The Impact of the 2005 CAP First Pillar Reform as a Multivalued Treatment Effect

Alternative Estimation Approaches

Roberto Esposti

Department of Economics and Social Sciences – Università Politecnica delle Marche
Piazzale Martelli 8 – 60121 Ancona (Italy)
e-mail: r.esposti at univpm.it; phone: +39 071 2207119

Paper prepared for presentation at the EAAE 2014 Congress
'Agri-Food and Rural Innovations for Healthier Societies'

August 26 to 29, 2014
Ljubljana, Slovenia

The Impact of the 2005 CAP First Pillar Reform as a Multivalued Treatment Effect

Alternative Estimation Approaches

Abstract

This paper aims at evaluating the impact of the 2003/2005 CAP reform on farm production choices. The outcome of "market orientation" is measured by considering both the short-term production choices and the long-term investment decisions. The Treatment Effect (TE) is estimated through recent alternative multiple/continuous TEs estimators based on the Generalized Propensity Score (GPS). Instead of looking at non-treated counterfactuals these approaches take advantage of the different intensity with which the first pillar support is delivered to treated units. These alternative estimators are implemented and their statistical robustness assessed and results compared. Results show that the 2003/2005 reform of the first pillar of the CAP actually had an impact more in (ri)orienting short-term farm production choices then investment decisions and this effect is significantly more evident for farms with a limited contribution of the CAP on their own Gross Production Value.

Keywords: *Treatment Effects, Common Agricultural Policy, Farm Production Choices, Matching*
JEL classification: *Q18, Q12, C21, O13*

1. Introduction: Objectives of the paper

In the last decade the empirical research on policy evaluation has paid increasing attention to the so-called Treatment-Effects (TE) literature (Imbens and Wooldridge, 2009). A rich toolkit allowing the estimation of these TEs under different and complex circumstances has progressively taken form. Nonetheless, though equipped with this powerful toolkit, practitioners often encounters serious problems in adapting it to the peculiar real-world circumstances into which policy measures have been actually implemented. As a consequence, the recent empirical literature has also focused on assessing the robustness of the estimated policy impacts to identify those that are strongly dependent on the specific limitations of the adopted methods with respect to the real context under study (Chabé-Ferret, 2010).

In the case of the quantitative evaluation of Common Agricultural Policy (CAP) measures and reforms, these two apparently contrasting tendencies of an increasingly powerful toolkit and the need of robustness of results clearly emerge. On the one hand, the growing interest in such approaches is evident for the second pillar of the CAP (the Rural Development Policy, RDP) whose measures seem particularly suitable for these empirical methodologies (European Commission, 2006; EENRD, 2010; Pufahl and Weiss, 2009; Lukesch and Schuh, 2010; Salvioni and Sciulli, 2011; Michalek, 2012; Chabé-Ferret and Subervie, 2013). On the other hand, it is widely agreed that the same does not hold true for the first pillar of the CAP (still the largest part of CAP budget) as the way it is designed and delivered makes these methods not particularly helpful, or simply useless, to achieve a proper impact evaluation (Esposti, 2014).

The objective of this paper is to critically re-consider the apparent infeasibility of the TE econometrics' toolkit in the case of the First Pillar of the CAP by pursuing the evaluation of the impact of its 2005 Reform (FPR henceforth).[1] In this respect, the paper adopts an alternative empirical strategy to indentify and estimate the TE compared to that usually followed in the evaluation of second pillar's measures. This strategy does not look for non-treated units (or *counterfactuals*)[2] but takes advantage of the

[1] Though approved in 2003, the FPR implementation actually started in 2005. In addition, for some productions (cotton, hop, olive oil, tobacco, sugar, fruits and vegetables, wine) the decoupling of the support formerly delivered through the individual OCMs was actually approved and accomplished in the following years (2004-2008) (OECD, 2011, pp.64-65). For the sake of simplicity, we refer here to all these reform steps as the "2005 First Pillar Reform" or FPR (see also section 4).

