A Cautionary Note on Designing Discrete Choice Experiments: A Comment on Lusk and Norwood's "Effect of Experiment Design on Choice-Based Conjoint Valuation Estimates"

Richard T. Carson, Jordan J. Louviere, and Nada Wasi

Recent Monte Carlo work on choosing experimental designs for discrete choice experiments seemed to greatly simplify this choice for applied researchers. It suggested that (*a*) commonly used designs can generate unbiased estimates for indirect utility function specifications with main effects only and main effects plus higher order terms, and (*b*) random designs are more efficient than main effects designs. We show that these results are very specific to the particular indirect utility specifications studied and do not generalize well. We further show that conclusions drawn concerning random designs are problematic and potentially dangerous for applied researchers.

Key words: choice experiments, experimental design, non-market valuation.

Discrete choice experiments (DCEs) are widely used in many fields (Louviere, Hensher, and Swait 2000) including environmental economics, health policy, marketing, and transportation. However, economists typically do not receive formal training in either design of surveys or statistical design of experiments. In this article, we look at experimental design issues that have wide implications for the types of models that can be estimated and the cost of conducting DCE projects via the sample sizes required. As a branch of statistics, the principles of optimal design are likely to be accessible to those with strong backgrounds in econometrics. Work on experimental design is rapidly growing; recently, highly statistically efficient designs, new evaluation criteria, and new generation algorithms have been developed that offer more design choices (see, e.g., Street and Burgess 2007; Rose and Scarpa 2008). This poses important questions for applied researchers, such as, does the experimental design chosen by a particular researcher matter? And, if so, how much does it matter? Evaluation and choice of designs is a complex multi-criteria decision, dependent to a large degree on the assumptions analysts are willing to make about the nature of the underlying data generating process.

J.L. Lusk and F.B. Norwood (hereafter "L&N") tried to address these important issues in their 2005 article "Effect of experimental design on choice-based conjoint valuation estimates." Their main conclusions were *"... All experimental designs considered in this study generated unbiased valuation estimates. However, random designs or designs that explicitly incorporated attribute interactions generated more precise valuation estimates than main effects only designs. A key result of our analysis is that a large sample size can substitute for a poor experimental design..."*. While comforting to applied researchers, these findings would appear to be at odds with decades of work by statisticians (e.g., Box, Hunter, and Hunter 1978; Box and Draper 1987; Cochran and Cox 1992) which suggests that (*a*) incorrect experimental designs often produce biased parameter estimates, (*b*) random designs are inherently inefficient, and (*c*) a large sample size is a poor substitute for a good design.

The purpose of this article is to point out that L&N's conclusions should be viewed as potentially misleading, or at best, limited to a narrow range of situations. The intent of L&N's work was admirable and may be useful in particular circumstances; however, it is unlikely to generalize well due to caveats associated with the nature of L&N's Monte Carlo experiments.

Carson is professor at the Department of Economics, University of California, San Diego. Louviere is professor at the School of Marketing and executive director of Centre for the Study of Choice, University of Technology Sydney. Wasi is a research fellow at the Centre for the Study of Choice and at School of Finance and Economics, University of Technology Sydney.

The authors would like to acknowledge support from Australian Research Council grant DP 0774142.

Amer. J. Agr. Econ. 91(4) (November 2009): 1056–1063

DOI: 10.1111/j.1467-8276.2009.01309.x

We note and discuss several issues with L&N's simulation work, and suggest that researchers should be cautious about taking their conclusions for granted. We think caution is warranted as a number of researchers already have based design choices on L&N's recommendations.[1]

There are two main limitations with L&N's work:

1) The choice of indirect utility function parameters greatly limits their conclusions.
2) The conclusion that random designs give more precise estimates than other designs is likely to stem from the particular way they performed Monte Carlo experiments on random designs.

L&N tried to evaluate the performance of six different designs (random design, main effects design, main effects + two-way interaction effects design, two Kuhfeld et al. designs (SAS designs), and a "FULL/BIN" design) under different experimental conditions controlled by a $3^2 \times 2$ "master" experiment.[2] The master experiment is defined by 2 sample sizes (n1, n2) $\times$ 3 indirect utility functions (hereafter "IUFs") where all attributes are inherently continuous and three IUFs where some attributes are qualitative. The IUFs differ in degrees of nonlinearity.

