# Do Random Coefficients and Alternative Specific Constants Improve Policy Analysis? An Empirical Investigation of Model Fit and Prediction

H. Allen Klaiber*
*The Ohio State University*

Roger H. von Haefen**
*North Carolina State University*

**Abstract**

Concerns about unobserved heterogeneity – either in preference or attribute space – have led environmental economists to turn increasingly to discrete choice models that incorporate random parameters and alternative specific constants. We use four recreation data sets and several empirical specifications to show that although these modeling innovations often lead to substantial improvements in overall model fit, they also generate poor in-sample predictions relative to observed choices. Given the apparent tradeoff between fit and prediction, we then propose and empirically investigate a series of 'second-best' strategies that attempt to correct for the poor prediction we observe.

**Keywords:** Discrete choice, recreation demand, revealed preference, stated preference, alternative specific constants, random parameters

**JEL codes**: C25, Q51, L83

*\* Department of Agricultural, Environmental and Development Economics, The Ohio State University, 333 Ag Admin Bldg., 2120 Fyffe Rd, Columbus OH 43210. Email: klaiber.16@osu.edu.*

*\*\* Department of Agricultural and Resource Economics, North Carolina State University, Box 8109, Raleigh, NC 27695-8109. Email: roger_von_haefen@ncsu.edu.*

## 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 general framework, two econometric innovations that applied researchers are using with increasing regularity are random coefficients (McFadden and Train, 2000) and alternative specific constants (Berry, 1994). Random coefficients introduce unobserved preference heterogeneity and more plausible substitution patterns by relaxing the independence of irrelevant alternatives (IIA) assumption. Including a full set of alternative specific constants allows the analyst to control for unobserved attributes that may be correlated with observed covariates (Train, 2009; Moeltner and von Haefen, 2011).

Across a wide range of empirical applications, researchers typically find that models with random parameters and alternative specific constants generate substantial and statistically significant improvements in overall model fit as measured by log-likelihoods and information criteria (e.g., von Haefen and Phaneuf, 2007; Murdock, 2006). In our own empirical investigation of four recreation data sets, we also find large gains in model fit. However, our results also imply a less appreciated and more troubling result: the introduction of random parameters often results in model predictions that poorly match the in-sample aggregate choice patterns implied by the data used in estimation. This empirical regularity generates important implications for the credibility of policy analysis – why should one believe welfare measures derived from models that cannot, at a minimum, approximate 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 document this phenomenon using four recreation data sets and several empirical specifications. Two of the four applications combine revealed and stated preference (RP/SP) to identify all demand parameters (Adamowicz et al., 1997; Haener et al., 2001) as previously shown by von Haefen and Phaneuf (2007). The others exploit only revealed preference (RP) data (Parsons et al., 1999; von Haefen, 2003) and employ a variation of the two-step estimator first proposed by Berry, Levinsohn, and Pakes (2004) that Murdock (2006) used in the recreation context. With all four data sets, we find the introduction of random coefficients and alternative specific constants (ASCs hereafter) substantially and significantly improves overall model fit. Virtually any statistical test or information criteria would lead one to adopt models with these innovations over more restrictive versions. We also find that in-sample trip predictions often (but not uniformly) deteriorate with these richer empirical specifications, particularly with the introduction of random parameters.

To develop the intuition for why these poor predictions arise in practice, we review a key theoretical result from Gourieroux, Monfort, and Trognon (1984) about the properties of the linear exponential family of distributions of which the fixed parameter logit model is a member. The upshot of our discussion is that: 1) fixed parameter logit models with a full set of ASCs will generate in-sample aggregate trip predictions for each alternative that *perfectly* match the data, and 2) random coefficient models with ASCs will not predict perfectly in-sample in a maximum likelihood setting. Based on Monte Carlo evidence, we argue that these models should generate predictions that are reasonably close if the analyst has correctly specified the underlying data generating process. By implication, the poor in-sample predictions that we find in our empirical

applications arise because of model misspecification. This finding represents a cautionary tale to researchers about recent econometric innovations. These models may fail to account for important features of the data that are masked by focusing exclusively on in-sample statistical fit.

