A Comparison between the Traditional Contingent Valuation Methodology and Prediction Mechanism

Lava P. Yadav*, Thomas H. Stevens†, and James J. Murphy‡

University of Massachusetts, Amherst

March 2007

Job Market Paper

Abstract: *The traditional contingent valuation methodology (CVM) asks individuals to provide their personal opinions or values for some commodity. This stated preference format has resulted in the continual appearance of hypothetical bias. Literature in psychology suggests that such bias is prominent when individuals are asked to provide their own values, but significantly diminishes when asked to predict others' values. In this study we test to see if hypothetical bias disappears when traditional 'willingness to pay' questions are replaced by prediction based questions. According to the results, we observe hypothetical bias in both the traditional CVM and the prediction mechanism. Although prediction estimates were significantly lower than self-estimates, it was observed that individuals are able to correctly predict the magnitude of hypothetical bias in the traditional CVM.*

Key Words: contingent valuation, hypothetical bias, prediction mechanism

JEL codes: Q5, C91, C92, H41

---

*Lava P. Yadav, Department of Resource Economics
University of Massachusetts, Stockbridge Hall, 80 Campus Center Way, Amherst, MA 01003
E: lava@resecon.umass.edu P: 315-559-4271 F: 413-545-5853

† Thomas H. Stevens, Department of Resource Economics
University of Massachusetts, Stockbridge Hall, 80 Campus Center Way, Amherst, MA 01003
E: tstevens@resecon.umass.edu P: 413-545-5714 F: 413-545-5853

‡ James J. Murphy, Department of Resource Economics
University of Massachusetts, Stockbridge Hall, 80 Campus Center Way, Amherst, MA 01003
E: murphy@resecon.umass.edu P: 413-545-5716 F: 413-545-5853

1. Introduction

The traditional contingent valuation methodology (CVM) typically requests individuals to state their hypothetical payments for hypothetical provision of a non-market good. There is substantial evidence that these hypothetical values often exceed actual payments (Harrison and Rutstrom 2002; List and Gallet 2001; Murphy, et al. 2005). In recognition of this problem, the National Oceanic and Atmospheric Administration (NOAA) guidelines state that contingent valuation (CV) value estimates should be halved unless adjusted using some form of calibration technique (Arrow, et al. 1993). Several attempts to elicit unbiased responses through the use of calibration techniques such as cheap talk scripts (Cummings and Taylor 1999), uncertainty adjustments (Champ and Bishop 2001; Poe, et al. 2002), and statistical bias functions (Hofler and List 2004) have been made. However, none have succeeded in providing convincing results.

Theories in the psychology literature regarding human behavior suggest that the format of the CV technique itself may be responsible for inducing various biases including hypothetical bias.[1] It is suggested that the methodology places the respondent under the 'spot light' or in psychological terms, under the 'locus of control' (Rotter 1954). While this causes stated values to be vulnerable to various types of personal biases, some respondents may also be motivated to fabricate them. However, if the locus of control is shifted, that is, if the respondent is asked to make a prediction of other individual's (or group of individual's) behavior, then the response is less likely to be biased, hence providing his/her true predictions.

[1] Hypothetical bias is defined as the difference between stated and actual willingness to pay (WTP) for some commodity.

According to the ‘false consensus hypothesis,’ predictions are in fact representations of the respondent’s personal values. So in theory, prediction estimates are representations of an individual’s personal values without many of the biases.[2] In addition, the process of averaging individual predictions tends to cancel positive and negative prediction errors (Surowiecki 2004; Chen, et al. 2005; Lusk and Norwood 2005) and thus the outcome is a bias free value estimate.

There are however several possibilities that the line of reasoning favoring the prediction mechanism overlooks. Although it is theorized that personal biases are ignored when placing predictions, the false consensus theory suggests otherwise. According to this theory individuals believe their own preferences and values as a strong representative of others’. Thus several biases encountered with the traditional CVM, including hypothetical bias, could be carried over to prediction estimates. Furthermore, it has been observed that individuals tend to underestimate the ‘moral goodness’ of other people while overestimating their own (Epley and Dunning 2000). Hence in addition to including biases pertaining to the CVM, the predicted values may also be underestimated.

In this paper we compare the traditional CVM with a prediction format to test if prediction estimates are effective against hypothetical bias. In a previous study conducted by Lusk and Norwood (2005), hypothetical bias was successfully eliminated using a prediction format. However, according to our results, in addition to the existence of hypothetical bias in both formats, the prediction estimates were observed to be significantly lower. Furthermore, the analysis of the survey questionnaire revealed a

[2] We define prediction estimate to be a predicted value or behavior of another individual or a group of individuals.

disparity in the thought processes between individuals participating in the two valuation formats.

2. Hypothetical Bias

Hypothetical bias is defined as the difference between stated and actual willingness to pay (WTP) for some commodity. Since Bohm's 1972 paper which compared actual values with hypothetical values, a considerable amount of literature about hypothetical bias has been published. Meta-Analysis conducted by Harrison and Rutstrom (2002), List and Gallet (2001), and Murphy, et al. (2005) all reveal the continual existence of hypothetical bias.

Examples of studies that have demonstrated the existence of hypothetical bias for provisions of public goods include Cummings and Taylor (1999), Champ, et al. (1997), Cummings, et al. (1997), Bishop and Heberlein (1979), and Bishop and Heberlein (1986). Cummings and Taylor (1999) focused on funding a Nature Conservancy Project to protect the rain forest in Costa Rica. Champ, et al. (1997) examined contributions to remove dirt roads in the north rim of the Grand Canyon and Cummings, et al. (1997) valued a bilingual (English and Spanish) citizen's guide about contaminated ground water for residents of Albuquerque, New Mexico.

Several factors may be responsible for the hypothetical bias observed in these and similar studies. One possible explanation is that some respondents pay low "actual" values in an attempt to free ride on others (Carson, et al. 1999). This result has been demonstrated empirically by Seip and Strand (1992) and Champ, et al. (1997). On the other hand, others may overstate their WTP in hypothetical situations to increase the probability of provision of the good (Harrison and Rutstrom 2002). Also, in hypothetical situations, some may overestimate their WTP because they fail to consider their budget constraints (Brown, et al. 2003) while some may view the valuation question simply as a game (Bredhal, et al. 2003). Moreover, many respondents may not answer the question being asked. Instead, these respondents may state their opinion regarding whether the good being valued is worthy of support (Brown, et al. 2003). List, et al. (2004) observed that some subjects seek social approval in hypothetical situations and therefore overestimate what they are actually willing to pay.

Recognizing hypothetical bias as a major drawback of the CVM, the NOAA panel guidelines stated that values obtained using the CVM should be halved unless they can be calibrated in some way. This resulted in the development and experimentation with several types of calibration techniques including cheap talk scripts (Cummings and Taylor 1999), uncertainty adjustments (Champ and Bishop 2001; Poe, et al. 2002), and statistical bias functions (Hofler and List 2004). These attempts have provided very mixed results.

Cummings and Taylor (1999) were able to successfully eliminate hypothetical bias using the cheap talk script, however, Murphy, et al. (2005) and Brown, et al. (2003) show that this is only effective for subjects facing relatively high payments. List (2001) found that cheap talk reduced hypothetical bias associated with some subjects but was ineffective for subjects who had experience in the market. Lusk (2003) and Aadland and Chaplan (2003) reach a similar conclusion.

Champ, et al. 1997, Champ and Bishop 2001, Ethier, et al. 2002, and Poe, et al. 2002 have used uncertainty treatments where uncertain WTP ‘yes’ responses are recoded as ‘no’. Although effective in several situations, the cutoff point at which this practice works is arbitrary and varies among studies. Moreover, Wang (1997) found that treating uncertain responses as ‘no’ can understate mean WTP.

Although statistical calibration functions that use laboratory experiments involving real payments to calibrate field data show promise (see, for example, Fox, et al. 1988), these functions are often thought to be commodity, context, and/or individual specific (List and Shogren 2002). In conclusion, thus far, continual inconsistent results have led to no firm consensus regarding a reliable calibration technique.

3. Background and Literature Review

The literature on CVM has focused almost exclusively on the study and development of new ways to overcome various kinds of biases including hypothetical bias. While many measures have been adopted and several types of calibration techniques have been developed to mitigate such biases, the underlying cause and nature of the biases have not been well understood. The literature in psychology however, provides some insights into several aspects of human nature that could be responsible for the continual appearance of some of these biases. To the extent that these biases are conscious strategic responses, it may be possible to design an incentive compatible mechanism that discourages intentional fabrication of responses. However, it is also possible that these biases are unconscious acts and are hence beyond the control of even the primary actors.

While trying to understand the various sources of bias, researchers seem to have overlooked a fundamental assumption upon which the CVM is based. According to this assumption, individuals have consistent preference orderings and that errors are simply ‘measurement biases’ which can be adjusted (Vatn 2004). Hence the problem related to the CVM can be considered to be twofold. The first is related to whether people have well defined preference orderings and whether they are capable of using this knowledge to accurately place a value on a commodity. Second, even if preferences are well defined, it may be very difficult to elicit them without bias.

