

*Incorporating Observed Choice in the Construction of
Welfare Measures from Random Utility Models*

Roger H. von Haefen
Research Economist
Division of Price and Index Number Research
U.S. Bureau of Labor Statistics
Room 3105
2 Massachusetts Ave, NE
Washington, DC 20212-0001
Phone – (202) 691-6593
Fax – (202) 691-6583
Von_Haefen_R@bls.gov
May 10, 2000

Selected Paper, 2000 AAEA Annual Meetings in Tampa, FL

Copyright © 2000 by Roger H. von Haefen. All rights reserved. Reader may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies. Please do not cite. Updated versions of this paper may be obtained from the author upon request.

I. Introduction

Traditional approaches to constructing Hicksian welfare measures from random utility models employ the unconditional distribution for the unobserved determinants of choice along with the specified structure of preferences and the relevant observable individual and commodity specific characteristics (e.g., Small and Rosen [1981], Hanemann [1981]). This paper proposes an alternative strategy employing the *conditional* distribution for the unobserved determinants of choice that incorporates the implications of an individual's observed choice. The traditional unconditional and proposed conditional welfare measures can be motivated by alternative assumptions about the factors that give rise to randomness in probabilistic choice models. These differing interpretations can imply welfare measures that diverge significantly for a chosen individual. However, these differences tend to cancel out as one aggregates across a random sample from the target population if the data generating process is correctly specified. As a result, one would expect the weighted sample means of unconditional and conditional welfare estimates to be roughly equivalent. In applied work, however, data limitations and/or the restrictive econometric features of commonly used empirical models imply that some form of model misspecification is likely present. As a result, a comparison of the traditional unconditional and proposed conditional welfare estimates can serve as a metric for evaluating the degree to which model misspecification impacts welfare measurement.

A subsample of outdoor recreators from the 1994 National Survey of Recreation and the Environment (NSRE) is used to empirically assess the proposed approach to welfare measurement. The seasonal trip demands for 157 residents of the Lower Susquehanna River Basin are examined within a repeated discrete-continuous modeling framework developed in von Haefen [1999]. The framework assumes that the individual's seasonal demand can be decomposed into separable choice occasions. On each choice occasion, the individual makes a discrete choice of whether and which site to visit as well as a quasi-continuous choice of the number of trips to take to a chosen site. The unconditional and conditional seasonal welfare estimates for the loss of a 40 mile reach of the Lower Susquehanna River diverge by roughly \$3.50 to \$5.00 (1994 dollars), but the difference between the two estimates for the loss of Raystown Lake are substantially larger. Although a single influential observation explains much of the divergence between the conditional and unconditional estimates for the latter scenario, large differences still exist for the remaining observations. This finding implies that the specified model fails to account for unique attributes of Raystown Lake. Similar welfare estimates for the two policy scenarios arise from a repeated discrete choice specification (e.g., Morey, Rowe, and Watson [1993]). Together, these findings point to a general result - if the specified model fails to incorporate important characteristics of the objects of choice, significant differences between unconditional and conditional welfare measures can arise.

The paper is organized as follows. Section II lays out the theory of the conditional approach to welfare measurement and discusses its economic and statistical properties in relation to traditional approaches. Section III describes the Lower

Susquehanna recreation data set used in the empirical analysis, and Section IV describes the repeated discrete-continuous specification used to model consumer choice. Section V then discusses the two welfare scenarios considered. Section VI reports the parameter and welfare estimates, and Section VII concludes.

II. Theory

This section clarifies the connections and distinctions between the traditional unconditional approach to welfare measurement developed by Small and Rosen [1981] and Hanemann [1981] and the proposed conditional approach. Although the unconditional and conditional approaches can be applied to all choice models employing the random utility hypothesis, this section employs a discrete-continuous framework first suggested by Chiang and Lee [1992] to structure the discussion. In addition to simplifying exposition, focusing on this specific choice scenario is natural given the empirical application that follows.