Limitation 1: Choice of Indirect Utility Function (IUF) Parameters

There are three major caveats associated with L&N's choice of IUF parameters. The first caveat is that both nonlinear IUFs that L&N considered are "nearly linear." Specifically, the three IUFs do not differ much, which can be seen by comparing differences in choice outcomes produced by these IUFs. We begin by noting that only differences in utility matter (i.e., only a ranking of choice options matters), and consider L&N's three continuous IUFs, "Linear," "Nonlinear 1," and "Nonlinear 2."

Table 1. L&N IUF Parameters

	Linear	Nonlinear 1	Nonlinear 2
β_1	−1	−1	−1
β_2	2	2	2
β_3	3	3	3
β_{22}	0	0.2	0.5
β_{33}	0	0.2	0.5
β_{23}	0	−0.5	−1.25

To wit, the utility of person i from choice j is given by

(1)
$$U_{ij} = \alpha_j + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_{22}(x_{2ij})^2 + \beta_{33}(x_{3ij})^2 + \beta_{23}(x_{2ij} * x_{3ij}) + \varepsilon_{ij}$$

where $j = 1, \ldots, 3$; $i = 1, \ldots,$ N; $\{x_{1ij}, x_{2ij}, x_{3ij}\}$ take on three discrete values (1, 2 or 3); α_j is an alternative-specific constant (ASC) for choice j, $\alpha_3 = 0$. The three IUFs differ by specifying different parameters as shown in table 1.

Table 2 gives utility values calculated for all possible options and their ranking based on calculated utility values associated with each IUF. If the rankings are the same for any pair or triple of profiles in a design, the observed choice outcomes will be the same.

Table 2 reveals that, while the IUF parameters produce different latent utility values, the rankings are very similar. For example, the Spearman correlation between linear utility and "Nonlinear 1" utility is 0.997; the Spearman correlation between linear utility and "Nonlinear 2" ranking is 0.926.[3] Thus, it is unlikely that these three "nearly linear" IUFs differ sufficiently so as to permit general conclusions about ignoring versus including interaction terms.[4] L&N noted that the parameters that they chose were consistent with case studies; yet if one views this set of parameters as a "typical" case study, it is unclear why one

[1] For example, Goldberg and Roosen (2007) use a random design for their choice experiment based on L&N's recommendation in their study of food safety.

[2] "FULL/BIN" design was created by generating the full factorial for a single 3^3 design. Identical copies of this design were placed in three "bins." Choice sets were created by randomly pulling one profile from each "bin" (see Lusk and Norwood 2005, p. 776, for further information).

[3] Part of the problem with the Monte Carlo design is choosing the interaction terms to be five to ten times smaller than the main effects. This is said to be based on case studies in Louviere, Hensher, and Swait (2000). While many of the interaction terms are small and insignificant, there are some interaction parameters that are fairly large relative to main effects and statistically significant. Further, one has to be careful about the interpretation of the interaction terms in logit models (Ai and Norton 2003) as the effect of the attribute interaction on choice probabilities is not the same as the effect of the attribute interaction on latent utility. The interaction effects on choice probability can be very important and highly significant even when the point estimate on the interaction parameter is zero.

[4] A similar example can be given for their cases of discrete utility function.

