

**The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in  
Fifteen Ethiopian Villages: Response to William Bowser**

Stefan Dercon

Daniel O. Gilligan

John Hoddinott

Tassew Woldehanna

January 12, 2015

## Abstract

In this reply, we explain why we welcome the principle of replication studies. We document how we have approached cooperation with the replicator but also express disappointment in how this process seems to have operated in practice; specifically the extent to which it created incentives to go beyond replication until methods and data were found that yielded different results. We are glad to note that in terms of pure replication, our results are confirmed beyond a minor coding error that did not matter for either the results or their interpretation. We are disappointed, however, that the replication study is selective in reporting our own cautious discussion on method and robustness in both our original *AJAE* paper, and a subsequent paper in *Journal of Development Studies*. We quote our own papers on how we already addressed a number of the concerns raised in this study and why we judge these innovations as being difficult to consider as 'superior' both in principle and in the way they are applied. The study places considerable weight on the robustness of our results on agricultural extension but ignores that we have highlighted as much in both papers before. We are not convinced that there is much value added in the part of the study that investigates robustness rather than just replicability.

## 1. Introduction

We appreciate the opportunity to respond to William Bowser's (WB's) replication of our 2009 *American Journal of Agricultural Economics* article, "The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages." Our note has two objectives: a) to describe, from our perspective as the replicatees, how this process operated; and b) to comment on WB's claim that his work has produced a novel findings, namely that our results on the impact of agricultural extension are not robust. While we disagree with this claim, we want to stress that our interactions with William, along with our interactions with 3ie, have been courteous and professional.

We begin with a chronology of this replication process; this is important in understanding some of our concerns regarding WB's work. We assess his findings before ending with some summary remarks.

## 2. Chronology of a replication

In June 2012, we received an email from Benjamin Wood indicating that our 2009 AJAE paper might be selected for 3ie's replication programme. He wrote, "3ie's replication program is designed to increase the conduct of *internal* [emphasis ours] replication studies, specifically of impact evaluations of development programs. In order to mitigate the incentive to replicate studies only to expose mistakes, the 3ie replication program provides incentives for replication by funding researchers to undertake replications and publishing the completed replications through the 3ie website and working paper series." A subsequent email in October 2012 indicated that our paper had been selected. 3ie kindly offered modest financial assistance (which we declined) to assist in making our data available for replication but did add that failing to do so would mean that our study would be placed on a list of studies described as "Unable to replicate." Following some introductory emails, WB wrote to us in December 2012 formally requesting our data and codes. While, at that time, we had already sent many of the necessary data files, WB added "I am still in possession of these files, however there are elements missing in these files indicating the mode of construction for important variables (e.g. household adult equivalent units) suggesting that there may be other instrumental data files and codes from earlier stages of the study. Therefore, I kindly request access to any additional files used in this study for the ensuing replication." On behalf of the author team, John Hoddinott replied (13 December 2012), copying in 3ie, as follows:

Thank you for this message. We are all pleased that you are trying to replicate our published results. As you've noted in your message, we've already sent you a considerable amount of data as well as the Stata do file used to generate the results reported in our AJAE paper. As you'll have seen if you've looked at the do file, we have annotated it in such a way that you can see which commands generate results reported in specific tables in the paper.

We are committed to being helpful and cooperative with your replication efforts. That said, we can only be helpful if you provide us with specific, focused requests. As you know, having read our paper, we use panel data collected over a ten year period. We have been working with these data since 1994; a number of the key variables such as the consumption aggregates and the measures of household size and composition have been constructed iteratively over time

