Modelling Annual VKT Considering Heterogeneous Approximation Biases

Toshiyuki Yamamoto

Noritaka Nakagawa

NU TREND
Nagoya University
TRansport and ENvironment Dynamics

Outline

- *Background*
- Objective
- Data
- Analysis
- Future tasks

Annual vehicle kilometres travelled

VKT (vehicle kilometres travelled)

- has been used as an index of car use
  - The strongest indicator of car dependencies and household's travel patterns
- There have been many studies to make use of VKT for various purposes
  - Gasoline consumption, vehicle emissions, and crashes

Difficulty in modelling VKT

Generally, goodness-of-fit is low

- R^2: 0.11 (Train, 1986), 0.15 (Kockelman, 1997), 0.17 (Yamamoto et al., 2001)

Reason might be

- Variability among household's vehicle use
  - Factors to affect car use are not fully incorporated
- Inaccuracy in observation
  - Annual VKT reported by respondents
  - Short-period odometer readings

Outline

- Background
- *Objective*
- Data
- Analysis
- Future tasks

Objectives

- Inaccuracy in observation is examined
- Annual VKT model is developed considering inaccuracy in observation
  - Efficiency is compared with conventional models
- Heterogeneity among respondents in inaccuracy of observation is also examined

Incomplete data

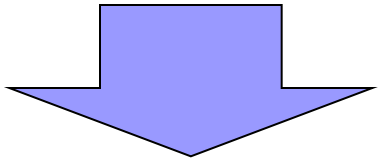
- *Missing data*: each data value is either perfectly known or entirely unknown
- *Coarse data*: only a subset of the complete-data sample space is observed
  - *Rounding*: data value is observed only to the nearest integer
  - *Censoring*: in failure time data, if an item has not failed by the time observation ends, failure time is known only to lie beyond the last observation point

Heaping

- one of the coarsening. related with rounding.
- includes the phenomenon known as digit preference.
- includes items reported with various levels of coarseness
- E.g., histograms of age often exhibit heaps at common ages such as integral multiples of ten years with adults, or integral multiples of six or twelve months with children.

Coarseness in VKT data

- Annual VKT reported by respondents includes some level of approximation
- Level of approximation may vary among respondents



VKT data is regarded as heaped

Outline

- Background
- Objective
- *Data*
- Analysis
- Future tasks

Parc-Auto

- French households' car ownership panel data
- Conducted yearly since 1976, and continues today
- Sample size is maintained at about 7,000 households each year
- Includes characteristics of up to 3 cars in the household, vehicle use, general attitudes concerning transportation, etc.

VKT data in Parc-Auto

2 types of information

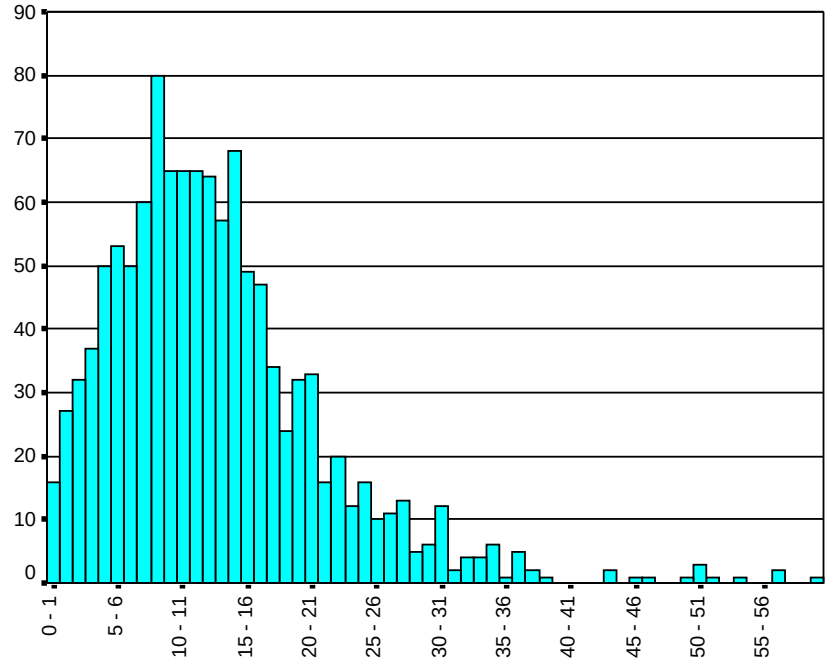
- Difference in odometer readings at 2 successive years -> *Calculated VKT*
- Annual mileage in kilometres reported by respondent -> *Reported VKT*

