimäki et al. (2013), for multi-class classification problems, let us consider a simple model for the latent utilities that reproduces the IIA property (equivalent to the one discussed previously):

$$u_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{q}_i^T \boldsymbol{\gamma}_j + f_i^j. \tag{25}$$

With c alternatives, f_i^j is an element from the random vector f with size nc:

$$\boldsymbol{f} = [f_1^1, f_2^1, \dots, f_n^1, f_1^2, f_2^2, \dots, f_n^2, \dots, f_1^c, f_2^c, \dots, f_n^c]^\top \tag{26}$$

that follows the multivariate normal distribution,

$$f \sim \mathcal{N}(0, \boldsymbol{K}), \tag{27}$$

with an $nc \times nc$ block-diagonal variance–covariance matrix generated using kernel $k(s, s')$ for each $n \times n$ block. That is, matrix $\boldsymbol{K}$ has c blocks $\boldsymbol{K_1}, \dots, \boldsymbol{K_c}$ where each block is generated using the kernel $k(s, s')$ associated with a Gaussian process Z. As discussed before, these Gaussian processes do not need to be identically parameterized across alternatives, but we consider that to be the case for the sake of simplicity.

By manipulating the off-diagonal blocks of matrix K, one can model correlation structures between alternatives at both the inter-individual and intra-individual levels. In this study, we assume that correlations between the latent utilities across alternatives are only at the intra-individual level. This means that $\text{cov}(u_{ij}, u_{lk}) = 0$, but both $\text{cov}(u_{ij}, u_{lj})$ and $\text{cov}(u_{lj}, u_{lk})$ could have non-zero values. To adhere to this assumption, the off-diagonal blocks for $\boldsymbol{K}$ need to have a diagonal structure as shown below (Eq. (28)).

$$\mathbf{K} = \begin{pmatrix} \mathbf{K_1} & \mathbf{cov}(f^1_{\cdot}, f^2_{\cdot}) \cdot \mathbf{I}_n & \cdots & \mathbf{cov}(f^1_{\cdot}, f^c_{\cdot}) \cdot \mathbf{I}_n \\ \mathbf{cov}(f^2_{\cdot}, f^1_{\cdot}) \cdot \mathbf{I}_n & \mathbf{K_2} & \cdots & \mathbf{cov}(f^2_{\cdot}, f^c_{\cdot}) \cdot \mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{cov}(f^c_{\cdot}, f^1_{\cdot}) \cdot \mathbf{I}_n & \mathbf{cov}(f^c_{\cdot}, f^2_{\cdot}) \cdot \mathbf{I}_n & \cdots & \mathbf{K_c} \end{pmatrix} \tag{28}$$

Since scale and location for the latent utilities are irrelevant for individual decisions (and modeled choice probabilities), not all parameters in K can be identified. To ensure identification in this type of model, one can follow a similar approach to the one proposed by Train in Train (2009) for probit models. For this variance–covariance structure, only $[(c-1)c/2] - 1$ parameters are identifiable (taking $K_1, K_2, \dots, K_c$ as single parameters instead of blocks generated from a kernel function). In the multinomial case study presented in this paper, we consider three alternatives with $K_1 = K_2 = K_3$, so there is a single identifiable covariance between alternatives (i.e., there are $[(c-1)c/2] - 1 = 2$ identifiable parameters in K, where one is associated with the identical blocks K_1, K_2, K_3 and the other is a covariance between a pair of alternatives).

Table 1
Summary of the variables considered in the binary mode choice study.

Variable	Description	Mean
Trip cost difference	Transit cost minus private car cost (USD).	−3.36
Trip time difference	Transit travel time minus private car travel time (Minutes).	35.89
Vehicle availability	Indicator variable for car availability in the household. It takes the value of 1 if no cars are available in the household.	0.34
High income	Indicator variable for high income level (> 100 k USD per year).	0.33
Manhattan	Indicator variable for destinations in Manhattan.	0.47
Gender	Indicator variable for the male gender.	0.48
Mode choice (y)	Whether transit was selected as the travel mode. A choice indicator that takes the value of 1 if transit was selected.	0.61

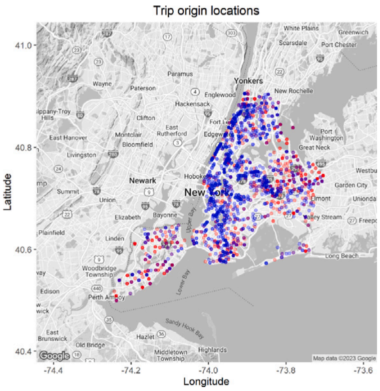

Fig. 3. Trip origins and mode choice in New York City.

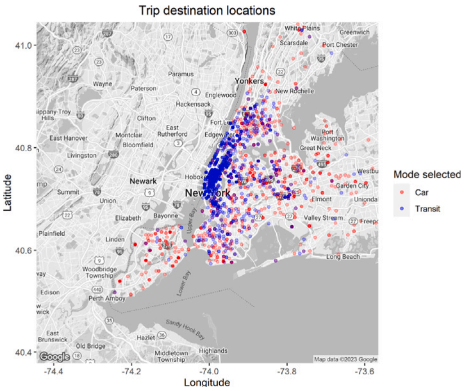

Fig. 4. Trip destinations and mode choice in New York City.

7. Binary mode choice in New York City

We use the 2010/2011 Regional Household Travel Survey by the New York Metropolitan Transportation Council (New York Metropolitan Transportation Council, North Jersey Transportation Planning Authority, 2014) to obtain mode choice data in New York City. The survey contains travel data from 18,965 households from areas in New York City, Long Island, the Hudson Valley, New Jersey, and Connecticut. The data is stored in a relational database with information on travel and mobility patterns, including socio-demographics, trip origins and destinations, and mode choices.

Following the reasoning presented in Goetzke (2008), we limit our analysis to the New York City areas. Additionally, we selected the home-based work (HBW) trips completed in a private car or transit longer than 1.5 miles, and we discarded observations with missing data. After the data selection and cleaning process, 2444 trips are left for our analysis.

We use the US census block data (Anon, 2023) to obtain the longitude and latitude for every trip's origin and destination. We also use the Google API to retrieve estimates for each trip's cost and travel time for various modes of transportation. This results in a novel binary mode choice dataset for NYC, containing trip origin and destination coordinates, trip costs, travel times, and socio-demographic variables. For our models, we consider the following variables: (i) trip cost difference (transit vs. car), (ii) travel time difference (transit vs. car), (iii) an indicator for access to a private car in the household, (iv) an indicator for the destination being in Manhattan, (v) an indicator for high-income level, and (vi) an indicator for declared gender. The description of these variables is presented in Table 1. Additionally, as done in Goetzke (2008), we assume that both alternatives are available for every individual.

Figs. 3 and 4 show the origin and destination locations (i.e. households and work locations), respectively. The red points represent the trips for which the individual selected a private car, and the blue points represent the ones for which the individual chose public transit. As shown in these plots, it is clear that there might be unobserved spatial effects that influence mode choice in New York City. For instance, for trips ending in Manhattan, individuals seem more likely to choose public transit over private cars.

In Figs. 5 and 6, we show trip examples for which public transit was selected, and in Figs. 7 and 8, we show trip examples for which private car was selected. The paths in blue correspond to the public transit routes, and the paths in red are for car. We show these figures to illustrate that mode choice is influenced by the alternative route characteristics, particularly the cost and time difference between modes.

7.1. Binary model estimation and diagnostics

We estimate three HGP models using the NNGP formulation, following the general structure presented in 5 discussed earlier. Additionally, we present the results for a standard logit model and a probit model with correlated error terms for comparison. The

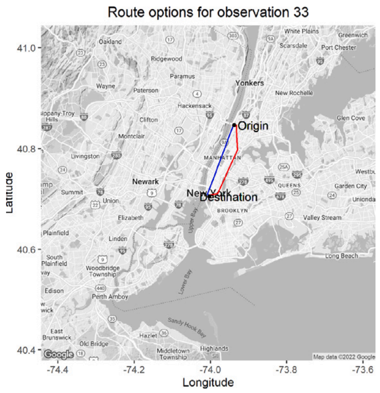

Fig. 5. Trip 33 alternative routes for public transit and private car. For this trip, the cost of using a private car is \$4.75 higher than the cost of using public transportation.

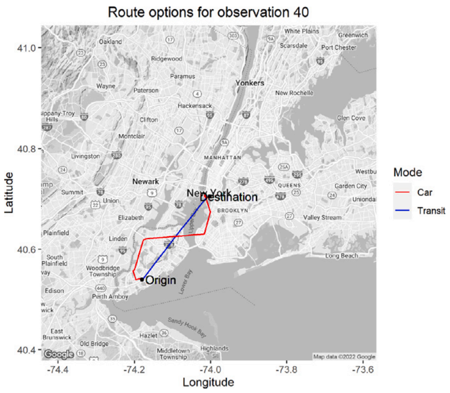

Fig. 6. Trip 40 alternative routes for public transit and private car. For this trip, the cost of using a private car is \$11.77 higher than the cost of using public transportation.

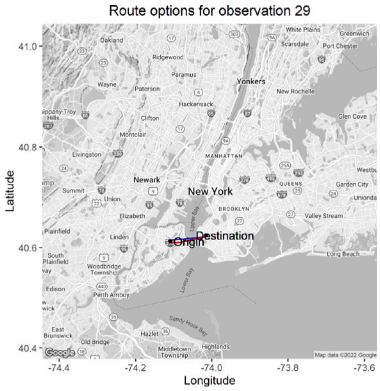

Fig. 7. Trip 29 alternative routes for public transit and private car. For this trip, the cost of using a private car is \$1.06 higher than the cost of using public transportation.

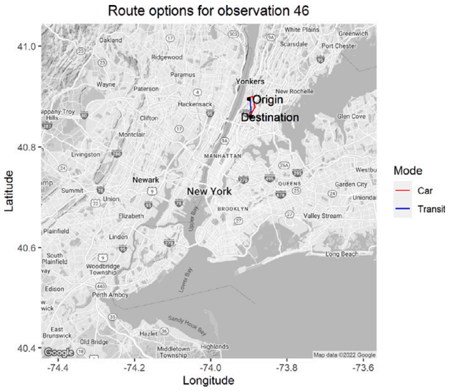

Fig. 8. Trip 46 alternative routes for public transit and private car. For this trip, the cost of using a private car is \$0.10 higher than the cost of using public transportation.

first two HGP models have latent utilities (Eq. (29)) with GPs that depend either on the trip origin ($s = s^o$) or the trip destination coordinates ($s = s^d$). While the third HGP model has a latent utility (Eq. (29)) with a GP associated with the trip direction and distance between the origin and destination coordinates, that is, $s = s^d - s^o$.

$$u_i = \beta_0 + \beta_c \text{Cost}_i + \beta_t \text{Time}_i + \gamma_{\text{Vehicle}} \text{Vehicle}_i + \gamma_{\text{Income}} \text{Income}_i + \gamma_{\text{Manhattan}} \text{Manhattan}_i + \gamma_{\text{Male}} \text{Male}_i + Z(\boldsymbol{s}_i). \tag{29}$$

In Eq. (29), $Z(\boldsymbol{s}_i)$ is an NNGP with a GP parent process that has the kernel defined by:

$$k(\boldsymbol{s}_i, \boldsymbol{s}_j | \sigma^2, \phi, \tau^2) = \sigma^2 \exp\left(-\phi \|\boldsymbol{s}_i - \boldsymbol{s}_j\|\right) + \tau^2, \tag{30}$$

where the parameter τ^2 is set to one to determine the general scale of the utility.

For the logit model, the error term from the latent utilities does not originate from a Gaussian process as in Eq. (29). Instead, it is assumed to be identically and independently distributed as a standard logistic random variable for all individuals (as mentioned in Section 2). For the probit model with correlated error terms, it is assumed that the latent utilities follow the form presented in Eq. (5) (as in the SAE model). The probit SAE model is estimated using the implementation in R as presented in Wilhelm and de Matos (2013) with four chains, a burn-in period of 5000, and 30,000 samples per chain.

For the HGPs and the logit models, we employed the No-U-Turn sampler in Stan for MCMC exploration of the desired posterior. We sampled four independent chains in parallel with 3000 samples each, and a burn-in period of 1000. We use normal priors for the model parameters ($\beta_0, \beta_c, \beta_t, \gamma_{\text{vehicle}}, \gamma_{\text{Income}}, \gamma_{\text{Manhattan}}, \gamma_{\text{Male}}$), a truncated normal prior for the kernel scale parameter σ, and a gamma prior for the kernel decay rate parameter ϕ.

