Integration of a Capacity-Constrained Workplace Choice Model

Recent Developments and Applications with an Agent-Based Simulation in Singapore

Basil J. Vitins, Alexander Erath, and Kay W. Axhausen

Destination choice models can be embedded in transport and land use models to understand travel and location choice behavior and to forecast scenarios. Utility-maximizing destination choice models can account for individual behavior and make them suitable for agent-based models, while processing destination capacities is also in line with agent-based modeling. This paper addresses the possibility and impact of introducing capacity constraints, their effect on choice behavior, and the feasibility of applying an approach like this in agent-based microsimulations with individual characteristics for each agent. Here, a comprehensive workplace choice model and its application in a large-scale simulation case study for Singapore are described; one technical and one methodological achievement are highlighted. Technical achievement benefits from recent computational advances; the workplace choice model is estimated with a comprehensive utility function on a large data set with 10^3 destinations. Reasonable model fit and robust parameters are achieved while obviating sampling techniques; resulting parameters are efficiently applied to the entire 5.4 million Singapore population and validated with survey data. For methodological innovation, capacity limitations are introduced at workplaces to avoid oversaturation. A robust optimization method based on shadow prices is proposed to accommodate capacity limitations at all workplaces during the choice model application defined above. The proposed method efficiently assigns commuters to unused workplaces while respecting individual commuter preferences. Validation of the simulation results, by comparing travel time distributions for commuting trips reported in travel diary data, shows that the model fits well with observed data.

Destination choice is a central and challenging problem in transportation and land use modeling. It can be characterized by the discrete choice of a specific person between different destination alternatives in a given situation; a person living or staying at a certain origin has to choose between a set of discrete destinations—geographically distributed in space—before conducting a trip. Specific destinations are assigned to all individuals considered at destination choice models, for example, for shopping, for commuting trips, or for determining home locations for household distributions. Assignment differs depending on personal characteristics, trip purpose, set of alternatives, and external environmental influences. In transportation, destination choice has often been applied in the traditional, aggregated four-step model [see, e.g., Ortúzar and Willumsen (*1*)]; however, disaggregated models also rely on destination choice methods, as well as various land use models and location assessments [see, e.g., McFadden (*2*) and Ben-Akiva and Lerman (*3*)]. Unlike other discrete choice situations, such as transport mode choice, destination choice has to deal with a very large number of choice alternatives. Multiple approaches exist to assign agents to specific destinations differing in data and computational requirements. The gravity approach basically mirrors Newton's physical gravitation law and serves as a robust solution for trip-based destination choice when given a certain trip impedance function, for example, for commuting trips (*1*). This approach spreads a given demand over all destination zones according to their attractiveness and travel impedance and has been implemented in numerous small- and large-scale real-world applications because of its robustness and low computational requirements. Destination choice can also be modeled with utility-maximizing approaches, which determine a probability for every traveling individual and destination according to destination quality, personal taste, and situation [see, e.g., Daly (*4*)]. This approach is especially suitable for agent-based frameworks able to capture individual behavior heterogeneity. Utility-maximizing approaches also have a long tradition in choice modeling; a major advantage lies in their underlying sensitivity to individual characteristics and travel behavior, for example, employment type or income. Both individual attributes can be used as independent variables when individuals are assigned to workplaces in a specific economic sector. Utility-maximizing approaches are, thus, more demanding about data and computational requirements in regard to destination choice. See Mishra et al. for a detailed comparison of gravity- and utility-based approaches (*5*).

One major difference between gravity- and utility-maximizing approaches is the consideration of capacity constraints at destinations. In demand modeling, alternatives' given capacity constraints are essential whenever market supply is limited for individuals in a choice situation. Capacity constraints should be considered in destination choice, as workplace locations have an upper limit for the number of workers accommodated. While capacity constraints are fully integrated in the doubly constrained, gravity-based approach and oversaturation is suppressed, capacity constraints are only partially considered in utility-maximizing models, such as the multinomial logit (MNL) model (*1*). Utility-maximizing models implement

B. J. Vitins and A. Erath, Future Cities Laboratory, 1 CREATE Way, 06-01 CREATE Tower, Singapore 138602. K. W. Axhausen, Institute for Transport Planning and Systems–IVT, ETH Zurich, CH-8093 Zurich, Switzerland. Corresponding author: B. J. Vitins, basil.vitins@ivt.baug.ethz.ch.

Transportation Research Record: Journal of the Transportation Research Board, No. 2564, Transportation Research Board, Washington, D.C., 2016, pp. 1–13.
DOI: 10.3141/2564-01

capacity limits to a certain extent with the utility function formula. However, this function consists of additional terms beyond capacity-related variables, and the parameter fitting does not exclusively consider the destination saturation, per se. Choice models allow for oversaturated destinations in their application. This paper proposes an efficient method to apply constraints in utility-based model applications.

Of the various destination choice models, the focus here is on workplace choice as part of the overall modeling process for a Singapore transport model. Commuting trips are very important because of their large share during peak hours; they often drive long-term investment in transportation infrastructure and spatial development.

Computational performance has improved steadily during the past few years, and recent advances allow parameter estimation for choice models on large (full) sample sets with sizable numbers of destination alternatives and complex utility functions. In addition, they allow model applications on populations as large as ~1.9 million commuters, such as the present example in Singapore.

The aim of this paper is twofold. First, the workplace choice model should account for capacity-constrained destinations. Second, recent computational and technical advances should be exploited to obtain robust parameters and a reasonable model fit and apply parameters in a large-scale transportation simulation. The main research issues are how to (*a*) introduce capacity constraints in a location choice application and (*b*) demonstrate the feasibility of applying such an approach in a large-scale application while accounting for destination choice heterogeneity.

These research questions are motivated by existing large, rich data sources, as well as increasingly limited land resources in growing cities. Singapore city recently conducted an extensive household travel diary survey containing relevant variables (*6*). In addition, detailed transport network graphs and travel times were available from previous work, along with information about population and workplace densities. These empirical data also allow comparison with modeling results. Singapore is a fast-growing city–state with a large population and significant workplace densities, generating a strong interest in efficient planning and spatial land use competition. It is therefore assumed that upper capacity constraints exist at certain travel destinations and that they should be included in a given transport model. Furthermore, Singapore is only one example of a city with scarce (land and other) resources; many other urban areas are experiencing similar developments, making the Singapore case even more interesting and topical.

DESTINATION CHOICE MODELING

This section highlights selected past achievements on (*a*) the methodological side and (*b*) the data and sampling side. In regard to the methodological side, many papers have explored different discrete choice methods; for example, Vovsha et al. applied an MNL model with an encompassing utility function containing diverse descriptive variables (*7*). An MNL model can be computationally efficiently estimated for destination choice purposes, thus accommodating a large number of alternatives and complex utility functions. *Modelling the Choice of Residential Location* was a seminal paper on residential choice with several model formulations (*2*); other studies applied complex choice methods and destination choice models to specifically account for unobserved similarities between destination alternatives. Bekhor and Prashker summarized and compared generalized extreme values models with relaxed independent and identically distributed assumptions with the MNL model to capture some unobserved similarities between the alternatives: the nested logit, cross-nested logit, generalized nested logit, spatial correlated logit model, and a combined model (*8*). All of these models have specific pros and cons for model characteristics, complexity, and computational burdens. As far as destination choice, the spatial correlated logit model deserves special attention because of its consideration of adjacent zone pairs, possibly spatially correlated (*9*). One could also add the generalized spatially correlated logit model, which considers spatial distances between all alternatives, compared with just adjacent alternatives in the spatial correlated logit model (*10*). Along with the model theory, solution algorithms are crucial for model applications; they depend strongly on model formulations. A closed-form model formulation can be solved with direct maximum likelihood techniques, compared with other formulations that require more challenging numerical- and simulation-based approaches [e.g., Koppelman and Sethi (*11*)].