We use for analysis 1167 sample cases

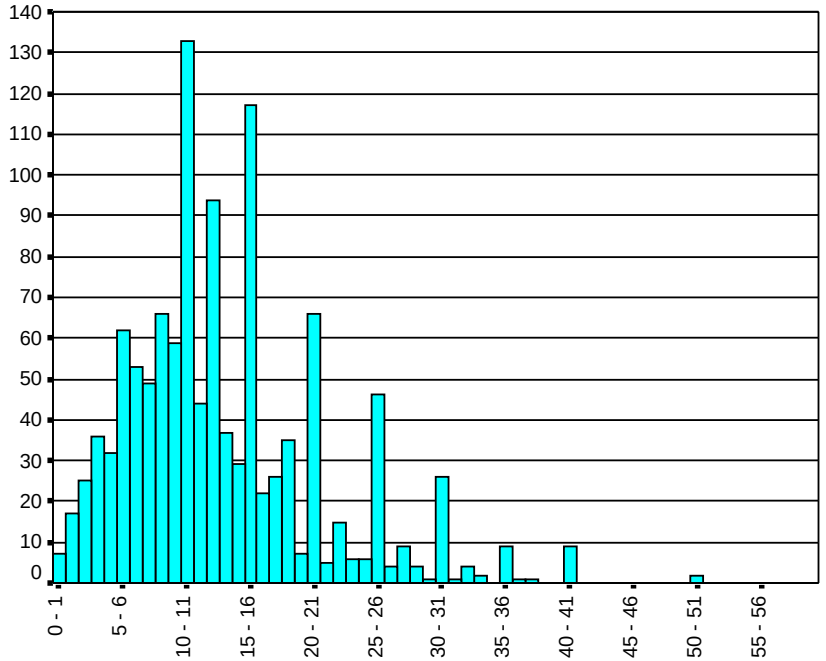
- 1998 VKT data
- Sub-sample who answered both 1997 & 1998 survey to get *Calculated VKT*