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 in-sample predictions that perfectly match aggregate data; 2) using less-efficient non-panel random coefficient models that, as we demonstrate, generate more plausible in-sample predictions; 3) employ a maximum penalized likelihood estimator that essentially penalizes combinations of parameters during the maximum likelihood search that predict poorly in-sample; and 4) following von Haefen (2003) and conditioning on observed choice in the construction of welfare measures. Our results suggest that each of these strategies is effective in terms of generating plausible in-sample predictions but differ considerably in terms of their implications for statistical fit.

The paper proceeds as follows. The next section documents the performance of random coefficient and fixed coefficient logit models with a full set of ASCs using four previously used 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 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.

**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 the problem is not an idiosyncratic feature associated with a single data set, we consider four recreation data sets that researchers have used in previously published studies with two combining revealed and stated preference (RP and SP, respectively) data and two using RP data alone. 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 our two RP studies, we use a two-step estimator originally outlined by Berry, Levinsohn, and Pakes (1995) and applied to a recreation model by Murdock (2006). This estimator recovers (J-1) alternative specific constants[1] and all individual varying covariates in the first stage of estimation, such as travel cost. Choice specific covariates are recovered in a second stage regression of the estimated ASCs on choice specific covariates and a constant term.

Our first RP/SP data set was used initially developed by Adamowicz et al. (1997) and consists of both revealed and stated 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 11 site attributes except travel cost in the RP and SP data are effects coded and interacted with three demographic variables. The second RP/SP data set was first used by Haener et al. (2001) and also consists of combined revealed and stated choice data for Canadian moose hunting. This data source, however, was collected in the neighboring province of Saskatchewan in 1994. The

---

[1] Only differences in utility enter into the logit model precluding estimation of the full set of ASCs. The normalized ASC is captured by a constant in the second stage of estimation.

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.

The first RP data set we consider examines Mid-Atlantic beach visitation and was first used by Parsons et al. (1999). This data set consists of seasonal trip data to 44 ocean beaches in 1997 for 375 individuals.[2] For each beach, we observe 14 site characteristic variables plus a constructed individual specific travel cost variable based on each recreator's home zip code and income. Our second RP data set focuses on recreation trips to the Susquehanna River basin for 157 nearby residents who take a combined total of 2,471 trips to one of 89 recreation sites. This data set was developed by von Haefen (2003) and includes a travel cost measure to each individual recreation site in addition to five site characteristics.

Table 1 summarizes our initial findings on model fit and prediction.[3] This table contains estimation results from four models associated with each of our empirical applications. These models include fixed and random parameter models with and without ASCs. All random parameter models assume all site characteristics have part worths that are normally distributed across the population,[4] but we assume the travel cost coefficient is fixed to avoid positive marginal utility of income estimates as well as numerical and computational difficulties (Thiene, Scarpa and Train, 2008). For the results reported in Table 1, our estimator differs from previous

---

[2] The authors collected information on day trips to 62 beaches from New Jersey to Maryland, but our analysis focuses on the 44 beaches that at least one respondent reported visiting.

[3] For compactness we do not report parameter estimates and standard errors for the different data sets and specifications, although these are available upon request. We note only that the parameters have intuitive signs that are consistent with published literature using these same data and are generally statistically significant.