7.1.1. Diagnostics

Figs. 9, 10, and 11 show the trace plots for the model and kernel parameters for the HGP models. As shown there, for the three models, the independent chains seem to converge to the same posterior mode, and there is no large auto-correlation after a

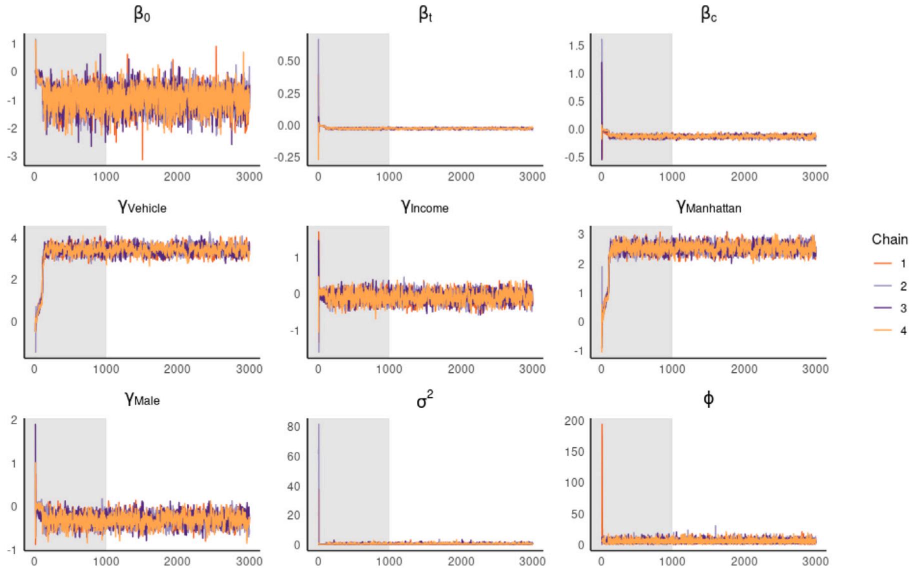

Fig. 9. Trace plots for HGP process model with origin locations s^o.

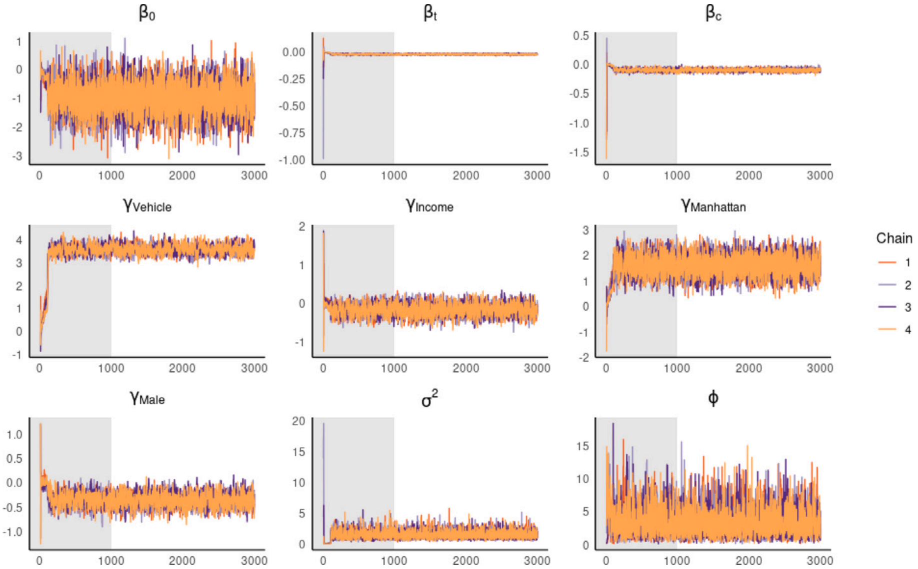

Fig. 10. Trace plots for HGP process model with destination locations s^d.

relatively low burn-in period. In Fig. 12, we show the chains associated with the standard logit model, and in Fig. 13 we show the ones associated with the probit SAE model.

Table 2 shows the effective sample sizes and $\hat{R}$ for the three HGP models described before, the logit model, and the probit SAE model. The effective sample size estimates the equivalent number of uncorrelated samples from the posterior for each variable chain, so large numbers are desirable. Whereas $\hat{R}$ is a diagnostic that compares multiple chains to assess the convergence of the Markov chains, usually values below 1.010 are a good indication of convergence (Vehtari et al., 2021).

Based on the MCMC diagnostics, we can conclude that the chains for all three HGP models have mixed well and converged. The largest $\hat{R}$ value of 1.006 belongs to γ_{Vehicle} in the model with origin coordinates, which should not be alarming since it is below the 1.010 threshold commonly used in the literature. The smallest sample size of 544 is associated with the same model parameter. As can be seen, the sample sizes for the logit model are larger, and the $\hat{R}$ values for that model are closer to 1.000. This is because the logit model has a lower complexity and a single mode posterior, making convergence with MCMC easier to attain. The probit SAE

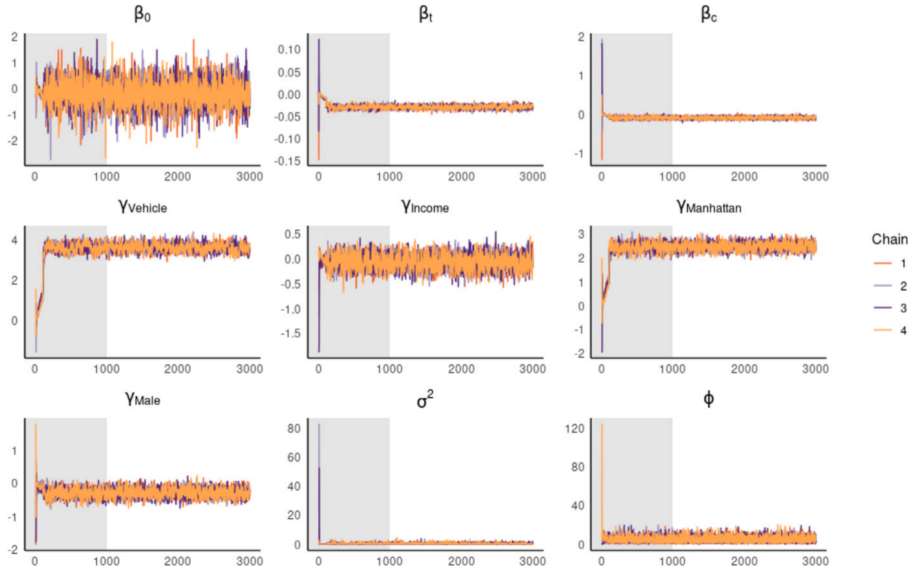

Fig. 11. Trace plots for HGP process model with direction vectors $s^d - s^o$.

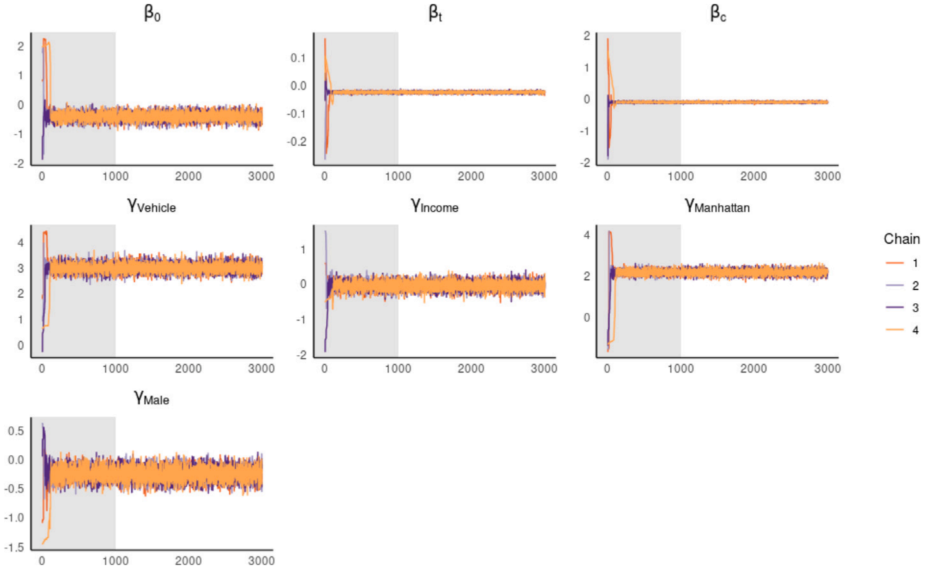

Fig. 12. Trace plots for the logit model.

model exhibits $\hat{R}$ values lower than 1.005 but effective sample sizes that are lower than any other model for some parameters. This was expected even for the large number of samples because our HGP models and the logit model are estimated using the No-U-Turn sampler from Stan, whereas the probit SAE implementation from Wilhelm and de Matos (2013) uses a basic Gibbs sampler.

7.2. Binary model results and discussion

This section presents and discusses the MCMC results based on the chain statistics for the three HGP models, the logit model, and the probit SAE model. We estimate the posterior means and medians and construct 95% high-density credible intervals for the model parameters ($\boldsymbol{\beta}, \boldsymbol{\gamma}$), the kernel parameters (σ^2, ϕ), the SAE importance parameter ρ, as well as the value of travel time savings (VOTT).

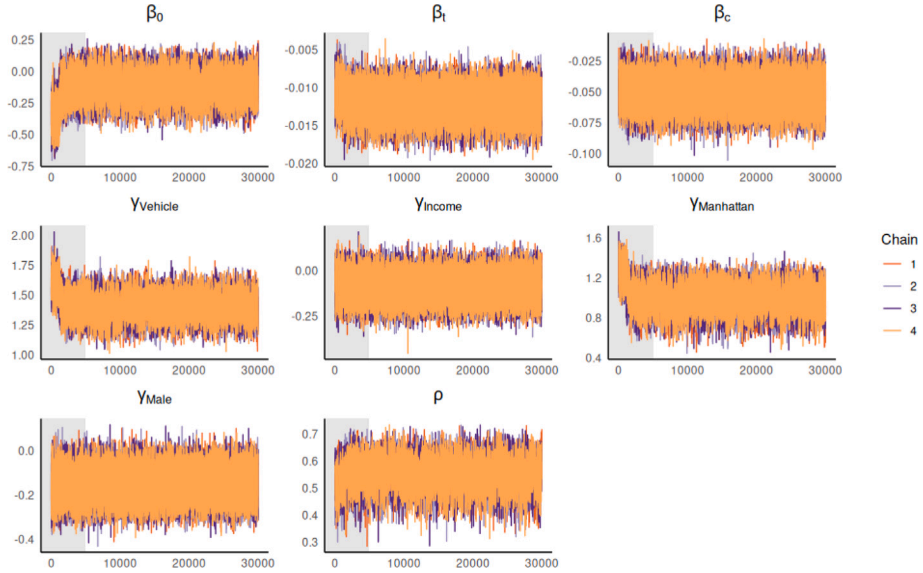

Fig. 13. Trace plots for the probit SAE model.

Table 2
Effective sample sizes (ESS) and $\hat{R}$ for the three HGP models, the logit model, and the probit SAE model.

	HGP models						Logit		Probit SAE	
	Trip Directions		Destination Locations		Origin Locations					
	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$
β_0	3589	1.000	4432	1.001	3229	1.001	5485	1.000	5983	1.000
β_t	1258	1.002	1172	1.002	1180	1.002	5981	1.000	368	1.002
β_c	2139	1.002	1457	1.001	1398	1.002	6522	1.000	608	1.001
γ_{Vehicle}	774	1.002	597	1.005	544	1.006	7233	1.000	138	1.004
γ_{Income}	1693	1.002	1434	1.001	1348	1.000	7935	1.000	7580	1.000
$\gamma_{\text{Manhattan}}$	1627	1.000	2401	1.002	911	1.002	7076	1.000	142	1.000
γ_{Male}	1488	1.002	1198	1.002	1433	1.002	7577	1.000	2886	1.000
σ^2	1568	1.001	3119	1.000	1706	1.002	–	–	–	–
ϕ	4517	1.001	2036	1.003	3057	1.001	–	–	–	–
ρ	–	–	–	–	–	–	–	–	5988	1.000

Table 3
Posterior mean and medians for β_c and β_t.

Parameter	Statistic	HGP models			Logit	Probit SAE
		Trip Directions	Destination Locations	Origin Locations		
β_c	Mean	−0.095	−0.092	−0.124	−0.092	−0.051
	Median	−0.095	−0.092	−0.124	−0.091	−0.051
β_t	Mean	−0.029	−0.021	−0.025	−0.023	−0.012
	Median	−0.029	−0.021	−0.025	−0.023	−0.012

7.2.1. Mode attribute parameters

The parameters $\boldsymbol{\beta}$ represent the marginal utilities for mode attributes, specifically travel time and trip cost, which are expected to have a negative effect on latent utilities. Table 3 summarizes the posterior means and medians for β_c and β_t. As presented there, the posterior means and median values for all models agree with behavioral intuition.

