Received: 6 July 2023 | Accepted: 4 January 2024

DOI: 10.1002/sta4.656

ORIGINAL ARTICLE

WILEY

Table inference for combinatorial origin-destination choices in agent-based population synthesis

Ioannis Zachos[1] | Theodoros Damoulas[2,3] | Mark Girolami[1,3]

[1]Department of Engineering, University of Cambridge, Cambridge, UK

[2]Departments of Computer Science and Statistics, University of Warwick, Coventry, UK

[3]The Alan Turing Institute, London, UK

Correspondence
Ioannis Zachos, Department of Engineering, Trumpington Street, Cambridge, CB2 1PZ, UK.
Email: iz230@cam.ac.uk

Funding information
Engineering and Physical Sciences Research Council; UK Research and Innovation; Ove Arup Foundation; Alan Turing Institute

A key challenge in agent-based mobility simulations is the synthesis of individual agent socioeconomic profiles. Such profiles include locations of agent activities, which dictate the quality of the simulated travel patterns. These locations are typically represented in origin-destination matrices that are sampled using coarse travel surveys. This is because fine-grained trip profiles are scarce and fragmented due to privacy and cost reasons. The discrepancy between data and sampling resolutions renders agent traits nonidentifiable due to the combinatorial space of data-consistent individual attributes. This problem is pertinent to any agent-based inference setting where the latent state is discrete. Existing approaches have used continuous relaxations of the underlying location assignments and subsequent ad hoc discretisation thereof. We propose a framework to efficiently navigate this space offering improved reconstruction and coverage as well as linear-time sampling of the ground truth origin-destination table. This allows us to avoid factorially growing rejection rates and poor summary statistic consistency inherent in discrete choice modelling. We achieve this by introducing joint sampling schemes for the continuous intensity and discrete table of agent trips, as well as Markov bases that can efficiently traverse this combinatorial space subject to summary statistic constraints. Our framework's benefits are demonstrated in multiple controlled experiments and a large-scale application to agent work trip reconstruction in Cambridge, UK.

KEYWORDS
combinatorial explosion, Markov bases, origin-destination matrix, population synthesis, spatial interaction models

1 | INTRODUCTION

Agent-based models (ABMs) are becoming increasingly popular policy-making tools in areas such as epidemic and transportation modelling (Bonabeau, 2002). The emergent structure arising from ABM simulations relies on the quality of the underlying agent population's demographic and socioeconomic attributes. In transportation ABMs, such as MATSim (Axhausen & Zürich, 2016), simulated travel patterns are predominantly governed by the location where agent activities take place (e.g. working and shopping). The trips between activities are summarised in origin-destination matrices (ODMs), which are often either partially or not available a priori. Therefore, *population synthesis* is performed to create artificial agents whose attributes (e.g. workplace location) have the same summary statistics as those described by population averages (e.g. regional job availability). Location choice synthesis translates to reconstructing integer-valued ODMs whose margins are summary statistics. To this end, coarse/aggregate agent activity surveys by geographical region and activity type are mainly leveraged (Fournier et al., 2021). This is because fine-grained individual/disaggregate profiles are scarce and fragmented due to privacy and/or data acquisition cost reasons. Therefore, a discrepancy arises between the spatial resolutions of the data and latent states. Inferring individual agent trips subject to population summary statistics necessitates the exploration of a combinatorial choice space. The size of this space induces identifiability issues since a unique set of agent location choices consistent with the data cannot be recovered.

A downsampling approach of sampling individual choices is computationally infeasible for any real-world application. Assuming that there are M agents with L available location choices, then computing the likelihood of the aggregate data given individual model parameters requires summing over L^M possible location configurations, many of which are inconsistent with the data. Computational and identifiability issues can be alleviated by appropriately constraining the discrete latent space. The problem of exploring a constrained combinatorial agent state space is pertinent to any agent-based inference setting where the latent state is discrete.

Although discrete choice models (Train, 2009) are popular candidates for disaggregating agent location choices, they cannot encode aggregate statistic constraints. Therefore, they either accrue errors when reconstructing ODMs or lead to factorially growing rejection rates (DeSalvo & Zhao, 2016) when forced to adhere to discrete constraints in a rejection-type scheme. A suite of greedy optimisation algorithms such as iterative proportional fitting (Bishop et al., 2007) and combinatorial optimisation (Voas & Williamson, 2000) was employed to assimilate summary statistic constraints in continuous and discrete spaces, respectively. These methods suffer from poor convergence to local optima, yielding solutions heavily dependent on good initialisations. Moreover, operating in a continuous probability/intensity space requires an additional sampling step to discretise the ODM, such as stochastic rounding (Croci et al., 2022). This is an ad hoc treatment of the problem and produces errors. The unidentifiable nature of disaggregating agent choices from aggregate data calls for uncertainty quantification in order to give practitioners the ability to interrogate and rank the sampled ODMs according to their probability.

Probabilistic methods have overcome some of the aforementioned limitations (Farooq et al., 2013; Sun & Axhausen, 2016) but remain approximate since they operate in the continuous intensity/probability space. In the case of location choice synthesis, ODMs are equivalent to two-way contingency tables of two categorical variables (e.g. origin residential population and destination workforce population), and the joint distribution of the two variables is explored using Gibbs sampling. Table marginal probabilities are elicited by normalising the discrete summary statistics. This approximation incurs information loss and may cause marginal class imbalances in high-dimensional tables (Fournier et al., 2021), meaning a growing divergence between ground truth and sampled marginal distributions. In addition, partially available data cannot be accommodated in a principled manner, and unreasonable conditional independence assumptions are imposed.

The work of Carvalho (2014) endeavoured to address these two problems by adopting a Bayesian paradigm that operates directly on the discrete table space. However, neither the most efficient proposal mechanism nor the available intensity structure was exploited. Instead, a Metropolis–Hastings (MH) scheme for sampling contingency tables was employed that proposes jumps of size at most one in $\mathcal{O}(\#\text{ origins} \times \#\text{ destinations})$, causing poor mixing in high-dimensional tables. Furthermore, the author argued for a hierarchical construction that jointly learns the constrained discrete ODM and the underlying intensity function. In doing so, they attempted to leverage a family of log-nonlinear intensity models known as *spatial interaction models* (SIMs) (Wilson, 1971). SIMs incorporate summary statistic constraints directly in the continuous intensity space. Despite this effort, a log-linearity assumption was imposed on the SIM to simplify parameter inference. Also, the known dynamics of competition between destination locations (Dearden & Wilson, 2015) were ignored, effectively stripping SIMs of all their embedded structure. Moreover, additional data were required to calibrate the intensity function, such as seed matrices, which are seldom available, as opposed to regularly observed data on the economic utility of travelling to destination locations. The works of Ellam et al. (2018) and Gaskin et al. (2023) alleviated this problem by constructing a physics-driven log-nonlinear SIM intensity prior. However, both approaches operated strictly in the continuous intensity space and could not explore the discrete table space where population synthesis is performed.

1.1 | Contributions

Our paper focuses on reconstructing origin-destination agent trip matrices summarising residence-to-workplace location choices. We offer an upsampling Bayesian approach that jointly samples from the discrete table ($\mathbf{T}$) and continuous intensity ($\boldsymbol{\Lambda}$) spaces for agent location choice synthesis. Our framework seamlessly assimilates any type of aggregate summary statistic as a constraint, which in turn regularises the space of admissible disaggregate/individual agent choices. We demonstrate an improved reconstruction and coverage of a partially observed origin-destination matrix summarising agent trips from residential to workplace locations in Cambridge, UK.

Contrary to the previous work, we perform a Gibbs step and sample tables in $\mathcal{O}(\#\text{ destinations})$ leveraging the Markov basis (MB) machinery in Diaconis and Sturmfels (1998) to design a Markov Chain Monte Carlo (MCMC) scheme with proposals that allow arbitrarily large jumps in table space without any accept/reject step. Hence, we bypass the problem of marginal distribution imbalances by respecting the exact margin frequencies rather than marginal distributions. We employ SIMs to understand the behavioural mechanism of aggregate location choice in continuous intensity space and relax previously adopted log-linearity assumptions on the intensity model. In the same fashion as Ellam et al. (2018) and Gaskin et al. (2023), we account for the stochastic dynamics of competition between destinations governing agent location choices and enforce an interpretable structure in the SIM intensity prior. A summary of our framework's capabilities relative to the previous works is depicted in Table 1.

TABLE 1 Comparison of our method's capabilities against previous works.

$\mathcal{C}$	Constrained ODM	This work	(Ellam)	(Gaskin)	$\mathbb{P}(\mathbf{T}\mid\boldsymbol{\Lambda},\mathcal{C})$
$\{\Lambda_{++}\}$	Totally	✓	✓	✓	-
$\{\boldsymbol{\Lambda}_{\cdot+}\}$	Singly	✓	✓	✓	-
$\{T_{++}, \Lambda_{++}\}$	Totally	✓	×	×	Multinomial (T_{++}, Λ_{++})
$\{\mathbf{T}_{\cdot+}, \boldsymbol{\Lambda}_{\cdot+}\}$ or $\{\mathbf{T}_{\cdot+}, \Lambda_{++}\}$	Singly	✓	×	×	Product Multinomial $(\mathbf{T}_{\cdot+}, \frac{\boldsymbol{\Lambda}_{\cdot+}}{\Lambda_{++}})$
$\{\mathbf{T}_{\cdot+}\mathbf{T}\mathbf{T}_{+\cdot}, \mathbf{T}\Lambda_{++}\}$	Doubly	✓	×	×	Fisher's non-central hypergeometric $(\mathbf{T}_{\cdot+}, \mathbf{T}_{+\cdot}, \frac{\Lambda_{++}\boldsymbol{\Lambda}_{\cdot\cdot}}{\boldsymbol{\Lambda}_{\cdot+}\boldsymbol{\Lambda}_{+\cdot}})$
$\{\ \mathbf{T}_{\mathcal{X}'}, \Lambda_{++},\ \}$ $\forall\ \mathcal{X}' \subseteq \mathcal{P}(\mathcal{X})$	Doubly and cell	✓	×	×	Constrained Fisher's non-central hypergeometric $(\mathbf{T}_{\cdot+}, \mathbf{T}_{+\cdot}, \frac{\Lambda_{++}\boldsymbol{\Lambda}_{\cdot\cdot}}{\boldsymbol{\Lambda}_{\cdot+}\boldsymbol{\Lambda}_{+\cdot}})$

Note: Agent choices are described by a discrete table ($\mathbf{T}$) or a continuous intensity ($\boldsymbol{\Lambda}$). Subscripts define summary statistics: the row and column sums/margins are indexed by $(\cdot,+),(+,\cdot)$, respectively. The cell universe $\mathcal{X} \supseteq \mathcal{X}'$ contains table/intensity indices of an $I \times J$ matrix.
Abbreviation: ODM, origin-destination matrix.

