Consumer Surplus Estimates and the Source of Regression Error

Timothy K.M. Beatty
University of British Columbia
Faculty of Agricultural Sciences
Rm. 335 - 2357 Main Mall
Vancouver, B.C., Canada V6T 1Z4
timothy.beatty@ubc.ca

Nicholas Brozovic
University of Illinois at Urbana-Champaign
Department of Agricultural and Consumer Economics
326 Mumford Hall, MC-710
1301 West Gregory Drive
Urbana, IL 61801
nbroz@uiuc.edu

Michael B. Ward
University of California, Santa Barbara
Donald Bren School of Environmental Science & Management
4408 Bren Hall
Santa Barbara, CA 93106-5131
ward@bren.ucsb.edu

Consumer Surplus Estimates and the Source of Regression Error[1]

Abstract

Contrary to widely held belief, we show that the source of regression error does not matter when calculating Marshallian surplus. A misspecified demand curve, not the assumed source of regression error, leads to differences in estimates of consumer surplus.

JEL Codes

D60, C24, Q51.

Keywords

Regression Error; Marshallian Surplus; Welfare Analysis.

Introduction

There are several potential sources of error when using regression analysis for applied welfare studies. Does our belief over the source of regression error – measurement error, heterogeneous preferences, and omitted variables – dictate how consumer surplus should be calculated? The applied welfare literature, notably work on valuing changes in environmental amenities, employs different methodologies for calculating consumer surplus based on the assumed source of regression error (Bockstael and Strand (1987); Adamowicz et al. (1989); Bockstael et al. (1990); Kling and Sexton (1990); Hellerstein (1992a); Hellerstein (1992b); Kling (1992); Adamowicz et al. (1994); Gautam et al. (1996); Haab and McConnell (1996); Kling (1997); Pendleton and Mendelsohn (2000); Whitehead et al. (2000); Haab and McConnell (2002)).

The studies cited above suggest that if the presumed source of regression error is omitted variables, consumer surplus calculations should be made using individual demand functions passing through each observed data point. Conversely, if measurement error or random preferences are the source of error, then consumer surplus should be calculated using the average demand curve passing through the mean of the data. Because studies consistently report differences in estimated consumer surplus based on which of these two procedures is used, the type of error present is thought to have economic significance.

In this paper, we show that so long as the demand function is correctly specified, both procedures for calculating consumer surplus must give identical results. If the error terms

are additive and orthogonal to the regressors, the source of the error term is unimportant
to the determination of expected consumer surplus. Observed differences in the literature
occur when the individual and average demand functions are assumed to have the same
functional form; this will occur when censoring of quantities demanded is not properly
accounted for.

Marshallian consumer surplus from cross-sectional data

We now derive expressions for Marshallian consumer surplus under competing
assumptions about the source of the error term. Consider a general individual demand
function given by

$$q_i = q(p_i, \mathbf{e}_i, \mathbf{f}_i), \tag{1}$$

where demand for the good of interest by individual i is a function of the price (such as
the cost of travel to an environmental amenity) p, other variables $\mathbf{f}$ and a random
preference term, $\mathbf{e}$. Assume

$$
\begin{aligned}
&\mathrm{E}(\mathbf{e}_i \mid p_i) = 0 \\
&\mathrm{E}(\mathbf{f}_i \mid p_i) = 0
\end{aligned}
\tag{2}
$$

Furthermore, assume that the observed quantities are subject to some measurement error
of $\mathbf{g}_i$. Thus, the quantity actually observed is given by

$$\tilde{q}_i(p, \mathbf{e}_i, \mathbf{f}_i, \mathbf{g}_i) = q_i(p, \mathbf{e}_i, \mathbf{f}_i) + \mathbf{g}_i. \tag{3}$$

For this source of error we also assume:

$$\mathrm{E}(\mathbf{g}_i \mid p_i) = 0. \tag{4}$$

Given an initial price of zero, the expectation of the Marshallian consumer surplus is

$$\mathrm{E}\left(CS\right) = \underset{e,f,g}{\mathrm{E}} \int_0^\infty \not{q_i}(p,\boldsymbol{e}_i,\boldsymbol{f}_i,\boldsymbol{g}_i)\mathrm{d}p$$

$$= \underset{e,f,g}{\mathrm{E}} \int_0^\infty \left[q_i\left(p,\boldsymbol{e}_i,\boldsymbol{f}_i\right) + \boldsymbol{g}_i \right]\mathrm{d}p \,. \tag{5}$$

$$= \underset{e,f}{\mathrm{E}} \int_0^\infty q_i\left(p,\boldsymbol{e}_i,\boldsymbol{f}_i\right)\mathrm{d}p$$