Interestingly, but not surprisingly, the posterior mean and median values associated with β_c for the logit model differ the most from the HGP model that assumes that the correlation in individual decisions is associated with trip origin locations — where the individuals live. This large difference could be attributed to the marginal utility of income following the same correlation structure. For mean and median values associated with β_t, the differences are not as dramatic. The marginal utilities for the probit SAE model are not directly comparable to those associated with the other models, as the link function for the probit model differs from that of the logit and HGP models.

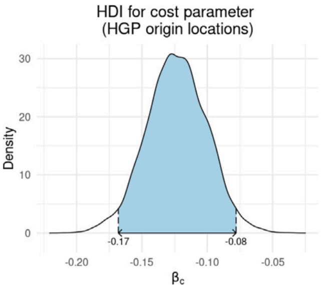

Fig. 14. β_c 95% high-density credible interval (HGP model with origin locations).

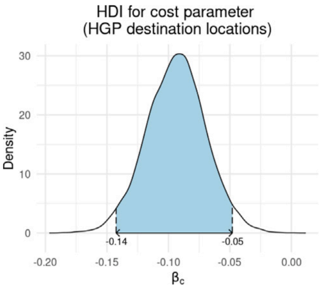

Fig. 15. β_c 95% high-density credible interval (HGP model with destination locations).

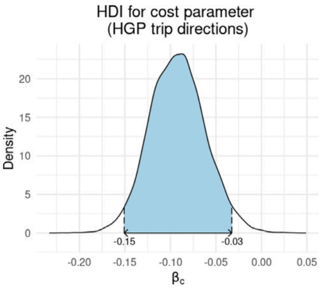

Fig. 16. β_c 95% high-density credible interval (HGP model with trip directions).

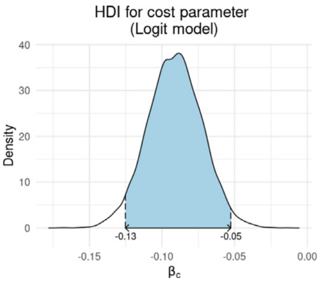

Fig. 17. β_c 95% high-density credible interval (Logit model).

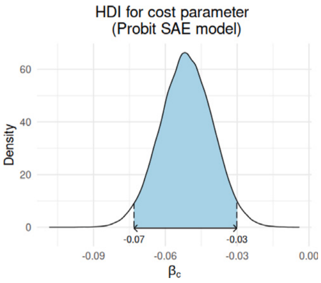

Fig. 18. β_c 95% high-density credible interval (Probit SAE model).

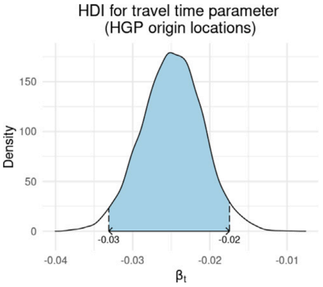

Fig. 19. β_t 95% high-density credible interval (HGP model with origin locations).

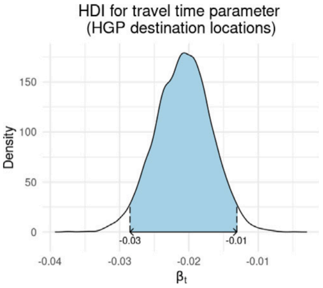

Fig. 20. β_t 95% high-density credible interval (HGP model with destination locations).

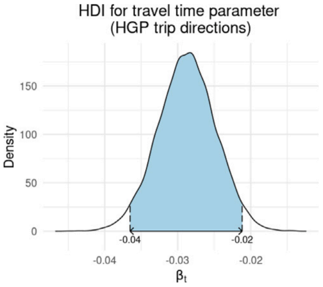

Fig. 21. β_t 95% high-density credible interval (HGP model with trip directions).

Figs. 14 through 18 display the 95% high-density credible intervals for β_c, while Figs. 19 through 23 show the credible intervals for β_t. These figures illustrate that the credible intervals cover similar parameter ranges for all models, but the posteriors associated with the logit model are more concentrated around their modes than the posteriors of the HGP models — which is expected. Additionally, there is no change of sign within any of the intervals, indicating with high posterior probability that travel time and cost have a negative influence on latent utilities for all models (equivalent to parameter significance in frequentist inference).

7.2.2. Socio-demographic parameters

As previously mentioned, we incorporated four socio-demographic variables into our models: (i) a dummy variable identifying individuals with no access to vehicles in their household (Vehicle), (ii) a dummy variable identifying high-income households (Income), (iii) a dummy variable for trips ending at Manhattan (Manhattan), and (iv) a dummy variable for individuals identifying as male (Male). Table 4 summarizes the means and medians for the parameters $\boldsymbol{\gamma}$ associated with this variables for the three HGP models, as well as for the standard logit model and the probit SAE model. We also provide estimates for the odds ratios of taking public transit. For the probit SAE model, the odds ratios are computed using the observed sample of socio-demographics and alternative attributes, setting the relevant dummy variable to zero for the base case and one for the comparison case.

The mean and median values for γ_{Vehicle} are similar across all models with a logistic link function. The empirical posterior mean ranges from 3.00 to 3.60 among these models, indicating that people without access to a vehicle at home are more likely to use

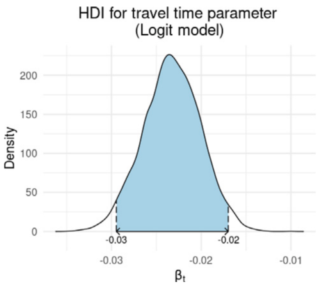

Fig. 22. β_t 95% high-density credible interval (Logit model).

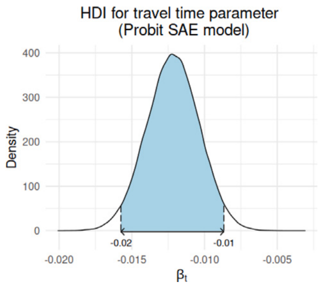

Fig. 23. β_t 95% high-density credible interval (Probit SAE model).

Table 4
Posterior mean and medians for γ.

Parameter	Statistic	HGP models			Logit	Probit SAE
		Trip Directions	Destination Locations	Origin Locations		
γ_{Vehicle}	Mean	3.60	3.59	3.41	3.00	1.40
	Median	3.60	3.59	3.41	3.00	1.40
γ_{Income}	Mean	−0.057	−0.160	−0.090	−0.044	−0.089
	Median	−0.057	−0.160	−0.090	−0.044	−0.089
$\gamma_{\text{Manhattan}}$	Mean	2.46	1.62	2.54	2.18	0.990
	Median	2.46	1.62	2.54	2.18	0.994
γ_{Male}	Mean	−0.280	−0.350	−0.312	−0.242	−0.148
	Median	−0.281	−0.350	−0.312	−0.242	−0.147

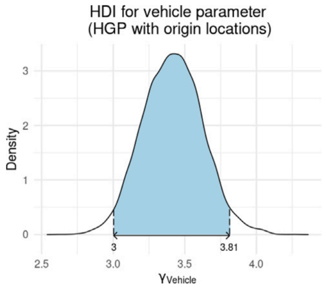

Fig. 24. γ_{Vehicle} 95% high-density credible interval (HGP model with origin locations).

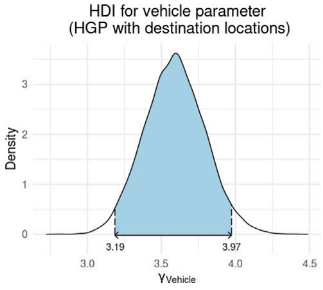

Fig. 25. γ_{Vehicle} 95% high-density credible interval (HGP model with destination locations).

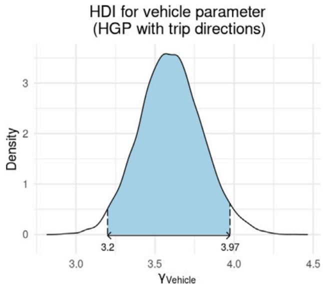

Fig. 26. γ_{Vehicle} 95% high-density credible interval (HGP model with trip directions).

public transit. According to these estimates, the odds of taking public transit are approximately 20 (logit model) to 37 (HGP models) times higher for people without access to a car in their household. For the probit SAE model, the mean odds ratio – computed using all observations in the dataset – is approximately 14. Figs. 24 through 28 display the high-density credible intervals for γ_{Vehicle}. These figures show that the credible regions are similar for all models with a logistic link, and there is no change in the parameter sign within the high-density credible intervals for any model — equivalent to significance in frequentist inference.

Regarding γ_{Income}, the results show that the mean and median values are similar in the HGP models with trip directions and origin locations. However, in the HGP model with destination locations, they have an absolute value twice as large than the ones from the other HGP models. For the logit model, the γ_{Income} posterior means and medians are roughly half as large as those from the HGP with trip directions and origin locations. Regardless of the model, the parameter posterior statistics are negative, indicating that high-income individuals are more likely to use a private car instead of public transit. Based on the empirical mean from the posterior, the odds for taking public transit are approximately 6% lower for high-income individuals in models with trip directions, 9% lower for the origin locations model, and 15% lower in the model with destination locations. For the logit model, the odds are 4% lower, and for the probit SAE model, on average they are 15% lower.

Nevertheless, Figs. 29 through 33 display the high-density credible intervals for γ_{Income}, which show a change of sign across all models. This suggests insufficient evidence of a negative marginal effect on the latent utility of taking public transit, equivalent to a non-significant parameter in frequentist inference.

The mean and median values for $\gamma_{\text{Manhattan}}$ are similar in the HGP models with trip directions and origin locations. However, in the HGP model with destination locations, they are approximately 35% lower than the other HGP models. The $\gamma_{\text{Manhattan}}$ posterior

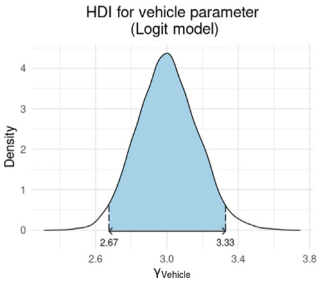

Fig. 27. γ_{Vehicle} 95% high-density credible interval (Logit model).

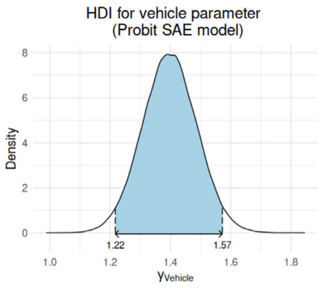

Fig. 28. γ_{Vehicle} 95% high-density credible interval (Probit SAE model).

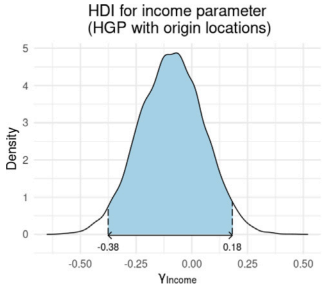

Fig. 29. γ_{Income} 95% high-density credible interval (HGP model with origin locations).

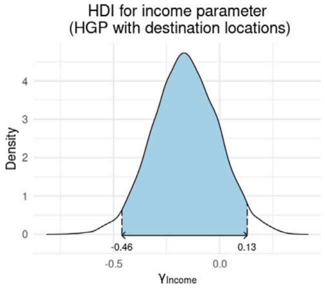

Fig. 30. γ_{Income} 95% high-density credible interval (HGP model with destination locations).

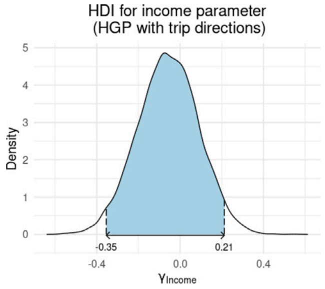

Fig. 31. γ_{Income} 95% high-density credible interval (HGP model with trip directions).

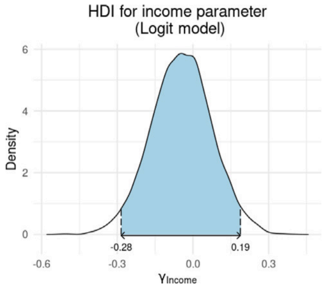

Fig. 32. γ_{Income} 95% high-density credible interval (Logit model).

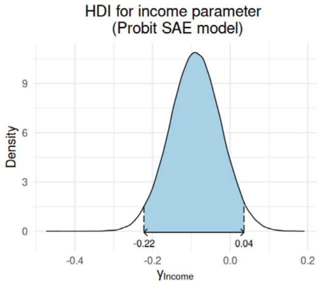

Fig. 33. γ_{Income} 95% high-density credible interval (Probit SAE model).

mean and median values for the logit model are close to the HGP models with trip directions and origin locations. This is a sensible result, considering that the Manhattan attribute is spatially correlated for the destination locations.

