Article:

Ali, A., Bhaduri, E. and Choudhury, C.F. (Accepted: 2026) Comparing econometric and machine learning algorithms for modelling daily time-use patterns. Transportmetrica A: Transport Science. ISSN: 2324-9935 (In Press)

Comparing econometric and machine learning algorithms for modelling daily time-use patterns

Azam Ali[1], Eeshan Bhaduri[1], and Charisma F. Choudhury*[1]

[1]Choice Modelling Centre & Institute for Transport Studies, University of Leeds, Leeds, United Kingdom

Abstract
Time-use decisions are a critical input for quantifying travel demand in activity-based models. These decisions are typically modelled using econometric methods grounded in utility maximisation theory - multiple discrete-continuous extreme value (MDCEV) models in particular. Since time-use decisions are quite complex and may not necessarily be underpinned by a utility-maximisation framework, we propose to use a data-driven machine learning technique (Recurrent Neural Networks, RNN) to model time-use. We compare and contrast time-use models developed using a long short-term memory (LSTM), a type of RNN model, with an MDCEV model based on the predictive performance and marginal effects. Using time-use data collected during the COVID-19 pandemic in the UK, we observe that there are no significant differences in prediction accuracy between the two approaches, both at the aggregate (sample) and disaggregate (individual) levels. Thus, we find no evidence that data-driven methods outperform traditional econometric models in predicting time-use behaviour. We also observe that the marginal effects derived from both models are broadly similar. However, in the absence of benchmark or ground-truth values, it is not possible for us to conclude the output of which model is closer to reality. This limits the depth of the comparison. Overall, our findings suggest that RNN models can serve as viable alternatives to MDCEV models for modelling time-use with the added capability of generating activity schedules. In contrast, MDCEV models offer greater transparency about the assumptions and provide more interpretable, policy-relevant outputs.

Key words: time-use; machine learning; recurrent neural network; multiple discrete continuous extreme value model

*Email: C.F.Choudhury@leeds.ac.uk

List of Abbreviations

- ANN: Artificial Neural Network
- BPNN: Back-propagation Neural Network
- CM: Choice model
- CNL: Cross Nested Logit
- CPNN: Counterpropagation Neural Network
- DNN: Deep Neural Network
- GBT: Gradient boosting trees
- KNN: K Nearest Neighbour
- LSTM: Long short-term memory
- MDCEV: Multiple Discrete Continuous Extreme Value
- ML: Machine learning
- MMNL: Mixed Multinomial Logit or Mixed Logit
- MNL: Multinomial Logit
- NB: Naive Bayes
- NL: Nested Logit
- NN: Neural Network
- PNN: Probabilistic Neural Network
- RBFN: Radial Basis Function Network
- RNN: Recurrent Neural Network
- RUM: Random Utility Maximisation
- SVM: Support Vector Machine
- XGB: Extreme Gradient Boosting

1 Introduction

Time-use decisions are fundamental in shaping travel and activity patterns, and therefore are key components of activity-based travel demand models. Time-use decisions are typically modelled using variants of the multiple discrete continuous extreme value (MDCEV) model based on the random utility maximisation (RUM) theory. The basic assumption of these models is that individuals choose the type of activities and the associated time durations that provide them the maximum utility (subject to a budget constraint) [Bhat, 2008]. Nevertheless, there are some issues with adopting a random utility maximisation framework to model time-use. Firstly, it must be noted that comprehending time-use patterns and activity schedules is a multifaceted endeavour. There exists no overarching or universally accepted behavioural theory that provides a comprehensive explanation of how individuals determine the appropriate time duration for various activities. [Axhausen and Gärling, 1992]. In terms of discrete choices such as mode choice, it is quite intuitive that people face limited options with different alternative specific attributes [Habib, 2011]. However, time-use is radically different as there are some activities which are mandatory, such as working and education, whereas some activities are discretionary, such as leisure and social activities. People also have time-space constraints, which have the potential to open up or restrict opportunistic trips or activities [Buliung and Kanaroglou, 2007]. Secondly, in time-use diaries, we only know the observed outcomes. But it is difficult to ascertain the choice sets, limitations, and penalties which may be associated with selecting the schedule and duration of activities [Habib, 2011]. Finally, a major problem is deriving a mathematical model that can incorporate the dynamics of activity scheduling under time constraints [Habib, 2011]. All these questions regarding the validity of using the MDCEV model and random utility maximisation framework for modelling time-use have led to researchers developing more complex models [Khaddar et al., 2024]. For instance, Castro et al. [2012] proposes to incorporate multiple budgets (such as both monetary and temporal budgets) instead of only a temporal budget, which is normally used to estimate MDCEV models. Saxena et al. [2021] proposes constraints in the MDCEV model structure to ensure that activities have an upper bound on time-use consumption, etc. However, an alternative data-driven approach exists, namely, machine learning models.

Over the past few years, there has been an increase in studies that have used machine learning (ML) algorithms to model travel behaviour as an alternative to using econometric models, which are normally embedded in behavioural theories [Hillel et al., 2021, van Cranenburgh et al., 2022]. Machine learning models are data-driven, can easily model non-linear relationships, and are not dependent on the modeller's judgement. Whereas, in econometric models, there is a need to specify the model structure, which leaves limited flexibility in adding complexity to the model. Further, the process of specifying the utility function depends on the modeller's intuition, thereby increasing the risk of model misspecification. This has led to the rise in the usage of machine learning (ML) algorithms to model travel behaviour, including mode choice [Hillel et al., 2021], car ownership decisions [Ali et al., 2023], travel mode and purpose [Wong and Farooq, 2020], route choice [Yao and Bekhor, 2022], airline choice [Lhéritier et al., 2019]. Nevertheless, there has been very limited application of machine learning techniques to model time-use or activity-based models [Koushik et al., 2020]. For instance, Arentze and Timmermans [2004] developed ALBATROSS, which is a rule-based or computational process model that predicts activity type, duration, location, and modes, where they use a decision tree, an ML algorithm, to cluster activity choices such as travel times, starting time, and duration time. Hafezi et al. [2021] developed a series of models (activity type, start time, and duration) to predict daily activity schedules, Drchal et al. [2019] developed a generative machine learning modelling framework which uses sequential models to model activity type, duration, destination, and mode to model daily activity schedules. Koushik et al. [2023] model activity schedule and time-use using a long short-term memory (LSTM) recurrent neural network (RNN) in the context of India.

However, there are some limitations in the previous research that utilise ML algorithms for modelling activity choice and time-use. Firstly, these studies do not compare the performance of their model to a similar econometric model for modelling time-use, namely the MDCEV model. Therefore, it is not clear whether or not RNN outperforms the widely used MDCEV. Especially, previous research in the comparison of econometric and machine learning modelling reveals that there is often not a substantial difference in the predictive performance of the ML and econometric models [Ali et al., 2023, Hillel, 2019, Zhao et al., 2020] and the transferability of the ML models are often not acceptable in cases where the application scenario is radically different from the training scenario [Ali et al., 2023]. Secondly, a major barrier in using ML models is that they are typically used as 'black boxes' and it is difficult to extract reliable economic and behavioural outputs such as elasticities, value of travel time savings, etc. from the ML outputs, making them unsuitable for underpinning planning and policy decisions [van Cranenburgh et al., 2022]. The previous researchers who have used ML models for activity and time-use did not address this issue. Therefore, there is scope to investigate in further detail if the outputs of the ML models provide behaviourally consistent outputs. In particular, there are some explainable ML techniques which can be used to explain ML model results, such as elasticities, marginal effects, and feature importance plots, which would give confidence to transport practitioners to use ML algorithms instead of econometric models. The current paper aims to address this research gap by applying ML models, specifically recurrent neural networks (RNN), to model time-use, compare it with the MDCEV model and interpret the behavioural insights gathered from the two types of models. The detailed research questions are as follows:

1. Assess whether data-driven time-use models have a better predictive performance compared to econometric or utility-based models.

2. Investigate the similarities and differences in the policy insights derived from the two types of models in different forecasting scenarios.

In this study, we model time-use using data collected in the UK during and before the COVID-19 Pandemic (2016-2020) [Gershuny et al., 2022] using both RNN and MDCEV models. We compare the predictive performance of both models at an aggregate and individual level on a testing dataset. To check if both models have similar performance under different policy forecasting scenarios, we compare the marginal effects. The rest of the paper has been organised as follows: Section 2 gives a brief introduction about the background theory of recurrent neural networks and multiple discrete continuous extreme value models, along with the comparison methodology. Section 3 describes the data, Section 4 describes the model development and results, followed by the conclusion in Section 5.

2 Literature review and methodological framework

2.1 Comparison of econometric and machine learning models in travel behaviour modelling

In the last decade, machine learning (ML) models have emerged as a powerful alternative to traditional econometric approaches in travel behaviour modelling. Their ability to capture complex, non-linear relationships and leverage high-dimensional datasets has positioned ML techniques as a promising complement, and in some cases, even better counterpart to RUM theory-based econometric models, which have long dominated the field. Therefore, there has been a growing interest in comparing the ML and RUM-based approaches from the perspective of prediction performance for discrete outputs as highlighted in Table 1.

For instance, Lee et al. [2018] conducted a comparative analysis of non-parametric and parametric methods for mode choice modelling using travel diary data. The results demonstrated that artificial neural networks (ANNs) consistently outperform the traditional multinomial logit (MNL) model, with probabilistic neural networks performing best, particularly for underrepresented travel modes. Golshani et al. [2018] found that neural networks (NN) outperformed traditional statistical models in terms of prediction accuracy and ease of implementation. Although NN lacks direct interpretability of explanatory variables, sensitivity analysis has revealed that it can effectively capture non-linearities and offers faster estimation than econometric models. Building on these, Zhao et al. [2020] assessed the predictive accuracy and behavioural interpretability of ML models (including Naive Bayes, tree-based algorithms, SVM, and neural networks) and logit models (MNL and MMNL) using stated preference data. Random Forest emerged as the best-performing machine learning model, surpassing MNL in predictive accuracy. However, the study underlined that while ML models excel in prediction, econometric models remain more behaviourally consistent for mode-choice interpretation. In a similar vein, Martín-Baos et al. [2023] conducted a systematic evaluation using three real-world and twelve synthetic datasets to compare ML classifiers (SVM, XGB, RF, NN, DNN) with MNL. Despite ML models achieving better individual-level predictions, they performed poorly in estimating behavioural indicators and aggregate mode shares. In contrast, the study by Salas et al. [2022] observed the ML approach (specifically, ANN) as the most effective, outperforming both conventional logit. Moreover, permutation feature importance analysis revealed that variable contributions in ML models align closely with logit estimates, suggesting consistency in behavioural interpretation. Javadinasr et al. [2023] noted that ML models achieved higher overall accuracy than MNL, though their performance declined for minority classes in unbalanced datasets. In contrast, MNL provided interpretable alternative-specific variables valuable for policy insights. Another study investigated the simultaneous modelling of mobility tool ownership by applying shallow and deep neural networks, along with nested and cross-nested logit models [Püschel et al., 2024]. The results revealed that shallow neural networks achieved higher accuracy and robustness than discrete choice models, while also offering easier interpretability compared to deep architectures with multiple hidden layers. Expanding the scope, Wang et al. [2024] developed an empirical benchmark for comparing 105 ML and choice models from 12 families across multiple contexts. Findings highlighted that ensemble and deep learning techniques generally outperform traditional choice models in statistical performance. Importantly, contextual factors such as data sources, input variables, and choice categories were found to exert a stronger influence on predictive performance than the choice of model type itself.

