Economics Bulletin

Volume 39, Issue 3

Recovering the counterfactual as part of ex-ante impact assessments: an application to the PASIDP – II project in Ethiopia

Manuela Coromaldi
Università degli Studi Niccolò Cusano

Alessandra Garbero
International Fund for Agricultural Development (IFAD)

Marco Letta
Sapienza Università di Roma

Abstract

Real-world ex-ante impact assessments are far from the ideal experimental design, where the eligible population is supposed to be randomly assigned to treatment and control groups. Often, many surveys in developing contexts do not even collect data from a comparison group. We propose a methodology that recovers the counterfactual for ex-ante impact assessments of policy interventions under the conditions of distance decay in the exposure to continuous treatments and lack of control groups. We test this approach on data from a large-scale irrigation project in Ethiopia.

Citation: Manuela Coromaldi and Alessandra Garbero and Marco Letta, (2019) "Recovering the counterfactual as part of ex-ante impact assessments: an application to the PASIDP – II project in Ethiopia", *Economics Bulletin*, Volume 39, Issue 3, pages 1844-1854
Contact: Manuela Coromaldi - manuela.coromaldi@unicusano.it, Alessandra Garbero - a.garbero@ifad.org, Marco Letta - marco.letta@uniroma1.it.
Submitted: May 27, 2019. **Published:** August 11, 2019.

1. Introduction

Randomized experiments are considered the gold standard in causal inference,because random assignment into treatment programs allows to overcome any sort of selection bias. In real-word situations, however,there are many reasons for which random assignment is often infeasible in practice. In such settings, in order to assess the causal effect of policy interventions, one must rely on results fromnon-randomized studies. When assignment into treatment is not random,quasi-experimental techniques can be exploited to neutralize as much as possible the selection that stems from the non-random allocation of individuals into programs.

In the case of prospective or ex-ante impact assessments, a rigorous evaluation of policy interventions requires that treatment and control groups are drawn from an eligible population determined in such a way to exhibit sufficiently similar characteristics at baseline. But, in the majority of cases, monitoring surveys fail to meet the gold standard conditions of an experimental design, thus requiring the use of control approaches and quasi-experimental methods to construct a valid counterfactual.

To make things worse, however, survey design sometimes even fails to identify an appropriate comparison or control group at baseline. Lack of information on untreated units (i.e., surveys that did not collect data on non-beneficiaries) can prevent the execution of a rigorous ex-ante impact assessment even if using quasi-experimental techniques.On top of all this, there are many situations in which the nature of the policy implementedis continuous rather than dichotomous. When the treatment is continuous, it is not only the status of 'treated' or 'untreated' which matters, but alsothe intensity of the treatment received.

Using data from the baseline survey of the second phase of the Participatory Small-Scheme Irrigation Development Project (PASIDP-II) in Ethiopia, we propose a methodology for reconstructing thecounterfactual under these special conditions, i.e., non-random assignment, lack of control group and continuous treatment, while preserving the quasi-experimental nature of the impact assessment ex-ante design.

2. Data and methods

2.1 Data

Data come from the baseline survey of the second phase of the Participatory Small-Scheme Irrigation Development Project (PASIDP) implemented by the International Fund for Agricultural Development (IFAD) and the Government of Ethiopia in four Ethiopian regions: Amhara, Oromia, SNNPR and Tigray. PASIDP-II is a large-scale irrigation programme that consists of the development of 150 modern irrigation schemes and improvement of the adjacent watersheds. The baseline survey collected data on 3000 beneficiary households from 74 out of the scheduled 150 schemes during the period May-June 2018. See Figure 1 for a map of the surveyed schemes.The new irrigation schemes and improved watersheds are spread across the four regions targeted by the project.Specifically, out of the 74 surveyed schemes, 11 are located in Tigray, 23 in Amhara, 30 in Oromia and 10 in SNNPR.[1]

[1]Cf. Table 1 for sample size allocation by region.

The sample was selected in stages to identify the locations where interviews took place and in order to choose the respondents efficiently. The design considered stratifications in such a way that the sample actually selected was properlyover geographic sub areas and population sub-groups.The sample size for the survey was selected by considering key parameters including target populations, precision and confidence level required, estimation domains, and allowance for non-response. Table 1 shows the sample size allocation by region compared to the number of beneficiary households. The sample size is distributed over the four regions with weights proportional to the number of beneficiary households of the regions for both groups of beneficiaries (irrigation schemes and watersheds) and to the investment budget (87% for irrigation systems and 13% for the adjacent watersheds).