In regard to the data and sampling side, sampling techniques in destination choice models have been widely discussed because of their large choice sets. McFadden provided techniques for sampling (*2*). Frejinger et al. (*12*) and Neralla and Bhat (*13*), as well as others, proposed sampling methods and found that a large number of observations are needed to achieve reasonable model parameter values. Neralla and Bhat suggested drawing ⅛ of the full choice set size as a minimum and ¼ as a desirable sample share in the case of their MNL models, stating that non-MNL models are even more demanding in regard to the required sample size (*13*). Recently, a trend has begun to calibrate model parameter for destination choice models with the entire set of alternatives and to avoid sampling techniques of the destination alternatives. This approach has the advantage that sampling techniques are avoided; even complex models can be applied without worrying about sampling strategy. Parameter calibration without data sampling is done, for example, in school choice models in which pupils select between different schools (*14*); the choice set here is the potential school set in a given area. In regard to workplaces, the potential workplace set is considerably larger; these are then aggregated into zones to achieve reasonable calculation times, as calculation time linearly (approximately) increases with the number of alternatives (*3*).

CAPACITY CONSTRAINTS AND SHADOW PRICES

Assuming a utility-based workplace choice model accounting for generalized travel costs, destination quality, and personal and situational variables, it can be argued that workplace choice has to incorporate market competition for more centrally located workplaces, for example. However, competition and market equilibrium for attractive workplaces and low rents take place on a firm level with a different utility function. Therefore, one cannot, a priori, assume a market clearing situation for workplace location choice, as might be the case in housing and residential location [see, e.g., McFadden (*2*) and de Palma et al. (*15*)]. Certain issues might lead to over- and undersaturated workplace locations after a utility-based destination choice model is applied.

1. Destination choice models obviously depend heavily on the survey data. Whenever survey data are stratified or biased on spatial attributes, it can be assumed that certain (commuting) behaviors are not well captured for a given stratum in the model. It is possible to overcome this problem by estimating parameters for the underrepresented strata, but a certain parameter bias can be expected to

persist after parameter estimation, meaning a bias in destination choice model applications.

2. Recent, very elaborate discrete choice models can deal with complex choice situations, for example, spatial correlation (see also section on modeling destination choice). However, it is still possible that not all correlations are actually captured in the model, making systematic errors possible.

3. The perception of generalized travel costs might not be proportional to actual costs; for example, Vovsha et al. included a differentiated distance function composed of six terms (*7*). One can assume that the perception could be approximated with complex functions, but residual errors might still remain and bias modeling outcome.

4. Even if a perfectly fitted model is assumed on the basis of a representative data set, parameter values might have to be changed in scenario applications. For scenarios with changing destination capacities, commuters might not be completely reassigned to all changing destinations because of the influence of the generalized travel costs (an alternative specific variable but one that completely ignores capacity).

5. Assuming an imperfect model for one of the above reasons, even a small oversaturation can lead to errors when results are evaluated and interpreted, for example, for business developments in over- and undersaturated destinations, as well as related traffic flows.

Shadow prices are proposed to account for many of these issues in destination model applications. Shadow prices are applied differently in various studies. In linear programming, they derive from the dual solution, while in economics, they are used to estimate unknown costs of a certain good or alternative. In the following, shadow prices serve as an additional impedance for attractive, but limited, capacity alternatives. Shadow prices should reflect alternatives' constraints, as explained in the following example based on microeconomics and productivity optimization. A price $p > 0$ can be assumed, as well as a stock of X with sold units x, where $0 \leq x \leq X$ and the objective $\max(px)$. Customers buying units x optimize their utility u, with respect to their time and budget limits, giving one (at least) two dependent optimization problems. Any resource is considered a constraint (e.g., time and units) if the number that customers would like to use exceeds the number available. In an inefficient market, it might be possible that demand exceeds supply because of distributional effects; then, a shadow price must be implemented to allow solution of the optimization problem (*16*). This idea of shadow prices is transferred and adjusted for destination choice situations.

In destination choice, shadow prices can be assigned to designated destinations and regarded as additional impedance for people choosing those destinations. Shadow prices can thus account for these alternatives' constraints, which cannot be captured with model parameters, or the error term (ε) choice model distributions. Higher prices also indicate scarcity, which means that shadow prices also uncover future spatial development potential of underdeveloped, or even missing (but potentially valuable), alternatives. The current literature provides little discussion of shadow prices and their effects in choice models. Shadow prices are applied in some studies, but a detailed method and references are missing, as well as definition and price effects. Gupta et al. (*17*) used shadow prices for park-and-ride lots to identify the best parking lot, as did Davidson et al. (*18*). Hammadou and Papaix applied shadow prices in a mode choice model for carbon dioxide pricing (*19*). De Palma et al. modeled the housing market and provided a detailed definition of shadow prices (*15*). They stated that supply and prices in the housing market might clear the demand, depending on the specific situation; they presented algorithms for constant and variable demand in this context (*15*). Workplace models described in this paper are different [compared with, e.g., de Palma et al. (*15*)]; office rents, in particular, are not fully mirrored in employees' utilities. Therefore, an employee's utility is not directly affected when choosing a workplace, compared with an otherwise identical workplace with lower rents.

Some researchers focused on joint travel and residential costs and formulated mathematical programs. Beckmann and Wallace introduced shadow prices, similar to the proposed approach, for welfare maximization of home location changes with infrastructure changes, including housing rents (*20*). Los modified the utility function, transportation, and residential costs to improve residential choice and the usage and impact of transportation (*21*). Earlier issues discussed remain mostly open according to the literature review; more detailed definition and knowledge about shadow price effects will be relevant for future applications.

DATA AVAILABILITY IN SINGAPORE CASE STUDY

In this study, a destination choice model was estimated according to a revealed preference household travel survey conducted for the entire island of Singapore, a city–state located in Southeast Asia, encompassing an island 43 km long and 23 km wide, with a permanent 2014 resident population of 3.8 million and a total population of about 5.4 million. In gross domestic product per capita, Singapore is one of the wealthiest countries in Asia, but specific regulation directly controlling the number of registered vehicles, as well as high taxes, keeps the car population in Singapore comparably low; 35.1% of all households have one or more privately registered cars available, and 48% have some type of motorized vehicle available.

Household Interview Travel Survey

The Household Interview Travel Survey is conducted every 4 to 5 years (*6*). The last was conducted between June 2012 and May 2013 (further referred to as "HITS 2012") by the Land Transport Authority of Singapore; Singaporean citizens, permanent residents, and legal immigrants residing there were included in the survey. Data of 9,635 households, 35,714 people, 70,984 trips, and 85,880 legs were collected in HITS 2012, providing a substantial sample of the population and travel behavior. Selected households were drawn systematically within the country of Singapore.

There are 12,292 trips to workplaces reported in HITS 2012, while weekend trips and trips to schools by students are excluded in this number. Of all commuters, 20.5% drove to work by car, 5.9% were car passengers, and 64.2% took public transportation; the remaining share includes company buses (5.6%), motorcycles (3.9%), taxis (3.6%), and other modes. In HITS 2012, origins and destinations were identified by their zip codes. In Singapore, nearly every building has its own zip code (generally a single high-rise building), resulting in approximately 163,000 zip code entries, including spatial coordinates. This system is advantageous; coordinates are known for every trip origin and destination.

For the workplace choice model, the model is generated for use with a synthetic population [see Sun and Erath for more details (*22*)], with a limited number of descriptive variables. Therefore, certain variables are ignored during model estimation because they are unavailable for the synthetic population. Occupation and individual income are considered essential variables for workplace choice modeling. Both person-related variables are available in the synthetic population and could be included in the utility function; however, individual income was favored as it allows potential future income changes and future scenario adaption. The authors are also aware that occupation might result in substantially important parameters; however, the overall purpose and future model application (see the section on the method) were reasons to consider only the income variable. Income is subdivided into 12 categories from $0 to above $8,000 a month per person.

Spatial Data

Spatial data in the destination choice model provide mainly information about destination characteristics, such as number of workplaces. In addition, spatial attributes can be linked with personal attributes in generalized utility functions.

The Singaporean Master Plan contains detailed information on current and planned land use on the island and includes medium-term spatial plans with a 10- to 15-year planning horizon, compared with the concept plan, with a longer horizon and feasibility studies with shorter time horizons (*23*). The master plan divides the entire island into ~11,000 zones comparable with a parcel size and assigns different land use types; it also defines transport infrastructure such as roads and stations. Land use types used in the master plan are aggregated to reduce the number of model parameters. There are 1,169 zones used for spatial aggregation of workplaces into a reasonably large choice set; however, not all zones contain workplaces (e.g., forests and bodies of water). The following list details different land use types:

- Business,
- Commercial,
- Residential,
- Transportation,
- Services,
- Recreation,
- School, and
- Open.

