Incorporating Random Coefficients and Alternative Specific Constants into Discrete Choice Models: Implications for In-Sample Fit and Welfare Estimates *

H. Allen Klaiber
Department of Agricultural and Resource Economics
North Carolina State University

Roger H. von Haefen
Department of Agricultural and Resource Economics
North Carolina State University

This Draft: March 2008

Abstract

In recent years, several innovative econometric methods have been employed in non-market valuation applications of discrete choice models. Two particularly attractive methods are random parameters (which introduce more plausible substitution patterns) and alternative specific constants (which control for unobserved attributes). In this paper, we investigate the properties of these methods along several dimensions. Across three recreation data sets, we consistently find large improvements in model fit arising from the inclusion of both methods; however, these gains often come concomitant with significant degradations in-sample trip predictions. We then show how poor in-sample predictions correlate with welfare estimates. Using econometric theory and Monte Carlo evidence, we illuminate why these perverse findings arise. Finally, we propose and empirically evaluate four 'second-best' modeling strategies that attempt to correct for the poor in-sample predictions we find in our applications.

Keywords: Discrete Choice, Recreation Demand, Revealed Preference, Stated Preference, Welfare Estimation, Alternative Specific Constants, Random Parameters

---

*Early draft that is currently under major revision. Comments welcome, but please do not cite. We thank George Parsons, Bill Breffle, Dan Phaneuf, Chris Timmins, and W2133 Annual Meeting participants for helpful comments.

Section I. Introduction

Discrete choice models have become one of the most frequently used modeling frameworks for recreation demand and locational equilibrium models (Murdock, 2006; Bayer and Timmins, 2007). Within the framework, two econometric innovations that applied researchers are using with increasing regularity are random coefficients (McFadden and Train, 2000) and the inclusion of alternative specific constants (Berry, 1994). Random coefficients are an attractive mechanism for relaxing the restrictive implications of the independence of irrelevant alternatives (IIA), thus introducing more plausible substitution patterns. Including a full set of alternative specific constants allows the analyst to control for unobserved attributes that may be correlated with observed attributes.

In applications of these modeling innovations to discrete choice models, researchers have found that they generate substantial and statistically significant improvements in fit (von Haefen and Phaneuf, 2008; Murdock, 2006). In an empirical investigation of three recreation data sets, we also find large gains in model fit. However, we also find that the models with alternative specific constants and random coefficients often fail to replicate the in-sample aggregate visitation patterns implied by the data. This empirical regularity generates important implications for the credibility of welfare analysis – why should one believe welfare measures derived from models that cannot replicate in-sample aggregate choice behavior?

Our goal in this paper is to shed light on the counterintuitive empirical regularity of improved statistical fit combined with poor in-sample prediction. We begin by documenting this phenomenon with three recreation data sets that have been used in previously published research. Two of the three applications combine revealed and stated preference (RP-SP) to identify all demand parameters (Adamowicz et al., 1997; Haener et al., 2001) as previously done by von Haefen and Phaneuf (2008). The other exploits only revealed preference (RP) data (Parsons et al., 1999) and uses a variation of the two-step estimator proposed by Berry, Levinsohn, and Pakes (2004) and used recently by Murdock (2006) in the recreation context. With all three data sets, we find the introduction of random coefficients and alternative specific constants (ASCs hereafter) substantially and significantly improves statistical fit as measured by the log-likelihood. We also find that in-sample trip predictions often (but not uniformly) deteriorate with these richer empirical specifications, and we document how these poor predictions correlate with welfare estimates for a range of policy scenarios.

We then explore why the poor predictions arise in practice. Here we use theoretical results from Gourieroux, Monfort, and Trognon (1984) about the properties of the linear exponential family of distributions as well as some Monte Carlo findings. The upshot of our discussion is that: 1) fixed coefficient logit models with a full set of ASCs will generate in-sample trip predictions for each alternative that *perfectly* match the data, and 2) random coefficient logit models with or without ASCs may not predict perfectly in-sample, but should generate reasonably close predictions if the analyst has correctly specified the underlying data generating process. An implication of this finding is that the poor in-sample predictions that we find in our three applications arise because of model misspecification. Thus, logit models with random coefficients and ASCs fit the data better than models without these econometric innovations, but they nevertheless fail to account for important features of the data.

We conclude by exploring a number of 'second best' strategies for dealing with poor in-sample predictions. These range from: 1) abandoning random coefficient specifications and using fixed coefficient models with ASCs that generate perfect in-sample predictions; 4) using less-efficient non-panel random coefficient models that, as we demonstrate, generate more plausible in-sample predictions; 3) using the Berry (1994) contraction mapping or maximum penalized likelihood (Montricher et al., 1975, Silverman, 1982; Huh and Sickles, 1994; Shonkwiler and Englin, 2005) with ASCs to force the in-sample predictions to match the data perfectly; and 4) conditioning on observed choice in the construction of welfare measures following von Haefen (2003). Our preliminary results suggest that each of these strategies is effective in terms of generating plausible in-sample predictions but they differ considerably in terms of their implications for statistical fit.

