UNBIASED ESTIMATION OF DESTINATION CHOICE MODELS WITH ATTRACTION CONSTRAINTS

Vince Bernardin, PhD
Steven Trevino
John Gliebe, PhD

APRIL 14, 2014

RSG
the science of insight

WHAT'S WRONG WITH ESTIMATING DOUBLY CONSTRAINED DESTINATION CHOICE MODELS WITHOUT SHADOW PRICES

Vince Bernardin, PhD
Steven Trevino
John Gliebe, PhD

APRIL 14, 2014

Introduction

The Issue – Bias from Inconsistency

DESTINATION CHOICE MODELS INCREASINGLY COMMON

- 5% of MPOs in 2005
- At least 10% by 2013, probably 15% or more

DOUBLY CONSTRAINED MODELS COMMON IN APPLICATION

- Primarily for work, but also NHB, etc.

USUALLY SINGLY CONSTRAINED VERSION IS ESTIMATED

- Then calibrated for doubly-constrained application

THIS CAN LEAD TO BIASED PARAMETERS

- Proven for constrained choice models generally (Satsuma et al., 2011)
- Demonstrated by de Palma et al., 2007 for residential location choice

The Reason

(DOUBLY) CONSTRAINED MODELS ARE DIFFICULT

- Standard logit estimation software cannot estimate models with constraints
- On the one hand, just a generalization of doubly constrained gravity model (as Daly, 1982, nicely demonstrated)
- On the other hand, this turns out to be a difficult type of model, not GEV, a universal or mother logit model (McFadden et al., 1977)
- Some recent formulations in academia, but more focused on choice set formation (Zheng and Guo, 2008; Pagliara and Timmermans, 2009; Martinez et al., 2009)
- Without general theoretical structure, estimation algorithms relying on analytic gradients are not possible

A Solution

A GENETIC ALGORITHM (GA)

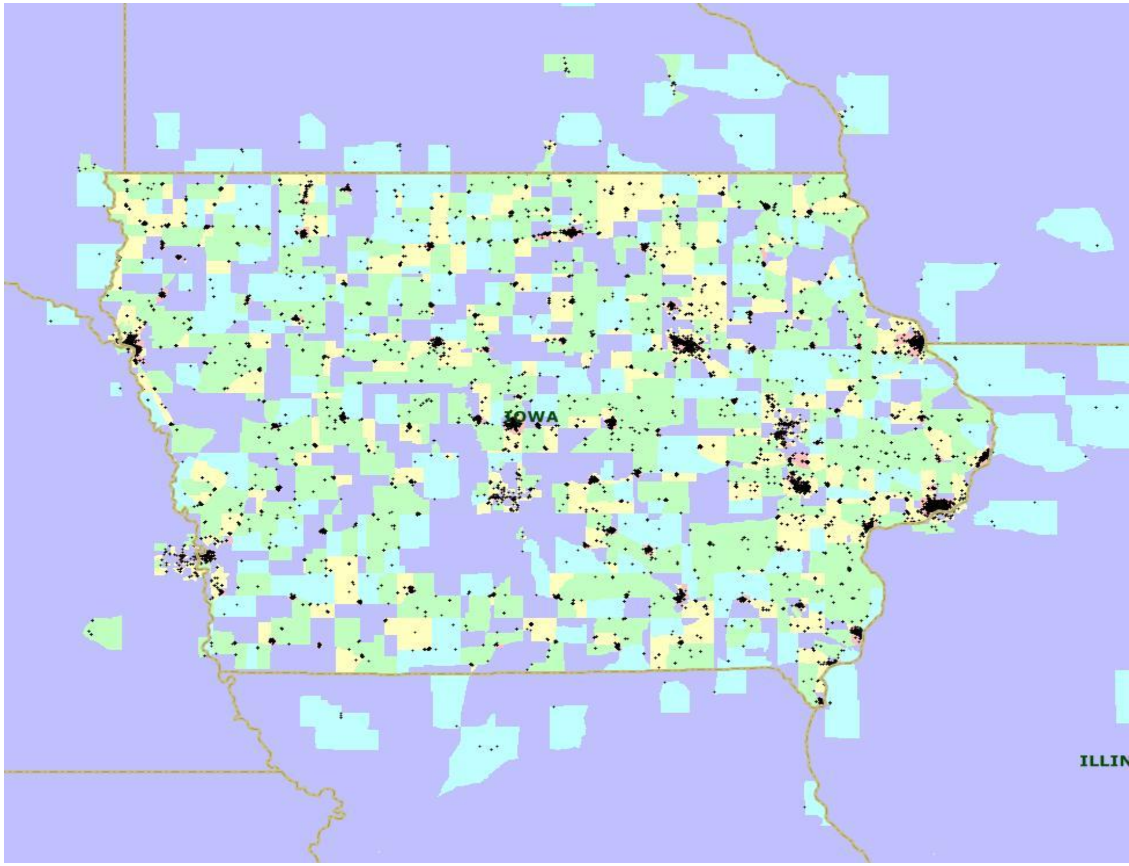
- Applied to estimate destination choice models for the new Iowa statewide model (iTRAM)
- GA used the model's application code for estimation
  - Reduces possibility for inconsistencies between estimation and application in general
- Both constrained and unconstrained versions of HBW model were estimated
- Results are compared to demonstrate the significance of parameter bias from estimating constrained model as if it were unconstrained