Table 2. Utilities for and Rankings of Choice Options Based on L&N IUFs

	Attributes			Linear		Nonlinear 1		Nonlinear 2	
Profile	x1	x2	x3	Utility	Rank	Utility	Rank	Utility	Rank
1	1	1	1	4	3.5	3.9	3	3.75	3
2	1	1	2	7	11	7	11	7	10.5
3	1	1	3	10	20	10.5	21	11.25	26
4	1	2	1	6	8	6	7.5	6	8
5	1	2	2	9	17	8.6	16.5	8	13.5
6	1	2	3	12	24.5	11.6	25	11	25
7	1	3	1	8	14	8.5	14.5	9.25	18.5
8	1	3	2	11	22.5	10.6	22.5	10	21.5
9	1	3	3	14	27	13.1	27	11.75	27
10	2	1	1	3	2	2.9	2	2.75	2
11	2	1	2	6	8	6	7.5	6	8
12	2	1	3	9	17	9.5	18	10.25	23
13	2	2	1	5	5.5	5	5.5	5	5.5
14	2	2	2	8	14	7.6	13	7	10.5
15	2	2	3	11	22.5	10.6	22.5	10	21.5
16	2	3	1	7	11	7.5	12	8.25	15
17	2	3	2	10	20	9.6	19.5	9	16.5
18	2	3	3	13	26	12.1	26	10.75	24
19	3	1	1	2	1	1.9	1	1.75	1
20	3	1	2	5	5.5	5	5.5	5	5.5
21	3	1	3	8	14	8.5	14.5	9.25	18.5
22	3	2	1	4	3.5	4	4	4	4
23	3	2	2	7	11	6.6	10	6	8
24	3	2	3	10	20	9.6	19.5	9	16.5
25	3	3	1	6	8	6.5	9	7.25	12
26	3	3	2	9	17	8.6	16.5	8	13.5
27	3	3	3	12	24.5	11.1	24	9.75	20

would worry about ignoring interaction effects anyway. Further, given the well-known difficulties with estimating interaction effects in general (McClelland and Judd 1993) and the failure of many if not most designs used in the economics literature to clearly identify such effects, it is unclear how much is known about what the "typical" case looks like.

The second caveat associated with L&N's choice of parameters seems to be accidental, and whether it impacts their main conclusions will depend upon a particular application. L&N used marginal willingness to pay of moving from $(x_2 = 2, x_3 = 2)$ to $(x_2 = 3, x_3 = 2)$ as a criterion to evaluate design performance. This can be expressed as $MWTP = -(\beta_2 + 5\beta_{22} + 2\beta_{23})/\beta_1$. Using $MWTP$ as a criterion to evaluate design choices would not be problematic had they examined a sufficiently large range of true parameters, but they did not do that. Instead, L&N assigned true parameters to β_{22} and β_{23} as 0.2 and –0.5, respectively, for "Nonlinear 1," and β_{22} and β_{23} as 0.5 and –1.25, respectively, for "Nonlinear 2." This implies that the term $(5\beta_{22} + 2\beta_{23})$ actually equals zero in both cases. If a design yields biased estimates such that $(\hat{\beta}_{22}, \hat{\beta}_{23}) = (\lambda\beta_{22}, \lambda\beta_{23})$ where λ can be any number, then the term $(5\beta_{22} + 2\beta_{23})$ always will equal zero, and $MWTP$ will be unbiased.[5] We suggest that one should also check whether estimates of individual βs are biased as unbiased βs imply unbiased $MWTP$, but not vice versa.[6]

The third caveat involves an identification issue. All parameters in the IUF in equation (1) are identified in a main effects design, but this is also a fortunate accident, which will not be

[5] For example, one can think of λ as a scale parameter. If the variance of the unobserved component is smaller (larger) than 1.67, all parameters will be proportionally biased upward (downward).

[6] There are a number of other issues that one would want to take into account in examining the property of experiment designs since more complicated models such as mixed logit and models allowing for heterogeneous error variances across agents are now commonly employed (see, e.g., Ferrini and Scarpa 2007). L&N do not examine the influence of their recommended designs on such models, further limiting the scope of their conclusions.

Table 3. Complete Factorial Design of 23

Profiles	Fraction	x_1	x_2	x_3
1	1	−1	−1	−1
2	1	−1	1	1
3	1	1	−1	1
4	1	1	1	−1
5	2	−1	−1	1
6	2	−1	1	−1
7	2	1	−1	−1
8	2	1	1	1

true in general.[7] Consider the following simple example of a binary choice problem with three attributes, each with two levels that gives all possible combinations in table 3 (full factorial). Two possible "main effects only" designs are either the first or last four combinations.

We conducted three Monte Carlo experiments to interact two different designs with two IUFs. The results are in table 4. The first experiment assumes that the true IUF is linear and uses one main effects only design. The mean coefficient estimates are unbiased as expected; the average bias in absolute value is 0.09 for all three parameters. The second experiment uses the same main effects design, but assumes that the true IUF has attribute interaction effects. The estimates are seriously biased from the true values in the latter case. The last experiment involves the same nonlinear IUF as the second experiment, but now uses the full factorial design; the estimates are unbiased in this case. This example clearly provides a case where a main effects design cannot identify interaction effects: the parameters are estimable, but the design matrix is very ill-conditioned. It is also worth noting that all two-way interactions are perfectly confounded with main effects in this case.

