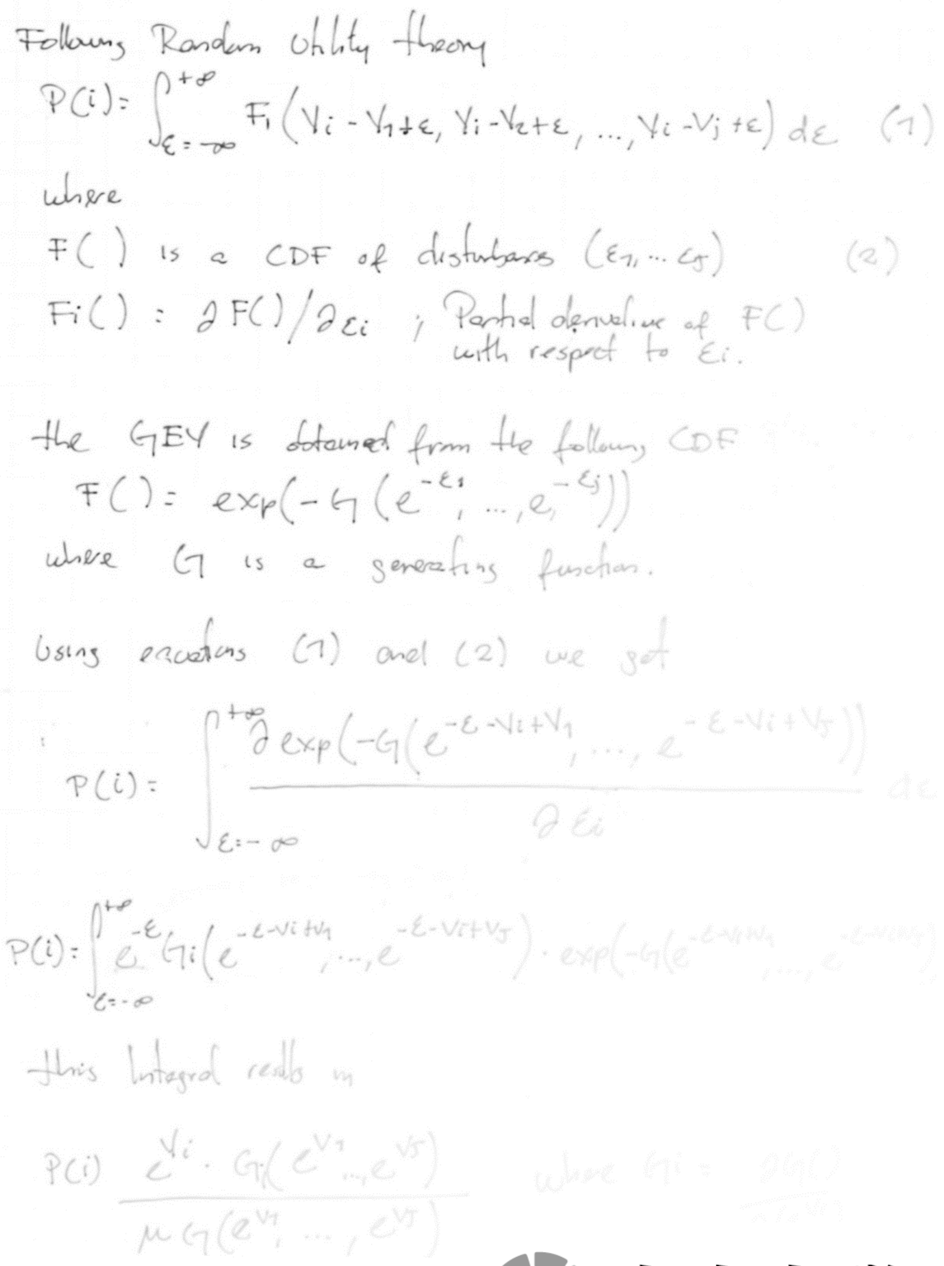
The 20th Behavior Modeling Summer School

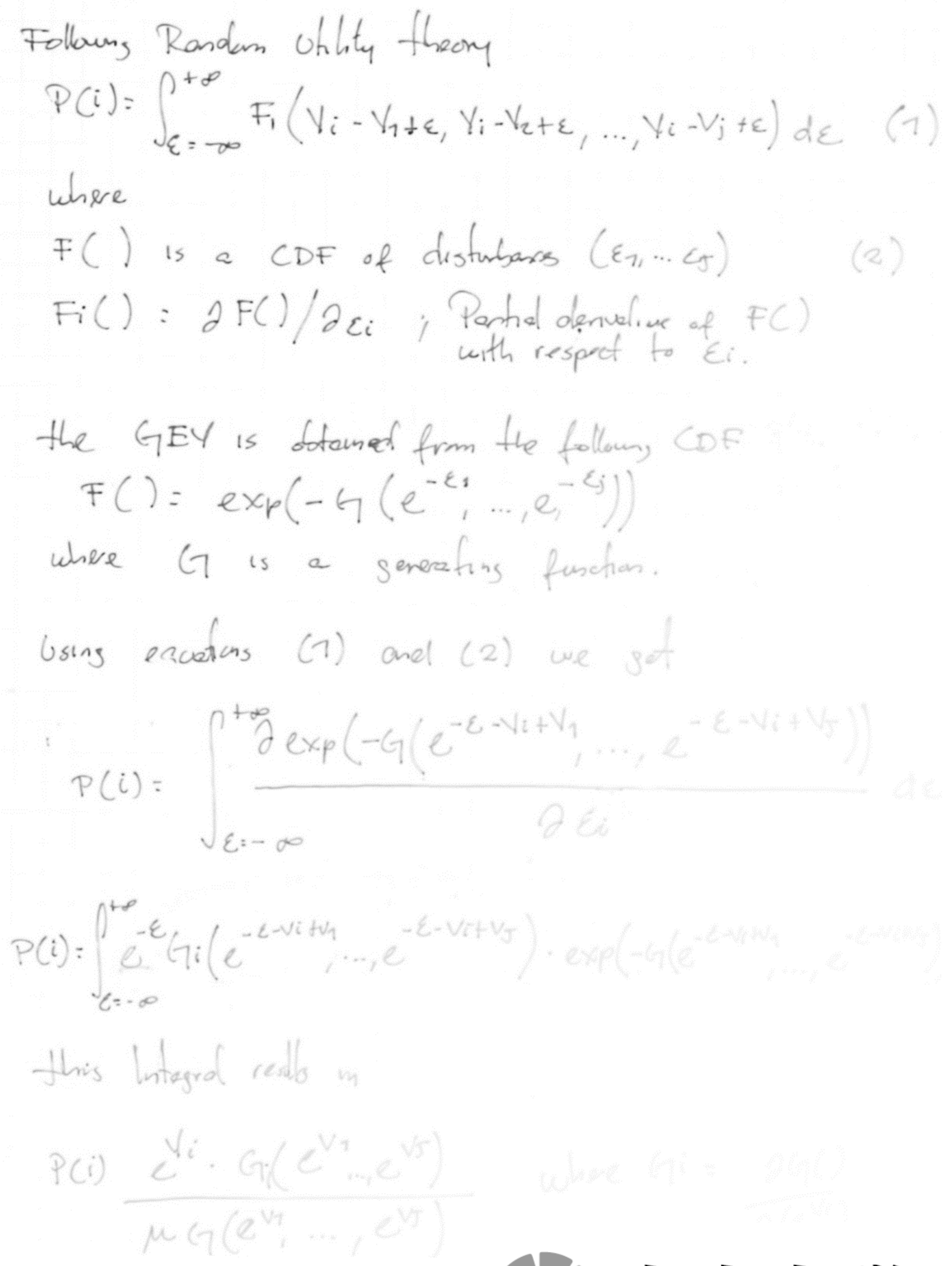
Sep. 15 – 19 , 2020 @ The University of Tokyo