Data / Application

Iowa

ITRAM AND NHTS

- 3,314 zone trip-based statewide model
- 2,439 (1,745 weekday) household add-on sample to 2009 NHTS
- 1,992 HBW observations



Methodology

Bi-Level Formulation

UPPER LEVEL – LOG-LIKELIHOOD

$$Max_{\vec{\beta}} \sum_{obs} w_{ij} ln(P_{ij})$$

LOWER LEVEL – CONSTRAINED DESTINATION CHOICE

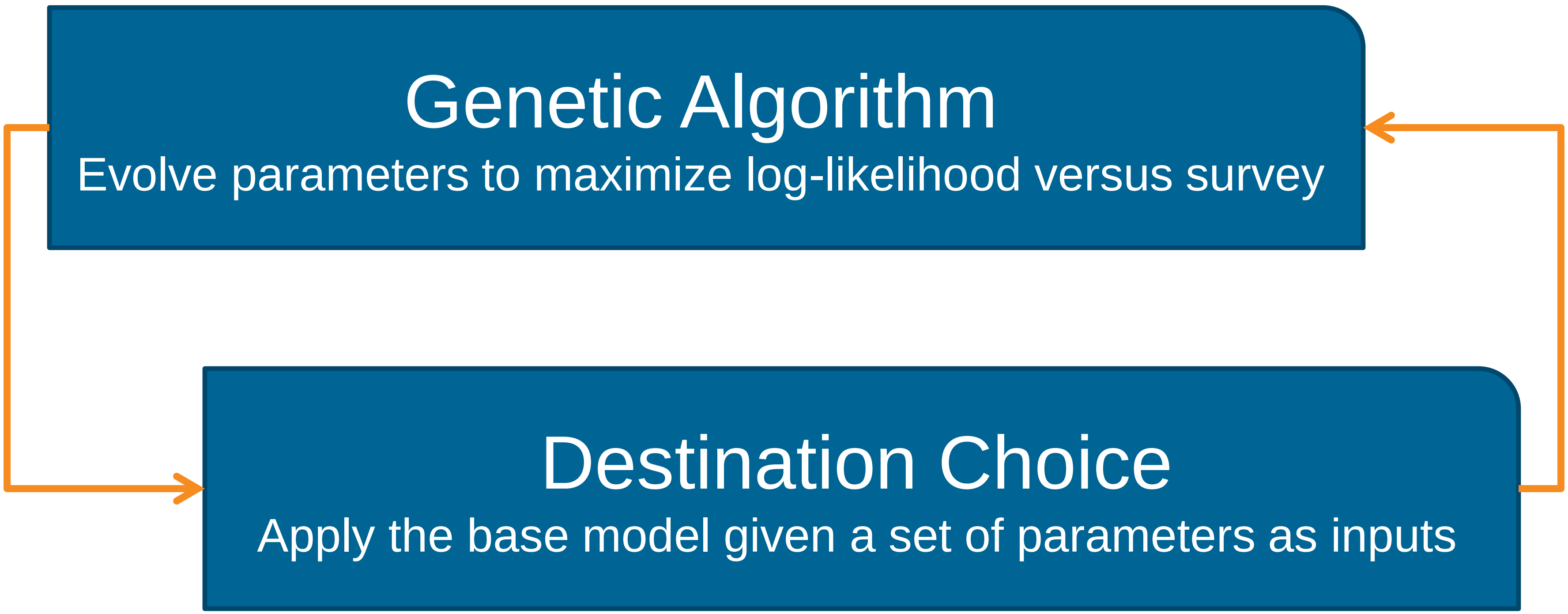
SUCH THAT:

