BDM Variance Pre-Analysis Plan

Horlick, Palma, Drichoutis

November 2025

1 General Information

1. Trial Title: Payoff variance in preference elicitation with the BDM mechanism
2. Location: United States of America
3. Primary Investigator: Dr. Marco Palma
4. Other Primary Investigators: Dr. Andreas Drichoutis & Benjamin Horlick
5. Keyword(s): Behavior
6. Additional Keyword(s): Underbidding, Overbidding, Game form recognition, Misbidding, BDM mechanism, induced value
7. JEL Code(s): D81, C90
8. Abstract: Under expected utility theory, the range of the bid distribution in the Becker-DeGroot-Marschak (BDM) mechanism should not influence bidding behavior. However, there is overwhelming empirical evidence showing it does. We examine the relationship between the upper bound of the range for the randomly drawn price in the BDM mechanism and the variance of the payoff. Treating payoff variance as a measure of risk, we develop a framework that causally explains empirical under/overbidding behavior shifts due to changes in the support set distribution. We then present results from an experiment eliciting willingness-to-accept after endowing subjects with an induced value card worth a fixed amount and varying the random offer support set's upper bound. We analyze within-subject effects among three BDM rounds and explore heterogeneous responses due to varying risk preferences that we hypothesize will result in aggregate under/overbidding or more symmetrical bids around the induced value.

2 Dates

1. Trial Start Date: December 1, 2025
2. Intervention Start Date: December 1, 2025
3. Intervention End Date: December 5, 2025
4. Trial End Date: December 5, 2025

3 Sponsors & Partners

1. Sponsor(s): N/A
2. Partner(s): N/A

4 Experimental Details

1. Intervention (Public): We implement the BDM mechanism in three repeated rounds to elicit willingness-to-accept for an induced value item. The value of the item is held constant at $3. The distribution of possible offers ranges from $0 to an upper bound that is exogenously changed each round to $4, $6, and $12 in random order. Shifting the upper bound changes the variance of the payoff, and we hypothesize it impacts subject behavior to explain overbidding (in $12), underbidding (in $4) and more symmetrical bidding (in $6) around the induced value.

2. Intervention (Hidden): N/A

3. Primary Outcome (End Point): Individual willingness-to-accept values in each treatment

4. Primary Outcomes (Explanation): We use the elicited values to test whether moving the upper bound of the support set results in changes to the distribution of WTA values. Please see the Analysis Plan in Section 6 for more details.

5. Secondary Outcomes (End Points):

   (a) Individual risk preference measures elicited via the Bomb Risk Elicitation Task (BRET) from Crosetto and Filippin (2013).

   (b) Individual risk preference measures elicited via self-evaluation.

6. Secondary Outcomes (Explanation): We utilize the secondary outcomes to conduct robustness checks of the primary hypothesis and explore heterogeneous effects by risk preferences. Please see the Analysis Plan in Section 6.

7. Experimental Design (Public): Our experiment is divided in two parts. In the first part, subjects are asked to submit offers to sell a card worth $3 to the experimenter. We vary on a within-subjects basis the upper bound of the support set at $4, $6, and $12 in three repeated rounds in random order with the lower bound remaining at $0. At the end of the experiment, one of the three rounds is selected for realization. Following the BDM mechanism task, we ask respondents about their strategy selection process and give them the opportunity to explain a rationale in an open response format.

   The second part of the experiment elicits subjects' risk preferences. We implement the static version of the BRET (Crosetto and Filippin, 2013). Subjects complete the task once with no practice rounds. The task presents participants with 100 boxes arranged in a 10x10 grid. One randomly selected box holds a bomb while the other 99 contain a reward of $0.10. Each subject then chooses how many boxes to collect. If the bomb is among the selected quantity of boxes, the participant receives no additional earnings from the task; otherwise, we add the amount of money inside the collected boxes to the subject's total payoff.

   As a secondary measure of risk preferences, we also employ the self-reported risk question from the German Socio-Economic Panel survey (Dohmen et al., 2011).

8. Experimental Design (Hidden): N/A

9. Randomization Method: Computer; all randomizations are performed within Qualtrics.

10. Randomization Unit: Treatment order is randomized at the individual level.

11. Was the treatment clustered? No.

12. Sample Size

    (a) Planned Number of Clusters: 300 individuals

    (b) Planned Number of Observations: 900 observations (300 individuals and 3 rounds)

    (c) Sample size (or number of clusters) by treatment arms: N/A

(d) Power calculation: We utilize the data from similarly parameterized treatments in Drichoutis et al. (2025) to estimate the inputs to our power calculations. With an induced value of \$3.00, the average offer shifted from \$2.701 with a support set upper bound of \$4.00 to \$3.158 with a support set upper bound of \$6.00. Each subject participated in both treatments. The pooled standard deviation of the two offer distributions was 0.895. In our power analysis presented below, we conservatively assume a between-treatment correlation of 0.3, compared to the actual observed correlation of 0.501. The effect size, d_z, found by Drichoutis et al. (2025) is then 0.429, given by the formula:

$$d_z = \frac{\mu_1 - \mu_2}{\sigma_{diff}} \tag{1}$$

where μ_1 and μ_2 represent the mean offer value by treatment, and σ_{diff} is the correlation-adjusted pooled standard deviation:

$$\sigma_{diff} = \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2} \tag{2}$$

where σ_1 and σ_2 are the standard deviations by treatment and r is the coefficient of correlation (Cohen, 2013; Lakens, 2013).

We employ the asymptotic relative efficiency (ARE) method in our power calculation which estimates the sample size required under a parametric t-test at a given power level and converts the result to the sample size required by the nonparametric Wilcoxon signed-rank test that we use to test our primary hypothesis (Faul et al., 2009).