[2] Within observational datasets counterfactuals are analogous to control units of randomized experiments. However, aassessing treatment effects within experimental or quasi-experimental situations presents substantially different methodological issues and solutions (Duflo et al., 2006).

multivalued nature of the first pillar support. It is based on the concept of Generalized Propensity Score Matching (GPSM) where matching is performed among farms with different treatment intensity. This seems particularly appropriate for the first pillar of the CAP where most farms receive the treatment (the payment) but its intensity largely varies. In recent years, several alternative estimators have been developed for the case of multivalued or continuous treatments[3] (Hirano and Imbens, 2004; Bia and Mattei, 2008; Cattaneo, 2010; Cattaneo et al., 2013). They will be implemented here to compare results and assess their robustness.
This approach is here applied to a balanced panel of Italian FADN farms observed over years 2003-2007. Therefore, these farms are observed before and after the FPR, whose implementation, in Italy, actually started in 2005, and the period under analysis is symmetric with respect to the treatment.

2. THE CONCEPTUAL BACKGROUND: WHY EVALUATING THE IMPACT OF THE FPR IS DIFFICULT?

In 2003 the first pillar of the CAP underwent what can be considered the most radical reform of its half-century history; the so-called Fischler Reform (Sorrentino et al., 2011). With decoupling the FPR substantially changed the way support is delivered to farms. One of the purposes of decoupling was (ri)orientation to market, that is, to eliminate the distorting effect on production decisions of the precedent coupled support and leave the farmers free to produce what they consider more profitable in the market. Therefore, an *ex post* evaluation of the effectiveness of the FPR should assess whether and to what extent the decoupling of first pillar's support really oriented farmers to market. Despite the wide empirical literature produced on the impact of the FPR at both the farm and several aggregate levels (OECD, 2011; Sorrentino et al., 2011), such evaluation has never been performed using the rich toolkit of TE econometrics, mainly because it is considered unsuited, or of too difficult adaptation, to this case.
One possible limitation of the FPR when considered as a "treatment" is that it should be associated to a clearly recognized *ex-ante* target, i.e. an outcome variable with respect to which the treatment/policy impact can be evaluated. Although we may agree that the main objective of the FPR was to reorient farms' choices towards the market, it remains questionable whether and how this objective can be univocally expressed by a single variable, a target/outcome variable, that eventually expresses the effect of the reform. Such outcome variable should measure to what extent farms have changed their production orientation (e.g., their production mix) after the FPR, but computing such variable isn't easy for production units that naturally are multi-output and, therefore, systematically change their production mix on a yearly basis.
More importantly, a second limitation really represents the key limitation that apparently prevents the application of TE methods to the evaluation of the FPR. To apply the TE logic counterfactual observations must exist; that is, observations where the outcome variable(s) is (are) observed without the treatment (*counterfactuals*). In the case of the FPR, however, finding a proper strategy to identify counterfactuals and compare them to treated units represents an often unsolved research challenge. On the one hand, the non-treated units (that is, farms receiving no support under the first pillar of the CAP and, therefore, not involved by the FPR) are relatively few compared to the treated ones and this represents a problem for a proper matching. On the other hand, they are not treated just because of their peculiar production and managerial choices. Even though a non-treated sample can be observed, it can be hardly considered a proper counterfactual sample because of this peculiarity, that is, those unobserved characteristics that affect, at the same time, the outcome and the treatment assignment. In the TE jargon, the specificity of the FPR as a treatment makes almost impossible to get rid of the *selection-on-unobservables bias*.
One final issue is that the treatment under study must be clearly identified and observable. Apparently, the FPR itself has substantially increased the clarity and the identifiability of the treatment associated to the first pillar of the CAP. Many and heterogeneous coupled payments (associated to about 30 Common Market Organizations, CMOs) have been transformed into a unique and decoupled Single Farm Payment (SFP) directly delivered to the farmer (European Commission, 2011; OECD, 2011). This makes the FPR treatment easily measurable as the amount of money that, at the farm level, have been

[3] Some texts refer to this case as "multilevel treatment" while with "multivalued treatment" it is designated the case of more than one treatment simultaneously administered to the same units (StataCorp, 2013). Here, following the prevailing terminology in the TE literature (Imbens and Wooldridge, 2009; Cattaneo, 2010) the former case is called "multivalued treatment", the latter "multiple treatment".

