Market Access and Quality Upgrading: Evidence from Four Field Experiments

A Replication Study of Bold et al. (*American Economic Review, 2022)*

Ryan McWay*　　Karim Nchare†　　Pu Sun‡

Journal of Comments and Replications in Economics, Volume 3, 2024-2, DOI: 10.18718/81781.32

JEL: L14, L15, O13, Q12, Z00
Keywords: Reproducibility, Replication, Farm Productivity, Economic Development

Data Availability: The data and Stata code to reproduce the results of this replication can be downloaded at JCRE's data archive (DOI: 10.15456/jl.2023270.1716216104).

Please Cite As: McWay, R., Nchare, K., Sun, P. (2024). Market Access and Quality Upgrading: Evidence from Four Field Experiments: A Replication Study of Bold et al. (2022b). *Journal of Comments and Replications in Economics*, Vol 3(2024-2). DOI: 10.18718/81781.32

Abstract

Bold et al. (2022b) investigate the effect of providing access to a larger, centralized market where quality is rewarded with a premium on farm productivity and framing incomes from smallholder maize farmers in western Uganda, using a series of randomized experiments and a difference-in-differences approach. We successfully reproduce the results of this study using the publicly provided replication packet. Then test the robustness of these results by re-defining treatment and outcome variables, testing for model misspecification and the leverage of outliers, and testing for non-random selection in the Fisher-permutation process. Our results show that the findings in Bold et al. (2022b) are robust to a variety of decisions in the research process. This evokes confidence in the internal validity of the findings.

*Corresponding author, University of Minnesota, mcway005@umn.edu
†Vanderbilt University, karim.nchare@vanderbilt.edu
‡University of Ottawa, psun032@uottawa.ca.

Acknowledgements: This replication is the result of the Institute for Replication's (I4R) Toronto Games, 2023. We would like to thank Abel Brodeur for the invitation to these games and the inception of this work. We acknowledge Lucienne Talba for her contribution at the Toronto Games. McWay is the corresponding author. The authors declare they have no conflict of interest in assessing Bold et al. (2022b) nor any personal relationships with the authors. Finally, we thank the authors of Bold et al. (2022b) for publicly providing the replication data and code for their study.

Received September 27, 2023; Revised February 06, 2024; Accepted February 07, 2024; Published March 12, 2024.

1 Introduction

Bold et al. (2022b) investigate the effect of providing access to a market where quality is rewarded with a premium on farm productivity and farming income amongst smallholder maize farmers in western Uganda using a series of randomized experiments and a difference-in-differences approach. The authors collect data on these farmers through a series of baseline and endline surveys of smallholder farmers in 20 community clusters centered on spot markets.[1] 100 farmers were enrolled in the first experiment to determine maize quality prior to the subsequent three experiments. Another 100 farmers were enrolled in the second 'quality upgrading' experiment. This study crosses randomized the third 'market access experiment' and the fourth 'extension service experiment'. The third experiment encompassed 20 clusters of 180 households for 1,198 household-season observations. The fourth experiment encompassed 18 clusters of 164 households for 931 household-season observations. The study ranges over five farming seasons from the years 2017 to 2020. Since smallholder farming in developing countries is characterized by low productivity and low-quality output, the aim of the study is to understand the benefits of quality upgrading and providing market access where market premiums might increase productivity and household incomes.

To study both quality upgrading and providing market access to a market with a market premium, the authors device four separate experiments through randomized controlled trials (RCT). The first experiment confirms the poor quality of maize production by measuring the quality of maize sold at the farm gate. This confirms both that the quality of maize is poor and that maize quality is partially observable by buyers. The second experiment randomly assigns farmers into treatment and control groups in which the treatment group is offered a service package that improves the quality of their maize. This experiment finds that both high and low quality maize sells for the same price in local markets suggesting that quality upgrading does not improve the seller's price. The authors conjecture that the lack of improved prices is due to a missing market for local buyers of high quality maize.

Therefore, they device a third experiment in which they randomize farmers into treatment and control groups again with the treatment group offered access to larger markets where presumably higher quality buyers of maize will purchase at a market premium (and thus be a benefit to the smallholder farmers). Given this opportunity to sell higher quality maize at higher prices, the authors observe a marked increase in both farm productivity, higher seller prices, and increased profits for the farmers. To distinguish again between access to market premiums and quality upgrading in the new scenario where some rural farmers have access to larger, centralized markets, the authors devise a fourth experiment where some farmers are placed in a treatment group that receives up-to-date knowledge pre- and post-harvest about high quality farming practices. This extension is a form of learning-by-doing experiment investigating if farmers are improving the quality of their maize with market premiums, or simply exploiting higher prices through the arbitrage of a larger, centralized market. Again, the authors report that they find no evidence that the farmers change their farming practices as a result of this supply-side intervention. Rather, all changes to production, seller prices, and farmer profits are driven by access to market premiums. Given these findings, the authors claim that for agricultural transformation (e.g., an important component of rural economic development) market access is a necessary condition, while quality upgrading is not a sufficient

[1] Bold et al. (2022b) describe a spot market as "The farmer and the buyer agree right before the sale, usually after a short visual inspection of the maize bags by the buyer, about the amount and the price. The farmer is paid directly and the transaction takes place at the farm gate."

condition.

