Appalachian
STATE UNIVERSITY

Department of Economics Working Paper

Number 17-08| October 2017

Validity and Reliability of Contingent Valuation and Life Satisfaction Measures of Welfare: An Application to the Value of National Olympic Success

Brad R. Humphreys
West Virginia University

Bruce K. Johnson
Centre College

John C. Whitehead
Appalachian State University

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Validity and Reliability of Contingent Valuation and Life Satisfaction Measures of Welfare: An Application to the Value of National Olympic Success[1]

Brad R. Humphreys, West Virginia University

Bruce K. Johnson, Centre College

John C. Whitehead[2], Appalachian State University

October 12, 2017

[1] This research was funded by the Social Sciences and Humanities Research Council of Canada, Award #410-2009-0920. We thank Christine Young of Consumer Contact for her help with the survey and data collection and Dan Mason for his contributions to this paper. This paper has benefitted from presentation at the Annual Meeting of the Southern Economic Association in Washington, DC, November 20, 2011, especially from the comments of Craig Depken. The authors declares that they have no relevant or material financial interests that relate to the research described in this paper.

[2] Corresponding Author: Department of Economics, Appalachian State University, Boone, NC 28608, phone: (828) 262-6121, e-mail: whiteheadjc@appstate.edu.

Validity and Reliability of Contingent Valuation and Life Satisfaction Measures of Welfare: An Application to the Value of National Olympic Success

Abstract. The contingent valuation method (CV) has long been used to estimate nonmarket values of environmental and other public goods and amenities. Recently, life satisfaction (LS) measures have been used to estimate nonmarket values. This paper empirically compares CV and LS measures of welfare. We elicit willingness-to-pay (WTP) estimates for medals won by Canadian athletes and LS measures using Canadian survey data collected before and after the 2010 Winter Olympic Games. These data permit comparative analyses of reliability and validity of CV and LS measures. Both exhibit econometric reliability. CV and LS WTP estimates for medals increases after the Olympics. CV measures of WTP exhibit temporal reliability but LS measures of welfare lack temporal reliability and are significantly greater than CV measures.

Keywords: contingent valuation method; life satisfaction method; willingness-to-pay; validity; reliability
JEL Codes: C18, C52, D12, D62, I15, Q51

1. Introduction

Valuing nonmarket goods or outcomes represents an important goal in a number of areas of economic research. In environmental economics, a large literature focuses on valuing environmental quality (Ferreria and Moro, 2010). In health economics, a large literature focuses on valuing health outcomes (Groot and Brink, 2006; Howley, 2017). Other applications include valuing local public goods (Levinson, 2012), sports public goods (Johnson and Whitehead, 2012; Kavetsos and Szymanski, 2010), natural disasters (Luechinger and Raschky, 2009; Kountouris and Remoundou, 2011), drought (Carroll, Frijters and Shields, 2009) and other nonmarket outcomes.

The literature contains three complementary approaches for valuing nonmarket goods and outcomes: revealed preference methods, stated preference methods, and the life satisfaction (LS) method. Revealed preference methods include hedonic pricing that relates property values or wages to observable nonmarket outcomes and the travel cost method that relates recreation visits to time and distance travelled. Stated preference approaches include the contingent valuation (CV) method that elicits willingness to pay (WTP) estimates. The life satisfaction method uses self-reported life-satisfaction data from surveys and relates these to variation in nonmarket goods. Each approach has strengths and weaknesses; each also has adherents and antagonists.

Though the three methods are complementary, few studies employ more than one, which makes direct comparisons of them difficult. This paper undertakes a CV and LS valuation of a sports-related public good, success by Team Canada in the 2010 Vancouver Winter Olympic Games, to evaluate the validity and reliability of the CV and LS methods. It uses data from a nationally representative sample of Canadians surveyed before and after the 2010 Games about their willingness to pay for improved Canadian success in winning Olympic medals. Most survey respondents have experience with Olympic success, or lack thereof. Canada hosted the Summer Olympic Games in 1976 and failed to win a single gold medal. It hosted the Winter Olympics in 1988, and though Canada's love of winter sports traditionally translates into success in the Winter Olympics, it once again failed to win a single gold. After Canada was awarded the 2010 Winter Games, the federal government funded "Own the Podium," a program to identify and train world-

class winter athletes with the ambitious goal of winning more medals than any other nation at the 2010 Games.

We use the survey results to compare the CV and LS methods in terms of reliability -- the extent to which a valuation method consistently generates similar estimates, and validity -- the extent to which a valuation method provides an estimate centered on the true value. Both methods exhibit econometric reliability, based on standard regression diagnostic statistics; only estimates from the CV method exhibit temporal reliability, i.e., consistency from one time period to the next. The results indicate low convergent validity based on WTP estimates for Olympic medal success. While CV generates consistent, small, and statistically significant WTP estimates, the LS estimates vary widely. Many are statistically indistinguishable from zero; one is positive and three orders of magnitude larger than comparable CV estimates and another is negative and very large.

The results contribute to the growing literature focused on alternative methods of estimating the value of nonmarket goods or outcomes in economics. CV methods have been criticized for presenting survey respondents with hypothetical situations they have no experience with (Groot and Brink, 2004), requiring respondents to perform difficult cognitive tasks (Kountouris and Remoundou, 2011), and being plagued by issues of strategic respondent behavior (Luechinger and Raschky, 2009). While our CV estimates may also suffer from these limitations, the results here indicate that CV WTP estimates have superior validity along some margins, and equal reliability when compared to LS WTP estimates based on data from the same survey.