Based on empirical statistics, the odds for taking public transit for trips that end in Manhattan are around 5 times higher in the HGP model with destination locations and approximately 12 times higher in the other two HGP models. With the logit model, the odds are almost 9 times higher if the trip ends in Manhattan. For the probit SAE model, the mean odds ratio is approximately equal to 6 –which is close to the one associated with the HGP model with destination locations. Figs. 34 through 38 display the high-density credible intervals for $\gamma_{\text{Manhattan}}$, which do not show a change of sign for any model. This indicates that the positive marginal effect of $\gamma_{\text{Manhattan}}$ on the latent utility holds with 95% posterior probability. The credible region for $\gamma_{\text{Manhattan}}$, as shown in Figs. 34 through 36, is much wider for the HGP model with destination locations than for the other two HGP models. Still, the upper value of the credible region for the model with the destination locations is lower than the other models.

The empirical statistics for γ_{Male} are negative, ranging from −0.28 in the HGP model with trip directions to −0.35 in the model with destination locations. These results suggest that individuals who identify as male are less likely to take public transit. Specifically, in the HGP models with trip directions and origin locations, the odds of using public transit are around 25% lower for males compared to other genders, while in the HGP model with destination locations, the odds are approximately 30% lower. For the logit model, the odds for males are over 21% lower than for other genders. In the probit SAE model, the odds of taking public transit are, on average, 24% lower for males. Figs. 39 through 43 illustrate the high-density credible intervals for γ_{Male}, which do

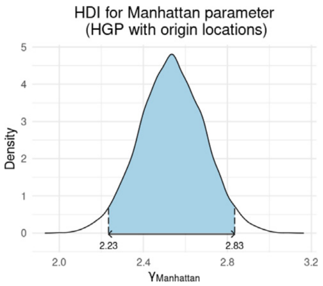

Fig. 34. $\gamma_{\text{Manhattan}}$ 95% high-density credible interval (HGP model with origin locations).

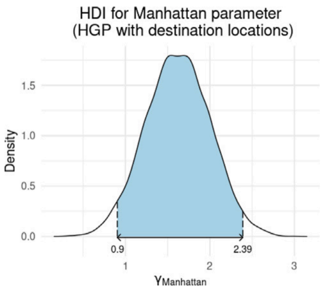

Fig. 35. $\gamma_{\text{Manhattan}}$ 95% high-density credible interval (HGP model with destination locations).

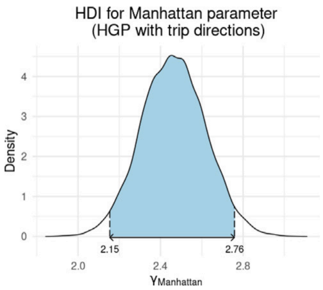

Fig. 36. $\gamma_{\text{Manhattan}}$ 95% high-density credible interval (HGP model with trip directions).

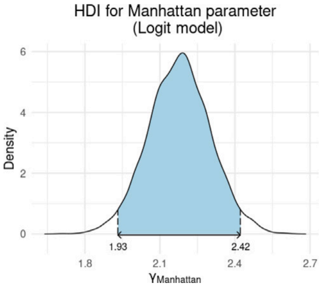

Fig. 37. $\gamma_{\text{Manhattan}}$ 95% high-density credible interval (Logit model).

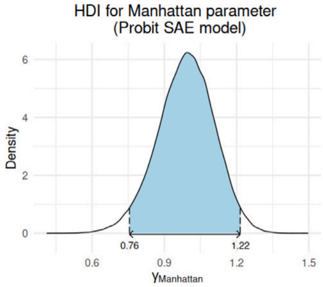

Fig. 38. $\gamma_{\text{Manhattan}}$ 95% high-density credible interval (Probit SAE model).

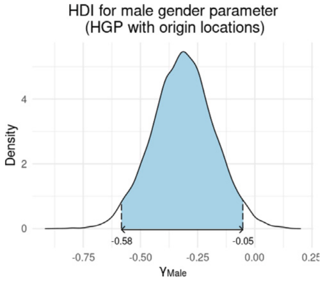

Fig. 39. γ_{Male} 95% high-density credible interval (HGP model with origin locations).

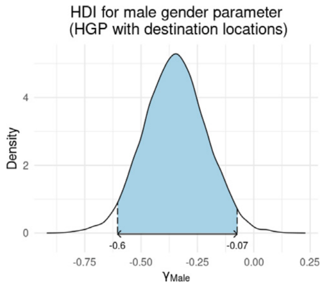

Fig. 40. γ_{Male} 95% high-density credible interval (HGP model with destination locations).

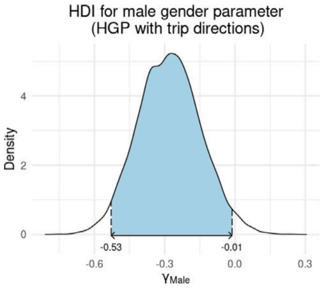

Fig. 41. γ_{Male} 95% high-density credible interval (HGP model with trip directions).

not show a change in the parameter sign. Therefore, there is a high posterior probability of a negative marginal effect on the latent utility associated with this socio-demographic characteristic.

7.2.3. Kernel parameters σ^2 and ϕ, and SAE importance parameter ρ

Up to this point, we have presented results for parameters that could be estimated using standard discrete choice models, albeit with some caveats mentioned before (e.g., biased parameter estimates). Now, we will delve into the kernel parameters σ^2 and ϕ (Eq. (30)), which offer insights into the underlying spatial process of the latent utilities, and ρ, which determines the importance of the spatially correlated part of the error term in the probit SAE model. As previously mentioned, σ^2 is associated with the scale of the spatially correlated unobservables, whereas ϕ represents the decay rate for the correlation of those unobserved effects. To put it in more concrete terms for the models presented here, σ^2 relates to how large are the spatially correlated unobserved effects in comparison to those that do not exhibit a correlation structure. Whereas ϕ is related to how distant two individuals could be and still exhibit correlated decisions. Table 5 displays the empirical means and medians from the posterior for these parameters.

The mean and median values for σ^2 differ between the model with destination locations and the other two models. Specifically, the mean value for the model with destination locations (1.48) is more than three times higher than that for the model with origin locations (0.510) and for the model with trip directions (0.501). To interpret these parameters, it is worth noting that the kernel nugget parameter τ^2 (Eq. (30)) represents the scale of the unobservables without spatial correlation and is set to one for all models

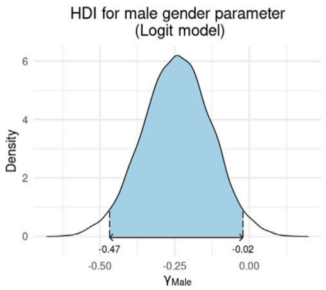

Fig. 42. γ_{Male} 95% high-density credible interval (Logit model).

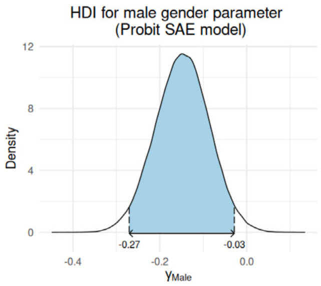

Fig. 43. γ_{Male} 95% high-density credible interval (Probit SAE model).

Table 5
Posterior mean and medians for kernel parameters σ^2 and ϕ, and for spatially-auto-correlated error importance ρ.

Parameter	Statistic	HGP models			Logit	Probit SAE
		Trip Directions	Destination Locations	Origin Locations		
σ^2	Mean	0.501	1.48	0.510	–	–
	Median	0.392	1.37	0.439	–	–
ϕ	Mean	5.11	3.18	5.39	–	–
	Median	4.57	3.80	4.87	–	–
ρ	Mean	–	–	–	–	0.556
	Median	–	–	–	–	0.558

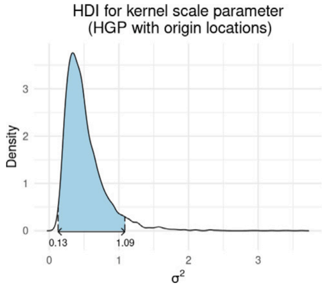

Fig. 44. σ^2 95% high-density credible interval (HGP model with origin locations).

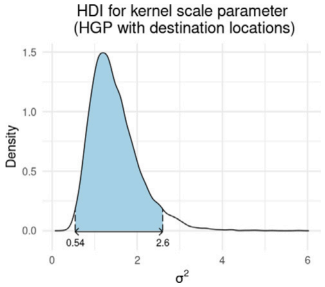

Fig. 45. σ^2 95% high-density credible interval (HGP model with destination locations).

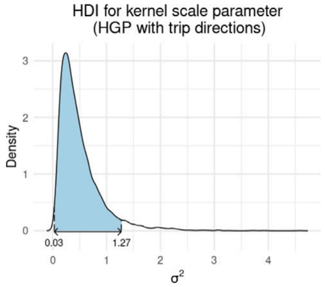

Fig. 46. σ^2 95% high-density credible interval (HGP model with trip directions).

to establish the general scale of the utility. Consequently, in the model with destination locations, spatially correlated unobservables appear more relevant than unobserved effects that are not spatially correlated. Conversely, the opposite is true for the other two models. Figs. 44 through 46 depict the high-density credible intervals for σ^2. As illustrated, the credible interval for the model with destination locations is wider than those of the other two models, but the lower end of the interval is higher than the mean for the other models.

As for the decay rate parameter ϕ, the findings reveal that spatial correlation in the model with destination locations decays slower than in the other two models. The mean value for that model is 3.18, compared to 5.39 for the model with origin locations and 5.11 for the model with trip directions. Figs. 47 through 49 display the high-density credible intervals for ϕ. The width of the credible region is similar for all models, but the lower and upper ends for the model with destination locations are smaller than those of the other two HGP models.

Finally, the spatially correlated importance parameter ρ – from the probit SAE model – has a mean value of 0.556, and as shown in Fig. 50, the 95% high-density credible interval suggests that the spatially correlated unobserved effects are significant.

7.2.4. Value of travel time savings

To estimate the posterior means, medians, and high-density credible intervals for the value of travel time savings, we use the chains associated with β_c and β_t to derive a post-processed sample from the posterior. This post-processing is achieved using Eq. (31), where VOTT_k, β_{t_k}, and β_{c_k} are samples from the posterior.

$$\text{VOTT}_k = \frac{\beta_{t_k}}{\beta_{c_k}}. \tag{31}$$

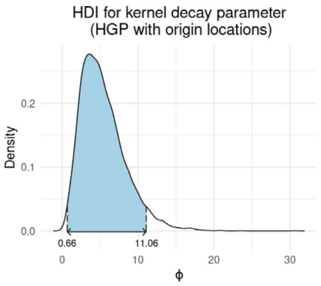

Fig. 47. ϕ 95% high-density credible interval (HGP model with origin locations).

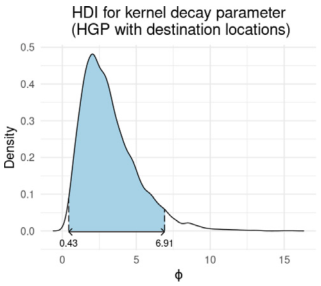

Fig. 48. ϕ 95% high-density credible interval (HGP model with destination locations).

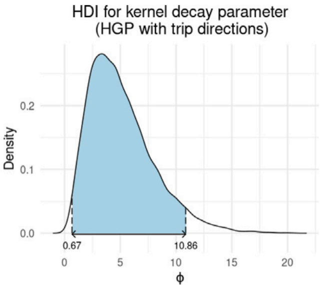

Fig. 49. ϕ 95% high-density credible interval (HGP model with trip directions).

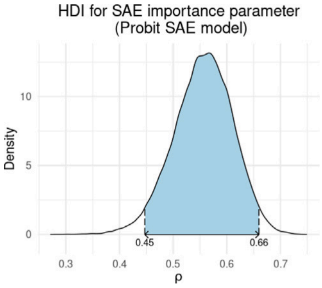

Fig. 50. ρ 95% high-density credible interval (Probit SAE model).

Table 6
VOTT posterior sample statistics for the HGP models.

Model	Mean [$/h]	Median [$/h]
HGP with Origin locations	12.29	12.09
HGP with Destination locations	13.80	13.34
HGP with Trip directions	21.15	18.54
Logit	15.71	15.27
Probit SAE	14.68	14.26

Table 6 summarizes the VOTT estimated means and medians from the posterior samples for the three HGP models, the logit and the probit SAE models. That table shows that the VOTT estimates are relatively similar, with a mean value ranging between $12 to $21 for an hour reduction in travel time. However, the VOTT posterior sample mean for the model with trip directions is over 45% higher than the ones associated with the other two HGP models. This difference is due to the heavy right tail of the VOTT posterior for the trip directions model, which can be identified in the high-density credible interval plots presented in Figs. 51 through 55.

