"A Replication of Willingness-to-pay Estimates in 'An Adding Up Test on Contingent Valuations of River and Lake Quality' (Land Economics, 2015)" (*Manuscript Number 2092; Discussion Paper Number 2017-55*)

Editor:

Your replication produces tests for four approaches: The nonparametric *Turnbull* procedure and three parametric procedures (*Linear Logit: Mean WTP*, *Linear Logit: Mean WTP* > 0, and *Log Linear Logit: Median WTP*). In two of the four cases (*Turnbull*, *Linear Logit: Mean WTP* > 0) you reject the adding up test, while failing to reject in the other two cases (*Linear Logit: Mean WTP*, *Log Linear Logit: Median WTP*). But even in the two cases where you fail to reject, the numerical gap between the two WTP values are very large.

For *Linear Logit: Mean WTP*, the Whole and Sum of the Parts are 250 and 418, respectively. For the *Log Linear Logit: Median WTP*, they are 201 and 359. Those are fairly large differences, even if they are not statistically significant. And I don't have to remind you that failure to reject the null is not the same thing as accept.

So there are several additional things I would like you to do. First, I would like you to evaluate the four models and discuss the pros and cons of the respective approaches. In particular, is there any reason to prefer one of the models more than the others?

Author's Response:

The Turnbull estimator was developed with the intention to provide a lower bound WTP estimate. This is useful when conducting sensitivity analysis in benefit-cost analysis (i.e., worst case scenarios) or in natural resource damage assessments where conservative estimates may be desired as a starting point in litigation.

A standard error can be produced for these estimates but any hypotheses tested with these standard errors should be done cautiously. First, the Turnbull truncates the WTP distribution at the highest bid.

Second, where the votes are not monotonically decreasing in the bids, pooling of the votes is performed in order to estimate the Turnbull. This pooling (in effect, recoding the dependent variable) will artificially reduce the standard errors and bias the test in favor of finding differences. For example, with the DMT (2015) data, the standard error of the WTP estimates are biased downward because of the recoding in the Turnbull. This may contribute to finding differences between WTP of the whole and the sum of the parts. In other words, standard errors that are biased downward due to pooling lead to a bias towards rejection of the adding up test, as found by DMT.

Hanemann (1984) was the first to describe the differences in the parametric WTP estimates that are presented here. The linear logit model produces median and mean WTP estimates that are equal. The weakness of the linear logit is that where the estimated Pr(Yes | WTP >

cost) survival function crosses the vertical axis and rises towards probability = 1 in the second quadrant, the area above the curve is included as a negative contribution to the mean WTP estimate, biasing these estimates downwards. In the extreme, where the estimated probability of the vote is less than 50% at a zero bid the WTP estimate will be negative (as in DMT's second scenario).

Hanemann (1989) described the truncated (at zero) WTP estimate (linear logit, mean WTP > 0) as one approach to avoiding the negative WTP problem. The truncated WTP estimate loses accuracy when the bids do not cover the range of WTP and the bid curve tail is flat, as in the DMT data. Hanemann (and Haab and McConnell) caution that researchers who use this estimate should make clear that it arbitrarily truncates WTP at zero. The linear logit model in this case is mis-specified.

A second approach to negative WTP is imposing the log-linear functional form to the data. A drawback of this approach is that the log-linear form cannot be worked back to an explicit utility function (Hanemann 1984) and the mean WTP is often undefined in a logit model (Haab and McConnell 2002), as in the DMT and Chapman et al. data. Yet, the log-linear functional form often fits the data better with higher Model χ^2 and McFadden's R^2 statistics (as with the DMT data) and testing differences between median WTP when the WTP distribution is right skewed with a long tail increases the power of the test (see Mitchell and Carson, 1989, Appendix C).