Further, the authors examine the impact of an influx of new suppliers on these larger, centralized markets. Using a difference-in-differences approach to adjust for selection into treatment (i.e., participation in the marketplace), they find that the higher prices by the buyers of maize drove up the seller's price for the new, smallholder farmers of maize by 30%. The authors claim that the increase in maize production by the treatment group of the third experiment (i.e., access to market premiums) is entirely driven by the exposure to the higher buyer prices in this market.

Using these findings, the authors claim that while quality upgrading has no causal impact on improved quality of production, providing access to markets with market premiums for quality does remove a demand-side constraint that was limiting the incomes of rural smallholder farmers and productivity growth of maize farmers.

In the present paper, we directly reproduce the results using the replication packet publicly provided by the authors through the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. We successfully reproduce the results presented in the manuscript identifying no significant variation in the reported magnitudes nor statistical significance in either the published manuscript as well as the online appendix for this article.

We complement this reproduction by testing the robustness and replicability of the results. We do this through three categories of robustness. The first is by redefining both treatment status and outcome measures in this context. The second is by measuring the leverage of outliers on the reported results and testing for model misspecification for the reported spillover effects. Finally, we test the robustness of the Fisher-permutation process for bootstrapping confidence intervals in the study due to small sample sizes. In all three sections, we report no significant deviations from the reported results when testing these robustness measures. This lends confidence in the internal validity of the reported findings.

2 Reproducibility

We conduct a computational reproducibility[2] of Bold et al. (2022b). The study being reproduced is a 2022 publication in the American Economic Review[3] aimed at providing insights into how market access influences agricultural production and household incomes within a developing context. This study was pre-registered through the American Economic Association's Randomized Controlled Trail (RCT) Registry as trail $AEARCTR - 0002812$[4] Svensson et al. (2018). The data and code used in this reproduction come from the publicly published replication packet by the authors of the original study through the Inter-university Consortium for Political and Social Research (ICPSR)

[2]Computational reproducibility is defined by Brodeur (2023) as "The ability to duplicate the results of a prior study using the same data and procedures as were used by the original investigator. Reproducibility is done using the same computer code, but can be achieved using a different software package."

[3]The original paper can be found at: https://www.aeaweb.org/articles?id=10.1257/aer.20210122.

[4]The pre-registered report can be found at: https://www.socialscienceregistry.org/trials/2812/history.

at the University of Michigan as project 158401.[5] Bold et al. (2022a) The replication packet was downloaded on February 3rd, 2023. The reproduction, along with the replication, was conducted in Stata 16.

The catalyst for this reproduction and replication is the result of the Institute for Replication's Toronto Games[6] hosted by the University of Toronto held on February 22nd, 2023. A team of four members was established for this one-day event to reproduce and replicate the results for Bold et al. (2022b). Prior to the Toronto Games, all members of the team successfully replicated all tables and figures produced in both the published manuscript as well as in the appendix.[7] To ensure our confidence in this reproduction, on the day of the event we double-checked for any discrepancies by dividing the exhibits amongst the members[8] and closely examined each section of results.[9] Again, all members of the team confirmed the exact reported magnitudes and statistical significance of results as those produced in the original publication. No coding errors were identified, and the reproduction of results matches those reported in the published manuscript (as well as in the published appendix).

3 Three Replication and Robustness Checks

Noting the successful reproduction of the study's findings, we examine the robustness replicability[10] of the study. Generally, we test the sensitivity of the results in three domains. Firstly, we redefined the terminology for what is considered the treatment status in quality upgrading as well as the outcome measures of interest. Second, we estimate the influence that outliers in the data have on the statistical significance of the results, as well as a formal test for any model misspecification for the spillover effects. And finally, due to the avid use of Fisher-permutations to bootstrap confidence intervals for several of the results, we test the sensitivity of these permutations to the initial seed generated. The decision to conduct these three robustness checks was taken after reading the paper but prior to observing the replication packet. These sensitivity analyses were not pre-registered, and are vulnerable to ad-hoc decision-making in the replication process.