2. Context

The data come from a survey designed to elicit WTP estimates for elite sporting success in the Olympic Games. Sports represent an interesting setting for comparing different econometric methods (Gu and Koenker, 2017). National, state and local governments heavily subsidize the sports industry. Sports mega-events, such as the Olympic Games, the Super Bowl and the World Cup, generate substantial economic activity, but it is not clear the tangible economic benefits generated by the mega-events ever exceed, or even equal, the costs of staging them (Siegfried and Zimbalist, 2007; Baade and Matheson, 2016). Nevertheless, subsidies for sports mega-events have

the potential to generate positive net benefits since they are more unique experiences and may generate local, state and national public goods such as civic pride (Matheson and Baade, 2004). Maennig and Du Plessis (2007) argue that the value of civic pride and other "feel-good" measures of value be included in economic assessments of subsidies to mega-events.

The CV method can be used to estimate the civic pride and "feel-good" values by constructing hypothetical policy scenarios and eliciting willingness to pay in survey settings. The CVM is increasingly being used to estimate the benefits of sports-related amenities (Johnson and Whitehead, 2012). Several prior studies estimated willingness to pay for mega-events. Atkinson, et al. (2008) conducted a CVM survey and asked people in London, Manchester, and Glasgow about their willingness-to-pay to attract the Olympics to London. They found that people across the country were willing to pay to host the Olympics, with Londoners willing to pay more than non-Londoners. Another survey in Bath and Southwest England also found that non-Londoners valued the London Games (Walton et al., 2008). Even though British aggregate willingness-to-pay was high, it fell billions of dollars short of the cost of hosting the 2012 Olympics. Wicker et al. (forthcoming) found that German willingness to pay to host the Olympics varied across regions, with no region willing to pay the full cost.

Sussmuth, Heyne and Maennig (2010) found that ex post German willingness to pay for hosting the 2006 FIFA World Cup rose substantially from the ex ante willingness to pay. They suggest that this is the result of resolving the uncertainty around experience goods. Humphreys et al. (forthcoming) estimate the intangible benefits generated by medal success by Canadian athletes in the 2010 Vancouver Winter Olympic games. They find that benefits of Canadian spending on elite athletes far exceed the costs. This finding stands in stark contrast to those of nearly every other published CVM analysis of the benefits of sports public goods, which almost invariably find that costs exceed benefits.

The LS method exploits spatial and/or temporal variation in outcomes such as public good amenities, environmental characteristics, or individual outcomes such as health. It correlates these with self-reported life satisfaction and assesses the tradeoff between income and public goods, amenities or health outcomes (Welsch, 2009; Welsch and Kühling 2010). Some argue that the LS

method may be preferred to revealed preference methods since it does not require correlation of actual choice with amenity changes. The LS method may be preferred to stated preference methods since it does not require hypothetical policy scenarios and incentive compatible valuation questions.

The LS method has been extensively applied to the valuation of environmental amenities and health outcomes. In the first environmental valuation application of the LS method Welsch (2006) used country-level data and finds that the value of air quality improvements made in Europe in the 1990s varied between 2 percent and 10 percent of per capita income. More recently, Ferreira and Moro (2010) used individual level data and compare the HP and LS methods for valuing air quality in Ireland and argue that LS WTP estimates are "too large." Other environmental applications include Kountouris and Remoundou (2011) who estimate the value of avoiding forest fires. Ferreira and Moro (2010) and Luechinger (2009) and others argue that LS measures of welfare may be biased upwards due to a downward biased parameter estimate on the income variable in the life satisfaction equation. The income coefficient may be biased downwards due to measurement error, because relative, not absolute, income influences happiness, or due to omitted variables such as lagged income. In contrast, Levinson (2012) finds that LS WTP estimates for air quality compare favorably to estimates from revealed and stated preference methods.

Kavetsos and Szymanski (2010) first to applied the LS approach to sports. They investigated the effects of hosting the Olympics, World Cup or the European soccer championships on LS in 12 European countries from 1974 to 2004. The LS question read, "On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?" They found that hosting soccer championships had a positive and significant short-term effect on life satisfaction. National sport success had no effect.

In this paper we provide the first direct empirical comparison of CV and LS measures of welfare using data from the same survey. Using telephone survey data of Canadians before the 2010 Winter Olympic Games we elicited expectations of gold and overall medals won by Canadian athletes, willingness-to-pay for future Olympic training and measures of life satisfaction. Following the 2010 Winter Olympics we re-surveyed respondents with the same willingness-to-

pay and life satisfaction questions. With these data we conduct a comparative analysis of the validity and reliability of both measures of welfare. We consider internal validity by estimating the effects of income and other determinants of utility, e.g., realized expectations of medals, on welfare. We test for convergent validity by comparing the magnitudes and correlation between the CV and LS welfare measures for individual respondents. We also test for the temporal reliability of both measures of welfare across the ex-ante and ex-post surveys.

3. Survey and Data Description

The Canadian government, through Sport Canada, operates three programs designed to develop and support elite athletes: the Sport Support Program, the Hosting Event Program, and the Athlete Assistance Program. In 2007-08, Sport Canada provided the programs a total of C$120 million in support. In addition, Canada spent C$110 million on its Own the Podium program, geared specifically toward enhancing Canadian performance in the 2010 Vancouver Winter Olympics.