From Rosner et al. (2006), Cohen (2013), and Lakens (2013), the estimated sample size n to achieve the maximum significance level (probability of Type I error) under the Holm-Bonferroni step-down procedure with three joint hypotheses of $\alpha = \frac{0.05}{3} \approx 0.0167$ and a power level (probability of Type II error) of $\beta = 0.80$ under a one-sided t-test is given by:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{d_z^2} \tag{3}$$

where $z_{1-\alpha}$ and $z_{1-\beta}$ take the values of 2.13 and 0.84, respectively (Holm, 1979). The sample size required under a t-test is then 51. The ARE of the Wilcoxon test is $\frac{3}{\pi} \approx 0.955$, implying a sample size of 53 (Faul et al., 2009). Figure 1 presents the sensitivity of the sample size calculation to the assumed between-treatment correlation. Utilizing the actual observed correlation value of 0.509 yields a sample size of 39. In order to ensure our tests analyzing the subpopulation of risk-tolerant subjects are adequately powered, we reviewed three BRET studies reporting the proportions of risk-averse, risk-tolerant, and risk-neutral responses and found that, on average, 24.3% of participants exhibited risk-tolerant behavior (Crosetto and Filippin, 2013; Gioia, 2017; Soetevent and Romensen, 2017). Our analysis, therefore, indicates a total sample size of 218 individuals provides adequate power to test our secondary hypotheses regarding heterogeneity by risk preferences. To account for attrition and unusable responses, we plan to target 300 total respondents which would yield approximately 199 risk-averse subjects, 73 risk-tolerant subjects, and 29 risk-neutral subjects. In the event that our initial sample does not reach the required threshold of 53 risk-tolerant subjects, we intend to extend our data collection to additional subjects until the quota is met.

References for Experimental Details

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Routledge.

Crosetto, P., & Filippin, A. (2013). The 'bomb' risk elicitation task. *Journal of Risk and Uncertainty, 47*(1).

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association, 9*(3).

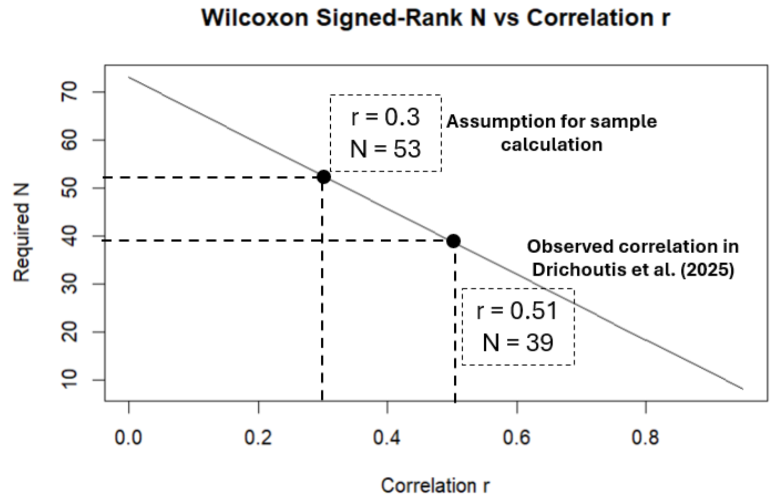

Figure 1: Since a higher correlation reduces σ_{diff}, our assumption of $r = 0.3$ means our sample size estimate is conservative.

Drichoutis, A. C., Palma, M. A., & Feldman, P. (2025). *Incentives and payment mechanisms in preference elicitation* [unpublished].

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–1160.

Gioia, F. (2017). Peer effects on risk behavior: The importance of group identity. *Experimental Economics*, *20*(1).

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. *Frontiers in psychology*, *4*, 863.

Rosner, B. A., et al. (2006). *Fundamentals of biostatistics* (Vol. 6). Thomson-Brooks/Cole Belmont, CA.

Soetevent, A., & Romensen, G. (2017). Tailored feedback and worker green behavior: Field evidence from bus drivers. *Tinbergen Institute Discussion Paper*, *No. 17-073/VII.*

5 Institutional Review Board

1. IRB Name: Texas A&M Institutional Review Board
2. IRB Approval Date: August 26, 2025
3. IRB Approval Number: STUDY2025-0840

6 Analysis Plan

6.1 Theoretical Considerations

The BDM mechanism is one of the most widely used methods in experimental economics to elicit willingness to pay (WTP) or willingness to accept (WTA) for goods and services (Becker et al., 1964). Under the standard expected utility framework, changes in the upper bound of the support set should not impact a subject's WTA (or WTP) for a fixed induced value (IV). In the BDM mechanism, submitting an offer equal to the IV, i.e., providing a truthful response, is a weakly dominant strategy and always maximizes the expected payoff. In practice, however, systematic deviations from truth-telling are well documented (Bohm

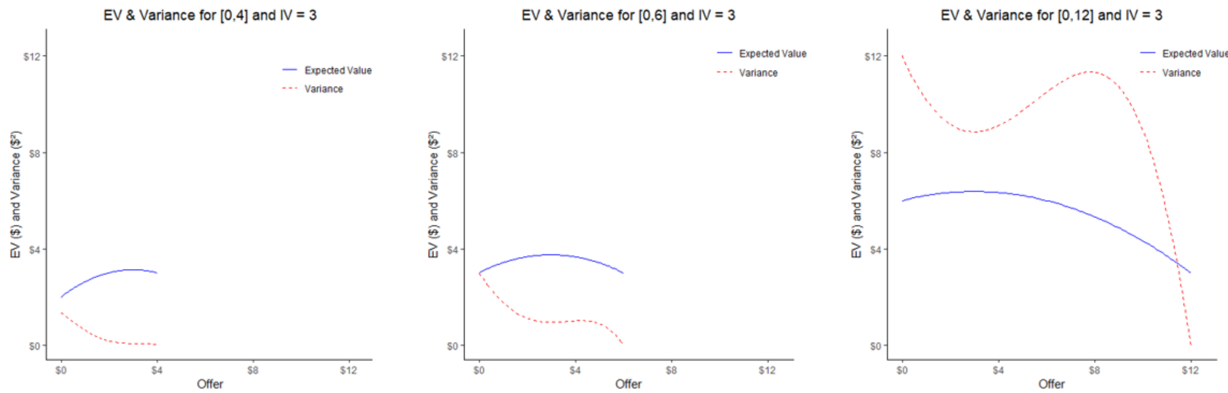

Figure 2: Expected value and variance plots across all offer values for each of the three treatments

et al., 1997; Irwin et al., 1998; Noussair et al., 2004; Banerji and Gupta, 2014; Cason and Plott, 2014; Flynn et al., 2016; Mamadehussene and Sguera, 2023; Drichoutis et al., 2025). We propose that the selection of the support set range influences subjects' under/overbidding behavior through determining the curvature of the variance function which establishes the risk/reward profile of a given bid.