3.1 Regression Models

For this analysis, we rely on the same model specifications defined in the article by Bold et al. (2022b). The causal estimation follows the same underlying principles and assumptions as laid out for the randomized controlled trial and difference-in-differences design used in the original paper. The econometric methods are the same as those used in the original paper for our replications. The deviation for our replication relies either on the definition of the variables in the dataset or through

[5]The replication packet can be found at: https://www.openicpsr.org/openicpsr/project/158401/version/V1/view?path=/openicpsr/158401/fcr:versions/V1&type=project.

[6]Details on the Toronto Games can be found at: https://i4replication.org/games.html.

[7]The online appendix can be found at: https://www.aeaweb.org/content/file?id=17133.

[8]The seven tables, five figures, fourteen appendix tables, and three appendix figures were divided into four portions of seven exhibits each.

[9]This design is inspired by the double data entry approach used for manual data entry.

[10]Robustness replicability is defined by Brodeur (2023) as "The ability to duplicate the results of a prior study using the same data but different procedures as were used by the original investigator. Robustness replicability can be done using the raw, intermediate or final data sets used by the original authors."

Table 1. Table 2 Column(1): Redefining Treatment Status

	(1) Original	(2) Below median	(3) Above median	(4) 25% Quantile	(5) 50% Quantile	(6) 75% Quantile
Treatment	-2.16	-1.41	-0.00	-1.03	-2.00	-3.00
	(0.212)	(0.114)	(0.000)	(0.379)	(0.153)	(0.011)
	[0.000]	[0.000]	[1.000]	[0.007]	[0.000]	[0.000]
Observations	622	463	159	622	622	622
Households	99	83	23	99	99	99

Table 1. *Notes:* This is a replication of Table 2 Column (1) from Bold et al. (2022b). To test the heterogeneity of the treatment effect, the outcome of interest below and above the median level has been regressed separately in columns (2) and (3). Besides, estimates at the 25th, 50th, and 75th quantiles are also reported. Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

the application of robustness checks as described in the following sections. The analysis is pooled at the household level consistent with the original study. For more details on the model specifications and causal identification strategies, refer to the original study Bold et al. (2022b).

3.2 Re-defining 'Quality' Maize, Loss of Production, & Maize Prices

Table 2 Bold et al. (2022b) presents the treatment effect of the quality-upgrading intervention on returns. In column (1), the average treatment effect on visually verifiable defects is -2.16. However, we believe that such an effect can be heterogeneous amongst farmers. So, we shrank the sample to those below the median level of the defects and then compared how the results differ about the median and the mean level of quality. Using the exact same econometric method, the result with the below median level sample shows an even smaller magnitude, while the above median level sample does not change significantly (See Table 1). In fact, the original paper also performs a robustness check as suggested by the outcome of interest in column (2) which is the grams of defects per 200g maize measured in the laboratory. The average treatment effect then is -0.20. Similarly, we employed a quantile regression method to test the heterogeneity of the treatment effect at the 25th, 50th, and 75th quantiles of the outcome of interest by repeating 500 bootstrapped permutations about each quantile. Table 2 reports these estimates which are significantly negative, though their magnitudes are smaller than the mean estimate. To supplement the robustness check, Figure 1 and Figure 2 plot the coeffects estimated at each 5th quantile with their 95% confidence interval. The larger treatment effect exists with the higher quantile of the outcome of interest while being significantly negative across the majority of the distribution. In summary, the existence of heterogeneity does not alter the claims of the original findings – that the treatment effect of the quality-upgrading intervention on the number of defects is robustly a negative effect (e.g., experiment two does improve quality despite a change in farmer's income or production).

Further, we examined column (4) of Table 2 in Bold et al. (2022b). The authors defined "deduction" as the percentage gap between the weights between the sold by enumerators and the agreed by buyers. In other words, "deduction" is a loss of production. We re-defined the outcome of interest in column (4) as the net sales in percentage: the percentage of the agreed by buyers to the sold by enumerators. Mathematically, this re-definition only alters the sign of the estimated effect in the regression (which remains insignificant). As anticipated, Table 3 compares the original estimate and the new estimate which are identical except for the sign. Following this re-definition, we re-

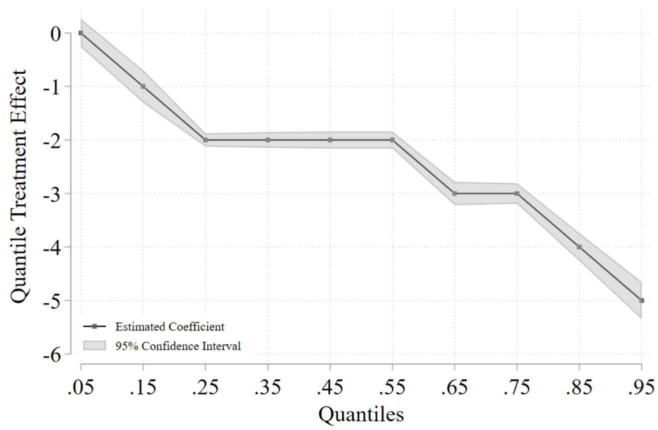

Figure 1. Table 1 Col 1 Redefine Quantiles

Figure 1. *Notes:* This is a replication of Table 2 Column (1) from Bold et al. (2022b) by estimating the same model at each 5th quantile. A bootstrap process with 500 times repeats is used for estimating sample standard errors. The coefficients with their 95% confidence interval are plotted.