Upon reviewing the approaches employed and corresponding findings in the studies mentioned above, the following research gaps could be identified:

- There is a lack of consensus regarding the predictive performance of machine learning and econometric models, especially for imbalanced datasets or out-of-context predictions.
- The consistency of behavioural outcomes from ML predictions compared to econometric models is still under-explored.
- While numerous studies have examined discrete outcomes such as mode choice and vehicle ownership, few have compared multiple discrete outcomes in the context of time-use decisions using both econometric and machine learning models.

Table 1: **Summary of recent studies using econometric and ML techniques in travel behaviour modelling**

Study	Aim	Data source	Methodology	Key findings
Lee et al. [2018]	Comparative analysis of non-parametric and parametric methods for mode choice modelling	Travel diary data	Non-parametric: BPNN, RBFN, PNN, CPNN; Parametric: MNL	1. ANNs outperform the traditional MNL model, with PNNs performing best, especially for under-represented travel modes. 2. ANNs enhance prediction accuracy and provide greater insights compared to MNL.
Golshani et al. [2018]	Comparative analysis of statistical and ML methods for mode choice modelling and departure time choice	Household travel survey	Statistical: MNL and Hazard model; ML: NN	1. NN demonstrated superior accuracy for both prediction outcomes, and significantly reduced estimation time compared to statistical models. 2. Sensitivity analysis confirmed better ability of NN to capture nonlinear decision behaviour.
Zhao et al. [2020]	Comparative analysis between ML and logit models in terms of both predictive accuracy and behavioural interpretability	SP survey dataset within an academic campus setting	ML models: NB, Tree-based models, SVM, NN; Logit model: MNL, ML	1. RF is the best-performing ML model and outperforms the predictive accuracy of MNL and ML models. 2. ML models offer higher predictive accuracy, but logit models provide more behaviourally consistent interpretations in mode-choice modelling.
Salas et al. [2022]	To perform a comparative evaluation of ML classifiers and CMs in the context of travel mode choice	Synthetic dataset generated from multinomial framework	5 ML classifiers: SVM, XGB, RF, ANN and KNN and 2 Choice Models: MNL and MMNL	1. ANN outperforms conventional logit and other ML models, highlighting the value of flexible modelling approaches. 2. Permutation feature importance analysis shows that the variable contributions to the ML model's predictive accuracy align with the logit model estimates.

Study	Aim	Data source	Methodology	Key findings
Martín-Baos et al. [2023]	To perform a systematic evaluation of ML and discrete choice models in the context of travel mode choice	3 real-world and 12 synthetic datasets	5 ML classifiers: SVM, XGB, RF, NN and DNN and 1 choice Model: MNL	1. XGBoost and RF achieve strong individual-level predictions but perform poorly in estimating behavioural indicators and aggregate mode shares. 2. The MNL model shows consistent performance across contexts, while ML methods can enhance behavioural metrics such as Willingness To Pay (WTP).
Javadinasr et al. [2023]	To model individual mode choice decisions	Household travel survey	CM: MNL and ML: NN and GB	1. ML models achieved higher overall accuracy than MNL, though their precision declined for minority classes in unbalanced datasets. 2. MNL model provided interpretable parameters compared to ML model, making it more relevant for policy implications.
Ali et al. [2023]	Vehicle ownership decision	Household travel survey	CM: MNL (Piecewise Linear) and ML: NN and GB	1. CM outperforms ML on the testing dataset in terms of LL and MAPE.
Wang et al. [2024]	To develop an empirical benchmark for comparing several ML models and CMs while accounting for contextual variations	RP survey, transit agency survey with simulation-based travel information, and SP survey	105 individual models from 12 model families (both ML and CM)	1. Ensemble techniques and deep learning approaches outperform CM models, including MNL, NL, and Mixed Logit. 2. The contextual factors play a more critical role in determining predictive performance than the differences between model types.
Püschel et al. [2024]	Simultaneous modelling of mobility tool ownership	Household travel survey	CM: shallow NN and DNN and CM: NL and CNL	1. Shallow NNs outperformed discrete choice models, offering greater accuracy, robustness, and easier interpretation than deeper neural networks.

Recognising the importance of these issues, the present study directly contrasts machine learning and econometric approaches to analyse time-use decisions, focusing on two key aspects: (i) their predictive performance at both aggregate and disaggregate levels; and (ii) the behavioural consistency of their predictions under different policy contexts.

2.2 Background Theory

This subsection discusses the fundamental theories of recurrent neural networks (RNNs) and multiple discrete continuous extreme value (MDCEV) models.

2.2.1 Recurrent neural networks

The core concept of supervised machine learning techniques is to train an algorithm that maps inputs to outputs by minimising a loss function. In standard classification tasks, the loss function is often categorical cross-entropy or log-likelihood, and the outputs are typically discrete and independent [Goodfellow et al., 2016]. However, many real-world problems, such as speech recognition, text generation, involve sequences of correlated outputs [Graves, 2012]. In these cases, predicting one output depends not only on the input but also on the context provided by previous outputs. Time-use decisions are similar to such decisions as they are inherently sequential and spatio-temporally interdependent, i.e., the time spent on one activity (or travelling for doing an activity) can influence the likelihood and duration of subsequent activities.

To model such dependencies, recurrent neural networks (RNN) are particularly well-suited as they are designed to handle labelled sequences and time-series data by maintaining an internal state that captures information from previous time steps [Goodfellow et al., 2016]. This structure allows RNNs to share information across time, making them ideal for capturing underlying dynamics in activity scheduling. This implies RNN can learn how different activities are chained throughout the day, taking into account constraints such as available time and activity patterns. It must be noted that there are other recent machine learning models which may be used to model time-use sequences, such as transformers and large language models (LLMs); however, we restrict our scope to RNNs as they are more suited to numerical data, whereas transformers and LLMs are more suited for textual data.

Before, understanding how RNN operates, it is essential to know how a basic neural network (NN) works. There are three layers in neural networks: the input layer, the hidden layer, and the output layer. The hidden layer has a set of neurons, where each neuron is considered to have a weight W associated with explanatory variables x and a bias b, that is transformed by a non-linear activation function f as mentioned in Equation 1. The parameters, i.e., the weights and biases, are estimated by backpropagation, i.e., by maximising the log-likelihood (LL), which is constructed from the probability in Equation 2, where i denotes the different alternatives.

$$Z_i = f(W * x + b) \tag{1}$$

$$P(class\ i|x) = \frac{e^{Z_i}}{\Sigma_i\ e^{Z_i}} \tag{2}$$

Figure 1 shows the structure of a recurrent neural network[1]. The difference between simple neural networks and RNN is that there are multiple outputs, and hidden layer at each time step depends on the previous time step as well. Thus, back-propagation through time is carried out. Therefore, the hidden layer at time t is also impacted by the time layer at

[1] We have plotted a one-to-many configuration for the RNN, which implies there is one input and multiple outputs. This configuration is most suitable for predicting activity sequences. There are other configurations as well, such as many-to-one, many-to-many, etc.

time $t-1$. In order to estimate the RNN, the loss function is similar i.e., to maximise LL of all of the choices depending on the previous choices. It is often beneficial to process the output sequences in both directions, i.e. information is passed through backwards and forwards, which leads to the bidirectional recurrent neural network.

Images/RNNs.png

Figure 1: Structure for a recurrent neural network

One of the issues in standard RNN is that there is a vanishing gradient problem, meaning the network forgets the initial association in outputs or inputs when the sequence is long. This has led to the development of long short-term memory (LSTM) networks, which are widely favoured because of their ability to retain historical information and handle long-term dependencies [Hiriyannaiah et al., 2020]. LSTM networks consist of three units inside a neuron, i.e., input, output, and forget gates. The purpose of the forget gate is to store information for a long period of time, even though the information is not passed directly to the output of the next layer. The LSTM unit also includes a cell that stores the accumulated information over arbitrary time intervals. As a result, they are more capable of capturing non-linear and dynamic patterns in sequential data, leading to strong predictive performance. [Graves, 2012]. Therefore, we employed RNN models while empirically evaluating the predictive performance of various RNN architectures, including standard RNN, LSTM, and bidirectional LSTM. For the sake of brevity, we have avoided adding details of LSTM and RNN; however, interested readers are encouraged to refer to works by Graves [2012] and Koushik et al. [2023].