Individual i is assumed to have preferences over $N + 1$ commodities that can be represented by the following strictly increasing, strictly quasi-concave, and twice-continuously differentiable utility function:

$$U \left[\sum_{k=1}^N \psi(\mathbf{q}_k, \varepsilon_{ik}) x_{ik}, \tau(\mathbf{w}_i, \varepsilon_{i0}), z_i \right] \quad (1)$$

where x_{ik} is the k th of N related but quality differentiated goods, z_i is a Hicksian composite good, $\psi(\mathbf{q}_k, \varepsilon_{ik})$ indexes the k th good's quality attributes included in the vector \mathbf{q}_k and the scalar ε_{ik} , and $\tau(\mathbf{w}_i, \varepsilon_{i0})$ indexes individual specific characteristics in

the vector \mathbf{w}_i and scalar ε_{i0} . Although the individual is assumed to know all factors that enter $\psi(\mathbf{q}_k, \varepsilon_{ik})$ and $\tau(\mathbf{w}_i, \varepsilon_{i0})$, the analyst does not observe the ε_{ik} 's ($k=0, \dots, N$) and treats them as random variables from her perspective. As a result, the above specification is consistent with the random utility hypothesis (McFadden [1974a, 1974b]).

The structure of preferences in (1) suggests that the N quality differentiated commodities are perfect substitutes, and the rational individual will consume only the good with the lowest quality adjusted price, $p_j / \psi(\mathbf{q}_j, \varepsilon_{ij})$. If the N quality differentiated commodities are nonessential, however, it is possible that the individual will choose not to consume any of them. Following Hicks [1940], this outcome would arise if the virtual price, ξ_i^* , i.e., the marginal willingness to pay for the right to consume the N goods, is less than all N quality adjusted prices:

$$\xi_i^* < \frac{P_k}{\psi(\mathbf{q}_k, \varepsilon_{ik})}, \quad \forall k \in 1, \dots, N \quad (2)$$

Conversely, the individual will choose to consume commodity j if its quality adjusted price is the lowest among all N commodities and less than the virtual price, i.e.:

$$\frac{P_j}{\psi(\mathbf{q}_j, \varepsilon_{ij})} < \frac{P_k}{\psi(\mathbf{q}_k, \varepsilon_{ik})}, \quad k \in 1, \dots, N \quad \& \quad \frac{P_j}{\psi(\mathbf{q}_j, \varepsilon_{ij})} < \xi_i^* \quad (3)$$

Conditional on the individual choosing commodity j , one can solve for the derived demand for good j by solving the following Lagrangian:

$$L = U(\psi(\mathbf{q}_j, \varepsilon_{ij})x_{ij}, \tau(\mathbf{w}_i, \varepsilon_{i0}), z_i) + \lambda(y - p_j x_{ij} - p_z z_i) \quad (4)$$

where y is income and p_j and p_z are prices. The optimal consumptive solutions for x_{ij}^* takes the general form:

$$x_{ij}^* = \frac{1}{\psi_j(\mathbf{q}_j, \boldsymbol{\varepsilon}_{ij})} x \left[\frac{p_j}{\psi_j(\mathbf{q}_j, \boldsymbol{\varepsilon}_{ij})}, \boldsymbol{\tau}(\mathbf{w}_i, \boldsymbol{\varepsilon}_{i0}), p_z, y \right] \quad (5)$$

Plugging (6) and the optimal solution for z_i^* into the conditional direct utility function allows the analyst to solve for the conditional indirect utility function that takes the general form:

$$V_{ij} = V \left[\frac{p_j}{\psi(\mathbf{q}_j, \boldsymbol{\varepsilon}_{ij})}, \boldsymbol{\tau}(\mathbf{w}_i, \boldsymbol{\varepsilon}_{i0}), p_z, y \right] \quad (6)$$

Alternatively, if the individual chooses not to consume any of the N commodities, the virtual price replace the quality adjusted price in (6) and the conditional indirect utility function takes the general form:

$$V_{i0} = V(\xi_i^*, \boldsymbol{\tau}(\mathbf{w}_i, \boldsymbol{\varepsilon}_{i0}), p_z, y) \quad (7)$$