Turning to the concern that there are large differences in the <u>point estimates</u> between WTP for the whole and the WTP for the sum of the parts, the major reason for that is the striking lack of divergent validity between the WTP for the whole scenario and the WTP for each of the four incremental parts. The scope test literature examines the consumer axiom that more is better. Formal scope hypotheses with these data are:

HO: $WTP_{whole} \geq WTP_1$
HO: $WTP_{whole} \geq WTP_2$
HO: $WTP_{whole} \geq WTP_3$
HO: $WTP_{whole} \geq WTP_4$

There are three sets of WTP estimates in Table 5. With each of these three sets of estimates there are 4 scope tests for a total of 12 WTP comparisons. The willingness to pay for the whole is greater than the WTP for each of the parts. But, the measurement of the WTP for the whole is so imprecise that the point estimate of the WTP for 10 of the 12 increments is included in the confidence interval of the WTP whole scenario. In the other two the 95% confidence intervals overlap. These data do not pass the scope test, in contrast to the Chapman et al. (2009) data estimated parametrically with the log-linear model:

		95% Confidence Interval	
	Median WTP	LB	UB
Whole	167	127	206
Second	46	27	66

Similar results are found for the mean with the linear model. The regression results are available upon request.

In general, comparisons of WTP across scenarios is inefficient since WTP is formed by a ratio of coefficients. A more efficient test, but one that is less recognized in the literature, is to test for differences in the votes themselves. In other words, if a dummy variable for a scope scenario is statistically (and economically) significant in a regression analysis of votes then this is evidence that the CVM data pass the scope test. A similar test on the DMT data is presented in Appendix B, Figure 4 of the paper where the best model constrains the cost coefficients to be equal across each of the scenarios and the constants are constrained to be equal for the whole, first and fourth scenarios. Implicitly, the DMT data pass the scope test for only the second and third scenarios in this more efficient test.

Whatever the reason for the lack of validity in the DMT data, it is not surprising that the adding-up test presented in DMT (2015) is not robust to parametric estimation given the variability of the data.

<u>Changes to the paper:</u>

I have added the paragraph to the end of section 4:

> "There are benefits and costs of each of the alternative WTP estimators presented here. Haab and McConnell (2002, page 106) suggest that when "there are concerns about the distribution of response data," as here, researchers should estimate the Turnbull mean and the log-linear median to present conservative willingness-to-pay estimates. Haab and McConnell do not expand on the appropriate measure of central tendency when conducting hypothesis tests although it is clear they prefer the log-linear. Their concern is that the linear functional form is mis-specified when negative willingness-to-pay is allowed and truncation at zero is arbitrary. Similarly, the Turnbull mean is problematic due to arbitrary truncation at the highest bid and the pooling of data. The most reliable adding-up test is conducted with the medians estimated from the log-linear data."

Editor:

In this context, you need to address the points made by the original authors. Namely, that the estimated coefficients for the *Linear Logit: Mean WTP* model imply "that 35% of the population has a negative WTP for the whole program, and over half of the population (55%) has a negative WTP for the program's 2nd increment". Likewise, using the *Log Linear Logit* model, "the estimated parameters imply that 24% of the population has a WTP of more than $10,000 for the program, and that 7.3% has a WTP of more than $1 million". If the original authors are correct in these statements, these should be explicitly acknowledged in your revision and discussed in the context of whether we should believe the WTP numbers from these parametric models.

Author's Response:

I do not dispute the calculations made by DMT based on the parametric models and agree that these extreme values are troubling. But, I do have a different interpretation of the source of the problem. DMT present these as a refutation of the WTP estimates from the parametric models. I suggest that these extreme values are a refutation of the parametric model estimates of WTP from the authors' data. The parametric WTP estimates are standard in the CVM literature and extreme value estimates, such as these, based on extrapolations from parametric models are not the norm.

These extreme value estimates are an artifact of the DMT (2015) data which illustrates its weakness and unsuitability for conducting an adding-up test. For example, I've estimated the log-linear models in R which allows construction of Krinsky and Robb confidence intervals (Aizaki, Nakatani, and Sato, Stated Preference Methods Using R, CRC Press, 2015) using both the DMT and Chapman et al. data. The 95% confidence interval around the median ($201) for the whole scenario is [$59, $12,791] with the DMT data. The WTP of $12,791 as the upper bound is troubling. In contrast the 95% confidence around the median of $167 using the Chapman et al. data (n=1093) is a more reasonable [$135, $214]. Again, it is not the parametric models that are the problem, it is the DMT data.

For a more detailed example of data that does not have these troublesome properties, see the Appendix to this response where I estimate similar models as in the paper with the Chapman et al. data simulated to have the same sample sizes as in DMT.

Changes made to the paper:

I have come to the realization that my analysis and interpretation of DMT's data may be perceived by some as supporting the CVM's ability to pass the adding up test. This is not the case. To make this clear I have added the following to the conclusions.

