Referee Report on "Impact Evaluation Findings after One Year of the Productive and Business Services Activity of the Productive Development Project, El Salvador" (July 3 2012 version)**"**

Michael Carter
University of California, Davis

July 2012

This evaluation employs a randomized rollout design, with early-treated producer households entering the Productive and Business Services (PBS) program approximately 12 months prior to the entry of the late-treated households. The study evaluates three areas of economic activity (dairy, horticulture and handicrafts). The study notes that the RBP could be reasonably expected to generate impacts in each of these activities within the time frame of a year. The study separately analyzes each activity, employing standard binary treatment estimators (ITT and TOT).

While the study is well-executed, I have several comments that may help sharpen our understanding of what has been learned from this evaluation.

1. *Sample size & Statistical Power of the Design*
   Each of the three activities being evaluation is studied separately with approximately 600 observations per activity (300 early-treatment units; 300 control/late-treatment units). These sample sizes seem a bit slender given the outcome variables being analyzed. This intuition is consistent with the fact that some very large point estimates of treatment effects are statistically insignificant. For example, the dairy program shows an ITT that implies a 50% increase in employment, and yet it is statistically significant. Similarly the point estimates for treatment effects for household income and consumption are both large (indicating increases in the range of 50%) and yet they too are insignificant. Did the original power calculations indicate that these relatively small samples would be large to reliably detect reasonable effects. I say this as I would think that any program like this that increased income or expenditure by "only" 15% in a year would be considered successfully (and hence the I would have thought that MDEs would have been in this range.

   There is of course nothing to be done at this time about the sample sizes. However, if it is indeed fair to characterize the study design as 'underpowered,' then I would at least report 95% interval estimates for these key outcome variables. Those intervals are obviously very wide, but equally obviously make it impossible to reject the hypothesis that the program impacts were huge.

   Finally, an alternative approach would be to pool the analysis across activities. This is in fact the approach followed in the Nicaragua evaluation, not because it would not be interesting to know activity-specific impacts, but because project

budget did not allow a large enough sample to allow reliable inference at the activity level.

2. *The Family Income/Consumption Puzzle*
Similar to other MCC evaluations, this study useful looks at both activity outcomes (activity specific income and input use) as well as household level outcomes (in this case, total household income and total household expenditures). The study finds that total household income went up by 2608 dollars, while consumption rose only 1221 dollars (both with p-values of ~0.12)

One concern might be that the change in total income measures are biased upwards. Imagine that early-treated households shifted labor towards their newly enhanced dairy activity and away from prior, say, informal work. Dairy presumably generates a regular cash flow, making it likely that income earned in this activity is properly reported. However, returns and earnings from sporadic informal work are notoriously hard to measure. Might it be possible that measured household income is systematically understated or control/late treated households relative to early-treated households?

It is of course entirely possible that the gap between the estimated income and consumption increases reflects cash siphoned off for investment. Perhaps akin to the Nicaragua evaluation, this can examined directly by looking at investment.

It is also possible that there is some very interesting heterogeneity in how households spend income increases (consumption versus investment). It would certainly be consistent with the intra-household bargaining literature to expect that when benefiting women, programs boost measured consumption more than when they target men. In the three activities, women make up ~15-25% of the beneficiaries. Endogeneity aside (at least initially), it would be very interesting to see if the pattern of consumption impacts is mediated by the sex of the producer. The Nicaragua evaluation has some provocative (but insignificant evidence on this point). Given that MCC programs are designed to impact poverty and living standards, then this idea might merit further investigation.

3. *What's Next?*
The strong design parallels between this study and the Nicaragua evaluation raises the issue about next steps. In particular, there would appear to be substantial synergy from learning from (and hopefully improving upon!) the double complier and continuous treatment approaches used in that latter evaluation. This steps would seem especially important as the PBS program was just moving into its phase 2 (with more investments coming) at end of the midline study. Unlike the Nicaragua study, it may be possible to define a better measure of treatment duration, capturing not only time in the program, but also the monetary value of resources transferred under the two phases of the PBS. I write this on the assumption that all participants graduate to phase 2. If that is

not correct (and phase 2 benefits are endogenous to phase 2 success), then a continuous treatment effect that relies on temporal duration in the program would be the more reliable approach.