The $\boldsymbol{g}_i$ term drops out because the conditional expectation of $\boldsymbol{g}_i$ is zero by (4). As integration is a linear operator, we may pass the expectation into the integral, yielding

$$\mathrm{E}\left(CS\right) = \underset{e,f}{\mathrm{E}} \int_0^\infty q_i\left(p,\boldsymbol{e}_i,\boldsymbol{f}_i\right)\mathrm{d}p$$

$$= \int_0^\infty \underset{e,f}{\mathrm{E}}\, q_i\left(p_i,\boldsymbol{e}_i,\boldsymbol{f}_i\right)\mathrm{d}p \tag{6}$$

In words, this states that the expectation of individual consumer surplus and the consumer surplus of the expected demand are equivalent. The extension to a finite sample is straightforward. This result holds irrespective of whether any or all of the possible disturbance terms are present, as the conditional expectations of all the disturbance terms are zero by standard regression assumptions.

Equation (6) runs contrary to much of the literature which suggests that both theoretically and empirically, the source of the error term is meaningful. Below we explain the discrepancy between our result and previous work on the subject. For ease of presentation, we use a linear demand function, but the results are completely general.

An implicit assumption in the studies cited is that demands for negative quantities are not observed. Except in unusual circumstances, it seems implausible that people should want to consume a negative quantity of a good when they could consume zero instead. Thus, it

is common practice to censor demand specifications at a quantity demanded of zero. For example, for the linear demand function, the actual individual demand function may be given by

$$q_i = \max[a + bp_i + e_i, 0],\tag{7}$$

where for simplicity, there is a single additive disturbance term, with a conditional expectation of zero. For normal goods, $b < 0$. With a linear demand function, the censoring defines an expected choke price $-a/b$, above which the good is no longer demanded.

From (6), the welfare measure given by the expectation of individual consumer surplus is

$$
\begin{aligned}
\mathrm{E}(CS_{\mathrm{Ind}}) &= \mathrm{E} \int_0^{-\frac{a+e_i}{b}} (a + bp + e_i)\,\mathrm{d}p \\
&= \mathrm{E}\left(\int_0^{-\frac{a}{b}} (a + bp)\,\mathrm{d}p + \int_{-\frac{a}{b}}^{-\frac{a+e_i}{b}} (a + bp)\,\mathrm{d}p + \int_0^{-\frac{a+e_i}{b}} e_i\,\mathrm{d}p \right) \\
&= \int_0^{-\frac{a}{b}} (a + bp)\,\mathrm{d}p + \mathrm{E}\left(\frac{e_i^2}{2b} \right) + \mathrm{E}\left(-\frac{e_i(a+e_i)}{b} \right) \\
&= -\frac{a^2}{2b} + \mathrm{E}\left(-\frac{e_i(2a + e_i)}{2b} \right)
\end{aligned}\tag{8}
$$

Following (6), the consumer surplus calculated from the correctly specified average demand curve must also equal (8).

However, if we use the individual demand function passing through the expected mean of the data then the analogous welfare measure, denoted CS_{Mean}, is

$$\mathrm{E}\left(CS_{\mathrm{Mean}}\right)= \int_{0}^{-\frac{a}{b}} \left(a + bp\right)\mathrm{d}p$$

$$= -\frac{a^2}{2b}$$ (9)

$$< \mathrm{E}\left(CS_{\mathrm{Ind}}\right)$$

The difference between the two welfare measures may be substantial. For example if we assume that $e_i \sim N\left(0, s^2\right)$, then

$$E(CS_{\mathrm{Ind}}) = -\frac{a^2 + s^2}{2b},$$ (10)

yielding a difference of $E(CS_{\mathrm{Ind}}) - E(CS_{\mathrm{Mean}}) = -s^2/2b > 0$, for normal goods. [2]

Above, we proved that calculating the average of the individual consumer surplus estimates using demand functions through each observed data point is exactly equal to the consumer surplus of the expected demand function. The assumption that negative quantities are never demanded means that there is idiosyncratic truncation along the price axis. This implies that the average demand will not have the same functional form as the individual demands. In the example used above, idiosyncratic truncation means that the average demand will not be linear if the individual demands are linear. As a result the welfare measure embodied by the estimator CS_{Mean}, calculated using the individual demand specification through the mean of the data, is misspecified and underestimates consumer surplus.

