Methodological Developments in Activity-Travel Behavior Analysis

David Brownstone
Email: dbrownst@uci.edu Web:
http://www.economics.uci.edu/~dbrownst/

Special thanks to: : Chandra Bhat, David Hensher, Stephane Hess, and Juan de Dios Ortúza

Typical Transportation Applications

- Mode/destination or mode/time of day choice
- Multiple responses to SP experiment by same respondent
- Household vehicle type choice
- Characterized by large number of discrete choices and complex correlation and heteroskedsticity patterns

10 Years ago

- Need better software for Mixed Logit
- Worried about empirical identification
- Stressed advantages of Bayesian methods (particularly Hierarchical Bayes)
- Multiple Imputation for measurement errors.
- Importance of proper inference for WTP and forecasts (including model uncertainty)

Mixed Logit

- Many advances in modeling heterogeneity – especially scale heterogeneity and intra-respondent heterogeneity.
- But there are serious problems using Maximum Simulated Likelihood to estimate these complex ML models – identification and inference.

Generalized ML (Hensher & Greene)

$$P_{i,j,t} = \frac{\exp\left(V_{itj}\right)}{\sum_{j=1}^{J}\exp\left(V_{itj}\right)}$$

$$V_{itj} = \boldsymbol{\beta}_i'\mathbf{x}_{itj}$$

$$\boldsymbol{\beta}_i = \sigma_i\left[\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{z}_i\right] + \left[\gamma + \sigma_i\left(1-\gamma\right)\right]\boldsymbol{\Gamma}\mathbf{v}_i$$

$$\sigma_i = \exp\left[\bar{\sigma} + \boldsymbol{\delta}'\mathbf{h}_i + \tau w_i\right]$$

$\mathbf{v}_i$ are iid $N\left(\mathbf{0},\mathbf{I}\right)$ and w_i are iid $N\left(0,1\right)$

Generalized ML 2

Note that the "hyperparameters" to be estimated are: $\boldsymbol{\beta}, \boldsymbol{\Delta}, \gamma, \boldsymbol{\Gamma}, \bar{\sigma}, \boldsymbol{\delta}, \tau$. It is possible to recover the distributions of $\boldsymbol{\beta}_i$ and σ_i , and this is a standard Bayesian calculation that has been used by Train, Hensher, and co-authors.

γ controls the relative importance of scale heterogeneity, and $\sigma_i = 1$ or τ=0 implies the standard random parameters ML model

WTP Space

When ML models are fit with random coefficients for time and cost, the implied distribution of WTP is frequently unrealistic. If ,$\boldsymbol{\Delta} = \gamma = 0$ and the element of $\boldsymbol{\beta}$ corresponding to cost is normalized to 1, then the model can be reparametrized as:

$$\boldsymbol{\beta}_i = \sigma_i\beta_c\begin{pmatrix}1\\ \left(\frac{1}{\beta_c}\right)\left(\boldsymbol{\beta}+\boldsymbol{\Gamma}\mathbf{v}_i\right)\end{pmatrix} = \sigma_i\beta_c\begin{pmatrix}1\\ \psi_c\left(\boldsymbol{\beta}+\boldsymbol{\Gamma}\mathbf{v}_i\right)\end{pmatrix}$$

This allows the distribution of WTP to be directly specified by the modeler.

ML Estimation

$$\log L = \sum_{i=1}^{N}\log\left(\int\prod_t\prod_j P_{ijt}\left(\Theta_i\mathbf{x}_{itj}\right)^{d_{ijt}} f\left(\Theta_i\right)d\Theta_i\right)$$

where Θ_i includes all slope and variance parameters.
MSL simulates draws of Θ_i and maximizes the simulated log likelihood function. Hierarchical Bayes simulates draws from the posterior distribution of the underlying hyperparameters conditional on the observed choices d_{ijt}. Both are trying to approximate the nasty integral, but the Bayesians don't need to maximize anything depending on the quality of this approximation.

Problems with MSL

- Simulation noise makes it hard to estimate weakly-identified models (Chiou and Walker, 2007)
- Need "sandwich estimator" for covariance, and this requires computing hessian of log likelihood (numerical derivatives don't work well)
- Not clear how to forecast – should we condition on observed choices?