The paper proceeds as follows. The next section documents the performance of fixed and random coefficient logit models with and without a full set of ASCs with three recreation data sets. Section III explores the factors that give rise to the perverse empirical findings reported in the previous section using econometric theory and a set of Monte Carlo simulations. Section IV investigates a number of 'second best' empirical strategies that applied researchers may find attractive in future applications. We then conclude with some final observations and recommendations.

Section II. Nature of the Problem

We begin by illustrating the poor in-sample prediction problem that serves as the motivation for this research. To demonstrate that this problem is not an idiosyncratic feature associated with a single data set, we consider three recreation data sets that researchers have used in previously published studies. The first data set was first used by Adamowicz et al. (1997) and consists of both revealed preference (RP) and stated preference (SP) choice data for moose hunting in the

Canadian province of Alberta. The RP data consists of seasonal moose hunting trips for 271 individuals to 14 wildlife management units (WMUs) throughout Alberta in 1993. The SP data consists of 16 choice experiments that were generated with a blocked orthogonal, main effects design. All eleven site attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. The second data set was first used by Haener et al. (2001) and also consists of combined RP-SP data for Canadian moose hunting. This data source, however, was collected in the neighboring province of Saskatchewan in 1994. The RP data consists of seasonal moose hunting trips for 532 individuals to 11 wildlife management zones (WMZs) throughout Saskatchewan. The SP data consists of 16 choice experiments that were generated with a blocked orthogonal, main effects design. All nine attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. As discussed in von Haefen and Phaneuf (2007), the fusion of RP and SP data is attractive in both data environments because the inclusion of a full set of ASCs confounds identification of the site attribute parameters given the relatively small number of sites in each application. For both data sets, we control for differences in scale across RP and SP data sources and use empirical specifications, estimation strategies, and welfare scenarios that match those used by von Haefen and Phaneuf (2008).

The third data set we consider looks at Mid-Atlantic beach visitation and was first used by Parsons et al. (1999). This data set consists of seasonal trip data to 62 ocean beaches in 1997 for 375 individuals. For each beach, we observe 14 site characteristic variables plus we construct individual-specific travel costs based on each recreator's home zip code. Because we use only RP data with this application, we use a two-step estimation strategy for those models that includes a full set of ASCs (Berry, Levinsohn, and Pakes, 2004; Murdock, 2006). For the results reported in Table 1, our two-step estimator differs from previous two-step estimators in the following way. Similar to Murdock, we use maximum likelihood techniques in the first step to estimate the travel cost parameter and a full set of ASCs that subsume all 14 site characteristics that do not vary over individuals (note: we do not include any demographic interactions in this model because preliminary testing suggested that they did not improve model fit). In contrast to Murdock, our first step estimator does not employ the Berry (1994) contraction mapping algorithm, an issue we return to in a later section. Thus, our first step estimator relies entirely on traditional maximum likelihood techniques, not the combination of maximum likelihood and Berry contraction mapping techniques that Murdock employs.[i] Our second-stage estimator is identical to Murdock's approach in that we regress the estimated ASCs from the first stage on the 14 site characteristics and a constant term. Importantly, this approach assumes that the unobserved site attributes are uncorrelated with observed site attributes.

Table I summarizes our findings.[ii] All random coefficient models assume that the main effects for the site attributes (excluding travel cost) are normally distributed with no correlations. In on-going work, we are exploring truncated normal and latent class mixing distributions. Arrayed across columns 2-5 are results from four alternative specifications that differ in terms of the inclusion/exclusion of ASCs and random coefficients. In particular, column 2 contains results from models with neither ASCs nor random coefficients, column 3's results contain ASCs but no random coefficients, column 4's results contain random coefficients but no ASCs, and column 5's results contain both. Note that all random parameter specifications assume that all main effects for the various site attributes vary randomly across the population but are common for a given individual, so we refer to these specifications as 'panel' random coefficient specifications following Train (1998). Beginning first with the Alberta results, we note that relative to our baseline model without ASCs and random coefficients, the addition of these modeling innovations generates substantial improvements in fit. The largest gains seem to come from the addition of random coefficients that introduce correlations across an individual's multiple trips, although likelihood ratio tests suggest that ASCs also improve model fit significantly (p value < 0.0001).

To ascertain how well these models predict aggregate trip taking behavior for each site, we construct the following summary statistic for each model:

(1) Percentage absolute prediction error = $$100 \times \sum_{i=1}^{J} s_i^S \frac{\text{abs}(s_i^S - s_i^M)}{s_i^S} = 100 \times \sum_{i=1}^{J} \text{abs}(s_i^S - s_i^M),$$

where s_i^S and s_i^M are the in-sample share of trips to site *i* and the model's prediction of the share of trips to site *i*, respectively, and *J* is the number of sites. The prediction error statistic can be interpreted as the share weighted in-sample prediction error for each site and thus can be used to rank order the models in terms of in-sample predictions that match the observed data. Intuitively, a model that can replicate aggregate trip predictions well for each site would generate a low prediction error value, whereas a model with poor in-sample aggregate predictions for each site would score a relatively high value. For the Alberta data, we see that the fixed coefficient specification with ASCs has the lowest prediction error statistic (effectively zero), whereas the random coefficient without ASCs has the highest. Interestingly, the substantially better fitting random coefficient with ASCs model has a prediction error statistic that is similar in magnitude to the more parsimonious fixed coefficient without ASCs specification.