The number of Singapore workplaces is not captured by any census; it is derived separately through reported travel patterns and destinations' designated land use types. Ordóñez Medina and Erath described the applied workplace distribution model to determine workplace distribution on a building level and, finally, on an aggregated zonal level (*24*). Travel time serves as a major variable for generalized travel cost estimation; calibrated car travel times and transit travel times are available from the current detailed, but aggregated, model of the Land Transport Authority. They could be replaced with updated travel times from this simulation when it is completed.

METHOD

This section has three parts; most relevant background information is provided initially, including the overall model's aim and its prerequisites. The second part explains the workplace choice model and its underlying utility function. The third addresses capacity constraints at destinations in the workplace choice model application defined in the second part.

The workplace choice model is part of an overall transport model for real-world model applications in Singapore and should be capable of reproducing future scenarios with future transport projects, for example, toll stations, new mass rapid transit lines, and large-scale urban and industrial developments. Besides infrastructure changes, Singapore is interested in behavior changes resulting from changing personal factors, such as income or travel time. Therefore, the model should be applicable for those purposes, with appropriate sensitivity, calculation times, and reasonable complexity.

The model is embedded in a demand generation process, which includes additional models to generate the entire synthetic population, household choice of owning a car, and having a driver's license. The decision on how to go to work is also made by applying an MNL model based on available synthetic population data (35.2% of the population go to work on weekdays according to the census). Choices on going to work and location are modeled sequentially. Descriptions of secondary location choice and additional trips for leisure and shopping are determined after workplace location choice and are not reported in this paper. In regard to trip chaining, reported trip chains in HITS 2012 contain fewer trips than would be the case in other countries. Shares of home–work–home tours are approximately 43%, home–education–home are 27%, and home–leisure–home are 9%. Currently, the work destination model and secondary location choice are applied consecutively because of reported data. After demand generation, MATSim, an agent-based microsimulation platform, simulates physical movements in the given infrastructure network and optimizes travel utility related to departure time, secondary location, travel mode, and joint trips (*25*).

Workplace Choice Model Method

The literature contains discussion of various methods (see above) outlining the current trade-off between complexity and low calculation time. An MNL model approach is chosen for the current study, largely because of the reasonable calculation times, even for a large data set. Moderate calculation times are also a factor in implementing and estimating a complex generalized utility function, with the (hopefully) small risk of biased parameter values for explanatory variables, resulting from unobserved spatial correlation. For more detail, Beckhor and Prashker estimated the potential parameter difference between different models (*8*). A rather large data set of 12,292 commuting trips is available for this study; the calculation time is quite high given the number of alternatives (*6*). Despite the large data set, sampling of the alternatives was omitted, and all model parameters were estimated with the entire data set and all destination alternatives. The resulting additional calculation time overhead was manageable, and sampling uncertainties were thus avoided. Technical specifications are described in the section on the results. If a sampling method had been chosen, a minimum sample size of at least 300 destinations would have led to reasonable parameter values. Since calculation time (approximately) scales linearly with the number of alternatives, the additional overhead of four times, at most, is considered reasonable (*3*). The sampling itself is also expensive in calculation time, as well as being dependent on the choice model.

Utility Function

The initially considered utility function consists of a personalized mode choice log sum term and generalized utility to combine personal attributes (e.g., income) with given land use types, similar to the method in Vovsha et al. (*7*). The utility function for origin i and destination j is defined in Equation 1, whereas all indicated components of Equation 1—weighted with γ—are explained in the subsections below. The utility function excludes an accessibility term similar to that in Vovsha et al. because it is assumed that accessibility describes mainly the distance-weighted quality of the surrounding area; not the destination alternative itself (*7*).

$$u_{i,j} = \underbrace{f_0\left(\gamma_{0,m}, c_m, c_m^2, \log(c_m)\right)}_{\text{1. generalized travel cost decay}} + \gamma_1 \cdot \underbrace{\ln\left(\sum_m e^{U_{i,j,m}}\right)}_{\text{2. mode utilities}} + \gamma 2 \cdot \underbrace{f_1(\text{workplaces} \cdot \text{income})}_{\text{3. generalized utilities}} \quad (1)$$

Generalized Travel Cost Decay

The generalized cost (c) decay function consists of a linear combination of mainly nonlinear travel cost modifications. Travel time is considered the main component of generalized travel costs and is, thus, included in the model. In addition, a mode-specific (m) estimation is conducted on the basis of car ownership, considered a major influence on travel behavior.

Mode Choice Log Sum

The mode choice log sum contains the utilities u of each mode m for a given origin–destination relationship i–j. Consideration of all modes can be justified because workplace choice is a long-term decision. Even if a person has a car-oriented travel behavior (which is reflected in the log sum), a specific workplace might still be more attractive if it is also accessible via public transport.

A separate mode choice model was estimated to determine different variable weights. Currently, an MNL model is implemented for car (including car passenger), transit, and others. Minor modes are not considered at this point, even though they might be relevant in some model applications.

Mode choice itself is a rather short-term decision because it determines the choice for a single specific trip and often includes travel time with a linear weight (in this research also). In addition, the mode choice term can correlate with the generalized travel cost decay function. Therefore, detailed experiments were conducted with the mode choice utility (but with a minor effect on the model fit, as described in the section on results).

Generalized Utilities for Land Use Types × Personal Variables

Generalized utility matches personal attributes with alternative specific attributes. For land use types, corresponding workplaces might be linked to a specific person type. Because of the envisaged future scenario applications, income is considered a more reliable prediction variable than occupation types. Income is thus further considered in the destination model. Generalized utilities replace to a certain extent the distance perception separation by income similar to in Mishra et al. (*5*).

Shadow Prices for Capacity Constraints at Destinations

In the following, shadow prices are defined as disutility added to destinations as a result of capacity restrictions. So, shadow prices are positive and are negatively perceived by choice makers. The following three assumptions are made for shadow price calculation: In Assumption 1, it is assumed that the number of workplaces at a given destination is known and fixed for the entire time; in Assumption 2, the demand at all origins is known and fixed and remains that way for different saturation levels at destinations; and in Assumption 3, choice model parameters are given beforehand. It is clear that the second assumption is critical in very specific destination choice models, for example, restaurant choice during evening peak hour. For workplace choice, Assumption 2 might be less critical but obviously depends on the study's economic situation. The third assumption is also critical because model parameters might be influenced by a given saturated market situation. For future research, it would be interesting to know by how much perceived weights or elasticities of generalized travel costs based on a hypothetic questionnaire differ from weights or elasticities of reported trips. Solutions might be found to overcome the additional complexity of Assumption 3, for example, the use of specific questionnaires.

In the following, it is assumed that there is a certain balance between the number of workplaces provided by companies and authorities and the number of commuters and their destinations. On the one hand, it is probably rare to have more workers at a designated place than indicated in the workplace survey data. On the other, it might be possible to have certain vacant workplaces. Therefore, an assignment balance is assumed with a certain upper restriction determined by the maximum number of workplaces, which should not be violated in the model. It can be assumed that this balance cannot be fixed deterministically; therefore, an iterative procedure is derived and proposed to approximate this balanced situation [similar to that in Spiess, who addressed parking lot capacities (*26*)].

Starting with the utility-maximizing approach, the probability $p_{i,j}$ of choosing an alternative destination j based on origin i is defined as

$$g_{i,j} = P_i \cdot p_{i,j} = P_i \cdot \frac{e^{u_{i,j}}}{\sum_j e^{u_{i,j}}} \quad (2)$$

where $u_{i,j}$ is the deterministic utility of alternative j and P_i is the number of people commuting from origin i (a building or a zone). Personal attributes and situational attributes are ignored in Equation 2 by omitting corresponding indexes. However, personal and situational attributes might be considered, as well.