Figs. 51 through 55 display the high-density credible intervals for the value of travel time savings. These plots reveal that the VOTT from all models lie within similar regions. The lower bound of the credible regions is very close for all models, ranging from $8.48/hour to $11.01/hour, whereas the upper bound is where the models differ the most, ranging from $16.06/hour to $38.31/hour. The tightest credible region is associated with the HGP model with origin locations, whereas the widest corresponds to the model with trip directions. This result indicates that the VOTT posterior distribution for the model with trip directions has the flattest surface and that model would require the most additional evidence from data to reduce the uncertainty associated with that estimate. It is also worth noting that the mean and median values for the probit SAE model are the ones closest to the statistics for the logit model that does not consider spatial effects.

7.2.5. *Out-of-sample prediction performance and model selection*

For comparing the different models, we used out-of-sample prediction performance measured by the models' accuracy and the leave-one-out information criterion (LOO-IC) (Vehtari et al., 2017). For the out-of-sample accuracy, we held out 20% of the data. For LOO-IC, which is defined as minus two times the expected log-predictive density, a lower value indicates better expected out-of-sample prediction performance. In Table 7, we provide estimates for train and test accuracy, as well as LOO-IC, for all models estimated in the binary case study, except for the probit SAE model, as the implementation from Wilhelm and de Matos (2013) does not provide the necessary per-observation chains. As shown in the table, the HGP model with destination locations outperforms all other models across all metrics.

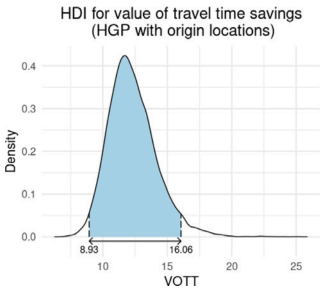

Fig. 51. VOTT 95% high-density credible interval (HGP with origin locations).

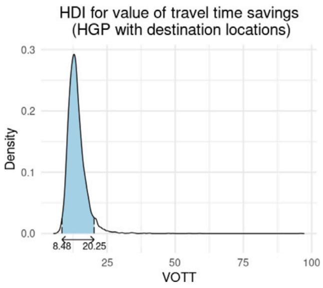

Fig. 52. VOTT 95% high-density credible interval (HGP with destination locations).

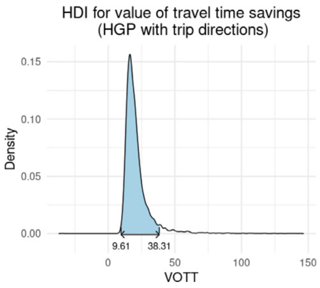

Fig. 53. VOTT 95% high-density credible interval (HGP with trip directions).

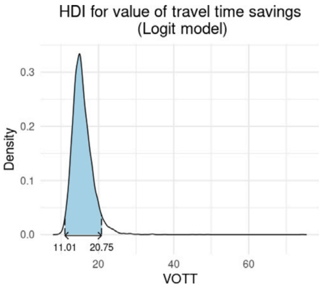

Fig. 54. VOTT 95% high-density credible interval (Logit model).

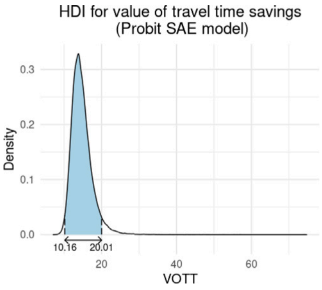

Fig. 55. VOTT 95% high-density credible interval (Probit SAE model).

Table 7
Model performance metrics.

Model	Test set accuracy	Train set accuracy	LOO-IC
HGP with origin locations	86.49%	87.06%	1952.5
HGP with destination locations	**86.70%**	**87.10%**	**1886.5**
HGP with trip directions	84.68%	85.63%	1972.5
Logit	84.07%	84.60%	1978.4
Probit SAE	83.87%	84.44%	–

8. Multinomial mode choice in NYC: Public transit, private car and non-motorized

For the multinomial case, we also limit our analysis to the New York City areas. We select home-based work (HBW) trips completed in transit, private car, or non-motorized modes (i.e. walking and bicycling) without limiting the trip length. Trip cost for the non-motorized alternative is set to zero and the travel time is computed using the Google API for the cases where it is not revealed. We discarded observations with missing data, so after the data selection and cleaning process, 3277 trips are left for our analysis. The description for the variables considered in the multinomial problem is presented in Table 8.

8.1. Multinomial model estimation and diagnostics

For the multinomial case study, we considered two HGP models: one with IIA and another without IIA. We followed the formulation presented in Section 6, where we assume that the blocks in the block-diagonal matrix K are identical. For the kernel function, we used trip destination locations. For the model without IIA, we considered an unknown but identifiable correlation between the public transit and the non-motorized alternatives. For easier model interpretation, we assume that the socio-demographics only have an effect on the latent utilities for private car. Additionally, we take the non-motorized alternative as the base alternative; hence, we only show the estimation results for alternative-specific constants associated with private car and public transit. We compare our multinomial HGP models with a standard multinomial logit model.

As for the binary case study, we employed the No-U-Turn sampler in Stan for MCMC exploration of the desired posterior. We sampled four independent chains in parallel with 6000 samples each, and a burn-in period of 1000. We use normal priors for the model parameters ($\beta_0, \beta_c, \beta_t, \gamma_{\text{vehicle}}, \gamma_{\text{Income}}, \gamma_{\text{Manhattan}}, \gamma_{\text{Male}}$), a truncated normal prior for the kernel scale parameter σ, and a gamma prior for the kernel decay rate parameter ϕ. For the correlation matrix associated with the public transit and non-motorized alternatives in the non-IIA case, we use the LKJ distribution (Lewandowski et al., 2009).

Table 8
Summary of the variables considered for the multinomial mode choice study.

Variable	Description	Mean
Trip cost transit	Transit cost (USD).	2.35
Trip cost car	Car cost (USD).	4.64
Trip cost non-motorized	Non-motorized cost (USD).	0.00
Trip time transit	Transit travel time (Minutes).	50.38
Trip time car	Car travel time (Minutes).	18.32
Trip time non-motorized	Non-motorized travel time (Minutes).	93.85
Vehicle availability	Indicator variable for car availability in the household. It takes the value of 1 if no cars are available in the household.	0.36
High income	Indicator variable for high income level (> 100 k USD per year).	0.32
Manhattan	Indicator variable for destinations in Manhattan.	0.46
Gender	Indicator variable for the male gender.	0.47
Car mode share	Proportion of trips completed by car	0.37
Transit mode share	Proportion of trips completed by transit	0.51
Non-motorized mode share	Proportion of trips completed by non-motorized means of transportation	0.12

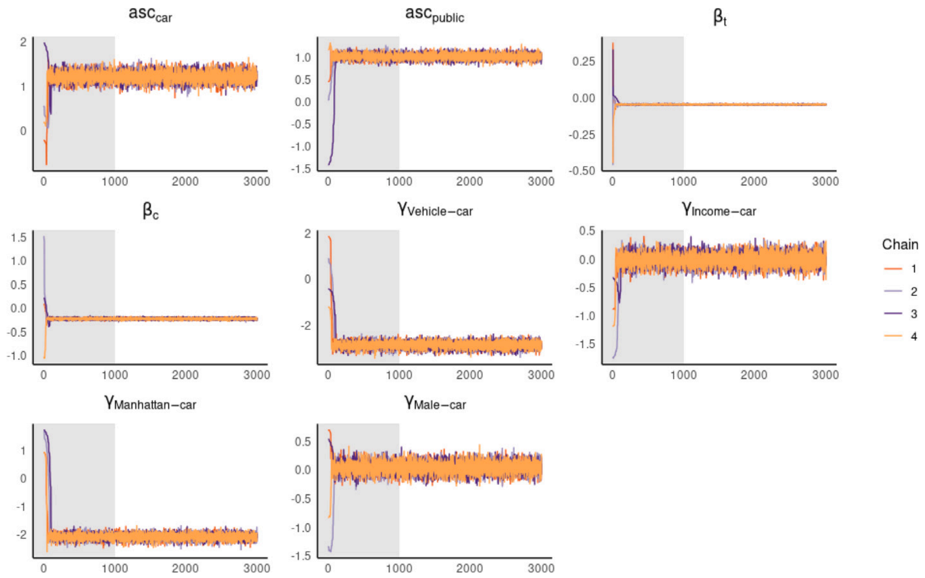

Fig. 56. Trace plots for the multinomial logit model.

8.1.1. Diagnostics

Fig. 56 shows the trace plots for the multinomial logit model. Figs. 57 and 58 show the trace plots for the model and kernel parameters for the HGP models. As shown, for the multinomial logit model and the HGP model with IIA, the independent chains seem to converge to the same posterior mode, and there is no large auto-correlation after a relatively short burn-in period. In contrast, for the HGP model with correlated public transit and non-motorized alternatives, the Markov chains seem to converge to the same posterior mode but for some parameters (e.g. $\gamma_{\text{Vehicle-car}}$) the autocorrelations seem to be higher than for the model with IIA.

In Table 9, we show the $\hat{R}$ and effective sample size for all the multinomial models. As presented, the HGP model without IIA has an $\hat{R}$ lower than 1.05 for most parameters, except for the correlation between alternatives, $\gamma_{\text{Vehicle-car}}$ and $\gamma_{\text{Manhattan-car}}$. The HGP model and the multinomial model have $\hat{R}$ values that are lower than 1.01. These diagnostics, along with the trace plots presented earlier, indicate that the model without IIA has a more complex posterior shape than the model with IIA and the multinomial model.

8.2. Multinomial model results and discussion

This section presents and discusses the MCMC results based on the chain statistics for the two multinomial HGP models and the multinomial logit model. We estimate the posterior means and medians and construct 95% high-density credible intervals for the model parameters ($\boldsymbol{\beta}, \boldsymbol{\gamma}$), the kernel parameters (σ^2, ϕ), the alternative correlation $\rho_{\text{transit-nonmotorized}}$, the alternative-specific constants ASC_{Car} and $\text{ASC}_{\text{Transit}}$, as well as the value of travel time savings (VOTT).

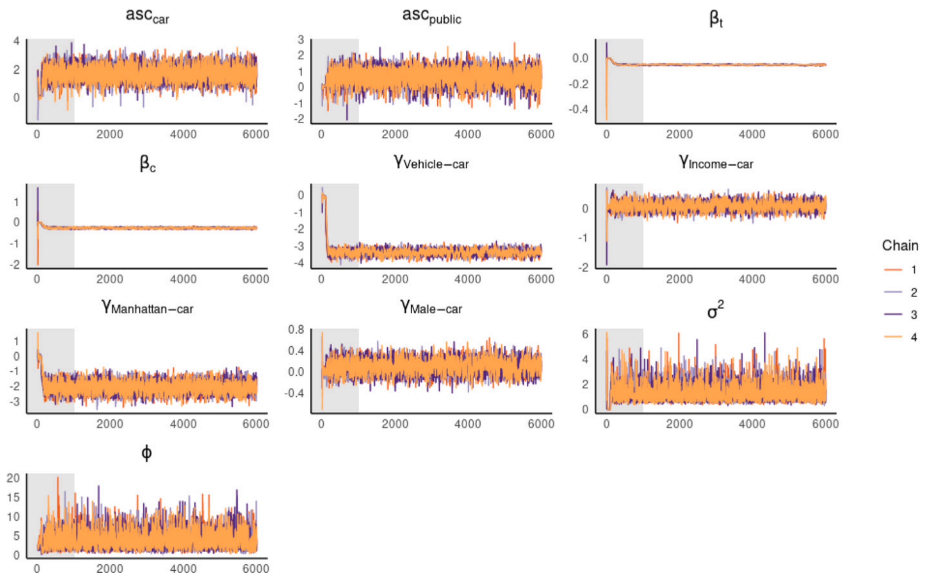

Fig. 57. Trace plots for the HGP multinomial model with IIA.

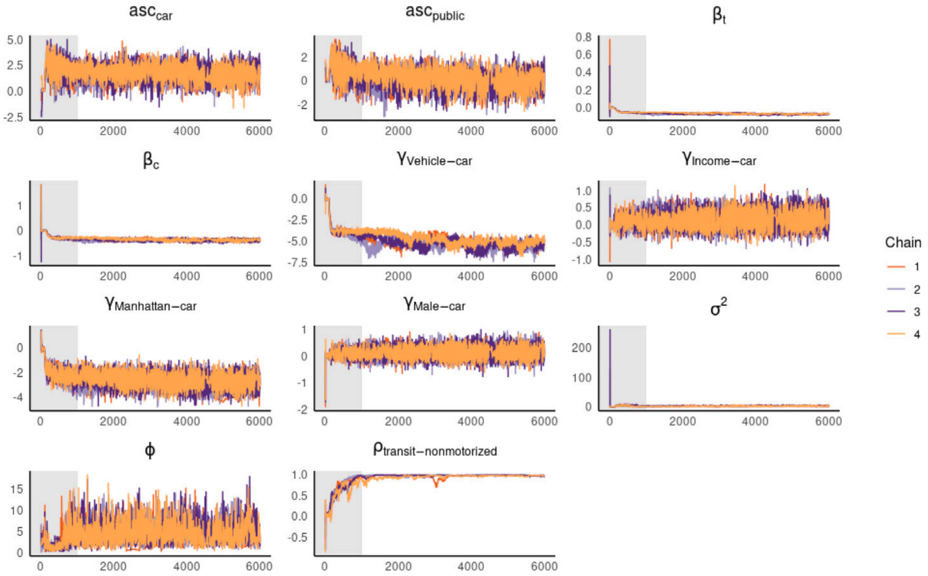

Fig. 58. Trace plots for the HGP multinomial model without IIA.