as new survey rounds have been collected. So your request “access to any additional files used in this study” is really a request saying, “Could you please go back and find all files that you constructed over a 14 year period that are relevant to the data used in your 2008 paper.” This will include files used to construct the aggregates and files used to clean the underlying data. My guess is that these will consist of both Stata and SPSS files and that the code used to clean and construct all these variables will easily run to 20,000 lines, if not more. Given the time frame you have for this replication study, my guess is that you could easily spend two or three times the allocated time you have for this entire replication study simply re-checking the way we have cleaned our data, deciding if our treatment of non-standard quantities is correct and so on. To give a specific example, quantities of grains consumed in different parts of Ethiopia are sometimes expressed in Madaberia. There are large Madaberia and there are small Madaberia and their conversion to kilograms depends on what part of the country the data come from (other conversions depend on location and survey round as the conversion factors sometimes change over time). So for example, rechecking the construction of the consumption aggregates means that you would research how these non-standard conversions should be calculated, then recheck our code for all five survey rounds to see whether changing the conversions materially affects our findings. You could do this – and if you’d like, we can send you the raw data so that you can construct these aggregates as you see fit – but given the sheer amount of data that you will end up working with, you should be aware ex ante that doing this work correctly will consume a vast amount of time.

WB, in consultation with 3ie responded (14 December 2012):

Your point about the overwhelming amount of files and code generated from such an extensive study on an evolving data set is well noted. I agree that much time could be lost in the replication by going back and examining the details of data manipulation over a 14 year period. I have indeed looked at the do file and agree that there is sufficient material to start the replication. My concern was ensuring my compliance with the expectations of 3ie.

I am glad that Ben is copied into the discussion so that perhaps he can advise on how well your suggestion to start with the data and code you’ve generously provided converges with the expectations of 3ie.

Thank you again for your willing cooperation and assistance with this replication.

We highlight this for a specific reason. We are entirely sympathetic with the decision to work with the summary data files and not the raw data. As we noted, starting from scratch would have required considerably more time to replicate our work. But this represents a significant difference from the type of replication work undertaken in the sciences where researchers focus as much (if not more so) on the construction of the underlying data as they do on the statistical analysis. While we perceive ourselves to have been careful on our data cleaning and preparation work, we are not under no illusions that doing this required making the type of judgments described in our email to WB. Different researchers might make reasonable, but different judgments which might, or might not, lead to different findings. We think that in economics, this aspect of replication has received insufficient attention.

In August 2013, we received a short message from 3ie with the results of “the report for the pure replication of your study”. This short report indicated that “These results show that under the key

assumptions of slow changes in levels of capital stock and access to technology, the original authors' conclusions are indeed robust and their work was carefully executed despite a slight programming oversight." The oversight lay in a minor programming (we missed a logarithm in one of our lines of code) for which we accept full responsibility; fortunately this had no substantive bearing on our results.

The report also noted that "The next stage of the replication will endeavor to further explore the validity of the assumption made in which enables the construction of p-period averages across the rounds of the survey. Additionally, the robustness of the main findings will be held against an alternative definition of all weather roads as well as a longer time horizon for the consumption growth model." While this seemed to us an activity that went beyond the internal replication of our findings (which was how the 3ie exercise had been originally described to us), we saw no reason to comment on this.

We received some subsequent queries about some of our data but there were no further communications with WB or 3ie until November 2014 when we received the "final report" and were invited to respond. This final report contained the report of the pure replication work as well as additional work by WB. Unlike the interim report, the statement that our original conclusions were robust was buried in the text with no mention of this in either the abstract or conclusion; instead, emphasis was placed on the statement that our findings on agricultural extension were not robust. We wrote to 3ie, commenting that "the abstract and summary both indicate that our results are not replicated. This seems to be a fairly serious factual misstatement of the findings with reputational costs for all concerned." In response, 3ie and WB agreed to re-instate the findings of the pure replication work in both the abstract and summary.

### **3. Findings regarding the robustness of the results on agricultural extension**

The core finding of the work undertaken by WB between August 2013 and November 2014 is that the agricultural extension results are not robust when new data – an additional survey round conducted after our original study (data which we had made publicly available) - and new econometric techniques by Andreou et al published a year after our AJAE study. There are several aspects of this that are worth noting.