[4] As is common in the recreation literature, we do not allow for correlation in the random parameters across attributes. As a sensitivity, we also considered truncated normal and triangular distributions for mixing distributions. These results were not qualitatively different, so we do not report them here.

two-step estimators in the following way. Similar to Murdock (2006) and Train and Winston (2007), we use maximum likelihood techniques to recover the travel cost parameter and a full set of ASCs that subsume all site characteristics that do not vary over individuals.[5] In contrast to Murdock and Train and Winston, our estimator does not employ the Berry (1994) contraction mapping algorithm, an issue we return to in a later section. Thus our estimator relies entirely on traditional maximum likelihood techniques, not the combination of maximum likelihood and Berry contraction mapping techniques that Murdock and Train and Winston employ. Because our focus in this article is on overall model fit and prediction, we do not need to use regression techniques to decompose the estimated ASCs into part worths associated with observable and unobservable site characteristics, so we do not implement the second step approaches proposed by these authors.

A comparison of log-likelihood values between fixed and random parameter specifications for all model combinations shows that inclusion of random parameters improves statistical fit significantly. For models that are nested (i.e., models with and without ASCs), likelihood ratio tests strongly reject the more restrictive models. And for non-nest models (i.e., models with and without random parameters), Akaike, Bayesian, consistent Akaike and corrected Akaike information criteria (see Hynes, Hanley and Scarpa (2008) for a discussion) all imply models with random parameters fit the data better than models with only fixed parameters. Overall, the highest log-likelihood values and information criteria are associated with models including both ASCs and random parameters. This general result holds regardless of application and is consistent with many other findings in the literature (e.g., von Haefen and Phaneuf, 2008).

---

[5] We do not include any demographic interactions in this model because preliminary testing suggested that they did not improve model fit.

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

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

where $s_i^S$ and $s_i^M$ are the in-sample share of trips to alternative $i$ and the model's prediction of the share of trips to alternative $i$, respectively, and $J$ is the number of alternatives. This prediction error statistic can be interpreted as the share weighted in-sample prediction error for each alternative and thus can be used to rank order the models in terms of their ability to generate in-sample predictions that match observed data. Intuitively, a model that can replicate aggregate trip predictions well for each alternative would generate a low prediction error, whereas a model with poor in-sample aggregate predictions for each alternative would score a relatively high value.

Examining the columns for absolute prediction error both with and without ASCs reveals an import result that further justifies including ASCs. For models employing fixed coefficients and ASCs, we see that the prediction errors are zero. This 'perfect' in-sample prediction follows from properties of the linear exponential family which we highlight in the following section. However, what is troubling with our results reported in Table 1 is that this pattern of perfect prediction disappears with random parameters specifications, despite their superior model fit. In the Susquehanna application for example, the fit deteriorates by nearly 50% with the introduction of random parameters.

Finally, it is interesting how these differences in fit and prediction vary across data sets. The best predictions are consistently associated with the Mid-Atlantic Beach dataset while the worst predictions are associated with the Susquehanna dataset. As the Mid-Atlantic Beach

dataset contains significantly more covariates, this finding suggests that a richer specification may capture aggregate site visitation patterns more accurately.

To assess the robustness of these empirical findings, we repeated the exercise using seasonal repeated discrete choice models which capture both participation and site choice decisions. These models assume the recreators have a fixed number of choice occasions (either 50 or 100 depending on the data set used) on which they chose whether and at what site to recreate. Results from these models are presented in Table 2. They qualitatively mirror those reported in Table 1, although they are generally more heterogeneous across models. Moreover, although the prediction errors can still be quite large (reaching nearly 50% in one case), the magnitude of these errors is lower for the most part relative to the trip allocation models. This later result may reflect the fact that with the seasonal repeated discrete choice model, the "no trip" alternative is consistently chosen on more than three-quarters of choice occasions and thus has significantly more leverage on the share weighted percentage error statistic than all other alternatives combined.