According to a study by Loewenstein and Adler (1995), individuals often fail to forecast their own behavior and tastes, which suggest that under some circumstances people may not know, or be able to accurately state their own preferences. A study by Tat, et al. (1988) shows that although a consumer’s decision to purchase some commodity is heavily dependent upon rebate offers, very few of them end up sending in the rebate forms. Ausubel (1991) shows that a large number of people expect to maintain a zero credit card balance, but very few manage to do so. According to Sherman (1980), inaccurate self-estimates could be a result of ambitious goals set by individuals to abide by social or moral norms.[3] It could also be a result of an attempt to impress the researcher or to avoid embarrassment. Even when people are extremely confident about self-estimates it is still possible that they do not comply with their stated intentions. In a study by Vallone, et al. (1990), which focused on the confidence level of individuals, although some individuals were a hundred percent certain about performing a particular behavior, it was performed only 77 percent of the time.

When people are asked to place values on goods that are not familiar to them, they struggle to understand the type of information that is presented and try to relate this

[3] We define a self-estimate to be an estimate of one’s own behavior or value. The value stated by an individual in a tradition CVM format is considered as a self-estimate.

information to previous personal encounters to construct a value for it (Vatn, 2004). It is at this point that even without the intention of fabricating opinions and values, respondents may often provide a biased response. For example, without knowing, some individuals may state a bid that is higher than their true bid in an attempt to impress the interviewer or because it is socially or morally acceptable. Similarly, certain aspects of the CVM such as a poorly designed survey is capable of picking up numerous biases even when the respondent has no intention of disguising his/her answers.

CVM practitioners often assume that ‘true preferences are there just to be uncovered’, and when faced with a choice, it is assumed that the response of a rational individual will be geared towards maximizing his/her own utility. The drawback however is that a utility maximizing response is not always a truthful response. The true answers can be fabricated by individuals to deliberately misrepresent their values in order to achieve a higher utility level. For example, with the intention of manipulating public policies that depend upon the survey results, the individual may be inclined to provide strategic responses.

It is often believed that the primary source of hypothetical bias is because the individual’s utility level is unaffected in an inconsequential CV survey (Carson, et al. 2004). So one challenge is to develop an incentive compatible survey that induces individuals to reveal their true values without requiring real payments. Although some survey designs have accomplished higher levels of incentive compatibility, many have not (Burton, et al. 2003).

The probability an individual engages in a particular behavior depends upon the value of the outcome to the individual and the probability that the particular behavior will lead to the desired outcome (Rotter 1954). Hence in the case of a CVM, this theory can be interpreted as, ‘the probability an individual uses a strategic response in a survey is a function of the level of increase in the individual’s utility from that particular strategy and the probability that the strategy would lead to the particular payoff or utility.

This theory can be written as

$$B = f(V, E) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equation (1)}$$

Where B is the probability of a strategic behavior, V is the payoff or the outcome and E is the expected probability that the particular behavior will lead to the desired outcome.

According to Rotter (1966), when an individual is asked to make a self-estimate, the question is directed towards the individual’s own preferences and values which cause the individual to expect a high value of E. In the psychology literature, this phenomenon is termed as ‘locus of control of reinforcement’ as labeled by Rotter (1966), or simply ‘locus of control.’

When the ‘locus of control’ is shifted away from the individual, there is a reduction in the perceived magnitude of E. According to equation (1), when the payoff V is constant, a decrease in E will lead to a lower probability of a strategic behavior. By switching the ‘self-estimate’ question to a question that concerns someone else, the incentive for strategic behavior may be reduced. So by replacing the traditional CVM with a value elicitation technique that asks individuals to predict other’s behavior rather than their own could be a strategy for achieving higher levels of incentive compatibility.

Higher levels of incentive compatibility may be achieved using the prediction based mechanism, however, several other biases including hypothetical bias are not exclusively due to the lack of incentive compatibility. So, how would the prediction mechanism perform in comparison to the traditional CVM when dealing with these sources of bias?

'Symbolic bias' occurs when a respondent doesn't answer the question being asked using his/her own preferences but instead in a manner that is morally or socially acceptable (Vatn 2004; Blamey 1998). The respondent may not intend to misrepresent her answer; however, because the preferred behavior according to society is different from what she believes, subconsciously she may be driven by social or moral norms. Essentially, when one is placed under the 'spot light', the feeling of being judged leads to a misrepresentation of actual values or opinions. This behavior may be reduced when the question is directed towards someone else.

Hoevenagel (1990) discusses the existence of a bias known as 'mental bias', which is mostly associated with environmental and charitable goods. The individual has an estimate that is formulated in advance regarding how much he/she plans to allocate for charitable or environmental goods. Hence, the bid placed by the individual may not represent his/her actual value but rather be dependent upon previous expenditures on goods of similar nature. Because 'Mental Bias' is related to one's personal budget, a prediction based question, which diverts the focus away from the individual may reduce this bias.

Thaler (1980) raises an issue regarding the 'endowment effect' which leads to inconsistencies in human behavior and is capable of distorting CVM values. The 'endowment effect' refers to a higher value placed on a commodity if the individual owns it as oppose to when it is not owned. This effect can be demonstrated by a simple experiment in which there are two groups of subjects; a group who are endowed with the commodity and are provided with the option of selling it for various amounts of cash, and another group that is not endowed with the commodity but is given the opportunity to choose between the commodity and various amounts of cash. The results are quite recurrent wherein the group that is endowed is willing to sell the commodity for an amount that is higher than the amount at which the individuals not endowed value it (Kahneman, et al. 1991; Loewenstein and Adler 1995). This phenomenon of endowment effect as evidence of loss aversion or prospect theory is challenged by Plott, et al. (2005) who successfully avoid such an effect in their experiment. According to Plott, et al. (2005) the endowment effect is a result of subject misconceptions which can be controlled through careful measures.

Marshall, et al. (1986) took these experiments a step further where they asked individuals to act as agents making recommendations on behalf of other individuals. Instead of reporting their own individual selling prices of the commodity in the event that they owned the commodity, the individuals were asked to act as an agent for some other individual that was endowed and asked to recommend the appropriate selling price for the commodity. Similarly, the individual was asked to act as an agent to an individual who was not endowed and to make recommendations in terms of choices between the commodity and various cash amounts. The disparity in the valuation of the commodity was highly significant when individuals acted for themselves, but this disparity disappeared when they were placing recommendations for others. Furthermore, when an identical experiment was repeated with real payments, the results were consistent with the hypothetical ones.

Marshall, et al. (1986) claim that this observation is a result of the disappearance of involuntary discriminatory treatment of gains and losses when making decisions on behalf of others. People are not aware that others may also be attached to objects that they own. It can be hypothesized that the manner in which individuals are prone to hypothetical bias is similar to the endowment effect. Stating higher values in a hypothetical treatment may not be intentional; instead it may be a phenomenon that occurs involuntarily. If individuals are asked to state or predict values placed by other individuals between the real and hypothetical treatments, they may not see the rationale of diverging values between the two treatments. Hence, in theory, the prediction mechanism may have an advantage over the traditional CVM in terms of its potential to mitigate hypothetical bias as well as other biases.

The ultimate goal of a non-market valuation technique is to assess the total value of a commodity by estimating individual values. Unlike the CVM which directly asks people for their personal values, the prediction mechanism inquires about others' values. So is there a link between an individual's true value for a commodity and his/her prediction of others' value?

According to the 'false consensus theory', there is a direct relation between predicted values and personal values (self-estimates) or opinions.

"False consensus refers to an egocentric bias that occurs when people estimate consensus for their own behaviors. Specifically, the false consensus hypothesis holds that people who engage in a given behavior will estimate that behavior to be more common than it is estimated to be by people who engage in alternative behaviors." (Mullen, et al. 1985)

‘False consensus’, which was first introduced by Ross, et al. (1977) has been observed in several psychological studies and experiments. A Meta analysis conducted by Mullen, et al. (1985) reports 115 studies with the false consensus effect. A consistent occurrence of the false consensus effect has been observed in an extensive range of preferences and beliefs. For example, questions and issues related to human behavior, opinions regarding various policies, politics, etc have all given results predicted by the ‘false consensus theory’. In a study conducted by Brown (1982), 179 psychology students were asked to choose between Anderson, Carter, or Reagan as their preferred candidate in the 1980 presidential election. Following these choices they were asked to estimate the share of students in the class that they believed indicated their preferences for each of the candidates. For all three of the candidates, the estimations were higher for the candidates of personal choice. A study by Katz and Allport (1931) demonstrated that the students who admitted to cheating predicted a higher percentage of other students that cheated.

Individual values and opinions play an important role when making predictions about others (Dawes 1989); when they are not provided with any information regarding other’s behavior, they consider their own preference to be a strong representation of the general public. Hence personal values and opinions are expressed through predictions of others.

The prediction mechanism is not a new concept; rather, its potential has been well recognized and employed in various areas. Prediction estimates have been used extensively in prediction markets such as the Iowa Electronic Markets. Since its establishment in 1988, it has constantly outperformed the results of polls in terms of political elections (Berg, et al. forthcoming; Forsythe, et al. 1992). Prediction markets, also referred to as information markets are designed to aggregate information and forecast future events. The mechanism behind opinion polls is similar to that of the CVM wherein they both ask individuals for their self-estimates (how they would act or respond themselves). As prediction markets seem to have done better than opinion polls, it shouldn't be a surprise that the prediction mechanism might also outperform the traditional CVM.