The rational individual will choose the alternative that maximizes her utility. Therefore, her unconditional indirect utility function takes the form:

$$V_i = \text{Max}_j \{V_j\}, j \in 0, \dots, N \quad (8)$$

Frequently applied analysts are interested in using random utility models such as the discrete-continuous framework described above to estimate the compensating surplus arising from a change in the observable quality attributes for one or a set of commodities. Individual i 's compensating surplus (CS_i) associated with an improvement in the quality characteristics of the N commodities from \mathbf{q}' to \mathbf{q}'' is implicitly defined as:

$$V_i(\mathbf{q}', y) = V_i(\mathbf{q}'', y - CS_i) \quad (9)$$

The above equations suggest that CS_i is in general a function of the ε_{ik} 's which are known only to the individual. As a result, CS_i is a random variable from the analyst's perspective which cannot be determined precisely. However, the analyst in general

knows or assumes the distribution for the unobserved determinants of choice. This information, along with the assumed structure of preferences and the observable good and individual specific characteristics, allows the analyst to construct a measure of the central tendency of the distribution of CS_i , such as the mean.¹

It is now possible to clearly delineate the traditional unconditional approach to welfare measurement from the proposed conditional approach. Although both approaches employ the assumed structure of preferences in (6)-(8), the traditional approach employs additional information about the unobserved determinants of choice implied by an individual's observed choices. In microeconomic applications, the analyst observes whether and which of the N quality differentiated commodities the individual consumes from the current menu of available goods. For chosen goods, the analyst also observes the quantity consumed. These observed choices, along with the inequalities in (2) and (3) as well as the equality in (5), imply restrictions on the support of the distribution of the unobserved determinants of choice that can be used in the construction of the compensating surplus estimates. The conditional approach to welfare measurement incorporates these additional restrictions while the traditional unconditional approach does not. That is, the proposed approach employs the conditional distribution for the unobserved determinants of choice that incorporates (2), (3) and (5), while the traditional approach employs the unconditional distribution for the unobserved determinants of choice.

The conditional and unconditional approaches to welfare measurement can be rationalized by alternative sets of assumptions about the factors that give rise to

¹ In some cases, the analyst may prefer an alternative summary measure such as the distribution's mode, median or some other percentile. In the discussion that follows, however, it will be assumed that the

randomness in probabilistic choice models. As stated above, the random utility hypothesis asserts that although the ε_{ik} 's are unobserved and random from the analyst's perspective, they are known to the individual and influence her choices. As suggested by Hausman and Wise [1978], one can interpret the ε_{ik} 's as arising from the "random firing of neurons" (p 407) or ephemeral factors such as one's state of mind at a particular point in time. Under this interpretation, the information about the ε_{ik} 's conveyed by the individual's observed choices would be largely uninformative about the individual's underlying preferences, leading the analyst to prefer the traditional unconditional approach to welfare measurement. Alternatively, one can interpret the randomness as arising from important unobserved commodity and/or individual specific characteristics that are not otherwise captured in the individual or commodity specific indexes. If the analyst believes that this unobserved commodity/individual specific heterogeneity would persist if the consumer faced the choice again, she would then prefer the conditional approach to welfare measurement.

Although the two approaches are conceptually different, the law of iterated expectations implies that they have a close statistical relationship. In particular, one can write the unconditional expected compensating surplus (i.e., $E(CS_i)$) for individual i as:

$$E(CS_i) = E_{C_i}(E(CS_i | C_i)) \quad (10)$$

where $E(CS_i | C_i)$ is the conditional expected compensating surplus and $E_{C_i}(\cdot)$ is the expectation operator with respect to the vector of observed individual choices, C_i . In words, equation (10) states that the expectation of the conditional expected compensating surplus is the unconditional expected compensating surplus. This relationship implies

analyst wishes to focus on the distribution's mean.

that the difference between $E(CS_i)$ and $E(CS_i | C_i)$ can be thought of as a random variable (call it a_i) with an expectation of zero (i.e., $E(a_i) = 0$). Although one might find differences between each individual's unconditional and conditional expected compensating surpluses, these differences should in some sense cancel out as one averages across a sample of individuals drawn from the target population. That is, if the difference between $E(CS_i)$ and $E(CS_i | C_i)$ is finite for each individual in the target population, the law of large numbers implies that the unconditional and conditional expected compensating surplus estimates for a random sample of individuals will converge in expectation. In other words, it should not matter whether the analyst employs the conditional or the unconditional expected compensating surplus estimates if the analyst is interested in using the sample to derive inference about the population's expected compensating surplus.