Our final set of empirical models considers latent class logit models employing panel random parameters with the results reported in Table 3. These models were estimated via the Expectation-Maximization algorithm (Dempster, Laird and Rubin, 1977) with ten randomly selected sets of starting values. Class membership probabilities were specified as functions of three to six demographic variables depending on the application, and the number of classes was chosen using corrected Akaike and Bayesian information criteria. Overall, the results in Table 3 suggest that these models maintain the general pattern of poor prediction, however, it appears that the ability of these models to better capture latent heterogeneity improve prediction over all of the random parameters models with normal mixing distributions. Nonetheless, prediction

errors can still be quite large – see, for example, the Susquehanna results. In all cases, inclusion of ASCs improves model prediction over models which do not include a full set of ASCs; however, in no case do we observe perfect prediction.

In summary, the results in Tables 1 through 3 suggest a somewhat counterintuitive result – including ASCs and especially random coefficients significantly improve overall model fit but generate in-sample trip predictions that poorly match the observed data. For many policy oriented applications, the failure to predict in-sample choice patterns accurately raises serious concerns about the credibility of derived welfare measures. For example consider an oil spill or acute environmental incident that impacts only a small number of sites. If predicted trips to affected sites under baseline conditions are larger than what is observed empirically, the estimated economic losses are also likely to be biased upwards as well.

### III. What explains these counterintuitive results?

In this section we use econometric theory to shed light on the counterintuitive results presented in the previous section. To motivate our main insight here, consider the conditional logit log-likelihood function for a sample of $N$ individuals each making separate choices from $J$ alternatives:[6]

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

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

---

[6] To avoid excessive notation we ignore the panel nature of most recreation data sets (i.e., the possibility that a given individual makes several discrete choices) although the results presented here generalize in a straightforward manor.

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

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

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

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

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

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. Other notable members include the Poisson and normal distributions. A defining characteristic of this family of distributions is that they are all mean-fitting, implying that with the inclusion of ASCs, aggregate predictions will match the estimation data perfectly. Moreover, 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. Standard errors will be biased, but this problem can be remedied by using robust formulas (White, 1981). Thus, if the analyst specifies the first moment correctly, consistent parameter estimates will result. This makes the fixed coefficient logit model with ASCs appealing for empirical work.

It is important to note, however, that adding random coefficients to the logit distribution results in a mixture distribution that falls outside the linear exponential family.[7] Random coefficient logit models, regardless of whether ASCs are included or how the mixing distribution is specified, 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 random coefficient logit model assuming a normal mixing distribution. The simulated likelihood function in this case is:

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

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

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

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

To assess how well in-sample predictions from correctly specified estimated logit models match aggregate shares from the data used in estimation, we conducted several Monte Carlo analyses where we know the underlying data generating process for the simulated data. Knowing the true data generating process allows us to ascertain the in-sample prediction performance of maximum likelihood estimators when model misspecification is absent. If the

---

[7] Because latent class models can be interpreted as random parameter with discrete mixing distributions, this result also holds for latent class models.

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 inherent properties of the estimator.

Our Monte Carlo worked as follows. Using summary statistics from Mid-Atlantic data set as a starting point, we simulated travel costs for a dataset of 50 alternatives with each alternative containing 5 site characteristics. We generated preference parameters for 1000 individuals who each faced 5 choice occasions resulting in a total of 5,000 choice occasions in the simulated dataset. In addition, we included in the "true" utility specification an alternative specific random normal term as well as a type I extreme value idiosyncratic term. Finally, we generated an individual and alternative varying travel cost measure. We simulated choice patterns treating the simulated preference parameters as the true population parameters. Using the simulated data and choice patterns estimation proceeded using simulated likelihood techniques.

We repeated this process 200 times and only report our main conclusions here for brevity. Across all specifications we assumed a fixed parameter on the travel cost term when simulating data while we allowed remaining site characteristic preference parameters to be normally distributed. Out initial mixed logit Monte Carlo estimation was purposefully misspecified by excluding ASC terms to document the importance of accounting for alternative specific unobservables to model prediction. Failure to account for these unobservables resulted in an expected absolute prediction error with an average across Monte Carlo iterations of 7.8%.