We derive the payoff variance for the IV WTA setting of the BDM mechanism, akin to the Cason and Plott (2014) experimental setup, and show that its behavior depends on the specific values chosen for the range of the support set for the randomly generated market offer. Intuitively, the payoff as a function of an offer can be considered a combination of the outcomes at the high and low end of the range. In the WTA case, an offer at the high end of the range results in the subject receiving the IV with certainty because a transaction never occurs. At the low end of the range, the payoff reduces to the underlying market offer distribution, which is typically uniform, since a transaction always occurs. For a uniform distribution, we show that an offer equal to the IV corresponds to a local variance minimum; however, an offer at the high end of the support set, with zero variance, is always the global minimum.

This result opens the possibility that offers away from the IV could maximize subject's utility by choosing a payoff profile exhibiting a relatively greater degree of certainty (or narrower outcome dispersion) at a lower expected value than the risk-neutral strategy. Since the selection of the upper bound of the support set establishes this expected value and variance tradeoff, we empirically test this theoretical insight in a controlled experiment. In our design, we utilize the BDM mechanism to elicit valuations for an item that subjects can redeem for $3, the IV, by asking for the minimum amount of money required to sell the card back to the experimenter. We aim to isolate the effect of changing the payoff variance on valuations by shifting the upper bound of the support set while holding all other factors constant (e.g., item redemption value, support set lower bound at $0). Figure 2 illustrates the expected value and variance in our three treatments with the $4, $6, and $12 upper bounds.

6.2 Primary Analysis

- Mean Bid by Treatment:

  Let μ_i, for $i \in \{4, 6, 12\}$ denote the mean WTA value for the treatment representing the upper bound of the support set distribution of $4, $6, and $12, respectively.

  $H1$: $\mu_{12} > 3$

  $H2$: $\mu_6 = 3$

  $H3$: $\mu_4 < 3$

  $H1$ and $H3$ are evaluated using one-sided, single-sample Wilcoxon signed-rank tests (Wilcoxon, 1945) with $H2$ evaluated with a two-sided, single-sample Wilcoxon signed-rank test. We adopt the Holm-Bonferroni step-down procedure for multiple hypothesis testing (Holm, 1979).

We first test whether exogenously changing the support set upper bound is associated with deviations in the offer distribution away from the IV. Our primary analysis tests this in terms of the central tendency of the WTA distributions across rounds while our robustness checks, outlined below, test for changes in behavior via regression analysis, difference in proportions, WTA distribution skewness.

Given that we expect a majority of the sample to be risk-averse, we anticipate the behavior of that group to determine the central tendency of the offer distribution. In the \$12 treatment, the payoff variance function declines steeply to zero for offer values toward the high end of the support set, indicating a large reduction in outcome variability for a small reduction in expected payoff value. If risk-averse subjects display a level of variance aversion, we would then expect a higher concentration of offers above the IV in the \$12 treatment (i.e. more overbidding) because they achieve utility maximization at an offer above the IV with a slight decrease in expected payoff and a larger decrease in payoff variance while risk-tolerant participants would bid below the IV because the variance function achieves its maximum at \$0 and is strictly decreasing over the $[0, IV]$ interval.

Similarly, the variance function in the \$6 upper bound treatment is decreasing over a relatively wider interval of offers greater than the IV ($[4.25, 6]$); however, the lesser degree of curvature created by a support set symmetric about the IV creates a correspondingly worse risk/reward tradeoff than the \$12 treatment, implying the bid distribution should be more symmetric about the IV as well. Accordingly, we predict $\mu_6 = 3$ as we do not expect the average to differ significantly from the IV.

In the \$4 treatment, the flatter curve results in the least favorable mean-variance profile above the IV, so risk-averse subjects should provide valuations closer to the expected-value maximum at the IV, while risk-tolerant participants should still tend to bid below the IV, causing the overall sample average to reflect underbidding.

- <u>Secondary Analyses</u>:

  1. <u>Alternative Mean Bid Tests by Treatment</u>:
     We also plan to conduct pairwise tests for statistically significant differences among the mean bids by treatment. This enables us to make direct comparisons among the treatments in addition to the means relative to the IV:

     $\mu_{12} > \mu_6 > \mu_4$

     These tests are evaluated using one-sided, paired Wilcoxon signed-rank tests (Wilcoxon, 1945).

  2. <u>Regression Analysis</u>:
     As a robustness check of the primary analysis, we will use an OLS regression including additional demographic control variables. Denote the \$4, \$6, and \$12 upper bound treatments by $i \in \{4, 6, 12\}$, respectively, and the j-th subject's WTA in the i-th treatment as $WTA_{i,j}$. We will estimate the following specifications:

$$WTA_{i,j} = \alpha_0 + \alpha_1 d_{4,j} + \alpha_2 d_{12,j} + \varepsilon_{i,j} \quad (4)$$

$$WTA_{i,j} = \alpha_0 + \alpha_1 d_{4,j} + \alpha_2 d_{12,j} + \delta \mathbf{X}_j + \varepsilon_{i,j} \quad (5)$$

     where $d_{i,j}$ is a binary variable taking the value of 1 for the j-th subject's offer in treatment i for $i \in \{4, 12\}$, omitting the \$6 round to serve as the baseline and $\mathbf{X}_j$ represents a vector of observable demographic controls for subject j including age, gender, ethnicity, and income. The coefficients in Equation (4) then represent the unconditional mean offers by treatment and, in Equation (5), the conditional average controlling for demographic factors. We anticipate the predicted results of our primary analysis to hold, controlling for demographic factors:

$\alpha_1 < 0$
$\alpha_2 > 0$

These hypotheses are evaluated first by F-tests for joint significance and then by t-tests with standard errors clustered at the individual level in both sets of tests.

3. Difference in Proportions:

   The use of only the WTA distributions' means, either unconditional or conditional, in identifying treatment effects could conceal offsetting effects above and below the induced value. As an additional robustness check, we test the hypothesis that the probability of an offer above the induced value is increasing across treatments. The approach is akin to the one adopted in Knetsch and Sinden (1984), which examines the change proportion of subjects willing to accept a given monetary offer in exchange for different endowed items. The difference in proportions test enables us to identify a shift in the number of subjects with offer levels above the IV (overbidding) in the case where the overall WTA distribution mean remains unchanged if the portion of the sample bidding below the IV moves closer to zero (underbidding).