By applying the Kuhn–Tucker conditions, it can be shown that the following convex minimization problem is equivalent to Equation 2; it is feasible because the function is partially differentiable (*27*):

$$\min \sum_{i,j} g_{i,j}\left(\ln\left(g_{i,j}\right) - 1 + u_{i,j}\right) \quad (3a)$$

subject to

$$\sum_j g_{i,j} = P_i \quad (3b)$$

Now, capacity restrictions, such as C_j for destination j (Equation 3c), are simply added to the constraints of the optimization problem above (Equations 3a and 3b).

$$\sum_i g_{i,j} \leq C_j \tag{3c}$$

The number of employees should not exceed the number of workplaces available (Equation 3d), or the problem above becomes infeasible. This requirement is assumed as given beforehand. However, the proposed approach is also applicable if Equation 3d is violated, resulting in oversaturated destinations.

$$\sum_k C_k \geq \sum_k P_k \tag{3d}$$

Vectors λ_1 and λ_2 are added as Lagrange multipliers to determine the Lagrangian L; λ_1 has a length of i, and λ_2 has a length of j:

$$L(g_{i,j}, \lambda_1, \lambda_2) = \sum_{i,j} \left(g_{i,j} \cdot \ln(g_{i,j}) - g_{i,j} + u_{i,j} \cdot g_{i,j}\right) + \lambda_1 \left(\sum_j g_{i,j} - P\right)$$

$$+ \lambda_2 \left(\sum_i g_{i,j} - C\right) \tag{4}$$

Now, optimality conditions of Equation 4 are ($\lambda_2 \geq 0$):

$$g_{i,j} = e^{-\lambda_{1,i} - \lambda_{2,j} - u_{i,j}} \qquad \text{for all } i, j \tag{5}$$

and

$$\sum_j e^{-\lambda_{1,i} - \lambda_{2,j} - u_{i,j}} = P_i \qquad \text{for all } i \tag{6a}$$

$$\sum_i e^{-\lambda_{1,i} - \lambda_{2,j} - u_{i,j}} \leq C_j \qquad \text{for all } j \tag{6b}$$

Vector λ_2 ensures that capacities are not exceeded and is therefore referred to as the shadow price. Unlike Equation 3a through 3c, the dual problem Equations 6a and 6b comes without constraints; λ_1 and λ_2 can be determined by solving Equations 6a and 6b iteratively.

For efficiency, all variables in Equations 6a and 6b are transformed: $\alpha = e^{-\lambda_1}$, $\beta = e^{-\lambda_2}$, $U_{i,j} = e^{-u_{i,j}}$, where α, β, $U > 0$ and $\beta < 1$:

$$\alpha_i \cdot \sum_j \beta_j U_{i,j} = P_i \tag{7a}$$

$$\beta_i \cdot \sum_i \alpha_i U_{i,j} \leq C_j \tag{7b}$$

Efficient Algorithm to Determine Shadow Prices

Algorithm 1 describes an iterative procedure to determine shadow prices. The threshold value, t, describes how much capacity should not be exceeded by a given destination. For example, $t = -2$ means that capacity can be exceeded by a maximum of 2. Algorithm 1 approximates a balanced situation within an adequate number of iterations, in which all commuters are assigned to a workplace. On one side, shadow prices λ_2 can be viewed as an additional (negative) utility for each person to respect capacity constraints. On the other side, λ_2 reflects the future spatial development potential of underdeveloped, or even missing, but valuable, alternatives. Overall, it might be possible that results similar to those achieved with Algorithm 1 could be obtained by randomly assigning weights to locations; however, it is definitely worth knowing how to efficiently approximate a balanced situation, in which all commuters find a designated working location.

Algorithm 1. Shadow price calculation:

$n \leftarrow 0$

$\beta_n \leftarrow 1$

while $C - P < t$ do

Calculate the demand $g_{i,j,n}$ for each pair i, j according to Equations 5 and 7a:

$$g_{i,j,n} \leftarrow \frac{P_i \cdot \beta_n \cdot U_{i,j}}{\sum_j \beta_n \cdot U_{i,j}} \tag{8}$$

Recalculate the β parameters according to Equations 5 and 7b:

$$\beta_{n+1} \leftarrow \min\left(\frac{c_j \cdot \beta_n}{\sum_i g_{i,j,n}}, 1\right) \tag{9}$$

$n \leftarrow n + 1$

end while

Shadow prices: $\lambda_2 \leftarrow -\log(\beta_{n+1})$

Terminate

RESULTS

The results section contains two parts. The first part describes the estimated destination choice model results, and the second part outlines the outcome of the applied shadow price method.

Destination Choice Model

An MNL model approach is chosen for this study; a comprehensive generalized utility function based on the above discussions is included. The MNL model results are described below for each utility function element listed in the section on utility function:

1. The nonlinear mode-specific (m) generalized cost (c) decay function $f_{0,m}$ consists of $f_{0,m} = \gamma_{0,m} (\eta_{0,m} \cdot c_m + \eta_{1,m} \cdot c_m^2 + \eta_{2,m} \cdot \log(c_m))$. In this study, m accounts for car availability, as mentioned above, and c includes travel time as a major component of the generalized costs. Experiments with additional components were conducted, such as

TABLE 1 Workplace Choice Model Parameters

Parameter	Value	t-Test	p-Value
Travel time decay$_{\text{Households with car}}$ ($\gamma_{0\text{ car}}$)	1.19	82.30	.00
$\eta_{2,\text{car}}{}^{a}$	–1.0	na	na
Travel time decay$_{\text{Households without car}}$ ($\gamma_{0\text{ no car}}$)	1.08	74.06	.00
$\eta_{0,\text{no car}}{}^{a}$	–0.121	–8.91	.00
$\eta_{2,\text{no car}}{}^{a}$	–0.688	–30.12	.00
Log (work capacities)	0.262	31.95	.00

NOTE: na = not applicable.
[a]The remaining η-parameters were irrelevant.

c^3 and $\sqrt{c}$. Not all of the η's were clearly significant; overfitting is avoided by selecting only robust elements for $f_{0,m}$. Perception of costs is clearly nonlinear; η values are listed in Table 1 and were kept fixed during the determination of all other parameters. Additional person- or household-based variables could be incorporated for m instead of car availability.

2. In all conducted experiments, the mode choice log sum term correlates with car and transit time (parameter correlation of 0.7 to 0.8) and does not add to the overall model fit; it is therefore ignored in the final model. It is proposed that the main influential variables in the mode choice model should be incorporated as personal variables, as shown above (as m) or in the generalized utility function explained below.

3. The workplaces are transformed with the logarithm because of a considerably higher ρ^2 for model fit and theoretical necessity (*28*). This transformation also reduces correlation with travel time and is in line with Vovsha et al. (*7*). Besides workplace as an isolated independent variable, the generalized utilities include the following combinations (see Table 2):

- Number of workplaces related to business activities × income categories,
- Number of workplaces related to commercial activities × income categories,
- Number of workplaces related to residential activities × income categories, and
- Number of workplaces related to service activities × income categories.

The remaining categories—"school," "recreation," "transport," and "open"—are not significant in most combinations; thus, they are removed from the model. Considered parameters for generalized utilities are significant through many different model estimations. Considered variables and parameters might contribute to a specific stratum, but their contribution to the overall model fitness is rather low because every parameter can contribute only in its very specific combination, for example, when the utility for a person with a high income traveling to a business zone is calculated.

Figure 1, *a* and *b,* shows travel time and cumulative travel time distributions after the workplace choice model is applied to the entire synthetic population of Singapore, comparing this with travel times reported in HITS 2012. Both figures show that overall population travel distributions match the survey distribution. Minor differences

TABLE 2 Workplace Choice Model Generalized Utilities Combinations

Income[a]	Value	t-Test	p-Value	Income[a]	Value	t-Test	p-Value
Log for Business Workplaces				Log for Residential Workplaces			
500	0.0289	2.59	.01	500	–0.101	–5.20	.00
1,250	0.0829	8.26	.00	1,250	–0.089	–4.82	.00
1,750	0.0859	11.38	.00	1,750	–0.169	–12.03	.00
2,250	0.0849	11.76	.00	2,250	–0.182	–13.38	.00
2,750	0.0841	9.59	.00	2,750	–0.251	–14.50	.00
3,500	0.0791	10.51	.00	3,500	–0.276	–18.45	.00
4,500	0.1010	10.67	.00	4,500	–0.252	–13.19	.00
5,500	0.0743	6.51	.00	5,500	–0.298	–12.77	.00
7,000	0.0924	7.19	.00	7,000	–0.257	–9.76	.00
Above 8,000	0.0639	5.78	.00	Above 8,000	–0.314	–14.46	.00
Log for Commercial Workplaces				Log for Service Workplaces			
500	0.060	5.17	.00	500	0.0568	4.32	.00
1,250	0.127	11.85	.00	1,250	0.0030	0.24	.81
1,750	0.126	15.62	.00	1,750	0.0644	7.04	.00
2,250	0.125	16.32	.00	2,250	0.0417	4.78	.00
2,750	0.112	12.22	.00	2,750	0.0586	5.51	.00
3,500	0.125	15.94	.00	3,500	0.0427	4.73	.00
4,500	0.129	12.99	.00	4,500	0.0327	2.80	.01
5,500	0.130	11.19	.00	5,500	0.0253	1.82	.07
7,000	0.163	12.40	.00	7,000	–0.0175	–1.09	.28
Above 8,000	0.194	17.73	.00	Above 8,000	0.0309	2.41	.02

[a]Monthly personal income in Singapore dollars.