Another example is in table 5.4 of Louviere, Hensher, and Swait (2000). In that example, an L^{MA} design is used to generate a paired DCE. Although the example focused on alternative-specific designs, we can treat it as a design for a generic problem. If we convert the numbers in table 5.4 to traditional 0, 1 codes for attribute levels, and calculate the attribute-by-attribute difference matrix, the main effects columns are orthogonal. If we also include the two- and three-way interactions, these columns are NOT orthogonal, and one cannot identify two of the two-way interactions. There are also very high correlations (>0.7) between some interactions and main effects; so identification is very fragile.

Limitation 2: Nature of Random Designs Implementation

The concern here is whether the finding that random designs give more precise estimates than other designs (including main effects only designs) is associated with conceptual flaws in the Monte Carlo experiments for random designs. L&N (pp. 777–78) described how they performed their Monte Carlo experiments:

Step 1: A particular experimental design was chosen and generated.
Step 2: The experimental design was replicated to achieve the desired sample size.
Step 3–7 [omitted] involves simulating choice outcomes from the given designs in Steps 1 and 2, estimating MNL model, and calculating welfare measures.
Step 8: Steps 1 through 7 were repeated 500 times. The end result of the exercise is a distribution of 500 marginal and total WTP estimates for each functional form, experimental design, and sample size.

By repeating Step 1, L&N produce a new design, being generated for each of the 500 iterations. L&N clearly note this on p. 775:

"The RAND [random] design was created by randomly drawing choices from the full factorial design consisting of $3^3 \times 3^3 \times 3^3 = 19{,}683$ choice sets. In the "low" sample size treatments, 243 choice sets were randomly drawn at each Monte Carlo iteration *and in the "high" sample size treatments, 729 choice sets were randomly drawn* at each Monte Carlo iteration."

By drawing 500 random designs, L&N's approach likely approximates a full factorial, or at least it ensures coverage of a large portion of the total design space. Of course, a single random design, which is what would be used in any particular study with a fixed sample size, will not have this property. The issue is straightforward: by averaging over different randomly chosen designs that result in biased estimates in different directions one obtains a reasonable

[7] L&N do state (p. 778) that "researchers must incorporate a priori information about interactions into experimental design of choice sets." It is, though, precisely the usual lack of a priori information about these interactions and the suggestion in L&N that they are identified in their main effects and random designs that is troubling. In some ways, the problem running through the L&N article is confusing the precision of parameter estimates under particular utility specifications with what parameters are clearly identified when they can take on unknown values.

Table 4. Results of Monte Carlo Simulation for Main Effects Versus Complete Factorial Designs Interacting with Linear and Nonlinear IUFs

True Utility Function	Experiment I Linear (Main Effects Only)			Experiment II Nonlinear (Main Effects + Two-Way Int.)			Experiment III Nonlinear (Main Effects + Two-Way Int.)		
Design	Main Effects Only			Main Effects Only			Complete Factorial		
Attributes	β_{true}	$\overline{\beta}$	$\overline{bias}$	β_{true}	$\overline{\beta}$	$\overline{bias}$	β_{true}	$\overline{\beta}$	$\overline{bias}$
x1	**−1.00**	−1.02 (0.12)	0.09	**−1.00**	−0.77 (0.07)	0.23	**−1.00**	−1.04 (0.12)	0.10
x2	**0.50**	0.50 (0.11)	0.09	**0.50**	0.15 (0.07)	0.35	**0.50**	0.51 (0.12)	0.09
x3	**0.50**	0.50 (0.11)	0.09	**0.50**	0.14 (0.07)	0.36	**0.50**	0.51 (0.12)	0.09
x1 × 2	–			**0.25**	−0.10 (0.05)	0.35	**0.25**	0.26 (0.12)	0.10
x1 × 3	–			**0.25**	−0.10 (0.05)	0.35	**0.25**	0.26 (0.12)	0.10
x2 × 3	–			**0.25**	0.50 (0.05)	0.25	**0.25**	0.25 (0.12)	0.09
x1 × 2 × 3	–			**0.00**	0.00 (0.12)	0.09	**0.00**	−0.01 (‘0.13)	0.10