Finally, it is interesting to see how these differences in fit and prediction play out in terms of welfare estimates. We consider two scenarios – a reduction in moose population at WMU #348 and an increase in moose population at WMU #344 – and calculate the partial equilibrium (i.e., ignoring changes in congestion) compensating surplus for both scenarios using the approach first suggested by Train (1998). In addition to point estimates and standard errors for the welfare measures, we also report the percentage in-sample prediction error for those sites directly affected by the different policies. Overprediction of the share of trips to these sites is likely to translate into larger welfare estimates, although variability in parameter estimates and the structure of substitution implied by the different models will also play a significant role. For the moose population reduction scenario, we find a range of point estimates from -$9.47 to -$25.00 with significant variation in these estimates' precision. The fixed coefficient with ASCs specification generates in-sample predictions for trips to WMU #348 that match the data well, whereas the other specifications overpredict trips to WMU #348 and generate larger (in absolute value) welfare estimates. For the moose population increase scenario, we find even larger variation in point estimates ($3.61 to $98.34) with significant variation in precision once again. In general, the smaller estimates correspond to specifications that underpredict the share of trips to WMU #344. Based on these results, we conclude that poor in-sample predictions play a significant role in explaining the variation of welfare point estimates in the Alberta data.

Similar results arise with Saskatchewan moose hunting data and the Mid-Atlantic beach data. With both data sets, adding ASCs and especially panel random coefficients improves statistical fit as measured by the log-likelihoods, but this improvement in fit does not necessarily generate lower prediction errors. The percentage absolute prediction errors for the fixed coefficient with ASCs models is once again near zero, but the percentage absolute prediction errors for the panel random coefficient models (with and without ASCs) are uniformly larger than the fixed coefficient models. For the Saskatchewan data, welfare point estimates and their precision vary significantly across the competing models. The variation in point estimates across the competing models seems to be correlated with the degree to which the models over- or underpredict trips to the affected sites. Finally, there appears to be considerably less variation in welfare point estimates for the Mid-Atlantic data, which may be explained by the fact that the alternative models seem to predict in-sample far better for the Mid-Atlantic data than the Alberta or Saskatchewan data.

In summary, the results in Table 1 suggest a somewhat counterintuitive result – including ASCs and especially random coefficients significantly improve overall statistical fit but do not generate in-sample trip predictions that match the observed data well. Welfare measures seem to be correlated with the degree of over- or underprediction implied by the different specifications, but other factors – parameter estimates, the structure of substitution implied by the models – certainly play a significant role. Overall, the results in Table 1 provide mixed evidence in favor of incorporating random coefficients and ASCs into discrete choice models, and cast doubt on the credibility of welfare estimates from models that predict in-sample poorly.

Section III. What explains these counterintuitive results?

In this section we use econometric theory and results from a Monte Carlo analysis to shed light on the counterintuitive results presented in the previous section. To motivate our main insight here, consider the log-likelihood function for a sample of *N* individuals each making separate choices from *J* alternatives:

$$\text{(2)} \qquad \ln L(\beta) = \sum_{i=1}^{N}\sum_{j=1}^{J} 1_{ij} X_{ij}\beta - \ln\left(\sum_{k=1}^{J}\exp(X_{ik}\beta)\right),$$

where 1_{ij} is an indicator function equal to 1 for individual *i's* chosen alternative and zero otherwise. The score condition associated with this log-likelihood is:

$$(3) \qquad \frac{\partial \ln L(\beta)}{\partial \beta} = \sum_{i=1}^{N}\sum_{j=1}^{J} X_{ij}\left[1_{ij} - \Pr_i(j \mid \beta)\right] = 0\,,$$

where $\Pr_i(j \mid \beta)$ is the logit probability for individual *i* choosing the *j*th alternative. If a full set of ASCs are included, then

$$(4) \qquad X_{ij} = \begin{cases} 1 & \text{if } j \text{ chosen} \\ 0 & \text{otherwise} \end{cases}, \forall j\,,$$

and the score conditions associated with the ASCs can be written:

$$(5) \qquad \sum_{i=1}^{N}\left[1_{ik} - \Pr_i(k \mid \beta)\right] = 0 \quad \text{or} \quad \frac{1}{N}\sum_{i=1}^{N} 1_{ik} = \frac{1}{N}\sum_{i=1}^{N}\Pr_i(k \mid \beta), \forall k.$$

Equation 5 implies that fixed coefficient logit models with a full set of ASCs will generate in-sample predictions that match the data perfectly, a result that is consistent with our empirical findings in Table 1 and well known in the discrete choice literature (see, e.g., Ben-Akiva and Lerman, 1985).