$$P_{ij} = \frac{A_j e^{\vec{\beta}_{ij} \vec{x}_{ij}}}{\sum_{j'} A_{j'} e^{\vec{\beta}_{ij'} \vec{x}_{ij'}}}$$

$$\sum_i T_i P_{ij} = A_j \;\; \forall j$$

Metaheuristic

ITERATIVE BI-LEVEL PROGRAM



Genetic Algorithm

OVERVIEW

- Initial “population” of solutions
- Evaluate “fitness” of each solution
- Kill least fit solutions
- Create new generation of solutions by
  - Randomly mutating fit solutions
  - Combining fit solutions

Fitness

LOG-LIKELIHOOD

- Model’s (PA) trip table matrix normalized by dividing by row sums
- Produces probability matrix in which each row sums to 1
- Log of this matrix multiplied by matrix of weighted survey observations
- Although partially aggregate, no information loss

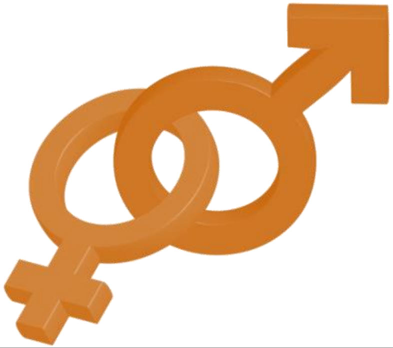
Mutation  and Combination

MUTATION

- Draw new parameter randomly from normal distribution around previous solution parameter
- Currently only mutating best solution
- A couple of ‘hyper-mutants’ (mutate all parameters) each generation

RE-COMBINATION

- ‘Mate’ two attractive solutions
- ‘Child’ solution has a 50% chance of getting each parameter from either parent solution



GA: Pros and Cons

PROS

- Robust to multiple optima – which are possible
- Reduces possibility for inconsistencies between estimation and application
- Allows inequality constraints on parameters $0 < \beta < \beta_{max}$
- Approach obviates need for sampling – improving the statistical efficiency of the estimator, better use of data
- Allows estimation of embedded decay parameters in accessibility variables

CONS

- Computationally intense
  - 23.9 days constrained
  - 20.7 days unconstrained
- (Need better distributed processing)

Results

Constrained Model is Better

RHO-SQUARED VS. ZEROS

- 0.216 constrained
- 0.189 unconstrained

CHI-SQUARED TEST

- could not reject unconstrained model with 3,314 degrees of freedom

Parameter Bias

MANY PARAMETERS SIGNIFICANTLY BIASED

Variable	Constrained	Unconstrained	Bias Sig.	Bias T
Total Employment	1.450	1.450	0.999	0.00
Theta	1.000	0.544	0.997	0.00
Accessibility to Employment	0.065	0.065	0.999	0.00
- decay	-0.571	-0.571		
Res. Accessibility x Impedance	-0.017	-0.020	0.000	6.52
Ln(Res. Accessibility x Impedance+1)	-0.381	-0.381	0.999	0.00
River Xing	-0.001	-0.019	0.000	5.19
RRD Xing	-0.222	-0.025	0.000	11.27
Interstate Xing	-0.005	-0.087	0.001	2.97
Different County	-0.690	-0.794	0.009	2.35
Intervening Rural Area	-0.003	-0.178	0.000	5.96
Intrazonal Constant	0.761	1.277	0.000	8.68
Intrazonal Gen. Accessibility	0.032	0.032	0.999	0.00
Intrazonal Gen. Accessibility Squared	-0.009	-0.009	0.999	0.00
Log-likelihood	-25334.0	-26197.0		
Rho-squared vs. zeros	0.216	0.189		

