What’s Missing From Policy Evaluation: Identification and Estimation of the Distribution of Treatment Effects

Il pezzo che manca alla valutazione delle politiche: identificazione e stima della distribuzione degli effetti del trattamento

Erich Battistin¹, Margherita Fort²
¹ Dipartimento di Scienze Statistiche, Università di Padova
e-mail: erich.battistin@unipd.it
² Dipartimento di Economia, Università di Bologna

Riassunto: La ricerca statistico-econometrica per lo studio degli effetti di un intervento pubblico ha sottolineato la rilevanza per l’analisi empirica dell’eterogeneità di tale effetto tra i partecipanti. La letteratura si è tradizionalmente concentrata sulla media dell’effetto, parametro che non sempre risponde ad importanti domande di policy: ad esempio, politiche pubbliche alternative che producono lo stesso effetto medio non sono da considerarsi equivalenti se tale uguaglianza in media è accompagnata da proporzioni molto diverse di partecipanti per i quali il beneficio è positivo. In questo lavoro ripercorriamo i (pochi) articoli che discutono le difficoltà operative che derivano dallo spostare l’attenzione dalla “media” alla “distribuzione” dell’effetto, discutendo le implicazioni di policy che ne conseguono per un importante intervento pubblico in Italia.

Keywords: impact heterogeneity, quantile treatment effects, rank invariance

1. Introduction

Countless theoretical and applied work from the early 70’s has addressed the evaluation problem, that is the identification of the causal effect of a policy intervention on one or more outcomes of interest (see Heckman et al., 1999, for a review). The evidence from almost all empirical studies points to heterogeneous effects, for example, in labor market and education programs.

The existence of such heterogeneity notwithstanding, drawing causal inference on the effects of a policy intervention has been traditionally concerned with the measurement of the mean effect of the intervention. The role played by heterogeneity is typically investigated by comparing the mean effect for different subgroups of the population identified by observable characteristics. However, this strategy does not allow to draw definitive conclusions about the distribution of the effects of the policy that can only be inferred by considering alternative policy parameters.

The common practise of looking at mean effects is mostly the result of a pragmatic approach to the evaluation problem. On the one hand, well established statistical techniques that have been developed in the literature over the years mainly focus on averages. Most importantly, identification of characteristics of the effect distribution other than the mean is precluded on a logical ground, since the assumptions required to achieve

¹(*) This paper benefited from comments by Alberto Martini and Enrico Rettore. Financial support from MIUR to the project “Evaluating the impact of public interventions: methods and case studies” is gratefully acknowledged.
point identification of the mean effect are not enough to retrieve the effect distribution. As it will be discussed in what follows, identification of the effect distribution needs additional assumptions whose validity is rarely plausible from an economic perspective.

The aim of this paper is twofold. First, we will discuss the policy relevance of parameters that are alternative to the mean effect. Second, we will review some recent developments that have been proposed in the literature to overcome the identification problem related to these parameters. The latter problem will be dealt with in a general non-parametric setting, thus without relying upon parametric restrictions.

The remainder of the paper is organized as follows. After introducing the basic notation in Section 2, in Section 3 we discuss a variety of policy relevant parameters which are able to capture heterogeneity of the effect across individuals. Section 4 discusses the identification problem that undermines direct estimation of policy parameters that require knowledge of the impact distribution and reviews alternative solutions. Estimation methods are presented in Section 5. Section 6 exemplifies some of the methods discussed to the case of the estimation of the returns to educational qualifications in Italy, while Section 7 concludes.

2. Notation

Throughout this paper $Y$ will denote the observed outcome, $D$ the intensity of the treatment received and $W$ a set of observable individual characteristics. For each unit we define a set of potential outcomes $Y_d$ representing the outcome that would be observed on each individual had the individual been exposed to level $d$ of the treatment. Though $\{Y_d\}_{d \in D}$ can be defined for each individual, only one outcome is observed depending on the level of the treatment actually received.