As Gourieroux, Monfort, and Trognon (1984) have shown, the logit distribution falls within the broad class of distributions known as the linear exponential family of distributions. Other notable examples include the Poisson and normal distributions. What defines this family of distributions is that they are all mean-fitting distributions, implying that with the inclusion of ASCs, predictions from these distributions will match the data perfectly. A notable advantage of using linear exponential distributions in empirical work is that if the analyst has correctly specified the conditional expectation function of the distribution (i.e., its first moment), higher order misspecification will not lead to inconsistent parameter estimates (it will, however, bias standard error estimates, but this problem can be addressed if the analyst uses robust standard errors (White, 1981) instead of traditional standard errors). Thus, if the analyst specifies the first moment correctly, consistent parameter estimates will result. This makes the fixed coefficient logit model with ASCs appealing.

What is important to note, however, is that adding random coefficients to the logit distributions results in a mixture distribution that falls outside the linear exponential family. Random coefficient logit models, regardless of whether ASCs are included, will not necessarily generate in-sample predictions that match the data perfectly. This can be seen by looking at the score conditions for the simulated nonpanel random coefficient models logit model. The simulated likelihood function in this case is:

$$L(\bar{\beta},\sigma)=\sum_{i=1}^{N}\sum_{j=1}^{J}1_{ij}\ln\left(\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j\,|\,\beta_i^r)\right)=\sum_{i=1}^{N}\sum_{j=1}^{J}1_{ij}\ln\left(\frac{1}{R}\sum_{r=1}^{R}\frac{\exp(X_{ij}\beta_i^r)}{\sum_{k=1}^{J}\exp(X_{ik}\beta_i^r)}\right), \quad (6)$$

where $\beta_i^r=\bar{\beta}+\sigma U_i^r$, $U_i^r \sim N(0,1)$, and the score condition is:

$$\frac{\partial L(\bar{\beta},\sigma)}{\partial \beta_i^r}=\prod_{i=1}^{N}\left[\frac{\prod_{j=1}^{J}\left[X_{ij}\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j\,|\,\beta_i^r)(1-\Pr_i(j\,|\,\beta_i^r)\right]^{1_{ij}}}{\prod_{j=1}^{J}\left[\frac{1}{R}\sum_{r=1}^{R}\Pr_i(j\,|\,\beta_i^r)\right]^{1_{ij}}}\right]=0\,. \quad (7)$$

With the inclusion of ASCs, this condition does not imply perfect in-sample predictions. Thus, some degree of imperfect in-sample prediction can be expected from random coefficient logit models, but the precise degree will vary across applications.

To assess how well in-sample predictions from estimated logit models will match the data, we conducted an extensive Monte Carlo analysis where we know the underlying data generating process for the simulated data. Knowing the true data generating process allowed us to ascertain the in-sample prediction performance of maximum likelihood estimators when model misspecification is absent. If the in-sample predictions generated from these correctly specified models match the observed data well, then we can conclude that poor in-sample predictions arise due to some form of model specification, and not due to an inherent property of the estimator.

For brevity, we only summarize the main conclusions of our Monte Carlo simulation here and leave for an appendix (to be written at a later date – apologies) the simulation details. Across a number of specifications, we consistently found that the in-sample predictions for panel and non-panel random coefficient models with and without alternative specific constants matched the simulated data very closely. Under none of our simulations did we find the degree of poor in-sample prediction that we observed with the Alberta, Saskatchewan, or Mid-Atlantic data – see Table 1. Based on these findings, we conclude that the poor predictions found in our three applications are a result of model misspecification.

The implications of the above discussion for how analysts should proceed are unclear. If the analyst estimates logit models with random coefficients and finds poor in-sample predictions, the obvious 'first best' solution would be to continue to search for empirical specifications that fit the data well and predict well in sample. In practice, however, finding empirical specifications that satisfy these two criteria will be computationally difficult, time-consuming, and in many cases infeasible. This suggests that 'second best' less demanding approaches that address these two concerns may be attractive alternatives to applied researchers. Perhaps the simplest second best approach would be to estimate a fixed coefficient logit model with ASCs where the in-sample aggregate predictions will match the data perfectly. One limitation with this approach is that it in practice employs models with substitution patterns that are consistent with the independence of irrelevant alternatives (IIA). These restrictive substitution patterns can be partially relaxed by using nested logit models, but the considerably more flexible substitution patterns that come with random coefficient models will not be realized.