In his summary, WB discusses several reasons why agricultural extension does not, in his view, contribute to consumption or income growth. He cites several well-regarded studies, noting that the role of extension agents as purveyors of agricultural inputs and credit may have hampered their ability to transfer new knowledge. However, he neglects to reference our AJAE paper which made precisely this point. In fact, we specifically cautioned against reading too much into the results on agricultural extension. We concluded the paper by writing (pp 1018-1019), "Understanding why agricultural extension has positive impacts is trickier ... Given this, drawing implications of our results for agricultural extension should be done cautiously. Some of the effect may represent transfers of technology or knowledge, while some of the effect may reflect the influence that extension agents have in terms of increased use of fertilizer and other inputs."

WB is concerned that in our basic estimates, we treat access to roads and extension as exogenous. We see this as a legitimate concern. However, WB does not reference our own explorations reported in the AJAE paper. When we treat access to extension as endogenous, we find an increased

estimate of impact of extension, just as WB does. However, unlike WB, we are more cautious about these. We write (p1016), “Because the number of agricultural extension officers changes only slowly over time—and in some localities does not change at all—it is possible that these results are picking up a local treatment effect (LATE) as opposed to an average treatment effect over the full sample”.

WB places considerable emphasis on the use of GMM estimation techniques. We agree that these are powerful tools. However, in our AJAE paper we noted that while GMM estimates are both consistent and efficient, they are vulnerable to the influence of outliers because the optimal weighting matrix that underpins them is a function of fourth moments. For this reason, we assessed the robustness of our GMM estimates in several ways, including trimming the top and bottom 1% of the sample and estimating using a LIML estimator that is both not vulnerable to this concern and has the added advantage of being a superior estimator when instruments are weak. WB does not consider these possibilities.

WB argues that we should only use the evenly spaced rounds of data (1994, 1999, 2004 and 2009) and not rounds that were not evenly spaced. He claims that our results on agricultural extension no longer hold up. Consider, for example, his finding on the impact of agricultural extension on consumption growth. He obtains a parameter estimate of 0.0723 with a standard error of 0.0891 (Table 3, column 2). In our AJAE paper, we report parameter estimates of 0.059 (se: 0.037) and 0.071 (se: 0.036), see our Table 3, columns 3 and 4. Subsequent models produce, with one exception, parameter estimates that lie between 0.088 and 0.189, all of which are higher than the estimate we report. WB does not test whether the difference between his parameter estimates and ours are statistically significant; our guess, based on his and our reported standard errors is that they are not.

WB places weight on the need to estimate models of consumption growth using evenly spaced panel data by drawing on a study by Andreou et al (2010) published after our AJAE paper came out. To do so, he drops two survey rounds that we had used in our AJAE paper (1995 and 1997) and adds an additional survey round collected by ourselves in 2009. Doing so is no longer an exercise in replication. In this extension to the original aims of his study, WB is now longer taking our original data and seeing if he can obtain our results – the original intention of the 3ie replication programme. Instead, he assesses whether the results use of new data and new methods confirm our findings. He notes that doing so was made difficult because of problems associated with coding of identifiers which reduced the size of the useable sample. Specifically, WB reports that his merging process reduces the sample size by at least 17.5 percent. This makes it difficult for us to compare our results to his. If the loss of 17.5 percent of the sample is non-random (and we have no way of knowing this because it is not discussed in his paper), it is entirely possible that WB’s results fundamentally differ from ours because he has chosen to work with a non-random sub-sample of our data. Further, the loss of these observations *ceteris paribus* increases his estimated standard errors. His comments, in several places, that the results on agricultural extension are no longer statistically significant could merely reflect this loss of statistical power. This is not considered by WB.

We agree with WB that assessing the robustness of our findings to the inclusion of the 2009 survey round is an important and interesting line of inquiry. In fact, we had already done so, see S. Dercon, J. Hoddinott and T. Woldehanna, 2012. “Growth, poverty and chronic poverty in rural Ethiopia: Evidence from 15 Communities 1994-2009,” *Journal of Development Studies*, 48(2): 238-253, <http://www.tandfonline.com/doi/pdf/10.1080/00220388.2011.625410>. This raises several issues.