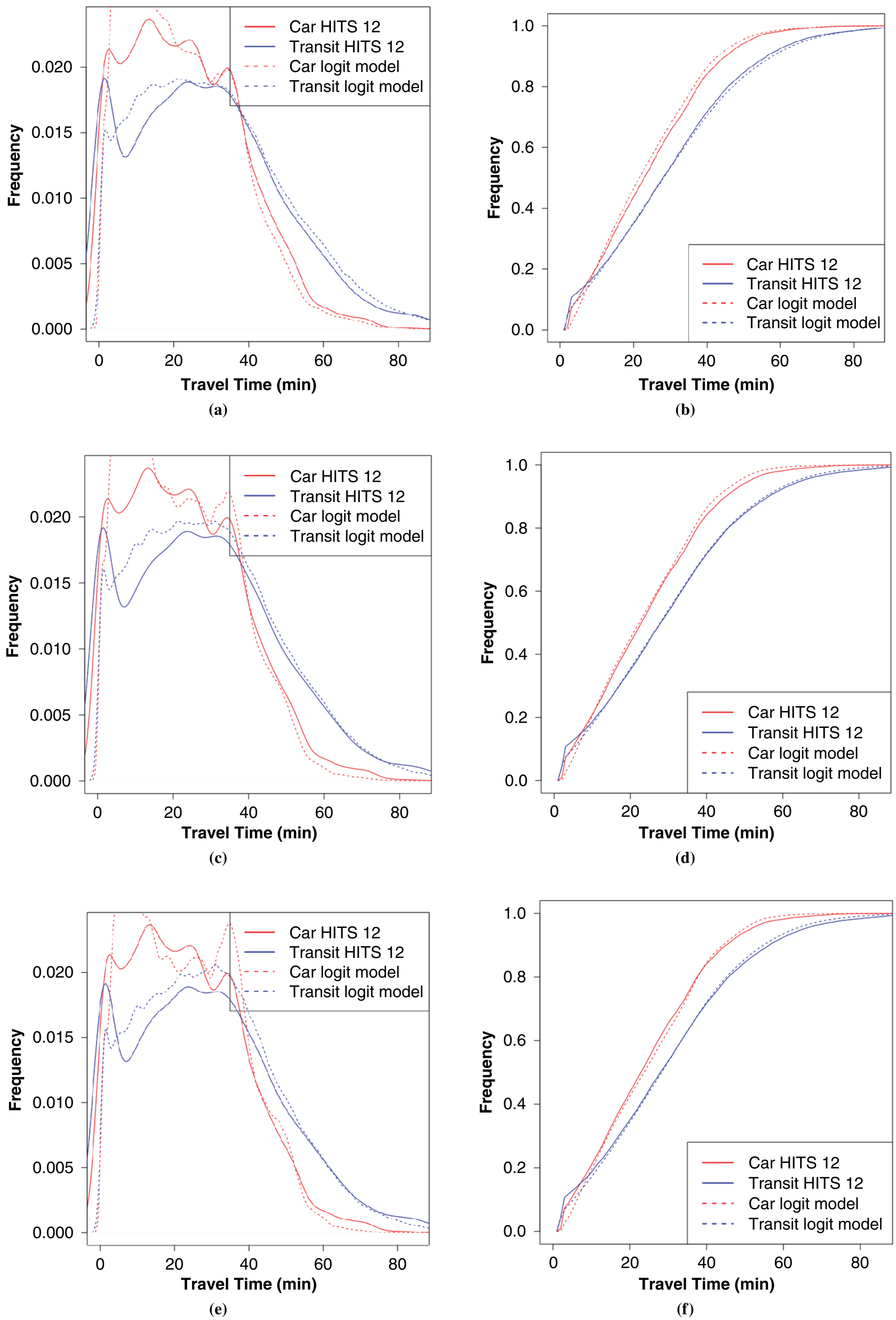


FIGURE 1 Kernel distributions of travel times and cumulative travel time distributions for entire synthetic Singapore population: (*a*) distribution after workplace choice model application, (*b*) cumulated distribution after workplace choice model application, (*c*) distribution after Iteration 1, (*d*) cumulated distribution after Iteration 1, (*e*) distribution after Iteration 10, and (*f*) cumulated distribution after Iteration 10.

can be observed for very short trips, which are not captured properly in the data source; destination zone granularity hampers precise value determination for these trips. Distance perception is not captured correctly in the model parameters for very short trips.

Figure 2*a* shows the scatter plot for all destinations and their saturation after the workplace choice model is applied. In addition, Figure 2*b* shows the destination density, with a given number of workplaces, to reflect the distribution of different-size destinations; those with smaller capacities are mostly oversaturated, whereas destinations with higher capacities are undersaturated, mainly because of the specific workplace distribution in Singapore. This observation is noteworthy because some bigger industrial zones are located in outlying areas and on smaller islands.

Model computation time was about 3 days (mainly because of generalized utilities) with Pythonbiogeme on a Linux computer featuring an Intel Xeon 3.07 GHz processor and 30 GB RAM (*29*). Various experiments with different parameter combinations showed robust parameter values.

Shadow Prices

Algorithm 1 is applied to determine the shadow prices of each zone *j,* relying on the number of workplaces in *j,* which is independently derived from separate data (see the section on spatial data). The algorithm converges to an optimal solution. Figure 1 shows estimated synthetic population travel time distributions compared with the travel time distribution of HITS 2012, during an iteration selection of Algorithm 1. Figure 1, *a* and *b,* shows results after Iteration 0, which basically applies the workplace choice model without shadow price modifications (Iteration 0 in Algorithm 1 has β_0-values = 1). Figure 1, *c* through *f,* displays the results after Iterations 1 and 10, respectively.

Figure 1, *c* through *f,* depicts minor changes in travel times compared with HITS 2012 and compared with Figure 1, *a* and *b,* because of the reassignment of the commuters to workplaces and because workplace data are estimated; a certain additional deviation from real workplace numbers might occur. Despite these changes, the statistical model fit improves during Algorithm 1 iterations. There is an