Sample distribution

- Reported VKT is obviously heaped

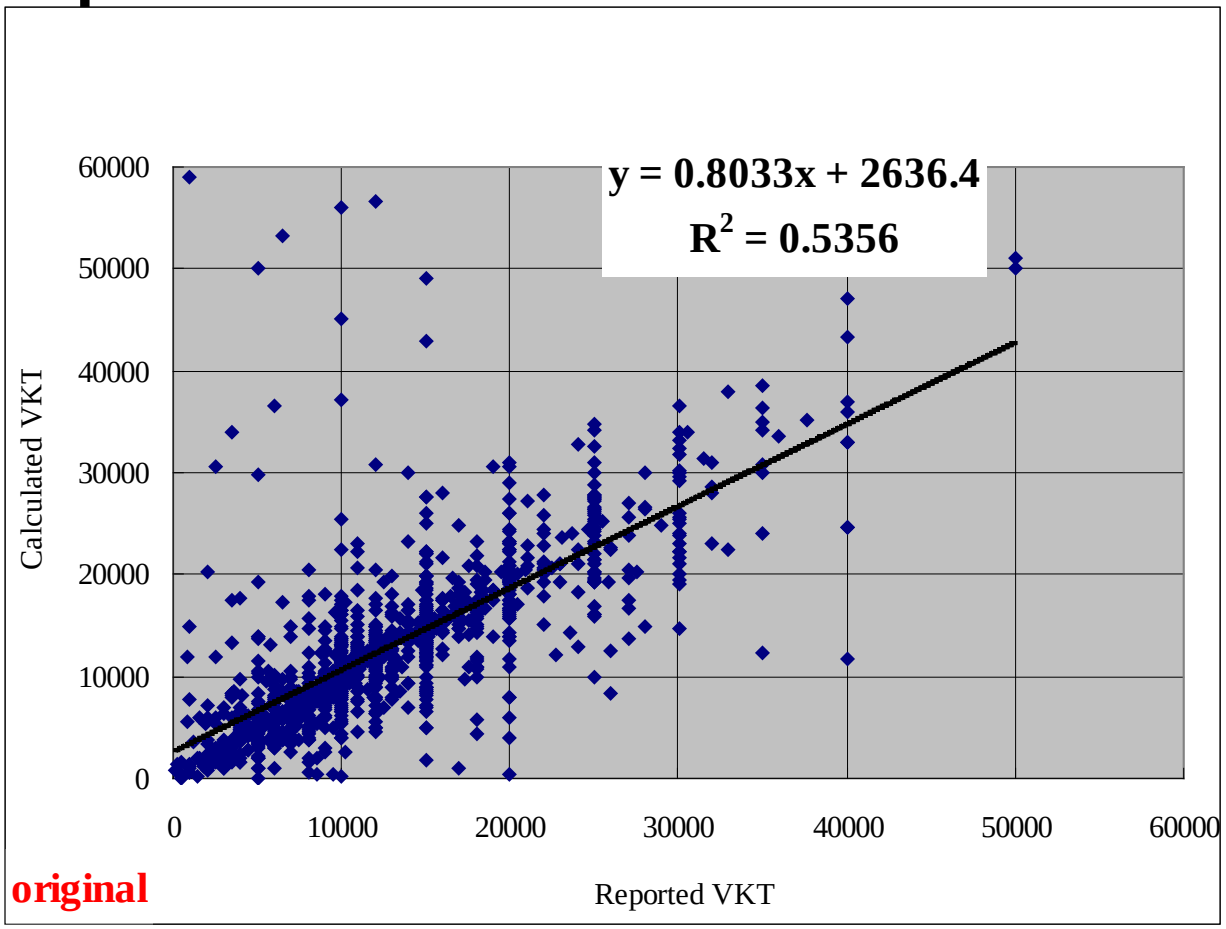


Calculated VKT



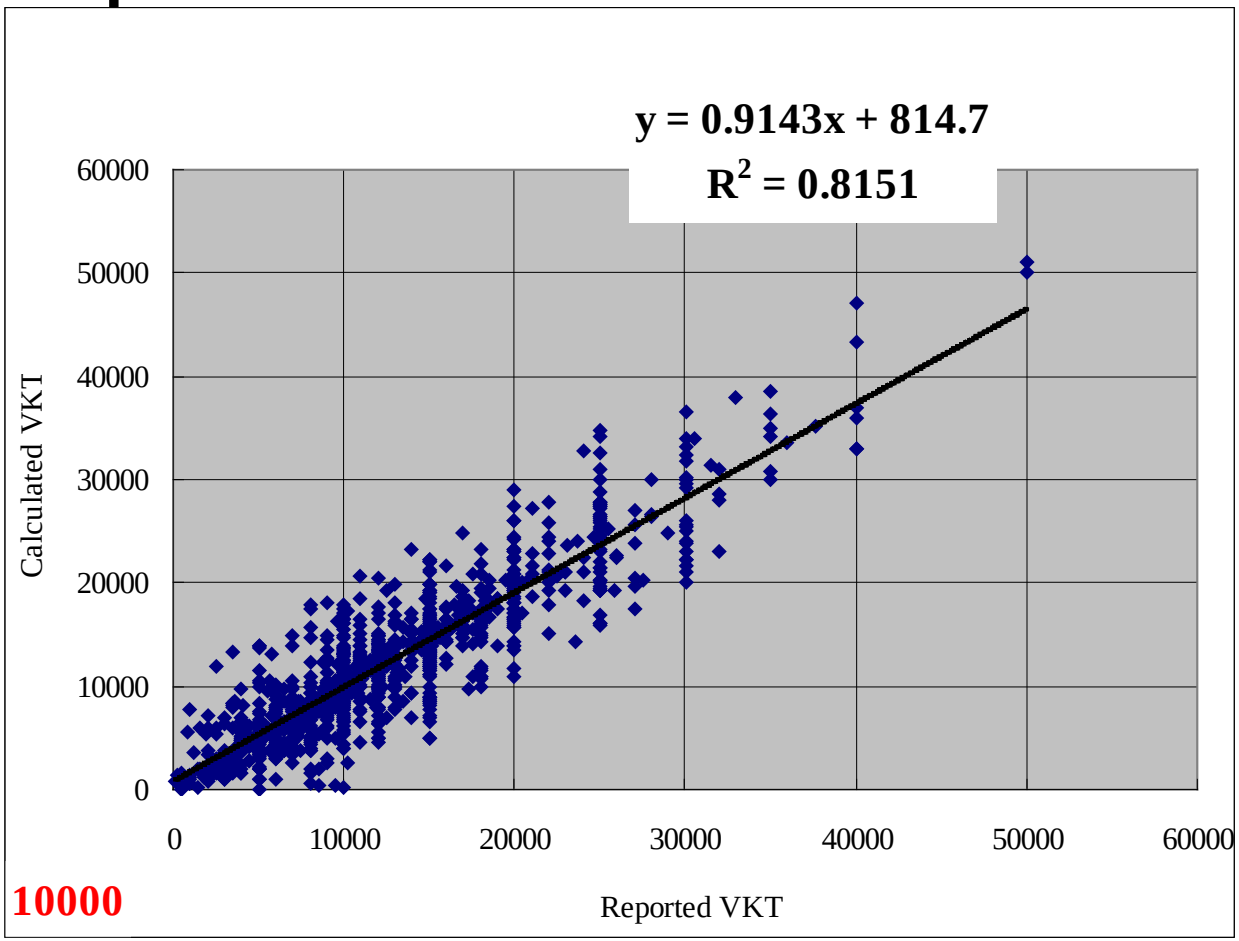
Reported VKT

Relationship between calculated and reported VKTs



- Some cases have large discrepancies

Relationship between calculated and reported VKTs



- Cases with 10000+ difference are discarded

Outline

- Background
- Objective
- Data
- *Analysis*
- Future tasks

Procedure of analysis

1. Regression models of both calculated and reported VKT
2. Regression models of error in reporting assuming that calculated VKT is true
3. Ordered-probit models of VKT using pre-determined coarseness
4. Ordered-probit model of possible maximum coarseness of the report by each respondent (heterogeneity among respondent)
5. Latent class model of VKT considering heterogeneity in coarseness among respondents

Explanatory variables

- Household's attribute
  - #children (15-), PT access., large city (300,000+), #cars, low income (F75,000-), high income (F200,000+)
- Personal attribute
  - Young (39-), old (60+), worker, male, car commute
- Car attribute
  - Diesel car, small car, large car, truck, car age

