Transportation Research Part A 82 (2015) 169–192

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

On finite sample performance of confidence intervals methods for willingness to pay measures

Valerio Gatta [a],*, Edoardo Marcucci [a], Luisa Scaccia [b]

[a] Dipartimento di Scienze Politiche, CREI, Università di Roma Tre, Via G. Chiabrera, 199, 00145 Roma, Italy
[b] Università di Macerata, Dipartimento di Economia e Diritto, via Crescimbeni 20, 62100 Macerata, Italy

ARTICLE INFO

Article history:
Received 3 March 2015
Received in revised form 15 June 2015
Accepted 2 September 2015
Available online 23 October 2015

Keywords:
Confidence interval
Willingness to pay
Choice modeling
Elasticity
Standard error

ABSTRACT

This paper systematically compares finite sample performances of methods to build confidence intervals for willingness to pay measures in a choice modeling context. It contributes to the field by also considering methods developed in other research fields. Various scenarios are evaluated under an extensive Monte Carlo study. Results show that the commonly used Delta method, producing symmetric intervals around the point estimate, often fails to account for skewness in the estimated willingness to pay distribution. Both the Fieller method and the likelihood ratio test inversion method produce more realistic confidence intervals for small samples. Some bootstrap methods also perform reasonably well, in terms of effective coverage. Finally, empirical data are used to illustrate an application of the methods considered.

1. Introduction

Willingness to pay (*WTP*) is the amount of money an agent would pay to obtain a desired good or service. Reliable *WTP* measures are fundamental in transportation economics. In a choice modeling framework, typically assuming linear-in-attributes utility functions, the *WTP* for a given attribute is obtained dividing its coefficient by that of cost. Since model estimation yields an estimate of the true coefficients, the computed *WTP* (i.e. $\widehat{WTP}$) is itself an estimate with a given probability distribution. Thus, it is desirable to calculate confidence intervals (CIs), in addition to point estimates. This is not trivial since the finite sample distribution of the *WTP* estimator is not known. When maximum likelihood estimates (MLEs) are used for the coefficients, the distribution of $\widehat{WTP}$ is the ratio of two correlated, asymptotically normal, distributions. The distribution of the ratio of two normal variables has been derived by Fieller (1932) and Hinkley (1969), and shown to be approximately normal when the coefficient of variation of the denominator variate is negligible (Marsaglia, 2006). More recently, Daly et al. (2012) showed that $\widehat{WTP}$ is itself a MLE, its distribution is asymptotically normal and the Delta method gives an exact measure of its standard error.

Notwithstanding the relevant results obtained by Daly et al. (2012) with respect to the asymptotic properties of $\widehat{WTP}$, its finite sample distribution can be substantially different from the normal distribution. While *WTP*s determined for policy making by governments generally use extensive stated choice studies with thousands of respondents, for which the asymptotic normality of $\widehat{WTP}$ is fulfilled, in many situations such large samples are not available and small sample CI

---

* Corresponding author.
E-mail addresses: valerio.gatta@uniroma3.it (V. Gatta), emarcucci@uniroma3.it (E. Marcucci), scaccia@unimc.it (L. Scaccia).

http://dx.doi.org/10.1016/j.tra.2015.09.003
0965-8564/

methods might be advocated. This is especially true under particular conditions. For example, very large samples are needed to compensate for a cost parameter approaching 0 (Bolduc et al., 2010) or for the greater uncertainty induced by random coefficients in a mixed logit context (Bliemer and Rose, 2013).

This motivates the development of different methods to calculate CIs for *WTP* in a finite sample context. For example, the Delta method relies on the asymptotic normality of $\widehat{WTP}$. While giving the exact standard error of $\widehat{WTP}$, a little thought is required when using this method to calculate *t*-ratios or CIs (Daly et al., 2012). In particular, for finite samples, the effective coverage rate of CIs obtained through the Delta method might be substantially different from the nominal one (Bolduc et al., 2010). In contrast to the Delta method, Fieller (Fieller, 1954; Bolduc et al., 2010) and likelihood ratio test inversion methods (Armstrong et al., 2001) only assume asymptotic normality of the coefficients involved in the ratio. Other methods use bootstrap sampling techniques, refraining from any distributional assumption (Efron, 1979, 1987; DiCiccio and Efron, 1996).