> "To be clear, I am not asserting that I have shown that the CVM will pass the adding-up test if data are properly analyzed. The only claim that I can confidently make is that the DMT (2015) data is not strong enough to provide credible evidence that the CVM does not pass the adding up test. An adequate adding-up test would require more resources devoted to the study than is apparent in DMT (2015). A survey instrument would need to be developed with extensive focus groups and pretesting to construct believable scenarios with income and substitution effects. Even if researchers devote the necessary resources to survey design a credible adding-up scenario would still impose an amount of cognitive burden on survey respondents that might make the conduct of adding-up tests difficult. Indeed, laboratory experiment studies have found it difficult to impose the adding-up condition for market goods (Bateman et al. 1997, Elbakidze and Nayga 2017). Considering this, self-guided online opt-in surveys are likely not conducive to satisfying the adding-up test. Researchers should consider devoting resources to the in-person survey mode for the conduct of adding-up tests."

Editor:

I would also like you to address the original authors' following statement: "…inadequate response to the cost prompts remains a fundamental issue that has not been investigated as extensively as inadequate response to scope, but is perhaps even more important". That should also be discussed in the context of how seriously we should take the results of these respective adding up tests.

Author's Response:

Inadequate, meaning unsatisfactory or unacceptable, is a subjective notion so I would really need to know how the authors define inadequate in this context before I could begin to address their assertion.

To me, "adequate" would be defined as a negative and statistically significant coefficient on the cost amount with effect sizes sufficient for welfare analysis. With this definition it is an empirical regularity that dichotomous choice CVM studies tend to produce adequate response to the cost prompts. With my definition of adequacy there are very few studies in the CVM literature that exhibit "inadequate response to the cost prompt." The fact that DMT (2015) is one of these outliers is the motivation for my replication efforts.

There are several ways to provide evidence to support my assertion that the CVM literature demonstrates "adequate responsiveness to the cost prompt." One approach would be to provide a reference list of CVM papers that have an "adequate response to the cost prompt". This would be a long list containing most every dichotomous choice / referendum CVM paper ever published. I would be happy to begin to construct such a list if it is necessary.

Another way to illustrate "adequate responsiveness to the cost prompt" is to appeal to the revealed preferences of CVM researchers. There have been a number of meta-analytic studies of the CVM literature considering the scope, income effects, the divergence between WTP and WTA and the divergence between actual and hypothetical WTP (i.e., hypothetical bias). To my knowledge, there is no meta-analysis of the "inadequate response to the cost prompts." This negative space is some evidence that this is not a known problem or concern in the literature.

Still another approach to illustrate why I don't consider "inadequate response to the cost prompt" to be a problem with the greater CVM literature is to provide a counter example. In the appendix to this response I show that the Chapman et al. data, simulated with the same sample sizes as DMT (2015), produces results that exhibit a more "adequate response to the cost prompt."

Another definition of "adequate response to the cost prompt" is data that do not suffer from non-monotonicity and fat tails (my first definition primarily addressed flat bid curves). Non-monotonicity is typically a problem that can be solved with larger sample size or better experimental design. For example, the Chapman et al. data have non-monotonically decreasing percentage yes responses in only one of twelve cost amounts. Fat tails, where the probability of a yes vote does not fall to zero as the cost amount increases, may be a larger issue because high cost amounts lack credibility. My view is that the fat tails problem can be addressed with double-bounded referendum data (trading off efficiency for bias) or sensitivity analysis.

Haab and McConnell (2002) spend a considerable amount of effort on ways to deal with these "difficult data." These approaches are suitable for benefit-cost analysis where sensitivity analysis should be conducted. These approaches are less suitable in the context hypothesis testing, such as with adding-up.

Changes to the paper:

In addition to all of the other robustness tests that I have performed with the data (see Appendix B of the paper), it is clear (to me, at least) that we should not take very seriously DMT's data and their adding up tests. That is the point that I'm trying to make throughout the paper. I have thoroughly rewritten the paper with this in mind.

I would not discourage researchers from conducting a meta-analysis over adequate response to the cost prompt but I believe it is beyond the scope of the current paper. Since this is a new issue brought up by the authors in their reply and not their original paper, I have not made any changes to the replication paper with specific reference to "inadequate responsiveness to the cost prompt".