Comparisons of potential outcomes for well defined populations of individuals define a variety of causal parameters. For the sake of exposition, we will maintain the binary treatments case as guiding example throughout, as the extension to the continuous case is straightforward on a logical ground but notationally more demanding. Either the individual is treated ($D = 1$) or is not treated ($D = 0$); for each individual we observe $Y = Y_0 + D\beta$, where $\beta \equiv Y_1 - Y_0$ is the causal effect of the treatment. This effect is in general individual specific, that is the return to participation may vary depending on individual characteristics which may or may not be observable to the analyst.

In what follows, $F_A[a]$ will denote the cumulative distribution function of the random variable $A$ calculated at $A = a$. A similar notation will be exploited for conditional distributions, where for example $F_{A|B}[a|b]$ will denote the conditional cumulative distribution of $A$ given $B = b$ calculated at $A = a$. Finally, $q_A[\tau]$ will denote the $\tau$-th quantile of $A$, and $q_{A|B}[\tau|b]$ the $\tau$-th quantile of the conditional distribution of $A$ given $B = b$.

In the next section two sets of causal parameters are introduced that can be defined from the joint and the marginal distributions of potential outcomes. In the former case, non-parametric identification without additional assumptions is precluded since potential outcomes are by their very nature never observed jointly for the same individual. In the latter case, identification crucially relies on the mechanism that assigns individuals to the treatment status.
3. Policy relevant parameters

The empirical relevance of the heterogeneity of treatment effects has been put forward by many authors (see among others Heckman et al., 1997). When the effect is heterogeneous in the population, average treatment effects provide a rather poor summary to evaluate the effectiveness of the intervention.\(^1\) One could rather consider two sets of measures instead: measures based on the distribution of the treatment effect \(\beta\) and measures which are functions of the marginal distribution of the potential outcomes.

Knowledge of the distribution of \(\beta\) allows to answer policy questions regarding, for instance, how widely treatment gains are distributed across recipients \((1 - F_{Y_1 - Y_0 | D}[0, 1])\), that is the proportion of participants for whom there is \(Y_1 - Y_0 \geq 0\) or to study the effect on recipients for specific values of the base state distribution \((F_{Y_1 - Y_0 | D, Y_0}[y_1, y_0])\). However, even when individuals are randomized into/out of the treatment identification of these parameters requires additional assumptions to retrieve the joint distribution of \(Y_0\) and \(Y_1\) (and thus the distribution of \(\beta\)) from the two marginal distributions of potential outcomes. The unrestricted set of joint distributions consistent with the marginals can be exploited to partially-identify the distribution of \(\beta\) via classical probability inequalities.\(^2\)

Measures based on the marginal distributions of \(Y_0\) and \(Y_1\) are useful to document the heterogeneity of the treatment across individuals; a collection of \(\tau\)th quantile treatment effects (QTE) for values of \(\tau \in [0, 1]\) describes how the treatment effect varies at different points of the outcome distribution. QTEs are defined as the horizontal distance between the distribution function under the effect of the treatment and the distribution function in the absence of the treatment\(^3\) (see Doksum, 1974), namely:

\[
q_{Y_1}[\tau] - q_{Y_0}[\tau].
\]

\(^1\) It is worth noting that potential outcomes can be represented through their quantile functions as:

\[
Y_0 = q_{Y_0}[U_0], \quad Y_1 = q_{Y_1}[U_1],
\]

where \(U_0\) and \(U_1\) are uniform random variables that can be referred to as ranks. Thus one could think of ranks as measures of proneness driving heterogeneity of outcomes amongst otherwise identical individuals and consequently see the QTE as a “measure of the interaction between the proneness property and the treatment” (see Doksum, 1974, pp. 274-275).

Ruling out the case of homogeneous returns to participation, the parameter in (1) is in general not informative about the treatment effect distribution.\(^4\) One exception worth

---

\(^1\) For example, one would be unlikely to assign the same weight to alternative interventions sharing the same mean outcome, some of which producing favorable outcomes for only a few participants while some other distributing gains more broadly.

\(^2\) However, the resulting identification set is generally far too large and the assumptions required to draw meaningful policy conclusions are rarely plausible from the economic point of view (see the discussion in Heckman et al., 1997). When potential outcomes are assumed conditionally independent given one or more common latent factors, the joint distribution can be retrieved from marginals through a factor analysis (see, for example, Carneiro et al., 2003).