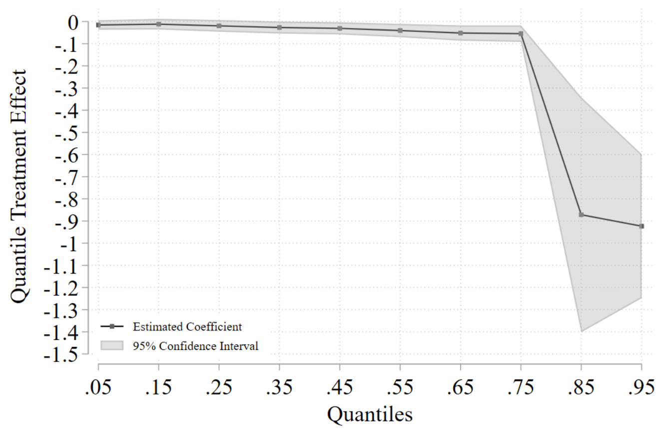

Figure 2. Table 1 Col 2 Redefine Quantiles

Figure 2. *Notes:* This is a replication of Table 2 Column (2) from Bold et al. (2022b) by estimating the same model at each 5th quantile. A bootstrap process with 500 times repeats is used for estimating sample standard errors. The coefficients with their 95% confidence interval are plotted.

placed the horizontal variable in Figure 3 Panel B of Bold et al. (2022b) with the net sales described above, certis paribus. This plots a symmetric curve to the original curve, as is presented in Figure 3. Again, the Kolmogorov-Smirnov D statistic (i.e., p-value) remains the same. This indicates that

Table 2. Table 2 Column(2): Redefining Treatment Status

	(1) Original	(2) 25% Quantile	(3) 50% Quantile	(4) 75% Quantile
Treatment	-0.20	-0.02	-0.03	-0.05
	(0.011)	(0.002)	(0.003)	(0.007)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	1640	1640	1640	1640
Households	82	82	82	82

Table 2. *Notes:* This is a replication of Table 2 Column (2) from Bold et al. (2022b). To test the heterogeneity of the treatment effect, estimates at the 25th, 50th, and 75th quantiles are also reported. Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

Table 3. Table 2 Column(4): Redefining Treatment Status

	(1) Original	(2) Redefinition
Treatment	0.18	-0.18
	(0.633)	(0.633)
	[0.780]	[0.780]
Observations	116	116
Households	94	94
R-squared	0.22	0.22

Table 3. *Notes:* This is a replication of Table 2 Column (4) from Bold et al. (2022b) with the modification of deductions to the net sales in percentage, while holding everything else the same. Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

the cumulative distribution function (CDF) curves in the treatment group and control group are not statistically different. These two checks indicate the intervention of the second experiment did not increase maize sales, as is established in the original manuscript.

Finally, we take this concept of re-definitions and apply it to the third and fourth experiments. Bold et al. (2022b) suggest the market access intervention increased farmers' profit by higher sales and price. By comparison, the extension service experiment only achieved trivial effects. Here we modified the price variable by multiplying it with the sold share of maize. The generated variable is the real achieved raw price of maize. Figure 5 Panel A plots the CDF curves of the raw price in the treatment group and control group after the two different interventions. Surprisingly, the two groups do show significantly different curves with the p-value of the Kolmogorov-Smirnov D statistic less than 0.05 after the market access intervention. However, this is no longer the case in the extension service intervention since both curves fit together showing little discrepancy (the p-value now is 0.219). Therefore, the main finding from the original paper also holds in this robustness check.

Together, these three re-definitions of outcome measures and treatment status have no impact on the reported claims in Bold et al. (2022b). This exudes confidence that the authors were careful to test the robustness of their variable definitions, and defends against any suggestion that the

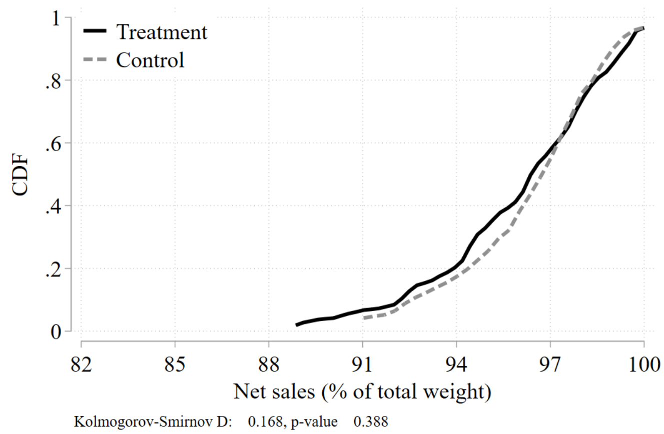

Figure 3. Figure 3 Redefine

Figure 3. *Notes:* This is a replication of Figure 3 from Bold et al. (2022b) by redefining net sales as its percentage of the total weight. The solid curve and dash curve represent the treatment group and control group. Besides, the Kolmogorov-Smirnov D statistic is reported to test their difference.