To explicitly model time-use using recurrent neural networks, we discretise daily time-use patterns (i.e. 24 hours in a day) into fixed time-length intervals, for instance, in 10-minute intervals. Mathematically, the time-length can be described as Δt^* (10 minutes), and the corresponding total number of time-steps t range from 1 to T (where $T = \frac{24 \text{ hours}}{\Delta t^*} = 144$ in the case of 10-minute fixed intervals). This has been visualised in Figure 2.The total number of possible activities is denoted by k, and the dependent variable is a sequence of activities performed in a day, i.e. $Y_{activity\ schedule} = (y_{1,k}, y_{2,k}, \ldots, y_{T,k})$, where $y_{t,k}$ is 1 if activity k is performed at time-step t and 0 otherwise. The RNN model, and its variants including LSTM, essentially represent a function $Y_{activity\ schedule} = f(X, \phi)$ where X are the explanatory variables which govern time-use, for example, occupation, and type of day (weekday and weekend), and ϕ are the estimated model parameters. The RNN therefore predicts the probability distribution $p_{t,k}$ over all possible activities k at each time step t, i.e. $p_{t,k} = P(\text{activity } k \text{ at time-step } t \mid X, \phi)$, where the probabilities for each activity is from 0 to 1, i.e. $0 \leq p_{t,k} \leq 1 \forall k$ and the probabilities sum to 1 for each time-step, $\sum_{k=1}^{k} p_{t,k} = 1$

Images/Time_steps.png

Figure 2: Dependent variables in RNN and MDCEV models

There are at least two ways in which the output from RNN could be further used. First, the predicted probabilities of observing each activity at each time step can be aggregated for the complete day, which yields the expected time-use or time duration for each activity across a day (as in MDCEV models). Mathematically, the expected time-use for activity k across a day would be the $\mathbb{E}(\text{time-use}_k) = \sum_{t=1}^{T} p_{t,k} \cdot \Delta t^*$. It must be noted that the total time-use for different activities would sum up to 24 hours as the discretised activities are for a complete day. Second, the predicted probabilities at each time step can be used to generate the daily activity schedules for individuals. To ensure these sequences are behaviourally realistic, a microsimulation framework combined with behavioural rules would be required. For example, the framework should include travel episodes between spatially distant locations (e.g., commuting between work and leisure activities), enforce minimum activity durations, and prevent implausible switching between activities. We believe that generating activity schedules merits a separate study; hence, in this study, we restrict our analysis to comparing the total activity durations predicted by RNNs with those from MDCEV models.

To train the recurrent neural network, there is a need to select hyper-parameters of the recurrent neural network. Hyper-parameters can be defined as tools which affect the learning mechanism of the neural network [Goodfellow et al., 2016] and are typically selected within a subset of the training data, which is called the validation dataset. The hyperparameters in neural networks include the neural network architecture, regularisation, dropout layer, type of optimiser, number of iterations, learning rate, etc. Further, ML models do not have an exact solution as in the case of econometric models. Therefore, it is a recommended practice in the literature to carry out multiple training under different initialisation settings to reduce the risk of model non-identification and local irregularities [Wang et al., 2020a].

Even though ML algorithms are black boxes, there are some methods which can be used to explain the results of a recurrent neural network, such as feature importance plots, Partial Dependence Plots, SHAP values, LIME [Molnar, 2025]. These techniques are model agnostic, i.e., can be used for any ML technique, and are applied after the ML model has been trained. However, most of the techniques are not readily transferable to recurrent neural networks, as instead of one output there is a sequence of outputs. Nevertheless, feature importance plots and marginal effects are used to explain the recurrent neural network model.

2.2.2 Multiple discrete continuous extreme value

Multiple discrete continuous extreme value (MDCEV) models, first introduced by Bhat [2005], are normally used to model choice scenarios where decision-makers can consume different alternatives or goods simultaneously. Typical cases include modelling time-use, where people decide to spend time on various activities, car ownership and usage, consumption of different brands, etc. MDCEV models are consistent with the random utility theory which states that people choose alternatives where they find the maximum utility.

In MDCEV models, the direct utility of consuming goods k can be written as Equation 3, where x_k is the consumption of quantity x for good k (i.e. time duration x_k spent on activity k), α is a satiation parameters, γ is a translation parameter, and ψ is the baseline marginal utility at zero consumption [Bhat, 2008]. The utility function is further subjected to a budget constraint mentioned in Equation 4 where E is the total expenditure, x_k is the amount consumed for good k, and p is the unit price associated with good k.

$$U(x) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \psi_k \left(\left(\frac{x_k}{\gamma_k} + 1 \right)^{\alpha_k} - 1 \right) \tag{3}$$

$$E = \sum_{k=1}^{K} x_k * p_k \tag{4}$$

$$\psi_k = \exp(\beta * z_k \cdot \varepsilon_k) \tag{5}$$

ψ_k is further parametrised to represent deterministic utility where z_k represents the explanatory variables and the corresponding β parameters, and ε_k represents a Gumbel (0,1) identically and independently distributed error term as mentioned in Equation 5. To find the optimal allocation of budget E, a Lagrangian is formed along with the Kuhn-Trucker conditions. The probability for consumption of goods x is derived in Equation 6 where $f_i = \frac{1-\alpha_i}{x_i+\gamma_i}$.

$$P(x_1, x_2, .., x_m) = \frac{1}{p_1} \cdot \frac{1}{\sigma^{M-1}} \left(\prod_{i=1}^{M} f_i \right) \left(\prod_{i=1}^{M} \frac{p_i}{f_i} \right) \left(\frac{\prod_{i=1}^{M} e^{V_i/\sigma}}{(\sum_{k=1}^{K} e^{V_i/\sigma})^M} \right) (M-1)! \tag{6}$$

In the case, where there is an outside good, i.e. a good which is always consumed, there can be adjustments in the utility form. Typically, time-use studies have an outside good, which is personal maintenance as people always carry out some maintenance activities such as sleeping, eating, etc. For the outside good, the gamma parameter is zero, which negates the possibility of a corner solution or zero consumption. Usually, there are issues in the empirical identification and typically only one satiation parameter, i.e. either the alpha or the gamma is estimated [Bhat, 2008]. While MDCEV models can predict both the probability that an individual participates in a time-use activity and the amount of time allocated to each activity, in our study, we focus solely on the predicted time-use for each activity over a day, $\mathbb{E}(\text{time-use}_k)$, which is equivalent to one of the outputs of the RNN model.

2.3 Comparison of RNN and MDCEV models

Despite the fundamental differences in how the dependent variable is modelled in RNN and MDCEV frameworks, it is possible to carry out a meaningful comparison of expected time-use durations based on predictive performance and marginal effects. As shown in Figure 2, the dependent variables in RNN models constitutes a sequence of activities performed at each time step, whereas the MDCEV models the total time-use for each activity across the complete day. MDCEV models provide distinct economic interpretations (in terms of satiation parameters, covariates, and baseline utility), but these can not be obtained for the RNN. In contrast, RNNs have the distinct ability to estimate the probability of observing an activity at each time step, which can further be used to generate individual activity schedules. Since, RNN can be used to predict time-use for each activity across a complete day (which is similar to the outputs of the MDCEV as observed in Figure 2), it is possible to compare both models based on two metrics - (1) predictive performance on a testing dataset and (2) marginal effects of the explanatory variables.

In terms of comparing the predictive performance, we compare the performance at a sample level and at an individual or observation level. On the sample level, we compare the differences in time-use across k different activities using the mean absolute percentage error (MAPE) as shown in Equation 7, where O_k and P_k is the observed and predicted time-use for each activity k.

$$\text{MAPE} = \sum_{k=1}^{K} \left| \frac{O_k - P_k}{O_k} \right| \times 100 \tag{7}$$

Additionally, the differences in root mean square error values (for time-use) between the models at an individual level and across different sub-samples in the testing dataset, can be assessed to understand the comparative predictive performances (See Equation 8).

$$RMSE_n = \sqrt{\frac{1}{K} * \sum_{k=1}^{K} (O_{k,n} - P_{k,n})^2} \quad (8)$$

where, $O_{n,k}$ and $P_{n,k}$ are observed and predicted time-use values across different activities k for each individual n, respectively. It must be noted the RMSE values are for each individual, whereas the MAPE values are at the sample level.

In the context of marginal effects, we ensured comparability by estimating identical exogenous variables, including covariates that influence time-use decisions across both models. We calculate the marginal effects for continuous variables to assess the variation in time-use patterns by introducing a 10% increase in the explanatory variables relative to the baseline prediction within the testing dataset. In contrast, for categorical variables, the marginal effect corresponds to the difference in time-use patterns (keeping the rest constant) if the entire testing dataset is assigned to a specific category of that variable vis-a-vis other categories. For more details regarding marginal effects for continuous and categorical variables, readers can refer to Hensher et al. [2015].

3 Data

In this study, a repeated cross-sectional time-use diary, which has been deposited in the UK Data Service and collected by Gershuny et al. [2022] between the year 2016 to 2021 across 6 waves has been used. The data includes pre-COVID data, i.e. data collected in 2016, during the first lockdown in May 2020, in August 2020 when the lockdown was relaxed, in November 2020 during the second lockdown, during the third lockdown in January 2021 and after the end of the lockdown in August 2021. The data includes demographic features, and diaries which note activities, location and enjoyment for every 10-minute interval in a day. All of the time-use diaries were filled by people over the age of 18 and were filled by each respondent for 2 days. The data is representative of the UK in terms of gender and social class. The time-use diaries have been collected based on a Click and Drag Diary Instrument Approach, which is a user-friendly and much easier approach compared to other time-use collection methods [Sullivan et al., 2020]. In a Click and Drag Diary Instrument Approach, people fill out the time-use diary digitally, using a cursor to select an activity and drag it till the time when they end the activity. In this survey, there are 36 predefined activity categories, which, for ease of analysis in this study, have been summarised into 9 categories. In this survey, the time slots have been defined in 10-minute intervals starting from 4 am for the next 24 hours. This led to a total of 144 different time steps (6 slots in an hour for 24 hours).

Before carrying out the modelling, the data were pre-processed. Respondents who did not disclose their household income, household size and marital status were removed. Further, people who reported less than 4 activity episodes in the original activity diaries were removed since they are considered as ‘lousy diaries’ [Sullivan et al., 2020]. Typically, such people carry out a single activity for a day. Further, there were a few observations where people did not perform a personal maintenance activity, such as eating, sleeping, or resting. These observations were also removed. All of these changes led to a reduction in the total number of observations from 6,896 to 6,305.

Table 2 shows the average time in each activity across the survey waves. Further, Figure 3 shows the COVID cases in the UK when the surveys were carried out. Evidently, the COVID-19 pandemic brought notable shifts in daily time-use patterns. Before the pandemic (Wave 1), average travel time was 1.18 hours per day, which dropped to just 0.51 hours during the first lockdown (Wave 2) and remained low in subsequent lockdowns (e.g., 0.53 hours in Wave 5), reflecting strict mobility restrictions. However, after the lifting of restrictions (survey wave 6), the time spent on travelling has increased. Work hours remained relatively stable, ranging from 2.36 to 3.12 hours, but education time declined from 0.13 hours pre-COVID to as low as 0.05 hours (Wave 3 and 6), likely due to school disruptions. The inferences regarding work hours have to be done with necessary caution, as it depends on several other attributes such as the employment status of the individuals, working vis-a-vis non-working days, etc. The other noteworthy observation is that leisure and social activities increased during lockdowns, peaking at 5.52 hours and 2.29 hours, respectively, suggesting a shift toward at-home entertainment and virtual socialising. Maintenance chores consistently rose, from 10.88 hours pre-COVID to 11.64 hours in Wave 5, reflecting increased time at home. Although some activities began to rebound as restrictions eased, patterns such as reduced travel and increased home-based tasks persisted post-lockdown.