Note: We assign the profiles given in table 3 as profiles of option 1, and treatment option 2 as status quo (all attributes are zeros). Equivalently, one can assign two options with nonzero attributes, provided that the differences in attribute levels are those values given in the table 3. For the complete factorial design, we replicate these eight choice sets sixty times. For the main effects design, we replicate the first four choice sets 120 times. As a result, there are 480 observations in all experiments. The number of Monte Carlo replications is 500. If β_{true} denotes the true parameter values, $\overline{\beta} = \frac{1}{500}\sum_{m=1}^{500}\beta_m$ and $\overline{bias} = \frac{1}{500}\sum_{m=1}^{500}|\beta_m - \beta_{true}|$ where m is the Monte Carlo replication index. The numbers in parentheses are empirical standard deviation, given by $sqrt(\sum_{m=1}^{499}(\beta_m - \overline{\beta})^2/499)$.

average estimate that obscures the risk of using any particular random design. In any empirical application, the researcher draws only one design chosen randomly from the allowable set.[8] L&N's procedure is conceptually equivalent to having many different researchers, each with a separate independent sample and experimental design drawn from the full factorial, pool their results to obtain a single set of point estimates but then calculating the standard errors as if the sample size used was only that employed by one of the researchers. Thus, L&N's procedure does not mimic what would be done in empirical field applications, and the Monte Carlo simulation results using this approach are potentially misleading.

Table 5 gives estimates from four different Monte Carlo experiments that use parameters from their "Linear" continuous IUF with a sample size of 243 with 500 replications, with *MWTP* calculated the same way as L&N. Experiment A uses the main effects design (nine replications of twenty-seven choice sets). For Experiment B, we display results of three different random designs with 500 more Monte Carlo replications on the error term.[9] L&N's procedure of drawing a new design each time a new vector of error terms was drawn is labeled as Experiment C. Experiment D picks a random design and performs 500 Monte Carlo replications with different random components, repeating this procedure 500 times. Experiment D shares the conceptual

[8] Technically it is possible to give each individual their own randomly chosen design, particularly in a computer administered choice experiment. However, this has not been the typical practice in applications in the economics literature, although it is sometimes seen in marketing. There are two serious drawbacks to randomly choosing a design for each individual. The first is that the resulting design matrix may be ill-conditioned and/or have poor efficiency properties. Second, use of this procedure makes it impossible to separate differences between and within individual error variability and, more generally, estimation of models incorporating preference heterogeneity become more difficult, if not impossible, due to the confounding of design assignment and individual preferences. A general discussion of these issues is beyond the scope of this article.

[9] We created random designs following L&N's procedure. We note though that because their procedure allows for the possibility of replicated alternatives and/or choice sets it is not what is usually called a random design in the experimental design literature. This effect appears to be small here as the number of potential designs is sizeable. The three individual designs displayed for experiment B are the first, 250th, and last designs from the suite of 500 designs that comprise Experiment D.

Table 5. Results of Monte Carlo Simulation for Main Effects Versus Random Designs

		Experiment A Main Effects Design		Experiment B One Fixed Random Design: Random Design #1		Experiment B One Fixed Random Design: Random Design #250		Experiment B One Fixed Random Design: Random Design #500		Experiment C 500 Random Designs (L&N Procedure)		Experiment D Average over 500 Random Designs	
		500 MC Iterations		500 MC Iterations		500 MC Iterations		500 MC Iterations		500 MC Iterations		500 $*$ 500 MC Iterations	
	True Value	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
α_1	1	1.03	0.27	1.07	0.29	1.06	0.32	1.06	0.35	1.06	0.30	1.06	0.32
α_2	2	2.06	0.31	2.10	0.37	2.10	0.34	2.11	0.38	2.09	0.36	2.09	0.36
β_1	−1	−1.03	0.18	−1.03	0.21	−1.04	0.19	−1.05	0.22	−1.03	0.21	−1.04	0.21
β_2	2	2.07	0.30	2.08	0.28	2.08	0.28	2.09	0.30	2.08	0.27	2.08	0.28
β_3	3	3.12	0.41	3.12	0.38	3.12	0.37	3.14	0.37	3.12	0.38	3.13	0.38
	True MWTP	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
$= -\beta_2/\beta_1$	2	2.05	0.38	2.07	0.41	2.05	0.35	2.09	0.52	2.08	0.40	2.06	0.40