We are baffled by the fact that WB never contacted us about the difficulties he experienced merging in the 2009 data. We would have gladly shared our code and the corrected data sets, saving him both time and trouble while allowing him to have used a larger sample. This simple request would have addressed both concerns regarding the non-random attrition he imposes on the data as well as increasing power. So while the focus of WB's work is on the robustness of our findings, there are unaddressed concerns regarding the robustness of his.

Perhaps more importantly, WB does not reference our 2012 *JDS* paper. This is unfortunate. We quote at length from pages 247 and 248:

Before continuing, we note a potential econometric concern with the results of Table 5 resulting from the fact that the ERHS rounds are not evenly spaced. As noted by Andreou et al. (2010), this can lead, under certain conditions, to biased and inconsistent parameter estimates. However, if we restrict ourselves to using data collected in 1994, 1999, 2004 and 2009, we have four evenly spaced rounds, which allows us to circumvent this problem. Results using only these rounds, together with village-round interaction terms (expressed as dummy variables), are reported in Table 5B. As seen by the R2 statistic, estimating the model with household fixed effects, village x round dummy variables and lagged consumption soaks up a lot of the variation in consumption growth rates and so it is not especially surprising that the household-level shock variables, *access to extension* [emphasis added] and death and illness shocks, are no longer statistically significant since the sample size is also smaller due to dropping two rounds of data, making precise estimation harder.

So it would seem that WB's major criticism has already been addressed by us. We had already found that his proposed approach eliminates much of the variability in our data and this, together with his decision to drop data from two rounds that we had included in our study meant that obtaining precise estimates was more difficult.

#### **4. Summary**

In his replication paper, WB makes two claims. First, he replicates the results of our 2009 *AJAE* paper. Second, he undertakes new analysis with additional data, arguing that this overturns our finding of the impact of agricultural extension on consumption growth. We welcome both his replication efforts along with his work assessing the robustness of our findings. We agree with his pure replication findings but take issue with the novelty of his findings regarding the impact of agricultural extension on consumption growth. Specifically, in our original study we noted that these results needed to be interpreted with care. WB echoes these concerns but fails to acknowledge that these were points that we had already raised. He criticizes our work for failing to address problems arising from the uneven spacing of our panel; however, he does not cite our subsequent work where we directly address this issue. Further, while we welcome the additional tests of robustness, these are not without their own problems including non-random sample selection bias, reduced statistical power, interpretation of IV estimates as ATE or LATE and the elimination of variability in the agricultural extension variable.

At the outset of this exercise, we were enthusiastic, but possibly naive participants. At its end, we find it hard to shake the feeling that an activity that began as one narrowly focused on pure

replication morphed – once our original findings were confirmed (save for a very minor programming error that we willingly confess to) - into a 14 month effort to find an alternative method/structure of researching the problem that would yield different results. While this is a legitimate research endeavor, it is not obvious to us that this is what is meant by replication. Further, while both WB and 3ie graciously agreed to amend the abstract and summary report to make clear our *AJAE* results were replicable, this only came after a specific request from us. So we perceive that WB's study only partially satisfies the objective of 3ie's replication program. While it increases the conduct of *internal* [emphasis ours] replication studies, it seems that it also incentivized efforts to obtain different results.

In this regard, our experience seems very much like that discussed by Robert Jensen and Emily Oster in their response to a similar 3ie replication exercise. They make several points about the conduct and consequence of replication studies, including: the need to have clear understandings of what replication activities will be undertaken; how and when replicators and replicatees interact with each other; how they will be reported and most importantly how findings will be discussed in a way that encourages healthy debate without trashing the reputations of all concerned. We can only echo these good suggestions, noting that had they been followed in this study, a number of the weaknesses in WB's work – most notably the needless introduction of non-random sample selection bias and the failure to recognize that many of his concerns had already been considered in our subsequent work – could have been easily avoided.