Table 2 also shows the association of selected explanatory variables on the average time-use. It can be observed that compared to males, females spend more time on chores and less time on paid work. There are obvious differences in the time spent on work, education, chores, and leisure on the weekend. We can also observe that employment status impacts the time duration in work, education, maintenance, and chores. People with jobs tend to spend more time at work. Similarly, homemakers and full-time carers spend more time on chores and maintenance activities.

Table 2: Average time duration (in hours)

Explanatory factors	Travel	Work	Education	Shopping	Maintenance	Chores	Leisure	Social	Other	Total diaries
Survey wave										
1 (pre-COVID data 2016)	1.18	2.84	0.13	0.58	10.88	1.87	4.72	1.59	0.21	911
2 (1st lockdown May 2020)	0.51	2.38	0.09	0.34	11.41	2.29	5.51	1.28	0.18	922
3 (Ease of restrictions August 2020)	0.74	3.12	0.05	0.53	11.04	2.07	4.72	1.57	0.16	901
4 (2nd lockdown November 2020)	0.64	3.06	0.09	0.45	11.18	1.86	5.47	1.11	0.15	1254
5 (3rd lockdown January 2021)	0.53	2.60	0.12	0.44	11.64	2.00	5.52	1.05	0.11	1145
6 (End of lockdown August 2021)	1.01	2.36	0.05	0.57	11.13	2.06	5.17	1.46	0.18	1172
Gender										
Male	0.82	3.09	0.05	0.48	11.04	1.59	5.61	1.17	0.16	3319
Female	0.70	2.32	0.14	0.49	11.43	2.49	4.77	1.49	0.17	2986
Day										
Weekday	0.77	3.49	0.12	0.46	11.01	1.91	4.88	1.21	0.16	4260
Weekend	0.74	1.13	0.04	0.53	11.67	2.25	5.91	1.56	0.18	2045
Employment Status										
Employed	0.79	3.91	0.06	0.42	11.05	1.76	4.73	1.17	0.11	4264
Student	0.58	0.89	1.38	0.33	11.85	1.19	5.79	1.85	0.14	209
Retired	0.86	0.11	0.00	0.68	11.43	2.25	6.85	1.57	0.26	1059
Unemployed	0.59	0.15	0.03	0.64	11.99	2.09	6.35	1.83	0.34	387
Homemaker or carer	0.48	0.42	0.02	0.60	11.51	4.56	4.59	1.55	0.27	386
Marital status										
Single	0.73	2.59	0.17	0.49	11.27	1.57	5.67	1.34	0.17	2270
Married/living with partner	0.78	2.80	0.04	0.48	11.20	2.27	4.96	1.32	0.16	4035
Presence of kids in household										
No	0.72	2.39	0.08	0.51	11.43	1.73	5.66	1.30	0.17	4367
Yes	0.85	3.47	0.11	0.43	10.76	2.65	4.20	1.37	0.15	1938
Age										
Less than 30	0.72	2.94	0.39	0.31	11.63	1.60	4.77	1.54	0.09	1007

Explanatory factors	Travel	Work	Education	Shopping	Maintenance	Chores	Leisure	Social	Other	Total diaries
Between 30 and 65	0.77	3.08	0.04	0.49	11.07	2.09	5.04	1.25	0.17	4486
Greater than 65	0.79	0.50	0.00	0.65	11.56	2.10	6.70	1.46	0.24	812
Annual household income										
Less than £10,000	0.72	0.77	0.24	0.66	11.17	2.32	6.22	1.47	0.42	316
£10,000 to £19,999	0.69	1.37	0.07	0.56	11.30	2.30	6.11	1.36	0.24	881
£20,000 to £29,999	0.75	2.40	0.10	0.51	11.34	2.02	5.39	1.32	0.16	1284
£30,000 to £39,999	0.78	2.80	0.07	0.52	11.18	2.11	4.97	1.43	0.14	1058
£40,000 to £49,999	0.72	3.45	0.06	0.45	11.13	1.85	5.01	1.21	0.11	865
£50,000 to £59,999	0.77	3.05	0.08	0.39	11.39	1.90	5.05	1.26	0.11	510
£60,000 to £69,999	0.84	3.51	0.13	0.39	11.24	1.91	4.61	1.26	0.09	416
£70,000 to £79,999	0.71	3.36	0.09	0.33	11.15	1.95	5.00	1.28	0.13	287
£80,000 to £89,999	0.84	3.31	0.01	0.42	10.90	2.69	3.90	1.69	0.23	208
£90,000 to £100,000	0.65	4.14	0.03	0.53	11.54	1.36	4.89	0.79	0.07	178
Over £100,000	1.01	4.30	0.17	0.37	10.80	1.34	4.60	1.32	0.10	302

4 Model development and analysis

This section first describes the development of the recurrent neural network and the multiple discrete-continuous extreme value model, followed by the comparison methodology mentioned in Section 2.3. To carry out the study, the dataset was randomly split into a training and testing dataset, where 80% of the data was for training (5,042 days of observations) and the rest for testing (1,263 days). The split was performed at the individual level to ensure that all observations from a given individual were assigned to the same subset.

4.1 Recurrent neural networks

In order to train the recurrent neural network models, 20% of the training dataset was set aside as a validation dataset to find the hyperparameters. Some of the hyperparameters were fixed e.g., the activation function was set to be hyperbolic tangent (tanh function), and the optimiser was set to Adaptive Moment Estimation (ADAM) as carried in most research in the transport literature [Wang et al., 2019, Han et al., 2022]. The last layer of the RNN was set to be a sigmoid function which enables the model to return the probability of the activities. The number of epochs or iterations was selected by early stopping, i.e. the iterations were stopped when the average log-likelihood did not decrease by 0.001 for 3 iterations. The hyperparameter were selected by comparing the validation scores (i.e. the average negative log likelihood scores) using Keras Random Search [O'Malley et al., 2019]. Each hyperparameter search was repeated 5 times under different initialisation settings to reduce the risk of local irregularity. All of the models were estimated using Keras [Chollet et al., 2015]. The results from the hyperparameter search can be viewed in Table 3.

The optimum neural network architecture was found out to be three hidden layers with 42 neurons in each layer and no dropout layer. Further, the architecture type for the neural network is bidirectional LSTM, which is similar to the best-performing model of Koushik et al. [2023]. Figure 4 shows the learning curve of the recurrent neural network. It was observed that by 10 iterations, the validation accuracy started to decrease, and there was no improvement in the validation dataset; hence, the total number of iterations was set to be 10. This value is quite less compared to other studies which use neural networks or recurrent neural networks; however, the number of iterations required depends on the dataset and the particular algorithm. To calculate the prediction error and marginal effects, results under 50 different initialisation settings were averaged. The input features considered in the model included the wave in which the data was collected, and socio-demographic features; this will later be discussed in the comparison of explanatory variables. The output of the RNN model is the probability at each time step, i.e. at every 10-minute interval for the 9 different time-use categories. Therefore, the output for the training and testing dataset is a matrix of $5042 \times 144 \times 9$ (observations$\times$timesteps$\times$activities) and $1263 \times 144 \times 9$, respectively.

Table 3: Hyperparameter grid

Hyperparameter name	Range	Selected
Type of layer	RNN, LSTM, Bidirectional LSTM	Bidirectional LSTM
Number of layers	1,2,3	3
Number of neurons	15 to 60 in steps of 3	42
Dropout layer	0, 0.1, 0.2, 0.3, 0.4, 0.5	0

Regardless the main objective of the study is to compare the time-use, the observed activity schedule and the predicted market shares, calculated by aggregating the predicted probabilities, for each time step in the testing dataset is plotted[2] in Figure 5. It can be observed

[2]It must be noted that it is not possible to construct a similar figure for the MDCEV model since MDCEV models only predict aggregated time-use or durations during a day.

Images/COVID_timeuse.png

Figure 3: Number of COVID cases in UK during survey waves [Wright, 2021]

Images/Learning_curve.png

Figure 4: Learning Curve for selected hyperparameter

that the recurrent neural network has managed to learn the main temporal relationship in activity scheduling. Both in the observed and the predicted market shares, it can be observed that most people work between 9:30 am to 5 pm, after which the probability of carrying out leisure activities increases. As expected in the observed diaries, the pattern is coarser. For instance, it can be observed that there is a dip in working activity between 1 pm and 2 pm as people tend to get lunch and hence perform other activities. Similarly, there is less variation in leisure activity.

Even though it is not possible to relate the impact of each explanatory variable on the choice probability in the RNN, the feature importance graph plotted in Figure 6 shows the relative size of how different explanatory variables impact the RNN model. It can be observed that the type of day is the most important factor which is intuitive. Other important variables include age and employment status. Gender and household size also impact RNN performance. Quite surprisingly the waves when the survey was carried out it is not an important explanatory factor.

Images/Test_Actual.png

(a)

Images/Test_RNN.png

Images/Feature Importance.png

Figure 6: Feature importance plot for the trained recurrent neural network

4.2 Multiple discrete continuous extreme value model

To estimate the MDCEV model, Apollo, the choice modelling library in R, was used [Hess and Palma, 2019]. To select the model specification, various specifications were tested. It was observed that the best goodness of fit and behavioural intuition was in the case where gamma (γ) parameters were estimated for all of the goods except for the outside good, i.e. personal maintenance. The alpha (α) parameter was set to approach zero, implying a gamma (γ) profile for the MDCEV model. Further, shifts (Δ) in the delta parameter or the intercepts of the baseline marginal utility (δ) for each wave were considered to indicate if a wave is statistically different compared to the base survey wave, i.e. the pre-COVID wave.

Table 4 shows the results of the estimated MDCEV. The sigma (σ) variable was observed to be 0.16. The γ parameter was found to be the highest for the work and education activities which indicates that individuals carry out this activity for the longest once selected. This is intuitive as people tend to work and spend most of the time at work or in education and previous studies also indicate that obligatory activities have a higher gamma parameter [Bhat et al., 2006]. The baseline marginal utility or the delta parameters δ for all activities are negative, which indicates that compared to the personal maintenance activity, there is a lower selection of all other activities.