When the survey was implemented, an updated and complete list of beneficiary households was not yet available. Therefore, to preserve randomness, random walk and a quota sampling technique (where 'quota' refers to the total number of households allocated to a given scheme) wereimplemented to select the respondent households from each scheme.The survey adopted a participatory approach aimed at including both qualitative and quantitative methods of data collection. To this end, survey instruments incorporated not only the typical household interviews, but also Key Informant Interviews (KIIs) and Focus Group Discussions (FGP) with potential beneficiaries from the schemes.Household questionnaires included modules on household demographics, education, labour participation, assets, food security, food and non-food expenditure, agricultural inputs and outputs, livestock activities, sources of income, access to information, market integration, resilience as well as on self-reported idiosyncratic and covariate shocks.

A key feature of the data collection was that only data on beneficiary households was collected. 'Beneficiary' households include both households living within the command area of the irrigation schemes as well as households living outside the command areas but within the range of the adjacent watersheds that will also be reached by project's interventions. A command area refers the area which can be irrigated from an irrigation system and is suitable for crop cultivation.

In such a context, it is reasonable to expect that households living inside or closer to the command areas will be exposed to a more intense treatment compared to households residing further away, who will benefit from the improvement of the adjacent watersheds but will be only indirectly affected by the effects of the scheme.

2.2 Methods

Consistently, we assume distance decay in the joint impacts of the irrigation schemes and improved watersheds and employ the inverse of the distance of each household from the respective scheme as a proxy for the intensity of the continuous treatment.

Our analysis consists of two steps: i) the analysis of preliminary descriptive statistics of key household-level variables, and their heterogeneity among groups of households that differ in terms of geographical distance from the respective schemes; ii) a proposal and empirical test for counterfactual building, to make up for the lack of data from non-treated households and the absence of control groups, and lie the foundation for the future quantitative assessment of the programme's effects on household welfare outcomes.

We divide the household sample in three groups, depending on the geographical distance from the respective scheme: i) household residing inside the command area of the respective scheme; ii) households residing outside the command area of their scheme but with a below than average distance among those who reside outside the command area; iii) households residing outside the command area of their scheme and with an above than average distance among those who reside outside the command area.

After detecting a statistically significant, and often large in magnitude, heterogeneity in the average levels of key household-level variables between the three different groupings determined based on distance, we conclude that a valid and unbiased counterfactual for an ex-ante impact assessment could not be established without a careful matching and balancing of household characteristics across the three groups. To this end, we implement step ii), by proposing and testing a quasi-experimental technique which exploits the continuous nature of the treatment to build validcontrol groups suitable fora rigorous quantitative assessment of the impacts of the small-scale irrigation schemes.

Weapply the generalized propensity score (GPS) matching technique to reconstruct the counterfactual ex-ante rather than estimate treatment effects ex-post, following the approach introduced by Hirano and Imbens (2004), as implemented by Bia and Mattei (2008) and subsequently extended by Guardabascio and Ventura (2014).[2] Differently from the traditional Propensity Score Matching (PSM), the GPS can handle cases in which the treatment takes continuous, not discrete, values (Imbens and Wooldridge, 2009). Additionally, it also solves a common problem in PSM estimation, namely the lack of the so-called '*common support*' between treated and non-treated units, i.e. the lack of correspondence of key characteristics between treatment and control groups. By design, the GPS technique only restricts impact evaluation to the group of treated units. More specifically, this technique allows building control groups based on a set of given covariates and tests for the validity of the balancing property (BP), a critical requirement for the reliability of the counterfactual thus created. Meeting the BP implies that assignment into treatment is unconfounded given the GPS. A necessary condition for the implementation of this method is that all the covariates used for counterfactual building refer to the 'pre-treatment' period: this is exactly the case inour dataset, since the baseline survey has been collected at a time when the schemes have been built and installed but their positive effects are not yet visible on households' characteristics and welfare indicators.