Once the predictions are obtained, the simple process of averaging the individual estimates is known to produce an accurate value estimate. Taking an arithmetic average, although very simple, has proved to be quite robust in opinion polls (Chen, et al. 2005). Lusk and Norwood (2005) also find that a simple average of stated predictions perform better than the prediction market mechanism. According to Surowiecki (2004) when prediction estimates are aggregated, the errors cancel out and all that is left is almost complete information. Surowiecki (2004) also emphasizes the importance of diversity along with individuality in the sample for aggregate results to be valuable.

The proponents of the prediction format seem to have strong arguments favoring the prediction mechanism over the traditional CVM. However, their claim for its superiority based upon the ability of predictions to evade several types of biases including hypothetical may have overlooked other possibilities. For example, the 'false consensus theory' discussed earlier, provides an explanation why the prediction estimates may in fact be similar to self-estimates of the traditional CVM in terms of its vulnerability to biases including hypothetical bias. The false consensus theory suggests that we consider most people to be similar to ourselves. So when asked to predict how other people perform, we tend to put ourselves in their shoes and think about how we ourselves would have responded if placed in a similar situation. Mental processing in this manner may then place the individual making a prediction back to the state where he/she is asked to provide a self-estimate and as a result, biases pertinent to the traditional CVM estimates such as hypothetical bias could also make its way into prediction estimates.

Another factor to consider, according to Epley and Dunning (2000), is that people often tend to underestimate the 'moral goodness' of others, and overestimate their own. Several experiments that involved contributing time or money to various causes showed that most individuals felt that others have lower moral values and are less generous than themselves (Goethals, et al. 1991; Allison, et al. 1989; Alicke 1985; Epley and Dunning 2000; Goethal 1986). Epley and Dunning (2000) in their experiment observed that 83% of the students predicted that they would buy at least one daffodil flower as a contribution to the annual campus day charity while they expected only 56% of their peers to do the same. Goethals, et al. (1991) showed that people believed they were more likely than their peers to give up their seat on a crowded bus for a pregnant woman. Another study by (Goethals 1986) showed that 60% of people said they would be willing to donate blood at a local charity if given a chance, while they believed only 39% of their peers would feel the same.

This is sometimes termed as the 'lake wobegon effect' or the feeling of being 'holier than thou'. According to Epley and Dunning (2000), individuals are excessively optimistic about their own behavior and consider themselves to be a member of 'an elite moral minority'. Although we have an idea of how other people normally behave in certain conditions and situations, when estimating our own behavior we tend to believe that such rules and patterns do not apply to the self. It appears that the feeling of being ‘holier than thou’ is due to an underestimation of others’ moral goodness and an overestimation of one’s own.

The behavior of individuals may include both the false consensus effect and the lake wobegon effect. That is, when asked to predict the behavior of others, many people may report what they themselves would have done in the same circumstance (the consensus effect), but, at least in a contribution setting, they might then subtract a little to account for being “above average” (the wobegon effect). On the other hand, by shifting the locus of control, the prediction format may reduce some types of strategic behavior, and this tendency might be reinforced if respondents are paid for making accurate predictions.

Providing monetary incentives for correct prediction estimates is not so simple for a setting where the true value is unknown. Prelec (2004) uses an ‘information scoring’ method designed to elicit truthful answers. In his experiment, the prediction question is preceded by a question asking the participants their own opinion and values to determine the accuracy of prediction estimates. The drawback of this strategy is that once the individuals possess the ‘locus of control’ when providing their own estimates, it is possible that the biases formed will get carried over to the prediction estimates. So how important is it to provide monetary incentives for eliciting accurate predictions?

Servan-Schreiber, et al. (2004) examined the consequence of monetary incentives on two information markets and concluded similar prediction accuracies between the two. Several other experiments however have shown that laboratory markets do not work well when the opportunity costs of making errors are very low (Harrison 1989). In a study conducted by Offerman, et al. (1996), a post-experimental questionnaire revealed that

50% of the individuals would have provided different answers if monetary incentives were not provided. Heijden, et al. (2003) in their study observe a reduction in the false consensus effect when monetary incentives are provided. Furthermore, a set of guidelines created by Vernon Smith (1982) to assist in conducting laboratory market experiments recommends salient monetary payoffs in order to elicit true preferences.

4. Experimental Design

The experiments were conducted using a 2 * 2 treatment design in which all sessions were conducted at the University of Massachusetts, Amherst. Almost all of the 193 participants were undergraduate students at the university. Participants were informed about the experiments by flyers that were placed around campus and also a recruiting table was set up in the campus center. They were informed about the amount of time the experiment would last (not more than 90 minutes) but were not given any information about possible earnings in any of the experiments. In the beginning of the experiment, each participant was paid a participation fee of $5 and provided with a copy of the instructions which was also read aloud. Each participant started with a $10 initial cash balance (in addition to the initial $5 participation fee) and was given the opportunity to withdraw from the experiment at any time.

A within-subject design was used where each experimental session began with an induced value experiment followed by a homegrown public good experiment. The same individual that participated in the induced value experiment was also asked to value a homegrown public good. In the induced value rounds, the induced value good was described to fit the nature of a public good.[4] The instructions were followed by two practice rounds so as to familiarize participants with the Becker-DeGroot-Marschak (BDM) auction mechanism that was used in all our experiments.[5]

The Becker-Degroot-Marschak (BDM) auction mechanism is one of the most widely used demand-revealing mechanisms in experimental economics. The BDM auction that we used worked in the following manner.

1. Each subject placed an auction bid between $0 and $15 for the commodity presented.
2. After the bids were collected we determined the 'project cost' by drawing a ball randomly from a basket containing 15 balls numbered 1 through 15.
3. The bids were collected and tallied. If more than 50% of the participants bid less than the randomly selected 'project cost' no one paid or received anything.
4. If more than 50% of the participants bid greater than or equal to the randomly selected 'project cost' then each participant paid the 'project cost' and received the 'project benefit' regardless of how he/she voted.

In theory, the BDM mechanism is incentive compatible (Becker, et al. 1964), and various studies have been conducted regarding its incentive compatibility in practice. "The maintained hypothesis is that the BDM is so transparent and involves so little cognitive effort that subjects will bid optimally, even when the rewards for doing so are small" (Irwin, et al. 1998). Hence a steep payoff schedule is not necessary to induce individuals to make optimal decisions. On the contrary, because the mechanism is highly transparent and the optimal strategy seems very obvious, some subjects may develop

[4] Although induced value goods were imaginary goods, they were described to fit the nature of a homegrown public good that followed it in the experiments.

[5] A modified version of the BDM mechanism was used in the experiments that involved a referendum mechanism.

doubts regarding their best strategy. This can be avoided by training subjects to increase their familiarity with the mechanism. Hence two practice rounds were conducted before each experiment.

One of the major criticisms of the BDM mechanism has been its reliance on the Independence Axiom of Expected Utility Theory (Irwin, et al. 1998).[6] Originally the BDM mechanism was designed so that subjects faced uncertain payoffs. They faced a 'Compound Lottery', which meant that they were purchasing lottery tickets with an uncertain payoff scheme. In our experiments, the commodities used did not have an uncertain payoff, hence this problem was eliminated. There is, however, another potential weakness of this mechanism. "The choice and knowledge of the upper and lower bounds of the randomly generated buyout or sellout prices are capable of distorting the incentive-compatibility of the BDM" (Bohm, et al. 1997). According to Bohm, et al. (1997), despite the fact that the participants are instructed that revealing the truth is the best strategy, providing them with extra information regarding the lower bound (in our case) of the 'selling price' may provide them with an incentive to act differently. In the experiment conducted by Bohm et al. (1997), where $20 gift certificates were given to participants, 25% of them stated a minimum selling price of $20 or more when they were told that the possible range of buyout prices was between $0 and $30. The fact that the range of possible buyers exceeded the $20 mark instigated the sellers to bid higher selling prices. There is also a possibility that higher minimum selling prices in Bohm's study were a result of an endowment effect.

[6] The Independence Axiom of Expected Utility Theory states that a risky prospect A is preferred to another risky prospect B if and only if the combination [A, C; p] is preferred to the combination [B, C; p] for all p and C. A combination [A, C; p] denotes a risky prospect which yields A with probability p and C with probability (1-p).

As can be seen in Table 1**,** for the first two sets of experiments (1a and 1b) participants were required to state their own level of contribution (Self-Estimates) to Heifer International to provide some needy families with a flock of chickens. In the third set of experiments the participants were asked to provide their prediction of the average level of contribution provided by the individuals in the first two sets of experiments. Before each of these experiments was conducted, there were two practice rounds to familiarize participants with the mechanisms of the experiment.[7]

4.1 Experiment 1 - Self-Estimation Experiments

The first stage of this experiment involved an imaginary group project where each participant was assigned a benefit from a project if it was implemented. The amount of benefit (project benefit) each subject received if the project was implemented was provided to them in monetary terms. If the project was implemented then they would receive a cash equivalent of their project benefit at the end of the experiment. This first stage consisted of two rounds in which each subject was assigned an induced value of $4.50 in the first round and $11.50 in the second.

After the first round was over but before the 'project cost' was randomly determined, we proceeded to round two. Following both induced value rounds, but before the 'project costs' for either of these rounds were declared, respondents were given another set of instructions for stage two. The second stage involved an actual public good which was a contribution to Heifer International. The format in which Heifer International was described and presented as the public good was as follows.

[7] The induced values in the two practice rounds were $6.25 and $11.60 respectively. Although each individual had the same induced value, they were not made aware of this.

