September 21-23, 2019
The 18th Summer Course for Behavior Modeling in Transportation Networks
@The University of Tokyo

Urban science and behavioral informatics with AI/machine learning

Application of AI for travel behavior modelling in urban networks

Makoto Chikaraishi
Hiroshima University

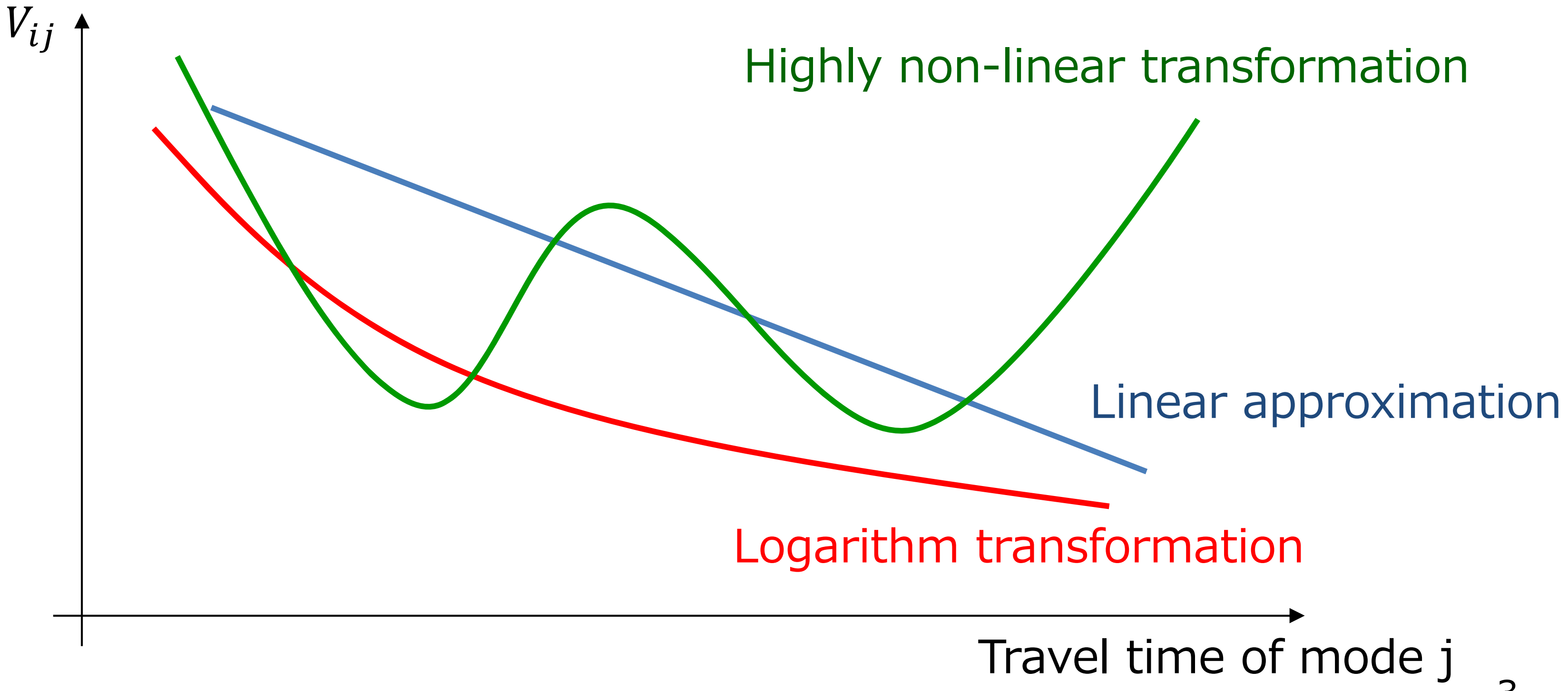
Problem setting

- Standard logit model: $P_{ij} = \frac{\exp(V_{ij})}{\sum_{j'=1}^{J} \exp(V_{ij'})}$
- The conventional form of V_{ij}:
  - Linear approximation (rooted to the Taylor's theorem)
  - Also known as a linear-in-parameter model
- Problem at hand:
  - Is there any better way to determine the functional form?
    - Obviously, taking into account the non-linearlity of V_{ij} would improve the goodness-of-fit.
    - What is the cost of doing that?

Problem setting

- Can we understand the non-linear transformation of V_{ij} logically?

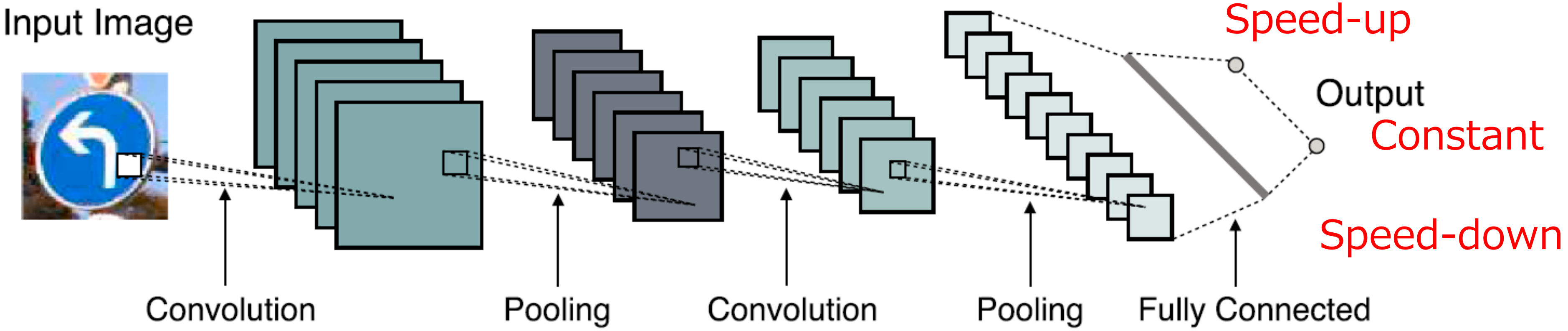
Example: contribution of travel time to mode/route choice model



Problem setting

- In what context we may NOT need to explain the impacts of each factor logically?

Example: contribution of eye movement to the speed choice



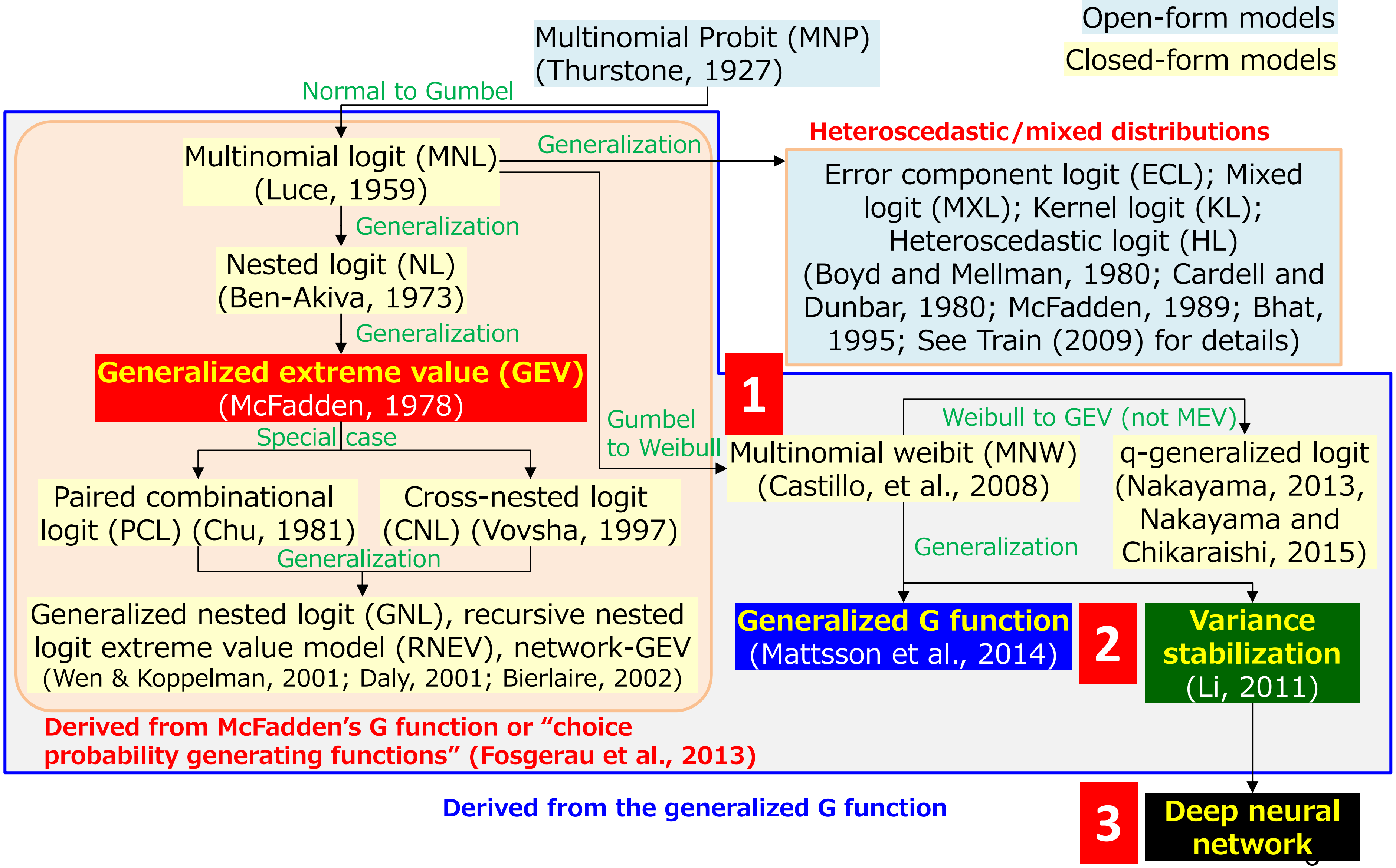
Adapted from Wang et al. (2019)