Two important qualifications should be made with respect to this last assertion. Frequently in applied work, the sample size used in estimation and the construction of welfare estimates is quite small. In these cases, the difference between the sample's estimated compensating surplus using the conditional and the unconditional approaches can be substantial at least in terms of its policy implications. Additionally, an important assumption underlying the assertion that a sample's average of unconditional and conditional welfare measures should be roughly equal is the assumption that the analyst has correctly specified the data generating process for the sample's observed choices. If the analyst has, for example, excluded important site or individual specific characteristics, substantial differences between the unconditional and conditional welfare measures may result. This point suggests that if the sample size is sufficiently large, a

comparison of the unconditional and conditional welfare measures can serve as a check of the structural model's ability to fit the data.

III. Data

Having laid out the economic and statistical relationships between the unconditional and conditional approaches to welfare measurement in the previous section, this section briefly describes the recreation data set used to empirically evaluate the two approaches. The interested reader should consult von Haefen [1999] for a more detailed discussion of the data construction. The recreation data comes from the 1994 National Survey of Recreation and the Environment (NSRE) conducted by the Economic Research Service.² 2,734 trips by the 161 sample members residing in the Lower Susquehanna River Basin who participated in water-based outdoor recreation during the past year are the focus of the analysis. These individuals visited a total of 219 geographically distinct lakes, rivers, and streams. Visited waterbodies were aggregated into 89 distinct recreation sites using an algorithm that exploited the region's natural watershed boundaries. Water chemistry variables collected by the EPA, Susquehanna River Basin Commission, Army Corps of Engineers, and Pennsylvania Fish and Boat Commission were attached to each of these defined sites using a similar watershed-based algorithm. These chemistries were used to construct two environmental variables that were used to describe the degree to which eutrophication – a significant environmental problem in the region (Frey et al. [1996]) - was a source of impairment at the 89 defined recreation sites. The first, *Lowdo*, is an indicator variable for whether water surface

dissolved oxygen levels fell below the EPA determined threshold for impairment of 6.5 mg/l for cold water fisheries and 5.5 mg/l for warm water fisheries. The second, TSI, is Carlson's [1977] Trophic State Index (TSI), which is a proxy measure for eutrophication in a waterbody based on phosphorus and secchi disk readings. Indicator variables for whether the site was located within or adjacent to a park (Park) or along the Susquehanna River (Susq) were also used as site characteristics.

For each individual in the sample and every defined recreation site, round trip travel distances from each recreator's home zip code to all 89 sites were estimated using the program *PCMiler* (Alt Associates [1997]). Travel costs estimates were estimated as the sum of the travel distance multiplied by \$0.30 per mile plus the travel time valued at one-third of each individual's wage rate. For each individual, an estimate of seasonal full income was also constructed as the sum of the individual's share of wage and non-wage family income plus her leisure time valued at her opportunity cost of time (wage). Dummy variables for whether the individual was female (Female) and/or participated in boating, fishing, or swimming in the past year (Water) were also included as individual specific shift variables in the econometric model.

IV. Econometric Specification

Most of the recreators in the Lower Susquehanna data set visited multiple sites. To allow for this type of behavior within the discrete-continuous framework described in Section II, it is necessary to assume that the recreation season can be decomposed into a series of separable choice occasions. On each choice occasion, the individual is assumed

² Thanks go to Peter Feather and Dan Hellerstein at the ERS for making this data set publicly available.

to make a discrete choice of whether and which recreation site to visit and a conditional quasi-continuous choice of the number of trips to take to the chosen site. This repeated discrete-continuous structure is conceptually similar to the repeated discrete choice model (e.g., Morey, Shaw, and Rowe [1993]) in that the recreation season is decomposed into separable choice occasions. However, it can be distinguished from the repeated discrete choice approach because it does not restrictively assume that the decisions of whether to visit a site and the number of trips to the site are uncoordinated (Bockstael, Hanemann, and Kling [1987]).

