ESTIMATING A VEHICLE OWNERSHIP MODEL FROM TARGETED MARKETING DATA

Gregory S. Macfarlane

School of Civil and Environmental Engineering
and School of Economics

MOTIVATION

There are many emerging sources of sociodemographic information which are similar to traditional sources such as Census products or regional surveys.

Are these products adequate substitutes (or even replacements) for traditional sources in travel behavior analysis?

Outline

TARGETED MARKETING DATA

Created by credit reporting agencies by examining household's

- Account volumes and transactions
- Debt obligations and repayment histories
- Loan applications

Distilled into commercial data products that contain

- Demographics (age, race, children, spouse)
- Economics (income, mortgage, home value)
- Home Address
- Spending and lifestyle characteristics

MISSING ITEMS

There are some things we are used to getting that we don't get in TM

- Labor force participation
- Education
- Work place

OUR DATA

Sampled 500,000 vehicles from the Georgia DMV registration database.

Collected TM records for those addresses.

Joined vehicles to TM records by address.

VARIABLES IN THE MODEL

- *Income* is given by a set of dummy variables separated at $20, $50, and $100 thousand dollars.
- *Density* is the jobs and population of a zone divided by its acreage.
- *Sufficiency* for workers is the number of household workers with a car available under each alternative. Other's sufficiency is the number of residual cars.
- *Importance* is the relative importance of cars to accessibility in a zone,

$$Import = \frac{\sum_j S_j / T^2_{ijHWY}}{\sum_k \sum_j S_j / T^2_{ijk}} \tag{1}$$

where S_j is the employment and population a zone, and T_{ijk} is the loaded network travel time from i to j by mode k.

PUBLISHED MODEL

$$
\begin{aligned}
0\ \text{Cars} &= 0 \\
1\ \text{Car} &= -1.992 + 0.619 * \mathit{Inc1} + 1.929 * \mathit{Inc2} + 2.205 * \mathit{Inc3} \\
&\quad + 1.869 * \mathit{Inc4} - 0.115 * \ln(\mathit{Density}) \\
&\quad + 3.38 * \mathit{WSuff} * \mathit{Import} + 1.27 * \mathit{OSuff} * \mathit{Import} \\
2\ \text{Cars} &= -3.314 - 0.450 * \mathit{Inc1} + 1.559 * \mathit{Inc2} + 3.085 * \mathit{Inc3} \\
&\quad + 3.582 * \mathit{Inc4} - 0.295 * \ln(\mathit{Density}) \\
&\quad + 3.38 * \mathit{WSuff} * \mathit{Import} + 1.27 \mathit{OSuff} * \mathit{Import} \\
\geq 3\ \text{Cars} &= -3.482 + 1.76 * \mathit{Inc1} + 0.352 * \mathit{Inc2} + 2.324 * \mathit{Inc3} \\
&\quad + 3.141 * \mathit{Inc4} - 0.803 * \ln(\mathit{Density}) \\
&\quad + 3.38 * \mathit{WSuff} * \mathit{Import} + 1.27 * \mathit{OSuff} * \mathit{Import}
\end{aligned}
$$

ESTIMATING IT

Because our sampling methodology excludes 0-car households, we need to re-adjust the alternative-specific coefficient estimates in the Atlanta model.

We also need to predict whether someone is a worker or not; we estimate this model on ACS data from the Atlanta region.

ESTIMATED

	ARC (Norm)	Estimates	(se)
2:(intercept)	-1.32	-0.22	(0.25)
3:(intercept)	-1.49	-0.67	(0.28)
WorkSuff * Import	3.38	6.45*	(0.89)
OtherSuff * Import	1.27	2.22*	(0.71)
2:Lower-middle Income	0.08	0.5	(0.24)
3:	1.33	0.46	(0.23)
2:Upper-middle Income	2.16	0.74*	(0.23)
3:	-3.37	0.71*	(0.23)
2:High Income	-1.64	1.06*	(0.25)
3:	-0.49	1.06*	(0.24)
2:log(*Density*)	-0.18	-0.18	(0.07)
3:	-0.69	−0.25*	(0.08)
N		1996	
LL($\hat{\beta}$)		-2081.14	
ρ^2_C		0.0285	

SUFFICIENCY

The *Sufficiency* variable eliminates a number of records. `NA` records cannot be 0, because this would imply all of the workers had a car.

	1 Car	2 Cars	3+ Cars	Total
0 Workers	3513	2244	1335	7092
1 Worker	4145	3075	2002	9222
2 Workers	1309	2947	2717	6973
3 Workers	219	447	817	1483
4+ Workers	22	56	152	230
Sufficiency	1	2	3	NA

Insufficiency

We can turn this variable inside-out by looking at the number of workers *without* a car.

	1 Car	2 Cars	3+ Cars	Total
0 Workers	3513	2244	1335	7092
1 Worker	4145	3075	2002	9222
2 Workers	1309	2947	2717	6973
3 Workers	219	447	817	1483
4+ Workers	22	56	152	230
Sufficiency	3	2	1	0

INSUFFICIENCY

	Sufficiency		Insufficiency	
2:(intercept)	-0.22	(0.25)	-0.39	(0.07)
3:(intercept)	-0.67	(0.28)	-1.26	(0.08)
WorkSuff * Import	6.45*	(0.89)		
OtherSuff * Import	2.22*	(0.71)		
WorkInsuff * Import			−8.12*	(0.22)
OtherInsuff * Import			−4.91*	(0.33)
2:Lower-middle Income	0.5	(0.24)	−0.33*	(0.06)
3:	0.46	(0.23)	−0.43*	(0.08)
2:Upper-middle Income	0.74*	(0.23)	−0.52*	(0.06)
3:	0.71*	(0.23)	−0.71*	(0.07)
	· · · Witheld for clarity · · ·			
N	1996		24010	
LL($\hat{\beta}$)	-2081.14		-26257.44	
ρ^2_C	0.0285		0.0645	

WORKERS

Our worker imputation may be shaky; can we get valid estimates using Senior/Not Senior?

SENIORS

	Workers		Seniors	
2:(intercept)	-0.39	(0.07)	-0.39	(0.07)
3:(intercept)	-1.26	(0.08)	-1.26	(0.08)
WorkInsuff * Import	−8.12*	(0.22)		
OtherInsuff * Import	−4.91*	(0.33)		
AdultInsuff * Import			−8.14*	(0.22)
SeniorInsuff * Import			−4.51*	(0.31)
2:Lower-middle Income	0.33*	(0.06)	0.32*	(0.06)
3:	0.43*	(0.08)	0.44*	(0.07)
2:Upper-middle Income	0.52*	(0.06)	0.52*	(0.07)
3:	0.71*	(0.07)	0.71*	(0.08)
· · · Witheld for clarity · · ·				
N	24010		24010	
LL($\hat{\beta}$)	-24563.95		-24554.64	
ρ^2_C	0.0645		0.0649	

NEXT STEPS

- Use insufficiency as a generic variable, or find some other way to express adequacy without sacrificing records.
- Try to recover coefficient estimates for zero-vehicle households (we're actually closer than you might think).
- Examine other variables unique to TM data.