Our next mixed logit specification correctly specified the model by including the previously omitted alternative specific constant terms. Unlike the empirical datasets presented previously, we consistently found in-sample predictions for random coefficient models that

matched the simulated data extremely closely. The percentage absolute prediction errors averaged less than 0.001% and were never more than 0.01%. Under none of our simulations did we find the degree of poor in-sample prediction that we observed with the random coefficient models and empirical data sets reported in Table 1. Based on these findings, we conclude that the poor predictions found in our empirical 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, relative to fixed parameter models, finds substantial improvements in fit but poor in-sample predictions, the obvious 'first best' solution would be to continue to search for empirical specifications that fit the data 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.

## IV. Improving Prediction in the Presence of Misspecification

Since obtaining the correct model specification is difficult if not impractical in many empirical settings, 'second best' approaches that address poor prediction concerns may be warranted. In this section we propose and empirically investigate four possible strategies for proceeding. In our view, each of the proposed strategies has merit and we do not aim to identify which one is uniformly best. Instead, our goal is to lay out the relative strengths and weaknesses of each strategy and leave it to researchers to decide which one works best in their applications given the data, computational and policy considerations they face.

Perhaps the simplest second best approach would be to drop the random parameters and estimate a fixed coefficient logit model with ASCs where the in-sample aggregate predictions will match the data perfectly. This approach restricts preference heterogeneity to entering

exclusively through additive, idiosyncratic errors, which may be a reasonable assumption in a given application. Although this approach implies restrictive substitution patterns, it can be partially generalized by using a nested logit model, which relaxes the independence of irrelevant alternatives (IIA) assumption across nests.[8] Moreover, it is relatively easy to implement this strategy even in applications with large choice sets.

Another second best approach for dealing with poor in-sample predictions involves estimating non-panel random coefficient models with ASCs within the maximum likelihood framework. This approach sacrifices 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, our experience is that it makes estimation more computationally intensive, and the statistical significance of many random parameters tends to be diminished. Nevertheless, if the true model is a panel random parameter model, this approach will generate consistent parameter estimates, albeit with less precision.

A third approach involves estimating random coefficient models with ASCs using a contraction mapping (Berry, 1994) that iteratively solves for the ASCs that match the aggregate model predictions with the data. This algorithm was first used in the industrial organization literature to estimate discrete choice models of product choice 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) and Train and Winston (2007) that the algorithm was used within the maximum likelihood framework to estimate

---

[8] It is worth noting that the nested logit falls outside the linear exponential family of distributions, so the exact mean fitting property does not carry over. Our experience with estimating nested logit models with ASCs, however, suggests that these models generate relatively accurate in-sample aggregate predictions that match observed choices closely.

random parameter logit models. It is worth pointing out that both Murdock and Train and Winston's approaches are conceptually similar in spirit to the penalized likelihood estimates that Shonkwiler and Englin (2005) and von Haefen and Phaneuf (2003) use.

The idea behind maximum penalized likelihood estimation is that one maximizes the likelihood subject to an additive function that penalizes the likelihood for some undesirable feature such as poor in-sample prediction. If the weight on the function is sufficiently large, the maximum penalized likelihood framework becomes observationally equivalent to both Murdock and Train and Winston's approaches. Although all of these maximum likelihood approaches directly address the poor in-sample prediction problem, their asymptotic properties are not well understood, so how to construct standard errors and conduct standard errors is unclear. In our second best results presented below, we follow the above referenced papers and naïvely use standard maximum likelihood formulas to construct standard errors. Another limitation with the maximum penalized likelihood approach is that it can be computationally difficult to implement in practice.[9]

Our final second-best strategy to improve prediction is to incorporate observed choice into the construction of behavioral and welfare measures as suggested by von Haefen (2003). The strategy is attractive because it relies on traditional maximum likelihood estimates and simulates the unobserved determinants of choice in a way that implies perfect prediction for every observation. The approach can be used with any set of model estimates, but it requires a somewhat more computationally intensive algorithm for calculating demand and welfare