Two related issues arise with the proposed repeated discrete-continuous structure: 1) the number of choice occasions; and 2) the specification of how the individual allocates her seasonal income to each separable choice occasion and how the allocation changes with changes in environmental quality. These same issues arise with the repeated discrete choice model.³ von Haefen [1999] develops a conceptual framework for thinking about these issues, but the information necessary to implement his framework is unfortunately absent in this study. As a result, it is assumed that each individual faces ten choice occasions, the individual allocates her seasonal income evenly across them, and this income allocation is unaltered with changes in environmental quality.⁴

³ The problem of determining the optimal income allocation is conveniently avoided in the repeated discrete choice framework if the analyst assumes that the marginal utility of income is constant on each choice occasion. A similar simplifying assumption is not available with the repeated discrete-continuous model.

⁴ von Haefen [1999] finds qualitatively similar welfare results to those reported in the subsequent sections with models employing alternative choice occasion specifications.

On choice occasion t , individual i 's conditional preferences for the chosen site j can be represented by the following Homothetic Indirect Translog specification (Chiang and Lee [1992]):⁵

$$V_{ij} = \ln y_{it} + \alpha(\ln p_{ij} - \ln \psi_{ij}) - (1 + \alpha)(\ln p_{iz} - \ln \tau_{it}) + \frac{1}{2} \beta (\ln p_{ij} - \ln \psi_{ij})^2 + \frac{1}{2} \beta (\ln p_{iz} - \ln \tau_{it})^2 - \beta (\ln p_{ij} - \ln \psi_{ij})(\ln p_{iz} - \ln \tau_{it}) \quad (11)$$

where α and β are estimable structural parameters and the individual and commodity specific indexes take the following log-linear form:

$$\begin{aligned} \ln \tau_{it} &= \mathbf{w}_i \boldsymbol{\delta} + \varepsilon_{it0} \\ \ln \psi_{ij} &= \mathbf{q}_j \boldsymbol{\gamma} + \varepsilon_{ij} \end{aligned} \quad (12)$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are vectors of structural parameters. Applying Roy's identity to (11) allows one to derive the conditional choice occasion expenditure share for site j that takes the form:

$$s_{ij} = -\alpha - \beta(\ln p_{ij} - \ln \psi_{ij} - \ln p_{iz} + \ln \tau_{it}) \quad (13)$$

As described in Chaing and Lee, one can set (13) equal to zero and solve for the implied virtual price of recreation on choice occasion t , i.e.:

$$\xi_{it} = \exp(-\alpha / \beta) p_{iz} / \tau_{it} \quad (14)$$

Because (11) is a flexible functional form, the analyst must verify that (11) is quasi-convex in prices in an open neighborhood of the relevant prices. The Homothetic Indirect Translog specification satisfies this restriction if the following restriction is satisfied on each choice occasion, i.e.:

$$s_{ij}(1 - s_{ij}) + \beta \geq 0 \quad (15)$$

⁵ von Haefen [1999] develops additional preference specification that fit within the discrete-continuous framework. These specifications were found to generate qualitatively similar results and are not reported here.

If one assumes that the ε_{ik} 's can be treated as independent and identically distributed draws from the Type I extreme value distribution, von Haefen [1999] has shown that the above model can be linked to a closed form likelihood function. In these cases, one can estimate the structural parameters by standard maximum likelihood techniques.

Because the repeated discrete-continuous model represents a novel approach to modeling seasonal recreation demand, a standard repeated discrete choice model is also estimated for comparative purposes. The specification employed assumes 365 choice occasions, a constant marginal utility of income, and an i.i.d. Type I Extreme Value distribution for the unobserved determinants of choice.

V. *Policy Scenarios*

Two policy scenarios are used in this study to compare the traditional unconditional and the proposed conditional welfare measures. The first involves the loss of a 40 mile reach of the Lower Susquehanna River from Columbia, PA, to the river's mouth at Havre de Grace, MD. Although water quality chemistries from 1994 suggest that the reach was not eutrophic, a significant increase in phosphorus and nitrogen loadings into the shallow and slow moving reach could overrun its natural assimilative abilities and result in the loss of subsurface dissolved oxygen, much flora and fauna life, and widespread algal blooms at the water surface. At present, the 40 mile reach is a widely used recreational resource with three state parks capable of supporting boating, fishing, swimming, hiking, nature viewing, and camping.