Another second best approach involves estimating random coefficient models with ASCs using a contraction mapping (Berry, 1994) that iteratively solves for the ASC values by matching the aggregate model predictions with the data. This algorithm was first used in the industrial organization literature to estimate discrete choice models of product differentiation using aggregate market share data (Berry, Levinsohn, and Pakes, 1995), but Berry, Levinsohn, and Pakes (2004) apply the algorithm to a disaggregate data context. Both of these applications employed generalized method of moments estimation techniques, and it was not until Murdock (2006) that the algorithm was used within a maximum likelihood framework with random coefficients. What is interesting to note, however, is that the use of this algorithm within a maximum likelihood estimation procedure *will not* generate maximum likelihood estimates. For this to be the case, the random coefficient maximum likelihood estimates would have to generate in-sample predictions which match the data precisely, but we showed above that in general this will not be the case. Thus, the estimates that one recovers from using the Berry contraction mapping to estimate random coefficient models within the maximum likelihood estimates are akin to maximum penalized likelihood estimates that Shonkwiler and Englin (2005) and von Haefen and Phaneuf (2003) have previously used. The idea behind maximum penalized likelihood estimation is that one maximizes the likelihood subject to a function that penalizes the likelihood for some undesirable behavior. Random coefficient logit models with ASCs that are estimated within the maximum likelihood framework using the Berry contraction mapping are observationally equivalent to estimating random coefficient logit models with ASCs within the maximum penalized likelihood framework with an infinitely weighted penalty function for poor in-sample predictions. A limitation with this approach is that the asymptotic properties of maximum penalized likelihood estimators are not well understood, but it does directly address the poor in-sample prediction problem. Moreover, due to plateaus and non-concavities in the penalized likelihood function, the choice of starting values and search algorithms can strongly influence the derived estimates.

Two other second best approaches for dealing with poor in-sample predictions involve estimating non-panel random coefficient models with ASCs within the maximum likelihood framework or incorporating observed choice into the construction of welfare measures as suggested by von Haefen (2003). As we demonstrate in the next section, the former approaches sacrifice the efficiency gains (which may be substantial) from introducing correlations across an individual's multiple trips for improved (but not perfect) in-sample predictions. Moreover, it makes estimation more computationally intensive. The idea of incorporating observed choice into welfare measurement construction is attractive because it simulates the unobserved determinants of choice in a way that implies perfect prediction for every observation and then uses the model's implied structure of substitution to ascertain how behavior and welfare change with changes in price, quality, and income. The approach can be used with any set of model estimates, but it does require a somewhat more computationally intensive algorithm for calculating welfare estimates (see von Haefen (2003) for details).

In the next section, we compare the sensitivity of welfare estimates to the use of these four second best strategies that address poor in-sample predictions. Our discussion will focus on the Mid-Atlantic application where all welfare measures have been generated. In future revisions to this paper, we will fill in the missing estimates for the Alberta and Saskatchewan data to see how the approaches fair in these alternative data environments.

Section IV. Sensitivity of Welfare Measures to Alternative Second Best Strategies

The bottom third of Table 2 reports welfare estimates from the Mid-Atlantic beach data for two policy scenarios – lost beach width at all Delaware, Maryland, and Virginia (DE/MD/VA) beaches and the closing of all northern Delaware beaches. We report the log-likelihood values as well as the percentage absolute prediction error for all sites in the first two rows to give the reader a sense of the relative statistical fit and in-sample prediction performance of the competing specifications. We also report unconditional (Train, 199?) and conditional (von Haefen, 2003) welfare measures for both scenarios as well as the percentage prediction error at the sites directly affected by the policy for all specifications.

In general, the results reported at the bottom of Table 2 have a number of qualitative implications, although the reader should interpret these implications cautiously until they have been confirmed with the Alberta and Saskatchewan data. First, all of the second best strategies suggested in the previous section for dealing with poor in-sample predictions – using fixed coefficients and alternative specific constants (column 3), using the Berry contraction mapping (columns 8 and 9), and using non-panel random coefficient specifications with alternative specific constants (columns 6 and 8), as well as incorporating observed choice into welfare measures (the conditional welfare measures in all columns) are effective tools for mitigating this problem. Second, the use of non-panel random coefficients results in a significant loss of statistical fit (compare the log-likelihoods in columns 4 and 5, 6 and 7, and 8 and 9). Because the non-panel random coefficient specifications generate smaller prediction error relative to the panel random coefficient models, there is a significant tradeoff between statistical fit and good in-sample prediction when specifying the correlation structure of random coefficients. Third, using the Berry contraction mapping in estimation modestly degrades statistical fit (compare the log-likelihoods in columns 6 and 8 as well as 7 and 9), but it does improve in-sample predictions, especially when panel random coefficients are used.

In terms of welfare estimates, the results in Table 2 imply that there is little difference between the conditional and unconditional welfare across *all* specifications and scenarios. This result is not surprising because the in-sample trip predictions for the affected sites are generally small. For the lost beach width at DE/MD/VA beaches, we see most of the point estimates are clustered in the range of -$3.34 to -$11.76, although the estimates that are based on non-panel random coefficient models with ASCs (columns 6 and 8) are positive in sign. As suggested above, the non-panel random coefficient models fit the data far worse than the panel random coefficient models, and thus we doubt the reliability of these estimates which also have rather large standard errors. For the welfare scenario simulating the closing of northern Delaware beaches, we see a general convergence of estimates between -$11.92 and -$23.69. We believe this interval represents a plausible range of welfare estimates that should be sufficiently informative for policy purposes.