Editor:

Further, I think you can do better with respect to testing the the adding up hypothesis. Currently, you simply test for zero difference between the Whole and the Sum of the Parts WTP estimates. The problem with this is that failure to reject does not mean that the differences are the same. To get at that, you should do an equivalence test. This is not so common in economics, but does get used in psychology. A good reference is: Daniël Lakens, "Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses", *Social Psychological and Personality Science*, Vol 8, Issue 4 (2017): 355 – 362.

Author's Response:

A formal adding up test is a two-tailed test:

HO: $WTP_{whole} = WTP_{sum\ of\ the\ parts}$
HA: $WTP_{whole} \neq WTP_{sum\ of\ the\ parts}$

If I understand correctly, an equivalence test is being suggested due to the large differences in point estimates between WTP for the whole and WTP for the sum of the parts. In effect, the alternative hypothesis is updated by the results of the experiment:

HO: $WTP_{whole} = WTP_{sum\ of\ the\ parts}$
HA: $WTP_{whole} < WTP_{sum\ of\ the\ parts}$

An equivalence test would specify an effect size, D, against which to test the observed effect size, $\Delta WTP = WTP_{sum\ of\ the\ parts} - WTP_{whole}$.

HO: $D < \Delta WTP$
HA: $D \geq \Delta WTP$

An equivalence test in this context is not appropriate because the survey design does not permit the conduct of an adding-up test for the two reasons I described in an unpublished comment on DMT: income and substitution effects. I added this to the paper in response to a referee comment and now have added a more formal version as section 2. In fact, economic theory suggests the appropriate statistical test considering the DMT survey design is a one-tailed test:

H0: $WTP_{whole} < WTP_{sum\ of\ the\ parts}$
HA: $WTP_{whole} \geq WTP_{sum\ of\ the\ parts}$

DMT (2015) implicitly acknowledges that theory suggests this test with their income effects simulation. Their claim is that income effects are typically so small in CVM studies that an appropriate survey design is not important. The authors do not address the potential for substitution effects in DMT (2015).

Finally, note that I am only surmising what is actually in the DMT surveys based on how they are described in DMT (2015). The authors claim that they used the Chapman et al. surveys with only minor modifications for the three new scenarios. In my experience and opinion, imposing an in-person survey on online opt-in panel respondents would not yield an agreeable outcome. Opt-in panel respondents are not well compensated and many are in a hurry to complete their survey task. In-person interviews tend to be longer and more complicated than an online survey. In-person interviews are led by a trained facilitator. Online surveys are self-guided and paced. This may be one reason for the low data quality in DMT (2015).

Note that I am only speculating. I have not been able to review the actual DMT surveys for their face validity. Face validity is achieved when a CVM survey instrument actually elicits the theoretically correct willingness to pay estimate. This sort of review is one of the guidelines for stated preference studies recommended by Johnston et al. (J. Assoc. Env. Res. Economists 2017) and it is common practice to share your surveys as part of the review process.

Changes to the paper:

I have not conducted an equivalence test for the reason described above. The survey design does not permit the conduct of an adding-up test except under unrealistic assumptions (the lack of substitution and income effects). Even given an appropriately designed adding up test, it is not clear to me what the appropriate effect size should be in the equivalence test.

Let me emphasize again that I do not believe that the DMT survey design is sufficient for eliciting data to conduct an adding up test. My focus on the statistical analysis in the original submission was my attempt at maintaining the spirit of a replication analysis instead of a broader comment on DMT (2015). In order to make the theoretical argument more clear, I have moved the appendix to section 2 of the paper.

I have added the following to the conclusions:

> "To be clear, I am not asserting that I have shown that the CVM will pass the adding-up test if data are properly analyzed. The only claim that I can confidently make is that the DMT (2015) data is not strong enough to provide credible evidence that the

CVM does not pass the adding up test. An adequate adding-up test would require more resources devoted to the study than is apparent in DMT (2015). A survey instrument would need to be developed with extensive focus groups and pretesting to construct believable scenarios with income and substitution effects. Even if researchers devote the necessary resources to survey design a credible adding-up scenario would still impose an amount of cognitive burden on survey respondents that might make the conduct of adding-up tests difficult. Indeed, laboratory experiment studies have found it difficult to impose the adding-up condition for market goods (Bateman et al. 1997, Elbakidze and Nayga 2017). Considering this, self-guided online opt-in surveys are likely not conducive to satisfying the adding-up test. Researchers should consider devoting resources to the in-person survey mode for the conduct of adding-up tests."