The second policy scenario involves the loss of Raystown Lake, the largest standing waterbody in central Pennsylvania. The lake is bordered by Trough Creek State Park, a large recreation facility that include beaches, boat launches, campgrounds, trails, and stocked fisheries. Raystown Lake is the most frequently visited recreation site by the sample of recreators and one of the most valuable water resources for outdoor recreation in the region. 1994 water quality chemistries suggest that the lake was not eutrophic.

VI. *Parameter & Welfare Estimates*

The first column of Table 1 reports parameter estimates and asymptotic z-statistics for the repeated discrete-continuous model employing the Homothetic Indirect Translog specification. All reported point estimates are statistically significant and have signs consistent with *a priori* expectations. The Lowdo water quality variable is found to negatively impact a site's probability of selection and derived trip demand. The quadratic specification of the TSI variable suggests that higher Trophic State Index levels decrease utility at an increasing rate. Moreover, the finding that the β structural parameter estimate is positive and highly significant implies that the economic consistency restriction in (15) is satisfied for all share values on the [0,1] interval.

The second column of Table 1 reports qualitatively similar estimates for the repeated discrete choice specification. Like the repeated discrete-continuous estimates, all point estimates are statistically significant and plausibly signed.

The parameter estimates from Table 1 were used to construct unconditional and conditional population welfare estimates for the two policy scenarios described in the

previous section. As discussed in von Haefen [1999], there is no closed form solution for either the unconditional or conditional expected compensating surplus arising from a change in quality or site access for the repeated discrete-continuous model. As a result, simulation techniques were required. The following algorithm was used to construct these estimates:

- Using a pseudo-random number generator, the Probability Integral Transformation and the unconditional and conditional Type I Extreme Value distributions for the unobserved determinants of choice, simulate separate vectors of ε_{ik} 's for all 10 choice occasions for each sample respondent (see Appendix A in von Haefen [1999] for further details).
- Using both simulated vectors and the preference specification in (11), construct each sample member's simulated unconditional and conditional compensating surplus for each policy scenarios.
- Using the sample weights, construct simulated population estimates of the compensating surplus associated with each policy scenario.
- Replicate these steps T times. The average across the T simulated unconditional and conditional population estimates are estimates of the population's unconditional and conditional compensating surplus for each policy scenario, respectively.

For the welfare results reported in this paper, experimentation with alternative values of T suggested that $T = 500$ resulted in population welfare estimates that were accurate to within \$0.03. For the repeated discrete choice specifications, a closed form solution exists for the unconditional expected compensating surplus (Small and Rosen [1981], Hanemann [1981]), but a simulation algorithm analogous to the one described above was necessary for the conditional welfare estimates.

Table 2 reports sample mean estimates for the seasonal compensating surplus arising from the loss of the 40 mile reach of the Lower Susquehanna River and Raystown Lake. Beginning with the Lower Susquehanna River results first, one finds that all reported welfare measures are of the same order of magnitude, but the estimates differ by roughly \$3.50 to \$5.00 across the conditional and unconditional welfare measures. Considering that there are an estimated 1.1 million recreators in the 25 county region of the Lower Susquehanna River Basin,⁶ these differences translate into \$3.8-\$5.5 million in aggregate welfare loss and are nontrivial from a policy perspective.

Turning to the loss of Raystown Lake scenario, one finds substantially larger differences between the unconditional and conditional welfare estimates for both the repeated discrete-continuous and repeated discrete choice models. There are also pronounced differences in conditional welfare estimates across the repeated discrete-continuous and repeated discrete choice models. Two factors help to explain these large discrepancies. Beginning with the differences across the conditional welfare estimates, a closer examination of the Lower Susquehanna recreation data set reveals that one individual in the sample took 40 trips to Raystown Lake at an estimated travel cost of \$49.51. The total expenditures involved in these trips represented a substantially larger share of his full income than any other recreator/site combination in the sample. These facts suggest that the individual is an outlier in a statistical sense. Indeed, the repeated discrete-continuous model implies that the individual's conditional expected compensating surplus is \$4,065, a sum eight times larger than any other recreator's in the sample. For the repeated discrete-choice model, the individual's conditional expected