> Heifer International is a nonprofit, humanitarian organization dedicated to ending world hunger and saving the earth by providing livestock, trees, training and other resources to help poor families around the globe become self-reliant. Since 1944, Heifer International has worked directly with millions of families in more than 125 countries to alleviate hunger and increase self-reliance. The livestock given to poor families (such as cows, llamas, goats, chickens, etc.), provide milk, eggs, plowing power and other benefits that can mean improved nutrition, higher income, education for children, health care, improved housing and a new way of life. Heifer requires all livestock recipients to "Pass on the Gift" by sharing offspring of their livestock with others in need. This common sense approach leads to sustainable development of one family at a time.
>
> You will vote to decide whether everyone in the group will contribute part of their earnings to provide needy families with a small flock of chickens. Each dollar per person will provide an additional needy family with a flock of chickens. For example, if the randomly selected Project Cost is $6 per person, and if a majority of people voted yes to $6, then the group will provide each of 6 families with a flock of chickens. A $3 per person Project Cost will provide each of 3 families with a flock of chickens.

To determine whether a project is implemented or not, a modified BDM with a referendum was used. For each round, only one random draw was made to determine the 'project cost' and it was the same for all participants. If more than half of the subjects in the room had a bid that was greater than or equal to the 'project cost' then the project would be implemented and so each participant paid the 'project cost' and received the project benefit. If more than half the participants had a bid that was lower than the 'project cost' then the project would not be implemented and so no one paid or received anything. The 'project costs' for stage one (Imaginary group project) and stage two (Contribution to Heifer International) were determined after respondents completed debriefing questionnaires. However, before moving on to subsequent rounds, bids of the participants were tallied and the range at which the referendum could pass was announced. For example, after tallying the level of bids, if $4 was the highest bid at which 50% or more of the participants bid yes then it was announced that for possible project costs between $1 and $4 the referendum would pass. If the project cost was between $5 and $15 then it would not pass. This additional information was provided so that before placing bids in the subsequent rounds, the individuals would have an idea regarding their potential payoffs. Furthermore, with this information, the provision rule is reduced to a simple BDM procedure where the respondents know their payoff for every possible project cost before it is drawn.

After stage two was completed, the participants were asked to fill out a short survey with some debriefing questions regarding their thoughts during the experiment, which was followed by a demographical survey. After collecting the surveys, random drawing of the project costs for both stages were conducted. The participants were well informed about which period during the experiment the random drawing would take place. At the end each subject was paid his/her earnings in cash. If the project (Contribution to Heifer International) was implemented then the project cost was subtracted from each participant's earnings and was placed separately to be mailed to Heifer International. The only difference between the real and hypothetical treatments was the subjunctive language that was used in the latter. In the hypothetical sessions each subject was paid a fixed amount of $13.50 (apart from the participation fee) regardless of the outcome of the experiments while the earnings in the real experiments varied.[8]

The project benefit from the imaginary group project as well as the actual group project can be considered to be pure public goods (at least for the participants involved in the experiment) because if the project is implemented then the benefit received is non-rival and non-excludable. However, there are a few points regarding the choice of public good that should be taken into account. The actual good being bought is a flock of chickens, which can be viewed as a private good as it is solely enjoyed by the household receiving them. But, even when viewed in this manner, the chickens can be regarded as being more than just a private good. This is because the Heifer International policy requires households to pass on the gift to other families once the chickens breed and increase in numbers. However, from the prospective of those paying for the chickens (the participants in the room) it is a public good. The chickens are being bought for some 'needy household' which means that the participants in the experiment do not benefit directly from them. The benefit from their contribution is in fact the 'value' they receive from knowing that they were able to help some needy family.

As two sets of the same experiment were conducted we refer to them as 1a and 1b. The only difference between experiments 1a and 1b is that they were conducted a week apart from one another and consisted of different participants. A time line of the order in which the different sessions took place is presented in Table 2a.

[8] In the real scenarios, given the probability of the possible outcomes, the average earning was calculated to be $13.50. To keep the average payments in real sessions consistent with the hypothetical ones, each of the participants in the hypothetical sessions received $13.50 in addition to the $5 participation fee.

4.2 Experiment 2 - Prediction Based Experiments

These experiments were very similar to the experiments above. The only difference between the two was a slight modification in stage two. After the participants placed their bids for the first two induced value rounds of stage one we moved on to stage two where instead of asking the participants for their level of contribution to Heifer International to provide needy families with a flock of chickens, we asked them to make predictions of the average contribution made by the individuals that participated in the first two sets of experiments.

At this stage (stage 2), in addition to the instructions for the prediction experiments, the participants were also provided with instructions that were given to the participants of the self-estimation experiments. The instructions at were as follows.

> In the past year, two experiments were conducted that were almost identical to the one today. Stage 1 of today's experiment, which is the part you have just completed, was identical to Stage 1 of those earlier experiments. In stage 2, (this stage) of the experiment, the participants were asked to indicate how much they were willing to contribute to Heifer International to fund needy families with a flock of chickens.
>
> Today, we won't be asking you for a contribution to Heifer International. Instead, we want your prediction of how much the participants in those two groups contributed.
>
> I shall now read to you the instructions that were provided to the participants in those two experiments.

In addition to these instructions, the participants were also provided with monetary incentives to provide accurate predictions. They were informed that ten of the participants with the most accurate predictions would receive extra cash.[9]

[9] Various amounts were provided to the participants with the ten most accurate predictions.

The difference between the real and hypothetical treatments was the subjunctive language that was used in the latter. In stage two, the participants in the hypothetical treatment were familiarized with instructions provided to individuals in the hypothetical treatments of experiments (1a and 1b). In the hypothetical sessions each subject was paid a fixed amount of $13.50 (apart from the participation fee) regardless of the out come of the experiments in the induced value sessions, while the earnings in the real experiments varied.

In both real and hypothetical treatments, monetary incentives were provided for accurate predictions. The prediction estimates provided by the individuals of stage two were compared to the average self-estimates from experiments 1a and 1b. Ten individuals with the most accurate predictions were then provided with some extra cash. A time line of the order in which the different sessions took place is presented in Table 2b.

5. Analysis and Results

There were three sets of experiments conducted in the spring of 2006 as can be seen in Table 1. The first two sets of identical experiments 1a and 1b were geared towards gathering respondents' self-estimates of the amounts they were willing to contribute to provide needy families with flocks of chicken. There were 29 participants each in the real and hypothetical settings of 1a and there were 28 and 29 participants in the real and hypothetical settings of 1b. In the third set of experiments (Experiment 2), individuals were asked to predict the average bid placed by the respondents in experiments 1a and 1b combined. When combined, there were a total of 57 individuals in the real and 58 individuals in the hypothetical setting of the self-estimate experiments.

There were a total of 39 predictors in the real and 38 predictors in the hypothetical scenarios.

5.1 Hypothetical Bias with Self-Estimates

The combined average of the self-estimates of the 57 individuals in the real setting was $8.46 and for the 58 individuals in the hypothetical setting was $10.40 as is shown in Table 3. The distribution of bids in the form of histograms displayed in figure 1 show that there are relatively more $10 and $15 bids in the hypothetical scenario. On average, the self-estimates in the hypothetical scenario were $1.94 higher than the real self-estimates. The Wilcoxon rank-sum test comparing the real and the hypothetical bids revealed a one sided p-value of 0.023 confirming the existence of hypothetical bias with the traditional CVM.

5.2 Hypothetical bias with Prediction Estimates

The average of prediction estimates was $5.21 in the real and $6.91 in the hypothetical treatments as is shown in Table 3. From the distribution of prediction estimates in Figure 2, it can be seen that the real predictions have a peak around the $4 mark while the hypothetical predictions are quite uniform. According to the Wilcoxon rank-sum tests, the difference of $1.70 between the real and hypothetical prediction treatments is statistically significant (two sided p-value = 0.014). Hence hypothetical bias is not only a phenomenon that occurs with individuals that are subject to the situation but is also predicted in other’s behavior.

These findings are consistent with the theory of false consensus effect. As discussed earlier, the false consensus theory argues that when asked to predict the behavior of others, individuals are likely to consider how they would behave themselves; since hypothetical bias is very common, it is not surprising that predicted real amounts were smaller than predicted hypothetical amounts.

Evidence supporting this interpretation was obtained from analysis of the survey debriefing questions. One of the questions in the prediction experiments asked the participants if they considered how they themselves would have performed when making their predictions.

"Did you think about how much you would be willing to contribute as a basis for formulating your prediction?"

66.67% of the respondents in the real prediction treatment and 60.53% in the hypothetical prediction treatment admitted to thinking about the amount they would have contributed if placed in such a situation.

Next, we analyze the prediction estimates to see if hypothetical bias is predicted accurately. In a study conducted by Ash, et al. (2004) using 15 different studies including 16 different types of goods, the mean level of bias, which is the ratio of the percent of respondents that stated a willingness to pay for the amount asked in the hypothetical scenario over the actual percent of respondents that agreed to pay that same amount, was observed to be 1.74. Using a similar approach, the level of bias with self-estimates is compared to that of the predicted estimates where the level of bias is considered to be the ratio of the mean willingness to pay in the hypothetical treatment and the mean willingness to pay in the real treatment. According to this measure, the bias is 1.23 with self-estimates and 1.33 with predicted estimates. The difference of 0.10 suggests that in fact people are not only capable of predicting hypothetical bias in other individuals; they are also able to make quite accurate predictions.