transformed from several coupled payments into a decoupled SFP. However, still a complication arises in considering the FPR as a "treatment" because, unlike other typical policy program evaluations (e.g., job training programs), the treatment associated to the FPR is not a binary treatment, (i.e., expressed as a dichotomous variable: 1=treated, 0=not treated). The FPR is a continuous treatment (*multivalued treatment*; Cattaneo, 2010) where only few units (farms) are not treated, while most units are treated but with different intensity. Therefore, in the case of the FPR the TE reasonably depends on the treatment intensity and not only on being treated or not. Nonetheless, the multivalued nature of the treatment, rather than being a limitation, actually represents the key opportunity to identify and estimate the TE of the FPR.

3. THE METHODOLOGICAL APPROACH: MULTIVALUED TREATMENT AND THE GENERALIZED PROPENSITY SCORE MATCHING (GPS)

Consider a sample of *N* observations (farms). Let Y_i indicate the outcome variable observed in the generic i-th farm (unit), $i = 1,..., N$ (where *N* is the sample size) and $T=0,1$ the binary policy treatment (T=0 if not treated, T=1 if treated). Let's assume that the attribution of a given treatment to the i-th farm does not affect the TE on the j-th farm, $\forall j \neq i = 0,...,N$. This assumption is called *stable-unit-treatment-value assumption (SUTVA)*. It seems plausible whenever micro data are used and the treatment assignment to single units may hardly have aggregate (or macro) effects (e.g., on partial and general equilibrium market adjustments) (OECD, 2011). In the present case, however, the SUTVA also implies the absence of diffusion or spillover effects like, for instance, imitation.[4]

By *Average Treatment Effect* (ATE) we intend the following expected value:[5]

(1) $$ATE_i = E(Y_{i1} - Y_{i0}),\ \forall i = 1,..., N$$

ATE expresses the difference that would be observed in the outcome in a purely experimental (or randomization) situation, that is, if the i-th farm were observed, in sequence, under treatment and non-treatment. In practice, with observational (or non-experimental) data, we really observe only the outcome under one of the possible states. The outcome in all other cases is, in fact, hypothetical or *potential* (Rubin, 1974).

In the case of binary treatment matching has become a popular approach to estimate ATE on treated units (or *Average Treatment Effect on the Treated*, ATT) (Imbens and Wooldridge, 2009). But, as mentioned, matching requires counterfactuals (non-treated units) to be compared with the treated units under the assumption that, once we control for all relevant and observable confounding factors (or covariates) **X**, the different outcome between the two groups only depend on the treatment. Such assumption is known as *Conditional Independence Assumption* (CIA) or *Unconfoundedness Assumption*. Vector **X** is expected to contain all the pre-treatment variables that affect, at the same time, the treatment assignment and to the outcome variable. Basically, it is the validity of the CIA that is seriously questioned when we apply conventional matching TE estimation to the FPR evaluation.

As shown in Esposti (2014), applying binary TE estimation to the FPR impact evaluation may raise severe objections since matching estimation can only ensure against the *selection-on-observables bias* but, evidently, can do nothing against the *selection-on-unobservables bias*. Evidently, these variables being unobservable, the presence of the bias they generate can not be tested: it is an hidden bias. At the same time, however, if there are unobserved variables that affect assignment to treatment and the outcome variable simultaneously, the respective hidden bias would make matching estimators not robust. As a consequence, checking the robustness and sensitivity of the matching estimates has become

[4] As the FPR concerns farms' market orientation, excluding such spillover effects of the treatment may seem a relevant assumption. Nonetheless, it is still hardly testable. Chabé-Ferret and Subervie (2013) actually suggest that a test on this assumption's validity can be still attempted by looking at neighboring farmers' outcome variable dynamics before and after the treatment. Such kind of statistical test, however, would imply making the spatial dimension explicit within the adopted panel sample, that is, to introduce spatial econometrics techniques. This solution seems computationally demanding when micro data are used (6542 units in the present case) and is here ignored. However, it may represent an interesting direction for future research.

