

Staff Papers Series

Staff Paper P88-44

November 1988

On Approximations to Expected Consumer Surplus

By

Theodore Graham-Tomasi

Wiktor L. Adamowicz

Jerald J. Fletcher



Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

On Approximations to Expected Consumer Surplus *

By

Theodore Graham-Tomasi

Wiktor L. Adamowicz

Jerald J. Fletcher**

November, 1988

* Minnesota Agricultural Experiment Station Publication No. 16476

** Respectively, Assistant Professor, Department of Agricultural and Applied Economics, University of Minnesota, Assistant Professor, Department of Rural Economy, University of Alberta, and Assistant Professor, Department of Agricultural Economics, Purdue University.

Staff Papers are published without formal review within the Department of Agricultural and Applied Economics.

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, religion, color, sex, national origin, handicap, age, veteran status, or sexual orientation.

I. Introduction

Consumer surplus and other welfare measures calculated from demand curves are random variables. This is so because these measures are functions of estimated (as opposed to known) demand parameters, which, of course, are random variables. Increasingly, this realization has been incorporated into studies which assess such benefit measures. It is typical in these studies that the expected value of the benefit measure is employed, in keeping with the common practice of assuming that benefit-cost analysts are risk neutral.¹

Unfortunately, consumer surplus measures usually involve the ratio of random variables, and the expected value of a ratio of random variables is not equal to the ratio of the expected values. The use of the ratio of the expected values leads to a biased estimate of the true measure, although it is a consistent estimator. Indeed, in small samples the expectation of consumer surplus often does not have a closed-form representation. In this case, one must resort to an approximation or a cumbersome Monte Carlo analysis. The former is the tack most often taken; for example, Bockstael and Strand and Kealy and Bishop have used a second-order Taylor series approximation to expected consumer surplus in their investigations.

The purpose of this note is to point out that this second-order approximation may not be very accurate, especially when the variance of the denominator of the ratio (typically an estimated demand parameter on a price variable) is relatively large and/or its mean is relatively small. This is because these statistics figure prominently in higher-order terms

of the approximation. We investigate this issue via a Monte Carlo analysis which (with a large enough number of trials) is able to give a direct estimate of the mean of the consumer surplus measure. Comparing this to the second-order approximation that has appeared in the literature shows that the magnitude of the higher-order terms can be substantial. The magnitude of this error due to truncation of the approximation varies markedly in our example across functional forms for the demand function. In this example the error is especially large for the popular linear functional form and is evident to a lesser extent in the semi-log form.

Our illustration of this phenomenon is conducted using consumer surplus estimates for a travel cost model of recreation demand. We investigate three functional forms for a model with the number of visits in a season regressed on a constant and travel cost. The recreational activity is hunting for Bighorn Sheep in Alberta, Canada. The exact model used in our analysis is quite simple; we do not intend to offer a convincing model of Bighorn sheep hunting as much as we wish to illustrate the size of the potential errors that can be made from truncating an approximation to expected consumer surplus at the second order.

II. Expected Consumer Surplus

The most commonly used functional forms for recreation demand functions are the linear and the semi-log. In our analysis we estimate both of these forms as well as the double-log. For the simple, one-variable demand equations we employ, these forms are:

- (1) Linear $Q = a + bP$
- (2) Semi-log $\ln Q = a + bP$
- (3) Double-log $\ln Q = a + b \ln P$

As Bockstael and Strand point out, the derivation of the consumer surplus functions for these forms requires an assumption regarding the source of error in the demand equation. We study the consumer surplus functions based on an assumption of "omitted variables", using the actual quantities for each observation in the surplus function. The alternative is to use the predicted quantity, which would be appropriate if the equation error is due to errors in measurement. The consumer surplus functions are

- (4) Linear $CS = Q^2/(-b^2)$
- (5) Semi-log $CS = Q/(-b^2)$
- (6) Double-log $CS = \begin{matrix} -PQ/(b + 1) & \text{if } b < -1 \\ (1/(b + 1))(\text{Max}(P)\text{Max}(Q) - PQ) & \text{if } b > -1 \end{matrix}$

The quantity used in these equations is the actual quantity, while the price P in (6) is the price corresponding to the sample average quantity. $\text{Max}(P)$ and $\text{Max}(Q)$ are the maximum price observed over the sample and the corresponding quantity; this is a practice commonly employed to render finite the consumer surplus for the double-log form.²

As can be seen in (4)-(6), the consumer surplus for these functional forms involve demand parameters in the denominator. If the analyst wishes to use the expected value of the consumer surplus functions, a closed-form

representation of the expected value is not available. As noted in the introduction, several investigators (e.g. Bockstael and Strand; Kealy and Bishop) have proposed using a Taylor series approximation to expected surplus. This requires that the random variables are continuous and their ratio has finite moments of all orders. The second-order approximation employed in the literature is given by

$$(7) \quad E[x/y] = E[x]/E[y] - \text{COV}[x,y]/E[y]^2 + E[x]\text{VAR}[y]/E[y]^3.$$