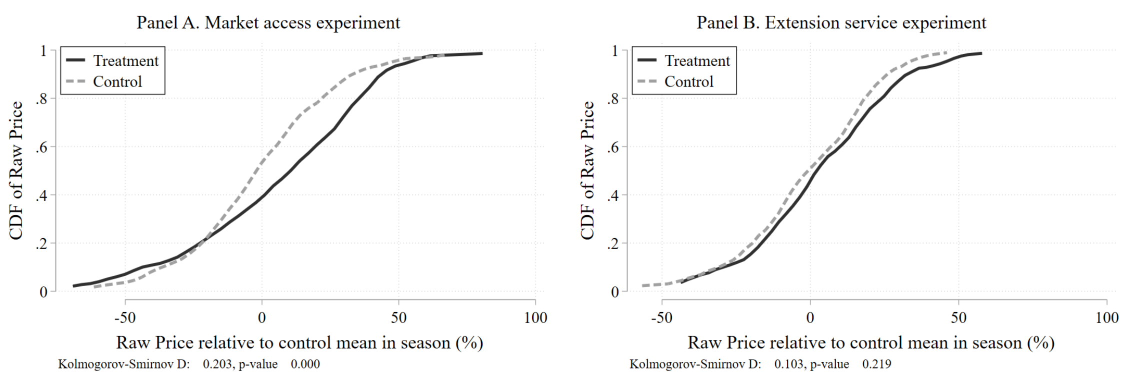

Figure 4. Figure 5 Redefined

Figure 4. *Notes:* This is a replication of Figure 5 from Bold et al. (2022b) with the modification of price by multiplying the sold share of maize. Panel A and B are the market access intervention and extension service intervention respectively. Besides, the Kolmogorov-Smirnov D statistic is reported to test their difference.

results are an artifact of a specific construction of the variables in the datasets.

Table 4. Table 7: Linktest for Model Misspecification

	(1) Price Change	(2) Price Change	(3) Price Diff.	(4) Price Diff.
$\hat{\beta}$	1	1	1	1
	(0.4553)	(0.3406)	(0.4239)	(0.4126)
	[0.028]	[0.003]	[0.019]	[0.016]
$\hat{\beta}^2$	0.0000	0.0000	0.0000	0.0000
	(1.7836)	(1.3680)	(2.0273)	(2.0073)
	[1]	[1]	[1]	[1]
Obs.	799	799	799	799
R^2	0.1744	0.1808	0.0881	0.0906
Trader	Local	Commerical	Local	Commerical

Table 4. *Notes:* This is a replication of Table 7 from Bold et al. (2022b) testing for model misspecification using the linkest proposed by Tukey(1949) and Pregibon(1980). Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

Table 5. Table 7: Testing Outliers via Li (1985)

	(1) Other Traders	(2) Local Traders	(3) Commercial Traders
Panel A. Difference in market shares and prices			
Difference in market shares	-0.379	-0.363	.
	(0.000)	(0.000)	.
Difference in prices vs. control	0.043	0.058	-0.012
	(0.001)	(0.000)	(0.596)
Panel B. Difference in prices adjusting for selection			
Difference in prices vs. control	0.061	0.071	0.025
	(0.001)	(0.000)	(0.417)

Table 5. *Notes:* This is a replication of Table 7 from Bold et al. (2022b) testing for outliers using the test developed by Li (1985). Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

3.3 Testing Model misspecification & the Leverage of Outliers

Table 7 of Bold et al. (2022b) describes results on the spillover effects of sales on the local market. That is, exposure to higher buyer prices in the centralized market might impact the local markets that smallholder farmers operate within. In these results, the authors claim that the entrance of high quality buyers of maize decreased the market share of other traders in local markets. This is due to the arbitrage shift in prices.