[5] If *i* indexes a randomly drawn unit in the population we can also write $ATE = ATE_i$, where ATE is also called *Population Average Treatment Effect* (PATE). In the sample, ATE is calculated averaging ATE_i across the sample units and it is also called *Sample Average Treatment Effect* (SATE) (Abadie et al., 2004). In the present application, whenever the ATE has to be intended as SATE, for simplicity, we drop the *i* index in the notation.

an increasingly important topic in the applied evaluation literature as a sort of indirect evidence of the presence of this hidden bias. Beside any helpful sensitivity analysis, it remains true that the major drawback of the matching methods consist in the possible and not testable presence of hidden biases caused by unobservable covariates. In particular, in the case of the FPR the treatment is applied to a large majority of farms, therefore it is a non-selective generalized policy. Farms not involved by this policy are very peculiar and, therefore, represent a self-selected exception, thus apparently not suitable counterfactuals in matching estimation.

One possible alternative consists in the *Difference-In-Difference* (*DID*) estimation approach (Smith and Todd, 2005) where the *selection-on-unobservables bias* is apparently circumvented by considering, as suitable counterfactuals, the same treated units observed before the treatment. After all, any treated unit can be compared with itself observed before the treatment (the FPR). In fact, the DID approach (conditional or not) still needs a control group of farms, that is a sample of non-treated farms to perform the ATT estimation. Therefore, it does not really solve the basic problem underlying matching estimation in the present case. Existing studies show that the application of the DID estimation to the FPR case shows little robustness across different before and after-treatment periods (Esposti, 2014). This may indicate that time strongly affects the outcome variables here considered and this effect of time is evidently not controlled by the set of covariates and differs between treated and non-treated units. In practice, this lack of robustness could be interpreted as an evidence against the validity, in the present case, of the *Conditional mean-Independence of Increments Assumption* (CIIA) (i.e., the key assumption underlying this estimation approach) due to the year-by-year strongly unpredictable and highly differentiated (across crops, territories, types of farm and farming) variations in market and environmental (e.g., weather) conditions. In addition, as emphasized by Chabé-Ferret and Subervie (2013), the apparent violation of the CIIA can be rather attributed to the elusive timing of the response to the treatment, that is, to the presence of anticipation effects (potentially occurring before the treatment) or of lagged effects (potentially occurring after the treatment).

Despite these challenging issues, however, an alternative approach to these binary-treatment estimators can be adopted any time the treatment is not binary ($T = 0,1$), but behaves as a multivalued (either discrete or continuous) variable ($T \in \mathbf{N}\,\text{or}\,\mathbf{R}^+$) and the response of the outcome variable (Y) to the treatment is itself continuous. In such circumstance, the intensity of the treatment can be correlated to the magnitude of this response and this allows the identification and estimation of the TE without using the non-treated units. In fact, these latter are no more needed to observe how the Y varies with T|X. Any treated unit can behave as a counterfactual case for units with a different treatment intensity.

The case of the FPR seems particularly suitable for a multivalued treatment approach. In Italy the SFP has been established on a historical basis (Povellato and Velazquez, 2005; Frascarelli, 2008), therefore the amount of support shifting from coupled to decoupled payment varies across treated units (farms) and this treatment clearly behaves as a continuous variable. We can here reasonably assume that for a farm with given characteristics (that is, conditional on **X**) the larger the coupled (thus constraining the production choices) support then converted into the SFP, the larger the expected change in the production mix (market reorientation). Nonetheless, whenever we have a multivalued treatment the critical issue shifts from finding appropriate counterfactuals to properly define the functional relationship between Y and $T|\mathbf{X}$.

3.1. The Hirano-Imbens approach: GPS and the Dose-Response Function (DRF)

The approach that follows this intuition has been originally proposed by Hirano and Imbens (2004) and it is based on the concept of Generalized Propensity Score (GPS). In a broad sense, it can be considered a generalization of the conventional binary-case matching estimation based on the Propensity Score (PSM) (Becker and Ichino, 2002). The Hirano-Imbens approach can be described as sequence of three steps.