\(^3\) Conditional versions of these parameters can be defined for subpopulations of individuals sharing the same characteristics \(W\).

\(^4\) Note, however, that knowledge of this parameter allows formal testing of a number of interesting hypothesis including the constant treatment effect hypothesis.
mentioning is the case of rank invariance (see Heckman et al., 1997), which amounts to assuming that the best in one distribution is the best in the other distribution. For example, individuals who are “highly ranked” earners without a training programme remain “highly ranked” earners after the training. Under rank invariance there is \( U_0 = U_1 = U \) in (2), and this implies that the \( \tau \)-th rank of each individual in the base state distribution corresponds to the same rank in the treatment distribution, so that the parameter in (1) is the individual treatment effect. Under rank invariance a single factor \( U \) is responsible for an individual’s ranking across treatment states. Rank invariance amounts to “transferring” all the mass at one percentile of the \( Y_0 \) distribution to a single percentile of the \( Y_1 \) distribution, implying that the potential outcomes \( Y_0 \) and \( Y_1 \) are not truly bivariate being jointly degenerate. Note however that it does not require the effect to be the same at all percentiles.

**Example.** Consider the binary treatment case where treatment status is defined by either having achieved the educational level of interest \((D = 1)\) or not \((D = 0)\). Examples include completing college compared to not doing so, or attaining any qualification compared to dropping out of high-school with no qualifications. The individual causal effect (or return) of achieving the qualification is defined as the difference between two potential outcomes: the wage if the individual were to achieve the qualification, \( Y_1 \), and the wage if the individual were not to achieve it, \( Y_0 \). The parameter that has traditionally received most attention by labor economists is the average return to a qualification for those who have obtained it (often called average treatment effect on the treated). This is the relevant parameter when the treatment is voluntary, which is the case for the achievement of (post-compulsory) educational qualifications, and is also the one needed for a cost-benefit analysis. Nonetheless, knowledge of this parameter is not conclusive about the likely effect distribution of achieving the qualification of interest, as educational achievement may have a different impact on individuals depending on their relative position in the wage distribution.

An alternative policy relevant parameter in thus the 
**distribution of returns** for those who actually obtained the qualification, that is \( F_{Y_1 - Y_0|D}[s|1] \), or the same parameter calculated at selected values of the base state wage distribution, that is \( F_{Y_1 - Y_0|D,Y_0}[s|1, y_0] \). Either parameter clearly rests upon knowledge of the joint distribution of potential outcomes and identification is therefore precluded without additional assumptions. Turning to the \( \tau \)-th QTE does not ease identification, as education is voluntary and individuals may decide to take further education depending on their innate ability. In fact, when individuals self select into the educational qualification of interest the marginal distributions of potential outcomes are no longer identified from raw data because of an endogeneity problem. In the context of educational achievement the representation (2) provides an economic foundation for using quantiles, as \( q_{y_1,a} \) can be interpreted as the earning function describing how an individual having education \( d \) and ability \( a \) is rewarded in the labor market (assuming that wages are increasing in ability).

---

\(^{(5)}\) Note that conditioning on covariates may help achieve this condition.
4. Identification

When assignment to the treatment group is through randomization, the marginal distributions of potential outcomes can be identified trivially exploiting the observed outcome distribution of treated individuals and non-treated individuals, respectively, and point identification of quantile treatment effect follows, provided that some regularity conditions are met. In what follows we shall therefore focus on the identification of QTEs when the treatment status is endogenous.\footnote{Semi-parametric estimation of quantile treatment effects under the conditional independence assumption is discussed in Firpo (2007).}

When individuals select into the treatment $D$ on the basis of expected gains, identification of QTEs becomes more challenging since endogeneity plagues quantile regression estimates. In what follows we will review strategies based on instrumental variables to solve the identification problem. For the purpose of the following discussion, let the covariates $W$ be split into two blocks $W = [X | Z]$, where $X$ collects a set of exogenous regressors and $Z$ is an instrumental variable for the treatment status $D$.