For this claim, we test two components. The first is to ensure that the model for spillovers is not misspecified. Misspecification error may falsely determine that there was a local effect when in fact there was none. Additionally, it may be of concern that outliers (i.e., very successful farmers in the treatment group) have a disproportionate leverage over the spillover effect. To test for model misspecification, we perform a linktest. Using an R^2 measure of model fit, the linktest Tukey (1949) and Pregibon(1980) uses the model prediction $\hat{\beta}$ and the squared prediction $\hat{\beta}^2$ to measure if they have explanatory power over the model. If there is model misspecification, then we would expect to have a statistically significant squared prediction $\hat{\beta}^2$. In Table 4, we report the results of the linktest

Table 6. Table 7: Testing Outliers via Koenker (2005)

	(1) Other Traders	(2) Local Traders	(3) Commercial Traders
Panel A. Difference in market shares and prices			
Difference in market shares	-0.379	-0.363	-0.071
	(0.000)	(0.000)	(0.280)
Difference in prices vs. control	0.040	0.064	-0.006
	(0.003)	(0.000)	0.799
Panel B. Difference in prices adjusting for selection			
Difference in prices vs. control	0.047	0.057	0.022
	(0.011)	(0.002)	(0.459)

Table 6. *Notes:* This is a replication of Table 7 from Bold et al. (2022b) testing for outliers using the test developed by Koenker (2005). Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

for the regressions in Table 7 of Bold et al. (2022b) for Panel A and Panel B of local traders and commercial traders. In all of these cases, the squared prediction $\hat{\beta}^2$ is not statistically significant. This implies that the spillover model is well specified to measure the impact of experiment 3 on prices and market share of local markets.

To test for the influence of outliers driving the effects the authors claim, we use two forms of least-absolute value models (LAV). The first performs an initial screening based on Cook's distance > 1 to eliminate gross outliers before calculating starting values and then performs Huber iterations followed by biweight iterations Li (1985) (See Table 5). And the second fits quantile regression models as a conditional distribution of the independent variables to measure the influence of each quantile's leverage on the mean effect (Chamberlain, 1994; Koenker, 2005) (See Table 6). In both tables, we find consistent results to those displayed in Table 7 of Bold et al. (2022b). In Table 5, the (average) price for sales to other traders in the treatment group is 4.3 percent higher than the average price in the control group ($pvalue = 0.001$). And in Table 6, the median price for sales to other traders in the treatment group is 4 percent higher ($pvalue = 0.003$). These two robustness checks for model misspecification and the influence of outliers provide confidence in the spillover results presented in the original manuscript.

3.4 Robustness of Permutation Process

As noted previously, the study for Bold et al. (2022b) involves only 1,198 household-season observations for 20 clusters of 180 households in the third experiment and 931 household-season observations for 18 clusters of 164 households in the fourth experiment. Considering the pivot importance of these two experiments at the heart of the claim by the authors that smallholder farmer productivity and incomes are driven by market access rather than quality upgrading, the original authors provide confidence in their results by using a Fisher-permutations test of 10,000 permutations of treatment assignment when calculating their p-values. This is helpful as it ensures that the results are not subject to the lower statistical power of the sample size.

We test this concern by considering if the authors' choice of initial seed in their code for the permutation tests created an artefactual confidence in their results. To test this, we create 100 randomly permutated versions of the initial seed for their Fisher-premutation tests. We then implement the Fisher-permutation tests for their reported results using each of the 100 randomly selected seeds for the results of the third and fourth experiments reported in Tables 4, 5, and 6 of Bold et al. (2022b). This 'bootstrapping' method is necessary to recover confidence intervals that accurately represent the underlying distribution of potential outcomes. In Table 7, we display the average β coefficient, standard error, and p-value for the impact of the maize upgrading intervention on maize quality. Compared to the original report of a 0.593 increase in graded maize with a $p-value = 0.001$, we also find an effect of 0.0593 with a $p-value = 0.001$. Again, we repeat this process for the results in Table 5 (Impact on investments) and Table 6 (Impacts on productivity and income) of Bold et al. (2022b), as shown in Tables 8 and 9. We note no noticeable deviation from our averaged results for the treatment effects as those reported in the original manuscript. This again exudes confidence in the claims by the original authors that despite a small sample size, the reported effects are robust to a variety of potential treatment assignment scenarios.