For any treated unit $i = 1, \ldots, N$, we observe the covariates $\mathbf{X}_i$, the treatment level T_i, the outcome variable Y_i. We define, $\forall i$, a set of *potential outcomes* $\{Y_i(T)\}_{T\in\Xi}$ where Ξ is the set of potential treatment levels and $Y_i(T)$ is a random variable that maps, for the i-th unit, a particular potential treatment, T, to a potential outcome. Evidently, of these potential outcomes only one is observed, that associated with the actual treatment T_i. Hirano and Imbens (2004) refer to $Y_i(T)$ as the *unit-level Dose-*

Response Function (*uDRF*). In fact, we are interested in the *average Dose-Response Function* (*aDRF*), $aDRF(T)=E[(Y(T)]$.

The first step of the Hirano-Imbens approach consists in the estimation of the univariate variable GPS. Hirano and Imbens (2004) define the GPS_i as the probability that the i-th unit is assigned the treatment level T given its observed characteristics $\mathbf{X}_i$: $GPS_i = r(T_i, \mathbf{X}_i)$, where $r(T, \mathbf{X})$ is the *propensity function*, i.e., the conditional density of the actual treatment given the observed covariates. For the GPS to be meaningful in the calculation of the ATE, the following condition must be satisfied within the sample: units with statistically equivalent values of **X** are expected to show, around a given interval of *GPS*, both treatment levels lower and higher than a given level *T*. This is the *balancing condition* in the continuous treatment case.[6] Hirano and Imbens (2004) demonstrates that if this condition is respected, and CIA assumed, the assignment to treatment is unconfounded, given the estimated GPS.[7] Therefore, the different Y observed across units showing the same estimated $GPS|\mathbf{X}$ can be fully attributed to the different treatment level *T*. Once the propensity function is estimated and the balancing condition is met, the second methodological step consists in estimating the conditional expectation of the potential outcome as a function of two scalar variables, the estimated *GPS* and *T*: $g(G\hat{P}S,T)=E\left[Y|G\hat{P}S,T\right]$. The third step estimates the *aDRF* as $aDRF(T)=E[(\hat{g}(T)]$, $T\in\Xi$, that is, by averaging the estimated conditional expectation $\hat{g}(G\hat{P}S,T)$ over the *GPS* at any level of the treatment we are interested in.

The estimation steps imply arbitrary specification assumptions whose validity can be only assessed *ex post* by checking for the robustness of results. The complexity of the overall estimation procedure together with this arbitrariness may explain why, though originally proposed some years ago by Hirano and Imbens (2004), this approach has been only recently applied to the evaluation of multivalued program or policies (Bia and Mattei, 2007, 2012; Flores and Mitnik, 2009; Kluve et al., 2012; Magrini et al., 2013) and never to the evaluation of the FPR.

The first arbitrary assumption implied by the method is the specification of the distribution of T_i conditional on $\mathbf{X}_i$ to compute its conditional density. The common practical implementation of the methodology, also followed here, assumes a normal distribution for the treatment given the covariates:

(2) $r(T_i,\mathbf{X}_i)=T_i|\mathbf{X}_i \sim N\left(\boldsymbol{\beta}'\overline{\mathbf{X}}_i,\sigma^2\right)$

where $\boldsymbol{\beta}$ is a vector of unknown parameters and $\mathbf{X}$ is the matrix of covariates. Therefore, the assumption is that the propensity function is linear in unknown parameters that can be thus estimated by OLS. Evidently, it is possible to assume other distributions, to adopt different (even non-parametric) specifications other than the linear regression and to estimate the GPS by other methods such as MLE.[8] In fact, while the normality assumption can be tested, the empirical specification of (2) remains arbitrary. In particular, it seems questionable here to assume a linear relationship between T and some set of conditioning variables **X**. Nonetheless, this problem can be prevented by using $\overline{\mathbf{X}}$ instead if $\mathbf{X}$, where $\overline{\mathbf{X}}$ includes transformations (e.g., polynomial terms) of $\mathbf{X}$ and/or interactions terms across variables in $\mathbf{X}$, in such a way that $\overline{\mathbf{X}}$ satisfies both the normality assumption and the balancing condition.