We test for the validity of the balancing property of the GPS estimator with respect to three key covariates: household size, gender of the household head and the Resilience Capacity Index (RCI), an aggregate indicator which broadly captures household resilience computed applying the methodology developed by the Food and Agriculture Organization (FAO, 2016).The RCI represents a field-tested and scientifically validated indicator of resilience to food insecurity. It is computed through a two-step procedure. In a first step, four resilience pillars (namely Access to Basic Services, Assets, Adaptive Capacity and Social Safety Nets), are generated starting from a set of either observed variables taken directly from the household raw data or other indicators generated from observed variables.[3] In a second step, a Multiple Indicators Multiple Causes (MIMIC) model is employed to generate resilience as a latent variable with respect to two food security indicators (food consumption and a subjective food security indicator taken by the questionnaire).[4]

---

[2]In particular, we use their Stata package *gpscore2*.
[3]A list of the variables used to construct the resilience pillars is available in Table 2. Technical details on the results of the RCI estimation and more information on the choice of variables used to construct the RCI are available upon request.
[4]For more details on the RIMA methodology, readers can consult FAO (2016).

As a robustness check to rule out sensitivity between different estimation procedures, we also re-estimate the GPS following the extension proposed by Bia *et al*. (2014) that allows to estimate the GPS parametrically under alternative distributional assumptions. In particular, we assume again a gamma distribution and assess the balancing property using a model-comparison approach with a likelihood ratio (LR) test in place of the standard two-sided t-tests.

3. Results and conclusions

Before implementing the BP test and GPS estimation, preliminary descriptive statistics of key household-level variables are provided. Specifically, we present summary statistics and t-tests of all the variables involved in the computation of the RCI (either via the pillars or as food security indicators acting as latent variables) as well as for the other two variables (gender of the household head and household size) employed as covariates to build the common support in GPSestimation. These statistics are reported in Table 2. As it is visible, many of these variables and household characteristics differ significantly between households residing at different distances from their respective scheme. Such tests confirm the need for control groups with similar characteristics at baseline to go beyond intrinsic household differences and, more generally, to conduct a sound evaluation of this large-scale irrigation programme and its targeting strategies.

Therefore, after identifying a statistically significant heterogeneity in the average levels of key household-level variables, we implement our GPS methodology that assumes a gamma distribution and exploits distance from the scheme as a proxy treatment intensity. The use of the GPS technique allows us to build control groups based on the set of our three keycovariates that proxy for targeting(household size, gender of the household head and the RCI) and test for the validity of the balancing property.

Results of the BP test and GPS estimation are presented in Table 3. A comparison of the t-tests of the differences-in-means of the three covariates before and after the implementation of the GPS shows that the methodology adopted leads to the creation of valid control units that are comparable to the treated ones. In practice, this means that the GPS allows to make up for the lack of data on control groups and the significant household heterogeneity detected in the descriptive statistics by creating an *ad hoc* counterfactual and matching beneficiary households with similar key characteristics but different treatment intensity levels.

The results of the LR tests using the alternative GPS estimator proposed by Bia et al. (2014) are shown in Table 4: the restricted model that excludes the three covariates (household size, gender of the household head and RCI) is not rejected (p-value is 0.82), whereas the restricted model that excludes the GPS is strongly rejected with a p-value of 0.This evidence confirms that the balancing property is satisfied and that the counterfactual generation process is statistically valid.

In conclusion, this quasi-experimental technique represents a reliable solution for reconstructing the counterfactual ex-ante in these special circumstances, i.e. continuous treatments with spillover effects and no data on comparison groups. In the follow-up monitoring surveys of the PASIDP-II project, it will thus be possible to carry out an ex-ante impact evaluation of the outcomes of this large-scale programmeeven without data on non-beneficiaries by re-surveying the matched groups and keeping the integrity of these groups in the analysis stage.

The methodology we propose here can be easily replicated and employed for other ex-ante impact assessments of projects with similar characteristics. A common feature of many policy interventions in developing contexts is the existence of budgets limitations time constraints and technical difficulties during the key stages of survey design and data collection in the field. In turn, this can sometimes prevent the collection of data on non-beneficiaries and/or lead to biased targeting strategies of beneficiary units. In such cases, our approach can provide a flexible solution to overcome household heterogeneity by establishing groups that are not systematically different in key targeting variables, ex-ante, e.g. at the beginning of the project. This approach can consequently guarantee the robustness of the subsequent quantitative evaluation of projects' impacts, provided that such groups can be re-surveyed to measure outcomes at the final stage of the project.

References

Bia, M. and A. Mattei (2008) “A Stata package for the estimation of the dose-response function through adjustment for the generalized propensity score” *The Stata Journal*, Vol.8 (3), 354-373.