The estimates indicate that compared to pre-pandemic levels there is a decrease in travelling in all of the survey waves at 95% confidence level which indicates that reduction in travelling due to work-from-home, changes in shopping behaviour have "sticked" throughout the pandemic [de Grignon Perez et al., 2022]. This is intuitive as travel was the most impacted activity during the COVID-19 pandemic. There is a lower shift parameter for the third and fourth survey waves compared to the second and fifth survey waves, which is intuitive as the first lockdown and the third lockdown were quite restrictive and severe compared to the second lockdown in the UK.

The baseline marginal utility for working has also decreased significantly in the second, third and the fifth survey wave at 95% confidence level compared to the first survey.

This can be attributed to the fact that in the second and the third survey wave, many businesses were closed. However, in the second lockdown or the fourth survey wave, there is no statistically significant difference compared to pre-pandemic levels possibly due to the decrease in the number of persons furloughed in the UK [de Grignon Perez et al., 2022]. In the fifth survey wave, or the third lockdown, there is a statistically significant difference in working compared to pre-pandemic levels which can be due to the increase in the number of cases (Figure 3) and the tighter restrictions in the UK. Finally, there is no statistically significant difference at 95% confidence level in working after the ease of COVID-related restrictions compared to pre-pandemic conditions. These results are similar to time-use survey analysed in Switzerland by Winkler et al. [2024]. It can be observed that education has not significantly altered compared to base during the pandemic except for the fourth survey wave which was held during typical summer breaks in the UK. The baseline marginal preference for shopping has decreased significantly in the lockdowns compared to pre-pandemic levels; however, shopping is not statistically different in the third and sixth survey waves when there were no restrictions.

It was observed in Table 2 that on average people have spent more time in chores across the pandemic, but it can be observed that except for the first lockdown, the delta δ parameter is not statistically significant. This is quite surprising as we would have expected that there would be more chores due to additional time spent at home, child caring responsibilities Giurge et al. [2021]. However, the results indicate that the utility of carrying out chores is impacted more by socio-demographic features rather than COVID-19 and lockdowns and is inline with de Grignon Perez et al. [2022]. It can also be established there has been an increased participation in leisure activities compared to pre-pandemic data, which is possible due to a reduction in travelling and socialising activity. Participation in social activities has decreased significantly across the pandemic; however, it is not statistically significant for the sixth wave. This indicates that at the end of the lockdown, with reductions in COVID cases and with the widespread vaccinations, people have increased their social interaction.

In terms of the socio-demographic variables, it can be observed that there is a gendered context in time-use as compared to males, females are less likely to travel and participate in paid work and leisure activities, and more likely to carry out chores and socialise at a 99% confidence level. This is in line with the findings of other researchers who have reported that due to social norms, the major shares of the chores are conducted by females [Collins et al., 2021, Giurge et al., 2021]. Age also has a significant impact on most of the activities. As expected, we observe a negative coefficient for the weekend variable to participate in work and education activities. Similarly, people are more likely to socialise and spend time relaxing on the weekends. An increase in the number of children in the household leads to an increase in chores and a reduction in leisure activities. With an increase in income, people tend to travel more and carry out less shopping and chores. It can also be observed that homemakers and people with caring responsibilities carry out more chores.

Table 4: MDCEV model estimates

Parameters	**Travel**		**Work**		**Education**		**Shopping**		**Chores**		**Leisure**		**Social**		**Others**	
	Estimate	Rob t stat	Estimate	Rob t stat	Estimate	Rob t stat	Estimate	Rob t stat	Estimate	Rob t stat	Estimate	Rob t stat	Estimate	Rob t stat	Estimate	Rob t stat
γ	1.96	22.60	17.17	16.42	7.49	7.49	2.21	20.22	2.62	19.39	4.24	19.21	3.81	20.95	3.80	9.63
δ base (wave 1)	-2.84	-76	-3.20	-51	-2.80	-36	-3.06	-42	-2.92	-57	-2.67	-60	-2.93	-76	-3.25	-64
Δ wave 2	-0.21	-9.48	-0.13	-5.47	-0.04	*-0.64*	-0.09	-4.02	0.01	*0.52*	0.01	0.35	-0.06	-3.24	-0.09	-2.08
Δ wave 3	-0.11	-5.47	-0.04	-2.03	-0.10	*-1.43*	-0.01	*-0.47*	0.02	*1.35*	0.00	*-0.17*	-0.01	*-0.48*	-0.04	*-1.05*
Δ wave 4	-0.14	-7.66	-0.03	*-1.50*	-0.02	*-0.29*	-0.04	-2.02	0.00	*0.13*	0.01	*0.97*	-0.07	-4.19	-0.04	*-1.22*
Δ wave 5	-0.21	-10.4	-0.07	-3.51	0.03	*0.56*	-0.10	-5.08	-0.03	-1.96	-0.01	*-0.68*	-0.12	-6.71	-0.12	-3.18
Δ wave 6	-0.05	-2.84	-0.02	*-0.93*	0.01	*0.14*	-0.02	*-0.74*	0.00	*0.26*	0.00	*-0.24*	-0.03	*-1.47*	-0.03	*-0.97*
β female	-0.05	-4.45	-0.09	-7.22					0.06	6.31	-0.06	-6.34	0.02	1.76		
β age			0.00	-6.52	-0.02	-8.57	0.00	6.83	0.00	6.50	0.00	4.99				
β hhsize									0.01	2.16			0.02	4.09		
β married											-0.03	-3.00	-0.01	*-1.01*		
β weekend	-0.06	-7.48	-0.26	-16.8	-0.23	-5.45					0.01	1.90				
β kids	0.03	5.13							0.03	3.95	-0.01	-1.97				
β income			0.03	3.68			-0.01	*-1.43*	-0.02	-2.76	-0.01	*-1.53*				
β employed	0.07	7.28	0.44	14.30												
β homemaker									0.11	6.01						

	Estimate	Rob t test
σ	0.16	26.59
Number of parameters	86	
LL (start)	-114,028.2	
LL (final)	-81,596.02	

Note: (1) The parameters which are statistically insignificant at 5% significance level have been *italicised*.

4.3 Comparison of RNN and MDCEV models

In this section, the outputs of the RNN and the MDCEV model are compared as per the methodology described in Section 2.3.

4.3.1 Predictive performance

Table 5 shows the average observed and predicted time-use in the testing dataset for the different activities, i.e. at the sample level. It can be observed that both the RNN-LSTM and MDCEV models exhibit similar predictive performance with a mean absolute percentage error (MAPE) of 7.18% and 7.15%, respectively. There are slight differences in the prediction accuracy as work and chores activities are over-predicted by the RNN, whereas the MDCEV model over-predicts maintenance activities and social activities. MDCEV model, on the other hand, under-predicts work activity. The models perform similarly on some activities, for instance, both models under-predict education activities compared to the observed time-use. Nevertheless, there seems to be no significant change in adopting either a data-driven approach or the econometric model based on a random-utility-maximisation framework to predict the average time-use at an aggregate level in this dataset. To enhance the robustness of our comparison and comprehensively comprehend the variability across distinct training and testing datasets, we conducted a sensitivity analysis of predictive performance employing a 5-fold cross-validation dataset, as detailed in Table A2 (Refer to Appendix 5).

To better compare the predictive performance of the two models, the prediction error is compared at an observation level and for different sub-groups based on explanatory variables. The latter is similar to checking if the RNN model has learned the distinct features in the input, i.e. the explanatory variables, compared to the MDCEV model. Table 6 shows the prediction error, where RMSE values were calculated for each observation between the observed and predicted time-use and then the average and standard deviations are calculated. It can be observed that, on average, the prediction error of RNN is slightly lower compared to the MDCEV model; however, based on a two-sample two-tailed t-test, it can be asserted that there is no evidence to state that the predictive performance of RNN and MDCEV is statistically different at a 90% confidence level. These results are similar to previous studies in the transport literature where the performance of discrete choice models is similar to machine learning models [Ali et al., 2023, Hillel, 2019, Nam et al., 2017, Wang et al., 2020b]. The table also compares the prediction error for different subgroups. The comparison indicates that for all socio-demographic variables, including the different survey waves, gender, type of day, and employment status, the predictive performance of both models is similar. Therefore, it may be inferred that the RUM-based approach does not necessarily have inferior predictive abilities (at both aggregate and disaggregate levels) as compared to its data-driven counterparts (RNN-LSTM in this case).

That being said, there are distinct benefits of adopting either of the two models. The RNN model can be used to generate the activity schedules of individuals which can be quite useful for activity-based models. Whereas, the MDCEV model which is embedded in an econometric and statistical framework can be used to estimate the significance of variables which is very useful in explaining how time-use patterns have altered over the course of the pandemic as highlighted in Section 4.2, and also be used to derive useful econometric values such as values of leisure time [Kuriyama et al., 2020]. The similar predictive performance of both modelling frameworks in the testing dataset also suggests that time-use which is a complex decision model can be equally explained by a random utility maximisation framework and a flexible data-driven technique.