Basic inference and validation in discrete choice modeling

Giancarlos Parady – The University of Tokyo

Basic inference discrete choice modeling



Why is inference important ?

When talking about policy, **Size matters!**

→ Ask not whether there is an effect, ask how big is it?

→ Is the effect big enough to make a policy successful?

→ Is the effect big enough to make invest millions of dollars?

→ 64.6% of studies in the transportation field report some sort of policy-relevant inference analysis (Parady, Ory and Walker, 2021)

Parady G., Ory, D., Walker, J. (2021)

Why is inference important ?

Variable name	Coefficient	S.E.	t statistic
Auto constant	1.45	0.393	3.70
In-vehicle time (min)	-0.0089	0.0063	-1.42
Out-of-vehicle time (min)	-0.0308	0.0106	-2.90
Auto out-of-pocket cost (c)	-0.0115	0.0026	-4.39
Transit fare	-0.0070	0.0038	-1.87
Auto ownership (specific to auto mode)	-0.770	0.213	3.16
Downtown workplace (specific to auto mode)	-0.561	0.306	-1.84
Number of observations	1476		
Number of cases	1476		
LL(0)	-1023		
LL(β)	-347.4		
-2[LL(0)-LL(β)]	1371		
ρ^2	0.660		
$\bar{\rho}^2$	0.654		

Table adapted from Ben-Akiva and Lerman (1985)

← Magnitudes are not directly interpretable. We can only interpret the effect direction, or use them to calculate utilities, and choice probabilities

To make some sense of these parameters we must calculate elasticities or marginal effects

Basic Inference in discrete choice models

MNL: Logit Elasticities (Point elasticities)

- **Direct elasticity:** measures the **percentage change in the probability** of choosing a particular alternative in the choice set with respect to a given **percentage change** in an attribute of that same alternative.

$$E_{x_{ink}}^{P(i)} = \frac{\partial P_n(i)}{\partial x_{ink}} \cdot \frac{x_{ink}}{P_n(i)} = [1 - P_n(i)]x_{ink}\,\beta_k$$

- **Cross-elasticity:** measures the **percentage change in the probability** of choosing a particular alternative in the choice set with respect to a given **percentage change** in a competing alternative.

$$E_{x_{jnk}}^{P(i)} = \frac{\partial P_n(i)}{\partial x_{jnk}} \cdot \frac{x_{jnk}}{P_n(i)} = -P_n(j)x_{jnk}\,\beta_k$$

← Because of IIA, cross-elasticities are uniform across all alternatives

Hensher, David A., John M. Rose, and William H. Greene (2015)

Basic Inference in discrete choice models

MNL: Logit Elasticities (Point elasticities)

- The elasticities shown before are **individual elasticities (Disaggregate)**
- To calculate sample (aggregate) elasticities we use the **probability weighted sample enumeration** method:

$$E_{x_{ink}}^{\overline{P(i)}} = \frac{\sum_{n=1}^{N} \hat{P}_n(i) E_{x_{ink}}^{P(i)}}{\sum_{n=1}^{N} \hat{P}_n(i)}$$

Sample direct elasticity

$$E_{x_{jnk}}^{\overline{P(i)}} = \frac{\sum_{n=1}^{N} \hat{P}_n(i) E_{x_{jnk}}^{P(i)}}{\sum_{n=1}^{N} \hat{P}_n(i)}$$

Sample cross-elasticity

Where $\overline{P(i)}$ *is the aggregate choice probability of alternative I, and* $\hat{P}_{in}(i)$ *is an estimated choice probability*

- Uniform cross-elasticities do not necessarily hold at the aggregate level
- Also note that elasticities for dummy variables are **meaningless!**

Hensher, Rose, and Greene (2015)

Basic Inference in discrete choice models

NL: Logit Elasticities (Point elasticities)

$$P(j) = \frac{e^{V_j/\tau}}{e^{IV(i)}} \cdot \frac{e^{\tau IV(i)}}{\sum_{i=1}^{I} e^{\tau IV(i)}}$$

← NL RUM2 specification

Direct Elasticity

When alternative j does not belong to any nest

$$E_{x_{jnk}}^{P(j)} = [1 - P_n(j)]x_{jnk}\,\beta_k$$

When alternative j belongs to nest i

$$E_{x_{jnk}}^{P(j)} = \left[\left(1 - P_n(j)\right) + \left(\frac{1}{\tau} - 1\right)\left(1 - P(j|i)\right)\right] x_{jnk}\,\beta_k$$

Cross Elasticity

When alternatives j and j' belong to different nests

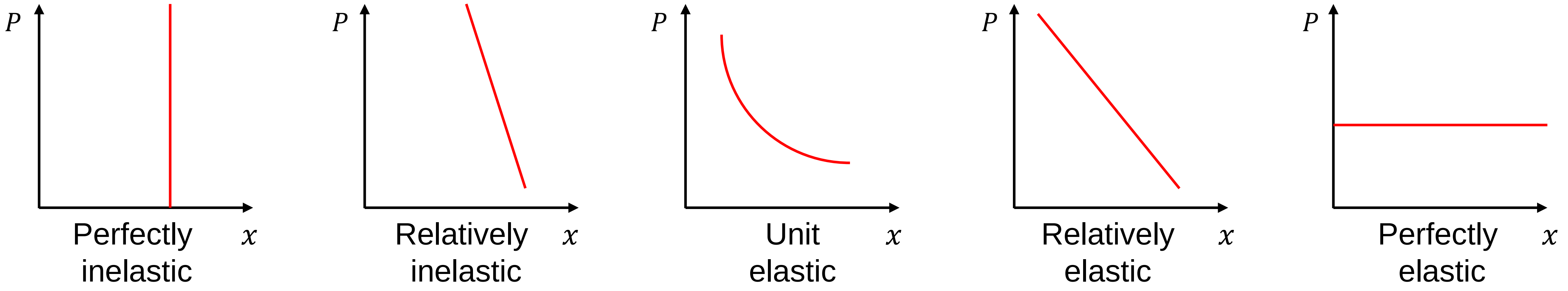
$$E_{x_{j'nk}}^{P(j)} = -P_n(j')x_{j'nk}\,\beta_k$$