This expression involves error due to truncation of the approximation. For example, the third-order term is

$$(8) \quad 2\text{VAR}[y]E[x]/E[y]^3 - E[x]\text{COV}[y,y^2]/E[y]^4 + (\text{COV}[x,y^2] - 2E[y]\text{COV}[x,y])/E[y]^3.$$

It is apparent that, if the price parameter in the demand function is fairly small, the third and higher order terms can be fairly large even if the price variable has a "significant" t statistic for a test with a null hypothesis of zero price effect. While the fit of the equation in terms of the t statistic on the price is related to the higher order terms, we do not know at what level of t these terms reasonably may be ignored. Moreover, this level will depend on the functional form of the demand equation.

In the case of consumer surplus measures most analysts have assumed that the numerator is nonrandom, in which case terms involving the covariances

of x and y and x and y^2 drop out.³ If the error terms are due to measurement error, then it is reasonable to use the predicted quantity in the numerator of (7), which is not a random variable (Bockstael and Strand). In this case the second-order approximation consists only of the first and third terms on the right hand side of (7), as exhibited in the literature.

Consider now the case in which errors arise due to omitted variables. Examination of (4)-(6) reveals that the variable Q , the individual number of trips, is in the role of x in (7). By hypothesis, Q is a random variable for each individual and the covariance term in (7) is not zero in general. The relevance of the covariance term in (7) depends on the functional form for the demand equation.

Suppose first that the demand equation is of the semi-log form. In this instance, the consumer surplus function is given by (5) and the covariance between x and y is a linear function of y . By Jensen's inequality, we can replace the mean (over the sample) of the covariances by the covariance of the sample mean quantity and the price coefficient. Since the sample mean of the (estimated) error terms is zero, the sample mean of the quantity variable is non-stochastic, and the covariance is zero. This will not necessarily be the case for functional forms, such as the linear demand curve, which yield a nonlinear covariance term.⁴

In order to investigate the potential magnitude of the error from truncating at the second term and how this might vary across functional

forms for the demand curve, we undertake a Monte Carlo analysis which yields a direct estimate of the expected value of consumer surplus. The difference between this direct estimate and the second-order approximation reveals the magnitude of the approximation error. The methods used and our results are reported in the next section of the paper.

III. Estimates and Results

We estimated the three functional forms given above for a travel cost demand function for recreation. The model employed was very simple; the number of trips over a season was the measure of quantity and travel cost was used as the price variable. The travel costs were measured as the round-trip out-of-pocket expense of the trip not including an opportunity cost of travel time. The data, collected by mail survey, concern hunting trips for Bighorn sheep in Alberta, Canada.⁵

The results of the estimations are reported in Table 1. Inspection of the results indicates that the linear and semi-log perform reasonably well in terms of t-statistics on the price variable and F-tests of the overall equation. The double-log did not perform as well by these criteria.

A Monte Carlo analysis was performed in order to estimate directly the expected value of consumer surplus using a technique described by Freedman and Peters. We derived an empirical distribution for the price parameter and used this distribution to describe the consumer surplus measure for each model. This was done in the following fashion. For each model we generated a new set of values for the dependent variable for each

observation using the non-random set of independent variables for that observation and an error term that was generated from a normal distribution with zero mean and variance equal to the variance of the error for the regression equation. This new set of dependent variables was then used to estimate a new set of regression parameters, which in turn were used to compute a new value of the consumer surplus. This was repeated 5,000 times. The mean of this distribution is used as a direct estimate of the true mean for the welfare measure. Then, this is compared to the second-order approximation in (7).

In addition to the estimated demand equations, Table 1 contains the point estimates of consumer surplus computed from equations (4)-(6), the second-order approximation to expected surplus as calculated from (7), and the expected surplus generated by the Monte Carlo procedure.

It is interesting that the magnitudes of the error due to truncation of the approximation at the second term can be as large as the differences between consumer surplus estimates across functional forms; much of the apparent similarity between the linear and double-log surplus estimates based on the approximation can be attributed to truncation error. As well, the truncation error is lowest for that model (the double-log) with the lowest t-statistic on the price variable.

IV. Discussion

Our analysis suggests that expected consumer surplus estimates obtained using a Taylor-series truncated at the second order may be misleading. In

our example, this is particularly severe for the linear demand curve, and moderately troublesome for the semi-log form.