To determine Canadian households' willingness-to-pay for Olympic medal success, we conducted a national survey of Canadians before and after the Vancouver Games. The survey began with a series of 17 questions to get respondents thinking about the 2010 Winter Olympic Games. The questions asked whether Olympic success makes them happy or proud, whether they think medal success is important, whether it enhances national prestige, and whether they thought Canada winning the most medals was a realistic goal.

Another section of the survey contained hypothetical scenarios and elicited willingness-to-pay for Olympic medal success. Before the Games, this section began by informing respondents that Canadians ranked third in total medals in the 2006 Winter Olympic Games and asked how satisfied they are with that performance. After the Games, this section informed respondents of Canada's 2010 performance. It then informed them that the federal government spends about C$120 million per year, or about C$10 per household to support athletes for both the Summer and Winter Olympic Games. Respondents were asked whether they support such spending. Then they were told that the Own the Podium program accounts for $3 of annual spending per Canadian

household and were asked whether they think Own the Podium can increase the number of Olympic medals won by Canadians in Vancouver. If yes, they were led through a series of questions to determine how many more gold and total medals Canadians might be expected to win as a result of the Own the Podium program.

These questions set up the following hypothetical scenario: "Suppose that continuing to use federal money to fund the training of elite athletes for the Olympic Games were put to a vote. If more than half of all voters were in favor of the proposal then it would pass. Remember, if the proposal passed a typical household would continue to pay about $13 per year. If the proposal does not pass, the typical household would have about $13 more to spend on other things each year. Do you think that you would vote for or against the proposal?" This scenario allows us to estimate willingness-to-pay for success in the Vancouver Games and conduct a simple benefit-cost analysis of Own the Podium.

Next the respondents were presented with a hypothetical scenario about expanded funding of Own the Podium for the 2014 Winter Olympic Games. They were told that the extended program would be financed by an annual income tax surcharge for three years of one of the following amounts: $10, $25, $35, $50 or $65. They were asked if they thought this could increase the number of medals won by Canadians in the 2014 Winter Games. If they said yes, they were asked a series of follow-up questions to determine how many more total and gold medals Canadians might win, and whether they would be satisfied with these increases.

Then the respondents were asked if they would vote for a referendum on the income tax surcharge in one of several amounts ($A): "Suppose continuing and expanding the Own The Podium program beyond 2010 were put to a vote. If more than half of all voters were in favour of the proposal then it would pass. Remember, if the proposal passed your household would have $A less to spend on other things each year. Do you think that you would vote for or against the proposal?" A referendum question with coercive funding that is considered by respondents to be consequential is most likely to be incentive compatible and generate truthful answers (Carson and Groves, 2007). Those respondents who voted "for" were asked to rate, on a scale of 1 to 10, how certain they were that they would really vote for the proposal in a referendum. Recoding uncertain

"for" votes to "against" is a conservative approach to mitigating hypothetical bias, the tendency of survey respondents to overstate their willingness-to-pay in hypothetical situations (Loomis, 2011).

In contrast to the survey infrastructure required for a contingent valuation scenario, the life satisfaction questions are relatively simple. We asked a number of life satisfaction questions, all of the same format. These related to satisfaction with Olympic medal counts, health, employment and overall life satisfaction. The overall question asked: "All things considered, on a scale of 1 to 10, with 1 being not satisfied at all and 10 being very satisfied, how satisfied are you with your life as a whole these days?"

Before the Olympics, 1,540 Canadians across the nation completed the survey. After the Olympics, 1,660 Canadians participated, including 758 who responded to the first survey. The sample was stratified by region and gender to be representative of the Canadian population. There was a small oversampling of residents of British Columbia to gain additional information about the use value of the Olympic Games. We drop cases that did not answer the referendum or life satisfaction questions and, because of the importance of the income coefficient in the life satisfaction models, we drop any respondent who did not report income. Missing values for other variables are imputed with the mean from the reporting cases. We employ 926 observations from the pre-Olympic survey and 1,236 observations from the post-Olympic survey (Table 1). Fifty percent of the pre-Olympic sample answered both the pre-Olympic and post-Olympic survey (Both=1). Forty-three percent of the post-Olympic sample answered both surveys.

The number of gold medals the respondent expects Canada will win in the 2014 Winter Olympics increases from 12 to 16 after the Olympics (GOLD). Assuming independent samples, this difference is statistically significant according to a t-test (t=18.51). The number of bronze and silver medals the respondent thinks Canada can win in the 2014 Winter Olympics decreases from 18 to 13 after the Olympics (OTHMEDALS). Assuming independent samples, this difference is statistically significant (t=23.90). While the total medal count is little changed, respondents think the percentage of gold medals will be improved.

Respondents were asked to state whether they strongly agree, somewhat agree, somewhat

disagree or strongly disagree with five questions about the importance of Canadians winning medals in the Olympics. For example, the first statement in the pre-Olympic survey is, "It is important that Canadian athletes win the most gold models." In the post-Olympic survey each statement was preceded by "In 2014." AGREE is a scale variable equal to the sum of these responses. AGREE falls from 15 to 11 after the Olympics. This change is statistically significant for those respondents in both surveys and for those respondents who appear in only one survey.

We use a few other control variables. EDUCATION is the number of years of formal education. MARRIED is equal to 1 if respondent is married and 0 otherwise. INCOME is respondent income in thousands of dollars. AGE is the respondent's age in years. HEALTH is the respondent's state of health, self-reported on a scale of 1 (bad) to 10 (excellent). EMPLOYED is equal to 1 if respondent is employed and 0 otherwise. Respondents are more likely to be married ($t = 2.72$) and older ($t = 1.73$) post-Olympics assuming independent samples. All other control variables are unchanged before and after the Olympics.