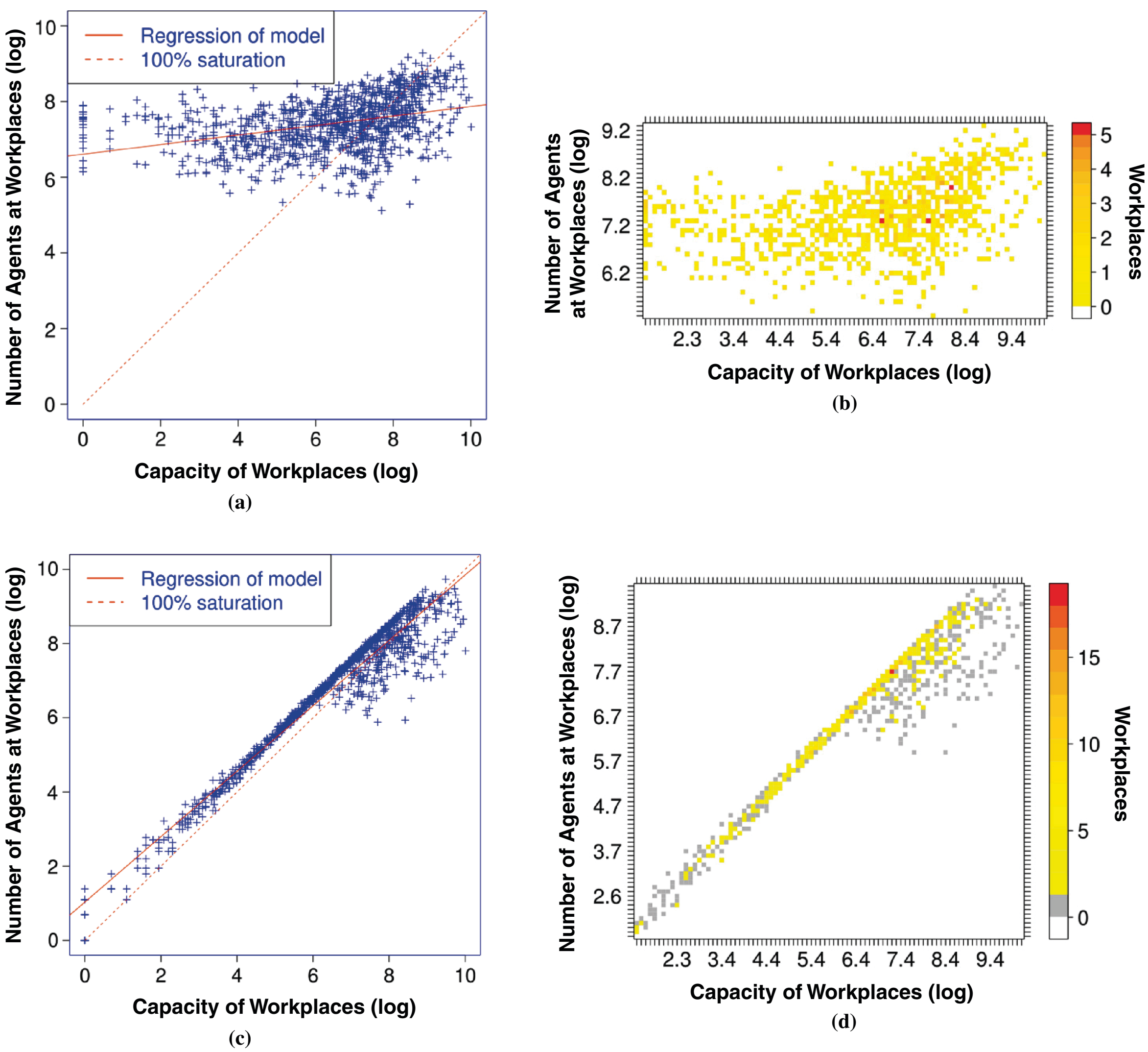


FIGURE 2 Regression and heat map with workplace capacities and saturations based on synthetic Singapore population: (*a*) saturation after workplace choice model application, (*b*) number of workplaces and saturation after workplace choice application, (*c*) saturation after Iteration 1, and (*d*) number of workplaces and saturation (Iteration 1).

(continued on next page)

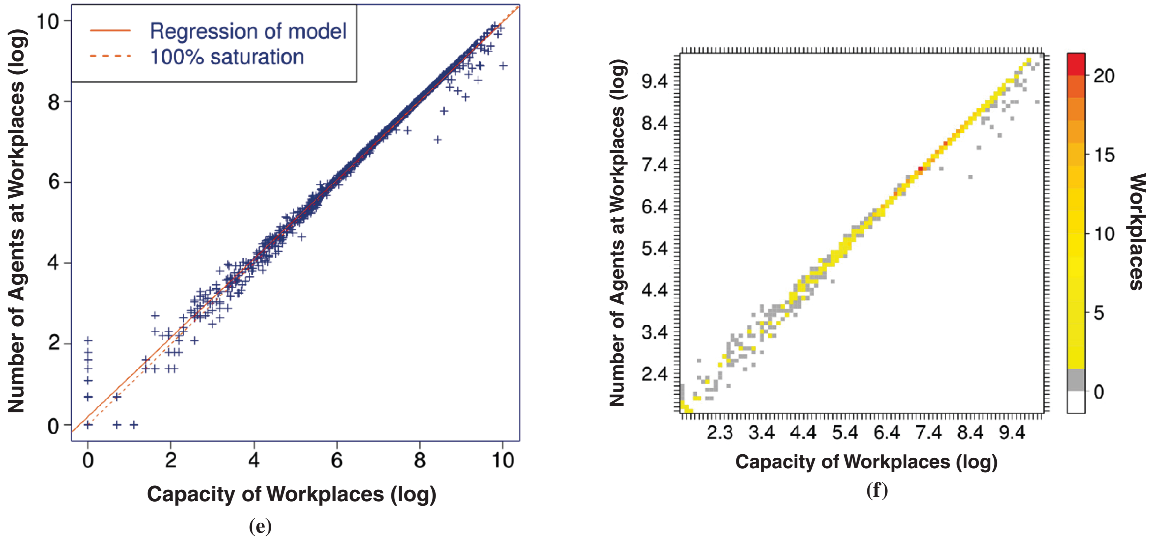


FIGURE 2 *(continued)* **Regression and heat map with workplace capacities and saturations based on synthetic Singapore population: (*e*) saturation after Iteration 10 and (*f*) number of workplaces and saturation (Iteration 10).**

increase in ρ^2 and a considerable decrease in the log likelihood, as summarized in Table 3. This result means that, statistically, model prediction is improved during Algorithm 1 application when compared with a standard MNL model application. It seems that Algorithm 1 is capable of further improving overall model prediction and even correcting certain data uncertainties, as mentioned above.

Figure 2, *c* through *f,* shows saturation of Iteration 1 and 10, similar to Figure 2, *a* and *b,* as explained above. Some over- and undersaturated workplaces can still be found in Iteration 1 (Figure 2, *c* and *d*), whereas workplaces are almost completely saturated in Iteration 10 (Figure 2, *e* and *f*).

Figure 3*a* shows the spatial distribution of the Singapore population, 3*b* shows work utilities, and 3*c* shows shadow prices ($\lambda_2 = -\log(\beta)$). It is clear that work utilities and shadow prices are often complementary. Knowledge about shadow prices can support location analysis and planning for spatial development.

Destination choice model application for the whole synthetic population (~1.9 million commuters) takes about 5.5 h for Iteration 0 and 1.2 h for each subsequent iteration, as well as considerable memory, about 50 GB (one thread on an Intel Xeon 3.07 GHz).

TABLE 3 Workplace Choice Model Statistic

Model Statistic	Value
Number of observations	12,292
Number of alternatives (excluding green spaces, bodies of water, and so forth)	1,087
Initial log likelihood	−85,416
Final log likelihood	−75,556
Final log likelihood (Iteration 10)[a]	−71,035
ρ^2	.115
ρ^2 (Iteration 10)[a]	.168