Bia, M., Flores, C., Flores-Lagunes, A., and A. Mattei (2014) “A Stata package for the application of semiparametric estimators of dose-response functions” *The Stata Journal*, Vol.14, 580-604.

Food and Agriculture Organization (FAO). 2016. Resilience Index Measurement and Analysis (RIMA) - II. Available at: http://www.fao.org/3/a-i5665e.pdf.

Guardabascio, B., and M. Ventura (2014) “Estimating the dose-response function through a generalized linear model approach” *The Stata Journal*, Vol. 14 (1), 141-158.

Hirano, K., and G. W. Imbens (2004) “The propensity score with continuous treatments” in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed.A. Gelman and X.-L. Meng, 73-84. West Sussex, England: Wiley InterScience.

Table 1: Sample size allocation by region

Region	Number of Beneficiary Households			Sample size allocation		
	Irrigation	Watershed	Total	Irrigation	Watershed	Total
Tigray	2421	4793	7214	250	49	299
Amhara	6943	10892	17835	710	118	828
Oromia	12877	16867	29744	1322	177	1499
SNNPR	3155	4849	8004	324	50	374
Total	25396	37401	62797	2606	394	3000

Table 2: Summary statistics of variables used in GPS Estimation by distance from the scheme

	Inside CA	Outside command area - Closer	Outside command area - Further	T-test p-values		
	Mean	Mean	Mean	1 vs 2	1 vs 3	2 vs 3
Socio-demographic characteristics						
Household size	5.302	5.593	5.610	0.037	0.025	0.863
HH head gender (Male=1; Female=2)	1.118	1.109	1.100	0.590	0.306	0.476
Access to Basic Services						
Road access (Yes=1)	0.763	0.806	0.809	0.059	0.050	0.857
Dwelling index	-0.006	-0.025	0.033	0.724	0.493	0.140
Nearest market (in kilometers)	11.163	12.226	12.648	0.297	0.085	0.533
Assets						
HH total area of all land per capita (Ha)	0.198	0.219	0.225	0.104	0.028	0.514
Livestock index - scaled (0-1)	0.029	0.029	0.035	0.770	0.069	0.005
Wealth index (95% variance)- scaled (0-1)	0.228	0.246	0.249	0.003	0.002	0.560
Social Safety Nets						
Relationship with a financial institution (Yes=1)	0.103	0.138	0.138	0.069	0.076	0.992
Access to financial services for saving (Yes=1)	0.081	0.123	0.118	0.019	0.038	0.726
Access to financial services for loan (Yes=1)	0.015	0.012	0.020	0.605	0.559	0.106
Adaptive Capacity						
Market participation (Yes=1)	0.098	0.177	0.179	0.000	0.000	0.898
Dependency ratio	0.544	0.532	0.523	0.372	0.134	0.332
Educational level of HH head	1.662	1.697	1.794	0.562	0.044	0.024
Food security						
Annual p.c. food consumption (US $)	196.503	189.706	173.321	0.588	0.020	0.032
Food security indicator (self-reported)	0.537	0.732	0.735	0.000	0.000	0.830
Resilience Capacity Index (0 to 100)	46.410	47.935	47.735	0.005	0.018	0.588
No. of observations	397	1435	1168			

Table 3: GPS estimation and Balancing Property

Variable	Mean difference	Standard deviation	t-value pre-matching	t-value post matching
Group 1				
Variable	**Mean difference**	**Standard deviation**	**t-value pre-matching**	**t-value post matching**
Household size (ln)	-0.00835	0.02135	**-2.1916**	-0.39078
Gender of the household head	-0.01102	0.01134	**1.6946**	-0.9714
Resilience Capacity Index	0.02954	0.35164	**-2.9226**	0.08399
Group 2				
Variable	**Mean difference**	**Standard deviation**	**t-value pre-matching**	**t-value post matching**
Household size (ln)	0.03544	0.02651	**3.2919**	1.3368
Gender of the household head	0.00048	0.01233	**-2.5630**	0.03867
Resilience Capacity Index	0.04419	0.44437	**2.9052**	0.09944
Group 3				
Variable	**Mean difference**	**Standard deviation**	**t-value pre-matching**	**t-value post matching**
Household size (ln)	-0.01247	0.04491	-0.7665	-0.27759
Gender of the household head	0.03917	0.02403	**2.2397**	1.6297
Resilience Capacity Index	1.2073	0.75452	1.0980	1.6
Group 4				
Variable	**Mean difference**	**Standard deviation**	**t-value pre-matching**	**t-value post matching**
Household size (ln)	-0.02441	0.04132	-0.3480	-0.59071
Gender of the household head	-0.00861	0.02033	-1.2259	-0.42333
Resilience Capacity Index	-0.35683	0.69173	-0.0005	-0.51586