Table 7: Table 4: Permutation of Initial Seeds

Seed	Beta	SE	P-Value
Average	0.592857	0.0003461	0.00123
166	0.592857122	0.000373904	0.0014
180	0.592857122	0.00028273	0.0008
265	0.592857122	0.000299865	0.0009
436	0.592857122	0.000346202	0.0012
535	0.592857122	0.00033148	0.0011
537	0.592857122	0.00033148	0.0011
595	0.592857122	0.000346202	0.0012
805	0.592857122	0.000423882	0.0018
1000	0.592857122	0.00041196	0.0017
1125	0.592857122	0.00031607	0.001
1186	0.592857122	0.00033148	0.0011
1493	0.592857122	0.00033148	0.0011
1682	0.592857122	0.00031607	0.001
1694	0.592857122	0.00019996	0.0004
1729	0.592857122	0.000264483	0.0007
1800	0.592857122	0.000387008	0.0015
1876	0.592857122	0.000423882	0.0018
2030	0.592857122	0.00033148	0.0011
2151	0.592857122	0.000223551	0.0005
2306	0.592857122	0.000360321	0.0013
2399	0.592857122	0.000373904	0.0014
2621	0.592857122	0.00039968	0.0016
2638	0.592857122	0.000435476	0.0019
2685	0.592857122	0.000373904	0.0014
2739	0.592857122	0.000373904	0.0014
2859	0.592857122	0.00028273	0.0008
2900	0.592857122	0.00031607	0.001
2929	0.592857122	0.00033148	0.0011
3012	0.592857122	0.00028273	0.0008
3254	0.592857122	0.00033148	0.0011
3545	0.592857122	0.000346202	0.0012
3573	0.592857122	0.00033148	0.0011
3582	0.592857122	0.000373904	0.0014
3672	0.592857122	0.000423882	0.0018
3740	0.592857122	0.00028273	0.0008
3746	0.592857122	0.00028273	0.0008
3788	0.592857122	0.000264483	0.0007
3803	0.592857122	0.000346202	0.0012
3825	0.592857122	0.000244875	0.0006
3871	0.592857122	0.000299865	0.0009
4172	0.592857122	0.00033148	0.0011
4205	0.592857122	0.000244875	0.0006

4234	0.592857122	0.000346202	0.0012
4287	0.592857122	0.000346202	0.0012
4399	0.592857122	0.00033148	0.0011
4576	0.592857122	0.00033148	0.0011
4668	0.592857122	0.000299865	0.0009
4713	0.592857122	0.00028273	0.0008
4746	0.592857122	0.000346202	0.0012
4807	0.592857122	0.000373904	0.0014
4841	0.592857122	0.00039968	0.0016
5065	0.592857122	0.000299865	0.0009
5276	0.592857122	0.00031607	0.001
5425	0.592857122	0.00041196	0.0017
5427	0.592857122	0.00031607	0.001
5452	0.592857122	0.000387008	0.0015
5540	0.592857122	0.00041196	0.0017
5692	0.592857122	0.000387008	0.0015
5693	0.592857122	0.000423882	0.0018
5727	0.592857122	0.00033148	0.0011
6066	0.592857122	0.000299865	0.0009
6093	0.592857122	0.000423882	0.0018
6326	0.592857122	0.000423882	0.0018
6464	0.592857122	0.000360321	0.0013
6570	0.592857122	0.00031607	0.001
6701	0.592857122	0.000360321	0.0013
6708	0.592857122	0.000244875	0.0006
6723	0.592857122	0.000387008	0.0015
6876	0.592857122	0.00039968	0.0016
7084	0.592857122	0.00039968	0.0016
7189	0.592857122	0.000387008	0.0015
7283	0.592857122	0.000360321	0.0013
7530	0.592857122	0.00041196	0.0017
7650	0.592857122	0.00031607	0.001
7656	0.592857122	0.00039968	0.0016
7752	0.592857122	0.000299865	0.0009
7789	0.592857122	0.000299865	0.0009
7931	0.592857122	0.000346202	0.0012
8218	0.592857122	0.00033148	0.0011
8274	0.592857122	0.000387008	0.0015
8409	0.592857122	0.00041196	0.0017
8429	0.592857122	0.000299865	0.0009
8572	0.592857122	0.000435476	0.0019
8627	0.592857122	0.000360321	0.0013
8766	0.592857122	0.000423882	0.0018
8866	0.592857122	0.000299865	0.0009
9027	0.592857122	0.000373904	0.0014
9032	0.592857122	0.000373904	0.0014

9105	0.592857122	0.00041196	0.0017
9130	0.592857122	0.00031607	0.001
9196	0.592857122	0.00033148	0.0011
9213	0.592857122	0.000299865	0.0009
9240	0.592857122	0.00028273	0.0008
9275	0.592857122	0.000360321	0.0013
9509	0.592857122	0.000387008	0.0015
9651	0.592857122	0.000264483	0.0007
9722	0.592857122	0.00033148	0.0011
9764	0.592857122	0.000360321	0.0013
9918	0.592857122	0.000446766	0.002
9990	0.592857122	0.000435476	0.0019

Notes: This is a replication of Table 4 from Bold et al. (2022b) duplicating the Fisher-permutation test for 100 random initial seeds. Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