2 | PROBLEM SETUP

Consider M agents that travel from I origins to J destinations to work. Let the expected number of trips (intensity) of agents between origin i and destination j be denoted by Λ_{ij}. The residential population in each origin (row sums) is equal to

$$\Lambda_{i+} = \sum_{j=1}^{J} \Lambda_{ij}, \quad i = 1,\ldots,I, \tag{1}$$

while the working population at each destination (column sums) is

$$\Lambda_{+j} = \sum_{i=1}^{I} \Lambda_{ij}, \quad j = 1,\ldots,J. \tag{2}$$

We assume that the total origin and destination demand are both conserved:

$$M = \Lambda_{++} = \sum_{i=1}^{I} \Lambda_{i+} = \sum_{j=1}^{J} \Lambda_{+j}. \tag{3}$$

This construction defines a totally constrained SIM. The demand for destination zones depends on the destination's attractiveness denoted by $\mathbf{w} := (w_1,\ldots,w_J) \in \mathbb{R}^J_{>0}$. Let the log-attraction be $\mathbf{x} := \log(\mathbf{w})$. Between two destinations of similar attractiveness, agents are assumed to prefer nearby zones. Therefore, a cost matrix $\mathbf{C} = (c_{i,j})_{i,j=1}^{I,J}$ is introduced to reflect travel impedance. These two assumptions are justified by economic arguments Pooler (1994). The maximum entropy distribution of agent trips subject to the total number of agents being conserved is derived by maximising

$$\mathcal{E}(\boldsymbol{\Lambda}) = \sum_{i=1}^{I}\sum_{j=1}^{J} -\Lambda_{ij}\log(\Lambda_{ij}) + \zeta\left(\sum_{i,j}^{I,J}\Lambda_{ij} - M\right) + \alpha\sum_{j=1}^{J} x_j\Lambda_{ij} - \beta\sum_{i,j}^{I,J} c_{ij}\Lambda_{ij}, \tag{4}$$

which yields a closed-form expression for the trip intensity:

$$\Lambda_{ij} = \frac{\Lambda_{++}\exp(\alpha x_j - \beta c_{ij})}{\sum_{k,m}^{I,J}\exp(\alpha x_m - \beta c_{km})}, \tag{5}$$

where α,β control the two competing forces of attractiveness and deterrence. A higher α relative to β characterises a preference over destinations with higher job availability, while the contrary indicates a predilection for closer workplaces. The destination attractiveness $\mathbf{w}$ is governed by the Harris-Wilson (Harris & Wilson, 1978) system of J coupled ordinary differential equations (ODEs):

$$\frac{\mathrm{d}w_j}{\mathrm{d}t} = \epsilon w_j\left(\Lambda_{+j} - \kappa w_j + \delta\right), \quad \mathbf{w}(0) = \mathbf{w}', \tag{6}$$

where $\kappa > 0$ is the number of agents competing for one job, $\delta > 0$ is the smallest number of jobs a destination can have and $\Lambda_{+j}(t) - \kappa w_j(t)$ is the net job capacity in destination j. A positive net job capacity translates to a higher economic activity (more travellers than jobs) and a boost in local employment, and vice versa. In equilibrium, the J stationary points of the above ODE can be computed using

$$\kappa w_j - \delta = \frac{\Lambda_{++} w_j^{\alpha}}{\sum_{k,m}^{I,J} w_k^{\alpha} \exp(-\beta c_{km})} \sum_{i=1}^{I} \exp(-\beta c_{ij}). \tag{7}$$

The value of κ can be elicited by summing the above equation over destinations, which yields

$$\kappa = \frac{\delta J + \Lambda_{++}}{\sum_{j=1}^{J} w_j}, \tag{8}$$

while δ corresponds to the case when no agent travels to destination j' ($\Lambda_{+j'} = 0$), that is,

$$\delta = \kappa \min_j \{w_j\}. \tag{9}$$

A stochastic perturbation of 6 incorporates uncertainty in the competition dynamics emerging from the randomness of agents' choice mechanisms. This gives rise to the Harris-Wilson stochastic differential equation (SDE) for the time evolution of log destination attraction $\mathbf{x}$

$$\mathrm{d}\mathbf{x} = -\epsilon^{-1} \nabla V(\mathbf{x}) \mathrm{d}t + \sqrt{2\gamma^{-1}} \mathrm{d}\mathbf{B}_t, \quad \mathbf{x}(0) = \mathbf{x_0}, \tag{10}$$

where the potential function $V(\mathbf{x})$ in the drift term is equal to

$$\epsilon^{-1} V(\mathbf{x}) = \underbrace{-\alpha^{-1} \sum_{i=1}^{I} O_i \log\left(\sum_{j=1}^{J} \exp\left(\alpha x_j - \beta c_{ij}\right)\right)}_{\text{utility potential}} + \underbrace{\kappa \sum_{j=1}^{J} \exp\left(x_j\right)}_{\text{cost potential}} - \underbrace{\delta \sum_{j=1}^{J} x_j}_{\text{additional potential}}, \tag{11}$$

and $\boldsymbol{\theta} = (\alpha, \beta)$ is the free parameter vector. The steady-state distribution of 10 is shown in Ellam et al. (2018) to be the Boltzmann–Gibbs measure

$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp(-\gamma V_{\boldsymbol{\theta}}(\mathbf{x})) \tag{12}$$

$$Z(\boldsymbol{\theta}) := \int_{\mathbb{R}^J} \exp(-\gamma V_{\boldsymbol{\theta}}(\mathbf{x})) \mathrm{d}\mathbf{x}. \tag{13}$$

The observed data $\mathbf{y}$ are assumed to be noisy perturbations of $\mathbf{x}$, where the error between the two satisfies $\log(\mathbf{e}) \sim \mathcal{N}(\mathbf{0}, \sigma_d^2 \mathbf{I})$, that is

$$\log(\mathbf{y}) = \mathbf{x} + \log(\mathbf{e}). \tag{14}$$

We introduce a data augmentation step to perform inference at the higher resolution origin-destination table space of agent trips as depicted in Figure 1. Assume that the $I \times J$ discrete contingency table $\mathbf{T}$ summarising the number of agents living in location i and working in location j is Poisson distributed as follows:

$$T_{ij} \sim \text{Poisson}\left(\Lambda_{ij}(\mathbf{x}, \boldsymbol{\theta})\right), \tag{15}$$

where the T_{ij}'s are conditionally independent given the Λ_{ij}'s. The contingency table inherits constraint 3. These hard-coded constraints can be viewed as noise-free data on the discrete table space. We abbreviate the vector of row sums, column sums and the scalar total of $\mathbf{T}$ by $\mathbf{T}_{\cdot+}$, $\mathbf{T}_{+\cdot}$

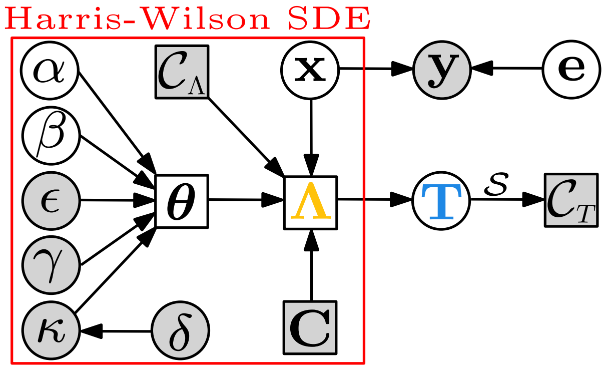

FIGURE 1 Plate diagram of our modelling framework. Rectangular and circular nodes are deterministic and random variables, respectively. Shaded nodes correspond to conditioned quantities.

and T_{++}, respectively. Note that $\mathbf{T}$ uniquely determines the rest of the aforementioned random variables and $T_{++} = \Lambda_{++}$. Moreover, the distribution of $\mathbf{x}$ in 12 coupled with a prior on $\boldsymbol{\theta}$ jointly induces a prior over the intensity function $\boldsymbol{\Lambda}$.

Performing inference in a discrete higher-resolution table space circumvents challenges associated with enforcing summary statistic constraints in the continuous intensity space. First, the doubly constrained intensity (see Table 1) admits solutions retrieved only through an iterative procedure that converges to poor local optima without any quantification of uncertainty, since the physical model in (10) becomes redundant. Second, maximising (4) subject to individual cell constraints induces discontinuities in the Λ space prohibiting SIM parameter calibration. To avoid dealing with discontinuities, a fully observable table is required, which is seldom available and defeats the purpose of ODM reconstruction. Alternatively, more parameters can be introduced, which entails identifiability problems as the number of free parameters becomes $\mathcal{O}(I+J)$ instead of $\mathcal{O}(J)$. Moreover, augmenting $\mathcal{C}_\Lambda$ to match $\mathcal{C}_T$ strengthens the dependence between $\mathbf{T}|\boldsymbol{\Lambda},\mathcal{C}$ and $\mathbf{y}|\mathbf{x},\mathcal{C}$. As a result, constraints are implicitly weighted (hard $\mathcal{C}_\Lambda$ and soft $\mathcal{C}_T$ constraints), which inflicts identifiability issues in $\boldsymbol{\Lambda}$.