When alternatives j and j' belong to the same nest

$$E_{x_{j'nk}}^{P(j)} = -\left[P_n(j') + \left(\frac{1}{\tau} - 1\right)P(j'|i)\right] x_{j'nk}\,\beta_k$$

Basic Inference in discrete choice models

Relation between elasticity of demand, change in price and revenue



Direct elasticity:

Perfectly inelastic	Relatively inelastic	Unit elastic	Relatively elastic	Perfectly elastic
1% increase in x_i results in a 0% decrease in $P(i)$	1% increase in x_i results in a less than 1% decrease in $P(i)$	1% increase in x_i results in a 1% decrease in $P(i)$	1% increase in x_i results in a more than 1% decrease in $P(i)$	1% increase in x_i results in a ∞ decrease in $P(i)$

Cross elasticity:

Perfectly inelastic	Relatively inelastic	Unit elastic	Relatively elastic	Perfectly elastic
1% increase in x_j results in a 0% increase in $P(i)$	1% increase in x_j results in a less than 1% increase in $P(i)$	1% increase in x_j results in no percent change in $P(i)$	1% increase in x_j results in a more than 1% increase in $P(i)$	1% increase in x_j results in a ∞ increase in $P(i)$

Adapted from Hensher, Rose, and Greene (2015)

Basic Inference in discrete choice models

MNL: Marginal Effects

- **Direct marginal effect:** measures the **change in the probability** (absolute change) of choosing a particular alternative in the choice set with respect to a **unit change** in an attribute of that same alternative.

$$M_{x_{ink}}^{P(i)} = \frac{\partial P_n(i)}{\partial x_{ink}} = P_n(i)[1 - P_n(i)]\beta_k$$

- **Cross-marginal effect:** measures the **change in the probability** (absolute change) of choosing a particular alternative in the choice set with respect to a **unit change** in a competing alternative.

$$M_{x_{jnk}}^{P(i)} = \frac{\partial P_n(i)}{\partial x_{jnk}} = P_n(i)(-P_n(j)\beta_k)$$

Hensher, Rose, and Greene (2015)

Basic Inference in discrete choice models

MNL: Marginal Effects

- We can also calculate sample (aggregate) marginal effects using the **probability weighted sample enumeration** method:

$$M_{x_{ink}}^{\overline{P(i)}} = \frac{\sum_{n=1}^{N} \hat{P}_n(i) M_{x_{ink}}^{P(i)}}{\sum_{n=1}^{N} \hat{P}_n(i)}$$

Sample direct marginal effect

$$M_{x_{jnk}}^{\overline{P(i)}} = \frac{\sum_{n=1}^{N} \hat{P}_n(i) M_{x_{jnk}}^{P(i)}}{\sum_{n=1}^{N} \hat{P}_n(i)}$$

Sample cross-marginal effect

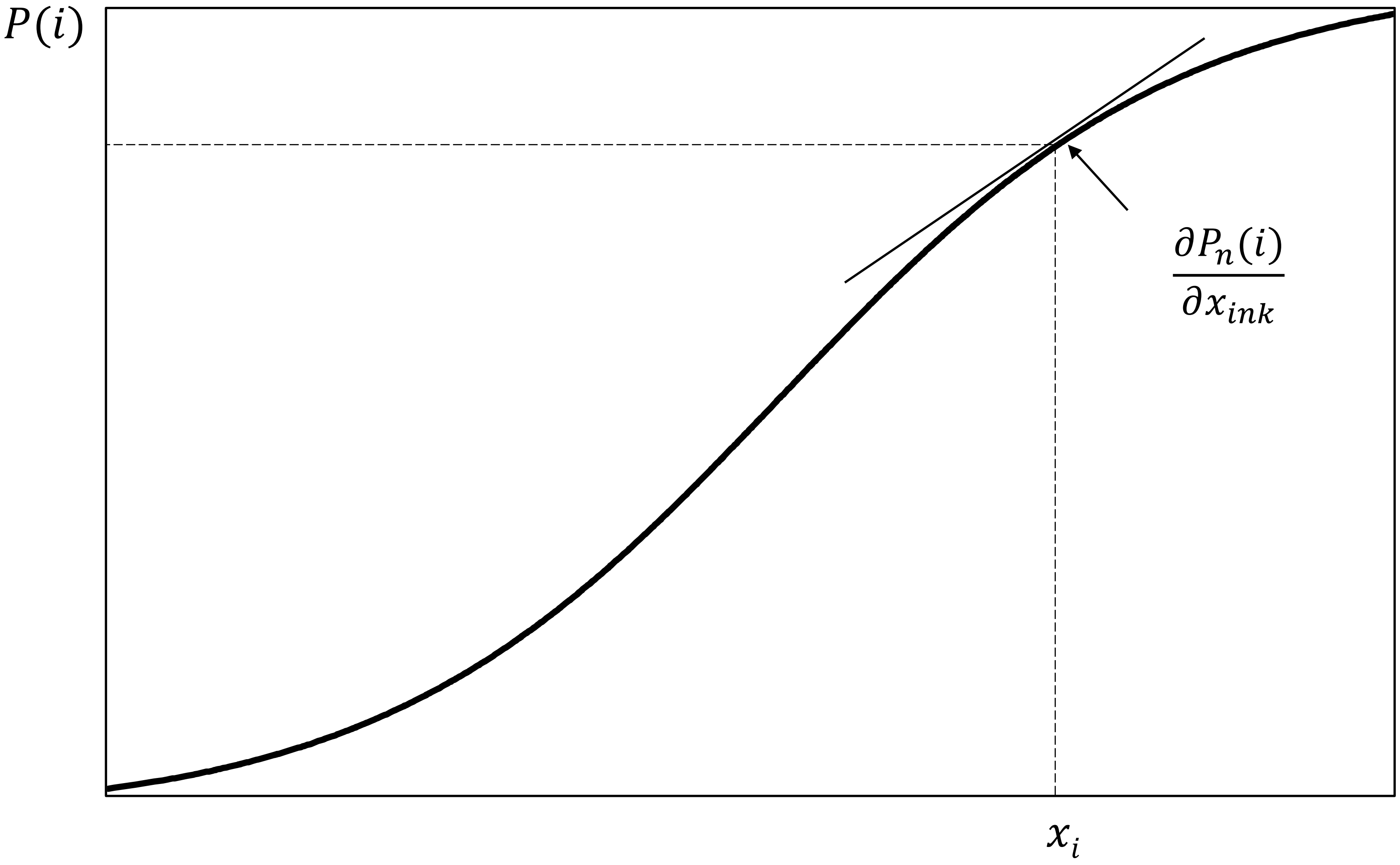
Where $\overline{P(i)}$ is the aggregate choice probability of alternative I, and $\hat{P}_{in}(i)$ is an estimated choice probability

- Marginal effects for dummy variables **do make sense** as we are talking about unit changes, but a different procedure is necessary to estimate marginal effects.