8.2.1. *Mode attribute parameters and alternative specific constants*

In Table 10, we show the summary statistics for β, ASC_{Car}, and $\text{ASC}_{\text{Transit}}$. As presented, the statistics for β_c and β_t follow behavioral intuition. The values for the alternative-specific constants indicate that, all other variables held constant, individuals are more likely to choose either driving or public transit than non-motorized modes of transportation.

In Figs. 59 to 61, we show the 95% high-density credible intervals for ASC_{car}. For all models, these credible intervals cover only the positive region of the parameter space, which can be interpreted similarly to parameter significance in frequentist statistics.

In contrast to the credible intervals associated with ASC_{car}, those for $\text{ASC}_{\text{transit}}$ (Figs. 62 to 64) exhibit an interesting difference between the HGP models and the logit model. For the HGP models, the high-density credible intervals cover regions that include zero, whereas the one associated with the multinomial logit model covers a strictly positive region. This means that for the multinomial model, with high posterior probability, individuals are more likely to choose public transit over non-motorized modes of transportation. In contrast, the HGP models do not provide evidence to support that claim.

Table 9
Effective sample sizes (ESS) and $\hat{R}$ for the two HGP multinomial models and the multinomial logit model.

Parameter	HGP models				Multinomial	
	IIA		non-IIA		logit	
	ESS	$\hat{R}$	ESS	$\hat{R}$	ESS	$\hat{R}$
$\mathrm{ASC}_{\mathrm{car}}$	4758	1.000	915	1.020	5592	1.000
$\mathrm{ASC}_{\mathrm{transit}}$	5495	1.000	478	1.011	7537	1.000
β_t	164	1.008	141	1.041	8795	1.000
β_c	1293	1.003	288	1.030	8440	1.001
$\gamma_{\mathrm{Income\text{-}car}}$	2125	1.003	5919	1.045	8278	1.000
$\gamma_{\mathrm{Male\text{-}car}}$	2168	1.001	4659	1.012	8732	1.000
$\gamma_{\mathrm{Vehicle\text{-}car}}$	1241	1.007	59	1.147	8653	1.000
$\gamma_{\mathrm{Manhattan\text{-}car}}$	3206	1.003	1671	1.140	9032	1.001
σ^2	3129	1.002	577	1.023	–	–
ϕ	2373	1.005	1195	1.048	–	–
$\rho_{transit-nonmotorized}$	–	–	51	1.238	–	–

Table 10
Multinomial models posterior mean and medians for β, ASC_{Car} and $\mathrm{ASC}_{Transit}$.

Parameter	Statistic	HGP models		Multinomial
		IIA	non-IIA	logit
$\mathrm{ASC}_{\mathrm{car}}$	Mean	1.71	1.71	1.22
	Median	1.70	1.72	1.22
$\mathrm{ASC}_{\mathrm{transit}}$	Mean	0.689	0.212	1.01
	Median	0.718	0.217	1.01
β_c	Mean	−0.269	−0.358	−0.219
	Median	−0.269	−0.356	−0.220
β_t	Mean	−0.052	−0.071	−0.045
	Median	−0.052	−0.073	−0.045

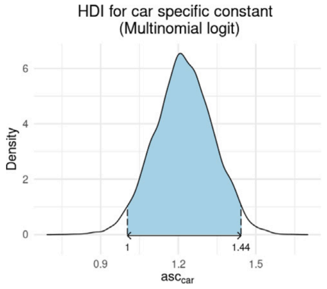

Fig. 59. $\mathrm{ASC}_{\mathrm{car}}$ 95% high-density credible interval (Multinomial logit).

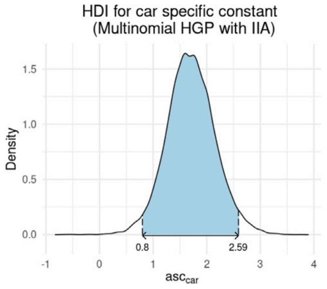

Fig. 60. $\mathrm{ASC}_{\mathrm{car}}$ 95% high-density credible interval (HGP with IIA).

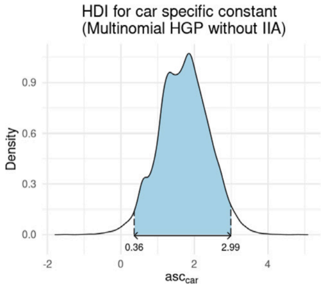

Fig. 61. $\mathrm{ASC}_{\mathrm{car}}$ 95% high-density credible interval (HGP without IIA).

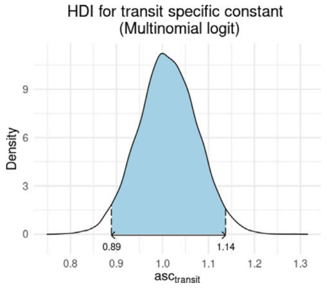

Fig. 62. $\mathrm{ASC}_{\mathrm{transit}}$ 95% high-density credible interval (Multinomial logit).

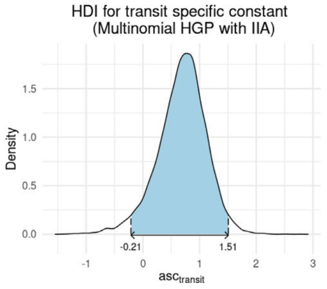

Fig. 63. $\mathrm{ASC}_{\mathrm{transit}}$ 95% high-density credible interval (HGP with IIA).

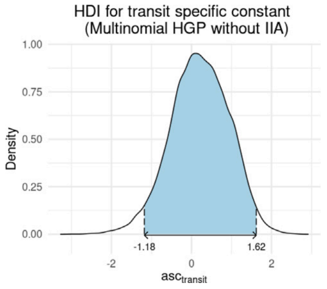

Fig. 64. $\mathrm{ASC}_{\mathrm{transit}}$ 95% high-density credible interval (HGP without IIA).

Finally, Figs. 65 to 67 show the high-density 95% credible regions for β_c, and Figs. 68 to 70 show the high-density 95% credible regions for β_t. As shown, the credible regions are strictly negative for both parameters in all models. Therefore, with high probability given the observed data, both travel time and trip cost have a negative influence on the latent utilities for all considered models.

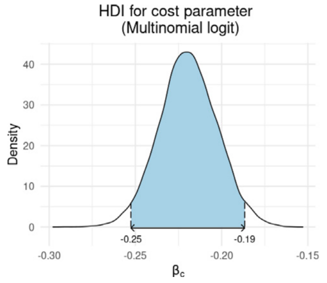

Fig. 65. β_c 95% high-density credible interval (Multinomial logit).

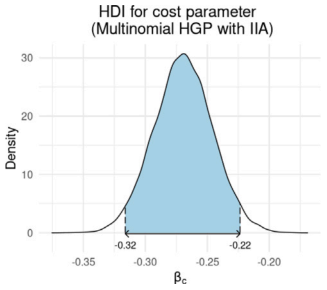

Fig. 66. β_c 95% high-density credible interval (HGP with IIA).

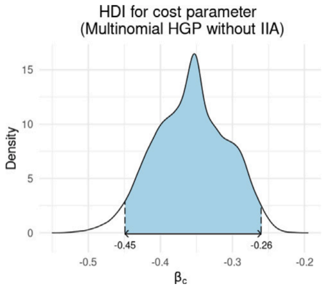

Fig. 67. β_c 95% high-density credible interval (HGP without IIA).

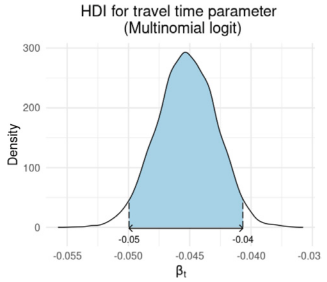

Fig. 68. β_t 95% high-density credible interval (Multinomial logit).

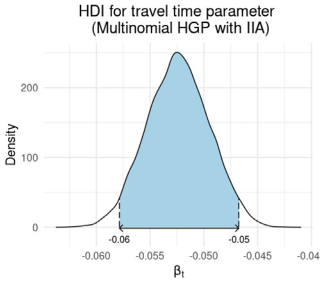

Fig. 69. β_t 95% high-density credible interval (HGP with IIA).

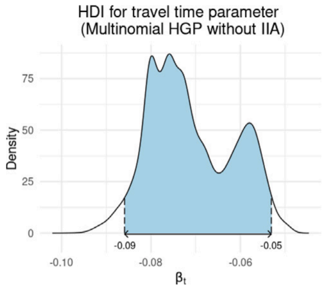

Fig. 70. β_t 95% high-density credible interval (HGP without IIA).

Table 11
Multinomial models posterior mean and medians for γ.

Parameter	Statistic	HGP models		Multinomial
		IIA	non-IIA	logit
$\gamma_{\text{Vehicle-car}}$	Mean	−3.37	−4.94	−2.88
	Median	−3.37	−5.11	−2.88
$\gamma_{\text{Income-car}}$	Mean	0.077	0.189	−0.009
	Median	0.080	0.182	−0.010
$\gamma_{\text{Manhattan-car}}$	Mean	−2.06	−2.64	−2.07
	Median	−2.06	−2.57	−2.07
$\gamma_{\text{Male-car}}$	Mean	0.088	0.145	0.040
	Median	0.087	0.138	0.040

8.2.2. Socio-demographic parameters

In Table 11, we summarize the results for the socio-demographic-related parameters γ shared by all considered multinomial models. Additionally, in this section, we provide estimates for the odds ratios and show 95% high-density credible intervals. As discussed previously, the socio-demographics only entered the latent utilities associated with the private car alternative, so the odds ratios are computed for that specific alternative.

As presented in Table 11, the mean for $\gamma_{\text{Vehicle-car}}$ ranges between −2.88 (multinomial logit) and −4.94 (non-IIA HGP). For the HGP model with IIA, the odds of taking a private car are 96% lower for people without access to a car in their household. For the HGP model without IIA, this value is 99.28%. For the multinomial logit model, the odds are 94% lower. Figs. 71 to 73 show the high-density credible intervals for $\gamma_{\text{Vehicle-car}}$. There is no change in sign for any credible interval presented there.

Regarding $\gamma_{\text{Income-car}}$, the mean values presented in Table 11 are positive for the HGP models and negative for the multinomial logit model. For the multinomial logit model, this result suggests that high-income individuals are less likely to drive private cars in NYC. Conversely, for the HGP models, high-income individuals are more likely to drive private cars. Specifically, for the multinomial logit model, the odds of driving a private car are 1% lower for high-income individuals. For the HGP model with IIA, the odds for high-income individuals are 1.08 times those for low-income individuals, and for the HGP model without IIA, the odds ratio is 1.20. However, as shown in the 95% high-density credible intervals (Figs. 74 to 76), the credible regions cover both positive and negative values for the three models, so this effect is not considered significant.

For all models, the mean and median values for $\gamma_{\text{Manhattan-car}}$ suggest a negative effect on the latent utilities for driving a private car if the destination is Manhattan. The high-density credible intervals presented in Figs. 77 to 79 support this negative effect with high posterior probability. Specifically, for the HGP model with IIA, the odds of driving a private car to work are 87% lower if the

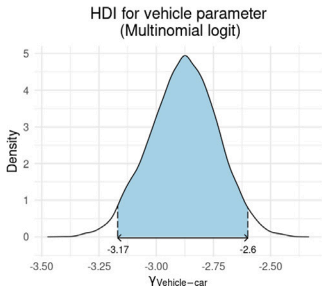

Fig. 71. $\gamma_{\text{Vehicle-car}}$ 95% high-density credible interval (Multinomial logit).

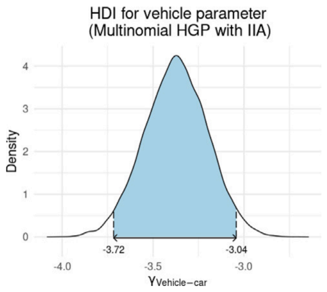

Fig. 72. $\gamma_{\text{Vehicle-car}}$ 95% high-density credible interval (HGP with IIA).