⁶ This estimate was generated based on the number of outdoor recreators in the full Lower Susquehanna subsample (378 respondents total) of the NSRE.

compensating surplus is approximately \$358, a large but comparable sum relative to other visitors of Raystown Lake. These findings suggest that conditional welfare measures from the repeated discrete-continuous model are considerably more sensitive to statistical outliers than the repeated discrete choice model. As a result, a more robust estimate of the central tendency of the sample's average compensating surplus for the loss of Raystown Lake may be appropriate. Table 2 also reports estimates of the sample's 1% α -trimmed mean conditional compensating surplus for the repeated discrete-continuous specification that essentially drops the outlier observation. This trimmed mean estimate is considerably smaller than the untrimmed estimate. However, the trimmed estimate nevertheless has substantively different policy implications than all estimates from the repeated discrete choice models.

The above discussion does not explain the large differences between the unconditional and conditional welfare measures. As noted in the theory section, such an empirical finding would suggest that model misspecification is present. Given the policy scenario's focus on a single site, the disparity between the unconditional and conditional welfare estimates suggests that important site attributes of Raystown Lake are not being captured. To evaluate this possibility, Table 3 reports parameter estimates from repeated discrete-continuous and repeated discrete choice models with dummy variables for Raystown Lake included. For both sets of results, the Raystown Lake dummy variables are positive and highly statistically significant while the remaining parameter estimates are qualitatively similar to those reported in Table 1. Table 4 reports unconditional and conditional welfare estimates for the loss of Raystown Lake from both models reported in Table 3. Compared to the results reported in Table 2, the gap between the unconditional

and conditional welfare measures is substantially narrower. This finding confirms the theoretical assertion in Section II that substantial differences between unconditional and conditional welfare estimates may arise if important site or individual specific characteristics are not captured in the econometric model.

An additional point that can be raised by a comparison of Tables 2 and 4 is the surprising robustness of the conditional welfare estimates to the exclusion or inclusion of the Raystown Lake dummy variable. This finding suggests that welfare measures that condition on an individual's observed choice may be more reliable estimates of the unknown population compensating surplus when some model misspecification exists.

VII. Conclusion

This paper has proposed an alternative approach to welfare measurement from random utility models that, in contrast to traditional approaches, incorporates observed choice in the construction of welfare estimates. Section II argued that the traditional and proposed approaches can be motivated by alternative assumptions about the sources of randomness in probabilistic choice models, but that both approaches generate sample welfare estimates that converge in expectation as the sample size grows if the model is correctly specified. The empirical application with the Lower Susquehanna River Basin data set reported in Sections III, IV, and V suggested that in applied situations, significant differences between unconditional and conditional welfare measures can arise when important site attributes are not captured in the empirical model. The empirical results

suggest that in these cases the conditional welfare measures may be a more reliable estimate of the population compensating surplus.

Selected References

- Alk Associates [1997]: *PC Miler User Guide*, Princeton, NJ: Alk Associates.
- Bockstael, Nancy, W. Michael Hanemann, and Catherine, Kling [1987]: "Estimating the Value of Water Quality Improvements in a Recreational Demand Framework," *Water Resources Research*, 23(5): 951-960.
- Carlson, R. [1977]: "A Trophic State Index for Lakes," *Limnological Oceanography*, 22: 361-369.
- Chiang, Jeongwen, and Lung-Fei Lee [1992]: "Discrete/Continuous Models of Consumer Demand with Binding Non-Negativity Constraints," *Journal of Econometrics*, 54(1-3): 79-93.
- Frey, Robert et al. [1996]: *Commonwealth of Pennsylvania 1996 Water Quality Assessment*. Report Submitted in Compliance with Section 305(b), Federal Clean Water Act.
- Hanemann, W. Michael [1981]: "Applied Welfare Analysis with Quantal Choice Models," Working Paper No. 173, Division of Agricultural Sciences, University of California, Berkeley.
- Hausman, Jerry [1978]: "Specification Tests in Econometrics," *Econometrica*, 46(6): 1251-1271.
- Hausman, Jerry, and David Wise [1978]: "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46(2): 403-426.
- Hicks, John [1940]: "The Valuation of the Social Income," *Economica*, 7: 105-126.
- McFadden, Daniel [1974a]: "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, Paul Zarembka, Editor. New York: Academic Press.
- McFadden, Daniel [1974b]: "The Measurement of Urban Travel Demand," *Journal of Public Economics*, 3(4): 303-328.
- Morey, Edward, Robert Rowe, and Michael Watson [1993]: "A Repeated Nested-Logit Model of Atlantic Salmon Fishing," *American Journal of Agricultural Economics*, 75(3): 578-592.
- Small, Kenneth, and Harvey Rosen [1981]: "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49(1): 105-130.
- von Haefen, Roger [1999]: *Valuing Environmental Quality in a Repeated Discrete-Continuous Framework*. Ph.D. Dissertation, Department of Economics, Duke University.