Contents

1. Logically explainable non-linear transformation
2. Spline-based non-linear transformation
3. NN (Neural network)-based non-linear transformation

Advanced discrete choice models

[based on Hato (2002)]



Open-form models

Closed-form models

Multinomial Probit (MNP)
(Thurstone, 1927)

Normal to Gumbel

Multinomial logit (MNL)
(Luce, 1959)

Generalization

Nested logit (NL)
(Ben-Akiva, 1973)

Generalization

Generalized extreme value (GEV)
(McFadden, 1978)

Special case

Paired combinational logit (PCL) (Chu, 1981)

Cross-nested logit (CNL) (Vovsha, 1997)

Generalization

Generalized nested logit (GNL), recursive nested logit extreme value model (RNEV), network-GEV
(Wen & Koppelman, 2001; Daly, 2001; Bierlaire, 2002)

Derived from McFadden's G function or "choice probability generating functions" (Fosgerau et al., 2013)

Generalization

Heteroscedastic/mixed distributions

Error component logit (ECL); Mixed logit (MXL); Kernel logit (KL); Heteroscedastic logit (HL)
(Boyd and Mellman, 1980; Cardell and Dunbar, 1980; McFadden, 1989; Bhat, 1995; See Train (2009) for details)

Gumbel to Weibull

1

Weibull to GEV (not MEV)

Multinomial weibit (MNW)
(Castillo, et al., 2008)

q-generalized logit
(Nakayama, 2013, Nakayama and Chikaraishi, 2015)

Generalization

Generalized G function
(Mattsson et al., 2014)

2

Variance stabilization
(Li, 2011)

Derived from the generalized G function

3

Deep neural network

Castillo, E., Menendez, J.M., Jimenez, P., Rivas, A. (2008) Closed form expressions for choice probabilities in the Weibull case. Transportation Research Part B 42, 373-380.

Chikaraishi, M., Nakayama, S. (2016) Discrete choice models with q-product random utilities, Transportation Research Part B

LOGICALLY EXPLAINABLE NON-LINEAR TRANSFORMATION

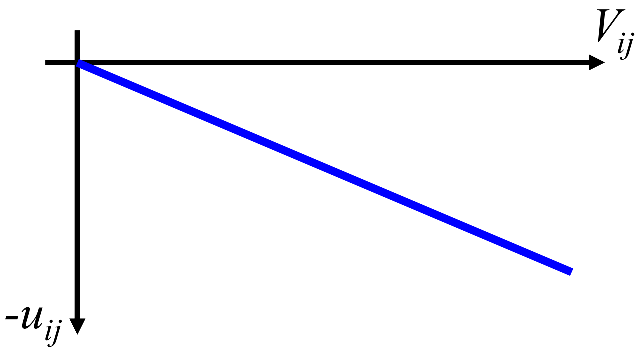
Nonlinearity of V_{ij}

(2) Difference in systematic utility

$$u_{ij} = \boxed{f\left(V_{ij}\right)} + \varepsilon_{ij}$$

ε_{ij}: Gumbel distribution

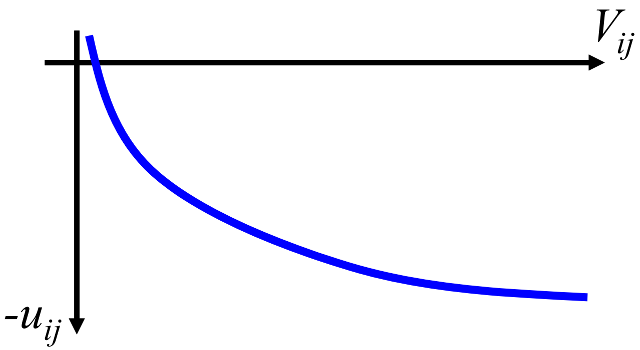
Linear systematic utility



Choice probability

$$p_{ij} = \frac{\exp\left(-\frac{1}{\theta}V_{ij}\right)}{\sum_k \exp\left(-\frac{1}{\theta}V_{ik}\right)}$$

Logarithm systematic utility



Choice probability

$$p_{ij} = \frac{V_{ij}^{-\frac{1}{\theta}}}{\sum_k V_{ik}^{-\frac{1}{\theta}}}$$

Distribution/linearity: an example

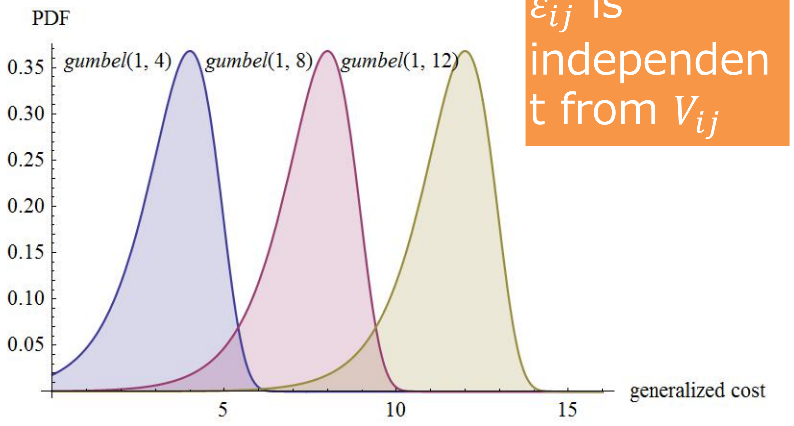
(1) Difference in distribution assumption

$$\boxed{u_{ij}} = g\left(V_{ij}, \varepsilon_{ij}\right)$$

u_{ij}: Random utility
V_{ij}: Systematic utility (linear in parameters)
ε_{ij}: Error term

Gumbel distribution

ε_{ij} is independent from V_{ij}

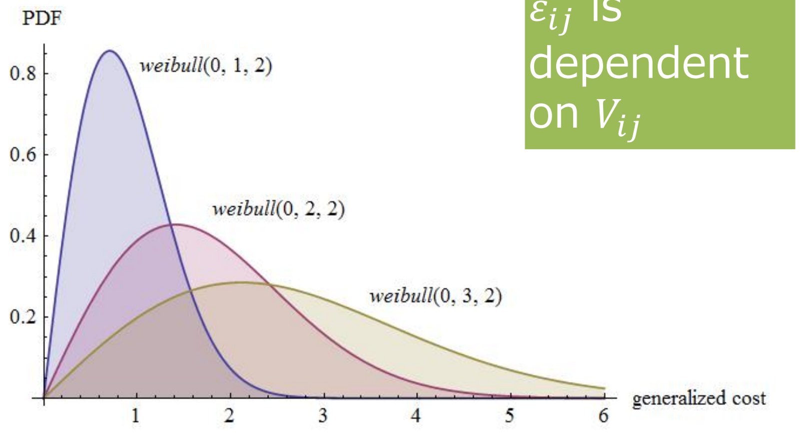


Logit model

$$p_{ij} = \frac{\exp\left(-\frac{1}{\theta}V_{ij}\right)}{\sum_k \exp\left(-\frac{1}{\theta}V_{ik}\right)}$$

Weibull distribution

ε_{ij} is dependent on V_{ij}

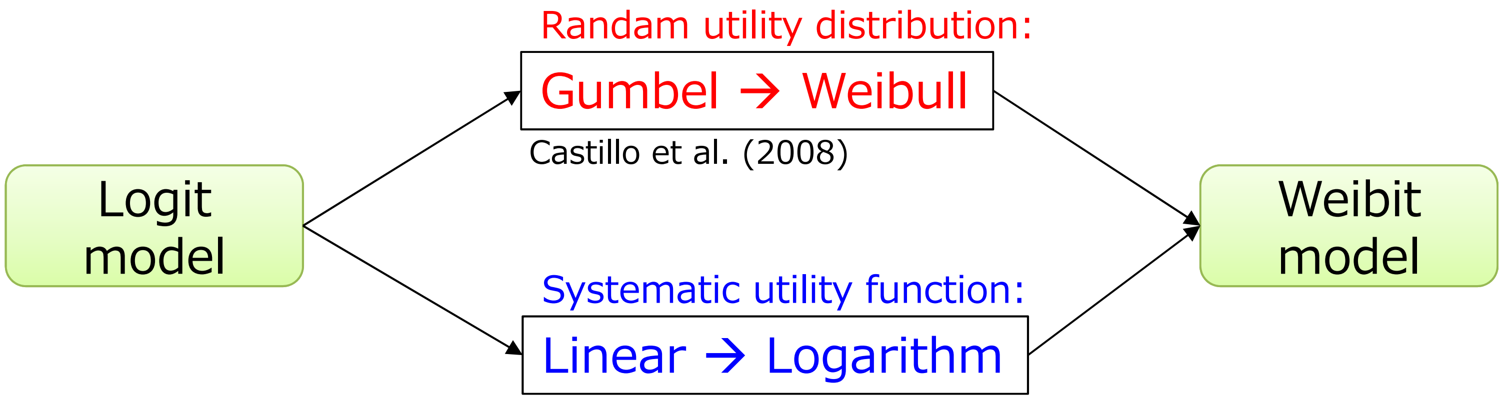


Weibit (or multiplicative) model

$$p_{ij} = \frac{V_{ij}^{-\frac{1}{\theta}}}{\sum_k V_{ik}^{-\frac{1}{\theta}}}$$

