A Bayes multinomial probit model for random consumer-surplus maximization

Esther Chiew and Ricardo A. Daziano*

September 2015

Abstract

Willingness to pay measures derived from discrete choice model not only have a clear economic interpretation, but also are a relevant input for welfare analysis, developing marketing strategies, and policy-making. Because inference on parameter ratios is associated with statistical issues, transforming the original utility maximization problem into a consumer surplus maximization model that provides direct inference on willingness to pay is thus desirable. Recent literature uses a parameter reparameterization from the original preference space to the desired willingness-to-pay space. However, we propose a slight variation to this reparameterization that is based on normalizing the marginal utility of income and that works particularly well for models with a general covariance matrix. (For logit-based models the normalization is equivalent to working with the standard reparameterization in willingness-to-pay space.) In fact, we show that Bayes implementation of the normalization of the marginal utility of income solves well-known problems of current multinomial probit samplers. The estimator that we propose effectively avoids defining proper priors on unidentified parameters as well as identification of priors for and making draws of a constrained covariance matrix, and improves the behavior of the predictive posterior of the choice probabilities. In addition, willingness-to-pay estimates from previous studies can easily be used as proper priors in the proposed Gibbs sampler.

JEL classification: C25, C51, D12.

Keywords: Discrete choice models; willingness-to-pay space; consumer surplus models

*School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853; Email: daziano@cornell.edu

1 A normalization approach to WTP space

The standard normalization of discrete choice models is to directly fix scale of the nuisance parameters. However, an alternative normalization is to fix the value of a structural parameter (i.e. a marginal utility). In fact, normalizing a coefficient is the standard procedure for fixing the scale in semiparametric estimators of binary outcome regressions (Ichimura, 1993; Klein and Spady, 1993).

Consider the following factorization of utility $U_{ij} = \mathbf{q}'_{ij}\boldsymbol{\beta} - \lambda p_{ij} + \varepsilon_{ij} = \lambda(\mathbf{q}'_{ij}\boldsymbol{\theta} - p_{ij}) + \varepsilon_{ij} \Leftrightarrow \text{CS}_{ij} = \frac{U_{ij}}{\lambda} = \mathbf{q}'_{ij}\boldsymbol{\theta} - p_{ij} + \frac{\varepsilon_{ij}}{\lambda}$, where i is the agent, j is the alternative, p_{ij} is price, q_{ij} are attributes, λ is the marginal utility of income, $\boldsymbol{\beta}$ is a vector of marginal utilities, $\boldsymbol{\theta}$ is a vector of willingness-to-pay (WTP) parameters, ε_{ij} is the error term, and CS_{ij} measures consumer surplus (cf. Jedidi et al., 2003). Thus, working with the consumer-surplus maximization model $\text{CS}_{ij} = \mathbf{q}'_{ij}\boldsymbol{\theta} - p_{ij} + \varepsilon_{ij}$ is equivalent to considering $\lambda = 1$. Working in preference ($\boldsymbol{\beta}$) or WTP ($\boldsymbol{\theta}$) space leads to estimators that are equivalent. This equivalence is found working either with the reparameterization of Train and Weeks (2005) ($U_{ij} = \lambda(\mathbf{q}'_{ij}\boldsymbol{\theta} - p_{ij}) + \varepsilon_{ij}$) under a standard normalization of scale or with the consumer-surplus maximization model. Under the normalization of the marginal utility of income, scale of $\boldsymbol{\epsilon}_i$ is identified (up to the normalized λ) and can be thus estimated. Since p_{ij} is a continuous variable, the normalization $\lambda = 1$ ensures identification of the model (see Klein and Spady, 1993).

In homogeneous-of-degree- μ^{-1} GEV models, where μ is the scale parameter, the general expression of the choice probability depends on $\mu^{-1}\lambda(\mathbf{q}'_{ij}\boldsymbol{\theta} - p_{ij})$. Thus, in GEV models normalization of λ or μ produces no difference in the implementation of the maximum likelihood estimator. This extends to the case with random parameters and a GEV kernel (mixed logit models). This identity does not extend to general non-IID models, such as multinomial probit models. The standard scale normalization for probit is to fix one nuisance parameter: usually the trace or the first element of the covariance of the model in differences. However, as discussed in the next section, all covariance elements can be identified when $\lambda = 1$. In fact, there are clear benefits in normalizing λ for the Bayes probit estimator, especially for the choice of appropriate priors, that are proper, have standard distributions, and are truly informative.