Hensher, Rose, and Greene (2015)

Basic Inference in discrete choice models

MNL: Marginal Effects



Marginal effects as the slopes of the Tangent lines to the cumulative probability curve

Hensher, David A., John M. Rose, and William H. Greene (2015)

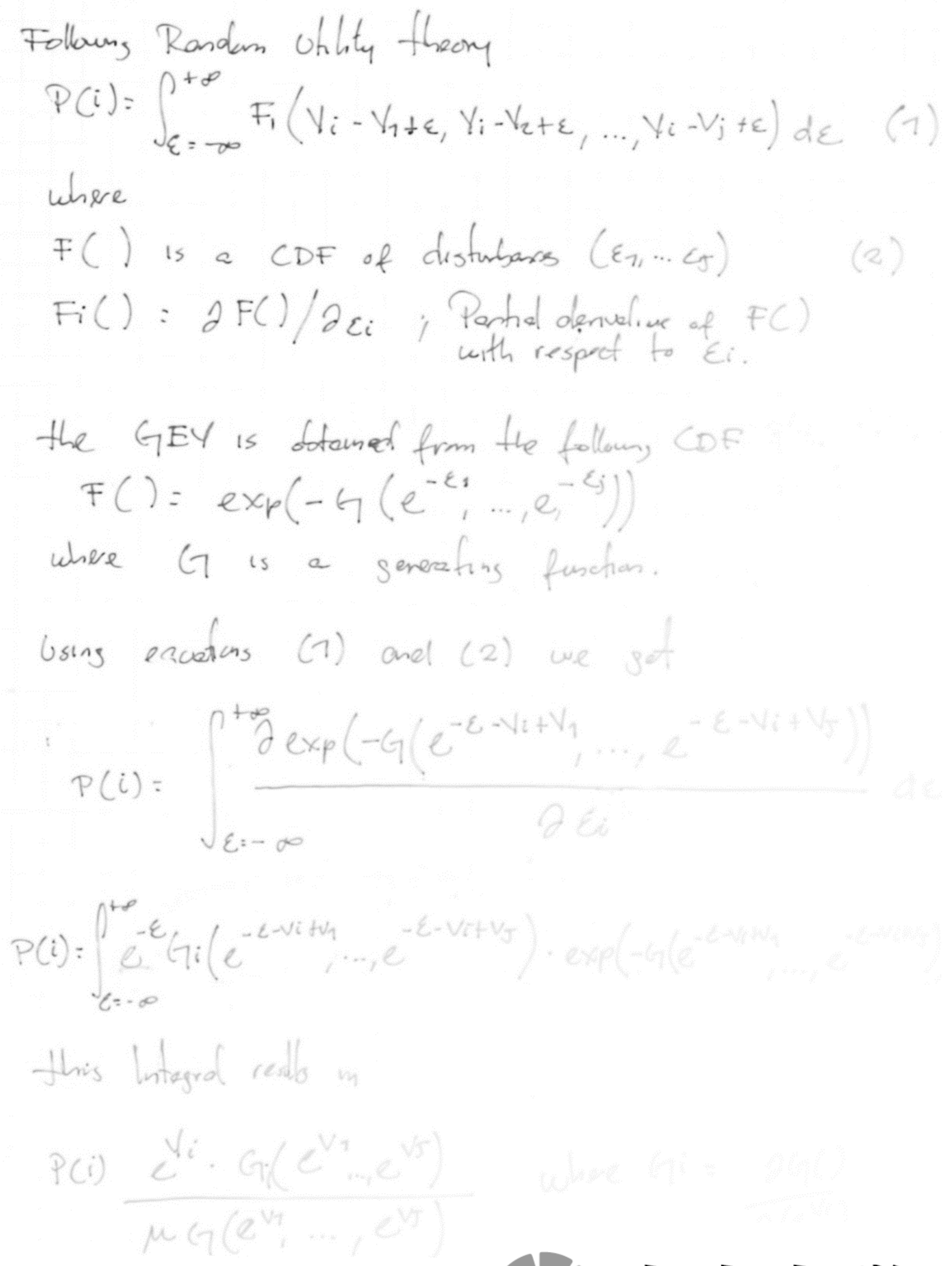
Basic Inference in discrete choice models

MNL: Marginal Effects

Calculating marginal effects for dummy variables

Calculated via simulation:

1. Set the values of the variable of interest to 0
2. Estimate base predictions (at the individual level)
3. Set the values of the variable of interest to 1
4. Estimate new predictions (at the individual level)
5. Calculate marginal effects by taking the mean of the difference in individual predictions



Validation practices in discrete choice modeling

東京大学
THE UNIVERSITY OF TOKYO

A credibility crisis in science and engineering?

Image removed due to copyright issues
See original article here:

M. Baker, D. Penny (2016) Is there a reproducibility crisis? Nature, 533 (7604) pp. 452-454

Baker and Penny (2016)

A credibility crisis in science and engineering?

Most published research findings are likely to be false due to factors such as lack of power of the study, small effect sizes, and great flexibility in research design, definitions, outcomes and methods.

Focused on experimental studies

(Ioannidis, 2005)

In the transportation field

Unlike the natural sciences

- Dependence on cross-section observational studies
- Classic scientific hypothesis testing is more difficult
- Impact evaluation of policies drawn based on model-based academic research is rarely conducted
- No feedback in terms of how right or how wrong are these models and the policy recommendations derived from them
- **These issues underscore the need for proper validation practices**

Term definitions

Predictive accuracy: The degree to which predicted outcomes match observed outcomes.

Predictive accuracy is a function of：

- **Calibration:** The degree to which predicted probabilities match the relative frequency of observed outcomes.
- **Discrimination ability:** The ability of a model or system of models to discriminate between those instances with and without a particular outcome.

Parady, Ory & Walker (2021)

Term definitions

Generalizability: The ability of a model, or system of models to maintain its predictive accuracy in a different sample.

Generalizability of a model is a function of：

- **Reproducibility:** The extent to which a model or system of models maintains its predictive ability in different samples from the same population.
- **Transferability:** The extent to which a model or system of models maintains its predictive ability in samples from different but plausibly related populations or in samples collected with different methodologies (sometimes called transportability.)