One could interpret the results from the Mid-Atlantic data as suggesting that the addition of ASCs and random coefficients has minor effects on policy inference. Indeed, the point estimates for the fixed coefficient model without ASCs are qualitatively similar to the mid-range values for the more complex specifications. Based on the incomplete set of results that are reported in Table 2 for the Alberta and Saskatchewan data, we doubt that this empirical finding will carry over to the other applications where prediction error is more extreme. However, one might conclude from the results presented in Table 2 that simple models that predict reasonably well in-sample might generate welfare estimates that are robust to the inclusion of alternative specific constants and random coefficients.

Section V. Conclusion

Our goal in this research has been threefold: 1) to document the somewhat counterintuitive in-sample prediction problems that arise with random coefficient logit models that include ASCs; 2) to explore the sources of these problems using economic theory and Monte Carlo analysis; and 3) to suggest and evaluate alternative, second best, strategies for dealing with the poor in-sample predictions that researchers might find attractive in future empirical work. Across three data sets, we document that the addition of ASCs and especially panel random coefficients generates significant improvements in statistical fit but do not uniformly improve model prediction. We also show how these poor predictions influence derived welfare estimates, with the degree of under- and overprediction at sites that are directly impacted by the policy being correlated with the magnitude of welfare estimates. We then argue that the fixed coefficient logit model falls within the larger family of linear exponential distributions, and thus the inclusion of a full set of ASCs will generate in-sample trip predictions for each site that match the data perfectly. The introduction of random coefficients, however, results in a mixture distribution that falls outside the linear exponential family and thus will not imply perfect in-sample predictions. Results from an extensive Monte Carlo analysis suggest that the poor in-sample predictions observed in our three applications are likely due to some form of misspecification. To account for these model shortcomings, the analyst may find attractive one of the second best strategies that we empirically evaluate for addressing poor in-sample predictions. Our preliminary empirical results with the Mid-Atlantic data suggest that all of these strategies are effective in controlling for poor in-sample predictions, but the use of non-panel random coefficients significantly degrades model fit and generates perverse signs for some of the policy scenarios. Otherwise, our results suggest that the other second best approaches imply qualitatively similar welfare estimates that fall within a narrow range.

Finally, it is worth stepping back and directly addressing the fundamental question that motivated this research: do random coefficients and alternative specific constants improve welfare analysis? With regard to random coefficients, we believe that the richer substitution patterns implied by random coefficients are quite attractive, but the poor in-sample predictions that often result from these models (especially panel random coefficient versions) need to be addressed in some way. If not, welfare estimates lack credibility. With regard to alternative specific constants, we believe that their ability to control for unobserved attributes that may generate endogenity concerns makes them extremely attractive. One limitation with their inclusion, however, is that one needs either an RP data set with many objects of choice (sites in recreation models, or neighborhoods in locational equilibrium models) or additional SP data to identify the part worths of the different site attributes. When these data are available, we believe that ASCs are an attractive modeling innovation.

References

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams, "Perceptions vs. objective measures of environmental quality in combined revealed and stated preference models of environmental valuation," *Journal of Environmental Economics and Management* 32 (1997): 65-84.

Bayer, P. and C. Timmins, “Estimating equilibrium models of sorting across locations,” *Economic Journal* 117(2007): 353-374.

Ben-Akiva, M., and S. Lerman. *Discrete Choice Analysis*. MIT Press: Cambridge, 1985.

Berry, S. “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25(2): 242-262.

Berry, S., J. Levinsohn, and A. Pakes, “Automobile prices in equilibrium,” *Econometrica* 63(1995): 841-890.

Berry, S., J. Levinsohn, and A. Pakes, “Differentiated products demand systems from a combination of micro and macro data: the new car market,” *Journal of Political Economy* 112(2004): 68-105.

Gourieroux, C., A. Monfort, and A. Trognon, “Pseudo-maximum likelihood methods: application to Poisson models,” *Econometrica* 52(1984): 701-720.

Haener, M., P. Boxall, and W. Adamowicz, “Modeling recreation site choice: Do hypothetical choices reflect actual behavior?” *American Journal of Agricultural Economics* 83(2001): 629-642.

Huh, K., R. Sickles, “Estimation of the duration model by non-parametric maximum likelihood, maximum penalized likelihood and probability simulators,” *Review of Economics and Statistics* 76(1994): 683–694.

McFadden, D. and K. Train, “Mixed MNL of discrete response,” *Journal if Applied Econometrics*, 15(2000): 447-470.

Montricher, G., R. Tapia, J. Thompson. “Nonparametric maximum likelihood estimation of probability densities by penalty function methods,” *Annals of Statistics* 3(1975): 1329–1348.