3 | DISCRETE TABLE INFERENCE

Let the set of table indices (cells) be $\mathcal{X} = \{(i,j) : 1 \leq i \leq I, 1 \leq j \leq J\}$ such that $T(x) = T_{ij}$ is the table value of cell $x = (i,j) \in \mathcal{X}$. For any subset $\mathcal{X}_k \subseteq \mathcal{X}$ let $S_k : \mathcal{X}_k \to \mathbb{N}^{I+J}$ be a bijective function that maps every cell $x \in \mathcal{X}_k$ to the $(I+J)$-dimensional binary vector with the i-th and $(I+j)$-th entries equal to one and the rest being zero. Define $\mathcal{S}_k : \mathcal{T} \to \mathbb{N}^{I+J}$ to be the summary statistic operator applying a uniquely defined $S_k(\cdot)$ to a table $\mathbf{T} \in \mathcal{T}$ over cells $\mathcal{X}'_k \subseteq \mathcal{X}$ such that $\mathcal{S}_k(\mathbf{T}') = \sum_{x \in \mathcal{X}} \mathbf{T}'(x) S_k(x)$. The ordered collection[1] of summary statistic operators $\{\mathcal{S}_1(\mathbf{T}'), \dots, \mathcal{S}_K(\mathbf{T}')\}$ is abbreviated by $\mathcal{S}(\mathbf{T}')$. Define a collection of discrete summary statistics $\mathcal{C}_T = \{\mathbf{s}_1, \dots, \mathbf{s}_K\}$ expressed as constraints on table space, where each $\mathbf{s}_k$ is a realisation of $\mathcal{S}_k$. We leverage the same convention to define continuous constraints $\mathcal{C}_\Lambda$ in the intensity space. The union of table and intensity constraints is summarised by $\mathcal{C}$. We sometimes refer to $\mathcal{C}_T$ by $\mathcal{C}$ to avoid notation clutter. In Table 1, the singly constrained ODM model corresponds to a given $\mathcal{C}$ as opposed to singly constrained tables and intensities that map to $\mathcal{C}_T$ and $\mathcal{C}_\Lambda$, respectively. Equivalently, constrained models are defined by combinations of constrained tables and intensities.

Definition 1. Consider an ordered[2] collection of constraints $\mathcal{C}_T$ and table summary statistics operators $\mathcal{S}$ with associated functions $\mathbf{S}$. A table $\mathbf{T}'$ is $\mathcal{C}_T$*-admissible* if and only if its summary statistics satisfy all the constraints in $\mathcal{C}_T$, that is, $\mathcal{S}_k(\mathbf{T}') = \mathbf{s}_k \in \mathcal{C}_T \ \forall \ k = 1, \dots, K$.

We denote the function space of all $\mathcal{C}_T$-admissible contingency tables (ODMs) of dimension $\dim(\mathbf{T}) = I \times J$ by $\mathcal{T}_{\mathcal{C}_T} = \{\mathbf{T} \in \mathcal{T} : \mathcal{S}(\mathbf{T}) = \mathcal{C}_T\}$ and drop the dependence on T for notational convenience. This space contains all agent location choices consistent with the aggregate summary statistics $\mathcal{C}_T$. The set $\mathcal{T}_{\mathcal{C}_k}$ contains all tables that when applied $\mathcal{S}_k$ over cells $\mathcal{X}_k$ satisfy the k-th constraint of $\mathcal{C}_T$. In the rest of the paper we set $\mathcal{C}_\Lambda = \{\Lambda_{++}\}$ unless otherwise stated. Our goal is to sample from $P(\mathbf{T}, \mathbf{x}, \boldsymbol{\theta} | \mathcal{C}, \mathbf{y})$[3], where

$$P\left(\mathbf{T}, \underbrace{\mathbf{x}, \boldsymbol{\theta}}_{\boldsymbol{\Lambda}} | \mathcal{C}, \mathbf{y}\right) \propto \mathbb{P}(\mathbf{T} | \mathbf{x}, \boldsymbol{\theta}, \mathcal{C}) p(\mathbf{y} | \mathbf{x}) p(\mathbf{x} | \boldsymbol{\theta}, \mathcal{C}) p(\boldsymbol{\theta}) \tag{16}$$

We achieve this by devising a MH-within-Gibbs scheme to sample from $\mathbb{P}(\mathbf{T} | \mathbf{x}, \boldsymbol{\theta}, \mathcal{C})$, $p(\mathbf{x} | \boldsymbol{\theta}, \mathbf{T}, \mathbf{y})$ and $p(\boldsymbol{\theta} | \mathbf{x}, \mathbf{T}, \mathbf{y})$. The conditional samplers for $\mathbf{x}$ and $\boldsymbol{\theta}$ have acceptance ratios similar to those in Ellam et al. (2018) and equal to

$$p(\mathbf{x}',\mathbf{m}'|\mathbf{x},\mathbf{m},\boldsymbol{\theta},\mathbf{T},\mathcal{C},\mathbf{y}) = \min\left(1,\frac{\mathbb{P}(\mathbf{T}|\mathbf{x}',\boldsymbol{\theta},\mathcal{C})p(\mathbf{y}|\mathbf{x}')\exp(-H_\theta(\mathbf{x}'))}{\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta},\mathcal{C})p(\mathbf{y}|\mathbf{x})\exp(-H_\theta(\mathbf{x}))}\right), \tag{17}$$

$$p(\boldsymbol{\theta}'|\boldsymbol{\theta},\mathbf{x},\mathbf{m},\mathbf{T},\mathcal{C},\mathbf{y}) = \min\left(1,\frac{\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta}',\mathcal{C})\exp(-\gamma V_{\boldsymbol{\theta}'}(\mathbf{x}))Z(\boldsymbol{\theta})p(\boldsymbol{\theta}')}{\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta},\mathcal{C})\exp(-\gamma V_{\boldsymbol{\theta}}(\mathbf{x}))Z(\boldsymbol{\theta}')p(\boldsymbol{\theta})}\right), \tag{18}$$

where $H_\theta(\mathbf{x}') = -\gamma V_{\boldsymbol{\theta}}(\mathbf{x}') - 1/2|\mathbf{m}'|^2$ is the Hamiltonian of state $\mathbf{x}'$ with associated momentum $\mathbf{m}'$. Although a singly constrained intensity can be leveraged here, enforcing hard constraints through $\mathcal{C}_\Lambda$ and potentially different soft constraints through $\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta},\mathcal{C})$ would cause identifiability issues in $\mathbf{x}$. We aim to provide a general construction for joint table and intensity inference and employ singly constrained SIMs only when $\mathcal{C}_T = \{\mathbf{T}_{\cdot+}\}$. In the following exposition, we show that the type of summary statistic data available determines whether the constrained table distribution $\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta},\mathcal{C})$ can be sampled directly or indirectly through MCMC.

3.1 | Tractable constrained table sampling

In this section, we offer closed-form contingency table sampling. Without loss of generality, assume that only one of the two table margins is known, namely, $\mathbf{T}_{+\cdot}$ (singly constrained table). Then, any subset $\mathcal{C}_T$ of the universe of summary statistic constraints $\{T_{++},\mathbf{T}_{+\cdot},\{\mathbf{T}_{\mathcal{X}_l}|\mathcal{X}_l\subseteq\mathcal{X},l\in\mathbb{N}\}\}$ yields a closed-form posterior table marginal as shown in Table 1. By the construction in (15), the case for $\mathcal{C}_T=\emptyset$ is equivalent to unconstrained table sampling conditioned on an intensity model, which in our case is a SIM. Cell constraints $\mathcal{C}_T = \{\mathbf{T}_{\mathcal{X}_l}|\mathcal{X}_l\subseteq\mathcal{X},l\in\mathbb{N}\}$ can be seamlessly incorporated in an unconstrained table without violating the posterior's tractability. Furthermore, leveraging that $\mathbf{T}$ uniquely determines both T_{++} and $\mathbf{T}_{+\cdot}$ and applying Bayes' rule, it follows that the models with $\mathcal{C}_T=\{T_{++}\}$ and $\mathcal{C}_T=\{\mathbf{T}_{+\cdot}\}$ yield Multinomial and product Multinomial distributions, respectively (See full derivations in supporting information). Equivalently,

$$\mathbb{P}(\mathbf{T}|\boldsymbol{\Lambda},T_{++}) = \prod_{i,j}^{I,J}\left(\frac{T_{++}!}{T_{ij}!}\left(\frac{\Lambda_{ij}}{\Lambda_{++}}\right)^{T_{ij}}\right), \tag{19}$$

and

$$\mathbb{P}(\mathbf{T}|\boldsymbol{\Lambda},\mathbf{T}_{\cdot+}) = \prod_{i,j}^{I,J}\left(\frac{T_{i+}!}{T_{ij}!}\left(\frac{\Lambda_{ij}}{\Lambda_{i+}}\right)^{T_{ij}}\right). \tag{20}$$

We obtain N independent samples $\mathbf{T}^{(1:N)}$ from (15), (19) (20) in closed-form. Samples from the Poisson and product multinomial distributions can be drawn in parallel. We note that the space complexity of table sampling is $\mathcal{O}(IJ)$ while the time complexity for (15), (19) (20) is $\mathcal{O}(IJ)$, $\mathcal{O}(1)$ and $\mathcal{O}(I)$, respectively. Moreover, coupling either constraint T_{++} or $\mathbf{T}_{\cdot+}$ with cell constraints leaves the target distribution unchanged but shrinks its support. Hence, the available table margin is updated by subtracting the value of fixed cell constraints from margin statistics and performing inference on the free cells. We present the joint intensity and table sampling algorithm for tractably constrained tables in Algorithm 1.