Bayesian Methods

- Do not require maximization, so identification is less of a problem
- Confidence regions derived from draws from posterior distribution are stable
- Bayes estimators are asymptotically equivalent to Maximum Likelihood.
- Sensitivity to prior is an issue, but usually related to weak identification

Prior distribution $\pi(\theta)$ Likelihood function :

$f(x \mid \theta)$ Observe data and get posterior distribution:

$$p(\theta \mid x) = f(x \mid \theta)\pi(\theta) \Big/ \int f(x \mid \theta)\pi(\theta)d(\theta)$$

In many cases the posterior mean is the optimal estimator. The key problem is computing high-dimensional integrals.

- Bayes confidence intervals
  - fixed regions containing θ with specified coverage probability
  - conditions on observed data
- Classical confidence intervals
  - region with random endpoints containing true θ over independent repeated replications of the data
  - depends on distribution of unobserved realizations of the data

Bayesian Model Uncertainty

Suppose there are M competing models, and let

π_m Be the prior probability that model m is correct

$$f_m(x) = \int f_m(x \mid \theta) p_m(\theta) \mathrm{d}\theta$$ Marginal density

$$\bar{\pi}_m = \pi_m f_m(x) \Big/ \sum_{j=1}^{M} \pi_j f_j(x)$$ Posterior probability that model m is correct

Unless there is a clear correct model, it is better to average over competing models :

$$\bar{p}(\theta \mid x) = \sum_{j=1}^{M} \bar{\pi}_j p_j(\theta \mid x)$$

The competing models do not need to be nested, and each model can be analyzed separately

•Bayesian Mixed Probit (Allenby and Rossi, 99)

$$p(\eta_i, \beta, \theta \mid x) \propto f(x \mid \eta_i, \beta) \pi(\eta_i \mid \theta) \pi(\theta) \pi(\beta)$$

f is likelihood for multinomial probit, η_i

is random effect for observation *i.* θ are the parameters of the distribution of the random effects over the sample. This Bayesian formulation permits inference for the individual effects (repeated SP experiments). Revelt and Train (1999) also give classical methods.

Bayesian Applications

- Scarpa, Thiene, and Train (2008)
- Kim, Kim, and Heo (2003)
- Alvarez-Daziano and Bolduc (2009). Look at a large hybrid choice model of “clean” vehicle choice
- Wang and Kockelman – spatial dynamic ordered probit
- None of these take advantage of Bayesian model selection or averaging

Bayesian Consistency

- Savage (1954) shows that the Bayesian paradigm is the only one consistent with Von-Neumann Morganstern “rationality”
- Modern SP design (Bliemer and Rose) is explicitly Bayesian and relies on prior updating. The only consistent way to update a prior (required for adaptive SP experimental design) is via Bayes rule.

Forecasting

Suppose we want to forecast the change in probabilities due to a change in x. This is given by:

$$\int P\left(\Theta_{\mathrm{i}}, x'_{ijt}\right) - P\left(\Theta_{\mathrm{i}}, x_{ijt}\right) f\left(\Theta_{\mathrm{i}} \mid x_{ijt}, d_{ijt}\right) d\Theta_{\mathrm{i}}$$

Bayesians simulate this using draws from the posterior distribution of Θ_{i} and this explicitly conditions on d_{ijt}. Although some classical analysts look at the conditional distribution of Θ_{i} given d_{ijt} they typically do not condition when doing forecasting.

Why condition on observed choices?

- We interpret choice probabilities as proportion of observationally identical respondents who choose that particular alternative, so this accounts for sampling variability and motivates sample enumeration
- We will improve forecasts by conditioning on all data, including (endogenous) choices

Forecasting Latent Class Models

- Process models, hybrid choice models, attribute processing models all have latent variables in them.
- Forecasting needs to integrate out the unobserved latent variables
- This integral should be taken over the posterior distribution given the observed data, and this may be hard to approximate.

Model Uncertainty

- There are a huge number of modeling choices when specifying a ML model (or flexible GEV model), but inference is always conditioned on the “best” model.
- Bayesians can handle model uncertainty by computing posterior odds and model averaging. This requires computing the marginal likelihood, which is frequently hard.