5.3 Accuracy of Prediction Estimates

The average self-estimate in the real and hypothetical treatments of $8.46 and $10.40 respectively were higher than the prediction estimates of $5.21 and $6.91 in the real and hypothetical scenarios. The predictions for the real and hypothetical treatments fell short of the self-estimates by $3.25 and $3.49 respectively. According to the two tailed Wilcoxon rank-sum tests, for real and hypothetical treatments, p-values of 0.018 and 0.000 suggest that the predictions are significantly lower than the self-estimates. These findings are consistent with the lake wobegon effect discussed earlier. As people often think of themselves as more generous than others, predicted payments could be less than CV format payments.

Evidence supporting this interpretation was obtained from analysis of the survey debriefing questions which shows that 20.78% (15.38% in the real and 26.32% in the hypothetical treatments) of individuals in the prediction experiments explicitly stated that they would have contributed more than what they expected others to contribute. Such statements include,

> “If I was asked to act as if the money would actually be given, I would give it all away, I don’t think other participants would.”

> “Yes, I asked myself what I would have done under the same circumstances, then subtracted a little from what I might have been willing to contribute.”

5.4 Analysis of Debriefing Questionnaire

The debriefing questionnaire towards the end of the experiment was used to gather information regarding the mental process people go through when making a valuation decision. In addition to assessing what people consider when formulating a value, a comparison between the prediction experiments and traditional CVM was made to see if these considerations vary between them.

The responses to the subjective questions of the questionnaire were content analyzed to extract and categorize them according to a few key details. Content Analysis is a technique in which qualitative data is analyzed and compressed into specific content categories so that it can be quantified and counted in an objective manner. Rather than eliciting every single bit of information, the goal is to capture the key ideas from the qualitative response and categorize them into the isolated concepts.

5.4.1 Content Analysis: Self-Estimates

Appendix A1 shows the several categories that were created from the qualitative data received from the experiments geared to elicit self-estimates (Experiments 1a and 1b) of contribution levels.

Table 4 provides a list of the various categories and the percent of individuals in each of the experimental treatments that expressed those views. It was observed that although there were small differences between how individuals stated their thought processes between these sets of experiments, there wasn't a notably significant difference between them in terms of any of the categories. According to Wilcoxon rank-sum tests, in none of the categories was the difference significant. This further suggests that the line of reasoning is not dependent upon whether experimental results are consequential or not.[10]

A factor considered quite often by the respondents was the welfare of families that would benefit from receiving the flock of chickens. 59.65% of respondents in the real and 50% of respondents in the hypothetical treatments expressed some form of *altruistic* motive for their contribution. Examples of such statements include,

> "The people the charity would be helping need it far more than I would need the $15"

> "The money would mean more to them than me and I would probably end up spending it on something I didn't really need anyways."

A significant number of respondents provided some form of a *calculated* response. According to the stated answers, 33.33% in the real and 48.28% in the hypothetical treatments seemed to have decided upon their bids through some form of rational or mathematical reasoning. Examples of such statements include,

> "Value to me vs. value to family. Amount of $ I wanted to give in relation to how much I thought I might be paid"

> "Reasonable amount. Retain enough money for myself. Give almost half of my earnings"

Only about a quarter of the total respondents, that is 22.81% in the real and 25.86% in the hypothetical treatments expressed profit maximization or self-interest as a significant factor when formulating their bids. Some of the statements expressing *self interest* were the following.

> "How much money I was going to have. Irresponsible to donate when you don't have enough to feed yourself."

> "How much money I have on me, how much money I will be able to make in the short term"

[10] The two tailed Wilcoxon rank-sum tests to examine the difference between Real and Hypothetical Treatments resulted in the following p-values for each of the categories: Altruistic: 0.301; Calculated: 0.105; Selfish: 0.704; Credibility: 0.178; No Trust: 0.089; Good Work: 0.149; No Force: 0.313.

When introducing Heifer International and their previous work to the respondents during the experiments, a brochure was also provided to each individual. This brochure seemed to be successful in gaining the trust of some of the individuals. Hence some of the respondents explicitly mentioned the *'credibility'* of Heifer International to be a factor they considered when placing their bids. Furthermore, the work performed by the organization in the past as presented by the brochure was another factor that was considered during the formulation of bids (*'goodwork'*). On the other hand, because some of the respondents were not familiar with the organization, they expressed concern regarding the credibility (*'no trust'*) of the organization. An example of such an expression include,

> "I was wondering if there is a real organization. Maybe it was a trick. I'm a little paranoid."

Because the provision mechanism included a referendum, there were some respondents who felt that it was wrong to force (*'no force'*) other individuals against their will to make a contribution. Such statements include,

> "I believe donations should be a matter of personal choice. We should not force the entire group to contribute to something they might not want to contribute to. I might support this charity but the person next to be might not. It isn't fair to him/her to take his/her money away from them to contribute to something I feel is a good cause."

A linear regression analysis was conducted to examine the relation between the various categorical and demographical variables with the self-estimates.[11] The results of the regression are presented in Table 5 (first column). The variable 'Hyp' which stands for hypothetical treatment has a coefficient of 2.1 and is statistically significant at the five percent level. This result confirms the existence of hypothetical bias with self-estimates.

[11] Normality tests of the linear regression analysis are presented in Appendix B

The rest of the independent variables in the linear regression are either demographic variables or categorical variables obtained from the survey analyses. In the debriefing questionnaire, a question was asked regarding how well the respondents understood the experimental procedures. The variable '*Understood'* represents the following question,

"On a scale of 1 to 10 where 1 means "I was lost" and 10 means "I understood everything," how well did you understand the Experimental Procedures Today?"

On average the respondents specified 8.53 in the real and 8.52 in the hypothetical treatments as to how well they understood the experimental procedures. According to the regression results, the level at which individuals understood the experiment was not a significant factor in explaining the magnitude of self-estimates. Another question was designed to measure how certain the respondents were regarding their self-estimates.

On a scale of 1 to 10, where 1 means "very uncertain" and 10 means "very certain," how certain are you of your level of contribution to Heifer International?

On average the respondents specified 7.96 in the real and 8.10 in the hypothetical treatments as their certainty level in terms of their stated self-estimates. This variable '*certain'* was also not significant, suggesting that the certainty level of respondents did not have an effect on the magnitude of contribution levels.

Among the demographical variables, women (represented by the variable '*female'*) were more likely to make higher contributions. The coefficient of 2.29 for the variable female was significant at the five percent level. Also, people who had participated in more experiments (represented by the variable '*numexp*') made lower contributions as can be seen from the negative coefficient which is marginally significant at the five percent level. It is possible that students who participate in more experiments simply view experiments as a way of earning some extra cash. This could have led to lower levels of contributions for these students while individuals coming for the first time may not have been as keen in terms of making a profit.[12]

A few of the categorical variables created from the content analysis were also able to explain the bidding behavior of the respondents. The individuals who considered the welfare of needy families (*altruism*), regarded Heifer International to be a credible organization (*credible*), or were impressed by the quality of their work (*goodwork*) made significantly higher contributions. On the other hand those who were more concerned about profit maximization or felt uncomfortable forcing others to make charitable donations (*noforce*), made significantly lower contributions.

5.4.2 Content Analysis: Prediction Estimates

In the debriefing questionnaire of the prediction based experiments, some of the categories created from the qualitative data using content analysis (see Table 6) were similar to the ones created for the self-estimation experiments as we wanted to test if people had similar thought processes between formulating self-estimates versus making predictions of other's. A significant difference was observed between the two in terms of the number of statements implying altruistic motives. In the prediction based experiments there were only 20.51% in the real and 26.32% in the hypothetical treatment that explicitly stated something regarding the welfare of needy families. In the self-estimation experiments these percentages were much higher, with 65% in the real and 50% in the hypothetical treatment stating welfare oriented concerns. A two tailed Wilcoxon rank-

[12] Because women and individuals who had participated in more experiments had a significantly different level of contribution, a test was conducted to see if the sample of individuals between the self estimation experiments and prediction experiments were the same. According to the two tailed Wilcoxon rank-sum tests a p-value of 0.8843 for '*female*' and 0.2993 for '*numexp*' suggests that the sample of participants were homogeneous between the two types of experiments.

sum test p-value of 0.000 shows that a significantly smaller number of individuals consider altruistic motives to be as significant when making predictions. This further supports our earlier observation of the lake wobegon effect where people regard other individuals to be less generous compared to themselves. There were however 20.51% in the real and 13.16% in the hypothetical treatment that did explicitly express their belief that other individuals were of good nature.

Although there were some differences between how people were thinking between the real and hypothetical treatments in the prediction based experiments, in none of the categories except one was this difference statistically significant.[13] For the category *'Self Interested'* there were 46.15% in the real and 15.79% in the hypothetical treatment that explicitly stated profit maximization (self-interested behavior) to be a significant motive. According to the two tailed Wilcoxon rank-sum test, a p-value of 0.004 suggests that this difference is highly significant. In addition, 21.05% of individuals in the hypothetical treatment explicitly acknowledged the hypothetical nature of the surveys. Some of the statements included,

"I figure it is not coming out of the earnings you are making, and the money goes to a good cause so $10 could provide a lot to the families who need it."

"Random, just a little high because no one was going to lose money"

There were two factors that may have significantly contributed to the prediction of hypothetical bias. First of all there were more participants that claimed to be aware of the hypothetical nature in the prediction based experiments in comparison to the self-estimation experiments; and second, predictors in the real treatment seemed to consider profit maximization to be less important.