   $P(WTA_{12,j} > 3) > P(WTA_{6,j} > 3)$
   $P(WTA_{6,j} > 3) > P(WTA_{4,j} > 3)$
   $P(WTA_{12,j} > 3) > P(WTA_{4,j} > 3)$

   These hypotheses are evaluated jointly with Cochran's Q test with pairwise comparisons using one-sided McNemar tests (Cochran, 1950; McNemar, 1947).

4. Skewness:

   Let $\gamma_i, i \in \{4, 6, 12\}$ denote the skewness of the bids in each treatment.

   $\gamma_{12} > 0$
   $\gamma_6 = 0$
   $\gamma_4 < 0$

   These hypotheses are evaluated using the skewness-specific statistic from D'Agostino's K^2 test for normality (D'agostino et al., 1990). We conduct one-sided tests in the \$12 and \$4 treatments and utilize a two-sided test in the \$6 treatment to align with our predictions.

   The expected payoff function is symmetric about the IV; however, payoff variance is not. Although several other theories offer explanations for observed over- and underbidding in BDM settings, the mean-variance framework suggests the extreme values of the variance function at the end points of the support set could result in an asymmetric offer distribution. At an offer of \$0, the variance simplifies to that of the underlying uniform posted price which is the maximum point in all 3 treatments. An offer equal to the upper bound of the support set reduces the variance to zero as a subject will always retain their item to be redeemed for the IV. If risk-aversion extends to variance in this manner, we would then expect a right-skewed offer distribution when the variance curve offers a sufficient risk/reward tradeoff. Alternatively, under a flatter variance curve, risk-tolerant subjects with offers between \$0 and the IV would cause a left-skewed WTA distribution. Correspondingly, we hypothesize $\gamma_4 < 0$, $\gamma_6 = 0$, and $\gamma_{12} > 0$.

6.3 Heterogeneity Analyses and Robustness Checks

1. Mean Bids by Risk Preference:

   In the primary analysis, we derive predictions from the assumption that subjects respond heterogeneously to changes in the upper bound of the support set according to their risk preferences. Using payoff variance as a proxy for the risk of a given bid, risk-averse subjects should exhibit a tendency to overbid by choosing offer values closer to the upper bound that yield a more certain outcome.

Conversely, risk-tolerant subjects are expected to show a preference toward the higher-variance payoff profiles of offers closer to the support set's zero lower bound, leading to underbidding on average.

We utilize the BRET responses to classify each subject as risk averse, risk tolerant, or risk neutral. The selection of 50 boxes is risk neutral, maximizing the task's expected value (Crosetto and Filippin (2013)). Responses below 50 then represent risk-averse behavior with responses above 50 corresponding to risk-tolerant preferences. Let k_j denote the number of boxes selected by the j-th subject in the BRET. We define subjects with $k_j < 50$ to be risk averse and subjects with $k_j > 50$ to be risk tolerant. Then, let $\mu_{i,RA}$ and $\mu_{i,RT}$ for $i \in \{4, 6, 12\}$ denote the mean bid by treatment level for each risk preference category. In all cases, we expect the average of the risk-averse subjects' offers to exceed that of the risk-tolerant subjects and test:

$$\mu_{12,RA} > \mu_{12,RT}$$

$$\mu_{6,RA} > \mu_{6,RT}$$

$$\mu_{4,RA} > \mu_{4,RT}$$

As discussed in the power analysis, we set our 300 subject target to conduct the secondary heterogeneity analyses. Based on our review of studies using the BRET that report response distributions, we expect approximately 199 risk-averse subjects, 73 risk-tolerant subjects, and 29 risk-neutral subjects which would provide enough risk-tolerant individuals to adequately power the heterogeneity analyses (Crosetto and Filippin, 2013; Gioia, 2017; Soetevent and Romensen, 2017). While the figure is included for reference, we exclude subjects with risk-neutral responses of $k = 50$ in the BRET from the heterogeneity analyses but retained in the aggregate treatment tests.

This hypothesis is evaluated using the Mann-Whitney U test (Mann and Whitney, 1947). For a robustness check, we employ an alternative risk preference scheme using the risk self-evaluation question from the German Socio-Economic Panel survey (Dohmen et al., 2011). Subjects are asked to report how willing they are to take risks on a scale of 0 to 10. We then classify responses below 5 as risk averse and responses above 5 as risk tolerant and perform a second set of tests on the resulting mean bids by treatment.

2. Heterogeneity Regression Analysis:

Similar to the primary analysis, we estimate a linear regression model using OLS, incorporating additional control variables to serve as a robustness check for the unconditional difference in means test. Accordingly, we estimate the specifications:

$$WTA_{i,j} = \alpha_0 + \alpha_1 d_{4,j} + \alpha_2 d_{12,j} + \alpha_3 k_j + \delta \mathbf{X}_j + \varepsilon_{i,j} \tag{6}$$

$$WTA_{i,j} = \alpha_0 + \alpha_1 d_{4,j} + \alpha_2 d_{12,j} + \alpha_3 k_j + \alpha_4 d_{4,j} k_j + \alpha_5 d_{12,j} k_j + \delta \mathbf{X}_j + \varepsilon_{i,j} \tag{7}$$

$$\begin{aligned} WTA_{i,j} = \alpha_0 &+ \alpha_1 d_{4,j} + \alpha_2 d_{12,j} + \alpha_3 (k_j - 50) + \alpha_4 d_{4,j}(k_j - 50) + \alpha_5 d_{12,j}(k_j - 50) \\ &+ \alpha_6 D_j + \alpha_7 D_j (k_j - 50) + \alpha_8 D_j d_{4,j} + \alpha_9 D_j d_{12,j} + \delta \mathbf{X}_j + \varepsilon_{i,j} \end{aligned} \tag{8}$$

with D_j is a binary variable indicating RT subjects i.e., $k_j > 50$ and all other variables maintain their definitions. We again use the \$6 treatment as the baseline for the binary variables. Equation (6) enables us to test for treatment and risk type effects as measured by the BRET on offer values while the inclusion of the $d_{i,j} k_j$ interaction terms in Equation (7) indicates whether that effect varies across treatments, holding other factors constant.