2 Random consumer-surplus maximization probit

Consider the random consumer-surplus probit $\mathbf{CS}_i = \mathbf{Q}_i \boldsymbol{\theta} - \mathbf{p}_i + \boldsymbol{\epsilon}_i$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_J)$, where $\mathbf{Q}_i$ is the attribute matrix with rows $\mathbf{q}'_{ij}$, $\mathbf{p}_i$ is the vector of prices, and $\boldsymbol{\epsilon}_i$ is a vector of multivariate normally distributed error terms. As with random utility, choice in a consumer-surplus model is determined by the discrete maximization process $y_i = j$ iff $\mathbf{CS}_{ij} > \mathbf{CS}_{ij'} \forall j' \neq j$. Discrete maximization means that location needs to be normalized. Therefore we work with the model in differences that focuses on consumer-surplus variations, independently from initial endowments. Let Δ_j be a matrix difference operator such that $\Delta_j \mathbf{CS}_i$ takes each element $\mathbf{CS}_{ij'}$ in $\mathbf{CS}_i$ ($j' \neq j$) and subtracts the base element $\mathbf{CS}_{ij}$. The model in differences is $\Delta_j \mathbf{CS}_i = \Delta_j \mathbf{Q}_i \boldsymbol{\theta} - \Delta_j \mathbf{p}_i + \Delta_j \boldsymbol{\epsilon}_i$, $\Delta_j \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}_{(J-1 \times 1)}, \Delta_j \boldsymbol{\Sigma}_J \Delta_j')$, where all parameters are identified, including all elements in the reduced-dimension covariance matrix $\boldsymbol{\Sigma} = \Delta_j \boldsymbol{\Sigma}_J \Delta_j'$. For the model in differences the decision rule becomes $y_i = j$ iff $\Delta_{jj'} \mathbf{CS}_{ij'} < 0$ and $y_i = j'$ iff $\Delta_{jj'} \mathbf{CS}_{ij'} \geq \max\{0, \Delta \mathbf{CS}_{i,-j}\}$, $\forall j' \neq j$, where $\Delta_{jj'}$ is row j' of the matrix Δ_j and $\Delta \mathbf{CS}_{i,-j}$ represents the set of all the elements in differences of $\mathbf{CS}_i$ with respect to and with the exception of element j .

2.1 A Bayes regression with a partially observed LDV

The Bayes estimator of a multinomial probit model in preference space is very intuitive. If utility were observable, then we could estimate the parameters using an ordinary regression. Because utility is a latent variable, we need to use the discrete choice modeling framework. From a Bayesian perspective, latent variables can be treated as additional parameters using data augmentation techniques. Following this premise and ideas by Albert and Chib (1993), McCulloch and Rossi (1994) derived a multinomial probit Gibbs sampler. Because in a probit model the utility function is normally distributed (subject to the feasibility region imposed by the choice indicators), a truncated normal distribution is used to sample realizations of utility. These utility draws are then used as observations of the dependent variable for estimating the model parameters using a Bayes ordinary regression. However, there are issues in the estimator of McCulloch and Rossi (1994), including the fact that the chains are built over unidentified parameters.

The Bayes estimator of a multinomial model in WTP space with $\lambda = 1$ follows the same principle. However, there are a couple of modifications that are needed and that solve the issues of the estimator of McCulloch and Rossi (1994) (see section 2.2). First, the model can be rewritten as $\Delta_j \mathbf{CS}_i + \Delta_j \mathbf{p}_i = \Delta_j \mathbf{Q}_i \boldsymbol{\theta} + \Delta_j \boldsymbol{\epsilon}_i$. Let $\mathbf{L}$ be the Cholesky root of $(\Delta_j \boldsymbol{\Sigma}_j \Delta_j')^{-1}$, then:

$$\mathbf{L}' \Delta_j \mathbf{CS}_i + \mathbf{L}' \Delta_j \mathbf{p}_i = \mathbf{L}' \Delta_j \mathbf{Q}_i \boldsymbol{\theta} + \mathbf{L}' \Delta_j \boldsymbol{\epsilon}_i, \quad \mathbf{L}' \Delta_j \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}_{(J-1 \times 1)}, \mathbf{I}_{(J-1 \times J-1)}). \quad (1)$$

Note that the limited dependent variable (LDV) is in this case partially observed ($\mathbf{p}_{ij}$ is known). Data augmentation is still needed to sample observations of $\Delta_j \mathbf{CS}_i$, which is a truncated normally distributed vector. Second, and unlike the model in preference space, no identification restrictions are required for $\boldsymbol{\Sigma}$, as the model already is specified over identified parameters. The steps for the Gibbs sampler of the consumer-surplus model are detailed below.