Distribution/linearity: an example

$$u_{ij} = g\left(f\left(V_{ij}\right), \varepsilon_{ij}\right)$$



Randam utility distribution:

Gumbel → Weibull

Castillo et al. (2008)

Logit model

Weibit model

Systematic utility function:

Linear → Logarithm

(See Castillo et al. (2008) for elegant explanations)

q-product random utility

q-product random utility → q-generalized Gumbel → Derivation of q-product logit → Behavioral implication

- Preliminaries: q-generalization (Tsallis, 2009)
  - Generalized Boltzmann–Gibbs statistical mechanics
  - The core concept is the so-called Tsallis entropy, where the “q-generalization” plays a central role

q-logarithm: $\ln_q(x) \coloneqq \frac{x^{1-q}-1}{1-q}$ $\qquad \lim_{q\to 1}(\ln_q(x)) = \ln(x)$

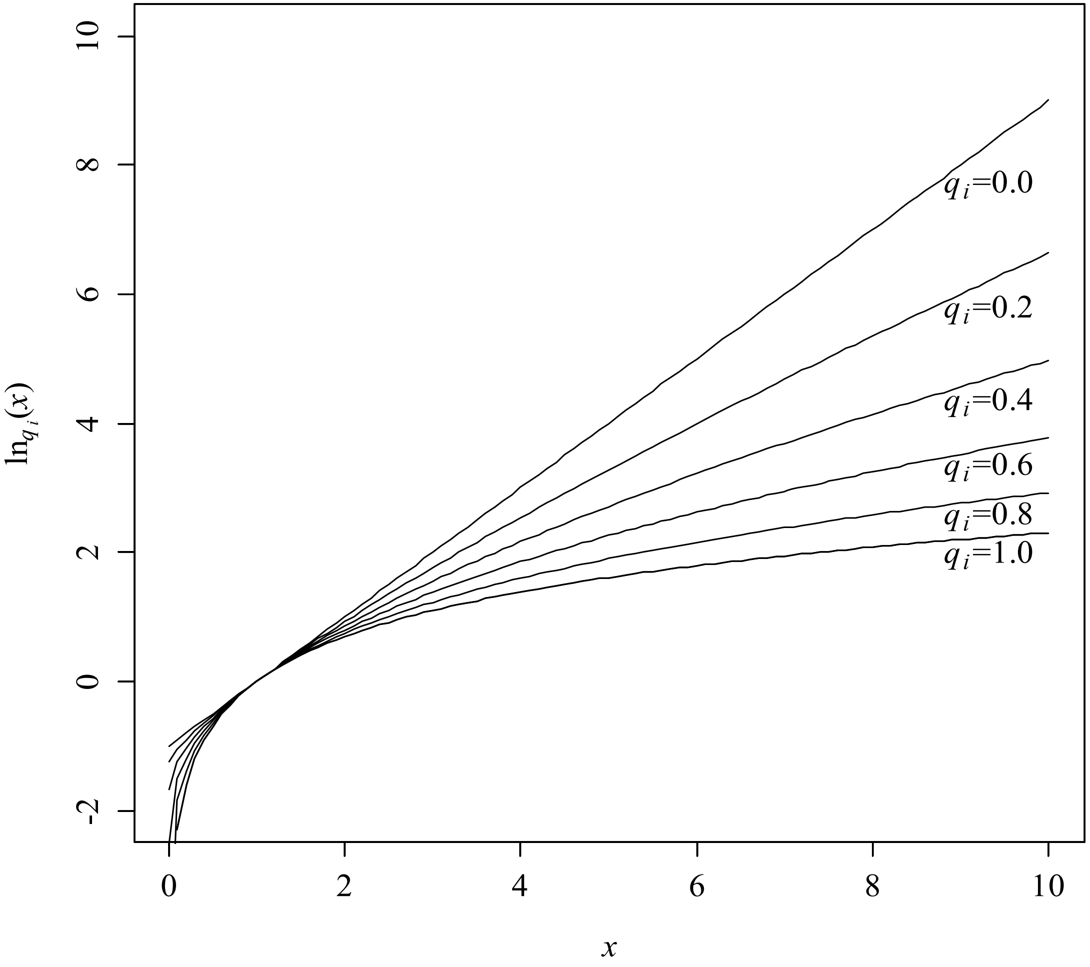
q-exponential: $\exp_q(x) \coloneqq [1+(1-q)x]^{\frac{1}{1-q}}$ $\qquad \lim_{q\to 1}(\exp_q(x)) = \exp(x)$

q-product: $x \otimes_q y {:=} [x^{1-q}+y^{1-q}-1]^{\frac{1}{1-q}}$ $\qquad \lim_{q\to 1}\left(x \otimes_q y\right) = xy$

Some properties:

$$\ln_q\left(\exp_q(x)\right) = x, \quad \ln_q\left(x \otimes_q y\right) = \ln_q(x) + \ln_q(y)$$

Generalization of logit and weibit



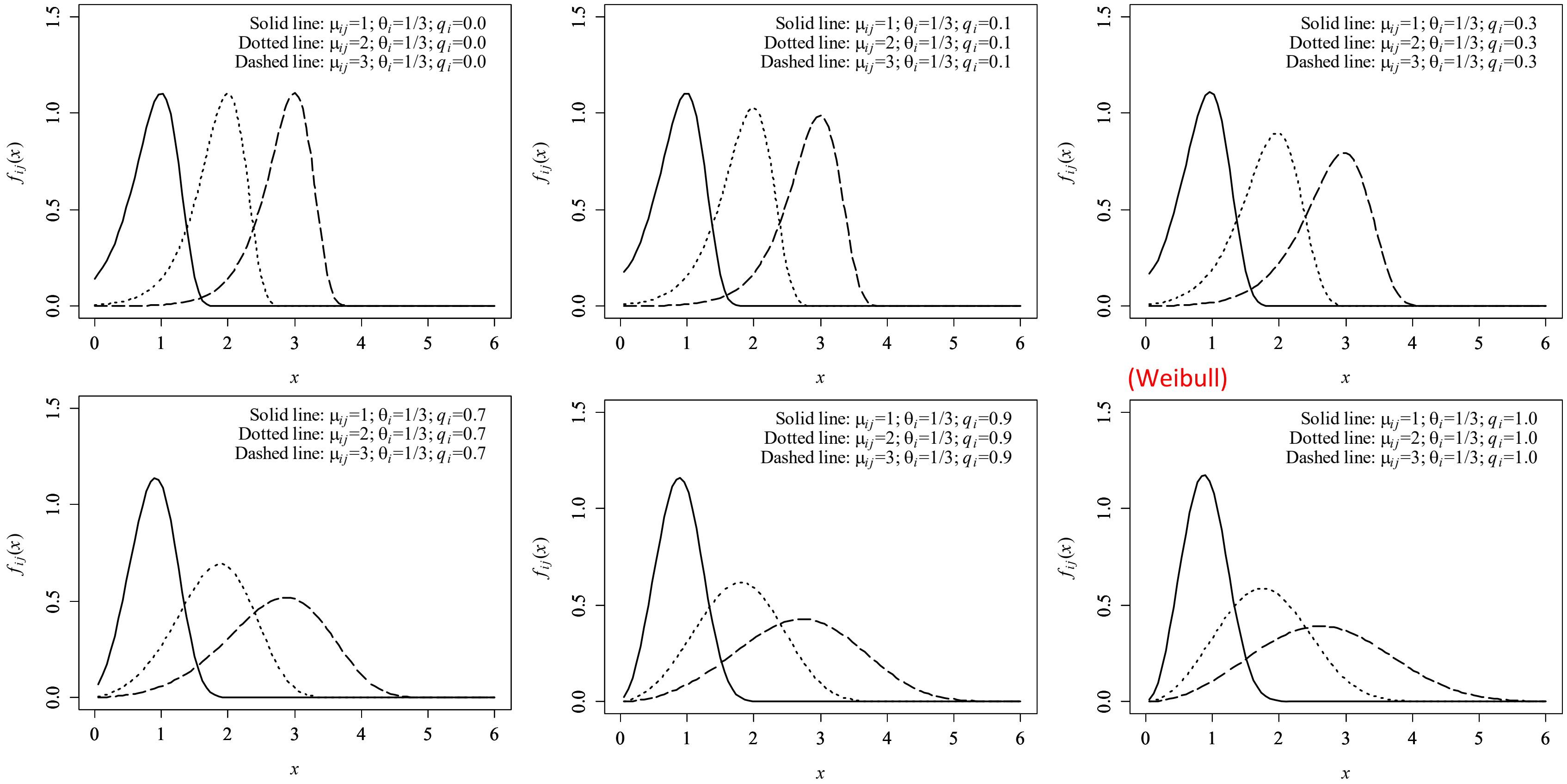
q-logarithm: $\ln_q(x) := \dfrac{x^{1-q}-1}{1-q}$ $\lim_{q\to 1}(\ln_q(x)) = \ln(x)$