The estimated GPS is thus calculated as:

(3) $$G\hat{P}S_i=\frac{1}{\sqrt{\left(2\pi\hat{\sigma}^2\right)}}\exp\left\{-\frac{1}{2\hat{\sigma}^2}\left(T_i-\boldsymbol{\beta}'\overline{\mathbf{X}}_i\right)^2\right\}$$

A second and, probably, more critical arbitrary assumption concerns the specification of the *uDRF*, $g(G\hat{P}S,T)=E\left[Y|G\hat{P}S,T\right]$, that is, the conditional expectation of the potential outcome with respect to *T* and the estimated *GPS*. The often adopted specification of the conditional expectation is a fully

---

[6] For more details on the balancing condition in the binary-treatment case see also Becker and Ichino (2002) and Abadie et al. (2004).

[7] In the multivalued treatment case, Hirano and Imbens (2004) actually call the CIA *weak unconfoundedness assumption* since it only requires conditional independence to hold for each value of the treatment, rather than joint independence of all potential outcomes. Such assumption is also called (*weak*) *ignorability* (Cattaneo, 2010).

[8] See next section for an Multinomial Logit specification of (2). In any case, following Bia and Mattei (2008) and also to test for the validity of the normality assumption, a MLE instead of a OLS estimation of parameters $\boldsymbol{\beta}$ is here performed.

interacted flexible function of its two arguments providing a good approximation of the underlying unknown relationship:

$$(4)\ g\left(G\hat{P}S,T\right)=E\left[Y|G\hat{P}S,T\right]=\alpha_0+\sum_{k=1}^{K}\alpha_k(T)^k+\sum_{h=1}^{H}\gamma_h\left(G\hat{P}S\right)^h+\sum_{k=1}^{K}\sum_{h=1}^{H}\lambda_{kh}(T)^k\left(G\hat{P}S\right)^h$$

where $\alpha_0, \alpha_k, \gamma_h, \lambda_{kh}$ are unknown parameters to be estimated.[9] For each i-th unit, the observed Y_i, T_i and the estimated GPS_i are used to estimate the unknown parameters of (4) by OLS. Our empirical approach (see next sections) start with the general form (4) and then adopts the best fitting specification according to the usual Akaike Information Criterion (AIC).

The final step thus consists in using these estimated parameters to compute the average potential outcome at a given treatment level, *T*. $aDRF(T) = E[(\hat{g}(T)]$ is estimated as:

$$(5)\ aDRF(T) = E[(\hat{g}(T)] = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\alpha}_0+\sum_{k=1}^{K}\hat{\alpha}_k(T)^k+\sum_{h=1}^{H}\hat{\gamma}_h\left(G\hat{P}S_i\right)^h+\sum_{k=1}^{K}\sum_{h=1}^{H}\hat{\lambda}_{kh}(T)^k\left(G\hat{P}S_i\right)^h\right]$$

The entire *aDRF* can be thus obtained by computing this average potential outcome for each level of the treatment, i.e. $\forall T \in \Xi$. Bootstrap methods can be used to obtain standard errors of the estimated $aD\hat{R}F(T)$ taking into account the estimation of parameters in (2) and (4) (i.e., the entire estimation process is bootstrapped).[10] Eventually, the first order derivative of (5) with respect to *T* represents the ATE of the various treatment levels and as such is here estimated.

In fact, large part of the empirical problem in properly performing policy evaluation within this approach consists in the limited or null knowledge about how the policy effect may vary with different intensity of the treatment. In the specific case of the FPR, it is reasonable to argue that the decoupling of the support may have led to a change in production mix but this response only partially, and not linearly, depends on the amount of decoupled payments. There are economic and physical limitations to the extent of the farm response to the treatment. When this limit is approached and reached a further increase of the treatment intensity (i.e., *T*) may be ineffective. Therefore, in this specific case it seems reasonable to argue that a *DRF* does exist but it is expected to be monotonous and non-linear with a decreasing slope (first order derivative). Accordingly, in the case of the FPR, the ATE is expected to be positive but decreasing in *T*.