---

[9] A key challenge is choosing the weight to put on the additive penalty function. Too large of a weight will prevent the maximum likelihood routine from taking any steps away from the starting values, while too small a weight will not result in a model with good in sample predictions. Our experience is that choosing the optimal weight requires several iterations where the researcher must balance both of these concerns, and the final weight chosen might imply less than perfect in sample predictions.

estimates (see von Haefen (2003) for details) and does not directly address the underlying source of model misspecification in estimation. It also requires that the analyst observes a choice to condition on, which may not be possible if the analyst is forecasting demand and welfare effects for individuals not in the estimation sample.

Table 4 provides trip allocation estimation results for each of our empirical applications using these four approaches. Relative to the traditional maximum likelihood results reported in Table 1, the first three second-best approaches result in models that fit the data worse as evidenced by the log-likelihood values. This is especially true for the fixed parameter and non-panel random parameter models. As can be seen, each of these approach implies significantly smaller prediction error relative to what is reported in Table 1.


## 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 econometric theory; 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 four data sets, we document that the addition of ASCs and especially panel random coefficients into site choice models improves model fit but degrades in-sample prediction. These findings persist when we consider seasonal recreation demand models as well as latent class models. We argue that poor in-sample prediction raises validity concerns for policy analysis.

Building on econometric theory, we show 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. Monte Carlo results suggest that the poor in-sample predictions observed in our four applications are likely due to some form of model misspecification.

To account for these shortcomings, we propose four second best strategies that applied researchers may find attractive. Our empirical results suggest that all of these strategies are effective in controlling for poor in-sample predictions, but the use of the fixed and non-panel random coefficient models significantly degrades model fit. The use of either a maximum penalized likelihood estimator, which is a close relative to Murdock (2006) and Train and Winston's (2007) approaches using Berry's (1994) contraction mapping, generates improved prediction, but at the cost of obtaining maximum likelihood estimates with unknown asymptotic properties.

It is worth emphasizing that our focus in this paper has centered on in-sample fit and prediction. These are certainly important considerations with model selection, but a more stringent and perhaps more relevant consideration for applications where forecasting demand and welfare effects is the focus is out-of-sample performance. Our view is that good in-sample performance is a necessary but not sufficient condition for credible policy inference. We hope data and methods in environmental economics will continue to evolve to the point where our policy recommendations have both in-sample and out-of-sample validity.

In closing, it is worth stepping back and directly addressing the fundamental question that motivated our research: do random coefficients and alternative specific constants improve policy analysis with discrete choice models? With regard to random coefficients, we generally believe

that the richer substitution patterns that comes with random preference heterogeneity are quite attractive and lead to more credible inference. However, the poor in-sample predictions that often result from these models (especially panel random coefficient versions) must be addressed in some way. If not, the credibility of policy implications derived from these models is suspect. First and foremost, we believe that researchers should invest more effort into finding specifications with better in-sample (and out-of-sample) prediction properties. Realistically, however, our current stock of empirical models and time constraints will make our proposed second best strategies attractive. With regard to alternative specific constants, we believe that their ability to control for unobserved attributes makes them extremely attractive. One limitation with their inclusion, however, is that for policy analyses involving changes in site attributes that do not vary across individuals, one needs either: 1) an RP data set with many objects of choice (sites in recreation models, or neighborhoods in locational equilibrium models) and credible instruments for potentially endogenous attributes (Bujosa et al., 2015), or 2) additional SP data to identify the marginal value of key attributes (von Haefen and Phaneuf, 2008). When these data are available, ASCs should be routinely included in discrete choice applications.

**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.

Bayer, P., N. Keohane and C. Timmins. 2009. "Migration and hedonic valuation: The case of air quality," *Journal of Environmental Economics and Management*, 58(1).

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.

Bujosa, A. A. Riera, R.L. Hicks and K.E. McConnell, "Densities rather than shares: improving the measurement of congestion in recreation demand models," *Environmental and Resource Economics*, 61:2(2015): 127-140.