q-generalized random term

- q-generalized reverse Gumbel distribution

$$f_{ij}(x) = g_{ij}(y)\left|\frac{dy}{dx}\right| = \frac{1}{\theta_i x^{q_i}} \exp\left(\frac{\ln_{q_i}(x) - \ln_{q_i}(\mu_{ij})}{\theta_i}\right) \exp\left(-\exp\left(\frac{\ln_{q_i}(x) - \ln_{q_i}(\mu_{ij})}{\theta_i}\right)\right)$$

(reverse Gumbel)



(Weibull)

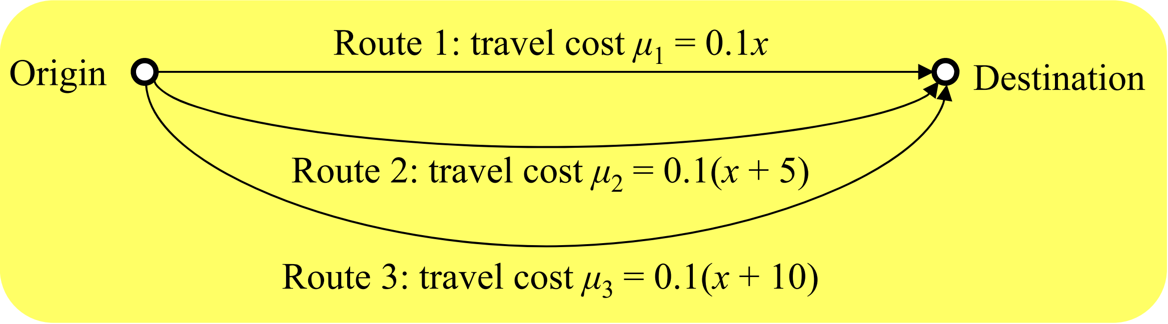
Behavioral implication

- Risk attitude
  - ✓ $\ln_{q_i}(\mu_{ij})$ is an isoelastic utility function (also known as a power utility function), which has been widely used in economics
  - ✓ This gives the parameter q_i a clear behavioral meaning: **q_i is equivalent to the Arrow–Pratt measure of relative risk aversion** when imposing Gumbel distribution on its error component

$$R(\mu_{ij}) = -\mu_{ij} \frac{\partial^2 \ln_{q_i}(\mu_{ij}) / \partial \mu_{ij}^2}{\partial \ln_{q_i}(\mu_{ij}) / \partial \mu_{ij}} = q_i$$

Behavioral implication

- Example



Origin

Route 1: travel cost $\mu_1 = 0.1x$

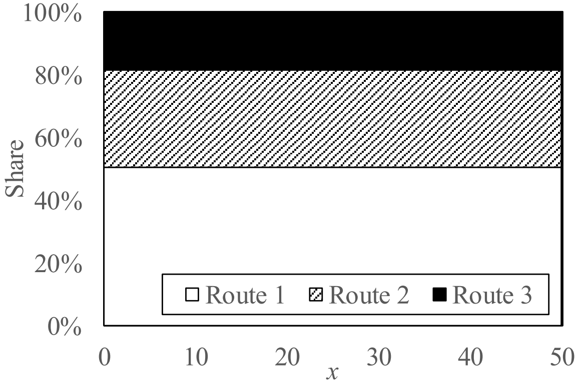
Destination

Route 2: travel cost $\mu_2 = 0.1(x+5)$

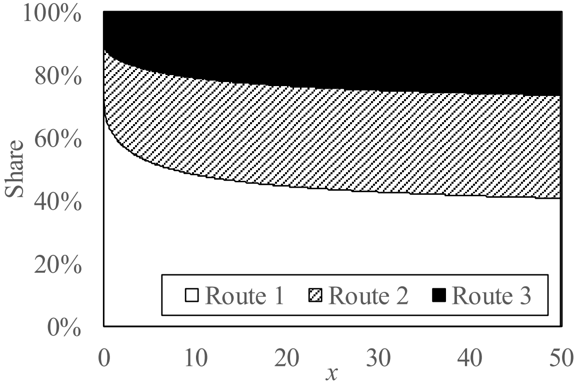
Route 3: travel cost $\mu_3 = 0.1(x+10)$

Choice probability:

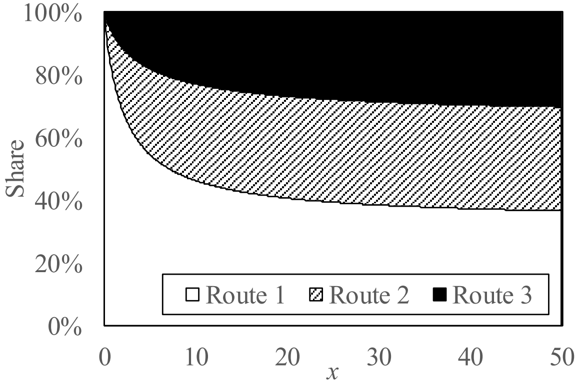
$$P_j = \frac{\exp(-\ln_q(\mu_j))}{\sum_{j'=1}^{3} \exp(-\ln_q(\mu_{j'}))}$$



(a) $q = 0.0$ (logit)

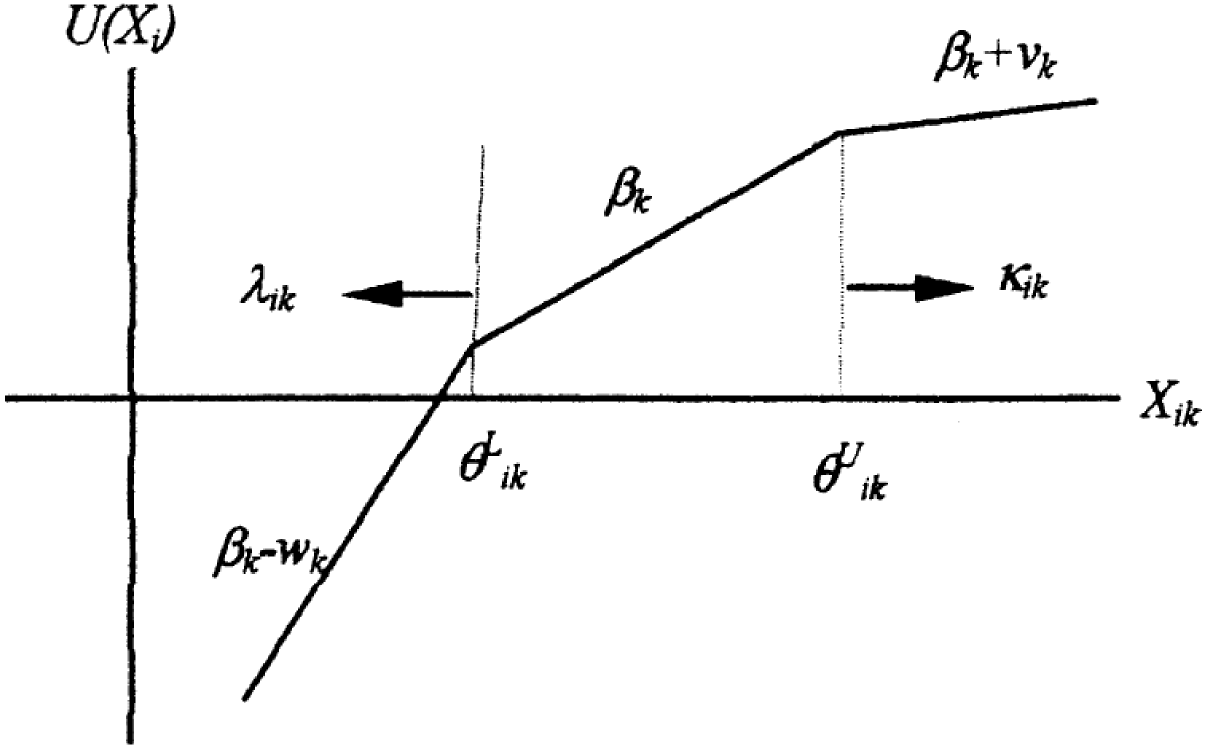


(b) $q = 0.5$



(c) $q = 1.0$ (weibit)

Other choice models with non-linearity transformation



Swait, J., 2001. A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological 35, 903-928.*

Li, B. (2011) The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis. Transportation Research Part B 45, 461-473.