Murdock, J., “Handling unobserved site characteristics in random utility models of recreation demand,” *Journal of Environmental Economics and Management* 51(2006): 1-25.

Parsons, G., D.M. Massey, and T. Tomasi, “Familiar and favorite sites in a random utility model of beach recreation,” *Marine Resource Economics*, 14(1999): 299-315.

Shonkwiler, J.S., and J. Englin, “Welfare losses due to livestock grazing on public lands: a count data systemwide treatment,” *American Journal of Agricultural Economics* 87(2005): 302-313.

Silverman, B. “On the estimation of a probability density function by the maximum penalized likelihood method,” *Annals of Statistics* 10(1982): 795–810.

Timmins, C. and J. Murdock, “A revealed preference approach to the measurement of congestion in travel cost models,” *Journal of Environmental Economics and Management* forthcoming.

Train, K., *Discrete Choice with Simulation.* Cambridge University Press: Cambridge, 2003.

Train, K., “Recreation demand models with taste variation over people,” *Land Economics*, 74:2(1998); 230-239.

von Haefen, R., “Incorporating observed choice into the construction of welfare measures from random utility models,” *Journal of Environmental Economics and Management* 45:2(2003): 145-165.

von Haefen, R., D. Phaneuf, “Identifying demand parameters in the presence of unobservables: a combined revealed and stated preference approach,” *Journal of Environmental Economics and Management* forthcoming.

Table 1 – Model Fits, In-Sample Predictions, and Compensating Surplus

Specifications				
Alternative Specific Constants?	No	Yes	No	Yes
Panel Random Parameters?	No	No	Yes	Yes
RP/SP Alberta moose hunting data from Adamowicz et al. (1997)				
Log-likelihood	-5,655.2	-5,376.7	-4,817.8	-4,521.6
Percentage improvement in log-likelihood	-	4.92%	14.8%	20.1%
Percentage absolute prediction error – all sites	30.0%	0.13%	45.6%	21.2%
CS for moose population reduction at WMU #348	-$14.11 (37.5)	-$9.47 (2.19)	-$25.00 (10.6)	-$20.91 (4.67)
Percentage prediction error at WMU #348	+10.8%	+0.12%	+16.3%	+6.60%
CS for moose population increase at WMU #344	$3.61 (2.50)	$98.34 (31.0)	$4.83 (3.19)	$73.02 (23.3)
Percentage prediction error at WMU #344	-48.4%	+0.07%	-88.1%	+26.3%
RP/SP Saskatchewan moose hunting data from Haener et al. (2001)				
Log-likelihood	-7,655.3	-7,482.3	-6,658.2	-6,547.5
Percentage improvement in log-likelihood	-	2.26%	13.0%	14.5%
Percentage absolute prediction error – all sites	26.3%	0.17%	56.8%	33.6%
CS for moose population reduction at WMZ #59	-$18.55 (7.52)	-$14.69 (2.99)	-$81.47 (11.5)	-$61.62 (9.77)
Percentage prediction error at WMZ #59	+31.5%	-0.02%	+82.9%	+30.7%
CS for moose population increase at WMZ #66	$27.54 (4.10)	$150.50 (36.9)	$22.30 (3.36)	$74.59 (14.3)
Percentage prediction error at WMZ #66	-39.2%	+0.20%	-31.3%	+13.2%
RP Mid-Atlantic beach data from Parsons et al. (1999)				
Log-likelihood	-13,160.2	-12,981.8	-11,015.8	-10,869.2
Percentage improvement in log-likelihood	-	1.36%	16.3%	17.4%
Percentage absolute prediction error – all sites	13.3%	<0.01%	27.0%	31.4%
CS for lost beach width at DE/MD/VA beaches	-$6.44 (1.16)	-$4.89 (4.62)	-$5.64 (1.41)	-$7.57 (3.26)
Percentage prediction error at DE/MD/VA beaches	-2.22%	<0.01%	-12.1%	-1.75%
CS for northern DE beach closings	-$19.56 (0.64)	-$21.88 (3.94)	-$14.97 (1.45)	-$16.83 (3.24)
Percentage prediction error at northern DE beaches	-5.04%	<0.01%	-6.01%	-8.28%

Robust standard errors in parentheses. All welfare estimates are per trip.