Before the Olympics, respondents were asked whether they would support increased funding of own the Podium for the 2014 Winter Olympics at levels of $5, $10, $20, $30, $35, $50 or $65 per year over three years through an income tax surcharge. After the Olympics the tax surcharge amounts were adjusted upward to $15, $25, $35, $50, $65, $75, $100, and $150 because early responses indicated an apparently much higher willingness-to-pay than before the Olympics. Those who answered yes to this question were also asked how certain they were of their answers to allow mitigation of hypothetical bias in the question. The percentage of those voting "for" drops as the bid amount increases (Table 2). Adjusted for certainty, the percentage of respondents who said they would vote for higher taxes rose from 52 percent before the Games to 58 percent after. The percentage in favor rose even though hypothetical tax increases rose after the Olympics. While most bid amounts from the first survey were not used again in the second, three were: $35, $50, and $65. The percent voting yes at each of these bid levels rose substantially after the Games. For instance, 39 percent were certain they would vote yes at $65 before the Games, while 55 percent would have voted yes after the Games.

The respondent's overall life satisfaction is reported on scale of 1 (low) to 10 (high). Due

to the relative infrequency of responses from 1 to 5, we pool these with 6 and recode responses from a low of 0 to 4, with 1-6 becoming 0, 7 becoming 1, etc. The average life satisfaction before the Olympics is 2.48 and 2.57 after the Olympics (SATISFY). Assuming independent samples, a t-test finds no significant difference in life satisfaction before and after the Olympics (t=1.57). The life satisfaction frequencies demonstrate little change from the pre-Olympic survey to the post-Olympic survey (Table 3). About one-third of respondents report life satisfaction at the highest level and about one-quarter report life satisfaction at the middle level (on the condensed scale). Less than 20 percent report life satisfaction at each of the other levels, with the lowest reporting at the lowest level of life satisfaction.

4. Assessing Validity and Reliability

In the absence of market prices, nonmarket valuation methods face questions of accuracy. Accuracy of a measure of a theoretical construct, e.g., willingness-to-pay, is comprised of validity and reliability (Whitehead and Haab, 2013). Validity is the extent to which a valuation method generates a measure that is unbiased, that is, provides an estimate centered around the true value, if it were known. Reliability is the extent to which a valuation method consistently generates the same measure. It is important that CV and LS studies demonstrate some degree of both validity and reliability.

Several types of validity have been considered in the CV literature. Theoretical validity is the extent to which a valuation measure changes in response to the changes in conditions under which it is evaluated. Theoretical validity is also known as internal validity because the tests conducted are internal to the data. Theoretical validity tests begin with economic theory. Comparative static results are derived from the willingness to pay function as derived from the assumed underlying preference structure. The so-called "scope test" has taken on extra significance. If willingness to pay for nonmarket goods is nondecreasing in quality or quantity, i.e., scope, it exhibits theoretical validity. (Siikamäki and Larson 2015).

To specify theoretical validity tests, consider the indirect utility function that depends on nonmarket goods and income, $v(q, y)$, where utility is increasing in q and y. If an expansion in

Own the Podium is expected to lead to more Olympic medals, willingness to pay for the resulting increase in utility is the reduction in income that would exactly offset the effect on utility arising from the better Olympic outcome.

(1) $$v(q, y) - v(q', y - WTP)$$

where $q' > q$ is the expected increase in Olympic medals that would result from expansion of the Own the Podium program.

Willingness to pay can also be expressed as the difference in expenditure functions

(2) $$WTP = e(q, u) - e(q', u)$$

Substituting the indirect utility function into (2) yields

(3) $$WTP = y - e(q', v(q, y))$$

Comparative static properties of the willingness-to-pay function can be used to test for the theoretical validity of WTP. Willingness-to-pay is increasing in medals, $\frac{\partial WTP}{\partial q'} = -\frac{\partial e}{\partial q'} > 0$, and increasing in income, $\frac{\partial WTP}{\partial y} = 1 - \frac{\partial v(q)}{\partial y} \Big/ \frac{\partial v(q')}{\partial y} > 0$ if income and medal success are substitutes.

The response, Y_{i1}, to the referendum valuation question depends on whether willingness-to-pay is greater than the tax amount:

(4) $$Y_{i1} = \begin{cases} For = 1 & if\ WTP_i \geq A \\ For = 0 & otherwise \end{cases}$$

The probability of a yes response is estimated with the probit model where $WTP_i = \alpha' x_i$

$$
\begin{aligned}
\Pr(Y_i) &= \Pr(WTP_i \geq A) \\
&= \Pr\left(\alpha' x_i + \varepsilon_i \geq A\right) \\
&= \Pr\left(\frac{\alpha' x_i - A}{\sigma} \geq \frac{\varepsilon_i}{\sigma}\right) \\
&= \Phi\left(\frac{\alpha' x_i - A}{\sigma}\right)
\end{aligned}
\tag{5}
$$

where x_i is a vector of independent variables including q' and a constant, $-1/\sigma$ is the coeffient on the tax amount, $\varepsilon_i \sim N(0,\sigma^2)$, and $i = 1, \ldots, n$ respondents.