Inference for WTP

- Daly and Hess (2009) give a general method (similar to delta method) to generate standard errors for functions of choice model parameters.
- They give an example showing that a bootstrap doesn’t work very well. WTP highly skewed in their example.
- Bootstrap percentile works as well as Daly and Hess (and is asymptotically equivalent to Bayesian HPD regions

Multiple discrete choice

- Bhat has developed the MDCEV model that allows for more than one choice.
- MDCEV is derived from constrained utility maximization so it can jointly model choice of activity type and duration. (or choice of vehicles and their utilization).
- Requires the total utilization to remain fixed – may be unrealistic for some policies.

$$U = \sum_{k=1}^{K} \exp(\max_{l \in N_k}\{W_k + Y_{kl} + \eta_{kl}\})(m_k + 1)^{sk} \qquad (16)$$

where W_k depends on household characteristics $\mathbf{x}$ that relates to the choice of class k and equals to $\mathbf{x}'\beta_k$, Y_{kl} depends on vehicle properties of a certain make/model l within class k, $\mathbf{z}_{kl}$, and equals to $\mathbf{z}'_{kl}\gamma$, and m_k denotes miles driven by a vehicle of class k. s_k is considered as a non-satiation

$$H_k = H_1, \quad \text{if} \quad m_k^* > 0$$

$$H_k < H_1, \quad \text{if} \quad m_k^* = 0$$

where

$$H_k = \mathbf{x}'\beta_k + \theta_k \ln \sum_{l \in N_k} \exp(\frac{\mathbf{z}'_{kl}\gamma}{\theta}) + lns_k + (s_k - 1)ln(m_k^* + 1) + \varepsilon_k \qquad (17)$$

The probability that the first Q of the K vehicles being chosen, $P(m_1^*, m_2^*, ..., m_Q^*, 0, ..., 0)$, is then derived from the above Kuhn-Tucker conditions (c.f. Bhat 2005).

The probability function for each household i can be written as

$$P_i(\beta,\tau,\gamma,\theta;\mathbf{m}_k^*,l) \sim \underbrace{\prod_{k=1}^{Q} r_k \sum_{k=1}^{Q} \frac{1}{r_k} \frac{\prod_{k=1}^{Q} e^{V_k}}{(\sum_{h=1}^{K} e^{V_h})^Q}(Q-1)!}_{P(m_1^*,m_2^*,...,m_Q^*,0,...,0)} \quad \prod_{\substack{l\in N_k \\ (m_k^*>0)}} \underbrace{\frac{\exp(\frac{\mathbf{z}_{kl}'\gamma}{\theta})}{\sum_{g\in N_k}\exp(\frac{\mathbf{z}_{kg}'\gamma}{\theta})}}_{P(l|m_k^*>0,l\in N_k)} \qquad (18),\ (19)$$

where

$$r_k = (\frac{1-s_k}{m_k^*+1}) \qquad (20)$$

$$V_k = \mathbf{x}'\beta_k + \theta_k \ln \sum_{l\in N_k} \exp(\frac{\mathbf{z}_{kl}'\gamma}{\theta}) + \ln\alpha_k + (s_k-1)\ln(m_k^*+1) \qquad (21)$$

Multiple discrete choice 2

- Fang and Jeliazkov have Bayesian ordered probit and tobit systems.
- These are reduced form but don't require any constraints on total utilization.
- Jeliazkov's blocking methods result in order of magnitude improvements in computational cost, so they can be fit on very large models

BMOPT Model

$$y_i^* = D_i\alpha + x_i\beta + \varepsilon_i \qquad (1)$$

$$D_i = z_i\gamma + \eta_i \qquad (2)$$

where y_i^* is a 4 by 1 vector of latent dependent variables for number of cars, number of trucks, mileage on cars, and mileage on trucks; D_i is a measure of density for households i at the census tract level, and is endogenous. The relation between the latent dependent variables and their observed values are:

$$y_j = 0, \text{ if } y_j^* \le \alpha_0, \ j = 1,2$$

$$y_j = 1, \text{ if } \alpha_0 < y_j^* \le \alpha_1, \ j = 1,2$$

$$y_j \ge 2, \text{otherwise}, \ j = 1,2$$

$$y_j = y_j^*, \text{ if } y_j^* > 0, \ j = 3,4$$

$$y_j = 0, \text{otherwise}, \ j = 3,4$$

Spatial Dependence

- Difficult due to curse of dimensionality
- Bhat has applied cupolas and composite marginal likelihood (CML) techniques to build some very flexible models.
- Cupolas are a way of building a joint distribution functions with specified marginal distributions. Bayesians use them to build models that are easy to simulate

Composite Maximum Likelihood

- Only estimate spatial dependence model on all possible pairs of spatial autocorrelated responses.
- Allows estimation of very complex spatial dependence models.
- Bhat shows that CML does not lose much efficiency relative to full Maximum Likelihood.