Table 1
Parameter Estimates from Repeated Discrete-Continuous & Repeated Discrete Choice Models

<i>Parameters</i>	<i>Repeated Discrete-Continuous Model</i>	<i>Repeated Discrete Choice Model</i>
Log-Likelihood	-904.86	-15,336.01
α	.01313 (2.707)	-
β	.03468 (19.282)	-
Price	-	-.1306 (65.600)
Water	-.4010 (-3.385)	.7477 (13.401)
Female	.2645 (3.869)	.9722 (20.511)
Lowdo	-.4579 (-3.100)	-2.234 (-11.702)
TSI	.04677 (6.978)	.0846 (16.019)
TSI ²	-7.378e-4 (-5.742)	-1.718e-3 (-16.169)
Park	.6329 (8.444)	.5119 (8.685)
Susq	.5057 (5.710)	.2310 (4.330)
μ (Scale Parameter)	.5710 (39.800)	-
Dummy Variable for Non- Recreation Alternative	-	2.502 (37.290)

¹ Asymptotic z-statistics in parentheses

Table 2
Weighted Sample Mean Welfare Estimates

	<i>Repeated Discrete-Continuous Model</i>	<i>Repeated Discrete Choice Model</i>
Loss of 40 mile reach of the Lower Susquehanna River		
Unconditional Weighted Sample Mean	-\$17.55	-\$10.30
Conditional Weighted Sample Mean	-\$12.43	-\$13.84
Loss of Raystown Lake		
Unconditional Weighted Sample Mean	-\$1.39	-\$1.20
Conditional Weighted Sample Mean	-\$39.18	-\$6.16
Conditional 1% α -trimmed Weighted Sample Mean	-\$16.25	-

Table 3
Parameter Estimates from Repeated Discrete-Continuous & Repeated Discrete Choice Models
Raystown Lake Dummy Variable Included

<i>Parameters</i>	<i>Repeated Discrete-Continuous Model</i>	<i>Repeated Discrete Choice Model</i>
Log-Likelihood	-870.15	-15,201.90
α	.01262 (2.586)	-
β	.03514 (19.290)	-
Price	-	-.1333 (-65.411)
Water	-.3896 (-3.318)	1.025 (21.396)
Female	.2711 (4.010)	1.767 (18.481)
Lowdo	-.3843 (-2.877)	-2.043 (-10.967)
TSI	.02778 (4.020)	.04921 (8.765)
TSI ²	-3.803e-4 (-2.877)	-1.039e-3 (-9.286)
Park	.7336 (9.919)	.6409 (11.077)
Susq	.4825 (5.608)	.2347 (4.409)
μ (Scale Parameter)	.5630 (39.667)	-
Dummy for Non-Recreation Alternative	-	2.499 (36.956)
Raystown Lake Dummy	1.198 (9.745)	.7795 (13.867)

¹ Asymptotic z-statistics in parentheses

Table 4
Weighted Sample Mean Welfare Estimates
Employing Models with Raystown Lake Dummy Variable

	<i>Repeated Discrete-Continuous Model</i>	<i>Repeated Discrete Choice Model</i>
Loss of Raystown Lake		
Unconditional Weighted Sample Mean	-\$7.67	-\$4.67
Conditional Weighted Sample Mean	-\$39.32	-\$6.18
Conditional 1% α -trimmed Weighted Sample Mean	-\$16.35	-