Table 5: Observed and predicted time-Use (in hours) for the testing dataset

Activity	Observed Time-Use	Predicted Time-Use	
		RNN	MDCEV
Travel	0.76	0.78	0.76
Work	2.81	2.90	2.70
Maintenance	11.12	11.05	11.30
Leisure	5.24	5.22	5.18
Education	0.12	0.08	0.08
Shopping	0.50	0.50	0.49
Social	1.31	1.28	1.34
Other	0.20	0.16	0.15
Chores	1.96	2.02	1.99
MAPE (%)		7.18	7.15

Table 6: Prediction error (root mean square error) for MDCEV and RNN model in testing dataset

Description	RNN		MDCEV		Observations	t value
	Mean	Std dev	Mean	Std dev		
Average	1.92	0.92	1.96	0.92	1263	-0.87
Subsamples/subgroups						
Employed individuals	1.96	0.93	2.00	0.91	899	-0.83
Home maker	1.78	0.95	1.85	1.02	78	-0.45
Retired	1.74	0.85	1.74	0.86	180	0.02
Student	2.17	1.06	2.19	1.08	41	-0.08
Female	1.87	0.88	1.90	0.88	599	-0.63
Weekend	2.01	0.93	2.01	0.94	402	-0.04
Survey wave 1	1.84	0.83	1.85	0.84	180	-0.12
Survey wave 2	2.04	0.89	2.11	0.86	183	-0.76
Survey wave 3	2.09	0.96	2.11	0.93	182	-0.25
Survey wave 4	1.82	0.85	1.85	0.84	252	-0.42
Survey wave 5	1.81	0.97	1.83	0.94	227	-0.20
Survey wave 6	1.98	0.99	2.02	1.02	239	-0.41

4.3.2 Forecasting under different scenarios: marginal effects

This section highlights the findings based on examination of all 117 marginal effects (9 activities × 13 attributes) and their standard deviations to understand any differences across results from MDCEV and RNN (See Table 7) Overall, most of the marginal effects, in terms of their signs, are similar for RNN and MDCEV. At the same time, the marginal effects of MDCEV are higher in magnitude compared to those of the RNN on 73 instances, while the opposite is true for the remaining instances. For instance, RNN predicts that in the survey wave 2, when COVID restrictions came into place, there is a 27.59% decrease in travelling, whereas the MDCEV model predicts a larger 32.96% decrease. However, it is difficult to ascertain which marginal effects are closer to the true data-generating process since we do not have access to the true marginal effects. Moreover, the close comparison indicates that both models have similar and intuitive substitution patterns for all statistically significant attributes. The rest 22 instances where the substitution pattern differs are all for statistically insignificant attributes, with 10 of them belonging to the 'Education' and 'Other' categories. This can possibly be due to these activities being the least chosen in the dataset (See Table 5).

Even in such cases where the substitution pattern differs, it can be seen that the standard errors of the effects are high, indicating that the signs could be similar. For instance, the RNN model predicts that, on average, time spent on work increased by 0.23% during Survey wave 3, which coincides with the easing of COVID-related restrictions in August 2020. However, the relatively large standard deviation of 3.33% suggests considerable uncertainty around this estimate, with the possibility that the true effect may be negative. In contrast, the MDCEV model estimates a 1.74% average decrease in work time during the same period, which appears to be more intuitive, as August is typically a holiday month for many people. Nevertheless, the MDCEV marginal effect has a large standard deviation of 3.86%, which highlights a high degree of uncertainty regarding the true impact. Similar trends are observed for other marginal effects for time spent on education, shopping, and other activities. Therefore, we conclude that both models have similar marginal effects, and we do not have any evidence to state which model yields more accurate marginal effects.

Table 7: Average marginal effects for RNN and MDCEV in percentage (standard deviations are noted in brackets)

	Travel	Work	Maintenance	Leisure	Education	Shopping	Social	Other	Chores
Recurrent neural network									
Survey 1	39.55 (4.02)	12.25 (3.59)	-4.36 (0.81)	-8.28 (1.07)	**5.64 (12.02)**	15.65 (3.75)	18.16 (4.38)	16.41 (5.92)	-4.48 (3.21)
Survey 2	-27.59 (2.56)	-21.67 (3.86)	2.26 (0.98)	8.54 (2.39)	-31.61 (10.36)	-10.57 (2.77)	**-4.97 (2.59)**	-1.22 (5.53)	14.83 (3.94)
Survey 3	-0.59 (2.43)	**0.23 (3.33)**	-0.86 (0.78)	-3.75 (1.47)	-26.37 (11.48)	7.02 (2.91)	18.35 (1.8)	4.25 (4.01)	1.88 (2.98)
Survey 4	-14.84 (2.91)	5.67 (2.34)	**0.76 (0.56)**	4.73 (1.47)	**29.03 (17.03)**	-9.76 (1.53)	-15.43 (1.98)	**-11.27 (5.19)**	-6.88 (1.82)
Survey 5	-23.24 (2.34)	4.21 (3.95)	2.22 (0.69)	6.16 (1.6)	14.61 (9.22)	-15.39 (2.38)	-20.84 (1.48)	-14.56 (4.09)	-7.7 (2.55)
Survey 6	30.28 (3.4)	5.78 (4.42)	-1.83 (0.71)	-8.44 (1.13)	**-12.14 (10.32)**	10.14 (1.13)	12.26 (2.67)	9.05 (4.1)	**1.23 (2.21)**
Female	-2.14 (2.72)	-12.18 (2.01)	1.15 (0.59)	-5.97 (1.07)	1.98 (5.34)	4.99 (2)	7.87 (2.59)	4.96 (2.6)	20.85 (1.65)
Number of kids	17.54 (3.57)	**7.37 (5.22)**	-1.69 (0.94)	-11.53 (2.38)	**2.52 (13.03)**	**3.47 (2.53)**	**2.99 (1.8)**	**8.22 (4.98)**	17.62 (5.05)
Household size	-5.01 (2.19)	-9.48 (2.95)	-0.14 (0.62)	**2.1 (0.97)**	-9.17 (4.3)	**2.69 (1.43)**	2.33 (2.03)	**0.15 (3.61)**	9.2 (2.39)
Income	**0.46 (0.44)**	1.85 (0.43)	**0.01 (0.09)**	-0.55 (0.17)	**1.24 (1.4)**	-1.28 (0.36)	0.31 (0.36)	**-1.65 (0.47)**	-1.23 (0.3)
Weekend	-2.48 (2.78)	-59 (4.18)	3.94 (0.88)	13.2 (1.97)	-48.77 (11.05)	12.48 (1.8)	20.1 (2.48)	12.84 (5.37)	14.85 (3.46)
Age	-1.05 (0.97)	-5.1 (1.07)	**0.2 (0.23)**	2.62 (0.52)	-13.79 (2.68)	3.82 (0.77)	-1.63 (0.78)	**0.22 (1.09)**	0.52 (1.32)
Employed	1.65 (3.21)	25.1 (1.68)	-0.94 (0.52)	-4.34 (1.31)	-26.8 (12.08)	-5.64 (4.56)	-8.78 (2.15)	-24.07 (3.59)	-10.52 (2.21)
Multiple discrete continuous extreme value									
Survey 1	58.49 (7.09)	13.02 (3.96)	-3.4 (0.87)	-9.31 (1.76)	**-7.52 (28.14)**	11.21 (7.29)	15.29 (5.46)	19.72 (17.27)	-10.83 (3.06)
Survey 2	-32.96 (4.66)	-27.87 (3.64)	2.83 (0.87)	9.69 (1.83)	-1.62 (28.71)	-17.25 (6.55)	**2.28 (5.09)**	-8.82 (15.17)	12.65 (3.27)
Survey 3	-2.2 (5.67)	**-1.74 (3.86)**	-1.14 (0.87)	-4.56 (1.78)	-44.74 (20.29)	14.35 (7.38)	19.3 (5.52)	2.58 (16.05)	6.68 (3.24)
Survey 4	-14.27 (5.31)	8.87 (3.9)	**-0.72 (0.87)**	3.29 (1.81)	**-7.06 (28.28)**	-0.72 (6.95)	-13.96 (4.77)	**4.12 (16.08)**	-1.46 (3.17)
Survey 5	-32.54 (4.72)	1.45 (3.81)	3.42 (0.89)	5.36 (1.79)	48.53 (36.41)	-19.17 (6.44)	-23.04 (4.49)	-21.63 (13.95)	-3.02 (3.16)
Survey 6	29.29 (6.45)	8.54 (3.92)	-1.71 (0.87)	-6.28 (1.77)	**5.8 (30.43)**	9.29 (7.25)	5.71 (5.23)	4.77 (16.16)	**-3.54 (3.14)**
Female	-5.01 (5.25)	-18.28 (3.72)	2.03 (0.88)	-5.99 (1.78)	5.4 (29.82)	7.55 (7.38)	11.91 (5.1)	7.66 (16.3)	20.16 (3.25)
Number of kids	12.78 (5.81)	**-1.62 (3.98)**	-0.43 (0.88)	-4.92 (1.76)	**-1.89 (28.7)**	**-2 (6.99)**	**-1.93 (4.9)**	**-2.16 (15.69)**	14.62 (3.25)
Household size	-1.51 (5.44)	-1.34 (3.97)	-0.39 (0.88)	**-1.09 (1.77)**	-1.72 (28.87)	**-1.72 (7.01)**	6.96 (5.12)	**-1.85 (15.69)**	3.4 (3.15)
Income	**-0.08 (5.48)**	1.47 (3.99)	**-0.02 (0.88)**	-0.39 (1.77)	**-0.07 (29.09)**	-0.57 (7.03)	0 (4.97)	**0.02 (15.8)**	-0.72 (3.12)
Weekend	-5.48 (5.18)	-63 (2.98)	4.35 (0.87)	14.48 (1.87)	-59.53 (16.17)	15.29 (7.6)	15.25 (5.31)	17.04 (16.83)	12.49 (3.11)
Age	-0.3 (5.46)	-6.77 (3.91)	**-0.09 (0.88)**	2.4 (1.78)	-25.02 (24.32)	5.43 (7.2)	-0.67 (4.95)	**-0.86 (15.81)**	3.81 (3.14)
Employed	3.73 (5.69)	29.07 (4.6)	-2.06 (0.87)	-5.48 (1.8)	-12.81 (26.14)	-8.41 (7.01)	-7.6 (4.89)	-8.02 (15.5)	-6.69 (3.09)

5 Conclusion

In this research, we use a repeated cross-section time-use data collected before and during the COVID-19 pandemic (2016-2020) across six survey waves to compare the recurrent neural network and multiple discrete-continuous extreme value models. It is useful to use time-use across the pandemic as there were major disruptions and variations across individual's time-use which makes it easier to check if the machine learning or data-driven model can adequately capture the differences observed across the survey waves. We find that for our dataset, the best architecture of the RNN model is the bidirectional LSTM model, which is similar to the study by Koushik et al. [2023]. It is observed that the RNN model can replicate the activity schedules on an unseen dataset as indicated in Figure 5 that clearly demonstrates the potential of using RNN to generate activity schedules.