Parady, Ory & Walker (2021)

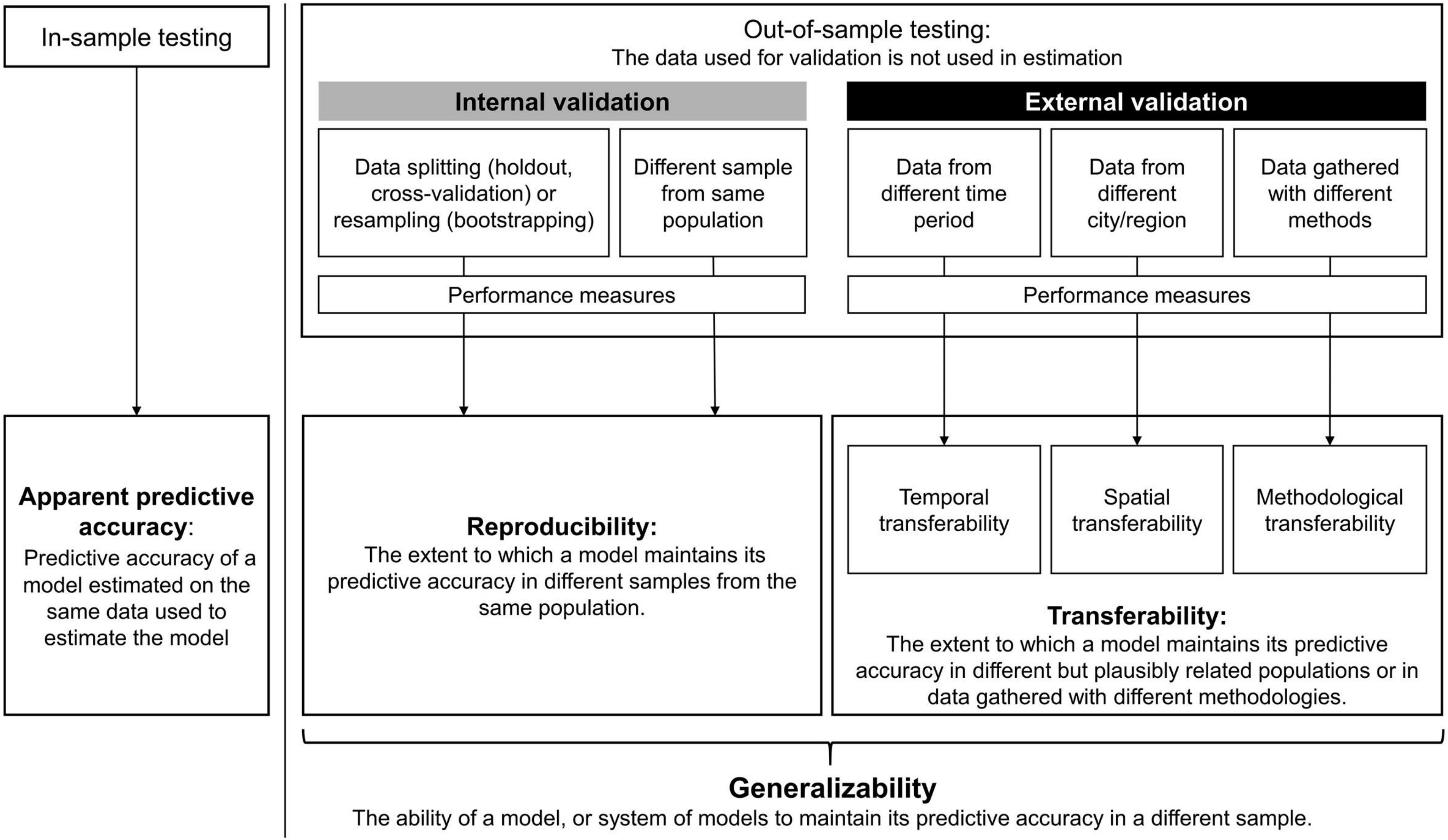
Term definitions

Model validation: The evaluation of the generalizability of a statistical model.

Types of model validation：

- **Internal validation:** The evaluation of the reproducibility of a model.
  - Data splitting, resampling methods
  - Different sample from the same population
- **External validation:** The evaluation of the transferability of a model.
  - Temporal transferability
  - Spatial transferability
  - Methodological transferability

Parady, Ory & Walker (2021)



Parady, Ory & Walker (2021)

A brief introduction to internal validation

- **Internal validation:** The evaluation of the reproducibility of a model.
- Due to the high costs of datal collection the most common approaches are
    - **Data splitting (Holdout validation, Cross validation)**
    - **Resampling methods (Bootstrapping)**

Parady, Ory & Walker (2021)

A brief introduction to internal validation (data splitting methods)

Holdout validation: Dataset is randomly split into an estimation dataset and a validation dataset.

For illustration purposes, let us define $\mathrm{Q}[y_n, \hat{y}_\mathrm{n}]$ as a measure of prediction correctness for the nth instance, for the binary choice case as:

$$\mathrm{Q}[y_n, \hat{y}_\mathrm{n}] = \begin{cases} 0 \; if \; y_n = \hat{y}_\mathrm{n} \\ 1 \; if \; y_n \neq \hat{y}_\mathrm{n} \end{cases}$$

where is y_n the observed outcome, and $\hat{y}_\mathrm{n}$ is the predicted outcome for instance *n*.

The holdout estimator is

$$HOV = \frac{1}{N_v} \sum_{n_v=1}^{N_v} Q[y_{n_v}, \hat{y}_{n_v}^e]$$

where $\hat{y}_{n_v}^e$ is the predicted outcome for instance n in sample v, using the model estimated with sample e, and N_v is the validation sample size.

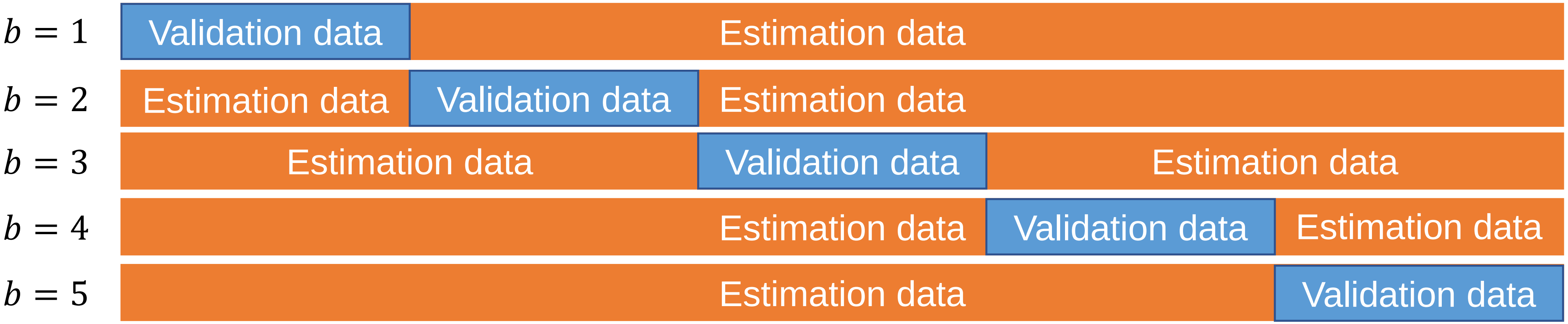
Parady, Ory & Walker (2021)