- Start at any given point $\mathbf{CS}_i^{(0)}$, $\boldsymbol{\theta}^{(0)}$, and $\boldsymbol{\Sigma}^{(0)} = (\mathbf{L}^{(0)} \mathbf{L}^{(0)')^{-1}}$ in the parameter space.
- For $g \in \{1, \dots, G\}$

1. If $y_i = j$, draw $\Delta_j \mathbf{CS}_i^{(g)}$ from the truncated normal distribution

$$\mathcal{N}(\Delta_j \mathbf{Q}_i \boldsymbol{\theta}^{(g-1)} - \Delta_j \mathbf{p}_{ij}, \Delta_j \boldsymbol{\Sigma}^{(g-1)} \Delta_j') 1(\Delta_{jj'} \mathbf{CS}_{ij'} < 0, \forall j' \neq j)$$

otherwise draw $\Delta_j \mathbf{CS}_i$ from the truncated normal distribution

$$\mathcal{N}(\Delta_j \mathbf{Q}_i \boldsymbol{\theta}^{(g-1)} - \Delta_j \mathbf{p}_{ij}, \Delta_j \boldsymbol{\Sigma}^{(g-1)} \Delta_j') 1(\Delta_{jj'} \mathbf{CS}_{ij'} > \max\{0, \Delta \mathbf{CS}_{i,-j}\}, \forall j' \neq j)$$

2. Draw $\boldsymbol{\theta}^{(g)}$ from the normal distribution

$$\mathcal{N} \left((\check{V}_{\boldsymbol{\theta}}^{-1} \check{\boldsymbol{\theta}} + (\mathbf{L}^{(g-1)'} \Delta_j \mathbf{Q})' \mathbf{L}^{(g-1)'} \Delta_j \mathbf{Q})^{-1} (\check{V}_{\boldsymbol{\theta}}^{-1} + \mathbf{Q}' \Delta_j' \mathbf{L}^{(g-1)} \mathbf{L}^{(g-1)'} \Delta_j (\mathbf{CS}^{(g)} + \mathbf{p})) , \right. \\ \left. (\check{V}_{\boldsymbol{\theta}}^{-1} + (\mathbf{L}^{(g-1)'} \Delta_j \mathbf{Q})' (\mathbf{L}^{(g-1)'} \Delta_j \mathbf{Q}))^{-1} \right),$$

where $\check{\boldsymbol{\theta}}$ and $\check{V}_{\boldsymbol{\theta}}$ are the parameters of the prior distribution of the willingness-

to-pay $p(\boldsymbol{\theta}) \sim \mathcal{N}(\check{\boldsymbol{\theta}}, \check{V}_{\boldsymbol{\theta}})$.

3. Draw $\boldsymbol{\Sigma}^{(g)}$ from the inverted-Wishart distribution

$$IW\left(\check{\nu} + N, \check{\boldsymbol{\Sigma}} + \sum_{i=1}^N \Delta_j(\mathbf{CS}_i^{(g)} + \mathbf{p} - \mathbf{Q}_i\boldsymbol{\theta}^{(g)})(\mathbf{CS}_i^{(g)} + \mathbf{p} - \mathbf{Q}_i\boldsymbol{\theta}^{(g)})'\Delta_j'\right),$$

where the $\check{\nu}$ and $\Delta_j\check{\boldsymbol{\Sigma}}\Delta_j'$ are the parameters of the inverted-Wishart prior $p(\Delta_j\boldsymbol{\Sigma}\Delta_j') = IW(\check{\nu}, \Delta_j\check{\boldsymbol{\Sigma}}\Delta_j')$. Update the Cholesky root $\mathbf{L}^{(g)}$ of $\boldsymbol{\Sigma}^{-1}$.

- Because the long-run distribution of the $\boldsymbol{\theta}^{(G)}$ is the true posterior of interest, each sample $\boldsymbol{\theta}^{(g)}$ can be treated as a random draw of the joint posterior.

2.2 Benefits of the normalization of the marginal utility of income in the Bayesian context

The standard probit estimators use normalization of scale for identification. The traditional normalization is $\sigma_{11} = 1$, which can be seen as a constraint imposed on $\boldsymbol{\Sigma}$ (cf. Burgette and Nordheim, 2012). In the Bayesian context, a constraint in the covariance matrix makes it difficult to define appropriate priors.¹ The scale normalization $\sigma_{11} = 1$ means either working with the Gibbs sampler of an unidentified model, where identification is achieved after estimation is completed (McCulloch and Rossi, 1994),² or making adjustments in the inverted-Wishart distribution for $\boldsymbol{\Sigma}$ that incorporate the scale constraint (McCulloch et al., 2000; Nobile, 2000). Postprocessing the results for identification implies defining proper priors on unidentified parameters, and the induced priors for the parameters that are identified are nonstandard (McCulloch et al., 2000; Nobile, 2000; Imai and van Dyk, 2005). Working with a constrained covariance matrix may result in poor convergence behavior of the estimator, as well as the generation of extreme posterior predictions for the choice probabilities (Burgette and Nordheim, 2012). The Gibbs sampler for the consumer-surplus

¹In fact, covariance constraints have an influence on posterior predictions (Burgette and Nordheim, 2012).