[13] The two tailed Wilcoxon rank-sum tests to examine the difference between Real and Hypothetical Treatments resulted in the following p-values for each of the categories: Altruistic: 0.55; Calculated: 0.916; Selfish: 0.004; Goodnature: 0.392; Selfbetter: 0.240; Thoughtofself: 0.578; Guess: 0.712.

A linear regression of prediction estimates with predicted values as the dependent variable is presented in Table 5 (second column). It can be seen that prediction estimates in the hypothetical scenario (represented by the variable '*hyp*') are significantly higher in comparison to the real. A positive coefficient of 1.64 which is statistically significant at the five percent level reinforces our previous result that individuals are capable of predicting hypothetical bias in others. Similar to the regression results of the self-estimates, the variables *'Understood'* and *'Certain'* were not statistically significant. In terms of the certainty question, the respondents in the prediction based experiments were asked how confident they were in terms of their predictions.

"For Stage 2 of today's experiment, on a scale of 1 to 10, where 1 means "not very confident" and 10 means "very confident," how confident are you of the accuracy of your Prediction today?"

The average level of confidence as expressed by the respondents was 6.23 in the real and 6.13 in the hypothetical setting. Although the confidence levels were not statistically different between the real and hypothetical scenarios, they were much lower compared to the stated certainty levels of individuals in the self-estimate experiments. A two tailed p-value of 0.000 using the Wilcoxon rank-sum test suggests that the stated levels of confidence for predictions are significantly lower than the stated levels of certainty for self-estimates. This shows that people believe they are capable of predicting their own behavior far better than predicting the behavior of a group of other individuals.

People who considered altruistic motives ('*altruism'*) while making their predictions made significantly higher predictions. Predictions were also higher for those who believed that human beings were good by nature ('*goodnature'*). Individuals who considered themselves to be more generous in comparison to others ('*selfbetter'*) also made significantly higher predictions. This result incorporates both the lake wobegon effect as well as the false consensus effect. Participants who considered themselves to be more generous, by making higher predictions, subconsciously believed others to be generous too.

In both the real and hypothetical prediction based experiments the individuals were offered monetary incentives for providing accurate predictions. Unlike the study conducted by Offerman et al. (1996), where over 50% of the individuals stated that they would have changed their answers if no monetary incentive was provided; according to our results, only 10% said that they would change their predictions. However, apart from what the participants stated, we do not have any other evidence regarding the effects of monetary incentives on accurate predictions.

5.4.3 Content Analysis: Self-Estimates and Prediction Estimates

The self-estimates and prediction estimates were pooled together to run a joint linear regression (see Table 5, Column 3). A positive coefficient of 1.35 for the variable ‘*hyp*’ suggests that hypothetical bias is a common phenomenon in both self-estimates and prediction estimates. A negative coefficient for the variable ‘*numexp*’ of -0.19 and a positive coefficient for the variable *predction*numexp* (an interaction term between variables *numexp* and *prediction*) of 0.30 suggests that for the self-estimation experiments, the participants who had participated in more previous experiments stated lower levels of contributions. As discussed earlier, this result could have been due to a profit-maximizing behavior which is likely to be more prevalent with individuals who participate in experiments more often. Moreover, because making significantly lower or higher predictions did not lead to profit maximization in the prediction based experiments, *numexp* was not significant in these experiments.

Women are more likely to provide higher levels of contribution as can be seen from the positive coefficient of 2.75 for ‘female’. However, a negative coefficient for *prediction*female* (an interaction term between *prediction* and *female*) of -2.62 suggests that their predictions aren’t significantly different from that of males.[14]

6. Summary and Conclusions

The success of prediction estimates in prediction markets has been demonstrated by its ability to outperform results of polls. This leads us to believe that prediction estimates may also outperform the traditional CVM as a tool for non-market valuation. Primarily because of the shift in locus of control when predicting, it is plausible that incentives for strategic bias are reduced. However, it is not apparent that a prediction format will actually do this. For example, just as an individual may state higher WTP estimates to increase the likelihood of the provision of a commodity, the individual may also find it strategically advantageous to predict that others will pay more than what he/she thinks they actually would.

In many situations, the prediction format may induce the types of behavior anticipated by the theory of false consensus and the lake wobegon effect. Since people expect others to be like themselves, their prediction estimates of others’ behavior may be heavily influenced by their own preferences, and as a result generate very accurate predictions of hypothetical bias. Coincidentally, the prediction estimates may be

[14] Although significantly higher levels of contribution were placed by women, the prediction estimates between men and women weren’t necessarily different. Prediction estimates between men and women only varied by a factor of \$2.75 – \$2.62 = \$0.13

underestimated as people generally think of themselves as being more generous. In accordance to both the false consensus and the lake wobegon effects, our results show that while accurately predicting hypothetical bias, the prediction estimates are also significantly lower than the estimates from the traditional CVM.

There are however certain limitations in our experiments that may be worth noting. We have compared the prediction estimates to the self-estimates with the assumption that the self-estimates from the real treatment are in fact the true values. It is quite possible that the predictions represent the true values and it is the self-estimates that are flawed. In addition, between the different experimental groups, a variation in group dynamics could have influenced the overall outcome. A solution could be to have only one big group of respondents where half the group are asked to place their personal self-estimates and the other half asked to predict the average self-estimate placed by the first half.

Strong conclusions cannot be made from the results of this study as the observations from our small sample size could be due to sampling errors. There is however tremendous potential for prediction mechanisms as has been experienced and employed in prediction markets. Hence more research is required to improve this technique and explore its benefits in order to use it as an alternative to the CVM for non-market valuation.

Bibliography:

Aadland, D., and A. J. Caplan. 2003. "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias." *American Journal of Agricultural Economics*. 85(2): 492-502.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. Journal of Personality and Social Psychology, 49, 1621-1630.

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. Social Cognition, 7, 275-295.

Arrow, K., R. Solow, P. Portnoy, E. Leamer, R. Radner and H. Schuman. (1993) Report of the NOAA Panel on Contingent Valuation. *Federal Register*. 4601-4614.

Ash, M., Murphy, J., Stevens, T., (2004) “Hypothetical Bias in Dichotomous Choice Contingent Valuation." (Under Review)

Ausubel, L.M. (1991). “The failure of competition in the credit card market.” *American Economic Review*, vol. 81, 50-81

Becker, G., DeGroot, M., and Marschak, J. (1964). “Measuring utility by a single-response sequential method,” *Behavioral Science*, 9, 226-232

Berg, J., R. Forsythe, F. Nelson, and T. Rietz. “Results from a Dozen Years of Election Futures Markets Research.” In Handbook of Experimental Economics Results. C. Plott and V. Smith, eds., Amsterdam: Elsevier, (forthcoming).

Bishop, R. C. and T. A. Heberlein (1979), “Measuring Values of Extra market Goods: Are Indirect Measures Biased?” *American Journal of Agricultural Economics*, 61, 926-930.

Bishop, R. C. and T. A. Heberlein (1986), “Does Contingent Valuation Work?,” in Cummings, R.,D. Brookshire and W. Schulze, eds., *Valuing Environmental Goods: A State of the Art Assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman and Allenheld.

Blamey, R.K,. (1998) “Decisiveness, attitude expression and symbolic responses in contingent valuation surveys.” *Journal of Economic Behavior and Organization* 34, 577-601.

Bohm, P., Linden, J., and Sonnegard, J. (1997). “Eliciting reservation prices: Becker-Degroot-Marshak mechanisms vs. markets.” *The Economic Journal*, 107, 1079-1089

Bohm, P. (1972), “Estimating the Demand for Public Goods: An Experiment,” *European Economic Review*, 3, 111-130.

Bredhal, M., Lee, J. G., Paarlberg, P. L, (2003) “Implementing Multifunctionality” Paper presented at the International Conference, Agricultural policy reform and the WTO: where are we heading? Capri (Italy), June 23-26.

Brown, C.E. (1982) “A False Consensus Bias in 1980 Presidential Preferences,” The *Journal of Social Psychology*, 118, 137-138.

Brown, T. C., I. Ajzen, and D. Hrubes. (2003) "Further Tests of Entreaties to Avoid Hypothetical Bias in Referendum Contingent Valuation." *Journal of Environmental Economics and Management*. 46(2): 353-361

Burton, A. C., K. S. Carson, S. M. Chilton, and W. G. Hutchinson. (2003) "An Experimental Investigation of Explanations for Inconsistencies in Responses to Second Offers in Double Referenda." *Journal of Environmental Economics and Management*. 46(3): 472-489.

Carson, R.T., T. Groves and M. Machina. (1999) "Incentive and Informational Properties of Preferences Questions," Plenary Address, European Association of Environmental and Resource Economists, Compatibility Issues in Stated Preference Surveys," Oslo Norway, June.

Carson, R., T. Groves, J. List, and M. Machina. (2004). “Probabilistic Influence and Supplemental Benefits: A Field Test of the Two Key Assumptions Underlying Stated Preferences,” paper presented at NBER Public Economics Workshop, Palo Alto.

Champ, P. A., and R. C. Bishop. (2001). "Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias." *Environmental and Resource Economics*. 19(4): 383-402.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum. (1997) "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management*. 33(2): 151-162.