SPLINE-BASED NON-LINEAR TRANSFORMATION

Choice probability

Li (2011) shows that we can derive a number of discrete choice models under different error tem distributions:

$$P_{ij} = \theta \alpha_{ij} \int_{z \in \Omega_i} \exp[-\alpha_{i0} \exp(\theta z)] \exp(\theta z)\, dz$$

$$= \frac{\alpha_{ij}}{\alpha_{i0}} = \frac{\alpha_{ij}}{\Sigma_{j\prime \in C_i} \alpha_{ij\prime}} = \frac{H\left(V_{ij}\right)}{\Sigma_{j\prime \in C_i} H\left(V_{ij\prime}\right)} = \frac{\exp\left(S\left(\boldsymbol{\beta}\mathbf{x_{ij}}\right)\right)}{\Sigma_{j\prime \in C_i} \exp\left(S\left(\boldsymbol{\beta}\mathbf{x_{ij}}\right)\right)}$$

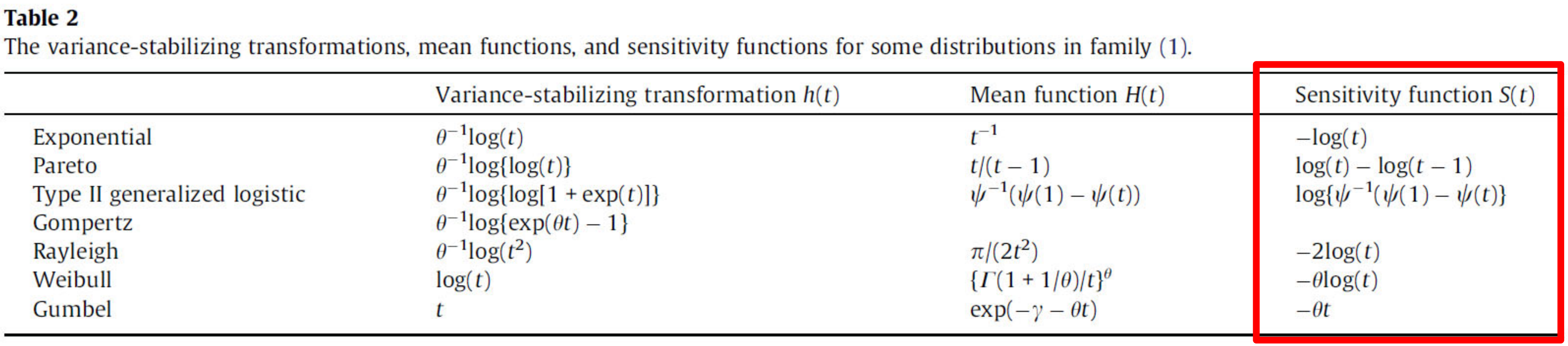


Table 2
The variance-stabilizing transformations, mean functions, and sensitivity functions for some distributions in family (1).

	Variance-stabilizing transformation $h(t)$	Mean function $H(t)$	Sensitivity function $S(t)$
Exponential	$\theta^{-1}\log(t)$	t^{-1}	$-\log(t)$
Pareto	$\theta^{-1}\log\{\log(t)\}$	$t/(t-1)$	$\log(t) - \log(t-1)$
Type II generalized logistic	$\theta^{-1}\log\{\log[1+\exp(t)]\}$	$\psi^{-1}(\psi(1) - \psi(t))$	$\log\{\psi^{-1}(\psi(1) - \psi(t))\}$
Gompertz	$\theta^{-1}\log\{\exp(\theta t) - 1\}$		
Rayleigh	$\theta^{-1}\log(t^2)$	$\pi/(2t^2)$	$-2\log(t)$
Weibull	$\log(t)$	$\{\Gamma(1+1/\theta)/t\}^{\theta}$	$-\theta\log(t)$
Gumbel	t	$\exp(-\gamma - \theta t)$	$-\theta t$

The above equation indicates the choice of error term distribution would be equal to the choice of non-linear transformation of systematic utility. (as we already confirmed)

Semi-parametric discrete choice models

Semi-parametric approach (such as P-splines approach) can be used as an approximation of any base distribution.

$$\frac{H(V_{ij})}{\Sigma_{j'\in C_i} H(V_{ij'})} = \frac{\exp\{S(\boldsymbol{\beta}\mathbf{x_{ij}})\}}{\Sigma_{j'\in C_i}\exp\{S(\boldsymbol{\beta}\mathbf{x_{ij}})\}}$$

$S(\boldsymbol{\beta}\mathbf{x_{ij}})$: Sensitivity function

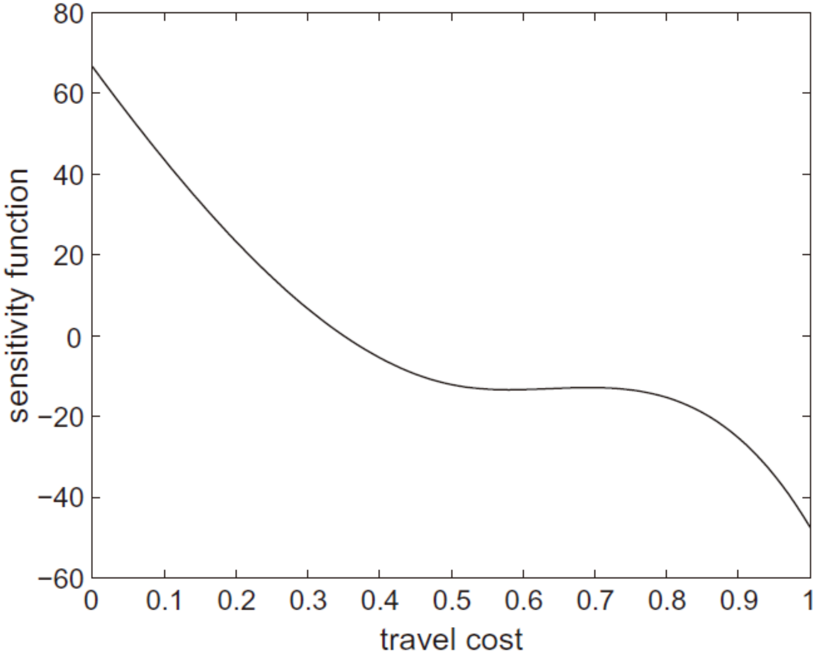


Fig. 2. The estimated sensitivity function $S(t)$ for the train data.

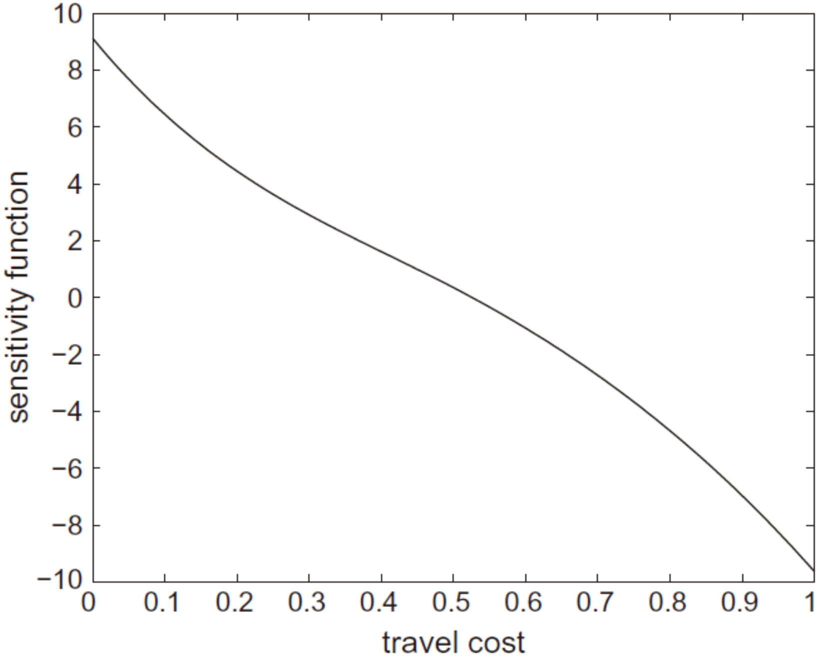


Fig. 3. The estimated sensitivity function $S(t)$ for the bus data.

Emergence of Deep Learning

- **Limitations of linear-in-parameter model**
  - No consideration of non-linearity
  - No consideration of interactions among variables

- **Possible solutions** (Goodfellow et al., 2016)
  1. Theory-driven (e.g., assuming non-compensatory, using BPR function, etc.)
  2. Use Kernel, Splines, etc.
  3. Learn the function from the data (e.g., deep learning) → produce more accurate results in many cases

Sifringer, B., Lurkin, V., Alahi, A., 2018. Enhancing Discrete Choice Models with Neural Networks. 18th Swiss Transport Research Conference, Monte Verità, May 16–18.