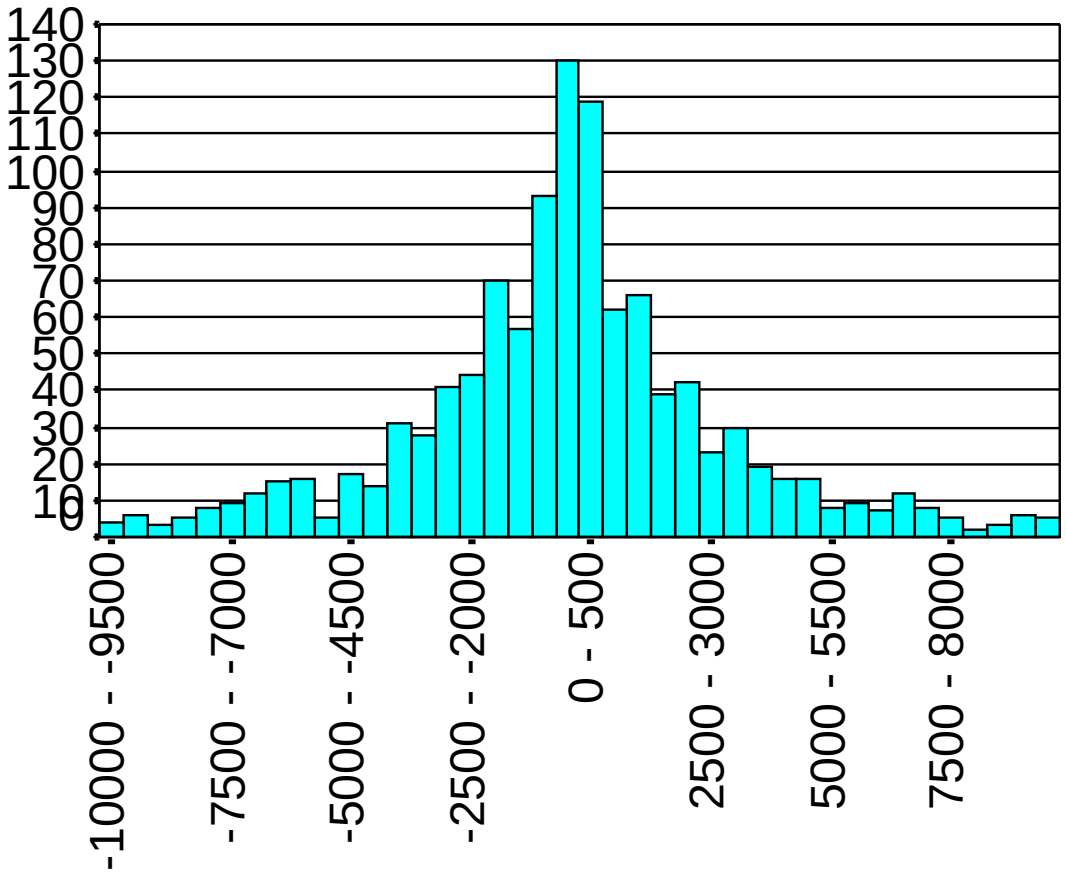
1. Regression models of VKT

- Unexpectedly, R2 is higher in reported VKT model
- Most var. have similar significance in both models

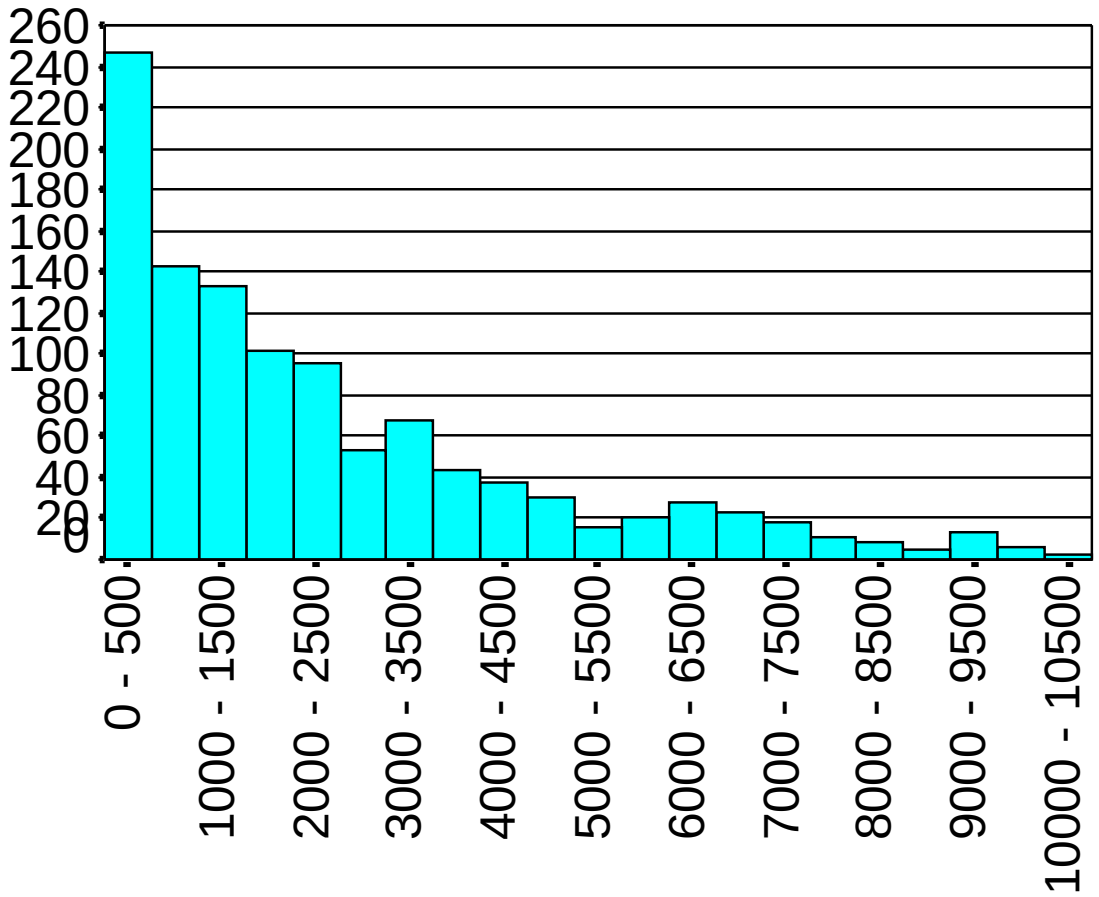
	Calculated		Reported	
	Coef.	t-stat.	Coef.	t-stat.
Intercept	9.380	70.9	9.345	81.5
#children	-0.015	-0.5	-0.014	-0.5
PT access	-0.070	-1.4	-0.061	-1.4
Large city	0.116	2.1	0.013	0.3
#cars	-0.038	-0.9	0.017	0.5
Low income	-0.115	-1.4	-0.202	-2.8
High income	0.041	0.8	0.057	1.3
Young	0.095	1.4	0.105	1.8
Old	-0.326	-3.8	-0.201	-2.7
Worker	-0.137	-1.7	-0.074	-1.0
Male	0.115	2.3	0.110	2.6
Car commute	0.390	6.3	0.357	6.6
Diesel car	0.389	8.2	0.379	9.2
Small car	-0.269	-5.3	-0.174	-3.9
Large car	0.163	2.1	0.160	2.4
Truck	0.536	1.5	0.579	1.8
Car age	-0.037	-6.6	-0.038	-7.9
s.e.	0.634		0.550	
R2	0.324		0.347	
Sample size	975		975	