In terms of the predictive performance of the estimated MDCEV and RNN model, we observe that there are no significant differences between the predictive performance of the models at a sample level on a testing dataset. We further compare the predictive error at an individual level but also observe that there is no evidence to state that one modelling framework is better than other. Further comparisons were also made to check if RNN model has better captured certain activities or socio-demographic features compared to the specified MDCEV model. However, for nearly all activities and socio-demographic features, there was no evidence to state that there is a statistically significant difference between the predictive performance of the models. This finding is in line with previous studies which assert that there is a limited improvement by adopting a data-driven approach in terms of predictive performance compared to an econometric approach [Ali et al., 2023, Hillel, 2019, Nam et al., 2017, Wang et al., 2020b]. It has been hypothesised that in the case of large number of explanatory variables and alternatives, machine learning models would capture the data-generating process quite efficiently and better compared to econometric models which rely on hand-specified utility functions [van Cranenburgh et al., 2022]. However, in our study, we do not find evidence to support such hypothesis, as the MDCEV model has a large number of explanatory variables (over 15) and 9 alternatives. The similar predictive performance also indicates that time-use decisions, which are a complex decision, can be equally predicted either using a data-driven approach, which relies on neural networks or using a multiple discrete continuous model, which is underpinned by a random utility maximisation framework. This is an interesting but expected finding, as there is only limited information about the data-generating process in the data, which can be exploited using any modelling framework. Furthermore, the study demonstrates that random-utility-maximisation approaches are not necessarily inferior to ML models in terms of predictive accuracy, contrary to popular belief. Nonetheless, this inference must not be interpreted as the 'usefulness' of the respective modelling techniques, since MDCEV and RNN offer distinct advantages in terms of interpretability and forecasting.

To assess whether the RNN model can reliably be used to forecast outcomes under different policy scenarios, we compare its marginal effects with those from the MDCEV model. We again find no significant differences between the two modelling frameworks. However, due to the absence of true marginal effects or a known data-generating process, it is not possible to determine which model provides more intuitive or accurate estimates. This is a limitation of our study. In other contexts, such as mode choice models, marginal effects are more interpretable; for example, the values of travel time can be extracted that have econometric and empirical benchmarks. In contrast, modelling time-use is inherently more complex, and such reference values or benchmarks are not readily available.

To summarise, MDCEV model relying on microeconomic theories offers high interpretability and estimates parameters with clear behavioural meaning, thereby enabling analysts to interpret preferences and trade-offs in time-use decisions. It is particularly useful for policy analysis due to its transparency and explanatory power. However, it relies on strong assumptions about non-chosen time-use alternatives, spatio-temporal budget constraints, and rational decision-making, which may not hold in complex real-world behaviours. RNN models, on the other hand, capture sequential and non-linear relationships in time-use data without predefined behavioural assumptions, excelling in practical implementation (prediction tasks) and also in generating intuitive activity schedules. However, their 'black-box' nature poses interpretation difficulties, requiring post-hoc analysis for extracting behavioural insights. While RNN save effort for manual model specification and may offer similar predictive performance and marginal effects compared to MDCEV models, their limited transparency can be seen as a drawback in policy-oriented applications where understanding behavioural mechanisms is crucial.

Nevertheless, to better generalise the findings, there should be more studies on different time-use datasets to compare machine learning algorithms and econometric models. It may be possible that with advanced variants of the MDCEV models, such as with random taste heterogeneity amongst individuals and inclusion of heteroscedastic error components, econometric models yield better prediction accuracies. In this study, we only compare the potential use of RNN for modelling time-use duration; however, a natural next step is to generate and assess activity schedules. Future work could also look at potential ways to embed the outputs from RNN in an agent-based modelling simulation such as MATSim. Further, the generated activity schedules can be compared with other activity scheduling models such as TASHA [Roorda et al., 2008], CUSTOM [Nurul Habib, 2018], and SALT [Hafezi et al., 2022]. Another aspect which can be further explored is the potential of using RNN to model dynamic activity or time-use patterns by considering the previous or partial time-use patterns of individuals. Future efforts to improve time-use or econometric models should aim to retain the interpretability offered by econometric models but use the extra flexibility offered by machine learning techniques or concepts such as long short-term memory.

Acknowledgement

This work was supported by ESRC AQM Grant Reference No ES/P000746/1 and UKRI Future Leaders Fellowship Next Generation Travel Behaviour Models' [Grant Number MR/T020423/1].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The time-use data used in this study is deposited in UK Data Service by Gershuny et al. [2022] as Time Use Survey 6-Wave Sequence across the COVID-19 Pandemic, 2016-2021 at http://doi.org/10.5255/UKDA-SN-8741-4.

References

Azam Ali, Arash Kalatian, and Charisma F. Choudhury. Comparing and contrasting choice model and machine learning techniques in the context of vehicle ownership decisions. *Transportation Research Part A: Policy and Practice*, 173:103727, 2023. doi:10.1016/j.tra.2023.103727.

Theo A Arentze and Harry JP Timmermans. A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7):613–633, 2004. doi:10.1016/j.trb.2002.10.001.

Kay W Axhausen and Tommy Gärling. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12(4):323–341, 1992. doi:10.1080/01441649208716826.

Chandra R. Bhat. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological*, 39(8):679–707, September 2005. ISSN 01912615. doi:10.1016/j.trb.2004.08.003.

Chandra R. Bhat. The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological*, 42(3):274–303, March 2008. ISSN 01912615. doi:10.1016/j.trb.2007.06.002.

Chandra R Bhat, Sivaramakrishnan Srinivasan, and Sudeshna Sen. A joint model for the perfect and imperfect substitute goods case: application to activity time-use decisions. *Transportation Research Part B: Methodological*, 40(10):827–850, 2006. doi:10.1016/j.trb.2005.08.004.

Ron N Buliung and Pavlos S Kanaroglou. Activity–travel behaviour research: conceptual issues, state of the art, and emerging perspectives on behavioural analysis and simulation modelling. *Transport Reviews*, 27(2):151–187, 2007. doi:10.1080/01441640600858649.

Marisol Castro, Chandra R Bhat, Ram M Pendyala, and Sergio R Jara-Díaz. Accommodating multiple constraints in the multiple discrete–continuous extreme value (MDCEV) choice model. *Transportation Research Part B: Methodological*, 46(6):729–743, 2012. doi:10.1016/j.trb.2012.02.005.

François Chollet et al. Keras. `https://keras.io`, 2015. Accessed: 2026-03-02.

Caitlyn Collins, Liana Christin Landivar, Leah Ruppanner, and William J Scarborough. COVID-19 and the gender gap in work hours. *Gender, Work & Organization*, 28:101–112, 2021. doi:10.1111/gwao.12506.

Joan L. de Grignon Perez, Oriel Sullivan, Marga Vega-Rapun, and Jonathan Gershuny. Big changes in daily activities across the course of the UK pandemic; but which of them will stick?, July 2022. URL `https://osf.io/preprints/socarxiv/h35cr_v1`. Preprint, Open Science Framework.

Jan Drchal, Michal Čertický, and Michal Jakob. Data-driven activity scheduler for agent-based mobility models. *Transportation Research Part C: Emerging Technologies*, 98:370–390, January 2019. ISSN 0968090X. doi:10.1016/j.trc.2018.12.002.

Jonathan Gershuny, Oriel Sullivan, Joan Lamote de Grignon Perez, and Marga Vega-Rapun. Centre for Time Use Research UK Time Use Survey 6-Wave Sequence across the COVID-19 Pandemic, 2016–2021. UK Data Service [data collection], 2022. URL `https://datacatalogue.ukdataservice.ac.uk/studies/study/8741#details`. Study Number: 8741.

Laura M Giurge, Ashley V Whillans, and Ayse Yemiscigil. A multicountry perspective on gender differences in time use during COVID-19. *Proceedings of the national academy of sciences*, 118(12), 2021. doi:10.1073/pnas.2018494118.

Nima Golshani, Ramin Shabanpour, Seyed Mehdi Mahmoudifard, Sybil Derrible, and Abolfazl Mohammadian. Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model. *Travel Behaviour and Society*, 10:21–32, 1 2018. ISSN 2214-367X. doi:10.1016/J.TBS.2017.09.003.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL http://www.deeplearningbook.org.

Alex Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer, Berlin, Heidelberg, 2012. URL https://link.springer.com/book/10.1007/978-3-642-24797-2.

Khandker M Nurul Habib. A random utility maximization (RUM) based dynamic activity scheduling model: Application in weekend activity scheduling. *Transportation*, 38(1): 123–151, 2011. doi:10.1007/s11116-010-9294-9.

Mohammad Hesam Hafezi, Naznin Sultana Daisy, Hugh Millward, and Lei Liu. Ensemble learning activity scheduler for activity based travel demand models. *Transportation Research Part C: Emerging Technologies*, 123:102972, 2021. doi:10.1016/j.trc.2021.102972.

Mohammad Hesam Hafezi, Naznin Sultana Daisy, Hugh Millward, and Lei Liu. Framework for development of the Scheduler for Activities, Locations, and Travel (SALT) model. *Transportmetrica A: Transport Science*, 18(2):248–280, 2022. doi:10.1080/23249935.2021.1921879.

Yafei Han, Francisco Camara Pereira, Moshe Ben-Akiva, and Christopher Zegras. A neural-embedded discrete choice model: Learning taste representation with strengthened interpretability. *Transportation Research Part B: Methodological*, 163:166–186, September 2022. ISSN 01912615. doi:10.1016/j.trb.2022.07.001.

David A. Hensher, John M. Rose, and William H. Greene. *Applied Choice Analysis*. Cambridge University Press, 2 edition, 2015. URL https://doi.org/10.1017/CBO9780511610356.

Stephane Hess and David Palma. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, 32: 100170, September 2019. ISSN 17555345. doi:10.1016/j.jocm.2019.100170.

Tim Hillel. *Understanding travel mode choice: A new approach for city scale simulation*. PhD thesis, University of Cambridge, 2019. URL https://www.repository.cam.ac.uk/handle/1810/293576.

Tim Hillel, Michel Bierlaire, Mohammed Z.E.B. Elshafie, and Ying Jin. A systematic review of machine learning classification methodologies for modelling passenger mode choice. *Journal of Choice Modelling*, 38:100221, March 2021. ISSN 17555345. doi:10.1016/j.jocm.2020.100221.

Srinidhi Hiriyannaiah, A. M.D. Srinivas, Gagan K. Shetty, G. M. Siddesh, and K. G. Srinivasa. A computationally intelligent agent for detecting fake news using generative adversarial networks. *Hybrid Computational Intelligence: Challenges and Applications*, pages 69–96, 1 2020. doi:10.1016/B978-0-12-818699-2.00004-4.

Mohammadjavad Javadinasr, Sina Asgharpour, Motahare Mohammadi, Abolfazl Mohammadian, and Joshua Auld. A comparative analysis between machine learning and econometric approaches for travel mode choice modeling. In *International Conference on Transportation and Development 2023*, pages 95–105, 2023. URL https://ascelibrary.org/doi/abs/10.1061/9780784484883.009.