DISCRETE CHOICE WITH NEURAL NETWORK

Background and objective

- **RUM model vs neural network**
  - Advantage of RUM model
    - Interpretability of the results.
  - Advantage of neural network
    - Better goodness-of-fit
- **Objective**
  - Bringing the predictive strength of Neural Networks, a powerful machine learning-based technique, to the field of Discrete Choice Models (DCM) without compromising interpretability of these choice models.

RUM: random utility maximization

RUM model and neural network (NN)

Discrete choice model as a Random Utility Maximization (RUM) model

Utility function: $$U_{in} = \beta_1 x_{1in} + \ldots + \beta_d x_{din} + \varepsilon_{in} \qquad \forall i \in C_n$$
$$= V_{in} + \varepsilon_{in}$$

Choice probability: $$P(i|C_n) = P(U_{in} > \max_{j \neq i}(U_{jn})) = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})}$$

(negative) log-likelihood: $$LL = -\sum_{n=1}^{N} \sum_{i \in C_n} y_{in} \log[P(i|C_n)]$$

A discrete choice model from the perspective of neural network

Softmax activation function: $$(\sigma(\mathbf{V}_n))_i = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})}$$

Cross-entropy: $$H_n(\sigma, \mathbf{y}_n) = -\sum_{i \in C_n} y_{in} \log[(\sigma(\mathbf{V}_n))_i]$$

The conventional MNL can be seen as a neural network model with a simple network structure.

Discrete Choice Model with NN

Utility function with non-linear component:

$$\mathbf{U}_n = \boldsymbol{\beta}\boldsymbol{\chi}^T + \mathbf{u}_n + \boldsymbol{\varepsilon}_n$$

Linear-in-parameters component

Non-linear component (via NN)

$\mathbf{U}_n$: $\{U_{1n}, U_{2n}, ..., U_{In}\}$

$\boldsymbol{\beta}$: A vector of parameters (1×*d*)

$\boldsymbol{\chi}$: A set of explanatory variables (I×*d*)

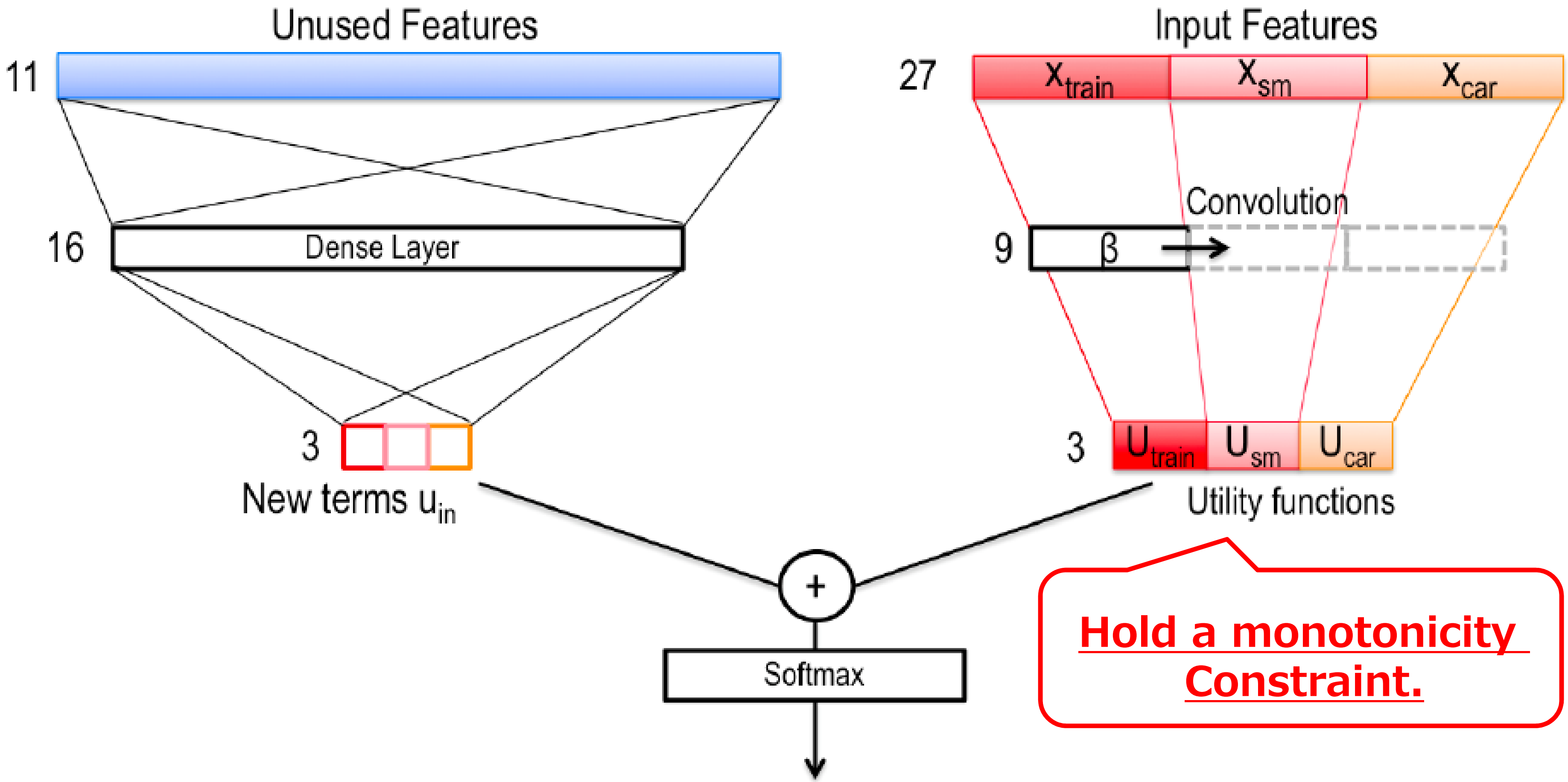
$$\mathbf{u}_n = \psi(\mathbf{Q})$$

where $\mathbf{Q}$ is the ensemble of input features, and, ψ is the function defined by multiple neural network layers and their corresponding activation functions.

Discrete Choice Model with NN

Non-linear component

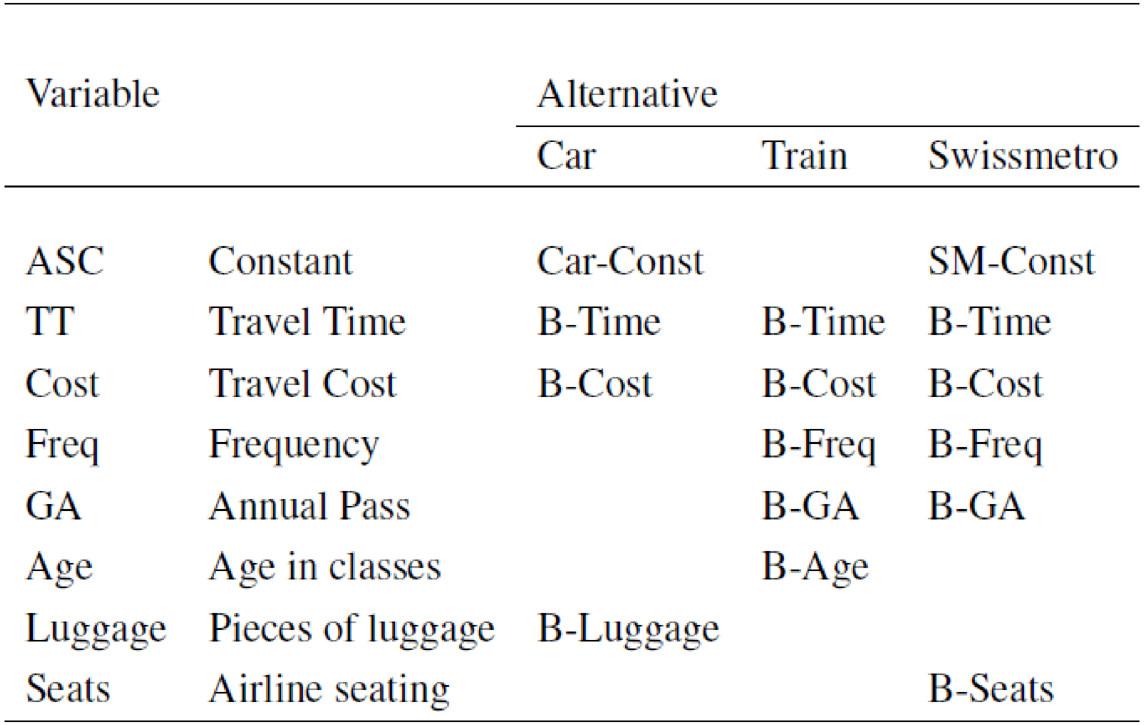
Linear-in-parameters component



Hold a monotonicity Constraint.