Table 8. Table 5: Permutation of Initial Seeds

Variable	Obs	Mean	Std. Dev.
Panel A			
Expenses: seeds & fertilizer	100	0.0496	0.0207
Expenses: all inputs	100	0.0875	0.0299
Proper drying	100	0.0007	0.0026
Sorting	100	0.0013	0.0039
Winnowing	100	0.0491	0.0208
Preharvest expenses	100	0.2908	0.0524
Postharvest expenses	100	0.272	0.0465
Postharvest labor expenses	100	0.1556	0.0359
Panel B			
Expenses: seeds & fertilizer	100	0.6655	0.0546
Expenses: all inputs	100	0.8847	0.0331
Proper drying	100	0.7714	0.0422
Sorting	100	0.4117	0.0523
Winnowing	100	0.0678	0.0237
Preharvest expenses	100	0.8753	0.0331
Postharvest expenses	100	0.7826	0.0389
Postharvest labor expenses	100	0.6217	0.0484

Table 8. *Notes:* This is a replication of Table 5 from Bold et al. (2022b) duplicating the Fisher-permutation test for 100 random initial seeds. Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

Table 9. Table 6: Permutation of Initial Seeds

Variable	Obs	Mean	Std. Dev.
Panel A			
Price	100	0.0036	0.0052
Maize acerage	100	0.8353	0.0395
Harvest	100	0.341	0.0483
Yield	100	0.0449	0.0224
Harvest value	100	0.1019	0.0311
Monetary expenses	100	0.3209	0.0499
Profit: monetary	100	0.0778	0.0289
Profit: incl. own hours	100	0.0285	0.0167
Panel B			
Price	100	0.4029	0.0504
Maize acerage	100	0.8075	0.0355
Harvest	100	0.6225	0.0529
Yield	100	0.6109	0.0513
Harvest value	100	0.7394	0.0489
Monetary expenses	100	0.6939	0.0454
Profit: monetary	100	0.9975	0.0054
Profit: incl. own hours	100	0.5195	0.0479

Table 9. *Notes:* This is a replication of Table 6 from Bold et al. (2022b) duplicating the Fisher-permutation test for 100 random initial seeds. Robust standard errors are clustered by enumeration area. Significant at the ***[1%] **[5%] *[10%] level.

4 Conclusion

We successfully reproduced the results through a computational reproduction via the publicly available replication packet provided by the authors of Bold et al. (2022b). Generally, we tested the sensitivity of the results in three domains. Firstly, we redefined the terminology for treatment and outcome measures. Second, we estimated the influence that outliers in the data have on the statistical significance of the results, as well as a formal test for any model misspecification. And finally, we tested the sensitivity of the Fisher-permutations to the initial seed generated. All three robustness replications found the results of the original manuscript are not sensitive to a variety of potential concerns. In conclusion, Bold et al. (2022b) is a soundly designed and implemented series of RCTs that provide insight into the economic development of smallholder farmers through market access but not through quality upgrading through a robust analysis. Their results are conceiving and should be confidently added to our knowledge of improving economic development for smallholder farmers.

References

Bold, T., Ghisolfi, S., Nsonzi, F., & Svensson, J. (2022a). "Data and Code for: Market Access and Quality Upgrading: Evidence from Four Experiments." *Open ICPSR*, 158401. DOI: 10.3886/E158401V1.

Bold, T., Ghisolfi, S., Nsonzi, F., & Svensson, J. (2022b). "Market Access and Quality Upgrading: Evidence from Four Field Experiments." *American Economic Review*, 112(8): 2518–2552. DOI: 10.1257/aer.20210122.

Brodeur, A. (2024). "Template Reproduction/Replication Report." *Institute for Replication*. DOI: 10.17605/OSF.IO/8DKXC.

Chamberlain, G. (1994). "Quantile Regression, Censoring, and the Structure of Wages." In: Sims C. A. (Eds.), *Advances in Econometrics: Sixth World Congress*, pp. 171-210, *Cambridge University Press*. DOI: 10.1017/CCOL0521444594.005.

Koenker, R. (2005). "Quantile Regression." *Cambridge University Press*. DOI: 10.1017/CBO9780511754098.

Li, G. (1985). "Robust Regression." In: Hoaglin, D. C., Mosteller, F., Tukey, J. W. (Eds.), *Exploring Data Tables, Trends, and Shapes*, pp. 281–340, *Wiley*. DOI: 10.1002/9781118150702.

Pregibon, D. (1980). "Goodness of Link Tests for Generalized Linear Models." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(1): 15–23. DOI: 10.2307/2346405.

Svensson, J., Nsonzi, F., Ghisolfi, S., & Bold, T. (2021). "Linking Farmers to Markets: A Clustered Randomized Trial on Agricultural Productivity and Technology Adoption." *AEA RCT Registry*, AEARCTR-0002812. DOI: 10.1257/rct.2812-3.0.

Tukey, J. W. (1949). "One Degree of Freedom for Non-Additivity." *International Biometric Society*, 5(3): 232–242. DOI: 10.2307/3001938.