Notes: values in bold indicate moderate, strong to very strong or decisive evidence against the balancing property. Treatment is the inverse of the distance (in kilometers) of each household from the respective scheme. Treatment intervals are the three followinggroups: 1) Household living inside the Command Area (CA); 2) Household living outside the CA but witha below than average distance among those who reside outside the CA); 3) Household living outside the CA with an above than average distance among those who reside outside the CA. Groups 1-4 are the GPS intervals. Mean difference stands for the post-matching difference-in-mean of each covariate between units that belong to the treatment interval and units that are in the same GPS interval but belong to another treatment interval. GPS estimated using the Stata package gpscore2 by Guardabascio and Ventura (2014).

Table 4

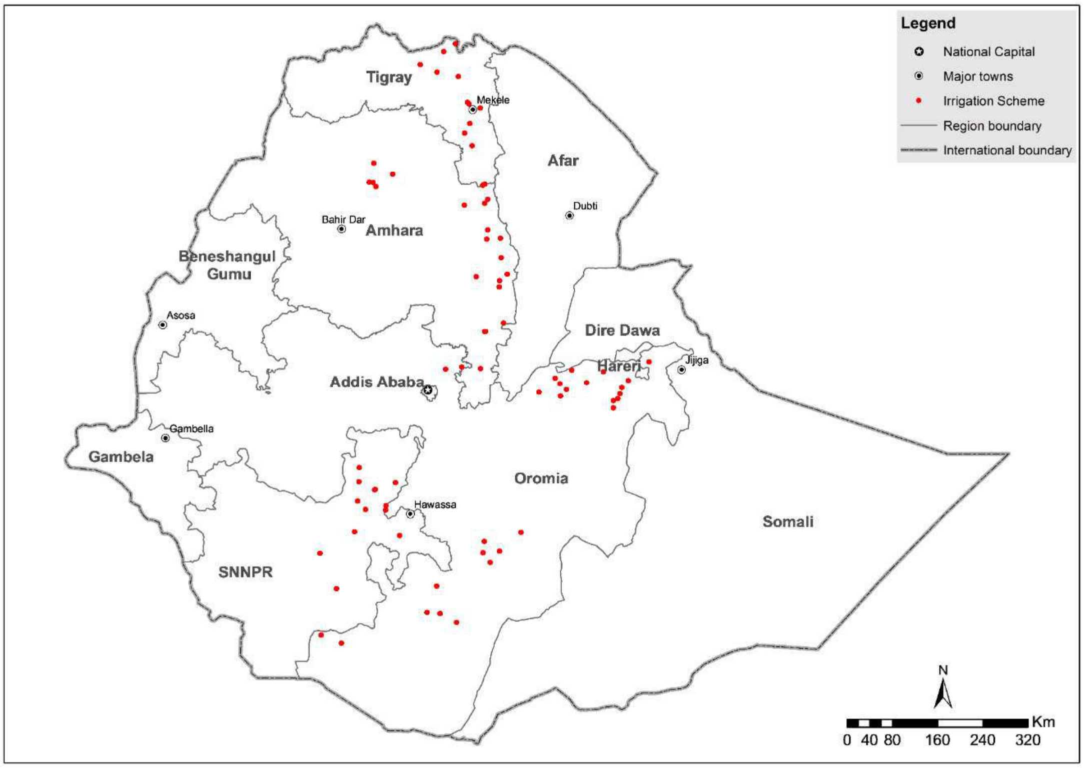
Likelihood-ratiotests:

Comparison between unrestricted and restricted models (Bia et al. 2014)

Models	**LR Test**	**T-statistics**	**p-value**	**Restrictions**
Unrestricted	320.330			
Covariates X	319.869	0.922	0.82	3
GPS terms	-2127.535	4895.73	0	3

Notes: GPS estimated using the Stata package *drf* by Bia *et al*. (2014).

Figure 1: Map of the irrigation schemes



Source: ITAB Consult