Chen, Y., Chu, C., Mullen, T., Pennock, D. M. (2005) “Information markets vs. opinion pools: an empirical comparison.” ACM Conference of Electronic Commerce : 58-67
Cummings, R. G., S. Elliott, G. W. Harrison, and J. Murphy. (1997) "Are Hypothetical Referenda incentive Compatible?" *Journal of Political Economy*. 105(3): 609-21.

Cummings, R. G., and L. O. Taylor. (1999). "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *The American Economic Review*. 89(3): 649 - 665.

Dawes, R.M., (1989) “Statistical Criteria for Establishing a Truly False Consensus Effect:; *Journal of Experimental Social Psychology* 25, 1-17.

Epley, N., & Dunning, D. (2000). Feeling “holier than thou”: Are self-serving assessments produced by errors in self or social prediction? Journal of Personality and Social Psychology, 79, 861-875.

Ethier, R. G., G. L. Poe, W. D. Schulze, and J. Clark. (2000). "A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Electricity Programs." *Land Economics*. 76(1): 54-67.

Forsythe, R., F. Nelson, G.R. Neumann, and J. Wright (1992). “Anatomy of an Experimental Political Stock Market.” *American Economic Review* 82:1142-1161.

Fox, J. A., J. F. Shogren, D. J. Hayes, and J. B. Kliebenstein. 1998. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics*. 80(3): 455-465.

Goethals, G. R. (1986). Fabricating and ignoring social reality: Selfserving estimates of consensus. In J. M. Olson, C. P. Herman, & M. P. Zanna (Eds.), Social comparison and relative deprivation: The Ontario symposium (Vol. 4, pp. 135-157). Hillsdale, NJ: Erlbaum.

Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls & T. A. Will (Eds.), Social comparison: Contemporary theory and research (pp. 149-176). Hillsdale, NJ: Erlbaum.

Harrison, G. W. and E. E. Rutström, “Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods,” *Handbook of Results in Experimental Economics*. New York: Elsevier Science. 2002

Harrison, G. (1989) “Theory and Misbehavior of First-Price Auctions,” *American Economic Review*, 79, 749-762

Heijden, E.v.d., Nelissen, J., and Potters, J., (2003). “Opinions on Tax Deductions and the Consensus Effect in a Survey-Experiment”; Mimeo, Tilburg University.

Hoevenagel, R.(1990) “The validity of the contingent valuation method: Some aspects on the basis of three Dutch studies,” paper to the European Association of Environmental and Resource Economists conference, Venice.

Hofler, R. A., and J. List. 2004. "Valuation on the frontier: Calibrating Actual and Hypothetical Statements of Value." *American Journal of Agricultural Economics*. 86(1): 213-221.

Irwin, J.R., McClelland, G.H., McKee, M., Schulze, W., and Norden, N. (1998). "Payoff dominance vs. cognitive transparency in decision making." *Economic Inquiry*, 36, 272-285.

Kahneman, D., J.L. Knetsch and R.H. Thaler, (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." *The Journal of Political Economy*. 98: 1325-1348.

Kahneman, D., J. Knetsch and R.Thaler (1991) "The endowment effect, loss aversion and status quo bias", *Journal of Economic Perspectives*, 5, 193-206.

Katz, D., and Allport, F. (1931) "Students' attitudes." Syracuse: Craftsman Press, *Journal of Higher Education*, Vol. 2, No. 9 p. 526

List, J. A. (2001). "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review*. 91(5): 1498-1507.

List, J. A., and J. F. Shogren. (2002). "Calibration of Willingness to Accept." *Journal of Environmental Economics and Management*. 43(2): 219-233.

List, J. A. and C. Gallet (2001), "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics*, 20, 241-254.

List, J. A., R. Barrens, A. Bohara and J. Kerkuliet. (2004). "Examining the Role of Social Isolation on Stated Preferences." *American Economic Review*. 94(3), 741-752.

Loewenstein, G. and D. Adler. (1995) "A Bias in the Prediction of Tastes." *Economic Journal*. 105: 929-37.

Lusk, J. L. 2003. "Willingness-to-pay for Golden Rice." *American Journal of Agricultural Economics*. 85(4): 840-856.

Lusk J. L., and Norwood F.B. (2005) "Do Your Neighbors Know You better than Themselves? A Prediction-Based Non-Market Valuation Method. (Forthcoming).

Marshall, J.D., Knetsch, J.L. and Sinden, J.A. (1986). "Agents' evaluations of the disparity in measures of economic loss." *Journal of Economic Behavior and Organization*, vol. 7, pp.115-27.

Mullen, B., Atkins, J. L., Champion, D.S., Edwards, C., Hardy, D. Story, J. E., and Venderklok, M., (1985). "The False Consensus Effect: A Meta – Analysis of 115 Hypothesis Tests: *Journal of Experimental Social Psychology*, 21, 263-83.

Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D. (2005). “A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation.” *Environmental and Resource Economics*, 30(3):313–325.

Offerman, T., Sonnemans, J., and Schram, A. (1996). “Value Orientations, Expectations and Voluntary Contributions in Public Goods.” *The Economic Journal*. 106, 817-845.

Plott, C. and Zeiler, K., 2005, “The Willingness to Pay-Willingness to Accept Gap, the “Endowment Effect”, Subject Misconceptions, and Experimental Procedures for Eliciting Valuations”, *American Economic Review*, Vol, 95(3), 530- 545.

Poe, G. L., J. E. Clark, D. Rondeau, and W. D. Schulze. (2002). "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics*. 23: 105-131.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science, 306*, 462-466.

Ross, L., Greene, D., and House, P., (1977). “The ‘False Consensus Effect’; An Egocentric Bias in Social Perception and Attribution Processes”; *Journal of Experimental Social Psychology* 13, 279-301.

Rotter. J. B, (1954) “Social Learning and Clinical Psychology.” New York: Prentice-Hall.

Rotter, J. (1966). “Generalized expectations for internal versus external control of reinforcement,” *Psychological Monographs*, 80, 1-28.

Seip, K. and Strand J. (1992), “Willingness to pay for environmental goods in Norway: a contingent valuation study with real payment,” *Environmental and Resource Economics.* 2, 91-106.

Servan-Schreiber, E., J. Wolfers, D.M. Pennock, and B. Galebach. “Prediction Markets: Does Money Matter.” Electronic Markets 14(2004):2443-251.

Sherman, S.J. (1980). “On the self-erasing nature of errors of prediction.” *Journal of Personality and Social Psychology*, 39, 211-221.

Smith V.L. (1982), “Microeconomic systems as an experimental science”, *American Economic Review* 72, 923-955.

Surowiecki, J. (2004) “The Wisdom of Crowds.” New York: Doubleday.

Tat, P., Cunningham, W.A. and Babakus, E. (1988). “Consumer perceptions of rebates.” *Journal of Advertising Research*, vol. 28, 45-50.

Thaler, R. (1980). “Toward a positive theory of consumer choice.” *Journal of Economic Behavior and Organization*, vol. 39 45-50.

Vallone, R.P., Griffin, D.W., Lin, S., and Ross, L. (1990). “Overconfident predictions of future actions and outcomes by self and others.” *Journal of Personality and Social Psychology*, 58, 582-592.

Vatn, A., (2004) “Environmental Valuation and Rationality” *Land Economics*, 80(1):1-18.

Wang, H. (1997). "Treatment of 'Don't-Know' Responses in Contingent Valuation Surveys: A Random Valuation Model." *Journal of Environmental Economics and Management*. 32(2): 219-232.

Table 1: Experimental Design

Experiment	Treatment	Number of Participants	Stage 1 Induced Value		Stage 2 Homegrown
			Round 1	Round 2	
1a Self-estimate	Real	29	$4.50	$11.50	Stated own level of Contribution to Heifer International
	Hypothetical	29	$4.50	$11.50	
1b Self-estimate	Real	28	$4.50	$11.50	
	Hypothetical	29	$4.50	$11.50	
2 Prediction Estimate	Real	39	$4.50	$11.50	Predicted the Average Contribution to Heifer International in Experiments 1a and 1b
	Hypothetical	38	$4.50	$11.50	

Table 2a: Timeline of the Experimental Procedures for Experiments 1a and 1b

Instructions	Practice Rounds (Induced Value)		Induced Value Rounds		Self Estimate of contribution level	Survey
	Practice Round 1 ($6.25)	Practice Round 2 ($11.60)	Induced Value Round 1 ($4.50)	Induced Value Round 2 ($11.50)		

TIME

Table 2b: Timeline of the Experimental Procedures for Experiment 2

Instructions	Practice Rounds (Induced Value)		Induced Value Rounds		Prediction Estimate of contribution levels	Survey
	Practice Round 1 ($6.25)	Practice Round 2 ($11.60)	Induced Value Round 1 ($4.50)	Induced Value Round 2 ($11.50)		

TIME

Table 3: Mean and Median bids and Prediction of Contribution to Heifer International

Experiment	Mean		Median	
	Real	Hypothetical	Real	Hypothetical
1a. Self Estimate	9.448	10.931	10	10
1b. Self Estimate	7.429	9.862	5	10
1a. and 1b. Combined	8.46	10.4	8	10
2. Prediction Estimate	5.208	6.908	4.25	7.2

Table 4: Percent of Individuals that explicitly specified the following: Self-Estimation Experiments