The coefficients on the willingness-to-pay model are estimated from the single-bound censored probit coefficients using the procedures described in Cameron and James (1987). Since the dollar amount varies across respondents, σ can be identified and mean willingness-to-pay is

$$
WTP_{CVM} = \sigma\left(\alpha' \bar{x}\right). \tag{6}
$$

Mean willingness to pay is calculated from the estimates of the regression coefficients at the mean of the independent variables, $\bar{x}$. The standard errors are obtained from the asymptotic covariance matrix by the Delta method (Cameron, 1991).

The LS method begins with the indirect utility function, Equation (1). Marginal willingness-to-pay for a change in Olympic medals resulting from the Own the Podium program is the marginal rate of substitution

$$
WTP_{LSM} = \frac{\partial v/\partial q}{\partial v/\partial y} \tag{7}
$$

The LS method assumes that statements related to satisfaction or happiness reflect ordinal (or cardinal in some cases) utility with a linear "utility" function

$$
LS = \beta_0 + \beta_1 q + \beta_2 y + \gamma' x \tag{8}
$$

where LS is the life satisfaction measure of utility. Theoretical validity tests would consider whether $\beta_1 > 0$ and $\beta_2 > 0$. Willingness-to-pay is

$$WTP_{LSM} = \frac{\beta_1}{\beta_2}. \tag{9}$$

Convergent validity is the extent to which estimates of an effect produced by two or more different methodologies are correlated. For example, if valuation estimates of improved medal success produced by CV and LS both increase, and if confidence intervals of the estimates overlap, then CV and LS display convergent validity, increasing the confidence in both valuation estimates. The null hypothesis for convergent validity is $H_0: WTP_{CVM} = WTP_{LS}$.

We use life satisfaction data where respondents state on a scale of one to ten the satisfaction with their current life situation, with ten being most satisfied. We assume cardinal ordering of life satisfaction measures are and use ordinary least squares to estimate the determinants.[3]

Reliability is the extent that willingness to pay is nonrandom. Reliability tests focus on variation in estimates within and across studies rather than whether studies accurately measure unbiased value. The lower the variability in estimates, the more consistent those estimates are and the less influenced by researcher decisions. High variability leads to the ability of researchers to significantly influence results in some direction. Unreliable results are indefensible against the criticism that the researcher is determining the result. There are several tests for reliability of the CV method including econometric and test-retest reliability. Econometric reliability is the ability to explain the overall variation in willingness-to-pay through observable variables included in the econometric specification. Econometric reliability can be tested with statistical measures of overall fit of a regression model. In a limited dependent variable model, a significant chi-squared statistic is a measure of econometric reliability.

Test-retest, or temporal, reliability is the consistency of willingness to pay estimated at two different times. Test-retest reliability involves conducting more than one CVM survey with time

[3] Similar results are found assuming that life satisfaction measures are ordinal and estimated with the ordered probit.

between surveys. The time period could be one month, one year or longer. If the magnitude of willingness to pay is consistent across time then willingness to pay is considered temporally reliable. Consistency in factors affecting willingness over time is also evidence of temporal reliability. However, a temporal difference in willingness to pay does not necessarily indicate unreliable results. If willingness to pay changes over time in response to changing factors that affect willingness to pay then the researcher may conclude that the survey results are reliable. The null hypothesis for temporal reliability is $WTP_1(x) = WTP_2(x)$ if the factors that affect willingness to pay remain the same or $WTP_1(x) = WTP_2'(x')$ if the factors change and willingness to pay in the second time period can be adjusted to hold those factors constant.

5. Empirical Results

We estimate four different CV and LS WTP models. We split the full sample into pre-Olympics and post-Olympics sub-samples and estimate separate models for each. We then split the full sample into sub-samples of respondents who answered both surveys and respondents who answered only one survey (either pre-Olympics or post-Olympics) and estimate panel and pooled models. Of the 2090 cases, 427 appear twice in the panel model. There are 1,936 respondents who appear in either the pre-Olympic or post-Olympic sample, including those who answered the survey twice but reported income only once.

Table 4 shows estimates from probit willingness-to-pay models. The coefficients on the tax amount variables are negative and statistically significant at conventional significance levels. The coefficients on the gold and other medals variables are each positive and statistically significant. In alternative models not reported here, the gold and other medals coefficients are constrained to be equal; the restriction is rejected in all but the post-Olympics model. This indicates that individuals value each gold medal more than each silver or bronze medal. Respondents who agree about the importance of winning Olympic medals are more likely to vote for the referendum. The income coefficient is positive and significant in each model. Income elasticities range from 0.77 to 1.13 and are not significantly different across models. Older respondents are more likely to vote for the referendum.

Other parameter estimates are not consistent across models. Males are more likely to vote for the referendum in the pre-Olympic and the panel models. Married respondents are less likely to vote for the referendum in the pre-Olympic and pooled models. Healthier respondents are more likely to vote for the referendum in the pre-Olympic model. In the panel model, respondents are no more likely to vote for the referendum after the Olympics when holding all other variables constant. In the pooled model, respondents who answered only the post-Olympic survey are more likely to vote for the referendum, although the coefficient is marginally significant.

Table 5 shows ordinary least squares LS model estimates. The only consistent results across the four models are for the parameter estimates on the income, health and marital status variables. Increases in income and improvements in health both increase LS. Married respondents report higher life satisfaction than non-married respondents in all four models. Surprisingly, employed respondents report lower life satisfaction in two of the four models. However, this result might be explained by considering that respondents would likely prefer not working holding income constant. Or, the result might reflect the impact of nonwage income. More educated respondents report lower life satisfaction in the post-Olympic and pooled models. Respondents who agree that earning Olympic medals is important report greater life satisfaction in the panel and pooled models. Males report lower life satisfaction in the panel and pooled models. Older respondents report greater life satisfaction in the post-Olympic model. Respondents who answered both surveys report greater life satisfaction. Respondents report higher life satisfaction with increases in gold medals in the post-Olympics and pooled models.