Distribution of error in reporting

- (Calculated VKT) – (Reported VKT)



Distribution of absolute error in reporting



2. Regression models of error in reporting

- R2 are low in both models
- Calculated VKT has a highly significant coef.

	W/O calculated VKT		W calculated VKT	
	Coef.	t-stat.	Coef.	t-stat.
Intercept	766	1.119	-1410	-2.0
#children	4	0.0	60	0.4
PT access	96	0.4	199	0.8
Large city	499	1.7	329	1.2
#cars	-140	-0.6	-132	-0.6
Low income	561	1.3	706	1.7
High income	-173	-0.7	-294	-1.2
Young	-75	-0.2	-264	-0.8
Old	-1396	-3.2	-888	-2.1
Worker	-1054	-2.5	-901	-3.6
Male	-210	-0.8	-762	-1.9
Car commute	136	0.4	-178	-0.7
Diesel car	-78	-0.3	-1018	-2.7
Small car	-508	-1.9	-2699	-1.5
Large car	-779	-2.0	-308	-1.3
Truck	-1184	-0.6	-619	-2.0
Car age	33	1.2	87	3.1
Calculated VKT			0.17	9.5
s.e.	3250		3106	
R2	0.016		0.101	
Sample size	965		965	

2. Regression models of absolute error

- Logarithm of absolute error is used as dependent variable
- R2 are low in both models
- No significant var. except #cars

	W/O calculated VKT		W calculated VKT	
	Coef.	t-stat.	Coef.	t-stat.
Intercept	7.513	25.4	6.546	8.9
#children	0.033	0.5	0.035	0.5
PT access	0.126	1.1	0.134	1.2
Large city	-0.007	-0.1	-0.020	-0.2
#cars	-0.225	-2.4	-0.221	-2.3
Low income	-0.037	-0.2	-0.025	-0.1
High income	0.002	0.0	-0.003	0.0
Young	-0.044	-0.3	-0.053	-0.4
Old	-0.142	-0.7	-0.107	-0.6
Worker	-0.020	-0.1	0.152	1.4
Male	-0.079	-0.7	-0.006	0.0
Car commute	0.163	1.2	-0.176	-1.5
Diesel car	0.192	1.8	0.088	0.5
Small car	-0.204	-1.8	0.074	0.1
Large car	0.105	0.6	-0.092	-0.8
Truck	0.129	0.2	0.123	0.9
Car age	-0.018	-1.4	-0.014	-1.1
Calculated VKT			0.103	1.4
s.e.	1.406		1.406	
R2	0.010		0.011	
Sample size	965		965	