flaw with Experiment C that averaging over designs that span much of the space is likely to obscure the larger average (absolute) bias associated with a random design. These two experiments differ in what they are implicitly simulating. As noted earlier, Experiment C is consistent with a researcher running 500 different projects each of which uses a different random design and draws its own vector of random components. Experiment D is consistent with 500 researchers each of whom has a particular randomly chosen design repeatedly running a study 500 times.

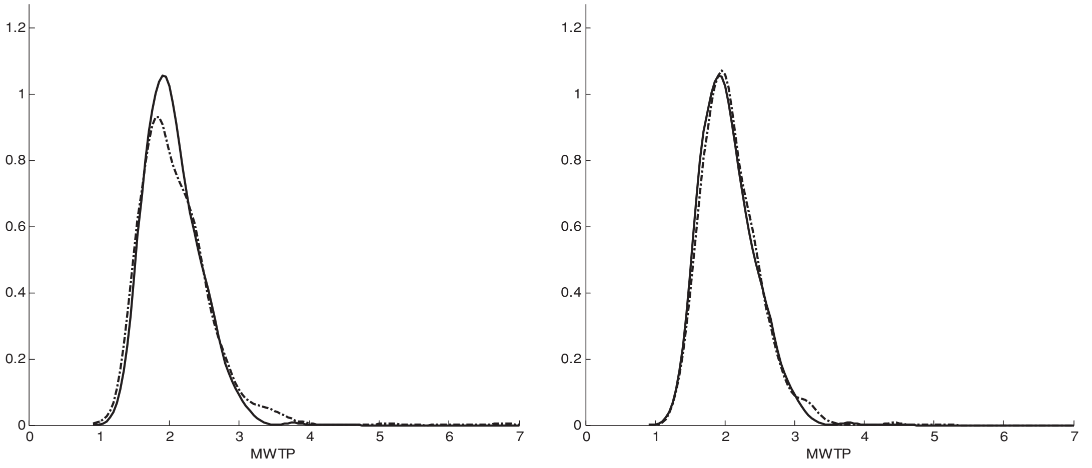
L&N's results suggested that random designs should outperform main effects designs, but this is not typical of our results.[10] First, we compare the three random designs (Experiment B) to the main effects design (Experiment A). These three designs taken from the Experiment D runs are typical of what one might see in practice. The first design is slightly worse than Experiment A in terms of the empirical standard deviations and bias in the parameters and *MWTP* estimates, while the second is slightly better, with the third noticeably worse.

Comparing L&N's approach in Experiment C to Experiment A, the results suggest only a modest increase in the bias and a standard deviation for the key *MWTP* statistic. This certainly obscures the risk of using a random design as the last random design in Experiment B has a standard deviation for *MWTP* that is 36% larger than Experiment A. We plot the density function for the standard main effects design versus a single random design, and versus L&N's procedure in figure 1. This helps make it clear (*a*) how dangerous a random design can be, which is why the random design approach is undesirable; and (*b*) how L&N's procedure obscures the risk of random designs. The most telling statistic is that only 1.4% of the Experiment A iterations yield a *MWTP* estimate > 3 (a 50% increase over the true value), while the comparable percentage for this particular random design and Experiment C are 4.6% and 2.8%, respectively.

Experiment D behaves somewhat better than Experiment C but sends a mixed message about the individual parameter estimates compared with Experiment A. The *MWTP* (average) estimate is closer to its true value than in Experiment C, but the standard deviation is of similar magnitude. Although the results of Experiment D are somewhat more favorable than Experiment C, it is still dominated by Experiment A, with Experiment D having over twice the density for *MWTP* > 3 than Experiment A.