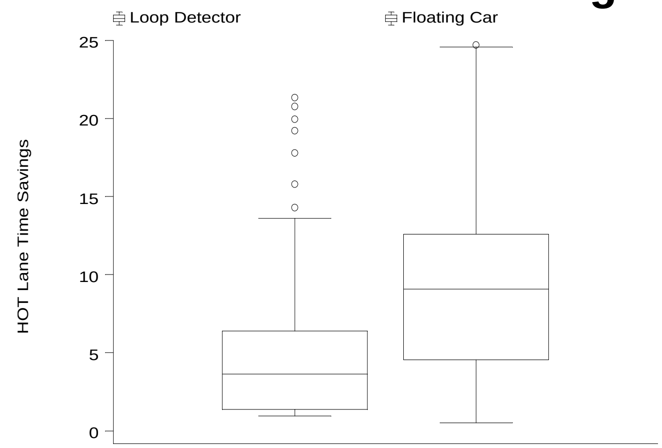
SP/RP

- Yanez, Cherchi, and Ortuzar point out that it is not clear how to forecast demand from SP/RP models with different heterogeneity specifications for SP and RP data.
- This is particularly difficult when some attributes are only identified from SP data.
- Brownstone and Small (2005) show VOT from RP are 2 times VOT from SP!

Measurement Errors

- May dominate parameter estimation error!
- Steimetz and Brownstone (2005) and Washington, Ravulaparthy et. al. (2009) are the only application of Multiple Imputations to travel demand (imputing attributes of non-chosen alternatives in RP surveys)
- Multiple Imputation is more widely used in other fields, and is available in recent versions of STATA and SAS

I15 HOT Lane Time Savings



Multiple Imputations

- Treat floating car data as missing for 7/8 of sample, and use data where we observe floating and loop detector data to impute missing data.
- Much easier than joint modeling of model choice and measurement error
- Asymptotically equivalent to full Bayesian approach:

$$h(\theta|Y_{obs}) = \int g(\theta|Y_{obs}, Y_{mis}) f(Y_{mis}|Y_{obs}) dY_{mis}$$

- 1.Draw predicted values from predictive density of missing observations (assumed to be linear model after logit transformation). Observed floating car data held fixed (must condition on all observables). Must match first 2 moments of missing data!

$$y^*_{mis} = X_{mis}\beta^* + \varepsilon^*$$

2. Transform predicted values to form used in choice model.
3. Estimate choice model using transformed predicted values.
4. Repeat steps 1 – 3 above many times.

- 5. Combine estimation results from 3 using:

$$\hat{\theta} = \sum_{j=1}^{m} \tilde{\theta}_j / m$$

$$U = \sum_{j=1}^{m} \tilde{\Omega}_j / m$$

$$B = \sum_{j=1}^{m} \left(\tilde{\theta}_j - \hat{\theta}\right)\left(\tilde{\theta}_j - \hat{\theta}\right)' / (m-1)$$

$$\hat{\Sigma} = U + \left(1 + m^{-1}\right)B$$

Hypothesis Testing

$$\left(\theta - \theta^0\right)' \hat{\Sigma}^{-1} \left(\theta - \theta^0\right)$$

Is $F_{K,\nu}$, $\nu = (m - 1)(1 + r_m^{-1})^2$

$r_m = (1 + m^{-1})$ Trace$(BU^{-1})/K$.

Choose m to make ν large

Conclusion

- Focus on forecasting and inference instead of getting large t-statistics
- We have too many poorly identified parameters. This is independent of whether model is “closed form” (GEV) or simulated.
- Inference from MSL estimators is still problematic
- Real forecast uncertainty is probably dominated by model uncertainty and measurement error!