Algorithm 1 Metropolis-within-Gibbs MCMC sampling algorithm for tractably constrained tables.

```
Inputs: $\mathbf{C}, \mathcal{C}, \mathbf{y}, N$.
Outputs: $\mathbf{x}^{(1:N)}, \theta^{(1:N)}, \text{sign}\left(\theta^{(1:N)}\right), \mathbf{T}^{(1:N)}$.
Initialise $\mathbf{x}^{(0)}, \theta^{(0)}, \mathbf{T}^{(0)}$.
for $n \in \{1, \dots, N\}$ do
    Sample $\mathbf{x}^{(n)}|\theta^{(n-1)}, \mathbf{T}^{(n-1)}$ using Hamiltonian Monte Carlo (Neal, 2011) with acceptance 17.
    Sample $\theta^{(n)}|\mathbf{x}^{(n)}, \mathbf{T}^{(n-1)}$ using Random Walk Metropolis–Hastings with acceptance 18.
    Construct intensity $\boldsymbol{\Lambda}^{(n)}$ using $\mathbf{x}^{(n)}, \theta^{(n)}$ using 5.
    Sample tables (in parallel) using the relevant closed-form distribution ( 15, 19, 20) $\mathbf{T}^{(n)} \sim \mathbb{P}(\mathbf{T}|\boldsymbol{\Lambda}^{(n)}, \mathcal{C})$.
end for
```

3.2 | Intractable constrained table sampling

In this section, we introduce an MCMC scheme for sampling tables subject to any subset of the power set $\mathcal{P}(\{\mathbf{T}_{\cdot+}, \mathbf{T}_{+\cdot}, \{\mathbf{T}_{\mathcal{X}_l} | \mathcal{X}_l \subseteq \mathcal{X}, l \in \mathbb{N}\}\})$ excluding those subsets contained in the constraint universe of the previous section. By conditioning on both table margins and leveraging the conditional distributions of $\mathbf{T}_{\cdot+}|T_{++},\boldsymbol{\Lambda}$ and $T_{++}|\boldsymbol{\Lambda}$, the induced conditional distribution becomes Fisher's noncentral multivariate hypergeometric (Agresti, 2002):

$$\mathbb{P}(\mathbf{T}|\boldsymbol{\Lambda}, \mathbf{T}_{\cdot+}, \mathbf{T}_{\cdot+}) \propto \frac{\prod_{i=1}^{I} T_{i+}! \prod_{j=1}^{J} T_{+j}!}{T_{++}! \prod_{i,j=1}^{I,J} T_{ij}!} \prod_{i,j=1}^{I,J} \left(\frac{\Lambda_{ij} \Lambda_{++}}{\Lambda_{i+} \Lambda_{+j}} \right)^{T_{ij}}, \tag{21}$$

where $\omega_{ij} = \frac{\Lambda_{ij}\Lambda_{++}}{\Lambda_{i+}\Lambda_{+j}}$ is called the odds ratio and encodes the strength of dependence between row i and column j. Complete independence is achieved if and only if $\omega_{ij} = 1$. Our choice of intensity model encodes this dependence in the travel cost matrix $\mathbf{C}$. Origin-destination independence is achieved if and only if the travel cost's effect on destination choice is irrelevant ($\beta = 0$). Moreover, the normalising constant of (21) is a partition function defined over the support of all tables satisfying the conditioned margins and can't be efficiently computed by direct enumeration. In Appendix A, we prove an extension of Chu–Vandermonde's convolution theorem for multinomial coefficients (Belbachir, 2014) that facilitates computation of the normalising constant in $\mathcal{O}(1)$. In particular, we show that the following identity holds:

$$\binom{T_{++}}{T_{+1} \ldots T_{+J}} \prod_{i,j}^{J} \omega_{ij}^{T_{+j}} = \sum_{\mathcal{S}(\mathbf{T}) = \mathcal{C}_T} \prod_{i,j}^{I,J} \binom{T_{+j}}{T_{1j} \ldots T_{Ij}} \omega_{ij}^{T_{ij}}, \tag{22}$$

where $\binom{T_{+j}}{T_{1j} \ldots T_{Ij}} = \frac{T_{+j}!}{T_{1j}! \ldots T_{Ij}!}$ is the multinomial coefficient. Shrinking the $\mathcal{T}_{\mathcal{C}}$ space using elements of the constraint universe above requires a MB MCMC sampling scheme (Diaconis & Sturmfels, 1998) due to the intractability of the induced table posterior.

3.2.1 | MB MCMC

We construct a $\mathcal{C}_T$-admissible table for initialising MB MCMC using a suite of greedy deterministic algorithms, such as iterative proportional fitting (Bishop et al., 2007). We concoct a proposal mechanism on $\mathcal{T}_{\mathcal{C}}$ as follows.

Definition 2. A *null-admissible* table $\mathbf{T}$ is a $\mathcal{C}_T$-admissible table with $\mathcal{C}_T \subseteq \{\mathbf{T}_{+\cdot}, \mathbf{T}_{\cdot+}\}$ and $\forall\ \mathbf{s} \in \mathcal{C}_T$ it follows that $\mathbf{s} = \mathbf{0}$.

Definition 3. A *MB* is a set of table moves $\mathbf{f}_1, \ldots, \mathbf{f}_L : \mathcal{X} \to \mathbb{Z}$ that satisfy the following conditions:

1. $\mathbf{f}_l$ is a null-admissible table for $1 \leq l \leq L$ and
2. for any two $\mathcal{C}_T$-admissible $\mathbf{T}, \mathbf{T}'$ there are $\mathbf{f}_{l_1}, \ldots, \mathbf{f}_{l_A}$ with $\eta_l \in \mathbb{N}$ such that $\mathbf{T}' = \mathbf{T} + \sum_{m=1}^{A} \eta_l \mathbf{f}_{l_m}$ and $\mathbf{T} + \sum_{m=1}^{a} \eta_l \mathbf{f}_{l_m} \geq 0$ for $1 \leq a \leq A$.

Condition (i) guarantees that all proposed moves do not modify the summary statistics in $\mathcal{C}_T$, while condition (ii) ensures that there exists a path between any two tables such that any table member of the path is $\mathcal{C}_T$-admissible. The collection of constraints $\mathcal{C}_T$ generates a MB $\mathcal{M}$. When $I \times J$ tables satisfy both row and column margins, $\mathcal{M}$ consists of functions $\mathbf{f}_1, \ldots, \mathbf{f}_L$ such that $\forall\ x = (i_1, j_1), x' = (i_2, j_2) \in \mathcal{X}$ with $i_1 \neq i_2, j_1 \neq j_2$,

$$\mathbf{f}_l(x) = \begin{cases} \eta & \text{if } x = (i_1, j_1) \text{ or } x = (i_1, j_2) \\ -\eta & \text{if } x = (i_2, j_2) \text{ or } x = (i_2, j_1) \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

The case for coupling individual cell constraints with table margins requires a minor modification. Let $\mathcal{X}' \subseteq \mathcal{X}$ and $\mathcal{C}_T'$ be the individual cell admissibility criteria. Then, $\mathcal{M}$ is updated to exclude all basis functions $\mathbf{f}_l$ with $\mathbf{f}_l(x) \neq 0 \quad \forall\ x \in \mathcal{X}'$. Moreover, $\mathcal{C}_T$ is revised so that $\forall\ \mathbf{s} \in \mathcal{C}_T$, $\mathbf{s}' \in \mathcal{C}_T'$, $\mathbf{s}$ is updated to $\mathbf{s} - \mathbf{s}'$ at every $x \in \mathcal{X}'$. In other words, the constrained cell values are deducted from the rest of the summary statistic constraints in $\mathcal{C}_T$. A MB Markov chain (MBMC) can now be constructed.

Proposition 1 Adapted from Diaconis and Sturmfels (1998). Let μ be a probability measure on $\mathcal{T}_{\mathcal{C}}$. Given a MB $\mathcal{M}$ that satisfies 3, generate a Markov chain in $\mathcal{T}_{\mathcal{C}}$ by sampling l uniformly at random from $\{1,\ldots,L\}$. Consider the Markov basis Metropolis–Hastings (MB-MH) and Gibbs (MB-Gibbs) proposals:

1. MB-MH: Let $\eta \in \{-1,1\}$ and choose η from this set with probability $\frac{1}{2}$ independent of l. If the chain is at $\mathbf{T} \in \mathcal{T}_{\mathcal{C}}$ it will move to $\mathbf{T}' = \mathbf{T} + \eta \mathbf{f}_l$ with probability

$$\min\left\{\frac{\mu(\mathbf{T}+\eta\mathbf{f}_l)}{\mu(\mathbf{T})},1\right\}$$

provided $\mathbf{T}' \geq 0$. In all other cases, the chain stays at $\mathbf{T}$.

2. MB-Gibbs: Let $\eta \in \mathbb{Z}$. If the chain is at $\mathbf{T} \in \mathcal{T}_{\mathcal{C}}$, determine the set of η such that $\mathbf{T} + \eta \mathbf{f}_l \geq 0$. Choose

$$\mathbb{P}(\eta) \propto \prod_{x \in \{x \in \mathcal{X} : \mathbf{f}_l(x) \neq 0\}} \frac{1}{\mu(n(x) + \eta \mathbf{f}_l(x))}$$

and move to $\mathbf{T}' = \mathbf{T} + \eta \mathbf{f}_l \geq 0$.

In both cases, an aperiodic, reversible, connected Markov chain in $\mathcal{T}_{\mathcal{C}}$ is constructed with stationary distribution proportional to $\mu(\mathbf{T})$.

The proof of Proposition 1 is provided in Diaconis and Sturmfels (1998). Theoretical guarantees of MB MCMC convergence on $\mathcal{T}_{\mathcal{C}}$ show that the MB-MH scheme in 1 mixes slowly and is not scalable to high-dimensional $I \times J$ tables for large T_{++}. Instead, a Gibbs sampler can be constructed as detailed in the same proposition (MB-Gibbs).

In doubly constrained tables, η is distributed according to Fisher's noncentral hypergeometric distribution for 2×2 tables. The derivation of this result is provided in the supporting information. The overhead of generating $\mathcal{M}$ for any constrained table is at most $\mathcal{O}(I^2 J^2)$ in both time and space. This overhead can be easily overcome by amortising the construction of $\mathcal{M}$ prior to sampling. The sampling procedure for a constrained model with an intractable table marginal distribution and underlying SIM intensity model is summarised in Algorithm 2. The time complexity of proposing a move in $\mathcal{T}_{\mathcal{C}}$ is $\mathcal{O}(1)$ and $\mathcal{O}(\max\{\max(\mathbf{s}) \mid \mathbf{s} \in \mathcal{C}\})$ for MB-MH and MB-Gibbs, respectively. The corresponding space complexities are both $\mathcal{O}(IJ)$.