A brief introduction to internal validation (data splitting methods)

Cross-validation: When the holdout process is repeated multiple times, thus generating a set of randomly split estimation-validation data pairs, we refer to the validation procedure as cross-validation (CV).

$$CV = \frac{1}{\mathrm{B}} \sum_{b} HOV_b$$

where B is the number of estimation-validation data pairs generated and is the holdout estimator for set *b.*



A 5-fold cross validation illustration

Parady, Ory & Walker (2021)

A brief introduction to internal validation (data splitting methods)

Cross-validation：Commonly used methods

$$CV = \frac{1}{B}\sum_{b} HOV_b$$

- Cross-validation **methods differ from one another in the way the data is split.**
- When the data splitting considers all possible estimation sets of size , the splitting is **exhaustive**, otherwise the splitting is **partial**. (Arlot and Celisse, 2009)。

Exhaustive splitting methods

- **Leave-one-out :** estimation set size is $\mathrm{N_e} = N - 1$, and $B = N$. The model is fitted leaving out one instance per iteration, and the outcome of that single instance is predicted based on the estimated model.
- **Leave-p-out :** $\mathrm{N_e} = N - p$. The model is fitted leaving out p-instances per iteration, and the outcome of the remaining instances is predicted based on the estimated model.

Parady, Ory & Walker (2021)

A brief introduction to internal validation (data splitting methods)

Cross-validation：Commonly used methods

$$CV = \frac{1}{B}\sum_{b} HOV_b$$

- Cross-validation **methods differ from one another in the way the data is split.**
- When the data splitting considers all possible estimation sets of size , the splitting is **exhaustive**, otherwise the splitting is **partial**. (Arlot and Celisse, 2009)。

Partial splitting methods (lower calculation cost)

- **K-fold cross-validation:** data is partitioned into *K* mutually-exclusive subsets of roughly equal size, and *B*=*K.*
- **Repeated learning-testing:** a B number of randomly-split estimation-validation pairs are generated. This method is also called repeated holdout validation.

Parady, Ory & Walker (2021)

A brief introduction to internal validation (data splitting methods)

Performance measures

Market share comparison

- Easy to execute
- Does not provide a quantitative measure to evaluate the level of agreement between predictions and observations

Image removed due to copyright issues
See original article here:

Hasnine, M. S. and Habib, K. N. (2018) 'What about the dynamics in daily travel mode choices? A dynamic discrete choice approach for tour-based mode choice modelling', Transport Policy. Elsevier Ltd, 71(August), pp. 70–80. doi: 10.1016/j.tranpol.2018.07.011.

Hasnine and Habib (2018)

A brief introduction to internal validation (data splitting methods)

Performance measures

Percentage of correct predictions: the alternative with the highest probability is defined as the predicted choice. However,

Model A：

- **Alt. A: 0.34＊**
- Alt. B: 0.33
- Alt. C: 0.33

Model B：

- **Alt. A: 0.50＊**
- Alt. B: 0.30
- Alt. C: 0.20

Model C：

- **Alt. A: 0.90＊**
- Alt. B: 0.05
- Alt. C: 0.05

＊Observed choice

Cannot discriminate differences in estimated probabilities.

A measures that accounts for “clearness” of prediction is necessary.

A brief introduction to internal validation (data splitting methods)

Performance measures

Clearness of prediction:

Percentage of clearly right choices: *"the percentage of users in the sample whose observed choices are given a probability greater than threshold t by the model"*

$$\%\mathrm{CR} = \frac{100}{N_v}\sum_{n_v=1}^{N_v} CR_{n_v} \quad where \quad \mathrm{CR}_{n_v} = \begin{cases} 1 \; if \; \hat{P}\left(y_{n_v}^{e}\right) > t \\ 0 \quad otherwise \end{cases}$$

Percentage of clearly wrong choices: *"the percentage of users in the sample for whom the model gives a probability greater than threshold t to a choice alternative differing form the observed one"*

$$\%\mathrm{CW} = \frac{100}{N_v}\sum_{n_v=1}^{N_v} CW_{n_v} \quad where \quad CW_{n_v} = \begin{cases} 1 \; if \; \hat{P}\left(! \, y_{n_v}^{e}\right) > t \\ 0 \quad otherwise \end{cases}$$

$\hat{P}\left(! \, y_{n_v}^{e}\right)$ is the estimated choice probability of an alternative other than the chosen one.

de Luca and Cantarella (2009)

A brief introduction to internal validation (data splitting methods)

Performance measures

Clearness of prediction: defining threshold t

- To be meaningful, the threshold t must be “considerably larger” than c^{-1}, where c is the choice set size.
- Values used in the literature:
  - Binary model： $t = 0.9$ (de Luca and Di Pace, 2015)
  - Trinary model： $t = 0.5$ (Glerum, Atasoy and Bierlaire , 2014)

Image removed due to copyright issues
See original article here:

de Luca, S. De and Cantarella, G. E. (2009) ‘Validation and comparison of choice models’, in Saleh, W. and Sammer, G. (eds) Travel Demand Management and Road User Pricing: Success, Failure and Feasibility. Ashgate publications, pp. 37–58.

See appendix for a list of commonly used indicators

Validation and reporting practices in the transportation academic literature

226 articles reviewed by Parady, Ory and Walker (2021)

92% reported a goodness of fit statistics

64.6% reported a policy-related inference

Marginal effects, elasticities, odds ratios, value of time estimates, marginal rates of substitution, and policy scenario simulations

18.1% reported a validation measure

Table 3

Internal validation methods reported in the literature by frequency.

Method	Abbvr.	Frequency	Percentage
Holdout validation	HOV	18	56.3%
Repeated learning-testing	RLT	8	25.0%
Validation against an independent sample	IS	4	12.5%
Repeated K-fold cross-validation	R-K-CV	1	3.1%
Other sample splitting methods	SS-O	1	3.1%

Towards better validation practices in the field

- **Make model validation mandatory:**
  - Non-negotiable part of model reporting and peer-review in academic journals for any study that provides policy recommendations.
  - Cross-validation is the norm in machine learning studies.

- **Share benchmark datasets:**
  - A fundamental limitation in the field is the lack of benchmark datasets and a general culture of sharing code and data.

- **Incentivize validation studies:**
  - Lot of emphasis on theoretically innovative models.
  - Encourage submissions that focus on proper validation of existing models and theories.

- **Draw and enforce clear reporting guidelines:**
  - In addition to detailed information of survey characteristics such as sampling method, discussion on representativeness of the data, validation reporting is required.
  - Efforts to improve reporting are well documented in other fields
    (i.e. STROBE statement (von Elm et al., 2007))

Wait a minute…

"I'm not validating my model because I'm not trying to build a predictive framework. I'm trying to learn about travel behavior"

The more orthodox the type of analysis conducted (such as the dimensions of travel behavior covered in this study), **the stronger the onus of validation.**

Wait a minute…

“Does every study that uses a discrete choice model should be conducting validation?”