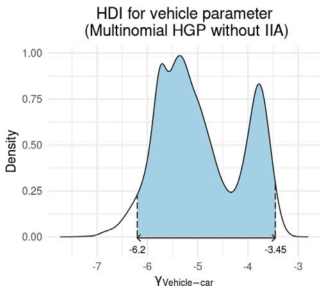

Fig. 73. $\gamma_{\text{Vehicle-car}}$ 95% high-density credible interval (HGP without IIA).

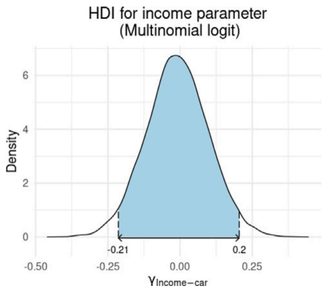

Fig. 74. $\gamma_{\text{Income-car}}$ 95% high-density credible interval (Multinomial logit).

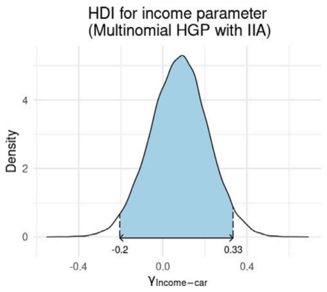

Fig. 75. $\gamma_{\text{Income-car}}$ 95% high-density credible interval (HGP with IIA).

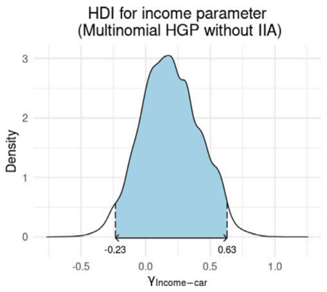

Fig. 76. $\gamma_{\text{Income-car}}$ 95% high-density credible interval (HGP without IIA).

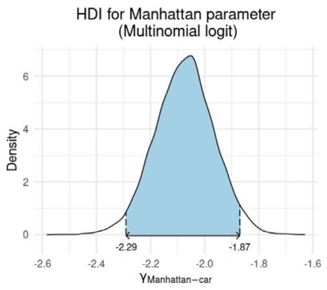

Fig. 77. $\gamma_{\text{Manhattan-car}}$ 95% high-density credible interval (Multinomial logit).

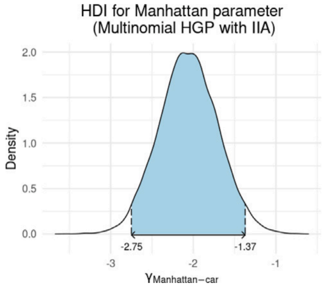

Fig. 78. $\gamma_{\text{Manhattan-car}}$ 95% high-density credible interval (HGP with IIA).

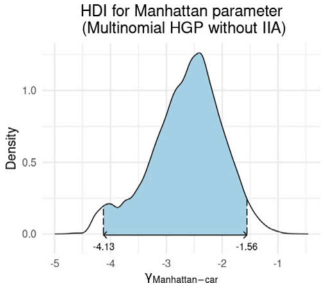

Fig. 79. $\gamma_{\text{Manhattan-car}}$ 95% high-density credible interval (HGP without IIA).

destination is Manhattan. For the HGP model without IIA, the odds are 92% lower. And, for the multinomial logit model, the odds are 87% lower.

Finally, according to the mean and median posterior values for $\gamma_{\text{Male-car}}$, there is a positive marginal effect on the latent utilities for driving a private car for people who declared their sex as male. However, as shown in Figs. 80 to 82, the high-density credible intervals cover parameter regions with both positive and negative values. Using the mean values presented in Table 11, the odds ratio associated with $\gamma_{\text{Male-car}}$ for the HGP model with IIA equals 1.09. For the non-IIA HGP model, the odds of taking a private car for males are 1.15 times those of others. For the multinomial logit model, that value is 1.04.

8.2.3. Kernel parameters σ^2 and ϕ, and alternative correlation $\rho_{\textit{transit-nonmotorized}}$

Here, we present the summary statistics and high-density credible intervals for the kernel parameter and the alternative correlation parameter from the HGP models. As previously mentioned, σ^2 is associated with the scale of the spatially correlated unobservables, whereas ϕ represents the decay rate for the correlation of those unobserved effects. In Table 12, we present the mean and median values for ϕ and σ^2, and the alternative correlation $\rho_{\text{transit-nonmotorized}}$.

The mean and median values for ϕ show that the spatially correlated effects for the model without IIA decay 16.62% faster than those for the model with IIA. In Figs. 83 and 84, we show the high-density credible intervals for that kernel parameter. As presented in those figures, for the model with IIA, the posterior density for ϕ is more concentrated around the mean than for the model without IIA.

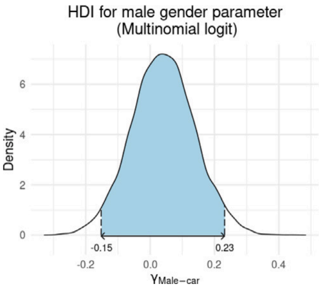

Fig. 80. $\gamma_{\text{Male-car}}$ 95% high-density credible interval (Multinomial logit).

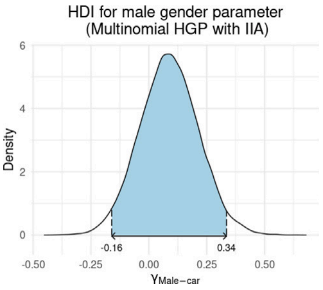

Fig. 81. $\gamma_{\text{Male-car}}$ 95% high-density credible interval (HGP with IIA).

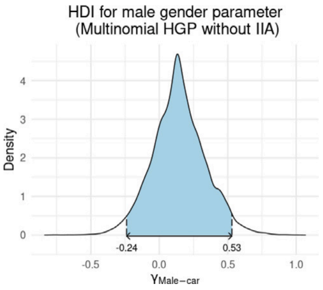

Fig. 82. $\gamma_{\text{Male-car}}$ 95% high-density credible interval (HGP without IIA).

Table 12
Multinomial models posterior mean and medians for kernel parameters ϕ and σ^2, and alternative correlation ρ.

Parameter	Statistic	HGP models	
		IIA	non-IIA
ϕ	Mean	3.67	4.28
	Median	3.28	3.88
σ^2	Mean	1.28	2.28
	Median	1.17	2.18
$\rho_{\text{transit-nonmotorized}}$	Mean	–	0.832
	Median	–	0.973

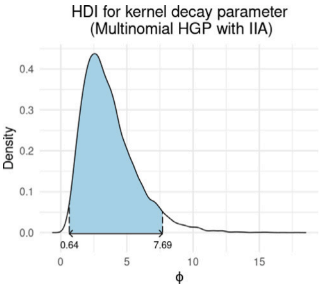

Fig. 83. ϕ 95% high-density credible interval (HGP with IIA).

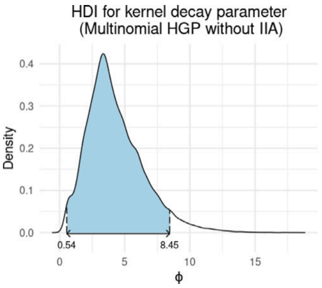

Fig. 84. ϕ 95% high-density credible interval (HGP without IIA).

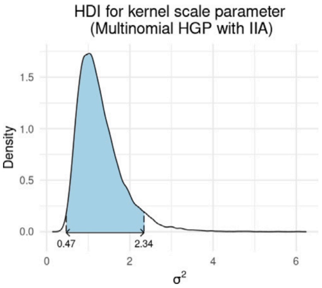

Fig. 85. σ^2 95% high-density credible interval (HGP with IIA).

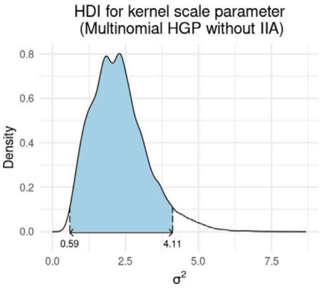

Fig. 86. σ^2 95% high-density credible interval (HGP without IIA).

Regarding the scale of the spatially correlated unobservables, σ^2, the mean and median values suggest a greater importance of these effects compared to the unobservables without spatial correlation. For the model without IIA, the scale of the correlated unobservables is almost twice that of the one for the IIA model. In Figs. 85 and 86, we show the high-density credible interval for σ^2 for both HGP models.

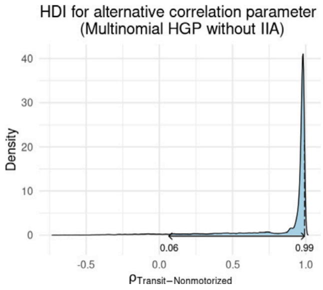

Fig. 87. $\rho_{\text{transit-nonmotorized}}$ 95% high-density credible interval (HGP without IIA).

Table 13
VOTT posterior sample statistics for the multinomial HGP and logit models.

Model	Mean [$/h]	Median [$/h]
HGP with IIA	11.72	11.67
HGP without IIA	11.96	11.86
Multinomial Logit	12.42	12.38

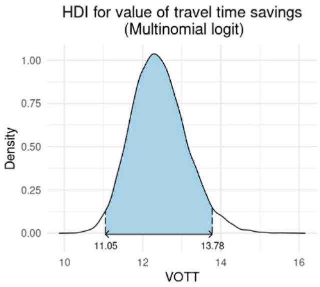

Fig. 88. VOTT 95% high-density credible interval (Multinomial Logit model).

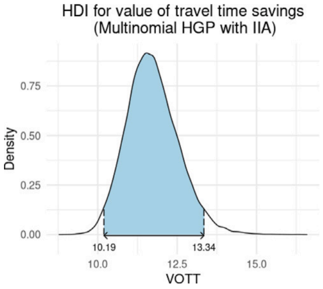

Fig. 89. VOTT 95% high-density credible interval (HGP with IIA).

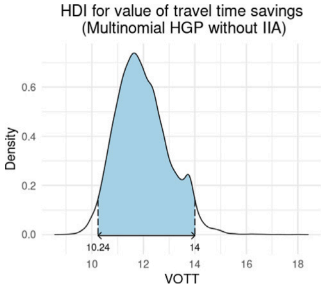

Fig. 90. VOTT 95% high-density credible interval (HGP without IIA).

For the correlation between alternatives, $\rho_{\text{transit-nonmotorized}}$, from the non-IIA HGP model, we found positive mean and median posterior values. This indicates that the unobserved effects for public transit and non-motorized modes are positively correlated. As shown in Fig. 87, this positive correlation holds with high posterior probability.

8.2.4. Value of travel time savings

Similar to the binary case study, we also computed MCMC samples for the value of travel time savings for all the estimated multinomial models. The VOTT means and medians are summarized in Table 13. As can be seen from the table, the VOTT values for the HGP models are lower than the value found using the multinomial logit model. The lowest mean VOTT is associated with the HGP with IIA ($11.72 per hour), while the highest value is associated with the multinomial logit model ($12.42 per hour).

The high-density credible intervals for the VOTT are shown in Figs. 88 through 90. These credible intervals cover very similar regions, suggesting with high posterior probability that the VOTT lies within $10 per hour and $14 per hour for all three models.

8.2.5. Out-of-sample prediction performance and model selection

To compare the performance of the multinomial models, we employ out-of-sample accuracy and the leave-one-out cross-validation information criterion, mirroring the approach used in the binary case study. Here, we also hold-out 20% of the data for computing the out-of-sample accuracy. Based on these out-of-sample prediction metrics, the HGP model with IIA exhibits superior performance compared to both the non-IIA HGP model and the multinomial logit model. The difference in prediction accuracy between the HGP models and the multinomial logit model is substantial, reaching almost 5 percentage points for the model without IIA and nearly 10 percentage points for the IIA model (see Table 14).

9. Conclusions and future work

In this work, we have introduced a novel Hierarchical Gaussian Process (HGP) model for discrete choice analysis, effectively addressing the challenge of spatially correlated unobservables in both the binary and multinomial settings. Our HGP model

Table 14
Multinomial models performance metrics.

Model	Test set accuracy	Train set accuracy	LOO-IC
HGP with IIA	**84.8%**	85.29%	**3966.8**
HGP without IIA	80.20%	**87.50%**	3968.5
Multinomial Logit	75.84%	73.35%	4153.8

generalizes the Spatially Autoregressive Error (SAE) model from spatial econometrics, demonstrating how Gaussian Processes can be seamlessly integrated into the latent utility specification while preserving economic interpretability. We showcase the practical value of HGP models for applied discrete choice analysis through a binary and a multinomial mode choice case study in New York City. For these case studies, we have curated a novel mode choice dataset for NYC, encompassing trip origin and destination coordinates, trip costs, travel times, and sociodemographic variables.