Table 2 – Alternative Strategies

Specifications								
Alternative Specific Constants?	No	Yes	No	No	Yes	Yes	Yes	Yes
Berry Contraction Mapping?	-	No	-	-	No	No	Yes	Yes
Random Parameters?	No	No	Panel	Non-Panel	Non-Panel	Panel	Non-Panel	Panel
RP/SP Alberta moose hunting data from Adamowicz et al. (1997)								
Log-likelihood	-5,655.2	-5,376.7	-4,817.8	-5,626.7	-5,368.3	-4,521.6		
Percentage absolute prediction error – all sites	30.0%	0.13%	45.6%	32.3%	0.14%	21.2%		
Unconditional CS for moose population reduction at WMU #348	-$14.11	-$9.47	-$25.00	-$23.09	-$9.56	-$20.91		
	(37.5)	(2.19)	(10.6)	(6.08)	(3.25)	(4.67)		
Conditional CS for moose population reduction at WMU #348	-$13.29	-$9.90	-$22.04	-$36.53	-$10.21	-17.01		
	(2.86)	(1.54)	(3.91)	(7.85)	(7.85)	(1.88)		
Percent. predict. error at WMU #348	+10.8%	+0.12%	+16.3%	+25.4%	+0.04%	+6.60%		
Unconditional CS for moose population increase at WMU #344	$3.61	$98.34	$4.83	$2.76	86.81	$73.02		
	(2.50)	(31.0)	(3.19)	(1.70)	(27.4)	(23.3)		
Conditional CS for moose population increase at WMU #344	$5.97	$98.00	$2.99	-$1.70	85.33	61.02		
	(7.17)	(29.8)	(2.37)	(1.62)	(23.5)	(20.6)		
Percent. predict. error at WMU #344	-48.4%	+0.07%	-88.1%	-45.2%	-0.20%	+26.3%		
RP/SP Saskatchewan moose hunting data from Haener et al. (2001)								
Log-likelihood	-7,655.3	-7,482.3	-6,658.2	-7,587.2	-7,472.9	-6,547.5		
Percentage absolute prediction error – all sites	26.3%	0.17%	56.8%	32.5%	4.84%	14.47%		
Unconditional CS for moose population reduction at WMZ #59	-$18.55	-$14.69	-$81.47	-$41.39	-$27.53	-$61.62		
	(7.52)	(2.99)	(11.5)	(16.3)	(12.1)	(9.77)		
Conditional CS for moose population reduction at WMZ #59	-$12.64	-$14.68	-$40.23	-$34.71	-$30.67	-$42.72		
	(5.31)	(2.65)	(4.57)	(12.4)	(12.1)	(4.58)		
Percent. predict. error at WMZ #59	+31.5%	-0.02%	+82.9%	+37.0%	+2.34%	+30.7%		
Unconditional CS for moose population increase at WMZ #66	$27.54	$150.50	$22.30	$25.14	$85.27	$74.59		
	(4.10)	(36.9)	(3.36)	(8.41)	(19.0)	(14.3)		
Conditional CS for moose population increase at WMZ #66	$32.82	$150.62	$26.57	$24.50	$79.56	$80.27		
	(4.18)	(36.3)	(3.68)	(8.85)	(17.0)	(14.7)		
Percent. predict. error at WMZ #66	-39.2%	+0.20%	-31.3%	-39.6%	+3.22%	+13.2%		
RP Mid-Atlantic beach data from Parsons et al. (1999)								
Log-likelihood	-13,160.2	-12,981.8	-11,015.8	-13,021.4	-12,856.5	-10,869.2	-12,874.5	-10,962.9
Percentage absolute prediction error – all sites	13.3%	<0.01%	21.8%	11.0%	+1.81%	+25.9%	<0.01%	<0.01%
Unconditional CS for lost beach width at DE/MD/VA beaches	-$6.44	-$4.89	-$5.64	-$3.34	$7.51	-$7.57	$1.83	-$11.76
	(1.16)	(4.62)	(1.41)	(0.60)	(7.72)	(3.26)	(3.96)	(4.27)
Conditional CS for lost beach width at DE/MD/VA beaches	-$6.58	-$4.89	-$7.15	-$3.59	$6.95	-$7.35	$1.78	-$11.53
	(1.18)	(4.27)	(1.11)	(0.70)	(7.16)	(1.79)	(3.70)	(1.95)
Percent. predict. error at DE/MD/VA beaches	-2.22%	<0.01%	-11.8%	-2.36%	+0.35%	-2.50%	<0.01%	<0.01%
Unconditional CS for northern DE beach closings	-$19.56	-$21.88	-$14.97	-$12.27	-$11.98	-$16.83	-$11.92	-$22.23
	(0.64)	(3.94)	(1.45)	(0.58)	(4.55)	(3.24)	(2.46)	(3.84)
Conditional CS for northern DE beach closings	-$20.75	-$22.04	-$16.58	-$13.54	-$13.51	-$19.34	-$13.27	-$23.69
	(0.69)	(2.34)	(2.09)	(0.69)	(1.76)	(1.88)	(1.09)	(2.54)
Percent. predict. error at northern DE beaches	-5.04%	<0.01%	-6.01%	-1.20%	-0.49%	-8.47%	<0.01%	<0.01%

Robust standard errors in parentheses. All welfare estimates are per trip.

[i] Using traditional maximum likelihood estimation techniques without the Berry contraction mapping is feasible in our application due to the relative small number of sites in the Mid-Atlantic data set. However, computational tractability requires the use of the Berry contraction mapping in random coefficient applications with many sites.

[ii] Parameter estimates are available upon request.