3.2. The Cattaneo Approach: IPW and EIF Estimation of the ATE

Evidently, beside the normality assumption in (2), the Hirano-Imbens GPSM approach strongly depends on the arbitrary and non-testable specifications adopted. A feasible empirical strategy to assess their reliability consists in comparing respective results with those obtained using alternative specifications or estimation strategies. Cattaneo (2010) proposes an alternative approach to the estimation of the ATE under multivalued treatment. Though it shares several points in common with the original Hirano and Imbens (2004) estimation, the method proposed by Cattaneo (2010) significantly differs in the way the treatment (and the treatment variable) enters the analysis, how functional relationships are specified and, eventually, how the ATE is estimated. Therefore, this novel approach seems particularly interesting not just to pursue an allegedly superior estimation but to investigate the robustness of the results, i.e. the estimated ATE, once these applicative variants are admitted.

In fact, both approaches are based on the estimation of the GPS and on the consequent estimation of the potential outcome given the treatment. However, Cattaneo (2010) does not estimate a parametric *DRF* but look for a non-parametric identification and estimation of some parameter (e.g., the mean) of the statistical distribution of the potential outcome. While the Hirano and Imbens (2004) approach applies to continuous treatment, Cattaneo (2010) estimation implies multiple but discrete treatment levels. Thus, continuous treatment variables must be converted in advance into a categorical variable. Though both approaches share the parametric GPS estimation, therefore, the different nature of the treatment variable (continuous vs. categorical variables) makes the specification of the propensity function differ:

[9] Hirano and Imbens (2004) emphasize that there is no direct meaning (i.e., economic interpretation) of the estimated coefficients in (2), except that testing whether all coefficients involving the GPS are equal to zero can be interpreted as a test of whether the covariates introduce any bias.

[10] As shown in Hirano and Imbens (2004), in principle, it is possible to derive the analytical calculation of the asymptotic standard errors of the estimated DRF. In practice, it may be unaffordable or highly computationally demanding.

a Multinomial Logit Model (MLM) in Cattaneo (2010) instead of a linear regression as in (2). Moreover, the Cattaneo (2010) approach is a three-step semi-parametric estimation where the first step (the parametric GPS estimation) is followed by two semi-parametric stages estimating the conditional distribution of the potential outcome given the treatment.

Cattaneo (2010) actually proposes two semiparametric-efficient estimation procedures and derives their large-sample properties: the Efficient-Influence-Function (EIF) and the Inverse-Probability Weigthed (IPW) estimators.[11] To shortly describe how these two estimations proceed, let's assume that there are J possible treatment levels $j = 0, \dots, J$; therefore, it is $T_i = j$. Also define a new indicator variable $d_i(j)$ taking either value 1 if the i-th unit has received the j-th treatment or value 0 otherwise. It is thus possible to define the J potential outcomes $Y_i(j)$, $\forall j \in J$. $F_{Y(j)}(Y)$ is the distribution function of the potential outcome $Y(j), j = 0, 1, \dots, J$, that is, the distribution of the outcome variable that would occur if individuals were administered the treatment level j. The only observed outcome variable, however, is given by $Y_i = d_i(0)Y_i(0) + d_i(1)Y_i(1) + \cdots + d_i(J)Y_i(J)$[12] where $\{Y_i(0), Y_i(1), \dots, Y_i(J)\}'$ is an independent and identically distributed draw from $\{Y(0), Y(1), \dots, Y(J)\}'$ for each i-th individual in the sample. The estimation method proposed by Cattaneo (2010) allows to estimate several parameters of the distribution function $F_{Y(j)}(Y)$. In particular, the interest here is in estimating the $Jx1$ vector of potential outcome means: $\mathrm{H}=(\mu_0, \mu_1, \dots, \mu_J)'$.[13]

Under some regularity conditions (Cattaneo 2010; Cattaneo et al., 2013), $F_{Y(j)}(Y)$ can be identified from observed data and, consequently, also its parameters in **H**. Assuming weak unconfoundedness, this identification starts from the fact that $F_{Y(j)}(Y) = E\{F_{Y(j)|\mathbf{X}}(Y|\mathbf{X})\} = E\{F_{Y(j)|\mathbf{X}}(Y|\mathbf{X}, T=j)\}$, where the latter term is identifiable from the observed data. The estimation methods proposed by Cattaneo (2010), therefore, estimate **H** exploiting the fact that the observed potential outcome distributions have been marginalized over the covariate distributions. Cattaneo (2010) proposes two Z-estimators,[14] one constructed using an inverse-probability weighting scheme and the other constructed using the full functional form of the EIF. Both estimators are shown to be consistent, asymptotically normal, and semiparametric efficient under appropriate conditions. Therefore, the two estimators are asymptotically equivalent.