Empirical analysis

- Dataset
  - Swissmetro dataset (Bierlaire et al., 2001)
  - A stated preference data on mode choice
  - 10700 entries from 1190 individuals
- Linear-in-parameters component:



Variable		Alternative		
		Car	Train	Swissmetro
ASC	Constant	Car-Const		SM-Const
TT	Travel Time	B-Time	B-Time	B-Time
Cost	Travel Cost	B-Cost	B-Cost	B-Cost
Freq	Frequency		B-Freq	B-Freq
GA	Annual Pass		B-GA	B-GA
Age	Age in classes		B-Age	
Luggage	Pieces of luggage	B-Luggage		
Seats	Airline seating			B-Seats

Empirical analysis

- Non-linear component:
  1. **Travel purpose**: Discrete value between 1 to 9 (Business, leisure, travel,...)
  2. **First class**: 0 for no or 1 for yes if passenger is a first class traveler in public transport
  3. **Ticket**: Discrete value between 0 to 10 for the ticket type (One-way, half-day, ...)
  4. **Who**: Discrete value between 0 to 3 for who pays the travel (self, employer, ...)
  5. **Male**: Traveler's gender, 0 for female and 1 for male
  6. **Income**: Discrete value between 0 to 4 concerning the traveler's income per year
  7. **Origin**: Discrete value defining the canton in which the travel begins
  8. **Dest**: Discrete value defining the canton in which the travel ends

Multinomial Logit as Benchmark

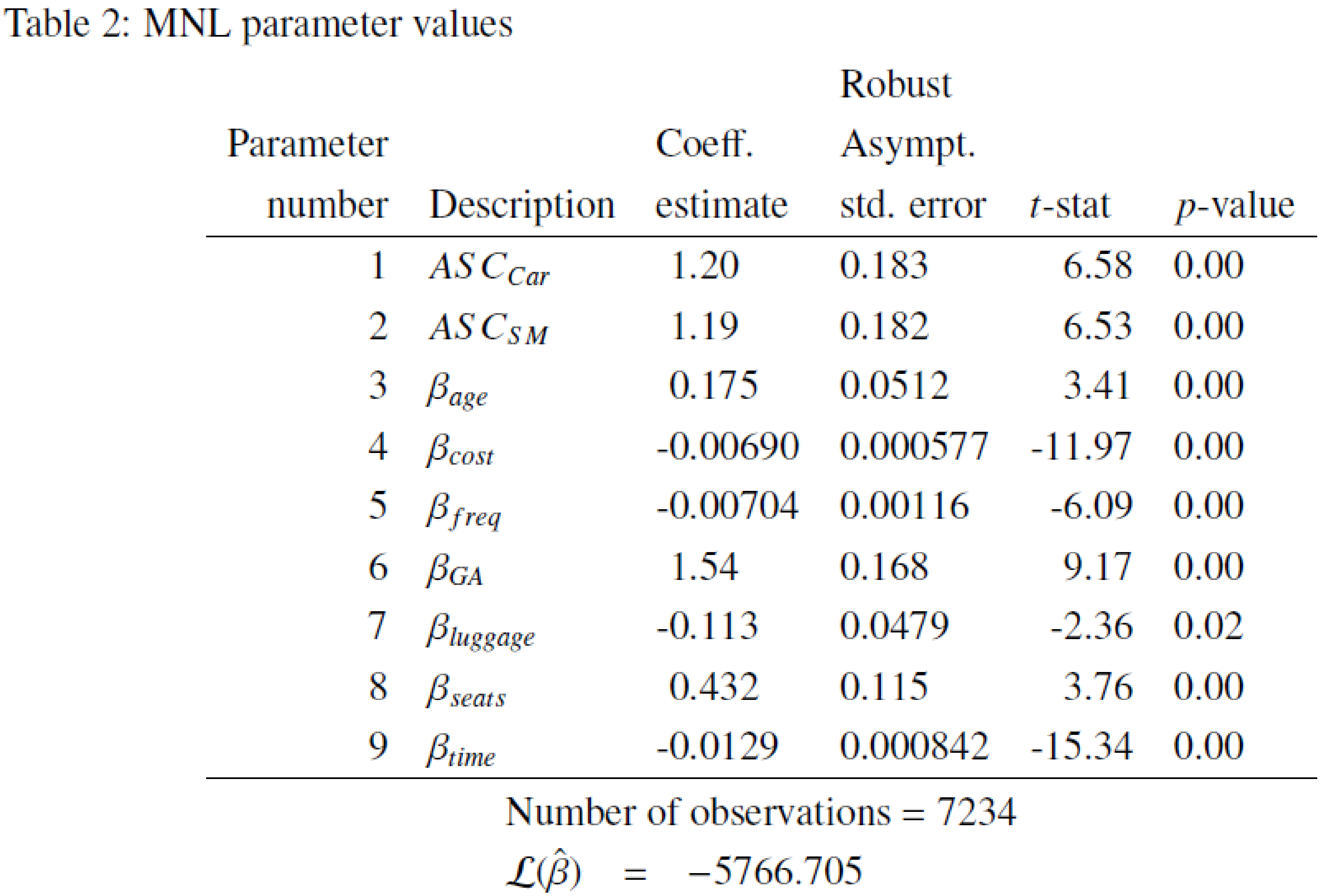


Table 2: MNL parameter values

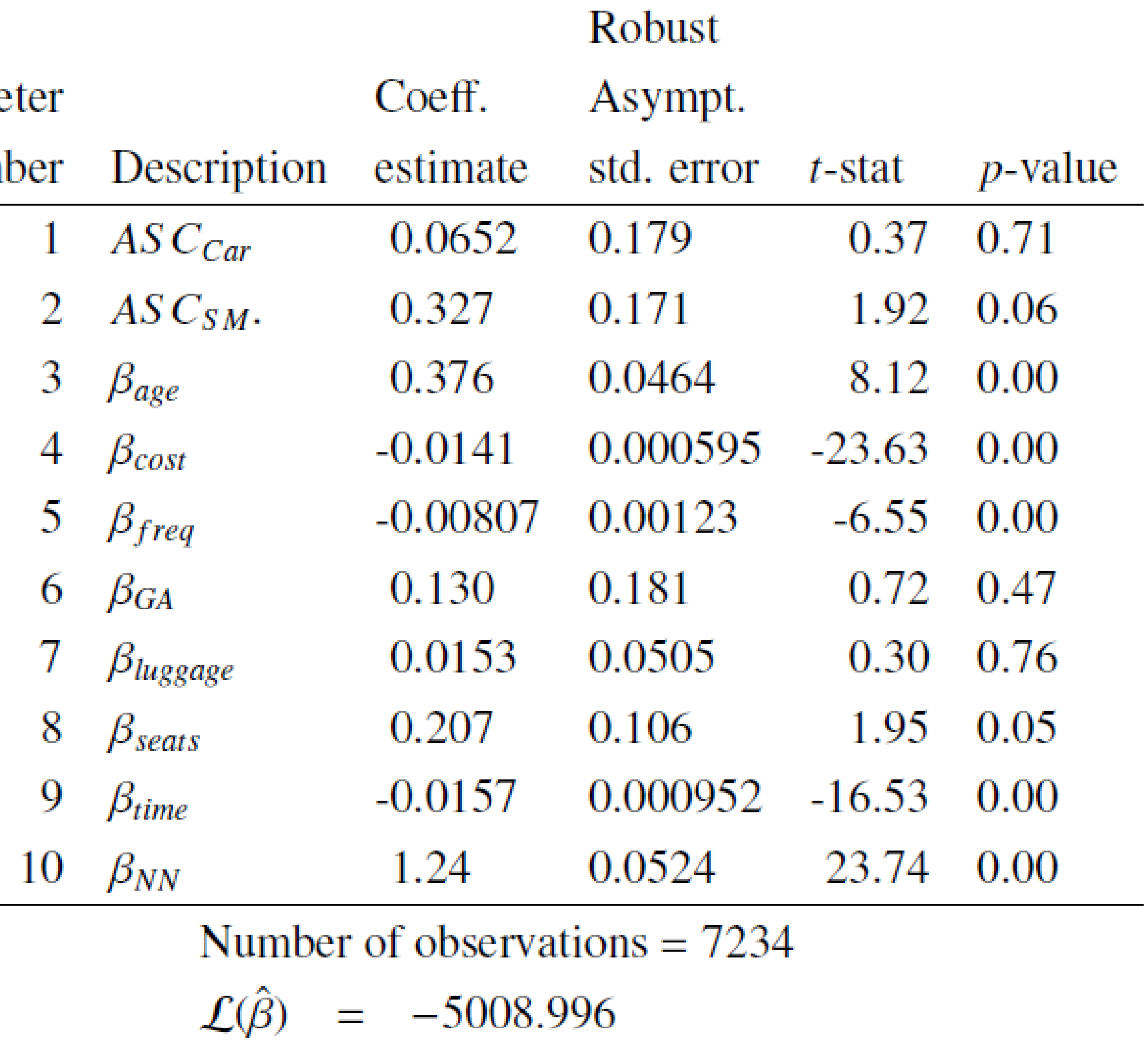
Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_{Car}	1.20	0.183	6.58	0.00
2	ASC_{SM}	1.19	0.182	6.53	0.00
3	β_{age}	0.175	0.0512	3.41	0.00
4	β_{cost}	-0.00690	0.000577	-11.97	0.00
5	β_{freq}	-0.00704	0.00116	-6.09	0.00
6	β_{GA}	1.54	0.168	9.17	0.00
7	$\beta_{luggage}$	-0.113	0.0479	-2.36	0.02
8	β_{seats}	0.432	0.115	3.76	0.00
9	β_{time}	-0.0129	0.000842	-15.34	0.00