In short, yes. At the very least, **any article that makes policy recommendations should be subject to proper validation** given the dependence of the field on cross-section observational studies, and the lack of a feedback loop in academia.

Wait a minute…

"Is what we learn about travel behavior from coefficient estimation less valuable if not conducted?"

There is a myriad of reasons why some **skepticism is warranted** against any particular model outcome. the most obvious one being model overfitting.

Finally

Better validation practices will not solve the credibility crisis in the field, but it's a step in the right direction.

Model validation is **no solution to the causality problem** in the field, but we want to underscore that **the reliance on observational studies inherent to the field demands more stringent controls to improve external validity of results**.

References:

1. Ben-Akiva, M. E., Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand. MIT press.

2. Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). Applied choice analysis: a primer. Cambridge University Press. 2nd Edition.

3. M. Baker, D. Penny (2016) Is there a reproducibility crisis? Nature, 533 (7604) pp. 452-454

4. Parady, G., Ory, D., Walker, J. (2021) The overreliance on statistical goodness of fit and under-reliance on validation in discrete choice models: A review of validation practices in the transportation academic literature . Journal of Choice Modelling 38, 100257 (Open Access)

5. de Luca, S. De and Cantarella, G. E. (2009) 'Validation and comparison of choice models', in Saleh, W. and Sammer, G. (eds) Travel Demand Management and Road User Pricing: Success, Failure and Feasibility. Ashgate publications, pp. 37–58. doi: 10.1017/cbo9780511619960.008.

6. Glerum, A., Atasoy, B. and Bierlaire, M. (2014) 'Using semi-open questions to integrate perceptions in choice models', Journal of Choice Modelling. Elsevier, 10(1), pp. 11–33. doi: 10.1016/j.jocm.2013.12.001.

7. Hasnine, M. S. and Habib, K. N. (2018) 'What about the dynamics in daily travel mode choices? A dynamic discrete choice approach for tour-based mode choice modelling', Transport Policy. Elsevier Ltd, 71(August), pp. 70–80. doi: 10.1016/j.tranpol.2018.07.011.

Following Random utility theory

$$P(i) = \int_{\varepsilon=-\infty}^{+\infty} F_i(V_i - V_1 + \varepsilon, V_i - V_2 + \varepsilon, \ldots, V_i - V_j + \varepsilon)\, d\varepsilon \quad (1)$$

where

$F(\,)$ is a CDF of disturbances $(\varepsilon_1, \ldots, \varepsilon_J)$ (2)

$F_i(\,)$: $\partial F(\,)/\partial \varepsilon_i$; Partial derivative of $F(\,)$ with respect to ε_i.

the GEV is obtained from the following CDF

$$F(\,) = \exp(-G(e^{-\varepsilon_1}, \ldots, e^{-\varepsilon_J}))$$

where G is a generating function.

Using equations (1) and (2) we get

$$P(i) = \int_{\varepsilon=-\infty}^{+\infty} \frac{\partial \exp(-G(e^{-\varepsilon - V_i + V_1}, \ldots, e^{-\varepsilon - V_i + V_J}))}{\partial \varepsilon_i}\, d\varepsilon$$

$$P(i) = \int_{\varepsilon=-\infty}^{+\infty} e^{-\varepsilon} G_i(e^{-\varepsilon - V_i + V_1}, \ldots, e^{-\varepsilon - V_i + V_J}) \cdot \exp(-G(e^{-\varepsilon - V_i + V_1}, \ldots$$

this integral results in

$$P(i) = \frac{e^{V_i} \cdot G_i(e^{V_1}, \ldots, e^{V_J})}{\mu G(e^{V_1}, \ldots, e^{V_J})}$$

東京大学
THE UNIVERSITY OF TOKYO

Appendix: Definition of model validation performance measures reported in the literature

Parady, Ory & Walker (2021)

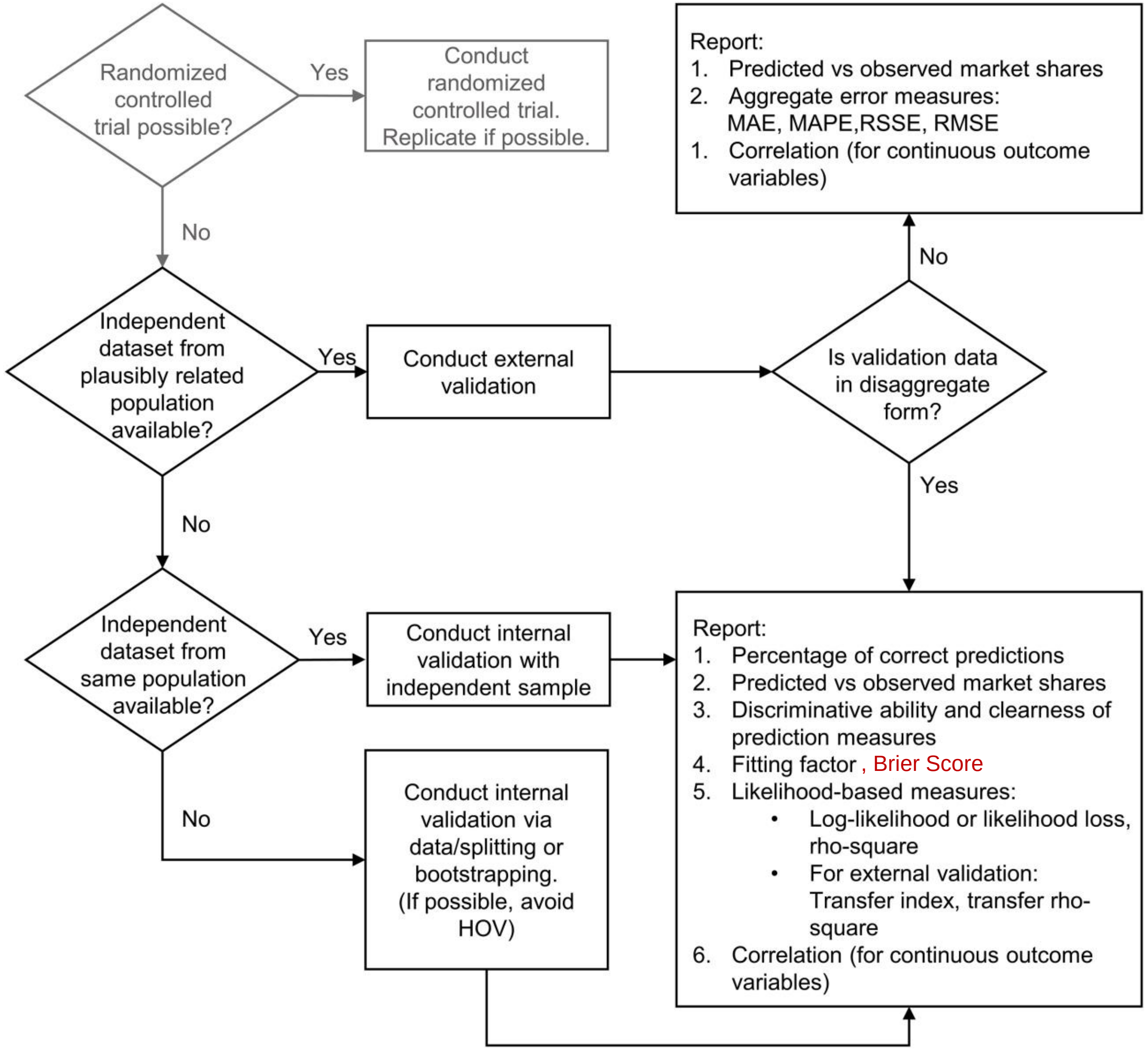
Index		Type	Formula	Notes
Mean absolute percentage error 平均絶対誤差率	MAPE	*Absolute*	$\frac{100}{M}\sum_{m=1}^{M}\left\lvert\frac{\hat{s}^{e}{}_{v,m}-s_{v,m}}{s_{v,m}}\right\rvert$	M is the number of alternatives in the choice set.
Root sum of square error 二乗平方根誤差和	RSSE	Relative	$\sqrt{\sum_{m=1}^{M}\left(\hat{s}^{e}{}_{v,m}-s_{v,m}\right)^{2}}$	$s_{v,m}$ is an aggregate outcome measure in sample *v*, such as the market share of alternative *m* (i.e. modal market share), choice frequency, etc.
Mean absolute error 平均絶対誤差	MAE	Aggregate: Relative Disaggregate: Absolute	$\frac{1}{M}\sum_{m=1}^{M}\left\lvert\hat{s}^{e}{}_{v,m}-s_{v,m}\right\rvert$	$\hat{s}^{e}{}_{v,m}$ is an aggregate outcome measure in sample *v*, such as the market share of alternative *m*, predicted from model estimated on sample *e*.
Mean squared error 平均二乗誤差	MSE	Aggregate: Relative Disaggregate: Absolute	$\frac{1}{M}\sum_{m=1}^{M}\left(\hat{s}^{e}{}_{v,m}-s_{v,m}\right)^{2}$	
Root mean square error 二乗平均平方根誤差	RMSE	Aggregate: Relative Disaggregate: Absolute	$\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\hat{s}^{e}{}_{v,m}-s_{mv,}\right)^{2}}$	$\hat{P}\left(y^{e}_{n_v,m}\right)$ is the predicted probability that individual *n* chooses alternative *m*, predicted from model estimated on sample *e*. y_{nm} is the actual outcome variable valued 0 or 1.
Brier Score ブライアスコア	BS	Absolute	$\frac{1}{N_v}\sum_{n_v=1}^{N_v}\sum_{m=1}^{M}\left(\hat{P}\left(y^{e}_{n_v,m}\right)-y_{n_v,m}\right)^{2}$	