Algorithm 2 Metropolis-within-Gibbs Markov basis MCMC sampling algorithm for intractably constrained tables.

```
Inputs: $\mathbf{C}, \mathcal{C}, \mathbf{y}, \mathcal{M}, \mu, N$.
Outputs: $\mathbf{x}^{(1:N)}, \theta^{(1:N)}$, sign $\left(\theta^{(1:N)}\right)$, $\mathbf{T}^{(1:N)}$.
Initialise $\mathbf{x}^{(0)}, \theta^{(0)}, \mathbf{T}^{(0)}$.
for $m \in \{1, \ldots, M\}$ do
    Sample $\mathbf{x}^{(n)} | \theta^{(n-1)}, \mathbf{T}^{(n-1)}$ using Hamiltonian Monte Carlo (Neal, 2011) with acceptance 17.
    Sample $\theta^{(n)} | \mathbf{x}^{(n)}, \mathbf{T}^{(n-1)}$ using Random Walk Metropolis–Hastings with acceptance 18.
    Construct intensity $\mathbf{\Lambda}^{(n)}$ using $\mathbf{x}^{(n)}, \theta^{(n)}$ using 5.
    Sample $l$ uniformly at random from $\{1, \ldots, L\}$.
    Find the valid $\eta$ support yielding $\mathcal{C}_T$-admissible tables.
    Use MB-Gibbs in case 2 of 1 to sample valid $\eta$ with specified $\mu$.
    Obtain $\mathbf{T}^{(n)} = \mathbf{T}^{(n-1)} + \eta \mathbf{f}_l$.
end for
```

The curse of dimensionality prohibits the use of any standard convergence diagnosis techniques, such as the Gelman and Rubin criterion (Gelman & Rubin, 1992). Therefore, we employ the l_1 norm to empirically assess the convergence of sample summary statistics and establish convergence in probability. Furthermore, we assume the underlying intensity function is known a priori, which acts as a ground truth. In the case of Fisher's noncentral hypergeometric distribution, exact moments are not available (McCullagh & Nelder, 2019). These are approximated by the moments of a product Multinomial kernel derived in the Supporting information.

4 | EXPERIMENTAL RESULTS AND DISCUSSION

We showcase table sampling convergence results based on a fixed synthetic intensity across different numbers of origins I, destinations J and agents $M = T_{++}$. Figure 2 depicts empirical convergence rates based on a total of 10^3 chains each run for 10^3 steps. Sparse tables (━━) induce multimodal distributions in $\mathcal{T}_{\mathcal{C}}$ and mix slowly compared with their dense counterparts (━━). Convergence is decelerated more by a larger number of agents rather than higher table dimensionality. The number of agents grows as fast as the diameter of the chain's state space and bounds the number of MCMC steps required to reach the stationary distribution. This observation agrees with the theoretical bounds obtained in Diaconis and Sturmfels (1998), although the latter bounds are derived based on a uniform measure over $\mathcal{T}_{\mathcal{C}}$ explored using MB-MH. Despite this discrepancy, theoretical results provide an upper bound for our case of direct sampling, as evidenced by Figure 3. Direct sampling from the closed-form table posterior achieves the fastest convergence, and we use it to benchmark against MB MCMC. Any doubly constrained table can be explored using either MB-MH (- - -) or MB-Gibbs (- - -). Encoding additional constraints in $\mathcal{T}$ to contract the posterior entails the overhead of using MCMC, introducing a trade-off between convergence rate and distribution contraction in the presence of more summary statistic constraints $\mathcal{C}_{\mathcal{T}}$.

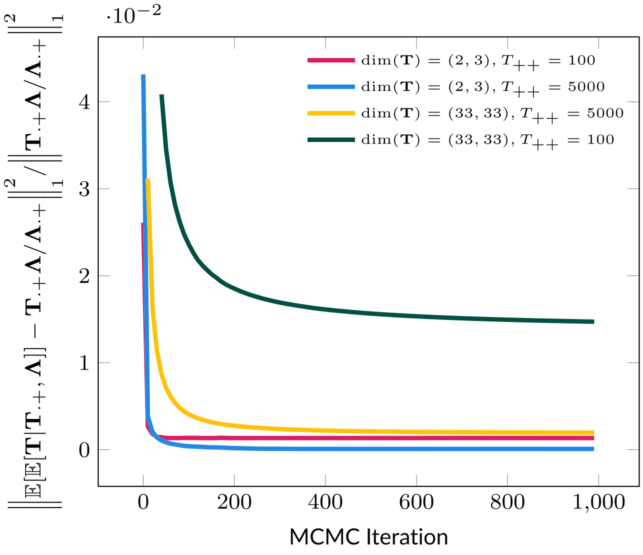

FIGURE 2 l_1 error norm of $\mathbb{E}[\mathbf{T}|\mathbf{y},\mathbf{T}_{\cdot+}]$ across table sizes $\dim(\mathbf{T})$ and number of agents T_{++} using Algorithm 1. Convergence is slower for sparse tables (━━) that induce multimodal distributions. As T_{++} grows (━━,━━) convergence is decelerated by a factor inversely proportional to the table size, which agrees with the theoretical bounds established in Diaconis and Sturmfels (1998).

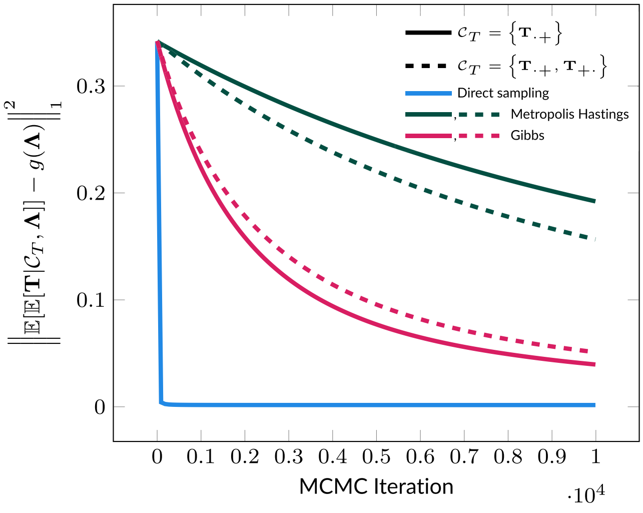

FIGURE 3 l_1 error norm of $\mathbb{E}[\mathbf{T}|\mathbf{y},\mathcal{C}_{\mathcal{T}}]$ for a 33 × 33 table with 5000 agents in the singly (━━,━━,━━) and doubly (- - -, - - -) constrained tables. Markov basis (MB)-Gibbs has a substantially faster convergence rate than MB-Metropolis–Hastings (MH) and mixes reasonably slower compared to direct sampling. Ground truth averages $g(\Lambda)$ are approximate for doubly constrained tables (- - -, - - -).

Furthermore, we present a large-scale application of discrete ODM reconstruction to Cambridge commuting patterns from residence to workplace locations, using the ODM models in Table 1. The precise experimental setup mimics that of (Ellam et al., 2018) and is provided in the supporting information. In light of new summary statistics $\mathcal{C}_T$ (e.g. ■, ⬟, ♦), the table posterior contracts and its high mass region concentrates around the ground truth table (★), as shown in Figure 4. The fact that the low-noise table samples (△, ▨, ⬠, ◇) are nearby their high-noise counterparts (▲, ■, ⬟, ♦) indicates a more dominant effect of the table likelihood on the posterior relative to that of the intensity SDE prior, which enforces the confidence in our reconstructed ODM. The intensity samples of Gaskin et al. (2023) (△, ▲, ▥) have the highest variance among the sampled intensities due to the random initialisations of the Neural ODE solver in Gaskin et al. (2023). Despite this, the intensity distributions in Ellam et al. (2018) and Gaskin et al. (2023) have insufficient $\mathcal{C}_\Lambda$ constraints and a higher divergence from the ground truth table region than table samples. Our intensity samples are also distant from the ground truth table (△, ▨, ⬠, ◇,▲, ■, ⬟, ♦) because they are informed strongly by $\mathcal{C}_\Lambda$ and weakly by $\mathcal{C}_T$ (See Figure 1), where the former set is smaller than the latter.

The ODM validation results summarised in Table 2 affirm that reasoning at the discrete table level accomplishes greater error reductions and enhanced ground truth coverage. Data fitness and posterior prediction errors are computed using the Sorensen similarity index (SSI), standardised root mean square error (SRMSE) and Markov basis distance (MBD). Uncertainty quantification is evaluated based on the coverage probability (CP) of ground truth table cells contained in the 99% highest posterior mass (HPM) region. We elucidate each of these metrics in the supporting information. The best error-coverage trade-off, lowest SRMSE, MBD and highest SSI are attained in the doubly and 20% cell constrained model due to it having the richest constraint set $\mathcal{C}$. Our doubly constrained models account for an SRMSE reduction of 16% relative to the singly constrained model while sustaining an acceptable ground truth coverage equal to approximately 90%. The apparent increase in the mean intensity SRMSE across all doubly constrained models potentially alludes to the SIM's lack of expressivity. This may be because $\mathcal{C}_T$ and $\mathbf{y}$ give rise to conflicting SIM parameter configurations in the limit of large $\mathcal{C}_T$. The MBD decrease in the growth of $\mathcal{C}$ indicates that the expected upper bound on the number of MB moves required to exactly match $\mathbf{T}^{\mathcal{D}}$ is reduced. In the totally and singly constraint models, our table posterior mean matches or outperforms the intensity mean of Ellam et al. (2018) and Gaskin et al. (2023) in terms of data fit (SSI) and SRMSE. The highest ground truth cell coverage probability (94%) is achieved by the most relaxed table, namely, the unconstrained table, but entails a high bias. A lower SRMSE (0.67 instead of 0.85) is attained by the intensity field of the totally constrained model in Gaskin et al. (2023), at the expense of a coverage probability drop from 94% to 89% and a discretisation error accrued for population synthesis.

Our framework's benefits also extend to SIM parameter estimation. In Figure 5, we show that the log destination attraction prediction R^2 increases for larger constraint sets $\mathcal{C}_T$ from 0.77 to 0.84. This allows us to explain the evolved destination employment by informing the data-generating process through $\mathcal{C}$ instead of increasing the diffusivity of the SDE prior in (12). Therefore, we mitigate the identifiability issues of the multimodal $\boldsymbol{\theta}$ posterior emerging in the high noise regime. The $\mathbf{x}$ predictions are further improved in the high noise regime ($R^2 = 0.99$) compared to the low noise counterpart ($R^2 = 0.84$), which favours the hypothesis of a stochastic growth in destination employment. In the high noise regime, unbiassed estimators of $\boldsymbol{\theta}$ are devised based on a more disperse SDE prior on $\boldsymbol{\Lambda}$ 12. Increased prior diffusivity steers the $\mathbf{x}$ posterior marginal towards a larger region of plausible SDE solutions in the vicinity of $\mathbf{y}$, which improves the quality of $\mathbf{x}$ predictions. Additionally, we recover the $\mathbf{x}$ and $\boldsymbol{\theta}$ posterior marginals obtained in Ellam et al. (2018) at a fraction of additional computational cost.