Number of observations = 7234

$\mathcal{L}(\hat{\beta}) = -5766.705$

Hybrid model (1)

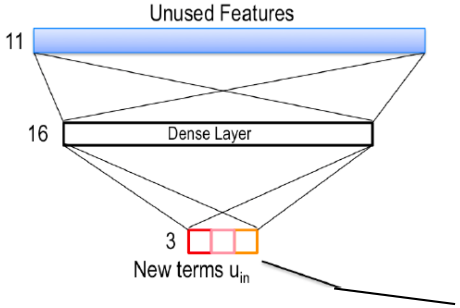
Table 3: Hybrid Model parameter values



Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_{Car}	0.0652	0.179	0.37	0.71
2	ASC_{SM}.	0.327	0.171	1.92	0.06
3	β_{age}	0.376	0.0464	8.12	0.00
4	β_{cost}	-0.0141	0.000595	-23.63	0.00
5	β_{freq}	-0.00807	0.00123	-6.55	0.00
6	β_{GA}	0.130	0.181	0.72	0.47
7	$\beta_{luggage}$	0.0153	0.0505	0.30	0.76
8	β_{seats}	0.207	0.106	1.95	0.05
9	β_{time}	-0.0157	0.000952	-16.53	0.00
10	β_{NN}	1.24	0.0524	23.74	0.00

Number of observations = 7234

$$\mathcal{L}(\hat{\beta}) \quad = \quad -5008.996$$



Note: Statistical properties of the parameters are obtained through Biogeme (Bierlaire, 2009)

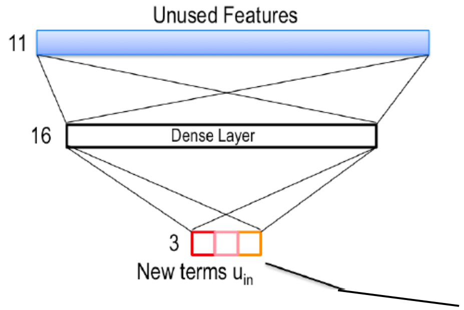
Simplified hybrid model (2)

Table 4: Hybrid model containing only values of greater interest

Parameter number	Description	Coeff. estimate	Robust Asympt. std. error	t-stat	p-value
1	ASC_{Car}	0.966	0.0977	9.89	0.00
2	ASC_{SM}	1.13	0.0941	11.97	0.00
3	β_{cost}	-0.0165	0.000666	-24.71	0.00
4	β_{freq}	-0.00820	0.00129	-6.38	0.00
5	β_{time}	-0.0171	0.000853	-20.05	0.00
6	β_{NN}	1.25	0.0854	14.65	0.00

Number of observations = 7234

$\mathcal{L}(\hat{\beta}) = -4894.539$



All remaining variables are used here

Conclusions & future works

Conclusions:

- Combining the advantage of linear-in-parameters RUM model and the advantage of neural network where highly non-linear impacts of explanatory variables

Future works:

- The selection of hyper parameters (it would change the results)
- Possibility of using the model for long-term demand forecasting (cross-validation may not be enough)
- Possibility of using different NN components (e.g., convolutional NN, recurrent NN, etc.)

Comparison of key parameters

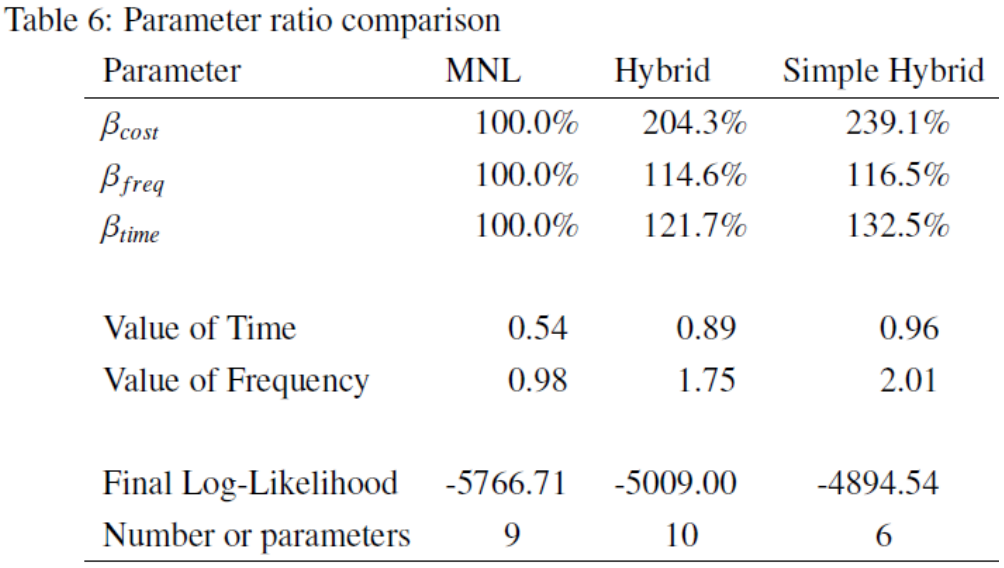


Table 6: Parameter ratio comparison

Parameter	MNL	Hybrid	Simple Hybrid
β_{cost}	100.0%	204.3%	239.1%
β_{freq}	100.0%	114.6%	116.5%
β_{time}	100.0%	121.7%	132.5%
Value of Time	0.54	0.89	0.96
Value of Frequency	0.98	1.75	2.01
Final Log-Likelihood	-5766.71	-5009.00	-4894.54
Number or parameters	9	10	6

References

- Ben-Akiva, M. (1973) Structure of passenger travel demand models. Ph.D. Thesis, Massachusetts Institute of Technology. Dept. of Civil and Environmental Engineering (http://hdl.handle.net/1721.1/14790).
- Bhat, C.R. (1995) A heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part B 29, 471-483.
- Bierlaire, M. (2002) The Network GEV model. Proceedings of the 2nd Swiss Transportation Research Conference, Ascona, Switzerland.
- Cardell, N.S., Dunbar, F.C. (1980) Measuring the societal impacts of automobile downsizing. Transportation Research Part A: General 14, 423-434.
- Castillo, E., Menendez, J.M., Jimenez, P., Rivas, A. (2008) Closed form expressions for choice probabilities in the Weibull case. Transportation Research Part B 42, 373-380.
- Chikaraishi, M., Nakayama, S. (2016) Discrete choice models with q-product random utilities, Transportation Research Part B.
- Daly, A. (2001) Recursive nested EV model. ITS Working Paper 559, Institute for Transport Studies, University of Leeds.
- Daly, A., Bierlaire, M. (2006) A general and operational representation of Generalised Extreme Value models. Transportation Research Part B: Methodological 40, 285-305.

References

- Fosgerau, M., McFadden, D., Bierlaire, M. (2013) Choice probability generating functions. Journal of Choice Modelling 8, 1-18.
- Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. Transportation Research Part B: Methodological 56, 70-80.
- Hato, E. (2002) Behaviors in network, Infrastructure Planning Review, 19-1, 13-27 (in Japanese).
- Koppelman, F.S., Wen, C.-H. (2000) The paired combinatorial logit model: properties, estimation and application. Transportation Research Part B: Methodological 34, 75-89.
- Li, B. (2011) The multinomial logit model revisited: A semi-parametric approach in discrete choice analysis. Transportation Research Part B 45, 461-473.
- Luce, R. (1959) Individual Choice Beahviour. John Wiley, New York.
- Mai, T., Fosgerau, M., Frejinger, E., 2015. A nested recursive logit model for route choice analysis. Transportation Research Part B: Methodological 75, 100-112.
- Mai, T., 2016. A method of integrating correlation structures for a generalized recursive route choice model. Transportation Research Part B: Methodological 93, Part A, 146-161.
- Mattsson, L.-G., Weibull, J.W., Lindberg, P.O. (2014) Extreme values, invariance and choice probabilities. Transportation Research Part B: Methodological 59, 81-95.

References

- McFadden, D., (1978) Modelling the choice of residential location, in: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Residential Location. North-Holland, Amsterdam.
- McFadden, D. (1989) A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. Econometrica 57, 995-1026.
- Nakayama, S. (2013) q-generalized logit route choice and network equilibrium model. Proceedings of the 20th International Symposium on Transportation and Traffic Theory (Poster Session).
- Nakayama, S., Chikaraishi, M., 2015. Unified closed-form expression of logit and weibit and its extension to a transportation network equilibrium assignment. *Transportation Research Part B 81, 672-685.*
- Thurstone, L.L. (1927) A law of comparative judgment. Psychological Review 34, 273-286.
- Train, K. (2009) Discrete Choice Methods with Simulation, 2nd Edition ed. Cambridge University Press.
- Vovsha, P. (1997) Cross-nested logit model: an application to mode choice in the Tel-Aviv metropolitan area. Transportation Research Board, Presented at the 76th Annual Meeting, Washington DC.
- Wen, C.-H., Koppelman, F.S. (2001) The generalized nested logit model. Transportation Research Part B 35, 627-641.