Table 6 reports the WTP estimates for gold and other medals. The CV estimates come from Equation (6). The LS estimates come from Equation (9). In the CV models, the willingness-to-pay for each additional gold medal ranges from $17 to $26. None of the differences are statistically different from zero since the 95% confidence intervals overlap for each pair of estimates. The willingness-to-pay for each additional silver or bronze medal ranges from $8 to $20. None of the differences are statistically different from zero. Estimates of willingness-to-pay for gold and other medals in each model are not significantly different. In contrast, the LS probit models' restrictions on equality were rejected in three out of four models because of the wide confidence intervals that result from constructing willingness-to-pay as the ratio of regression coefficients.

In contrast to the CV results, only one LS measure of welfare is statistically different from zero. The LS WTP estimate for an additional gold medal in the post-Olympic model is $14,094 per person. This estimate is substantially higher than the CV WTP estimate. We fail to find convergent validity for the WTP estimates across the CV and LS methods. We find a three-orders of magnitude difference in our only comparison between positive WTP estimates across the two methods. In a model that pools the survey respondents, we estimate a significantly lower, zero, WTP for gold medals from the LS method.

The CV estimates in Table 4 vary in expected directions with the price and scope of the policy. The income coefficients indicate that the Own the Podium program is a normal good. The income elasticity estimates are reasonable, based on those in the literature. The LS estimates in Table 5 also exhibit theoretical validity in terms of the estimated income and health effects, but not the scope effects for expected medals relative to the CV estimates. The estimated marginal utilities of income and health are positive and generally significant.[4]

Both the CV and LS estimates exhibit a basic degree of econometric reliability. The R^2, pseudo-R^2, chi-squared and F-tests from the empirical models indicate that a relatively large amount of the observed variation in the dependent variables is explained by the models, and that the explanatory variables in the models are jointly significant at conventional levels.

The CV estimates exhibit a high degree of temporal reliability. In a naïve model with only the tax amount and the POST variable indicating the post-2010 Games survey (results available upon request) we find that respondents are more likely to vote for the referendum after the Olympic Games. However, differences in medals expectations and opinions about medals before and after the Olympics can explain this result. After including these variables, temporal differences in referendum votes disappear in the panel model. CV WTP for medals estimates are not statistically different over time and the theoretically important regression coefficients are stable across models.

In contrast, while reported LS changes little between the two surveys, the temporal

[4] Models that impose a diminishing marginal utility of income functional form do not outperform the linear model presented here.

reliability of the LS method is questionable, since all regression coefficient estimates but two are inconsistently significant across models. Willingness-to-pay for a one unit change in health status is also highly variable with a range of $41,000 to $92,000. While none of the differences in health values are statistically significant, the magnitude of these estimates raises questions about the reliability of the LS (or the "face validity" of the method, where face validity refers to a measure reflecting the true underlying value).

6. Conclusions

Our results suggest the presence of potential limitations associated with using the LS method for valuing sports-related amenities. While the CV method passed several validity and reliability tests, the LS method performed worse. One concern in the current application is the lack of variation in LS method estimates of WTP with actual Olympic success, and success of the Own the Podium program. Team Canada was very successful at the 2010 Winter Games and the CV WTP estimates for medal success are positive and increase uniformly after the games. The LS WTP estimates are zero before the Games; the LS WTP for gold medal success estimate following the games is very large and positive while the LS WTP estimate for other medal success is very large and negative after the games.

Previous applications of the LS method used large data sets and spatial variation in environmental variables to value environmental quality or large data sets and variation in self-reported health status to value health outcomes. Applications in sports economics have used similar methods but have not attempted to value the sports-related amenities. Our results suggest that very large national surveys with spatial variation may be required to generate valid LS estimates of WTP.

Like previous applications that compared LS estimates to valuation estimates using revealed or stated preference methods, we find that the LS method generates unrealistically high estimates of the value of Olympic success, in terms of medals, relative to CV methods. Future research that addresses measurement of the marginal utility of income seems paramount in the LS valuation literature. While this problem is well-known and efforts are underway (Luechinger 2009,

Powdthavee 2011), LS research is presumptive in implicitly assuming the validity of unprecedentedly large estimates of environmental values.

Future sports applications should include the effects of professional team movement and construction of new stadiums and arenas on life satisfaction with large national datasets with spatial and temporal variation (e.g., General Social Survey). As emphasized here, future research that attempts to value nonmarket goods, amenities and individual health using the LS approach should consider an investigation of the validity and reliability of estimates. In particular, thorough assessment of the convergent validity of the LS method with revealed and stated preference methods should be a focus of future research.

References

Atkinson, Giles, Susana Mourato, Stefan Szymanski and Ece Ozdemiroglu, "Are we willing to pay enough to 'back the bid'? Valuing the intangible impacts of London's bid to host the 2012 Summer Olympic Games," Urban Studies 45(2):419-444, 2008.

Baade, Robert A., and Victor A. Matheson, "Going for the Gold: The economics of the Olympics," The Journal of Economic Perspectives 30(2): 201-218, 2016.

Cameron, Trudy Ann, and Michelle D. James, "Efficient estimation methods for 'closed-ended' contingent valuation surveys," Review of Economics and Statistics, 68, 269-276, 1987.