To the best of our knowledge, this work presents the first application of a fully Bayesian HGP discrete choice model. We addressed the computational challenges associated with running MCMC for HGPs by leveraging the Nearest Neighbor Gaussian Process formulation. This approach significantly reduces the computational complexity per iteration for computing variance–covariance matrix inverses, from $\mathcal{O}(n^3)$ to $\mathcal{O}(nm^3)$ — with $m \ll n$. In our empirical studies, we generated policy-relevant insights, including point estimates and credible intervals for the value of time as well as for odds ratios. Additionally, by formally interpreting the kernel parameters from the Gaussian process, we gained valuable insights into the underlying spatial process governing the latent utilities.

For the binary case study, we analyzed the kernel parameters, σ^2 and ϕ, in three distinct HGP models, namely: one focused on origin locations, another on destination locations, and the last on trip directions. Our study revealed differences between the destination locations model and the other two models regarding the mean, median values, and high-density credible intervals of the parameters of interest. The empirical results provided evidence that spatially correlated unobservables have greater importance, and their correlation decays slower in the model with destination locations compared to the other two specifications. Furthermore, in the model with destination locations, the mean and median values for the scale of spatially correlated unobservables (σ^2) were found to be higher than the scale of non-correlated unobserved effects (τ^2). These results support that, for mode choice selection in NYC, the spatially correlated unobservables at the destination locations are more relevant and decay slower than the ones at the origin or those associated with trip directions.

Intuitively, the result discussed above means that unobserved spatially correlated effects at the destination of the trip, which would include parking costs, have a higher relevance in mode choices than spatially correlated effects at the origin of the trip (e.g. public transit crowding), or the ones related to trip vectors (e.g. road congestion). Focusing on the HGP models with trip origin and destination locations, the correlation between latent utilities stays high for larger distances when comparing destinations to trip origins. That is, the latent utilities correlation for two individuals going to work 1 km apart (from the HGP model with destination locations) would be higher than the correlation for two agents departing from homes that are 1 km apart (from the HGP model with origin locations).

In the multinomial case study, we employed two HGP models: one with independence from irrelevant alternatives (IIA) and one without. For both models, we assumed that the kernel for each alternative shared the same attributes, ϕ and σ^2, and depended on trip destination locations. In the non-IIA HGP model, we specifically allowed for correlation between the public transit and non-motorized modes. Our analysis revealed a positive correlation between these alternatives, holding with high posterior probability.

In terms of point and credible-interval estimates for the parameters of mode attributes and socio-demographics, all models in both the binary and multinomial case studies consistently showed negative marginal effects on utility for trip cost and travel time. The binary case study revealed that individuals without vehicle access are more likely to choose public transit over a private car (including taxis). High-income individuals exhibited a lower propensity to use public transit, with this effect being more pronounced in the model considering destination locations. Trips ending in Manhattan were more likely to be completed via public transit, and males were less likely to use public transit. The multinomial case study yielded similar insights for most socio-demographic parameters across all models. However, a noteworthy difference emerged for the multinomial logit model, where high-income individuals were found to be more likely to take public transit.

Leveraging both the binary and multinomial datasets, we estimated the value of travel time savings (VOTT) for New York City and obtained credible intervals from our results. Based on the binary dataset, the empirical posterior mean for VOTT ranged from \$12.29 to \$21.15 per hour reduction in travel time. The multinomial dataset yielded VOTT mean values between \$11.72 and \$12.42. Interestingly, the mean VOTTs estimated using the binary HGP models that considered trip origins and destinations aligned more closely with their multinomial counterparts compared to the estimates derived from the binary logit and multinomial logit models. These VOTT values are close to the recommended values used by the US Department of Transportation (US Department of Transportation, 2016), which stipulates a range of 80% to 120% of the hourly wage (minimum wage of \$7.25). The VOTT estimates hold valuable insights for transportation planning and policy-making, such as evaluating the benefits of public investment in transportation infrastructure and congestion pricing strategies.

We empirically demonstrated that the posterior mean and medians associated with the parameters of a standard logit model can differ significantly from those obtained from models that account for correlations between individuals. This difference is particularly pronounced when the variables included in the latent utility model follow the same correlation structure as the unobservables, which we assumed to be spatially correlated. For instance, using the binary logit model, we estimated that the odds ratio of choosing public transit for people without access to a car in their household was 20 times higher than that for those with car access. However, in the HGP models, the multiplicative effect of not having access to a car on the odds for public transit reached values greater than 30. As another example, the binary logit model indicated that high-income individuals had odds of taking public transit 4% lower than that of less wealthy individuals. In contrast, the HGP models placed this percentage between 6% and 15%. Similarly, we found significant variations in the estimates for the posterior mean and medians associated with β between the logit and HGP models.

For the binary case study, we evaluated the out-of-sample performance of our HGP model by comparing it to a binary logit model and a probit SAE model. We used out-of-sample prediction accuracy and the leave-one-out cross-validation information criterion (LOO-IC) for this comparison. Notably, the HGP model that incorporated destination locations outperformed all other models across all metrics in the binary case study. In the multinomial case study, we similarly assessed the out-of-sample prediction performance of both the IIA and non-IIA HGP models relative to a multinomial logit model. This analysis revealed that the HGP model with IIA exhibited superior out-of-sample prediction performance, with prediction accuracy almost 10 percentage points higher than the multinomial logit model and better LOO-IC.

Our study underscores the critical importance of accounting for different sources of unobserved correlation within the latent utility specification. As demonstrated in this paper, a logit model that overlooks correlated unobservables can yield biased and inconsistent estimates if latent utility variables exhibit correlation structures similar to the unobserved effects. Consequently, for inference in the presented binary mode choice problem, we recommend employing one of the HGP models that incorporate trip coordinates. Particularly, based on the estimated kernel parameters, posterior shapes, and out-of-sample prediction performance, the HGP model with destination locations emerges as the preferred choice over the standard logit specification. Similarly, for the multinomial case study, the HGP model with IIA stands out due to its superior prediction performance, making it the recommended choice. For prediction purposes, practitioners may also consider exploring Bayesian Model Averaging (BMA) (Hoeting et al., 1999) that incorporates competing model specifications tailored to their specific use case.

Our current implementation leverages a kernel function within the HGP that enforces stationarity and isotropy for the latent utility's spatial process. However, these assumptions can be relaxed by adopting alternative, well-defined kernels that are anisotropic and non-stationary. Similarly, in the multinomial case study, we assumed that the blocks of the block-diagonal variance–covariance matrix K were generated using a common kernel, and we only considered correlations between alternatives at the intra-individual level. Both of these assumptions can be revisited with careful consideration for model identification.

We considered a single approach to incorporate non-IIA substitution patterns into our model, following a similar logic to that used in probit models, where substitution patterns are modeled through the per-individual covariance between the marginal utilities of alternatives. However, the flexible specification of our model allows for alternative approaches to modeling substitution patterns, potentially leading to more efficient estimation.

Regarding the limitations of our illustrative examples, we performed alternative aggregation. Whereas this is a common practice in discrete choice modeling, it can lead to biased parameter estimates depending on the aggregation strategy and dataset (Wong et al., 2019). In future empirical studies, we plan to construct datasets with elemental alternatives to examine this potential bias and compare its magnitude between our models that incorporate correlated error terms and those that do not.

We have explored how HGPs can generalize the SAE and SAL models, focusing on the SAE equivalence in this work. The SAL model counterpart could be particularly useful in scenarios where peer effects are likely to be a significant source of correlation, such as investigating the diffusion of new technologies. In this study, we employed three location vectors, s, for our HGP models: trip origin, trip destination, and trip direction. Future research could explore defining alternative location vectors for the travel mode choice problem's latent utility spatial process. Additionally, finding ways to efficiently incorporate both origin and destination locations, without information loss as encountered with the trip direction vector, is a promising avenue for future work. Furthermore, the structure of the variance–covariance matrix presented here for the multinomial non-IIA HGP warrants further investigation to potentially reduce the computational burden in model estimation.

CRediT authorship contribution statement

Daniel F. Villarraga: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ricardo A. Daziano:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Funding acquisition.

Acknowledgments

This research was supported by the National Science Foundation Award No. SES-2342215. We are also thankful for the financial support provided by the Fulbright Scholarship Program, which is sponsored by the U.S. Department of State, the Colombian Fulbright Commission, and the Colombian Science Ministry. This project is solely the responsibility of the authors and does not necessarily represent the official views of the Fulbright Program, the U.S. government, or the Colombian government.

Data availability

Data will be made available on request.

References

Anon, 2023. Census Tracts. U.S. Census Bureau, https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.2010.html#list-tab-264479560.

Anselin, Luc, 2001. Spatial econometrics. A companion to theoretical econometrics.

Anselin, Luc, Florax, Raymond, Rey, Sergio J., 2013. Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer Science & Business Media.

Bhat, Chandra R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transp. Res. B 45 (7), 923–939.

Bhat, Chandra, 2015. A new spatial (social) interaction discrete choice model accommodating for unobserved effects due to endogenous network formation. Transportation 42, 879–914.

Bhat, Chandra R., Pinjari, Abdul R., Dubey, Subodh K., Hamdi, Amin S., 2016. On accommodating spatial interactions in a generalized heterogeneous data model (GHDM) of mixed types of dependent variables. Transp. Res. B 94, 240–263.

Bhat, Chandra R., Sener, Ipek N., Eluru, Naveen, 2010. A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. Transp. Res. B 44 (8–9), 903–921.

Datta, Abhirup, Banerjee, Sudipto, Finley, Andrew O., Gelfand, Alan E., 2016. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. J. Amer. Statist. Assoc. 111 (514), 800–812.

Finley, Andrew O., Datta, Abhirup, Cook, Bruce D., Morton, Douglas C., Andersen, Hans E., Banerjee, Sudipto, 2019. Efficient algorithms for Bayesian nearest neighbor Gaussian processes. J. Comput. Graph. Statist. 28 (2), 401–414.

Goetzke, Frank, 2008. Network effects in public transit use: evidence from a spatially autoregressive discrete choice model for New York. Urban Stud. 45 (2), 407–417.

Goulard, Michel, Laurent, Thibault, Thomas-Agnan, Christine, 2017. About predictions in spatial autoregressive models: Optimal and almost optimal strategies. Spatial Econ. Anal. 12 (2–3), 304–325.

Guo, Shengbo, Sanner, Scott, Bonilla, Edwin V., 2010. Gaussian process preference elicitation. Adv. Neural Inf. Process. Syst. 23.

Hoeting, Jennifer A., Madigan, David, Raftery, Adrian E., Volinsky, Chris T., 1999. Bayesian model averaging: A tutorial. Statist. Sci. 14 (4), 382–401.

Lewandowski, Daniel, Kurowicka, Dorota, Joe, Harry, 2009. Generating random correlation matrices based on vines and extended onion method. J. Multivariate Anal. 100 (9), 1989–2001.

Li, Zheng, Hensher, David A., 2013. Crowding in public transport: a review of objective and subjective measures. J. Public Transp. 16 (2), 107–134.

MacKay, David J.C., Mac Kay, David J.C., 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press.

New York City Department of Transportation, 2023. Parking rates. Accessed: 2023, https://www.nyc.gov/html/dot/html/motorist/parking-rates.shtml.

New York Metropolitan Transportation Council, North Jersey Transportation Planning Authority, 2014. 2010/2011 Regional Household Travel Survey. United States. New York Metropolitan Transportation Council.

Riihimäki, Jaakko, Jylänki, Pasi, Vehtari, Aki, 2013. Nested expectation propagation for Gaussian process classification with a multinomial probit likelihood. J. Mach. Learn. Res. 14 (Jan), 75–109.

Sfeir, Georges, Rodrigues, Filipe, Abou-Zeid, Maya, 2022. Gaussian process latent class choice models. Transp. Res. C 136, 103552.

Train, Kenneth E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press.

US Department of Transportation, 2016. Revised departmental guidance on valuation of travel time in economic analysis.

Vecchia, Aldo V., 1988. Estimation and model identification for continuous spatial processes. J. R. Stat. Soc. Ser. B Stat. Methodol. 50 (2), 297–312.

Vehtari, Aki, Gelman, Andrew, Gabry, Jonah, 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27, 1413–1432.

Vehtari, Aki, Gelman, Andrew, Simpson, Daniel, Carpenter, Bob, Bürkner, Paul-Christian, 2021. Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC (with discussion). Bayesian Anal. 16 (2), 667–718.

Wilhelm, Stefan, de Matos, Miguel Godinho, 2013. Estimating spatial probit models in R. R J. 5 (1), 130–143.

Wilson, Andrew G., Izmailov, Pavel, 2020. Bayesian deep learning and a probabilistic perspective of generalization. Adv. Neural Inf. Process. Syst. 33, 4697–4708.

Wong, Timothy, Brownstone, David, Bunch, David S., 2019. Aggregation biases in discrete choice models. J. Choice Model. 31, 210–221.

Zhang, Lu, 2018. Nearest neighbor Gaussian processes (NNGP) based models in stan. (Accessed: 08 November 2023).