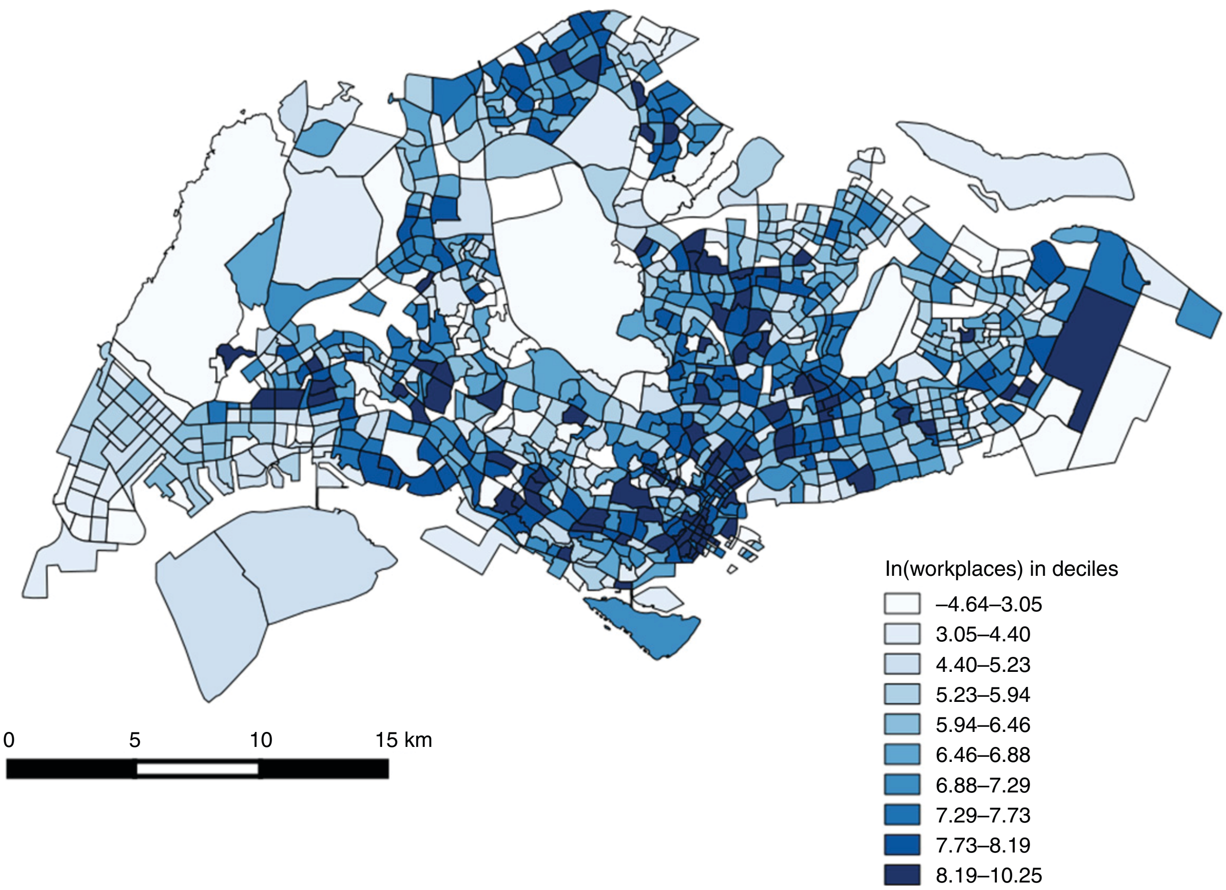
[a]Based on census data and application of Algorithm 1.

DISCUSSION AND OUTLOOK

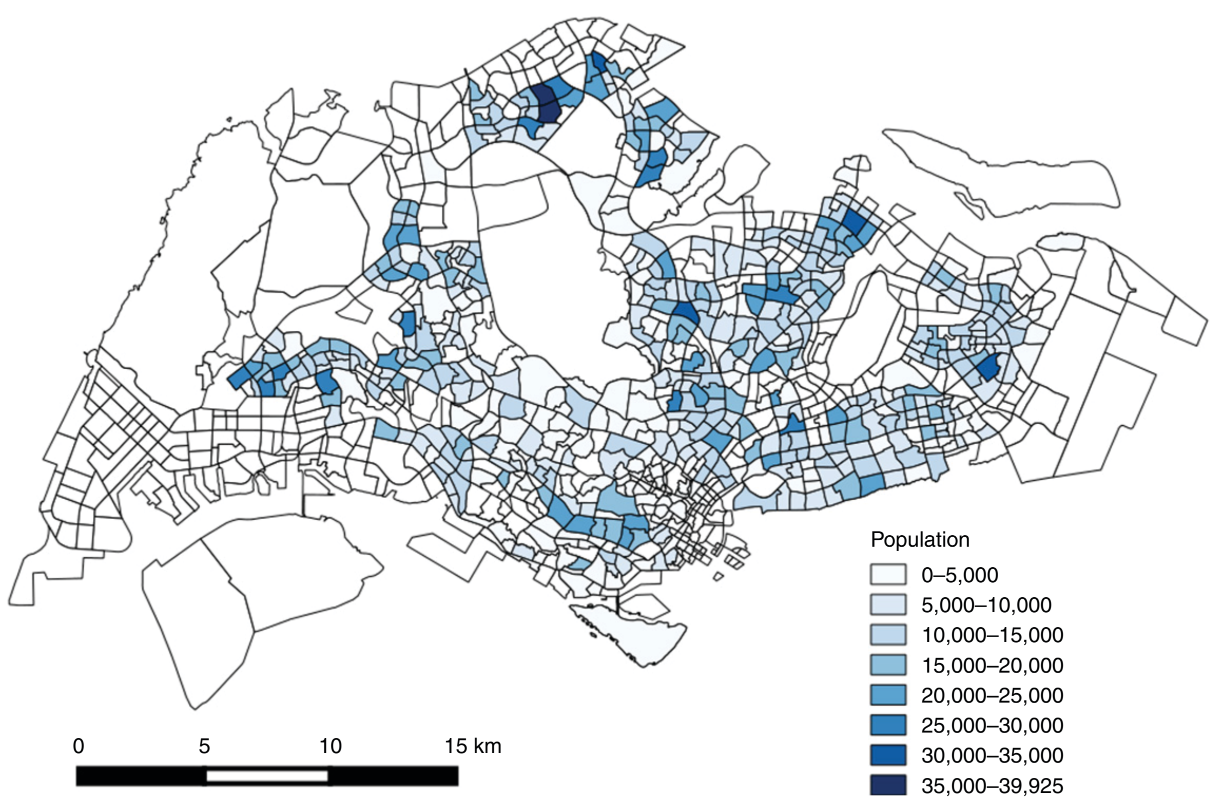
Today, destination choice models with elaborate utility functions can be applied on the basis of large data sets and many destinations. In addition, destination capacity constraints can be efficiently implemented in large-scale, agent-based models while maintaining choice heterogeneity. The proposed approach introduces shadow prices at destinations, which are able to avoid oversaturation at popular locations. The paper's proposed method is successfully applied to determine shadow prices for each zone in Singapore. Shadow prices can also help predict the future spatial development potential of underdeveloped, or unused and valuable, destination alternatives. According to the experiments and results, the travel time distribution of all trips changes only slightly during workplace assignment optimization compared with an empirical reference sample distribution.

The method is applied on data for all of Singapore island, which is isolated from the mainland except for two bridges to Malaysia, one international airport, and ferry connections. Geographic border effects might still occur in the island model or in potential future applications in mainland areas. Influence of people from areas outside the model perimeter might also affect shadow price calculations pertaining to outlying areas in the model and other areas easily accessible from outside the model perimeter. These external influences are difficult to capture quantitatively; however, the type of influence is probably similar to "border effects" occurring in other transport models, for example, at route choice. Overall, care is required in these areas when results are interpreted. Additional studies might also reveal the quantitative extent of these effects.

In addition to trip destination choice application, location choice for primary location (household and firm locations) can potentially benefit from the proposed method. Parking, battery-charging stations (e-mobility), and other capacity restraint activities might also profit from the proposed method, along with transportation modeling, spatial economics (as mentioned above), and resource distribution, in a wider sense (*26*). The proposed shadow price method, however, does not replace any alternative specific variable. Also, calculation



(a)



(b)

FIGURE 3 Singapore: (*a*) population distribution and (*b*) workplace distribution.

(continued on next page)