Cameron, Trudy Ann, "Interval estimates of non-market resource values from referendum contingent valuation surveys," Land Economics, 67, 413-421, 1991.

Carroll, Nick, Paul Frijters, Michael A. Shields, "Quantifying the costs of drought: new evidence from life satisfaction data," Journal of Population Economics: 22 (2): 445-461, 2009.

Carson, Richard T. and Theodore Groves, "Incentive and informational properties of preference questions," Environmental and Resource Economics 37(1):181-210, 2007.

Ferreira, Susan, and Mirko Moro, "On the use of subjective well-being data for environmental valuation," Environmental and Resource Economics 46:249-273, 2010.

Groot, Wim, and Henriëtte Maassen van den Brink, "A direct method for estimating the compensating income variation for severe headache and migraine," Social Science & Medicine 58(2): 305-314, 2004.

Groot, Wim, and Henriëtte Maassen van den Brink, "The compensating income variation of cardiovascular disease," Health Economics 15(10): 1143-1148, 2006.

Gu, Jiaying, and Roger Koenker, "Empirical Bayesball remixed: Empirical Bayes methods for longitudinal data," Journal of Applied Econometrics 32(3): 575-599, 2017.

Howley, Peter, "Less money or better health? Evaluating individual's willingness to make trade-offs using life satisfaction data," Journal of Economic Behavior & Organization 135: 53-65, 2017.

Johnson, Bruce K., and Whitehead, John C., "Contingent valuation of sports," in The Oxford Handbook of Sports Economics, Volume 2: Economics Through Sports, edited by Stephen Shmanske and Leo H. Kahane, 2012.

Johnson, Bruce K., Brad R. Humphreys, Daniel S. Mason, and John C. Whitehead, "Estimating the value of medal success at the 2010 Winter Olympic Games," Journal of Sports Economics, forthcoming.

Kavetsos, Georgios and Stefan Szymanski, "National well-being and international sports events," Journal of Economic Psychology, 31(2): 158–171, 2010.

Kountouris, Yiannis and Kyriaki Remoundou, "Valuing the welfare cost of forest fires: A life satisfaction approach," Kyklos 64(4):556-578, 2011.

Levinson, Arik. "Valuing public goods using happiness data: The case of air quality." Journal of Public Economics 96(9): 869-880, 2012.

Luechinger, Simon, "Valuing air quality using the life satisfation approach," The Economic Journal, 119 (March), 482–515, 2009.

Luechinger, Simon, and Paul A. Raschky. "Valuing flood disasters using the life satisfaction approach." Journal of Public Economics 93(3): 620-633, 2009.

Loomis, John, "What's to know about hypothetical bias in stated preference valuation studies?" Journal of Economic Surveys 25(2):363-370, 2011.

Maennig, Wolfgang and Stan Du Plessis, "World Cup 2010: South African economic perspectives and policy challenges informed by the experience of Germany 2006," Contemporary Economic Policy 25(4):578-590, 2007.

Matheson, Victor, and Robert A. Baade, “Mega-sporting events in developing nations: Playing the way to prosperity?” The South African Journal of Economics 72(5):1085-1096, 2004.

Siegfried, John and Andrew Zimbalist, “The economic impact of sports facilities, teams and mega-events,” The Australian Economic Review 39(4):420-427, 2007.

Siikamäki, Juha, and Douglas M. Larson. "Finding sensitivity to scope in nonmarket valuation." Journal of Applied Econometrics 30, no. 2 (2015): 333-349.

Süssmuth, Bernd, Malte Heyne, and Wolfgang Maennig, “Induced civic pride and integration,” Oxford Bulletin of Economics and Statistics 72(2):202-220, 2010.

Walton, Harry, Alberto Longo and Peter Dawson, “A contingent valuation of the 2012 London Olympic Games,” Journal of Sports Economics 9(3):304-317, 2008.

Welsch, Heinz, “Environment and happiness: Valuation of air pollution using life satisfaction data,” Ecological Economics 58:801–813, 2006.

Welsch, Heinz, “Implications of happiness research for environmental economics,” Ecological Economics 68(11): 2735–2742, 2009.

Welsch, Heinz and Kühling, Jan, “Using happiness data for environmental valuation: issues and applications,” Journal of Economic Surveys 23(2): 385–406, 2009.

Whitehead, John C., and Timothy C. Haab, Contingent Valuation Method, in Encyclopedia of Energy, Natural Resource, and Environmental Economics, ed. Jason Shogren, Elsevier, 2013.

Wicker, Pamela, John C. Whitehead, Daniel S. Mason, and Bruce K. Johnson, “Public support for hosting the Olympic Summer Games in Germany: The CVM approach,” Urban Studies, doi: 10.1177/0042098016675085

Table 1: Summary Statistics

	Pre Olympics		Post Olympics	
Variable	Mean	Std.Dev.	Mean	Std.Dev.
BOTH	0.46	0.50	0.37	0.50
FOR	0.59	0.49	0.60	0.49
SATISFY	2.48	1.35	2.57	1.34
TAX	31.88	19.99	55.56	35.19
GOLD	11.75	6.21	15.54	2.87
OTHMEDAL	18.20	6.01	13.40	2.94
AGREE	15.47	5.25	11.03	3.36
INCOME	69.83	42.11	72.05	41.77
MALE	0.54	0.50	0.52	0.50
EDUCATION	14.47	2.41	14.51	2.39
MARRIED	0.65	0.48	0.70	0.46
AGE	49.35	15.29	50.48	14.49
HEALTH	8.09	1.64	8.13	1.56
EMPLOYED	0.69	0.46	0.68	0.47
Cases	926		1164	