Editor:

Finally, related to the above, it would be good if you brought in some discussion of statistical power in the context of your analysis.

Author's Response:

The concern is that the lack of a statistically significant difference in the large effect sizes produced by parametric models could result in a Type II error: accepting the null of adding up when the alternative is true (as with the linear logit and log-linear logit). The power of the test is the probability of not making a type II error.

I believe that a formal discussion of statistical power is beyond the scope of my paper. I have suggested that the authors should have employed fewer scenarios in order to conduct their adding up test. Three scenarios, as suggested in the DMT (2012) paper, would have allowed for larger sample sizes and tighter confidence intervals. However, some simple sample size calculations suggest that very large sample sizes would be needed to find any significant differences in WTP for the whole and the sum of the parts given the large standard errors produced by the flat bid curves estimated from the DMT (2015) data.

For example, consider the summary statistics from the linear model where the standard deviation is estimated as $SE \times n$.

	Mean	SE	n	SD
Whole	250	81	172	1062
Sum of the Parts	418	127	708	3379

Holding the standard deviation constant the sample sizes would need to be 6 times larger for 95% confidence intervals to avoid crossing. With a cost per unit of $3.50 from the Survey Sampling International online panel used by DMT, the cost of the additional 4400 observations is $15,400.

Alternatively, researchers could obtain higher quality data from in-person interviews or an address-based online panel (e.g., GfK) at considerable additional cost. As demonstrated in the appendix to this response, the standard errors at the same sample size could be as much as 50% lower due to a more "adequate response to the cost prompt". In this case, sample sizes 1.5 times larger would be sufficient to produce statistically significant differences in WTP for the whole and sum of the parts with standard errors about the same magnitude as in the Chapman et al. simulated data. If the unit cost of a higher quality sample is $50, then these samples would cost $66,000 (n=1320).

<u>Changes to paper:</u>

I don't think that a "increase the power of the test" solution is feasible unless the researchers are willing to spend more money than they did in their original survey.

The problem of statistical power and sample size has been a well-known issue in CVM. Appendix C of Mitchell and Carson (1989) describes statistical power at length. I have provided a reference to this in the conclusions.

> "Future studies should attempt to address these data quality problems with experimental designs that lead to statistical tests with sample sizes large enough to provide sufficient power. This can be achieved in three ways. The most costly, of course, is simply increasing the overall sample size. Holding cost constant, researchers could reduce the number of cost amounts used in the experimental design or reduce the number of experimental scenarios presented. For example, DMT could have implemented their adding up test with three (instead of five) separate scenarios as they describe in Desvousges, Mathews and Train (2012). The interested reader is reminded that Appendix C of Mitchell and Carson (1989), the seminal CVM text, has an extensive discussion of sample size and statistical power."

Appendix to "Response to the Editor"

Throughout this response are criticisms of the DMT (2015) data. This might be interpreted as an empty assertion without contrast with higher quality data. To provide the contrast between low and high quality CVM data consider Chapman et al. (2009) who field "base" and "scope" surveys. DMT's "whole" and "second increment" scenario surveys are identical to the base and scope surveys. I construct data from Chapman et al. (2009) with similar sample sizes of DMT. I apply the bid curve probabilities from Table 6.27 of Chapman et al. (2009) to the sample sizes of DMT.

Chapman et al.'s sample sizes are much larger, $n = 1093$ and $n = 544$ than those used by DMT for the same scenarios ($n = 166$ and $n = 152$). In order to create the yes/no vote data we multiply the product of the sub-sample sizes from the DMT study to the percentage yes responses from Chapman et al. (2009) and round to the nearest integer (Table R1). The simulated percentage yes responses differ from those in the Chapman et al. (2009) data due to the small sample sizes and rounding.

The simulated bid curve for the whole (i.e., "base" in Chapman et al.) scenario is non-monotonically decreasing in one cost amount. The simulated bid curve for the second (i.e., scope) scenario is monotonically decreasing. Both of the simulated bid curves are steep relative to the DMT bid curves for the same scenarios. The overall percentage yes responses is similar in the whole scenario but 43% higher in the simulated second scenario compared to the DMT data.