Even if most of the methods to construct CIs for WTP are asymptotically equivalent, in a finite sample context their performances in terms, for example, of effective coverage rates might be substantially different. Only few studies try to investigate small sample performances of various CIs methods. Armstrong et al. (2001) investigate the potentialities of likelihood ratio test inversion and Fieller method compared to Delta and two bootstrap methods. Using only real data, they conclude that Delta method is the simplest to implement but is often inappropriate: not only are the intervals obtained too narrow but they are also symmetrical, by construction, with respect to $\widehat{WTP}$. On the other hand, they find Fieller, likelihood ratio test inversion and one bootstrap method providing very similar results and suggest using one of the first two, the last one being tedious and long to compute. Bolduc et al. (2010) make use of a Monte Carlo study to show the advantages of Fieller method over the Delta and a simple bootstrap methods, when the cost coefficient approaches 0. Chiew and Daziano (2013) extend the work of Bolduc et al. (2010) by including Bayesian post-processing method and solving the model in *WTP* space in the comparison. Their conclusion is that, under standard conditions, all methods perform similarly, apart from the Delta which appears to produce problematic CIs. Also working in WTP space seems to give narrower CIs, but with a poorer coverage. Under conditions of weak identification, instead, the Fieller method shows its superiority. Hirschberg and Lye (2010) compare the Delta and Fieller methods from a geometrical point of view and conclude that when the Fieller and Delta intervals differ, the Fieller interval results in better coverage, and in some cases, this advantage can be very large. Outside the transportation field, Hole (2007) proposes a Monte Carlo study to assess the performance of Delta, Fieller and two bootstrap methods in constructing CIs for WTP in health care. The Delta method is found to be the most accurate when data is well conditioned, while the bootstrap is more robust to noisy data and misspecifications of the model. Bernard et al. (2007) claim the merits of Fieller method over the Delta and an exact version of the likelihood ratio test inversion method in delivering CIs for elasticities in energy demand models.

As the above review illustrates, the conclusions reached by different studies are not always in accordance and, to the best of our knowledge, a comparison of all the existing methods does not exist. This paper provides some guidelines for choosing, under different conditions, an appropriate method to construct CIs for *WTP*, in finite sample contexts. It contributes to the literature by comprehensively and systematically comparing all the methods used in the choice modeling field, as well as proposing other methods borrowed from different research areas. The comparison is carried out through a Monte Carlo study, within a multinomial logit (MNL) framework which enables quick parameter estimates, a fundamental requirement in simulations. Data are generated under different scenarios mimicking real situations in which the finite $\widehat{WTP}$ distribution is potentially highly skewed and far from normal. Two real data sets are used to illustrate the practical relevance of the issues raised in the simulation study.

The paper is structured as follows: Section 2 describes *WTP* estimation within a choice modeling context; Section 3 illustrates assumptions, advantages and disadvantages of the methods used for CI estimation; Section 4 compares methods through a Monte Carlo study; Section 5 reports the results from real data applications; Section 6 concludes and suggests general guidelines.

2. Logit models and *WTP* estimation

Consider a sample of N decision makers, facing J alternatives, in T choice experiments. In a random utility framework, the choice of individual n, for $n = 1, \dots, N$, is modeled as:

$$y_{int} = \begin{cases} 1 & \text{if } U_{int} \geqslant U_{jnt} \text{ for } j = 1, \dots, J \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where

$$U_{int} = V_{int} + \epsilon_{int} \tag{2}$$

is the unobservable utility that individual n derives from alternative i (for $i = 1, \dots, J$), in choice experiment t (for $t = 1, \dots, T$), V_{int} is the observable utility and ϵ_{int} is an error term. Observable utility is generally assumed linear-in-the-attributes so that

$$V_{int} = X_{int}\beta, \tag{3}$$