Table 2: Willingness-to-pay Frequencies

	Pre Olympics Survey			Post Olympics Survey		
Tax	For	Total	% For	For	Total	% For
5	125	193	65%	--	--	--
10	6	9	67%	--	--	--
15	--	--	--	104	141	74%
20	140	214	65%	--	--	--
25	--	--	--	140	202	69%
30	4	7	57%	--	--	--
35	117	201	58%	119	191	62%
50	94	179	53%	113	199	57%
65	58	123	47%	11	18	61%
75	--	--	--	101	180	56%
100	--	--	--	83	175	47%
150	--	--	--	27	58	47%
Total	544	926	59%	698	1164	60%

Table 3: Life Satisfaction Frequencies

	Pre Olympics		Post Olympics	
Satisfy	Frequency	Percent	Frequency	Percent
4	303	33%	410	35%
3	154	17%	213	18%
2	252	27%	292	25%
1	116	13%	127	11%
0	101	11%	122	10%

Note: The respondent's overall life satisfaction is reported on scale of 1 (low) to 10 (high). Due to the relative infrequency of responses from 1 to 5, we pool these with 6 and recode responses from a low of 0 to 4.

Table 4: Probit Willingness-to-pay Models

	Pre-Olympics		Post-Olympics		Panel (Both=1)		Pooled (Both=0)	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Constant	-3.865	-7.23	-4.954	-8.89	-5.349	-6.04	-3.913	-8.39
TAX	-0.007	-2.93	-0.006	-5.33	-0.009	-3.47	-0.006	-5.10
GOLD	0.112	10.60	0.157	8.97	0.198	8.86	0.112	9.96
OTHMEDAL	0.054	5.22	0.120	7.24	0.108	6.77	0.068	6.08
AGREE	0.026	2.71	0.065	5.04	0.063	3.66	0.029	2.94
INCOME	0.007	4.95	0.008	6.17	0.009	3.92	0.008	6.54
MALE	0.162	1.75	0.018	0.22	0.352	2.23	-0.025	-0.32
MARRIED	-0.181	-1.73	-0.125	-1.29	-0.153	-0.79	-0.201	-2.25
EDUCATION	-0.004	-0.19	-0.014	-0.77	-0.037	-1.07	-0.008	-0.46
AGE	0.016	4.48	0.014	4.25	0.014	2.41	0.016	5.50
EMPLOYED	0.111	0.93	0.082	0.80	0.022	0.12	0.150	1.48
HEALTH	0.052	1.76	-0.019	-0.70	0.013	0.26	0.022	0.90
BOTH	0.135	1.44	0.022	0.26	--	--	--	--
POST	--	--	--	--	0.068	0.40	0.195	1.65
ρ	--	--	--	--	0.493	5.92	--	--
McFadden's R^2	0.200		0.181		0.033		0.164	
Chi-squared	250.68		284.21		29.45		275.78	
Cases	926		1164		854		1236	

Table 5: OLS Life Satisfaction Models

	Pre Olympics		Post Olympics		Panel (Both=1)		Pooled (Both=0)	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Constant	-.0254	-0.06	-0.525	-1.16	6.48	10.04	5.625	0.40
GOLD	0.002	0.22	0.044	3.26	0.0048	0.49	0.016	1.79
OTHMEDAL	-0.003	-0.41	-0.008	-0.61	-0.013	-1.46	0.038	0.47
AGREE	0.009	1.06	-0.004	-0.37	0.017	1.41	-0.006	-0.62
INCOME	0.006	5.45	0.003	2.88	0.006	4.49	0.0038	3.56
MALE	-0.030	-0.36	-0.082	-1.12	-0.172	-1.78	0.019	0.26
MARRIED	0.281	3.01	0.163	1.85	0.229	1.92	0.198	2.37
EDUCATION	-0.021	-1.15	-0.031	-1.87	-0.039	-1.76	-0.201	-1.26
AGE	0.002	0.71	0.007	2.56	0.003	0.73	0.0068	2.47
EMPLOYED	-0.278	-2.61	-0.078	-0.83	-0213	-1.75	-0.121	-1.28
HEALTH	0.260	10.00	0.285	11.70	0.263	7.24	0.277	11.98
BOTH	0.174	2.11	0.207	2.72	--	--	--	--
POST	--	--	--	--	0.066	0.64	0.0022	-0.02
R^2	0.178		0.148		0.168		0.145	
F	18.02		18.15		15.50		18.87	
Cases	926		1164		854		1236	

Table 6: Willingness-to-pay Estimates

	Contingent Valuation							
	Pre-Olympics		Post-Olympics		Panel (Both=1)		Pooled (Both=0)	
	WTP	S.E.	WTP	S.E.	WTP	S.E.	WTP	S.E.
GOLD	16.56	5.91	25.60	5.49	21.05	5.90	17.94	3.94
OTHER MEDAL	8.00	3.09	19.53	4.41	11.45	3.42	10.99	2.75

	Life Satisfaction (in 1000s)							
	Pre-Olympics		Post-Olympics		Panel (Both=1)		Pooled (Both=0)	
	WTP	S.E.	WTP	S.E.	WTP	S.E.	WTP	S.E.
GOLD	0.25	1.12	14.09	6.70	0.82	1.59	4.16	2.67
OTHER MEDAL	-0.47	1.13	-2.51	4.18	-2.35	1.66	1.02	2.20