Thus, it is unclear why one would want to choose a random design when a standard main effects design performs better on average, and has a lower risk of adverse outcomes. While one might get lucky with a random design in terms of matching up with true values of the underlying data generating process, this would not seem to be a good bet in expectation terms. That is, one might consider thinking formally about basing an experimental design on an informative prior as a better way to go, but this would involve the classic tradeoff of greater efficiency if the prior is close to being correct, or greater risk if it is not. Moreover, the latter requires much more design expertise, which is

[10] Surprised at getting a different result than L&N, we performed Experiment C several times with different seeds and occasionally found runs where Experiment C performed slightly better than Experiment A but this was not typical. Rose and Scarpa (2008) also find that random designs perform poorly.



Note: The solid line in both panels is the density plot for Experiment A (main effects design). The dash-dot line in the left panel is for the last Experiment B design. The dash-dot line in the right panel is for Experiment C.

Figure 1. MWTP density plots for Experiment A (main effects design), Experiment B, and Experiment C (L&N procedure)

contrary to the thrust of L&N's article, namely to simplify life for applied researchers.

Another rationale for using random designs is a common belief that higher order terms will be identified. This is true *ex ante* in a probabilistic sense before one chooses a particular experiment design and may be true *ex post* if one averages over a large number of random designs like Experiments C and D. However, *ex post* any single randomly chosen design *does not* necessarily identify any particular higher order term. This should be obvious because it almost always will be the case that a randomly chosen design from an allowable set can be a main effects design. Some interactions clearly will not be identified in such designs, while other randomly chosen designs in an allowable set may confound a very specific set of effects. Identification of the parameters of interest in a study is not something that should be left to chance.

Concluding Remarks

An issue in interpreting the results of any Monte Carlo experiment is whether a sufficiently broad range of conditions was covered to generalize so as to give useful guidance to applied researchers. This was not the case with L&N's experiment. In the case of main effects designs, it is easy to find examples where the parameters of interest are not statistically identified in the presence of unobserved but significant interaction effects. While there may be instances where interaction terms are unimportant, there may also be other instances where attribute interactions are a key research focus. Random designs often are very inefficient, and at best provide tenuous identification of key parameters if the IUF of interest is more complex than a main effects specification. Hence, bias can be a serious issue in such designs. It is true that larger sample sizes can compensate for inefficient designs if inefficiency is the only problem associated with a design. However, researchers should be aware of and cautious about potentially large costs associated with sample sizes needed to achieve given degrees of precision. Unfortunately, life for applied researchers who want to design and implement DCEs is unlikely to be as simple as choosing a random design anytime soon.

[Received July 2008;
accepted January 2009.]

References

Ai, C., and E.C. Norton. 2003. "Interaction Terms in Logit and Probit Models." *Economics Letters* 80:123–29.

Box, G.E.P., and N.R. Draper. 1987. *Empirical Model Building and Response Surfaces*. New York: Wiley.

Box, G.E.P., W.G. Hunter, and J.S. Hunter. 1978. *Statistics for Experiments: An Introduction to Design, Data Analysis, and Model Building*. New York: Wiley.

Cochran, W.G., and G.M. Cox. 1992. *Experimental Design*, 2nd ed. New York: Wiley.

Ferrini, S., and R. Scarpa. 2007. "Designs with A Priori Information for Nonmarket Valuation with Choice Experiments: A Monte Carlo Study." *Journal of Environmental Economics and Management* 53:342–63.

Goldberg, I., and J. Roosen. 2007. "Scope Insensitivity in Health Risk Reduction Studies: A Comparison Between Choice Experiments and the Contingent Valuation Method for Valuing Safer Food." *Journal of Risk and Uncertainty* 34:123–44.

Louviere, J.J., D.A. Hensher, and J.D. Swait. 2000. *Stated Choice Methods: Analysis and Applications*. New York: Cambridge University Press.

Lusk, J.L., and F.B. Norwood. 2005. "Effect of Experimental Design on Choice-Based Conjoint Valuation Estimates." *American Journal of Agricultural Economics* 87:771–85.

McClelland, G.H., and C.M. Judd. 1993. "Statistical Difficulties of Detecting Interactions and Moderator Effects." *Psychological Bulletin* 114:376–90.

Rose, J., and R. Scarpa. 2008. "Experimental Designs for Environmental Valuation with Choice Experiments: A Monte Carlo Investigation." *Australian Journal of Agricultural and Resource Economics* 52:253–82.

Street, D.J., and L. Burgess. 2007. *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. New York: Wiley.