Experiments	1a		1b	
Category	Real	Hypothetical	Real	Hypothetical
Altruistic	58.62	55.17	60.71	44.83
Calculated	41.38	48.28	25.00	48.28
Self Interested	17.24	24.14	28.57	27.59
Credibility	10.34	0	14.29	10.34
No Trust	3.45	6.90	3.57	17.24
Good Work	24.14	31.03	14.29	31.03
No Force	0	0	3.57	0
Hyp Nature	x	13.79	X	3.45
Number of Observations	29	29	28	29

Table 5: Regression Analysis of Self Estimates and Prediction Estimates

Variable	Estimated Coefficient (std error)		
	1. Self-Estimates	2. Prediction Estimates	3. Self Estimates and Prediction Estimates Combined
Intercept	61.229 (175.52)	-100.81 (111.55)	19.31 (110.70)
Hypothetical	2.10** (0.92)	1.64*** (0.64)	1.36** (0.60)
Understood	-0.03 (0.27)	-0.02 (0.18)	0.01 (0.18)
Certain[†]	0.09 (0.22)	-0.09 (0.19)	0.12 (0.15)
NumExp	-0.20* (0.10)	0.10 (0.10)	-0.19** (0.09)
Female	2.30** (0.91)	0.08 (0.78)	2.75*** (0.76)
Year Born	-0.03 (0.09)	0.05 (0.06)	-0.01 (0.06)
Altruistic	3.49*** (0.98)	2.01*** (0.78)	2.52*** (0.64)
Calculated	-0.78** (1.03)	-0.26 (0.68)	-1.09* (0.61)
Self Interested	-2.38** (1.11)		-2.81*** (0.68)
Credibility	3.50** (1.62)		
No Trust	-1.90 (1.69)		
Good Work	1.87* (1.10)		
No Force	-8.45 (4.68)*		
Good Nature		2.44*** (0.85)	
Self Better		1.58** (0.76)	
Thought of Self		0.42 (0.71)	
Guess		-0.94 (0.98)	
Prediction			-1.50* (0.87)
Prediction*Female			-2.63**

			(1.20)
Prediction*Numexp			0.30* (0.17)
Number of Observations	113	77	189
F	5.34 (p<0.0001)	3.13 (p<0.0015)	9.99 (p<0.0001)

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

† For the Prediction Experiments, instead of asking them their certainty level, they were asked their level of confidence.

The Standard Errors are provided in brackets

Table 6: Percent of Individuals that explicitly specified the following: Prediction Based Experiments

Experiment	**Prediction Based Experiment**	
Category	Real	Hypothetical
Altruistic	20.51	26.32
Good Nature	20.51	13.16
Calculated	46.15	47.37
Self Interested	46.15	15.79
Self Better	15.38	26.32
Guessed	12.82	15.79
Hyp Nature	x	21.05
Number of Observations	39	38

Figure 1: Distribution of Self-Estimates

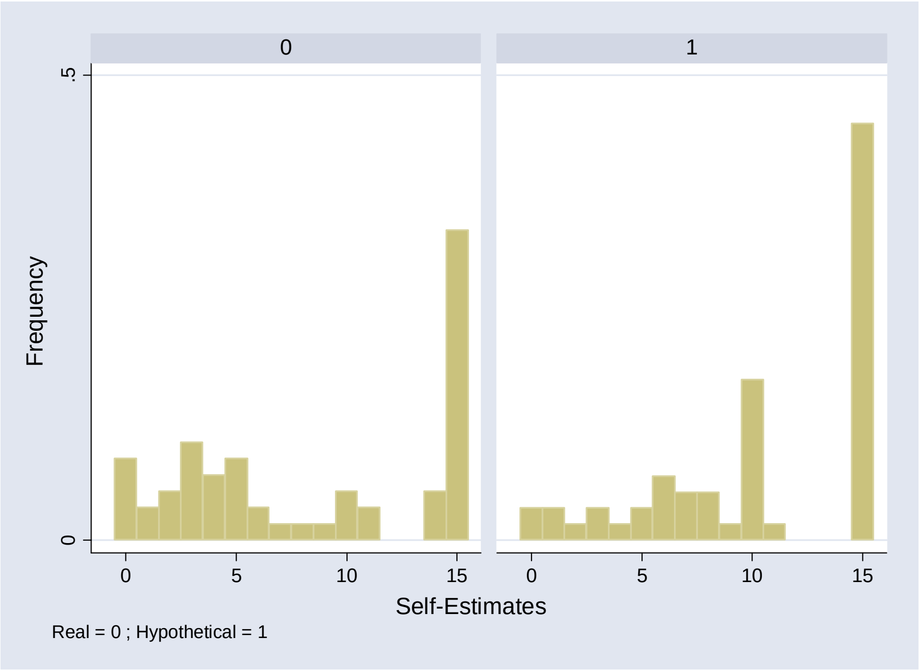
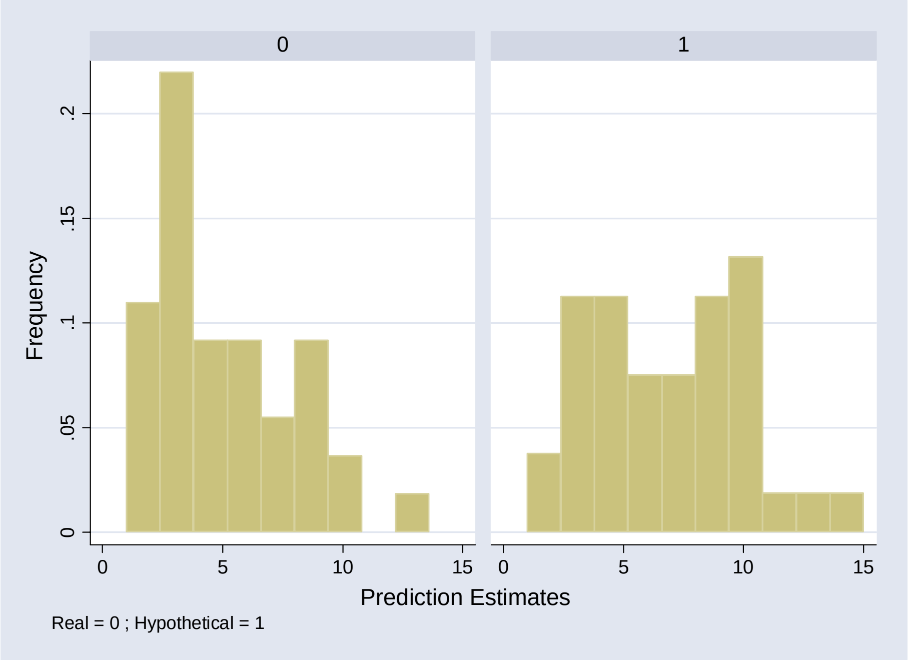


Figure 2: Distribution of Predictions



Appendix A1

Content Analysis Coding Frame: For Self-Estimates (1a and 1b)

CODE	DESCRIPTION
Altruistic	Expressed concern about the welfare of others, how his/her aid could make a difference, and their desire to contribute
Calculated	Provided a reasonable and calculated response (even mathematical), or considered how much others would contribute
Self Interested	Expressed the desire to make money, needed the money, or expressed self interest and concern for profit maximizing. Was concerned more with cost to self, whether the contribution was affordable and the possibility of incurring a loss
Credibility	Mentioned the credibility of the public good provision and the organization
No Trust	Expressed distrust or was skeptical about the organization, provision of the public good, or their methods
Good Work	Expressed their satisfaction or even how they were fascinated by the work/philosophy of the organization (Especially because of the information provided by the Brochure)
No Force	Felt Uncomfortable, or wrong forcing others to contribute
*Hyp Nature	Acknowledged the hypothetical nature of the about the Hypothetical nature
* Only for Hypothetical Treatments	

Appendix A2

Content Analysis Coding Frame: For Prediction Based Experiments

CODE	DESCRIPTION
Altruistic	Expressed concern about the welfare of others, how the aid could make a difference, and their desire to contribute
Calculated	Provided a reasonable and calculated response (even mathematical)
Self Interested	Expressed the desire of students to make money, how they needed the money. Mentioned that college students are poor in general and cannot afford to make such contributions.
Self Better	Expected others to be more self interested and contribute less while he/she was more generous and would contribute more.
Good Nature	Expressed faith in human beings (including students). Believed people to be good by nature.
Guess	Had no clue. Just placed a random guess

Appendix B

Tests of Normality for Regressions:

Regression 1

Test		Statistic	P-Value	
Shapiro-Wilk	W	0.98322	Pr<W	0.1693
Kolmogorov-Smirnov	D	0.068904	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.071017	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.432198	Pr > A-Sq	>0.2500

Regression 2

Test		Statistic	P-Value	
Shapiro-Wilk	W	0.969708	Pr < W	0.0658
Kolmogorov-Smirnov	D	0.082593	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.11472	Pr > W-Sq	0.0739
Anderson-Darling	A-Sq	0.711452	Pr > A-Sq	0.0637

Regression 3

Test		Statistic	P-Value	
Shapiro-Wilk	W	0.98857	Pr < W	0.1335
Kolmogorov-Smirnov	D	0.063684	Pr > D	0.0607
Cramer-von Mises	W-Sq	0.092541	Pr > W-Sq	0.1431
Anderson-Darling	A-Sq	0.577951	Pr > A-Sq	0.1362

At the 5% level of significance, all of the tests above fail to reject the null hypothesis that residuals are normal.