Equation (8) uses the D_j binary variable to explore heterogeneous changes in WTA based on their risk classification. We center the k_j variables in Equation (8) to reduce inflation in coefficient standard errors due to multicollinearity as D_j is a function of k_j. Running a comparable regression on the \$3 IV treatments in Drichoutis et al. (2025) with uniformly distributed random values used for simulated

BRET responses reduced the mean variance inflation factor (VIF) from 7.9 to 4.7 and decreased the individual coefficient VIFs related to the D_j and k_j from a range of approximately 23 to 28 to a range of 12 to 15. A simulated normal distribution of BRET responses reduced the VIFs even further.

Hypothesis
$\alpha_1 < 0$
$\alpha_2 > 0$
$\alpha_3 < 0$
$\alpha_4 < 0$
$\alpha_5 < 0$
$\alpha_6 > 0$
$\alpha_7 < 0$
$\alpha_8 > 0$
$\alpha_9 > 0$

Table 1: Hypotheses for the heterogeneity regression analysis

Table 1 presents our hypotheses by specification. We focus on Equation (7) as the richest model for inferential testing but will report results for all three regressions. The α_1 and α_2 predictions are consistent with the primary hypothesis in that $\alpha_1 < 0$ implies lower WTA values in the \$4 upper bound treatment than the \$6 treatment. Similarly, $\alpha_2 > 0$ implies higher WTA values in the \$12 treatment than the \$6 treatment. We also predict WTA to be decreasing in risk tolerance (i.e., larger values of k_j), reflected by α_3, α_4, and α_5, as offers closer to the zero lower bound result in larger payoff variance. In Equation (8), we expect $\alpha_6 > 0$, $\alpha_7 < 0$, $\alpha_8 > 0$, and possibly $\alpha_9 > 0$, reflecting higher baseline bids and stronger sensitivity to low- and high-variance treatments among risk-tolerant participants.

These hypotheses are evaluated first by F-tests for joint significance and then by t-tests with standard errors clustered at the individual level in both sets of tests.

3. Difference in Bid Distributions:

   A systematic behavioral pattern in response to a shift in payoff variance would be indicative of a connection between over- and underbidding in the BDM mechanism and risk preferences; however, identical distributions suggest some alternative causal mechanism. Thus, we test to determine whether offers from risk-averse and risk-tolerant subjects follow the same distribution. We adopt the same risk classification procedure using BRET responses. Denote the distribution of WTA values for risk-averse and risk-tolerant subjects for each treatment by $WTA_{i,RA}$ and $WTA_{i,RT}$, respectively, for $i \in \{4, 6, 12\}$. Similar to the mean bid tests, we will compare results under the alternative classification scheme based on the risk self-evaluation question as a robustness check. We test:

$$WTA_{12,RA} \neq WTA_{12,RT}$$
$$WTA_{6,RA} \neq WTA_{6,RT}$$
$$WTA_{4,RA} \neq WTA_{4,RT}$$

   These hypotheses are evaluated with the Kolmogorov-Smirnov test (Massey Jr, 1951).

4. Heterogeneity in Treatment Effect Size:

   Because the global maximum and minimum of the variance function occur at the lower and upper bounds of the support set, respectively, we expect subjects with greater risk preference magnitude to exhibit larger absolute treatment effects. We approach this hypothesis in two distinct manners. First, we test for a positive correlation between the each subject's distance from risk neutrality in the BRET and their observed treatment effects in the BDM task. For the j-th subject, let $|k_j - 50|$ represent the magnitude of the risk preference, and let $|TE_{i,j}| = |WTA_{i,j} - WTA_{6,j}|$ for $i = \{4, 12\}$ denote the absolute treatment effects, again using the \$6 round as the baseline. We test:

$Corr(|k_j - 50|, |TE_{4,j}|) > 0$

$Corr(|k_j - 50|, |TE_{12,j}|) > 0$

using the Pearson correlation test with the Spearman rank test as a robustness check against non-normal distributions of $|k_j - 50|$ and the absolute treatment effects (Pearson, 1895; Spearman, 1904/1987).

We then use an OLS regression-based approach to test the conditional relationship between risk preference magnitude and absolute treatment effect size. We estimate the model:

$$|TE_{i,j}| = \beta_0 + \beta_1 |k_j - 50| + \delta \mathbf{X} + \varepsilon_{i,j} \tag{9}$$

$\beta_1 > 0$

by a t-test with standard errors clustered at the individual level.

5. Consistency of Risk Preference Measures:

Adopting a utility functional form in the financial economics literature, we model subject utility as the second degree Taylor expansion of payoff moments (Markowitz, 1952; Pratt, 1978; Garlappi and Skoulakis, 2011). Let subject utility for the elicited WTA value in the i-th round by the j-th subject be represented by:

$$u(WTA_{i,j}) = \mu_{WTAi,j} + \frac{\lambda_{i,j}}{2}\sigma^2_{WTA_{i,j}} \tag{10}$$

where $\mu_{WTA_{i,j}}$ and $\sigma^2_{WTA_{i,j}}$ are the expected value and variance of the payoff when offering \$WTA, respectively, and $\lambda_{i,j}$ is the coefficient of interest. Assuming that a subject's offer represents their utility-maximizing choice across the set of feasible payoff mean-variance combinations in each treatment, the variance coefficient λ captures the impact of a change in variance on utility levels. When the offer is not equal to the IV (i.e., $WTA \neq \$3$), the value of λ implied by a given offer can be uniquely identified by the utility function's first order condition

$$\lambda^* = -2\frac{\partial E(\pi(WTA))}{\partial WTA} \cdot \frac{\partial WTA}{\partial Var(\pi(WTA))} \tag{11}$$

where $\pi(WTA)$ is the payoff function evaluated at WTA and $E(\cdot)$ & $Var(\cdot)$ are the typical expectation and variance operators. Thus, $\lambda_{i,j}$ becomes the Arrow-Pratt risk aversion measure in this setting. Given the parameters in each treatment, we proceed to calculate each $\lambda_{i,j}$. When $WTA_{i,j} = \$3$, $\lambda_{i,j}$ is undefined as risk-neutral observed behavior can be consistent with a range of actual risk preferences (Pratt, 1978; Rabin, 2000). Figure 3 presents the values of λ implied for every bid value in each treatment.