3. Ordered-probit models of VKT using pre-determined coarseness

- Reports are assumed as rounded as multiples of 500, 1000, or 5000 km regardless of reported value

$$\ln\left(VKT^{*}\right) = \beta X + \varepsilon$$

$$VKT \text{ is observed if } VKT - \delta \le VKT^{*} < VKT + \delta$$

$$\Pr\left(VKT\right) = \Phi\left(\frac{\ln\left(VKT + \delta\right) - \beta X}{\sigma}\right) - \Phi\left(\frac{\ln\left(VKT - \delta\right) - \beta X}{\sigma}\right)$$

δ: 500, 1000, or 5000

σ: standard error (estimated while thresholds are fixed in this model)

1000 km rounding

Compared with regression model,

- explanatory var. have similar coef. estimates
- s.e. has a smaller value

	Regression		1000 km rounded	
	Coef.	t-stat.	Coef.	t-stat.
Intercept	9.345	81.5	9.340	88.8
#children	-0.014	-0.5	-0.016	-0.6
PT access	-0.061	-1.4	-0.053	-1.3
Large city	0.013	0.3	0.026	0.6
#cars	0.017	0.5	0.011	0.4
Low income	-0.202	-2.8	-0.197	-2.9
High income	0.057	1.3	0.057	1.4
Young	0.105	1.8	0.103	1.9
Old	-0.201	-2.7	-0.187	-2.7
Worker	-0.074	-1.0	-0.067	-1.0
Male	0.110	2.6	0.100	2.5
Car commute	0.357	6.6	0.344	6.9
Diesel car	0.379	9.2	0.373	9.8
Small car	-0.174	-3.9	-0.173	-4.2
Large car	0.160	2.4	0.159	2.5
Truck	0.579	1.8	0.564	1.9
Car age	-0.038	-7.9	-0.036	-8.0
s.e.	0.550		0.508	43.1
Log-likelihood				-3048.7
Sample size		975		975

Comparison of goodness-of-fit

Standard error (Log-likelihood* in parenthesis)

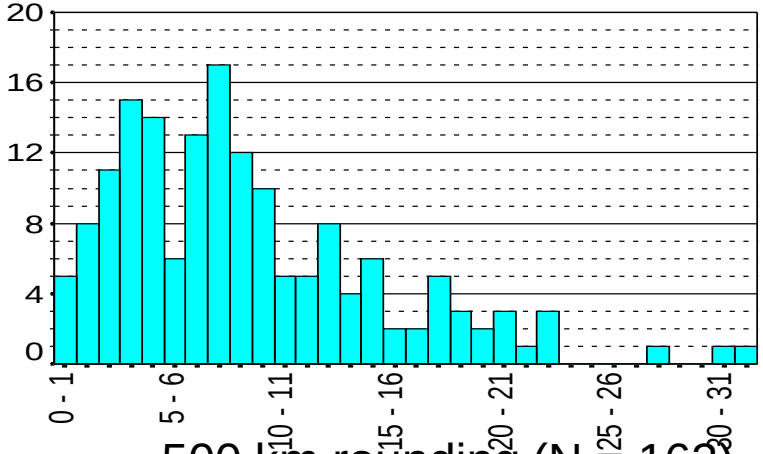
	Regression	500 km	1000 km	5000 km
Calculated VKT	0.634	0.603	0.583	0.526
	(0.324)*	(-3826)	(-3133)	(-1565)
Reported VKT	0.550	0.533	0.508	0.475
	(0.347)*	(-3753)	(-3049)	(-1492)