References

Alvarez-Daziano, R., D. Bolduc (2009). Canadian consumers’ perceptual and attitudinal responses toward green automobile technologies: an application of Hybrid Choice Models. Working paper accessed at http://www.feem-web.it/ess/files/alvarez-daziano.pdf on December 14, 2009.

Bhat, C.R. (2005) A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research-B*, 39(8) 679-707.

Bhat, C.R. (2008) The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. *Transportation Research-B*, 42(3), 274-303.

Bhat, C.R., C. Varin, and N. Ferdous (2009). A Comparison of the Maximum Simulated Likelihood and Composite Marginal Likelihood Estimation Approaches in the Context of the Multivariate Ordered Response Model System. U.T. Austin, Dept.of Civil Engineering working paper.

Brownstone, D. (2001). Discrete Choice Modeling for Transportation. in D. Hensher (ed.), Travel Behaviour Research: The Leading Edge. Amsterdam: Pergamon, pp. 97-124.

Brownstone, D. and K. Small (2005). Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. Transportation Research A, 39, 279-29.

Brownstone, D. and H. A. Fang (2009). A Vehicle Ownership and Utilization Choice Model With Endogenous Residential Density. Paper presented at TRB annual meetings.

Chiou, L. and J.L. Walker (2007). Masking identification of discrete choice models under simulation methods, *Journal of Econometrics* 141, 683–703.

Greene, W.H. and D.A. Hensher (2009). Does Scale Heterogeneity across Individuals Matter? An Empirical Assessment of Alternative Logit Models. Forthcoming in *Transportation*.

Hensher, D.A. and W.H. Greene (2003), The Mixed Logit model: The state of practice, *Transportation* 30: 133–176.

Hess, S. and A. Daly (2009). Calculating Errors for Measures Derived from Choice Modeling Estimates. Paper presented at TRB meetings Washington, DC, session 09-1656.

Hess, S. and D.A. Hensher (2008). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research B, forthcoming.

Jeliazkov, I., J. Graves, and M. Kutzbach (2008) Fitting and comparison of models for multivariate ordinal outcomes. *Advances in Econometrics*, 23, 115-156.

Kim, Y., Tai-Yoo Kim & Eunnyeong Heo (2003). Bayesian estimation of multinomial probit models of work trip choice. *Transportation* 30: 351–365.

Krinsky, I., and Robb (1986) A.L. On approximating the statistical properties of elasticities, *Review of Economics and Statistics*, 68, 715-719.

McFadden, D., and K. Train (2000) Mixed MNL models for discrete response. Journal *Of Applied Econometrics*, 15(5), 447-470.

Nobile, A, Bhat, C, Pas, E (1997) A Random Effects Multinomial Probit Model of Car Ownership Choice. Case Studies in Bayesian Statistics 3, eds. Gatsonis, C., Hodges, J, Kass, R., McCulloch, R., Rossi, P., Singpurwalla, N., Springer: New York, 419 – 434.

Pinjari, A.R., and C.R. Bhat (2009) A Multiple Discrete-Continuous Nested Extreme Value Model: Formulation and Application to Non-Worker Activity Time-use and Timing Behavior on Weekdays. *Transportation Research-B* (accepted).

Revelt, D., and K. Train. 1998. "Mixed Logit with Repeated Choices." *Review of Economics and Statistics* 80(4):647–57

Scarpa,R, Mara Thiene, And Kenneth Train (2008). Utility In Willingness To Pay Space: A Tool To Address Confounding Random Scale Effects In Destination Choice To The Alps. Amer. J. Agr. Econ. 90(4): 994–1010.

Steimetz, S., and D. Brownstone (2005). Estimating Commuters' "Value of Time" with Noisy Data: a Multiple Imputation Approach, Transportation Research B, 39, 865-889.

Train, K. (2003) Discrete choice methods with simulation. 1st edition, Cambridge University Press. (Note that there is a 2nd edition available published in 2009)

Wang, X. and K. Kockelman (2008). Application Of The Dynamic Spatial Ordered Probit Model: Patterns Of Land Development Change In Austin, Texas. Working paper.

Wang, X. and K. Kockelman (2008). The Dynamic Spatial Ordered Probit Model: Methods For Capturing Patterns Of Spatial And Temporal Autocorrelation In Ordered Response Data, Using Bayesian Estimation. Working paper.