To assess whether the variance preferences implied by subjects' choices are connected with our measures of risk preferences, we test the correlation between the number of boxes collected in the BRET by the j-th subject, k_j, and $\lambda_{i,j}$ for each treatment. In computing the average of λ, we omit observations where $WTA = \$3$. If a subject offered \$3 in all three rounds, then that subject is dropped entirely from the correlation test. We hypothesize that the correlation will be positive under the expectation that risk-averse subjects will select both small values of k and WTA such that $\lambda < 0$ while risk-tolerant subjects will select both large values of k and WTA such that $\lambda > 0$. Therefore, we test:

$Corr(\lambda_{12,j}, k_j) > 0$

$Corr(\lambda_{6,j}, k_j) > 0$

$Corr(\lambda_{4,j}, k_j) > 0$

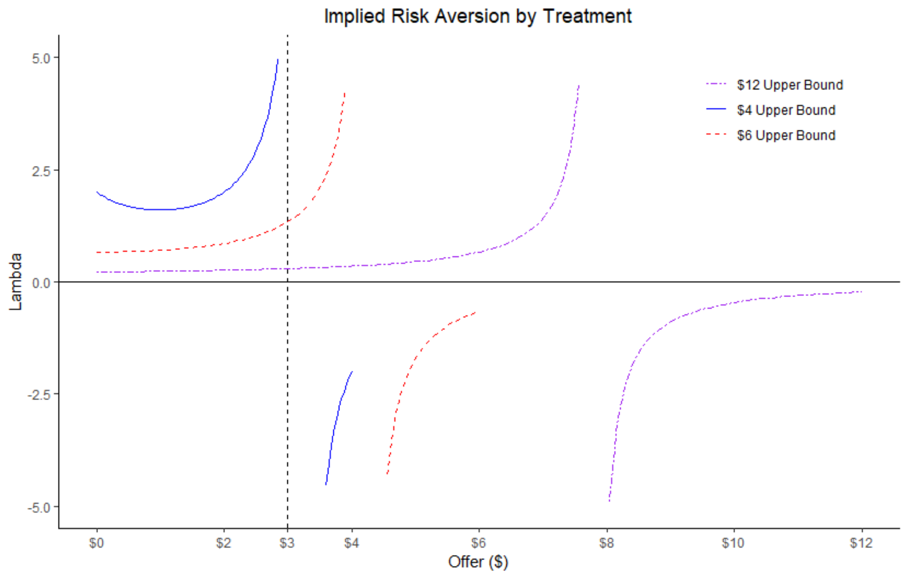

Figure 3: The vertical asymptotes reflect the point at which the variance function begins to slope downward toward zero. Positive values of λ reflect risk-tolerant behavior, and negative values indicate risk-averse behavior.

This is evaluated using the Pearson correlation test with the Spearman rank test as a robustness check against non-normal distributions of $\lambda_{i,j}$ and k_j (Pearson, 1895; Spearman, 1904/1987). As a robustness check, we conduct the same tests of correlation with $\lambda_{i,j}$ and the responses to the risk self-evaluation question.

6. Mean-Variance Regression Analysis:

   Let $\mu_{i,j}$ and $\sigma^2_{i,j}$ denote the expected value and variance of the payoff function evaluated at $WTA_{i,j}, i \in \{4, 6, 12\}$ for each subject j, and let k_j represent the j-th subject's BRET response. As $\mu_{i,j}$ and $\sigma^2_{i,j}$ are deterministic functions of $WTA_{i,j}$, the two variables along with the $k_j\sigma^2_{i,j}$ interaction term are endogenous in any regression specification unless addressed. We leverage the exogenous variation in the support set upper bound to enable parameter identification by using the interaction effects between the treatment dummy variables and k_j as a set of four instruments for the three endogenous variables. Adopting a two-stage fixed effects approach, our first stage regressions are:

$$\mu_{i,j} = \tau_0 + \tau_1 d_{4,j} + \tau_2 d_{12,j} + \tau_3 k_j d_{4,j} + \tau_4 k_j d_{12,j} + \eta_j + \varepsilon_{i,j} \tag{12}$$

$$\sigma^2_{i,j} = \tau_0 + \tau_1 d_{4,j} + \tau_2 d_{12,j} + \tau_3 k_j d_{4,j} + \tau_4 k_j d_{12,j} + \eta_j + \varepsilon_{i,j} \tag{13}$$

$$k_j\sigma^2_{i,j} = \tau_0 + \tau_1 d_{4,j} + \tau_2 d_{12,j} + \tau_3 k_j d_{4,j} + \tau_4 k_j d_{12,j} + \eta_j + \varepsilon_{i,j} \tag{14}$$

   where $d_{i,j}$ are indicator functions for offers in rounds with the upper bound of the support set at $i = \{4, 12\}$, treating the \$6 round as the baseline, and η_j represents individual-level fixed effects. The first-stage regressions yield estimates for $\hat{\mu}_{i,j}$, $\hat{\sigma}^2_{i,j}$, and $\widehat{k_j\sigma}^2_{j,k}$ to be used in the second stage. We then estimate the following fixed-effects model, employing $WTA_{i,j}$ as the dependent variable and controlling for fixed within-subjects factors:

$$WTA_{i,j} = \theta_0 + \theta_1\hat{\mu}_{i,j} + \theta_2\hat{\sigma}^2_{i,j} + \theta_3\widehat{k_j\sigma}^2_{i,j} + \eta_j + \varepsilon_{i,j} \tag{15}$$

where η_j represents individual-level fixed effects. While the impact of risk preferences is implicitly included in the fixed effects term, the interaction effect with the variance term is identifiable. The partial effects indicate how subject WTA changes with a shift in the first two moments of the BDM payoff, controlling for individual-specific factors via η_j. Other factors held constant, we expect WTA to be increasing in expected payoff and decreasing in payoff variance with larger values of k_j (i.e., more risk-tolerance) attenuating the negative variance effect. Accordingly, we test:

$\theta_1 > 0$

$\theta_2 < 0$

$\theta_3 > 0$

These hypotheses are evaluated first by F-tests for joint significance and then by t-tests with standard errors clustered at the individual level in both sets of tests. As a robustness check, we also estimate the model with a vector $\mathbf{X_j}$ of observable demographic factors (age, gender, ethnicity, income) in lieu of the fixed effect term.

6.4 Exploratory Analyses

Our study includes several questions aimed at gaining insights into subjects' decision-making processes in the BDM task. We plan to investigate potential relationships between WTA and focuses on increasing the probability of "winning," maximizing payoff value, and desiring to control the task outcome. While not the primary focus of this paper, responses to these questions may inform future work. One of the questions is an open-ended response box requesting subjects to outline their strategy explicitly. Depending on the richness of the answers, we plan to use the techniques outlined in Hassan et al. (2025) to identify trends in text responses and establish behavioral patterns.