Appendix: Definition of model validation performance measures reported in the literature

Parady, Ory & Walker (2021)

Index	Type		Formula	Notes
Log-likelihood 対数尤度	LL	Relative	$LL_v(\widehat{\boldsymbol{\beta}}^e)$	$LL_{v,r}(\widehat{\boldsymbol{\beta}}^e)$ is log-likelihood of the model estimated on data *e* applied to the validation data v_r.
Log-likelihood loss 対数尤度損失	LLL	Absolute	$\frac{1}{R}\sum_r -\frac{1}{N_{v,r}}\sum_{n_{v,r}} LL_{v,r}(\widehat{\boldsymbol{\beta}}^e)$ $\forall\ 1 \leq r \leq R$	$N_{v,r}$ is the size of the validation (holdout) sample r, and R is number of validation samples generated.
Rho-square σ^2	RHOSQ	Absolute	$\rho^2 = 1 - \frac{LL_v(\widehat{\boldsymbol{\beta}}^e)}{LL_v(\mathbf{0})}$	$LL_v(\mathbf{0})$ is log-likelihood of the model when all parameters are zero for data *v*. $LL_v(\widehat{\boldsymbol{\beta}}^v)$ is the likelihood of the model estimated on the validation data *v*.
Transfer rho-square 移転 σ^2	T-RHOSQ	Relative	$\rho^2_{transfer} = 1 - \frac{LL_v(\widehat{\boldsymbol{\beta}}^e)}{LL_v(\boldsymbol{MS}^v)}$	
Transfer index 移転指標	TI	Pass/Fail	$\frac{LL_v(\widehat{\boldsymbol{\beta}}^e) - LL_v(\boldsymbol{MS}^v)}{LL_v(\widehat{\boldsymbol{\beta}}^v) - LL_v(\boldsymbol{MS}^v)}$	$LL_v(\boldsymbol{MS}^v)$ is a base model estimated on validation data *v* (i.e. market share model.)
Transferability test statistic 移転性検定統計量	TTS	Relative	$-2\left(LL_v(\widehat{\boldsymbol{\beta}}^v) - LL_v(\widehat{\boldsymbol{\beta}}^e)\right)$	ρ^2_{local} is the local rho-square of the model.
χ^2 **test**	CHISQ	Pass/Fail	$\sum_{m=1}^{M} \frac{\left(f_m - E\left(f^e_{v,m}\right)\right)^2}{E\left(f^e_{v,m}\right)}$	f_m is the observed choice frequency of alternative m in sample *v*, and $E\left(f^e_{v,m}\right)$ is the expected choice frequency predicted from model estimated on sample *e*.

Appendix: Validation and reporting practices in the transportation academic literature



Heuristic to select validation method given available resources and recommended performance measures to report

Appendix: Validation and reporting practices in the transportation academic literature

Table 4
Predictive accuracy performance measures reported in the literature by frequency.

Performance measure	Abbrv.	Frequency	Percentage
Log-likelihood/log-likelihood loss	LL/LLL	19	46.3%
Percentage of correct predictions or First Preference Recovery	FPR	10	24.4%
Predicted vs observed market outcomes	PVO	10	24.4%
Mean absolute error	MAE	6	14.6%
Root mean square error	RMSE	4	9.8%
Error/Percentage error/Absolute percentage error	E/PE/APE	3	7.3%
Rho-Square	RHOSQ	3	7.3%
Transfer index	TI	2	4.9%
% clearly right (t)	%CR	1	2.4%
Brier Score	BS	1	2.4%
Chi-square	CHISQ	1	2.4%
Concordance index	C	1	2.4%
Correlation	CORR	1	2.4%
Fitting factor	FF	1	2.4%
Mean absolute percentage error	MAPE	1	2.4%
Sum of square error	SSE	1	2.4%
Transferability test statistic	TTS	1	2.4%
All other measures specified in Table 1	–	0	0%
Other measures not specified in Table 1	–	3	7.3%

Very similar measures are reported jointly.