Dempster, A., N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B*, 39:1(1977): 1-38.

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.

Hynes, S., N. Hanley and R. Scarpa, "Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreation demand models," *American Journal of Agricultural Economics*, 90:4(2008): 1011-1027.

Klaiber, H.A. and D.J. Phaneuf, "Valuing open space in a residential sorting model of the Twin Cities," *Journal of Environmental Economics and Management*, 60:2(2010): 57-77.

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

Moeltner, K. and R. von Haefen, "Microeconometric strategies for dealing with unobservables and endogenous variables in recreation demand models," In *Annual Review of Resource Economics*, *Vol 3*, G. Rauser, V.K. Smith, and D. Zilberman, 2011.

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.

Thiene, M., R. Scarpa and K. Train, "Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps," *American Journal of Agricultural Economics*, 90:4(2008): 994-1100.

Timmins, C. and J. Murdock, "A revealed preference approach to the measurement of congestion in travel cost models," *Journal of Environmental Economics and Management*, 53(2007): 230-249.

Train, K., *Discrete Choice Methods with Simulation, 2nd Edition.* Cambridge University Press: Cambridge, 2009.

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

Train, K. and G. Winston, "Vehicle choice behavior and the declining market share of U.S. automakers," *International Economic Review,* 48:4(2007): 1469-1496.

Tra, C., "A discrete choice equilibrium approach to valuing large environmental changes." *Journal of Public Economics*, 94(2010): 183-196.

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. and D. Phaneuf, "Identifying demand parameters in the presence of unobservables: A combined revealed and stated preference approach," *Journal of Environmental Economics and Management*, 56:1(2008): 19-32.

**Table 1. Baseline Performance using ASCs and Panel Random Coefficients for Trip Allocation Models**

| Empirical Example | Specification | Alternative Specific Constants | | | | | | | |
| | | NO | | | | YES | | | |
| | | Log-Likelihood | Abs Pred Err | # Observations | # Parameters | Log-Likelihood | Abs Pred Err | # Observations | # Parameters |
|---|---|---|---|---|---|---|---|---|---|
| Mid-Atlantic Beach RP | Fixed | -13,160 | 13.3% | 375 | 15 | -12,982 | 0.0% | 375 | 44 |
| | Random | -11,014 | 21.7% | 375 | 29 | -10,821 | 12.7% | 375 | 72 |
| Alberta RP/SP | Fixed | -5,655 | 29.9% | 271 | 50 | -5,377 | 0.0% | 271 | 63 |
| | Random | -4,823 | 33.9% | 271 | 62 | -4,537 | 19.6% | 271 | 75 |
| Susquehanna RP | Fixed | -5,303 | 61.5% | 157 | 6 | -4,385 | 0.0% | 157 | 89 |
| | Random | -4,381 | 86.4% | 157 | 11 | -3,604 | 48.9% | 157 | 94 |
| Saskatchewan RP/SP | Fixed | -7,655 | 26.3% | 532 | 42 | -7,482 | 0.0% | 532 | 52 |
| | Random | -6,673 | 59.7% | 532 | 52 | -6,567 | 35.2% | 532 | 62 |