We also placed hidden text in the HTML of one page that is invisible to subjects and de-emphasized for screen-reading software but able to be parsed by AI/LLM agents and other bots. That text provides specific instructions on how to answer one question in a manner that, if followed, will likely enable us to identify the respondent as non-human. We have also enabled Qualtric's Captcha bot scoring system. With these two metrics, we hope to provide insights into methods for improving data quality in online experimental surveys and the portion of responses flagged as likely automated.

Our experimental survey requires subjects to correctly answer 11 comprehension questions covering both the structure of the BDM mechanism as well as payoff calculations in a variety of scenarios. We included the extensive suite of questions to improve subject understanding of the task; however, we cannot ensure with certainty that correct responses indicate understanding as opposed to trial and error. As a consequence, we track the number of incorrect attempts at answering the comprehension questions and plan to explore the robustness of the analyses to varying exclusion criteria based on the comprehension questions.

6.5 Future Extensions

Experimental results in line with the hypotheses outlined above would provide evidence that the selection of the upper bound of the support set in BDM experiments predictably changes participants' tendencies to offer valuations above or below the IV. Pending the initial results with subsequent analysis and feedback, we may also perform a second iteration of this experiment wherein changes in the payoff variance function are achieved through varying the IV at \$3, \$6, and \$9 against a static \$12 upper bound which produces similar risk/reward profiles to this experiment as the slope of the variance curve is determined by the relative placement of the IV in the range from 0 to the upper bound. Holding the offer range constant between rounds would help to distinguish a variance-related treatment impact on WTA from alternative causal effects like anchoring and midpoint biases (Tversky and Kahneman, 1974; Thomas and Kyung, 2019; Crosetto et al., 2020).

References for Analysis Plan

Banerji, A., & Gupta, N. (2014). Detection, identification, and estimation of loss aversion: Evidence from an auction experiment. *American Economic Journal: Microeconomics*, *6*(1).

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3).

Bohm, P., Linden, J., & Sonnegard, J. (1997). Eliciting reservation prices: Becker-degroot-marschak mechanisms vs. markets. *The Economic Journal*, *107*(443).

Cason, T. N., & Plott, C. R. (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy*, *122*(6).

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3/4), 256–266.

Crosetto, P., & Filippin, A. (2013). The 'bomb' risk elicitation task. *Journal of Risk and Uncertainty*, *47*(1).

Crosetto, P., Filippin, A., Katuščák, P., & Smith, J. (2020). Central tendency bias in belief elicitation. *Journal of Economic Psychology*, *78*, 102273.

D'agostino, R. B., Belanger, A., & D'Agostino Jr, R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, *44*(4), 316–321.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3).

Drichoutis, A. C., Palma, M. A., & Feldman, P. (2025). *Incentives and payment mechanisms in preference elicitation* [unpublished].

Flynn, N., Kah, C., & Kerschbamer, R. (2016). Vickery auction vs bdm: Difference in bidding behaviour and other-regarding motives. *Journal of the Economic Science Association*, *2*(2).

Garlappi, L., & Skoulakis, G. (2011). Taylor series approximations to expected utility and optimal portfolio choice. *Mathematics and Financial Economics*, *5*(2).

Gioia, F. (2017). Peer effects on risk behavior: The importance of group identity. *Experimental Economics*, *20*(1).

Hassan, T. A., Hollander, S., Kalyani, A., van Lent, L., Schwedeler, M., & Tahoun, A. (2025). Text as data in economic analysis. *Journal of Economic Perspectives*, *39*(3).

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.

Irwin, J. R., McClelland, G. H., McKee, M., Schulze, W. D., & Norden, N. E. (1998). Payoff dominance vs. cognitive transparency in decision making. *Economic Inquiry*, *36*(2).

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics*, *99*(3), 507–521.

Mamadehussene, S., & Sguera, F. (2023). On the reliability of the bdm mechanism. *Management Science*, *69*(2).

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1).

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68–78.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157.

Noussair, C., Robin, S., & Ruffieux, B. (2004). Revealing consumers' willingness-to-pay: A comparison of the bdm mechanism and the vickery auction. *Journal of Economic Psychology*, *25*(6).

Pearson, K. (1895). Vii. note on regression and inheritance in the case of two parents. *proceedings of the royal society of London*, *58*(347-352), 240–242.

Pratt, J. W. (1978). Risk aversion in the small and in the large. In *Uncertainty in economics* (pp. 59–79). Elsevier.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, *68*(5), 1281.

Soetevent, A., & Romensen, G. (2017). Tailored feedback and worker green behavior: Field evidence from bus drivers. *Tinbergen Institute Discussion Paper, No. 17-073/VII.*
Spearman, C. (1987). The proof and measurement of association between two things. *The American journal of psychology*, *100*(3/4), 441–471. (Original work published 1904)
Thomas, M., & Kyung, E. J. (2019). Slider scale or text box: How response format shapes responses. *Journal of Consumer Research*, *45*(6), 1274–1293.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124–1131.
Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, *1*(6), 80–83.

7 Mathematical Appendix

In this section, we provide the derivations of (i) the payoff expected value and (ii) the payoff variance function and (iii) the formulae used for the implied risk aversion coefficients, λ, in Heterogeneity Analysis 5.