---

[2] If we assume that e has the standard logistic distribution, the difference between $E(CS_{\mathrm{Ind}})$ and $E(CS_{\mathrm{Mean}})$ is $-p^2/2b$.

As we have shown here, the two methods of calculating consumer surplus will yield identical results if the disturbance terms have a conditional expectation of zero. Discrepancies between the two measures are the result of the misspecification of one of the functional forms, not the result of some underlying economic significance of the source of the error term.

Conclusion

In welfare analysis there has been some concern that the source of the error term used in regression can affect the estimated consumer surplus. There exists a large literature suggesting that, based on the inferred source of error, one of two alternative consumer surplus estimation measures should be used: the expected individual consumer surplus, or the consumer surplus of the expected individual. Empirical studies imply that these two yield different welfare measures. However, in this paper we demonstrate that linearity of the integration operator requires that the two welfare measures yield identical results. The common assumption that negative demands are not observed implies that the average demand curve has a functional form different to that of the individual demand curves. Failure to take this into account by using the individual demand curve passing through the mean observed data is a common mistake and will lead to underestimates of Marshallian consumer surplus.

References

Adamowicz, W. L., J. J. Fletcher and T. Graham-Tomasi (1989). "Functional Form and the Statistical Properties of Welfare Measures." <u>American Journal of Agricultural Economics</u> **71**(2): 414-421.

Adamowicz, W. L., J. J. Fletcher and T. Graham-Tomasi (1994). "Functional Form and the Statistical Properties of Welfare Measures - Reply." <u>American Journal of Agricultural Economics</u> **76**(4): 958-959.

Bockstael, N., I. E. Strand, K. E. McConnell and F. Arsanjani (1990). "Sample selection bias in the estimation of recreation demand functions: an application to sportfishing." Land Economics **66**: 40-49.

Bockstael, N. E. and I. E. Strand (1987). "The effect of common sources of regression error on benefit estimates." Land Economics **63**: 11-20.

Gautam, A. B., I. Strand and J. Kirkley (1996). "Leisure/labor tradeoffs: The backward-bending labor supply in fisheries." Journal of Environmental Economics and Management **31**(3): 352-367.

Haab, T. C. and K. E. McConnell (1996). "Count data models and the problem of zeros in recreation demand analysis." American Journal of Agricultural Economics **78**(1): 89-102.

Haab, T. C. and K. E. McConnell (2002). Valuing Environmental and Natural Resources: The Econometrics of Non-market Valuation. Northhampton, Edward Elgar.

Hellerstein, D. (1992a). "Estimating Consumer Surplus in the Censored Linear-Model." Land Economics **68**(1): 83-92.

Hellerstein, D. M. (1992b). "Correcting for Bias When Average Values Are Used to Compute Changes in Consumer Surplus." Leisure Sciences **14**(4): 337-345.

Kling, C. L. (1992). "Some Results on the Variance of Welfare Estimates from Recreation Demand Models." Land Economics **68**(3): 318-328.

Kling, C. L. (1997). "The gains from combining travel cost and contingent valuation data to value nonmarket goods." Land Economics **73**(3): 428-439.

Kling, C. L. and R. J. Sexton (1990). "Bootstrapping in Applied Welfare Analysis." American Journal of Agricultural Economics **72**(2): 406-418.

Pendleton, L. and R. Mendelsohn (2000). "Estimating recreation preferences using hedonic travel cost and random utility models." Environmental & Resource Economics **17**(1): 89-108.

Whitehead, J. C., T. C. Haab and J. C. Huang (2000). "Measuring recreation benefits of quality improvements with revealed and stated behavior data." Resource and Energy Economics **22**(4): 339-354.