The simulated Chapman et al. data produces a Turnbull WTP estimate, $188, similar to that produced by the DMT data, $200, for the whole scenario. The simulated Chapman et al. data produces a Turnbull estimate different from that produced by the DMT data for the second scenario. The Turnbull mean WTP estimate is $139 with the simulated data compared to $84 with the DMT data.

In Table R2 we present regression models similar to those in Table 5 in the paper. Each of the regression coefficients is statistically different from zero except for the second constant in the linear logit model. The log-linear functional form provides a better statistical fit.

The WTP estimates calculated from the regression models in Table R2 are presented in Table R3. The mean WTP estimates for the whole scenario range from $228 to $291. The median WTP estimate from the log linear model is $170. The mean WTP estimates for the second scenario range from $64 to $228 with a median of $48 in the log-linear model. As in Table 5, the lower end of the range of mean WTP is from the linear logit model that allows negative WTP and the upper end of the range is from same model that constrains WTP to be positive.

Comparing these results to those in Table 6, the range of estimates is more narrow with the simulated Chapman et al. (2009) data. The ratios of the WTP range to WTP midpoint from the simulated data is 24% and 112% for whole and second scenario. The ratios of the WTP range to WTP midpoint are 58% and 665% for the whole and second scenario from the DMT data.

None of the WTP comparisons pass the scope test. The base and scope WTP estimates have overlapping confidence intervals. This is a result of the small sample size and relative inefficiency of tests for differences in WTP with dichotomous choice data (where WTP is a ratio of regression coefficients). A test for differences in votes is a more efficient scope test. Constraining the linear model slopes in Table R3 to be equal is not rejected by a likelihood ratio test. An equality constraint on the equality of the constants is then rejected ($\chi^2 = 8.52[1\ df]$). This indicates that the yes votes are higher for the whole scenario relative to the second part. Similar results are found for the log-linear model. Similarly, a specification with a constant and a dummy variable for the second (scope) scenario produces a negative and statistically significant coefficient in the linear and log-linear models (t = -2.89). These results are available upon request.

In conclusion, the Chapman et al. (2009) responses applied to the DMT sample sizes outperforms the DMT data. In contrast to the DMT data, the simulated Chapman et al. data exhibits sensitivity to scope in more efficient tests.

Table R1. Simulated Dichotomous Choice CVM Data from Chapman et al. (2009) Bid Curves

	Base (Whole)			
	Simulated		%Yes	
Cost	Yes	N	Simulated	Chapman et al.
10	20	24	83	82
45	22	31	71	70
80	16	26	62	60
125	16	26	62	62
205	12	28	43	44
405	11	31	36	34
Total	97	166	58	58
	Scope (Second)			
	Simulated		%Yes	
Cost	Yes	N	Simulated	Chapman et al.
10	17	24	71	71
45	13	28	46	48
80	10	23	43	41
125	11	28	39	39
205	8	24	33	31
405	7	25	28	29
Total	66	152	43	43

Table R2. Dichotomous Choice Probability Models with Chapman et al. (2009) Simulated Data

	Linear Logit			Log Linear Logit		
Intercept	Coefficient	S.E.	t-stat	Coefficient	S.E.	t-stat
Base (Whole)	1.070	0.254	4.21	3.23	0.78	4.14
Scope (Second)	0.226	0.247	0.92	1.86	0.69	2.70
Slope	Coefficient	S.E.	t-stat	Coefficient	S.E.	t-stat
Base (Whole)	-0.005	0.001	-3.76	-0.63	0.16	-3.87
Scope (Second)	-0.004	0.001	-2.56	-0.48	0.15	-3.17
χ^2	29.85			35.79		
McFadden R^2	0.07			0.08		
Sample size	318			318		

Table R3. Willingness to Pay Estimates with Chapman et al. (2009) Simulated Data

	Linear Logit						Log Linear Logit		
	Mean WTP			Mean WTP > 0			Median WTP > 0		
	WTP	S.E.	t-stat	WTP	S.E.	t-stat	WTP	S.E.	t-stat
Base (Whole)	228	40	5.66	291	54	5.43	170	49	3.48
Scope (Second)	64	54	1.18	228	66	3.44	48	19	2.55