The reason for this is that over repeated trials the Monte Carlo procedure will result in some estimates of the price variable that are very close to zero. Since this coefficient appears in the denominator of the consumer surplus function for the linear and semi-log forms, this can have a dramatic effect of the surplus estimate. This inflation of the variability of the surplus function occurs yields especially poor performance for a second-order approximation procedure. This effect is not as significant for the double-log form. The effects of the truncation are not necessarily signaled by the statistical insignificance of the price parameter. Our example exhibited a situation in which one functional form fits the data well, but approximation errors were large for this form.

One potential strategy to circumvent this problem is to truncate the approximation at a higher order. Unfortunately, this is not easily done since the third order term involves quantities that are not readily calculable. Of course, one always could carry out a Monte-Carlo analysis of the sort reported here. In the case where the numerator of the ratio is nonrandom, the distribution of $1/x$ can be integrated numerically, as long as the distribution of x is known, as it will be for a regression parameter (e.g. Kaylen and Preckel). This will not work for the case of a random numerator, since the joint distribution of x and y is not generally known. Another possibility is to truncate the consumer surplus function even in the case when it is finite (as with the linear form). For example,

it may be reasonable to impose that willingness to pay for each observation, as approximated by consumer surplus, cannot be negative and cannot exceed income. A full investigation of the imposition of these sorts of constraints is a topic for further research.

Footnotes

1. It is not necessarily the case that benefit-cost analysts act as if they are risk neutral. If they are not, then the variance of the welfare measure matters as well as its mean. On the relationship between functional form and the variance of the welfare measure see Adamowicz et al.

3. Another functional form of interest is the linear-log, i.e. $Q = a + b \ln P$. For this form the consumer surplus function is $CS = \text{Max}(P)(\text{Max}(Q) - b) - P(Q - b)$. Since this is a linear function of random variables, no approximation is needed. Hence, we do not consider this form here. The use of the double-log form can be questioned on the grounds that it is not integrable, which implies that recreation is essential (Bockstael et al.) and the manner in which it is truncated is ad hoc.

3. Both Bockstael and Strand and Kealy and Bishop adjust the ratio of expected values by an "expansion factor" $(1 + (1/t^2))$ to account for the other terms in (7), where t is the t -statistic for the price parameter estimate. Note that

$$\begin{aligned} E[x]/E[y](1 + (1/t^2)) &= E[x]/E[y](1 + (\text{VAR}[y]/E[y]^2)) \\ &= E[x]\text{VAR}[y]/E[x]^3 + E[x]/E[y] \end{aligned}$$

This is the approximation in (7) only if the second term in (7) is zero.

4. Let A be the $k \times n$ matrix $(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$. Then the covariance of the square of the dependent variable vector and the vector of parameters is $2\text{diag}[\sigma^2 \mathbf{x}_i \mathbf{b}] A'$, where σ^2 is the variance of the error terms, \mathbf{x}_i is the vector of explanatory variable values for observation i , and \mathbf{b} is the estimated parameter vector.

5. Further details regarding the data are available from the authors upon request.

References

Adamowicz, W., J. Fletcher, and T. Graham-Tomasi. "Functional Form and the Statistical Properties of Welfare Measures." American Journal of Agricultural Economics. In press.

Bockstael, N. and I. Strand. "The Effect of Common Sources of Regression Error on Benefit Estimates." Land Economics, 63(1987):11-20.

Freedman, D. and S. Peters. "Bootstrapping a Regression Equation: Some Empirical Results." Journal of The American Statistical Association, 79(1984):97-106.

Kaylen, M. and P. Preckel. "MINTDF: A Fortran Subroutine for Computing Parametric Integrals." Station Bulletin No. 519, Agricultural Experiment Station, Purdue University, May, 1987.

Kealy, M. and R. Bishop. "Theoretical and Empirical Specifications in Travel Cost Demand Studies." American Journal of Agricultural Economics, 69(1986):660-667.

TABLE 1

<u>Model</u>	<u>Regression Results</u>			<u>Consumer Surplus</u>			
	<u>Constant</u> <u>(a)</u>	<u>Slope</u> <u>(b)</u>	<u>F</u>	<u>Point</u> <u>Estimate</u>	<u>Monte</u> <u>Carlo</u> <u>Mean</u>	<u>Second-Order</u> <u>Approximation</u>	<u>Approx. Error</u> <u>As % of</u> <u>Approximation</u>
Linear	2.869 (0.181)	-.0025 (.00103)	5.84	\$1,367	\$2,200	\$1,601	37.4%
Semi-Log	0.864 (0.0677)	-.0010 (.00038)	6.79	\$2,593	\$3,161	\$2,971	6.4%
Double-Log	1.126 (0.238)	-.0885 (.0557)	2.52	\$1,532	\$1,548	\$1,538	0.65%