In conclusion, performing population synthesis directly on the discrete high-resolution space of agent attributes bears tangible empirical benefits. These include improved reconstruction and coverage of the ground truth ODM, as well as table posterior contraction in the limit of

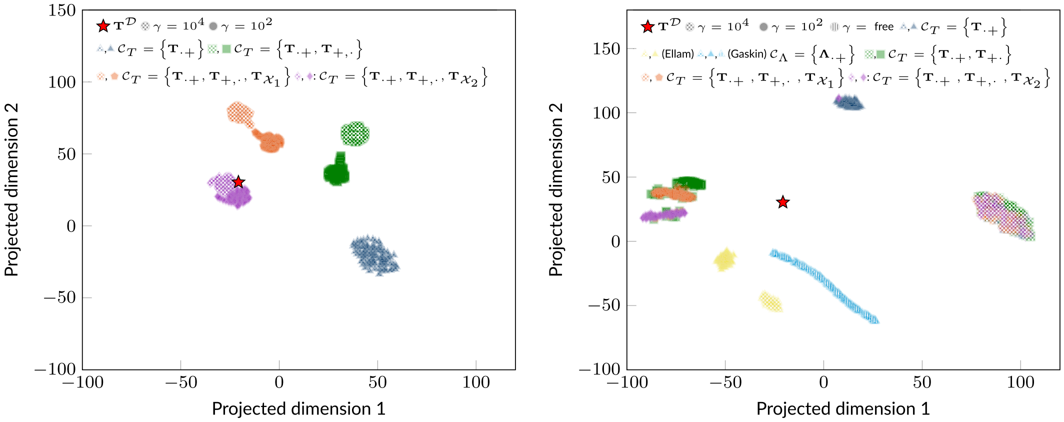

FIGURE 4 Visualisation of the table (left) and intensity (right) samples projected in 2D using T-distributed stochastic neighbour embedding (Hinton & Roweis, 2002). Samples are coloured by the constraint sets in Table 2 for low (▨), high (●) and variable (▥) noise regimes. The ground truth table (★) is better covered by the discrete table posterior regardless of $\mathcal{C}$, and the table distribution becomes increasingly concentrated around the ground truth table in light of more data $\mathcal{C}_T$. Intensity samples are weakly informed through $\mathcal{C}_\Lambda$ and $\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta},\mathcal{C})$ and more distant from the ground truth.

TABLE 2 Ground truth table validation metrics comparing our method against Ellam et al. (2018) and Gaskin et al. (2023) in the continuous intensity and discrete table levels across noise regimes γ and constraint sets $\mathcal{C}$.

Constrained ODM	$\mathcal{C}$	*Method*	γ	*Quantity*	$\lvert\mathcal{M}\rvert$	$\mathbb{E}\left[\mathrm{MBD}\left(\mathbf{T}^{(1:N)},\mathbf{T}^{\mathcal{D}}\right)\right]$	$\mathrm{SSI}\left(\mathbb{E}\left[\mathbf{T}^{(1:N)}\right],\mathbf{T}^{\mathcal{D}}\right)$	$\mathrm{SRMSE}\left(\mathbb{E}\left[\mathbf{T}^{(1:N)}\right],\mathbf{T}^{\mathcal{D}}\right)$	$\mathrm{CP}_{99}(\mathbf{T},\mathbf{T}^{\mathcal{D}})$
Totally				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	**0.72**	0.71	0.22
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	-	9986	0.54	0.95	0.72
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.48	0.83	0.43
	Λ_{++}		10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	-	12,684	0.70	0.85	**0.94***
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	**0.72**	0.73	0.21
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	0	**7385**	0.71	0.73	0.67
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.70	0.70	0.41
	T_{++}, Λ_{++}	This work	10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	0	7600	0.69	0.70	0.68
			10^4		-	-	**0.72**	0.74	0.21
	Λ_{++}	(Ellam)	10^2	$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.70	0.70	0.42
			10^4		-	-	0.70	**0.67**	0.79
			10^2		-	-	0.65	0.68	0.89
	Λ_{++}	(Gaskin)	Learned	$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.64	0.70	0.88
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.72	0.74	0.23
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	0	**6806**	**0.72**	0.69	0.68
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.70	0.71	0.40
	$\mathbf{T}_{\cdot+}, \Lambda_{++}$		10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	0	7067	0.71	0.63	0.71
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	**0.74**	0.69	0.23
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	0	6819	0.72	0.69	0.68
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.72	0.62	0.40
	$\mathbf{T}_{\cdot+}, \Lambda_{++}, \boldsymbol{\Lambda}_{+\cdot}$	This work	10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	0	6944	0.71	0.62	0.71
			10^4		-	-	**0.74**	0.69	0.23
	$\boldsymbol{\Lambda}_{+\cdot}$	(Ellam)	10^2	$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.72	0.62	0.42
			10^4		-	-	0.71	**0.61**	0.89
			10^2		-	-	0.66	0.62	**0.92**
Singly	$\boldsymbol{\Lambda}_{+\cdot}$	(Gaskin)	Learned	$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.65	0.65	0.86
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.73	0.71	0.21
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	182,988	**5912**	**0.76**	**0.59**	0.69
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.61	1.12	0.20

(Continues)

TABLE 2 (Continued)

Constrained ODM	$\mathcal{C}$	*Method*	γ	*Quantity*	$\lvert\mathcal{M}\rvert$	$\mathbb{E}\left[\text{MBD}\left(\mathbf{T}^{(1:N)},\mathbf{T}^{\mathcal{D}}\right)\right]$	$\text{SSI}\left(\mathbb{E}\left[\mathbf{T}^{(1:N)}\right],\mathbf{T}^{\mathcal{D}}\right)$	$\text{SRMSE}\left(\mathbb{E}\left[\mathbf{T}^{(1:N)}\right],\mathbf{T}^{\mathcal{D}}\right)$	$\text{CP}_{99}\left(\mathbf{T},\mathbf{T}^{\mathcal{D}}\right)$
Doubly	$\mathbf{T}_{\cdot+},\mathbf{T}_{+\cdot},\Lambda_{++}$	This work	10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	182,988	6370	0.73	**0.59**	**0.86**
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.72	0.71	0.22
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	119,421	**5314**	**0.78**	**0.55**	**0.89**
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.63	1.10	0.32
Doubly and 10% cell	$\mathbf{T}_{\cdot+},\mathbf{T}_{+\cdot},\mathbf{T}_{\mathcal{X}_1},\Lambda_{++}$	This work	10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	119,421	5449	0.77	0.56	0.88
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.72	0.71	0.26
			10^4	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	74,314	**4521***	**0.81***	**0.51***	0.89
				$\Lambda\lvert\mathcal{C},\mathbf{y}$	-	-	0.64	1.05	0.32
Doubly and 20% cell	$\mathbf{T}_{\cdot+},\mathbf{T}_{+\cdot},\mathbf{T}_{\mathcal{X}_2},\Lambda_{++}$	This work	10^2	$\mathbf{T}\lvert\mathcal{C},\mathbf{y}$	74,314	4543	0.80	**0.51***	**0.90**

Note: Our method achieves the best error-coverage trade-off in the doubly and 20% cell constrained origin-destination matrix (ODM) as well as the best reconstruction errors (Markov basis distance, MBD; standardised root mean square error, SRMSE), data fit (Sorensen similarity index, SSI) and ground truth coverage (coverage probability, CP) across all ODM models, as indicated by *. As the size of $\mathcal{C}$ increases, the ground truth table is better reconstructed (reduced SRMSE, increased SSI) and covered by the table posterior. In the cases of totally and singly constrained ODMs, we attain a similar SSI and coverage to Gaskin et al. (2023). We note that the last three ODM models cannot be handled by either Ellam et al. (2018) or Gaskin et al. (2023). Bold numbers correspond to the best metric value achieved by C (second column) and asterisks denote the best metric value across the entire table.

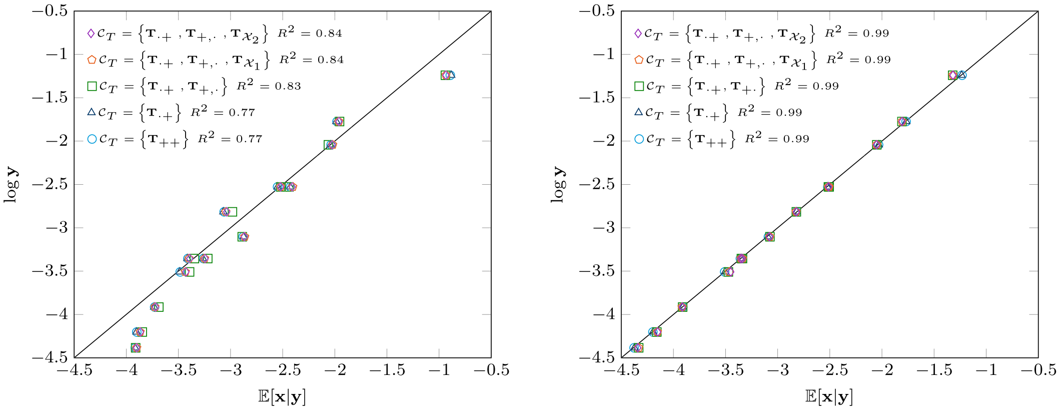

FIGURE 5 Posterior predictions of $\mathbb{E}[\mathbf{x}|\mathbf{y},\mathcal{C}]$ against observed log employment data $\mathbf{y}$ using Algorithms 1 and 2. The high noise regime (right) achieves a more plausible data fit than its low noise counterpart (left) due to improved uncertainty quantification. This is attributed to the unbiassed estimation of the normalising constant (13) obtained in the former case compared to biassed estimation and subsequent collapse of the $\boldsymbol{\theta}$ posterior to a Dirac mass in the latter case. Enriching $\mathcal{C}_T$ from ○,△ to □, ⬠, ◇ increases R^2, illustrating the added benefits of inference on a higher resolution table level.

constraint data $\mathcal{C}_T$. If population synthesis is not of interest, SIM parameters can be adequately estimated using competitive approaches such as Gaskin et al. (2023). Combining such optimisation methods with MB MCMC in a naive Bayes scheme can be promising, as it exploits the advantages of both optimisation and MCMC techniques. Regardless, the apparent shortcomings of SIMs call for a comparative study of various intensity model classes, such as discrete choice models (Train, 2009). Finally, the multifaceted nature of population synthesis opens up future avenues of research beyond ODM reconstruction, where more convoluted dependency structures can be exploited.