**Table 2. Baseline Performance using ASCs and Panel Random Coefficients for Seasonal Models**

| Empirical Example | Specification | Alternative Specific Constants | | | | | | | |
| | | NO | | | | YES | | | |
| | | Log-Likelihood | Abs Pred Err | # Observations | # Parameters | Log-Likelihood | Abs Pred Err | # Observations | # Parameters |
|---|---|---|---|---|---|---|---|---|---|
| Mid-Atlantic Beach RP | Fixed | -27,759 | 1.9% | 540 | 22 | -27,564 | 0.0% | 540 | 51 |
| | Random | -22,592 | 19.3% | 540 | 37 | -22,545 | 14.5% | 540 | 81 |
| Alberta RP/SP | Fixed | -8,890 | 3.2% | 271 | 54 | -8,683 | 0.0% | 271 | 67 |
| | Random | -7,792 | 10.5% | 271 | 67 | -7,549 | 15.4% | 271 | 80 |
| Susquehanna RP | Fixed | -11,599 | 9.3% | 157 | 10 | -10,743 | 0.0% | 157 | 93 |
| | Random | -9,392 | 40.6% | 157 | 16 | -8,526 | 49.0% | 157 | 105 |
| Saskatchewan RP/SP | Fixed | -10,565 | 1.1% | 532 | 46 | -10,172 | 0.0% | 532 | 56 |
| | Random | -9,258 | 4.8% | 532 | 57 | -9,194 | 3.0% | 532 | 67 |

**Table 3. Latent Class Panel Random Parameter Models**

| Empirical Example | Specification | Alternative Specific Constants | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | NO | | | | | YES | | | | |
| | | Log-Likelihood | Abs Pred Err | # Observations | # Parameters | Latent Classes | Log-Likelihood | Abs Pred Err | # Observations | # Parameters | Latent Classes |
| Mid-Atlantic Beach RP | Trip | -11,786 | 13.3% | 375 | 99 | 4 | -12,312 | 6.6% | 375 | 64 | 2 |
| | Seasonal | -24,443 | 1.9% | 540 | 66 | 3 | -24,733 | 1.6% | 540 | 85 | 3 |
| Alberta RP/SP | Trip | -4,865 | 31.0% | 271 | 113 | 7 | -4,678 | 23.0% | 271 | 110 | 6 |
| | Seasonal | -8,041 | 4.4% | 271 | 106 | 6 | -8,096 | 1.3% | 271 | 68 | 3 |
| Susquehanna RP | Trip | -4,370 | 66.6% | 157 | 42 | 5 | -3,561 | 33.3% | 157 | 125 | 5 |
| | Seasonal | -10,036 | 10.3% | 157 | 27 | 3 | -9,276 | 12.9% | 157 | 126 | 5 |
| Saskatchewan RP/SP | Trip | -6,662 | 47.4% | 532 | 113 | 8 | -6,627 | 29.5% | 532 | 95 | 6 |
| | Seasonal | -9,475 | 2.7% | 532 | 64 | 4 | -9,504 | 0.8% | 532 | 59 | 3 |

**Table 4. Potential Solutions to Poor Prediction (all with ASCs included)**

| Empirical Example | Specification | Log-Likelihood | Abs Pred Err | # Observations | # Parameters |
|---|---|---|---|---|---|
| Mid-Atlantic Beach RP | Fixed | -12,982 | 0.0% | 375 | 44 |
| | Non-Panel | -12,838 | 2.2% | 375 | 58 |
| | Penalty | -10,911 | 1.0% | 375 | 58 |
| | Conditional | -10,821 | 0.0% | 375 | 58 |
| Alberta RP/SP | Fixed | -5,377 | 0.0% | 271 | 63 |
| | Non-Panel | -5,368 | 0.1% | 271 | 75 |
| | Penalty | -4,572 | 0.8% | 271 | 75 |
| | Conditional | -4,537 | 0.0% | 271 | 75 |
| Susquehanna RP | Fixed | -4,385 | 0.0% | 157 | 89 |
| | Non-Panel | -4,384 | 0.9% | 157 | 94 |
| | Penalty | -3,747 | 3.6% | 157 | 94 |
| | Conditional | -3,604 | 0.0% | 157 | 94 |
| Saskatchewan RP/SP | Fixed | -7,482 | 0.0% | 532 | 52 |
| | Non-Panel | -7,467 | 5.1% | 532 | 62 |
| | Penalty | -6,598 | 1.9% | 532 | 62 |
| | Conditional | -6,567 | 0.0% | 532 | 62 |