* R2 is shown for regression model

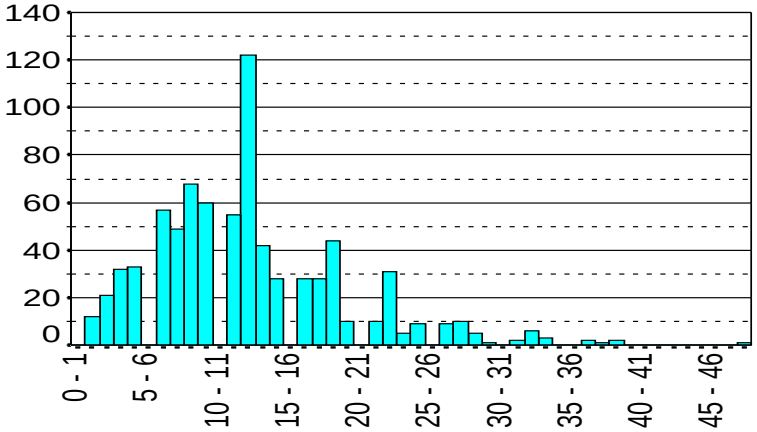
- Smaller standard errors are estimated when larger rounding is assumed

Sample distribution of possible maximum coarseness

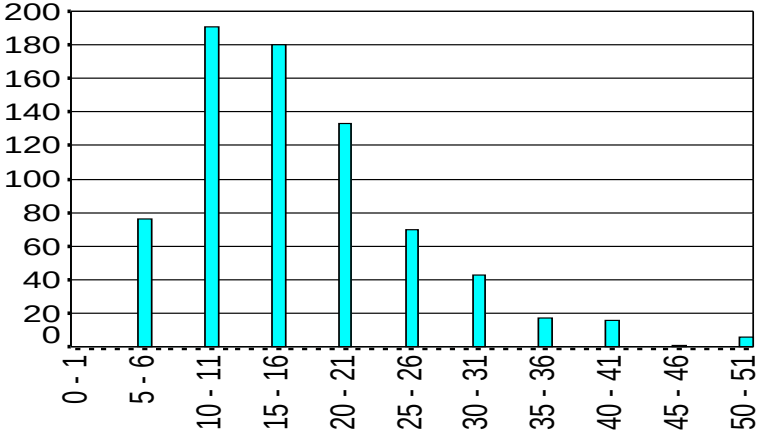
500 km rounding: 5500, 10500, …
1000 km rounding: 6000, 11000, …
5000 km rounding: 5000, 10000, …



500 km rounding (N = 163)



1000 km rounding (N = 786)



5000 km rounding (N = 732)

Sample distribution of possible maximum coarseness

	Pooled	500 km	1000 km	5000 km
Sample size	1919	163	786	732
Male	69.4%	71.8%	69.8%	68.7%
Young	19.5%	13.5%	18.1%	25.5%
Old	47.7%	62.6%	48.7%	37.8%
Low income	7.2%	7.4%	5.9%	8.2%
High income	25.7%	23.9%	25.4%	27.0%
Large city	82.6%	84.7%	81.2%	84.0%
Small car	28.8%	40.5%	28.9%	24.5%
Large car	8.5%	4.9%	8.9%	9.4%

Sample distribution of coarseness admitting multiple possibilities

	Pooled	500 km	1000 km	5000 km
Sample size	1919	1681	1518	732
Male	69.4%	69.5%	69.3%	68.7%
Young	19.5%	20.9%	21.7%	25.5%
Old	47.7%	45.3%	43.5%	37.8%
Low income	7.2%	7.0%	7.0%	8.2%
High income	25.7%	26.0%	26.2%	27.0%
Large city	82.6%	82.7%	82.5%	84.0%
Small car	28.8%	28.1%	26.7%	24.5%
Large car	8.5%	8.7%	9.2%	9.4%

Outline

- Background
- Objective
- Data
- Analysis
- *Future tasks*

Ordered-probit model of possible maximum coarseness of the report

- Conventional ordered-probit model with categories (500-, 1000, 5000+) will be estimated
- Sample distribution of possible maximum coarseness suggest that age, income, and car size may have significant effects

Latent class model of VKT considering heterogeneity in coarseness among respondents

- Latent class represents the level of coarseness in the report
  - Reported VKT of 6000 km belongs to both 500 km and 1000 km rounding classes
  - 12500 km only belongs to 500 km rounding class
- Ordered-probit model is applied to represent the class
- VKT is also developed as ordered-probit model