AUTHOR CONTRIBUTIONS

Ioannis Zachos and Theodoros Damoulas conceived of the presented idea. Ioannis Zachos developed the methodology, and performed the computations under the supervision of Theodoros Damoulas and Mark Girolami. Mark Girolami also provided the funding support. Ioannis Zachos wrote the paper with input from all co-authors.

ACKNOWLEDGEMENTS

I. Z. acknowledges support from the UK Research and Innovation (UKRI) Research Council and Arup Group via the PhD studentship. T. D. acknowledges support from the UKRI Turing AI Acceleration Fellowship EP/V02678X/1. M. G. acknowledges support from the Alan Turing Institute, the Royal Academy of Engineering and UKRI grants EP/T000414/1, EP/R018413/2, EP/P020720/2, EP/R034710/1 and EP/R004889/1.

DATA AVAILABILITY STATEMENT

Trip and employment data for Cambridge, UK, are obtained from the Office for National Statistics (2014,2015). Individual home and work facility locations are extracted from Geofabrik (2023). Our codebase has been released online (https://github.com/YannisZa/ticodm).

ORCID

Ioannis Zachos https://orcid.org/0000-0002-7503-2117
Theodoros Damoulas https://orcid.org/0000-0002-7172-4829
Mark Girolami https://orcid.org/0000-0003-3008-253X

ENDNOTES

[1] Summary statistics are arranged in increasing order of cell set sizes $|\mathcal{X}_k|$.

[2] Summary statistics are arranged in increasing order of cell set sizes $|\mathcal{X}_k|$.

[3] We denote probability distributions over discrete, continuous and mixed discrete-continuous supports by $\mathbb{P}, p, P$, respectively.

REFERENCES

Agresti, A. (2002). *Categorical data analysis*, Wiley Series in Probability and Statistics: John Wiley & Sons, Inc.http://doi.wiley.com/10.1002/0471249688

Axhausen, K. W., & Zürich, E. T. H. (2016). *The multi-agent transport simulation MATSim Edited by* Zürich, E. T. H., Horni, A., & Nagel, K. TU Berlin: Ubiquity Press. http://www.ubiquitypress.com/site/books/10.5334/baw/

Belbachir, H. (2014). A combinatorial contribution to the multinomial Chu-Vandermonde convolution. *Les Annales RECITS*, *1*, 27–32.

Berge, C. (1971). *Principles of combinatorics*: Elsevier Science.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis*: Springer.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*, 7280–7287. https://pnas.org/doi/full/10.1073/pnas.082080899

Carvalho, L. (2014). A Bayesian statistical approach for inference on static origin–Destination matrices in transportation studies. *Technometrics*, *56*(2), 225–237. http://www.tandfonline.com/doi/abs/10.1080/00401706.2013.826144

Croci, M., Fasi, M., Higham, N. J., Mary, T., & Mikaitis, M. (2022). Stochastic rounding: Implementation, error analysis and applications. *Royal Society Open Science*, *9*(3), 211631. https://royalsocietypublishing.org/doi/10.1098/rsos.211631

Dearden, J., & Wilson, A. G. (2015). *Explorations in urban and regional dynamics: A case study in complexity science*, Routledge Advances in Regional Economics, Science and Policy: Routledge, Taylor & Francis.

DeSalvo, S., & Zhao, J. Y. (2016). Random sampling of contingency tables via probabilistic divide-and-conquer. ArXiv e-prints. Retrieved from http://arxiv.org/abs/1507.00070

Diaconis, P., & Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, *26*(1), 363–397. https://projecteuclid.org/journals/annals-of-statistics/volume-26/issue-1/Algebraic-algorithms-for-sampling-from-conditional-distributions/10.1214/aos/1030563990.full

Ellam, L., Girolami, M., Pavliotis, G. A., & Wilson, A. (2018). Stochastic modelling of urban structure. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *474*(2213), 20170700. https://royalsocietypublishing.org/doi/10.1098/rspa.2017.0700

Farooq, B., Bierlaire, M., Hurtubia, R., & Flötteröd, G. (2013). Simulation based population synthesis. *Transportation Research Part B: Methodological*, *58*, 243–263. https://www.sciencedirect.com/science/article/pii/S0191261513001720

Fournier, N., Christofa, E., Akkinepally, A. P., & Azevedo, C. L. (2021). Integrated population synthesis and workplace assignment using an efficient optimization-based person-household matching method. *Transportation*, *48*(2), 1061–1087. https://doi.org/10.1007/s11116-020-10090-3

Gaskin, T., Pavliotis, G. A., & Girolami, M. (2023). Neural parameter calibration for large-scale multiagent models. *Proceedings of the National Academy of Sciences*, *120*(7), e2216415120. https://pnas.org/doi/10.1073/pnas.2216415120

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Geofabrik (2023). Great britain openstreetmap data. https://download.geofabrik.de/europe/great-britain.html

Harris, B., & Wilson, A. G. (1978). Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environment and Planning A: Economy and Space*, *10*(4), 371–388. http://journals.sagepub.com/doi/10.1068/a100371

Hinton, G. E., & Roweis, S. (2002). Stochastic neighbor embedding, *Advances in neural information processing systems*, Vol. 15: MIT Press. https://papers.nips.cc/paper_files/paper/2002/hash/6150ccc6069bea6b5716254057a194ef-Abstract.html

McCullagh, P., & Nelder, J. A. (2019). *Generalized linear models* (2nd ed.): Routledge. https://www.taylorfrancis.com/books/9781351445856

Neal, R. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (Eds.), *Handbook of Markov chain Monte Carlo*, Vol. 20116022: Chapman and Hall/CRC. http://www.crcnetbase.com/doi/abs/10.1201/b10905-6

Office for National Statistics (2014). Employment status (workplace population) data. https://www.nomisweb.co.uk/census/2011/wp601ew

Office for National Statistics (2015). Location of usual residence and place of work (OA level) data. https://www.nomisweb.co.uk/census/2011/wf01bew

Pooler, J. (1994). An extended family of spatial interaction models. *Progress in Human Geography*, *18*(1), 17–39.

Sun, L., & Axhausen, K. W. (2016). Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, *91*, 511–524. https://www.sciencedirect.com/science/article/pii/S0191261516300261

Train, K. (2009). *Discrete choice methods with simulation* (2nd ed.): Cambridge University Press.

Voas, D., & Williamson, P. (2000). An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata. *International Journal of Population Geography*, *6*(5), 349–366. https://doi.org/10.1002/1099-1220(200009/10)6:5<349::AID-IJPG196>3.0.CO;2-5

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space*, *3*(1), 1–32. http://journals.sagepub.com/doi/10.1068/a030001

AUTHOR BIOGRAPHIES

Ioannis Zachos (I. Z.) is a PhD student in the Interdisciplinary Centre for Doctoral Training in Future Infrastructure and Built Environment at Cambridge University.

Theodoros Damoulas (T. D.) is a Professor in Machine Learning and Data Science with a joint appointment in Computer Science and Statistics at Warwick University and a Turing Artificial Intelligence Fellow.

Mark Girolami (M.G.) is the Chief Scientist of the Alan Turing Institute, the Sir Kirby Laing Professor of Civil Engineering at Cambridge University and the Royal Academy of Engineering Research Chair in Data-Centric Engineering.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zachos, I., Damoulas, T., & Girolami, M. (2024). Table inference for combinatorial origin-destination choices in agent-based population synthesis. *Stat*, *13*(1), e656. https://doi.org/10.1002/sta4.656

APPENDIX A: AN EXTENSION TO CHU VANDERMONDE'S THEOREM FOR MULTINOMIAL COEFFICIENTS

Theorem 1. Let $\mathcal{T}_{\mathcal{C}}$ be the space of admissible tables satisfying $\mathcal{C}_T = \{\mathbf{T}_{+\cdot}, \mathbf{T}_{\cdot+}\}$ with fixed odds ratios $\boldsymbol{\omega} \in \mathbb{R}_{\geq 0}$. Denote the subsets of $\mathcal{C}_T$-admissible tables with $\mathcal{C}_T = \{T_{++}\}$, $\mathcal{C}_T = \{\mathbf{T}_{+\cdot}\}$ and $\mathcal{C}_T = \{\mathbf{T}_{\cdot+}\}$ by $\mathcal{T}_{++}$, $\mathcal{T}_{+\cdot}$ and $\mathcal{T}_{\cdot+}$, respectively. Then, for any $\mathcal{C}_T$-admissible $I \times J$ table $\mathbf{T}$ the following statement holds:

$$\binom{T_{++}}{T_{+1}\dots T_{+J}} \prod_{j}^{J} \omega_{+j}^{T_{+j}} = \sum_{\mathbf{T} \in \mathcal{T}_{+\cdot}} \prod_{j}^{J} \binom{T_{+j}}{T_{1j}\dots T_{Ij}} \prod_{i,j}^{I,J} \omega_{ij}^{T_{ij}}.$$

This is an extension of the Chu–Vandermonde theorem for multinomial coefficients (Belbachir, 2014) to polynomials with multinomial coefficients.