²Identified parameters are found in this case by dividing all elements in the chain by the posterior mean of $\sqrt{\sigma_{11}}$.

maximization model introduced in this paper solves these issues as priors are defined on identified parameters only, the prior distributions are standard, and the standard inverted Wishart is used directly for generating the desired covariance matrix as no restrictions are necessary. In fact, using diffuse priors or even improper priors is possible in the proposed estimator, whereas flat priors create problems in the standard sampler. Standard Bayes probit estimators can require more than 50,000 draws to achieve convergence for a standard choice model. Preliminary simulation tests show that our estimator can be as efficient as that of Imai and van Dyk (2005), converging in fewer than 2,500 draws.

Beyond convergence and because marginal utilities are affected by scale, it is unclear how to define economically meaningful prior distributions for these preference parameters that are truly informative. All the previous Bayes probit estimators exhibit the problem of defining priors for parameters that are affected by scale (including the statistically appropriate priors of Imai and van Dyk (2005)). The proposed consumer-surplus model does not face this issue. In fact, because WTP is not affected by scale, WTP estimates from previous studies can easily be used as priors. Furthermore, truly informative WTP-priors are key as they have an effect on predictive posteriors even for large samples. Note that, unlike the estimator we introduced in this paper, Bayesian implementation of the reparameterization of Train and Weeks (2005) with the standard normalization of scale still would require defining priors for λ (which is affected to scale) as well as restrictions on the covariance matrix.

3 Concluding remarks

In this paper we have proposed an alternative method to transform the parameters to WTP space based on the normalization of scale that fixes the marginal utility of income (λ). Although the normalization of a structural parameter is the standard assumption in semiparametric estimators of binary outcome regressions, this procedure has not been exploited for WTP inference. Even more, there are no practical applications of the normalization of a structural parameter to multinomial problems either.

When setting $\lambda = 1$ (instead of normalizing scale by fixing a covariance element),

the parameters of the model directly represent WTP rather than unitless marginal utilities. In fact, the whole choice model can be reinterpreted as a consumer-surplus maximization problem. Furthermore, in a model where $\lambda = 1$ there is no need for constraining any element of the covariance matrix, and Bayes estimators can be derived over the identified parameter space in a straightforward manner with effective gains in convergence. In fact, the Gibbs sampling scheme that we have derived easily solves two well known problems of Bayes probit estimators. First, we avoid defining priors on unidentified parameters. Second, we do not need to identify priors nor make draws of a constrained covariance matrix. In addition, the prior distributions for the structural WTP parameters are economically meaningful and estimates are easily transferrable across applications. Easily transferrable prior beliefs is a desirable feature for Bayesian efficient design of discrete choice experiments.

Finally, since the proposed method recasts the dependent variable to a direct measure of consumer-surplus – a result that is not obtained for the nonlinear WTP-space reparameterization with scale normalized, estimators based on the normalization $\lambda = 1$ potentially can transform the way welfare improvements are evaluated in choice modeling. Analysis of the predictive posterior of consumer-surplus, as well as a complete Monte Carlo study to test the empirical behavior of the theoretical expectations discussed in this paper, are both left as avenues of future research.

Acknowledgements

This research is based upon work supported by the National Science Foundation Faculty Early Career Development CAREER Award No. CBET-1253475.

References

- Albert J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88:669-679.
- Burgette, L.F. and E.V. Nordheim (2012). The trace restriction: an alternative identification strategy for the Bayesian multinomial probit model. *Journal of Business and Economic Statistics* 30(3), 404-410.
- Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models. *Journal of Econometrics* 58, 71-120.

- Imai, K. and D.A. van Dyk (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics* 124(2), 311-334.
- Jedidi, K., S. Jagpal and P. Manchanda (2003). Measuring heterogeneous reservation prices for product bundles. *Marketing Science*, 22, 107-130.
- Klein, R. W. and R. H. Spady (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica* 61(2), 387-421.
- McCulloch R., Rossi P. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* 64, 207-240.
- McCulloch R.R., Polson N.G., Rossi, P.E. (2000). Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99, 173-193.
- Nobile, A. (2000). Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics*, 99, 335-345.
- Train, K. and M. Weeks (2005) Discrete choice models in preference space and willingness to pay space, in R. Scarpa and A. Alberini, eds., *Applications of Simulation Methods in Environmental and Resource Economics*, Springer Publisher, Dordrecht, Chapter 1, 1-16.