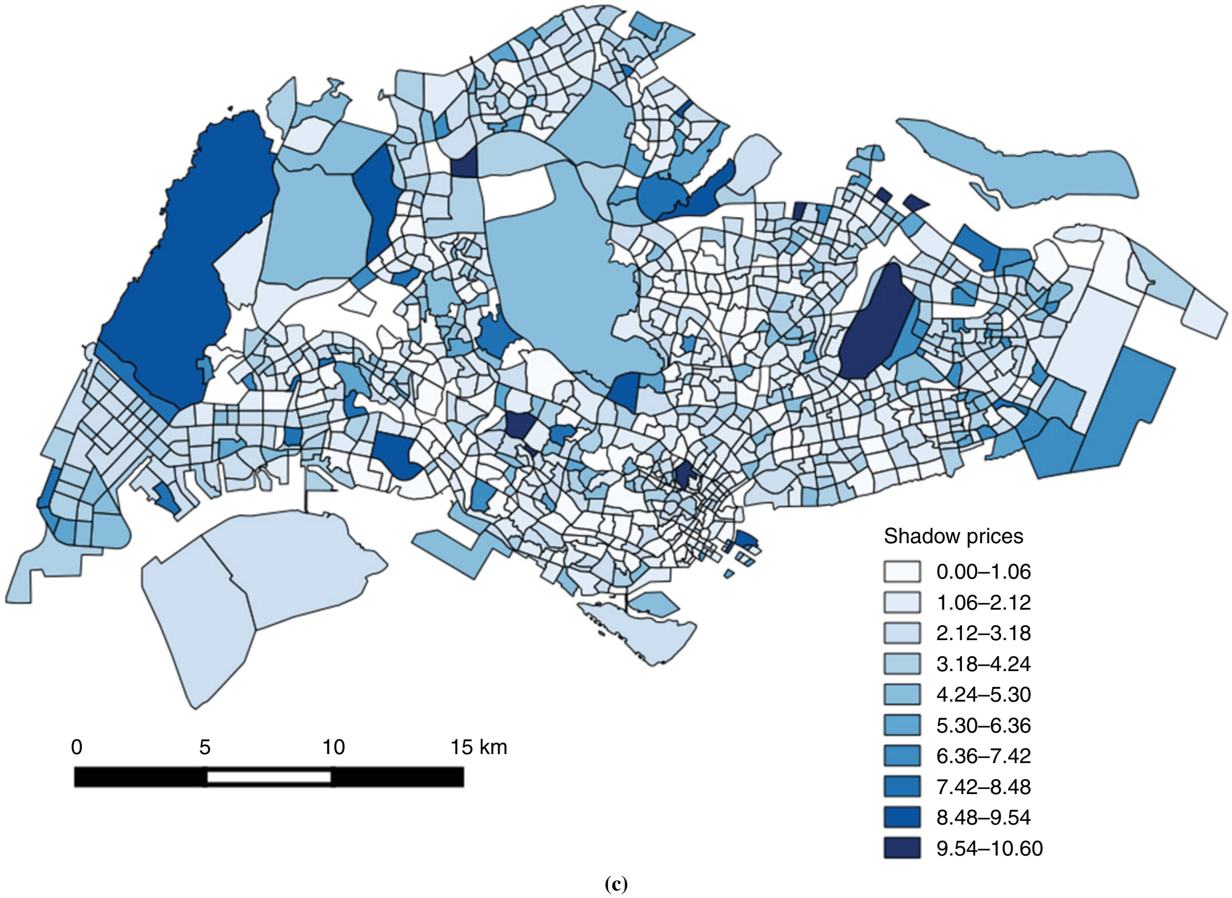


FIGURE 3 *(continued)* **Singapore: *(c)* shadow prices λ_2 distribution after Iteration 10.**

for shadow prices holds only for MNL models; applications for other choice model approaches have not been proposed yet.

ACKNOWLEDGMENTS

The authors thank Yoram Shiftan and Sriram Narayanamoorthy for support on the modeling side, Michel Bierlaire for providing Pythonbiogeme Version 2.4, four anonymous reviewers for their efforts and advice, and the reviewer of the draft manuscript for readability improvements. The Land Transport Authority and Urban Redevelopment Authority of Singapore contributed the data used for this research, which was conducted as part of the CREATE initiative of the National Research Foundation of Singapore.

REFERENCES

1. Ortúzar, J. d. D., and L. G. Willumsen. *Modelling Transport,* 4th ed. John Wiley & Sons, Inc., Chichester, United Kingdom, 2011.
2. McFadden, D. *Modelling the Choice of Residential Location.* Working Paper No. 477. University of California, Berkeley, Yale University, New Haven, Conn., 1977.
3. Ben-Akiva, M. E., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Mass., 1985.
4. Daly, A. J. Applicability of Disaggregate Models of Behaviour: A Question of Methodology. *Transportation Research Part A: Policy and Practice,* Vol. 16, No. 5, 1982, pp. 363–370.
5. Mishra, S., Y. Wang, X. Zhu, R. Moeckel, and S. Mahapatra. Comparison Between Gravity and Destination Choice Models for Trip Distribution in Maryland. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
6. Land Transportation Authority. *Household Interview Travel Survey 2012.* http://www.lta.gov.sg/. Accessed July 2, 2015.
7. Vovsha, P., S. Gupta, J. Freedman, W. Sun, and V. Livshits. Workplace Choice Model: Comparison of Spatial Patterns of Commuting in Four Metropolitan Regions. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
8. Bekhor, S., and J. N. Prashker. GEV-Based Destination Choice Models That Account for Unobserved Similarities Among Alternatives. *Transportation Research Part B: Methodological,* Vol. 42, No. 3, 2008, pp. 243–262.
9. Bhat, C. R., and J. Guo. A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling. *Transportation Research Part B: Methodological,* Vol. 38, No. 2, 2004, pp. 147–168.
10. Sener, I. N., R. M. Pendyala, and C. R. Bhat. Accommodating Spatial Correlation Across Choice Alternatives in Discrete Choice Models: An Application to Modeling Residential Location Choice Behavior. *Journal of Transport Geography,* Vol. 19, No. 2, 2011, pp. 294–303.
11. Koppelman, F. S., and V. Sethi. Closed-Form Discrete-Choice Models. In *Handbook of Transport Modelling* (D. A. Hensher and K. J. Button, eds.), Pergamon, Amsterdam, Netherlands, 2000, pp. 235–252.
12. Frejinger, E., M. Bierlaire, and M. Ben-Akiva. Sampling of Alternatives for Route Choice Modeling. *Transportation Research Part B: Methodological,* Vol. 43, No. 10, 2009, pp. 984–994.
13. Nerella, S., and C. R. Bhat. Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894,* Trans-

portation Research Board of the National Academies, Washington, D.C., 2004, pp. 11–19.
14. Müller, S., K. Haase, and F. Seidel. Exposing Unobserved Spatial Similarity: Evidence from German School Choice Data. *Geographical Analysis,* Vol. 44, No. 1, 2012, pp. 65–86.
15. de Palma, A., N. Picard, and P. A. Waddell. Discrete Choice Models with Capacity Constraints: An Empirical Analysis of the Housing Market of the Greater Paris Region. *Journal of Urban Economics,* Vol. 62, No. 2, 2007, pp. 204–230.
16. Albouy, D. Constrained Optimization, Shadow Prices, Inefficient Markets, and Government Projects. Working Paper No. 131. University of California, Berkeley, 2004.
17. Gupta, S., P. Vovsha, R. Kumar, and A. Subhani. Incorporating Cycling in Ottawa–Gatineau Travel Forecasting Model. Presented at 5th Conference on Innovations in Travel Modeling, Baltimore, Md., 2014.
18. Davidson, B., P. Vovsha, and J. Freedman. New Advancements in Activity-Based Models. Presented at Australasian Transport Research Forum, Adelaide, Australia, 2011.
19. Hammadou, H., and C. Papaix. Estimating Shadow-Prices of Urban Transport Mode Choice for a Second Best Pricing of CO_2 Emissions. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
20. Beckmann, M. J., and J. P. Wallace. Evaluation of User Benefits Arising from Changes on Transportation Systems. *Transportation Science,* Vol. 3, No. 4, 1969, pp. 344–351.
21. Los, M. Combined Residential–Location and Transportation Models. *Environment and Planning A,* Vol. 11, No. 11, 1979, pp. 1241–1265.
22. Sun, L., and A. Erath. A Bayesian Network Approach for Population Synthesis. Presented at 4th Symposium of the European Association for Research in Transportation, Copenhagen, Denmark, 2015.
23. Urban Redevelopment Authority. *Master Plan 2008.* https://www.ura.gov.sg/maps/. Accessed July 5, 2015.
24. Ordóñez Medina, S. A., and A. Erath. Estimating Dynamic Workplace Capacities by Means of Public Transport Smart Card Data and a Household Travel Survey in Singapore. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2344,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 20–30.
25. MATSim. Multi-Agent Transportation Simulation. 2015. http://www.matsim.org. Accessed July 30, 2015.
26. Spiess, H. *A Logit Parking Choice Model with Explicit Capacities.* Research report. EMME2 Support Center, INRO, Aegerten, Switzerland, 1996.
27. Horst, R. *Nichtlineare Optimierung.* Carl Hanser Verlag, Munich, Germany, 1979.
28. Train, K. E. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand.* MIT Press, Cambridge, Mass., 1986.
29. Bierlaire, M. BIOGEME: A Free Package for the Estimation of Discrete Choice Models. Presented at 3rd Swiss Transport Research Conference, Ascona, Switzerland, 2003.

The Standing Committee on Transportation Demand Forecasting peer-reviewed this paper.