Proof. We proceed with an algebraic proof. Let $[x]_{j=1}^{J} = 1 + x + \dots + x^J$ be a polynomial of order J. By writing the left-hand side in polynomial form and expanding it, we get

$$\begin{aligned}
\sum_{\mathbf{T}_{+\cdot} \in \mathcal{T}_{++}} \underbrace{\left(\binom{T_{++}}{T_{+1}\dots T_{+J}} \prod_{j}^{J} \omega_{+j}^{T_{+j}}\right)}_{\text{LHS}} \prod_{m}^{J} x_m^{T_{+m}} &= \sum_{\mathbf{T}_{+\cdot} \in \mathcal{T}_{++}} \binom{T_{++}}{T_{+1}\dots T_{+J}} \prod_{j}^{J} (\omega_{+j} x_j)^{T_{+j}} \\
&= \left(\sum_{j}^{J} \omega_{+j} x_j\right)^{T_{++}} \\
&= \prod_{j}^{J} \left(\sum_{j}^{J} \omega_{+j} x_j\right)^{T_{+j}} \\
&= \prod_{j}^{J} \sum_{\mathbf{T} \in \mathcal{T}_{+j}} \binom{T_{+j}}{T_{1j}\dots T_{Ij}} \prod_{i}^{I} (\omega_{ij} x_j)^{T_{ij}} \\
&= \prod_{j}^{J} \left(\sum_{\mathbf{T} \in \mathcal{T}_{+j}} \binom{T_{+j}}{T_{1j}\dots T_{Ij}} \prod_{i}^{I} \omega_{ij}^{T_{ij}}\right) x_j^{T_{+j}} \\
&= \sum_{\mathbf{T}_{+\cdot} \in \mathcal{T}_{++}} \underbrace{\left(\sum_{\mathbf{T} \in \mathcal{T}_{+\cdot}} \prod_{j}^{J} \binom{T_{+j}}{T_{1j}\dots T_{Ij}} \prod_{i}^{I} \omega_{ij}^{T_{ij}}\right)}_{\text{RHS}} \prod_{j}^{J} x_j^{T_{+j}},
\end{aligned}$$

where the exchange of product and sum in the last line is permitted due to the grouping of terms in the sum by column $\mathbf{T}_{\cdot j}$ $\forall j = 1, \dots, J$. The second and fourth equalities follow by direct application of the multinomial theorem (Berge, 1971). This completes the proof. □

The theorem above allows us to compute the normalising constant of Fisher's noncentral hypergeometric distribution in (See Lemma 3 in the in the supporting information). We note that summing over the support $\mathcal{T}_{+\cdot}$ yields

$$\begin{aligned}
\frac{\prod_i^I T_{i+}!}{T_{++}!} \sum_{\mathbf{T} \in \mathcal{T}_{+\cdot}} \prod_{i,j}^{I,J} \frac{T_{+j}!}{T_{ij}!} \omega_{ij}^{T_{ij}} &= \frac{\prod_i^I T_{i+}!}{T_{++}!} \prod_{j}^{J} \frac{T_{++}!}{T_{+j}!} \omega_{+j}^{T_{+j}} \\
&= \prod_{i,j}^{I,J} \frac{T_{i+}!}{T_{+j}!} \omega_{+j}^{T_{+j}}.
\end{aligned}$$

Subsequently, the normalised Fisher's noncentral hypergeometric distribution is equal to

$$\prod_{j}^{J}\frac{T_{+j}!T_{+j}!}{T_{++}!T_{ij}!}\prod_{i}^{I}\left(\frac{\omega_{ij}}{\omega_{+j}}\right)^{T_{ij}}=\prod_{j}^{J}\frac{\binom{T_{+j}}{T_{1j},\dots,T_{Ij}}}{\binom{T_{++}}{T_{+1},\dots,T_{+J}}}\prod_{i}^{I}\left(\frac{\omega_{ij}}{\omega_{+j}}\right)^{T_{ij}}, \tag{A1}$$

which is similar to the kernel of a product multinomial with $\mathbf{T}_{+\cdot}$ number of trials, $\frac{\omega}{\omega_{+\cdot}}$ event probabilities. We adopt this approximation for computing ground truth moments in our synthetic doubly constrained model experiments of Figure 3.

APPENDIX B: AUXILIARY RESULTS FOR THE CAMBRIDGE APPLICATION

In this section, we append additional experimental results of the large-scale application to Cambridge commuting patterns. Figure B1 compares the smallest ODM reconstruction errors from each method in Table 1. The doubly and 20% cell constrained table significantly improves ground truth coverage of free cells (row 2) over the singly constrained intensity of Ellam et al. (2018) (row 4) even though both use the same cost matrix. Therefore, augmenting $\mathcal{C}_T$ shrinks the support of free cells and diminishes their reconstruction error. Compared with the singly constrained intensity of Gaskin et al. (2023), the table posterior mean achieves a reduced SRMSE (0.51 versus 0.61) and does not overestimate the trips to popular destinations. Such destinations include 7 and 13, which correspond to areas where the city's university premises and hospital are located and are highly attractive. The neural network's low noise predictions (row 3) are characterised by stronger attraction effects, which translates to overestimation of trips to the aforementioned destinations. The higher noise regime has little effect on the neural network's trip prediction error as opposed to the table posterior's error, which is substantially reduced in the high noise regime. This suggests that continuous intensity is more susceptible to ODM overfitting than the discrete table. A diffuse SDE smooths the marginal distributions of free table cells without changing their support and allows for more contributions of the $\mathbb{P}(\mathbf{T}|\mathbf{x},\boldsymbol{\theta},\mathcal{C})$ term to the posterior mean.

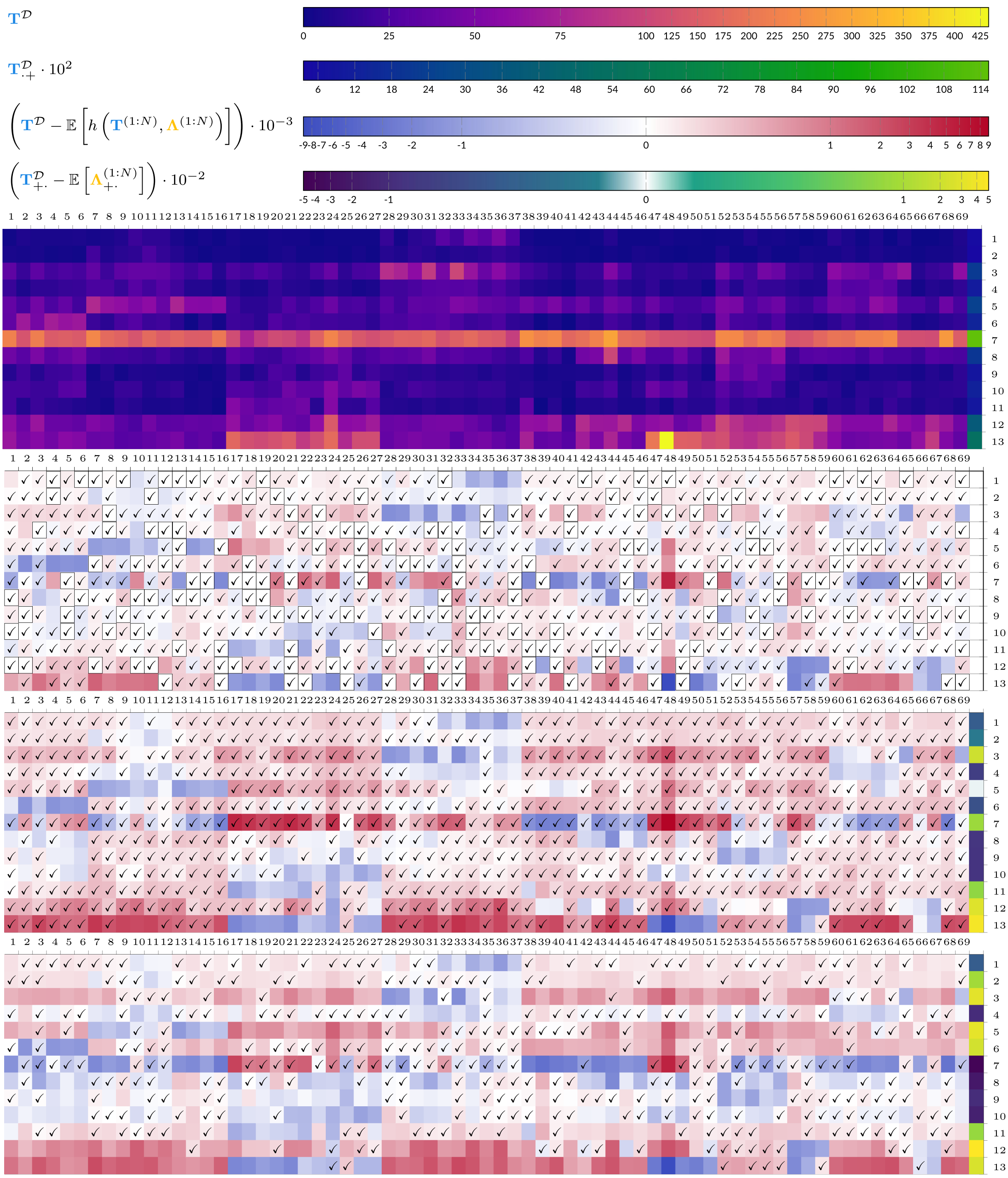

FIGURE B1 Ground truth table $\mathbf{T}^{\mathcal{D}}$ (row 1), $\mathbf{T}^{\mathcal{D}} - \mathbb{E}[\mathbf{T}|\mathbf{y}, \{\mathbf{T}_{\cdot+}, \mathbf{T}_{+,\cdot}, \mathbf{T}_{\mathcal{X}_2}, \Lambda_{++}\}]$ (row 2) and $\mathbf{T}^{\mathcal{D}} - \mathbb{E}[\boldsymbol{\Lambda}|\mathbf{y}, \{\boldsymbol{\Lambda}_{\cdot+}\}]$ (rows 3 & 4) using Algorithm 2 (row 2), the low noise Neural network in (Gaskin et al., 2023) (row 3) and the high noise MCMC scheme in (Ellam et al., 2018) (row 4). Each table's rows resemble the number of trips by destination, and vice versa. Cells with a black boundary correspond to fixed cells while ✓ cells cover the ground truth in the 99% HPM. The high noise table predictions interpolate missing cell data in the vicinity of fixed cells and produce a more accurate trip ODM. The error profile similarities across all methods are attributed to the use of the same cost matrix, which captures the spatial covariance of trips.