7.1 Payoff Expected Value

Consider the WTA BDM setting for an induced value item. Let IV represent the item's induced value, x represent a realization from the continuous uniform distribution on $[\alpha, \beta]$, or the experimenter's selected support range for the randomly generated market offer, and $b \in [\alpha, \beta]$ represent a subject's bid or offer. The probability density function of x is

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} \text{ for } \alpha \leq x \leq \beta \\ 0 \text{ otherwise.} \end{cases} \tag{16}$$

In the WTA version of the BDM mechanism, the subject redeems the item for IV if their offer is less than or equal to the randomly generated market offer, x. Otherwise, the subject sells the item and receives x. We can then define the payoff function as

$$\pi(b) = \begin{cases} IV \text{ for } x \leq b \\ x \text{ for } x > b. \end{cases} \tag{17}$$

The expected value of $\pi(b)$ is then

$$\mathbb{E}[\pi(b)] = \int_{\alpha}^{\beta} \pi(b) f(x) dx = \int_{\alpha}^{b} \frac{IV}{\beta - \alpha} dx + \int_{b}^{\beta} \frac{x}{\beta - \alpha} dx = \frac{IV(b-\alpha)}{\beta - \alpha} + \frac{\beta^2 - b^2}{2(\beta - \alpha)} \tag{18}$$

which simplifies to

$$\mathbb{E}[\pi(b)] = \frac{2IV(b-\alpha) + \beta^2 - b^2}{2(\beta - \alpha)}. \tag{19}$$

7.2 Payoff Variance

The variance of the payoff function at a given offer is defined by $\text{Var}(\pi(b)) = \mathbb{E}[\pi(b)^2] - (\mathbb{E}[\pi(b)])^2$. So we have

$$\mathbb{E}[\pi(b)^2] = \int_{\alpha}^{b} \frac{IV^2}{\beta - \alpha} dx + \int_{b}^{\beta} \frac{x^2}{\beta - \alpha} dx. \tag{20}$$

Then,

$$\int_{\alpha}^{b} \frac{IV^2}{\beta - \alpha} dx = \frac{IV^2(b-\alpha)}{\beta - \alpha}, \tag{21}$$

and

$$\int_{b}^{\beta} \frac{x^2}{\beta - \alpha} dx = \frac{\beta^3 - b^3}{3(\beta - \alpha)}. \tag{22}$$

Therefore,

$$\mathbb{E}[\pi(b)^2] = \frac{IV^2(b-\alpha)}{\beta-\alpha} + \frac{\beta^3 - b^3}{3(\beta-\alpha)} = \frac{3IV^2(b-\alpha) + \beta^3 - b^3}{3(\beta-\alpha)}, \tag{23}$$

implying

$$\mathrm{Var}(\pi(b)) = \frac{3IV^2(b-\alpha) + \beta^3 - b^3}{3(\beta-\alpha)} - \left(\frac{2IV(b-\alpha) + \beta^2 - b^2}{2(\beta-\alpha)}\right)^2. \tag{24}$$

7.3 Risk Aversion Coefficient

Consider a subject with initial wealth w_0 with $\pi(b)$ representing the random payoff in the previously described WTA BDM mechanism, and let $\mu_b = \mathbb{E}[\pi(b)]$ and $\sigma_b^2 = \mathrm{Var}(\pi(b))$. Assume there exists some C^3 utility function $u(\cdot)$ over final wealth states. The second-order Taylor expansion of $u(w_0+\pi(b))$ around $w^* = w_0+\mu_b$ is

$$u(w_0 + \pi(b)) \approx u(w^*) + u'(w^*)(\pi(b) - \mu_b) + \frac{u''(w^*)}{2}(\pi(b) - \mu_b)^2. \tag{25}$$

Then expected utility is approximated by

$$\mathbb{E}[u(w_0 + \pi(b))] \approx u(w^*) + u'(w^*)\mathbb{E}[\pi(b) - \mu_b] + \frac{u''(w^*)}{2}\mathbb{E}[(\pi(b) - \mu_b)^2] \approx u(w^*) + \frac{u''(w^*)}{2}\sigma_b^2. \tag{26}$$

Taking a first-order expansion of $u(w^*)$ around w_0 yields

$$u(w^*) \approx u(w_0) + u'(w_0)\mu_b \text{ (error } O(\mu_b^2)), \tag{27}$$

and the zeroth-order expansion of $u''(w^*)$ around w_0 is

$$u''(w^*) \approx u''(w_0) \text{ (error } O(\mu_b)) \tag{28}$$

so by substituting Equations (27) and (28) back into Equation (26) shows

$$\mathbb{E}[u(w_0 + \pi(b))] \approx u(w_0) + u'(w_0)\mu_b + \frac{u''(w_0)}{2}\sigma_b^2. \tag{29}$$

Dropping the $u(w_0)$ term which does not depend on b and normalizing by dividing through by $u'(w_0) > 0$, a positive affine transformation that does not impact the location of the utility maximizing bid, then gives

$$\mathbb{E}[u(w_0 + \pi(b))] \approx \mu_b + \frac{u''(w_0)}{2u'(w_0)}\sigma_b^2 \tag{30}$$

which shows λ in Equation (10) is equivalently $\frac{u''(w_0)}{u'(w_0)}$ or the negative of the Arrow-Pratt absolute risk aversion coefficient (Pratt, 1978).

Assuming a subject's bid indicates a preference over the available mean-variance combinations offered in a given BDM round, let $b^* = \arg\max_b \mathbb{E}[u(w_0 + \pi(b))]$. Then, the implied λ^* is the value that satisfies the first order condition in maximizing Equation (30):

$$\lambda^* = \frac{-2\mu_b'}{(\sigma_b^2)'} = -2\frac{\partial \mathbb{E}[\pi(b)]}{\partial b} \cdot \frac{\partial b}{\partial \mathrm{Var}(\pi(b))} \tag{31}$$

as given in Equation (11). The negative sign follows the convention that $u''(w_0) < 0$ under risk aversion, implying $\lambda^* > 0$ represents risk-tolerant preferences while $\lambda^* < 0$ represents risk-averse preferences. From Equation (19),

$$\frac{\partial \mathbb{E}[\pi(b)]}{\partial b} = \frac{IV - b}{\beta - \alpha} \tag{32}$$

and from Equation (24),

$$\frac{\partial \mathrm{Var}(\pi(b))}{\partial b} = \frac{IV^2 - b^2}{\beta-\alpha} - \frac{b^3 - 3IVb^2 + (2IV^2 + 2IV\alpha - \beta^2)b + IV\beta^2 - 2IV^2\alpha}{(\beta-\alpha)^2} \tag{33}$$

which simplifies to

$$\frac{(IV-b)(b^2+(\beta-\alpha-2IV)b+(IV\beta+IV\alpha-\beta^2))}{(\beta-\alpha)^2}. \tag{34}$$

Substituting Equations (32) and (34) back into Equation (31), we have

$$\lambda^* = \frac{-2(\beta-\alpha)}{b^2+(\beta-\alpha-2IV)b+(IV\beta+IV\alpha-\beta^2)}. \tag{35}$$

We use this Equation (35) to calculate the values of $\lambda_{i,j}$ for correlation testing in the Consistency of Risk Preference Measures analysis.