Shivam Khaddar, Varun Varghese, Mahmudur Rahman Fatmi, and Makoto Chikaraishi. Jointly modelling activity start time, travel mode, companionship, and destination location choices. *Transportmetrica A: Transport Science*, pages 1–27, 2024. doi:10.1080/23249935.2024.2372025.

Anil Koushik, M Manoj, N Nezamuddin, and AP Prathosh. Activity Schedule Modeling Using Machine Learning. *Transportation Research Record*, 2677(8):01–23, 2023. doi:10.1177/03611981231155426.

Anil NP Koushik, M Manoj, and N Nezamuddin. Machine learning applications in activity-travel behaviour research: a review. *Transport Reviews*, 40(3):288–311, 2020. doi:10.1080/01441647.2019.1704307.

Koichi Kuriyama, Yasushi Shoji, and Takahiro Tsuge. The value of leisure time of weekends and long holidays: The multiple discrete–continuous extreme value (MDCEV) choice model with triple constraints. *Journal of Choice Modelling*, 37:100238, 2020. doi:10.1016/j.jocm.2020.100238.

Dongwoo Lee, Sybil Derrible, and Francisco Camara Pereira. Comparison of Four Types of Artificial Neural Network and a Multinomial Logit Model for Travel Mode Choice Modeling. *Transportation Research Record*, 2672:101–112, 12 2018. ISSN 21694052. doi:10.1177/0361198118796971.

Alix Lhéritier, Michael Bocamazo, Thierry Delahaye, and Rodrigo Acuna-Agost. Airline itinerary choice modeling using machine learning. *Journal of Choice Modelling*, 31:198–209, 2019. doi:10.1016/j.jocm.2018.02.002.

Jose Angel Martín-Baos, Julio Alberto López-Gómez, Luis Rodriguez-Benitez, Tim Hillel, and Ricardo García-Ródenas. A prediction and behavioural analysis of machine learning methods for modelling travel mode choice. *Transportation Research Part C: Emerging Technologies*, 156:104318, 11 2023. ISSN 0968-090X. doi:10.1016/J.TRC.2023.104318.

Christoph Molnar. *Interpretable Machine Learning*. 3 edition, 2025. ISBN 978-3-911578-03-5. URL `https://christophm.github.io/interpretable-ml-book`.

Daisik Nam, Hyunmyung Kim, Jaewoo Cho, and R Jayakrishnan. A model based on deep learning for predicting travel mode choice. In *Proceedings of the transportation research board 96th annual meeting transportation research board, Washington, DC, USA*, pages 8–12, 2017.

Khandker Nurul Habib. A comprehensive utility-based system of activity-travel scheduling options modelling (CUSTOM) for worker's daily activity scheduling processes. *Transportmetrica A: Transport Science*, 14(4):292–315, 2018. doi:10.1080/23249935.2017.1385656.

Tom O'Malley, Elie Bursztein, James Long, François Chollet, Haifeng Jin, Luca Invernizzi, et al. Keras tuner. `https://github.com/keras-team/keras-tuner`, 2019. Software. Retrieved August 2024.

Jasper Püschel, Lukas Barthelmes, Martin Kagerbauer, and Peter Vortisch. Comparison of discrete choice and machine learning models for simultaneous modeling of mobility tool ownership in agent-based travel demand models. *Transportation Research Record*, 2678(7):376–390, 2024. doi:10.1177/03611981231206175.

Matthew J Roorda, Eric J Miller, and Khandker MN Habib. Validation of TASHA: A 24-h activity scheduling microsimulation model. *Transportation Research Part A: Policy and Practice*, 42(2):360–375, 2008. doi:10.1016/j.tra.2007.10.004.

Patricio Salas, Rodrigo De la Fuente, Sebastian Astroza, and Juan Antonio Carrasco. A systematic comparative evaluation of machine learning classifiers and discrete choice models for travel mode choice in the presence of response heterogeneity. *Expert Systems with Applications*, 193:116253, 5 2022. ISSN 0957-4174. doi:10.1016/J.ESWA.2021.116253.

Shobhit Saxena, Abdul Rawoof Pinjari, Ananya Roy, and Rajesh Paleti. Multiple discrete-continuous choice models with bounds on consumptions. *Transportation Research Part A: Policy and Practice*, 149:237–265, 2021. doi:10.1016/j.tra.2021.03.016.

Oriel Sullivan, Jonathan Gershuny, Almudena Sevilla, Pierre Walthery, and Marga Vega-Rapun. Time use diary design for our times-an overview, presenting a Click-and-Drag Diary Instrument (CaDDI) for online application. *Journal of Time Use Research*, 2020. doi:10.32797/jtur-2020-1.

Sander van Cranenburgh, Shenhao Wang, Akshay Vij, Francisco Pereira, and Joan Walker. Choice modelling in the age of machine learning - Discussion paper. *Journal of Choice Modelling*, 42:100340, March 2022. ISSN 17555345. doi:10.1016/j.jocm.2021.100340.

Feilong Wang, Jingxing Wang, Jinzhou Cao, Cynthia Chen, and Xuegang (Jeff) Ban. Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example. *Transportation Research Part C: Emerging Technologies*, 105:183–202, August 2019. ISSN 0968090X. doi:10.1016/j.trc.2019.05.028.

Shenhao Wang, Qingyi Wang, and Jinhua Zhao. Deep neural networks for choice analysis: Extracting complete economic information for interpretation. *Transportation Research Part C: Emerging Technologies*, 118:102701, September 2020a. ISSN 0968090X. doi:10.1016/j.trc.2020.102701.

Shenhao Wang, Qingyi Wang, and Jinhua Zhao. Multitask learning deep neural networks to combine revealed and stated preference data. *Journal of Choice Modelling*, 37:100236, December 2020b. ISSN 17555345. doi:10.1016/j.jocm.2020.100236.

Shenhao Wang, Baichuan Mo, Yunhan Zheng, Stephane Hess, and Jinhua Zhao. Comparing hundreds of machine learning and discrete choice models for travel demand modeling: An empirical benchmark. *Transportation Research Part B: Methodological*, 190:103061, 12 2024. ISSN 0191-2615. doi:10.1016/J.TRB.2024.103061.

Caroline Winkler, Aupal Mondal, Raphael Mesaric, Katherine Asmussen, Chandra Bhat, and Kay W Axhausen. Investigation of shifts in time use behavior during the COVID-19 pandemic: A mixed multiple discrete-continuous extreme value approach with panel GPS data. *Transportation Research Record*, 2678(12):197–208, 2024. doi:10.1177/03611981221099918.

Melvin Wong and Bilal Farooq. A bi-partite generative model framework for analyzing and simulating large scale multiple discrete-continuous travel behaviour data. *Transportation Research Part C: Emerging Technologies*, 110:247–268, January 2020. ISSN 0968090X. doi:10.1016/j.trc.2019.11.022.

Katie Wright. Covid-19: UK in early stages of third wave - scientist. BBC News, May 2021. URL `https://www.bbc.co.uk/news/uk-57304515`. Accessed August 2024.

Rui Yao and Shlomo Bekhor. A variational autoencoder approach for choice set generation and implicit perception of alternatives in choice modeling. *Transportation Research Part B: Methodological*, 158:273–294, April 2022. ISSN 01912615. doi:10.1016/j.trb.2022.02.015.

Xilei Zhao, Xiang Yan, Alan Yu, and Pascal Van Hentenryck. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society*, 20:22–35, July 2020. ISSN 2214367X. doi:10.1016/j.tbs.2020.02.003.

Appendix

Sensitivity analysis

In this section, we present the results of a 5-fold cross-validation on the time-use dataset to assess the predictive performance of the RNN and MDCEV models across different test set configurations. On average (across all 5-folds), the RNN model yields a MAPE of 7.40%, while the MDCEV model achieves a slightly lower MAPE of 7.28%, suggesting no substantial difference in overall performance. However, some folds show notable variations as highlighted in Table A2. For instance, in Fold 2, RNN performs poorer with a MAPE of 9.25%, compared to 7.04% for MDCEV. Conversely, in Fold 4, RNN performs better with a MAPE of 7.02%, whereas MDCEV records a higher error of 9.40%. These results highlight that while the choice of the test set can influence model performance, on average, both models perform comparably similar.

Table A2: Predictive performance in testing dataset across 5-folds

Dataset (observations in testing data)	Description	Time-use (in hours)									MAPE (%)
		Travel	Work	Maintenance	Leisure	Education	Shopping	Social	Other	Chores	
Fold 1 (n = 1263)	Observed	0.76	2.81	11.12	5.24	0.12	0.50	1.31	0.20	1.96	-
	RNN	0.78	2.90	11.05	5.22	0.08	0.50	1.28	0.16	2.02	7.18
	MDCEV	0.76	2.70	11.30	5.18	0.08	0.49	1.34	0.15	1.99	7.15
Fold 2 (n = 1268)	Observed	0.77	2.89	11.23	5.24	0.09	0.51	1.34	0.12	1.81	-
	RNN	0.78	2.87	11.13	5.26	0.08	0.46	1.30	0.19	1.92	9.25
	MDCEV	0.76	2.69	11.27	5.26	0.09	0.48	1.33	0.16	1.94	7.04
Fold 3 (n = 1253)	Observed	0.76	2.81	11.12	5.24	0.12	0.50	1.31	0.20	1.96	-
	RNN	0.78	2.90	11.05	5.22	0.08	0.50	1.28	0.16	2.02	7.18
	MDCEV	0.75	2.49	11.40	5.25	0.09	0.51	1.36	0.17	1.98	6.54
Fold 4 (n = 1268)	Observed	0.77	2.72	11.29	5.07	0.08	0.45	1.25	0.18	2.20	-
	RNN	0.70	2.55	11.26	5.28	0.10	0.48	1.31	0.16	2.17	7.02
	MDCEV	0.76	2.50	11.35	5.30	0.10	0.50	1.36	0.16	1.96	9.40
Fold 5 (n = 1253)	Observed	0.77	2.59	11.34	5.11	0.08	0.47	1.35	0.18	2.11	-
	RNN	0.77	2.83	11.08	5.12	0.09	0.50	1.31	0.16	2.13	6.36
	MDCEV	0.76	2.62	11.31	5.23	0.10	0.50	1.35	0.16	1.98	6.29