We distinguish between two groups of papers providing local and global identification results. Abadie et al. (2002) recover a causal parameter that refers to the sub-population of individuals whose treatment status is affected by values of the instrument (local identification). Chernozhucov and Hansen (2005) provide a global identification approach that retrieves parameters referring to the whole population.

Abadie et al. (2002) deal with the case where both $Z$ and $D$ are binary, extending the framework originally developed by Imbens and Angrist (1994) and proposing an approach that delivers causal parameters for the sub-population of compliers (see also Abadie, 2003). The intuition behind their result can be found in previous work by Imbens and Rubin (1997), where the identification of marginal distributions of the potential outcomes for compliers is provided in the same set up (thus implying identification of any functional of these distributions).

In what follows we will focus on the approach suggested by Chernozhucov and Hansen (2005), who propose a model where a single latent factor describes the effect heterogeneity as in (2).\footnote{Another approach worth mentioning is the causal chain model proposed by Chesher (2003), who describes effect heterogeneity using as many (mutually orthogonal) latent factors as endogenous regressors and requires joint (local) independence conditions. The approach is applicable when $Y$, $D$ and $Z$ are continuous random variables though it can be extended to the case where the outcome, treatment and instrument are discrete but with less identifying power.} The approach is applicable when the outcome variable $Y$ is continuous, while $D$ and $Z$ can be either continuous or discrete random variables. Their model implies that the conditional quantiles $q_{Y|X}[\tau|x]$ can be retrieved from the conditional quantiles of the observed outcome $Y$ given $X$ and $Z$. Their key assumptions are as follows: (i) $q_{Y|X}[\tau|x]$ is strictly increasing in $\tau$, thus ruling out non-continuous outcome variables; (ii) conditional on $X$, the rank variables $U_d$ are independent of $Z$; (iii) conditional on $X$ and $Z$, $D$ is determined by a random component $V$ whose correlation with the $U_d$’s drives endogeneity; (iv) conditional on $X$, $Z$ and $V$, the $U_d$’s are identically distributed. The latter condition is referred to as rank similarity, which essentially weakens rank invariance to allow for non-systematic differences in ranks between potential outcomes. In words, rank similarity implies that rank invariance holds allowing for “slippages” in one’s rank that reflect some random variation. Since rank
invariance \((U_0 = U_1)\) trivially implies rank similarity, identification holds true also in the former case. Finally note that the conditions required are conditional on a set of observable characteristics \(X\).\(^{(8)}\)

5. Estimation

The crucial result by Chernozhucov and Hansen (2005; see Theorem 1) can equivalently be stated by equating to zero the \(\tau\)-th quantile of the random variable \(Y - q_{Y_D|X}\):

\[
Pr[Y - q_{Y_D|X}[\tau|x] \leq 0|X, Z] = \tau.
\]

The latter formulation suggests an estimation procedure in two steps: first, compute the conditional quantiles of the random variable \(Y - q_{Y_D|X}[\tau|x]\) given \(X\) and \(Z\); then, choose as estimate of \(q_{Y_D}[\tau|X]\) the one that minimizes the absolute value of the coefficient associated with \(Z\) in the first step. Note that this procedure requires an estimate of \(q_{Y_D}[\tau|X]\) in the first stage. Chernozhucov and Hansen (2006) consider linear quantile regression models \(q_{Y_D}[\tau|X] = \alpha D + \beta X\) and suggest to take a grid over \(\alpha\) to compute \(Y - q_{Y_D}[\tau|X]\) in the first step. The grid should be centered around the two stage quantile regression estimates, that is the estimate of \(\alpha\) in the quantile regression of \(Y\) on \(\hat{D}\) and \(X\), where \(\hat{D} \equiv E[D|Z]\). This is the estimation procedure that will be exploited in the next section.

6. Empirical application

The aim of this section is to provide an application of the methods discussed above. To this end, we will estimate how the returns to education vary over the wage distribution in Italy using data from the 8\(^{th}\) wave (2001) of the European Community Household Panel Survey (ECHP). Junior high school (\(scuola media\)) became compulsory in Italy only after 1963, thus increasing compulsory schooling from 5 to 8 years. Compliance with the reform was not instantaneous: only in 1976 the proportion of children attending junior high school approached 100\%.