Other Findings

INTERVENING RURAL AREAS

- New psychological barrier
- Not especially significant in HBW but highly significant for HBO

ACCESSIBILITIES

- Confirmed findings on dual destination accessibilities (to substitutes and compliments)
- Confirmed value of residential accessibility and impedance interaction

Conclusion

Conclusions

DANGER OF ESTIMATION – APPLICATION INCONSISTENCY

- Most parameters biased by omission of constraints
- Difficult to correct for this with manual calibration
- Could draw wrong conclusions about parameter signficance
- Unconstrained destination choice model fit worse than doubly constrained gravity model
- Need for better estimation techniques

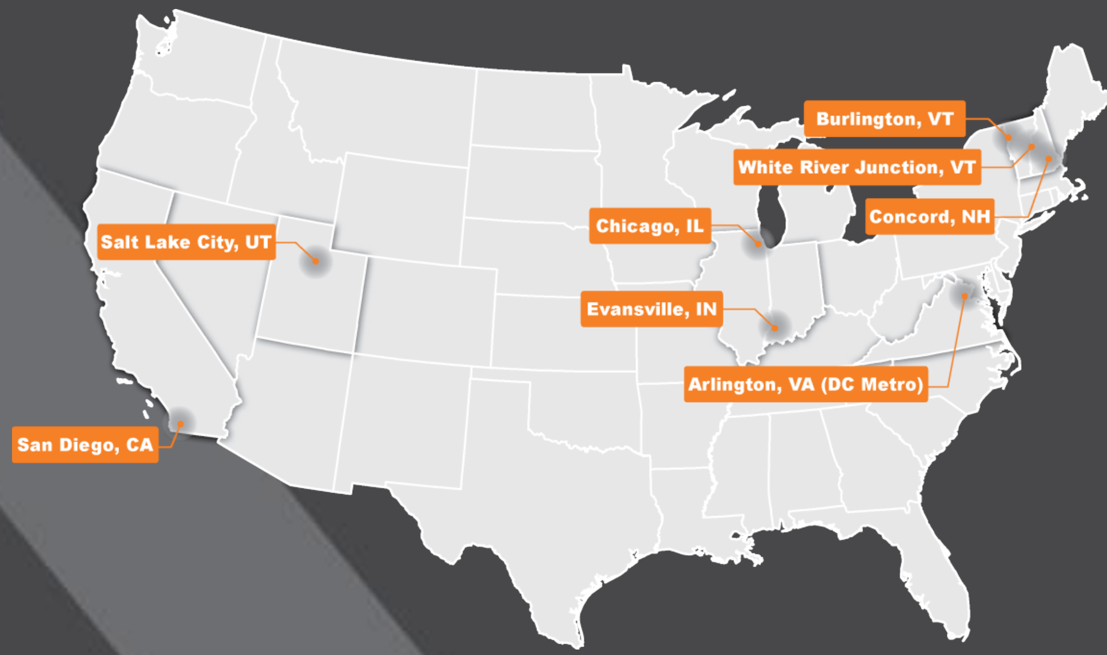
GENETIC ALGORITHM

- Many advantages
  - Robust, unbiased, statistically efficient estimator
  - Can handle constraints, embedded parameters, etc.
- Computationally challenging – need to improve implementation

Last Thoughts

WHAT ABOUT OTHER SIMPLIFYING ASSUMPTIONS?

- Established that sampling can lead to parameter bias in all but the simplest specifications
- Well known that estimating models without constants can bias parameter estimates
- Not enough survey data to estimate constants for destinations
  - Debate about whether we would want to / would they be stable
  - Model over-specification / saturation, identification issues
  - Some district constants on the other hand are not uncommon
- Now exploring simultaneous parameter estimation from household survey and traffic count data using the same genetic algorithm
  - May provide enough data to estimate constants
  - Still doesn't resolve whether or not we really want to
  - But may allow us to at least test if omission of constants leads to specification bias similar to omission of constraints



Contacts

www.rsginc.com

Vince Bernardin, PhD
Senior Consultant
Vince.Bernardin@RSGinc.com
812-200-2351

Steven Trevino, PhD
Analyst
Steven.Trevino@RSGinc.com
812-200-2351

John Gliebe, PhD
Senior Consultant
John.Gliebe@RSGinc.com
888-774-5986