We will estimate the returns to having at least \(scuola media\) vis-à-vis dropping out of school before. Data consist of a random sample \(\{y_i, d_i, w_i\}_{i=1}^{n}\) from \((Y, D, W)\), where \(Y\) denotes log hourly wages (in 2000 prices) and \(D\) an indicator for having completed \(scuola media\).\(^{(9)}\) In our application, the variables \(X\) include a quadratic polynomial in age and regional dummies. Moreover, following Brandolini and Cipollone (2002) we set \(Z\) to be a dummy variable for individuals born after 1949.\(^{(10)}\) The sample is limited to males

\(^{(8)}\) For more details see Chernozhucov and Hansen (2005), Conditions A1-A5. The approach suggested by Abadie et al. (2002) identifies the set of QTE’s for the compliers by assuming, in addition to relatively standard conditions, that \((Z, V, U_0, U_1)\) are mutually independent and that monotonicity as defined by Imbens and Angrist (1994) is met. Moreover, their strategy refers to the case in which both the instrument and the endogenous variable are binary.

\(^{(9)}\) The general evaluation setup can thus be described as a fuzzy discontinuity design: because of non-compliance, the probability of having \(D = 1\) is not one for all cohorts of individuals actually affected by the reform.

\(^{(10)}\) According to the new law (in force since October 1st, 1963), individuals should attend school at least until junior high school (\(scuola media\)) graduation but individuals who had been in school for at least 8
We start by showing that a standard approach provides an estimate of the average return to *scuola media* which is comparable to results from other studies. The estimated causal effect of having at least *scuola media* on wages using OLS is 38.8% with a standard error of 3.7%. Using IV we have 57% with a standard error of 24.4%, resulting from a first stage regression in which the effect of $Z$ on completing *scuola media* is highly significant and equal to 23%. Point estimates of the mean return are in line with those by Brandolini and Cipollone (2002), implying returns to one additional year of education at around 19%. In Figure 1 we report estimates as well as 95% confidence bands of the QTE at different values of $\tau$ obtained using ordinary regression quantile estimates (thus not taking endogeneity into account) and the procedure described in Section 5. It is clear that endogeneity affects the curvature of the QTE, which increases monotonically with ability in the former case and is u-shaped in the latter case. Take the difference in the QTEs at different quantiles of the wage distribution as a measure of conditional wage inequality induced by marginal increases in education: a profile of increasing QTEs implies that additional schooling would increase conditional wage inequality; an u-shaped pattern would indicate instead that additional schooling might reduce conditional wage inequality at the bottom and at the top of the wage distribution. These scenarios have different implications for the choice of the optimal public education policy. However, the low precision of the IV estimates does not allow to draw decisive conclusions.\(^{(11)}\)

### 7. Conclusions

In this paper we have reviewed the criticism made in the literature about drawing causal conclusions on the effectiveness of policy interventions considering only the mean returns

---

\(^{(11)}\) We also estimated the QTE for the *compliers* using the procedure described in Abadie et al. (2002). The resulting profile for the QTE overlaps with the solid line in Figure 1 after the 25\(^{th}\) percentile, but stays parallel to ordinary regression quantile estimates at the bottom end of the ability distribution. Note however that the *compliers* represent only 23% of the entire population (this follows directly from the first stage regression results discussed above).
to participation. A growing body of the literature has addressed the identification of aspects of the causal effect distribution different from the mean. We have shown that knowledge of features of the distribution of causal effects has implications for policy and equity considerations. If the returns to participation are heterogeneous across individuals, the conventional approach to evaluation is often unattractive in this respect.

We have focused on the identification and estimation of the quantile treatment effect, that is on the difference in quantiles of the distribution of potential outcomes under the “treatment” and the “no-treatment” states that is attributable to the intervention, when individuals self-select into the treatment, thus embedding identification within an endogeneity problem. A collection of these parameters at different quantiles allows formal testing of a number of interesting hypothesis of the effect distribution, including the constant treatment effect hypothesis. Finally, we have discussed identification and estimation issues related to quantile treatment effects in the likely scenario in which individuals self-select into participation.

References