The IPW estimation extends the idea of inverse-probability weighting, widely used in the case of binary treatment, to a multivalued treatment context. The estimator is based on a set of moment conditions implied by the analysis above and by the weak unconfoundedness assumption. In this sense, it can be considered a sort of GMM estimation. These moment conditions are all motivated by the fact that for each treatment level j, the following holds true:

$$(6)\quad E\left\{\frac{d(j)\,(Y-\mu_j)}{p_j(\mathbf{X})}\right\} = E\left[\frac{E\{d(j)|\mathbf{X}\}E\{Y(j)-\mu_j|\mathbf{X}\}}{p_j(\mathbf{X})}\right] = E\{Y(j)-\mu_j\} = 0$$

where $p_j(\mathbf{X})$ is the conditional probability function expressing the probability of receiving the j-th treatment conditional on $\mathbf{X}$.

From (6) it is possible to derive the following set of moment conditions based on observed data and from which the GMM estimation of **H** can be derived (those means that make all the moment restrictions hold true):

$$(7)\quad E\left[\psi_{IPW,j}\{\mathbf{z}_i; \mu_j, p_j(\mathbf{X}_i)\}\right] = 0 \;\text{ with }\; \psi_{IPW,j}\{\mathbf{z}_i; \mu_j, p_j(\mathbf{X}_i)\} = \frac{d_i(j)\,(Y_i-\mu_j)}{p_j(\mathbf{X}_i)}$$

[11] Efficiency has to be intended with respect to the ideal case of randomized treatment: in this case, these estimators are more efficient estimators than the usual parametric estimators. As the weak unconfoundedness assumption and the balancing condition mimic the randomization case, this result can be extended to the non-experimental cases as far as such conditions are valid.

[12] The *fundamental problem of causal inference* in a multivalued treatment context.

[13] Cattaneo (2010) also presents and discusses the quantiles' estimation.

[14] Z stands for "zero" and indicates those estimators based on some conditions imposing a function of the data and of the unknown parameter to be estimated to be = 0 (as (7)).

where $\mathbf{z}_i = (Y_i, T_i, \mathbf{X}_i')'$ simply represents the observed data (outcome variable, treatment level, covariates) and μ_j is the unknown element in **H** to be estimated. $p_j(\mathbf{X}_i)$ is not observed but can be estimated specifying a propensity function, as $\mathbf{p}(\mathbf{X}) = \{p_0(\mathbf{X}), p_1(\mathbf{X}), ..., p_J(\mathbf{X})\}$ simply is the GPS|**X**. Once the GPS is estimated, i.e. $\hat{\mathbf{p}}(\mathbf{X})$, and provided that the balancing condition is respected, all elements in **H** can be estimated imposing restrictions in (7):

(8) $\hat{\mu}_{IPW,j}$ s.th. $\frac{1}{N}\sum_{i=1}^{N} \psi_{IPW,j}\{\mathbf{z}_i; \hat{\mu}_{IPW,j}, \hat{p}_j(\mathbf{X}_i)\} = 0$

From (8), $\hat{\mu}_{IPW,j}$ can be also expressed in closed form as follows:

(9) $\hat{\mu}_{IPW,j} = \left\{\sum_{i=1}^{N} \frac{d_i(j)}{\hat{p}_j(\mathbf{X}_i)}\right\}^{-1} \sum_{i=1}^{N} \frac{d_i(j) Y_i}{\hat{p}_j(\mathbf{X}_i)}$

(9) clearly shows why this estimator is called IPW, as the GPS (the probability of receiving that treatment level) acts as a weighting factor of the observed outcome Y_i.

Compared to IPW, EIF estimation uses more information on the marginal potential-outcome distribution and, therefore, involves further functions to be estimated.[15] In particular, let's define the following function for each treatment level:

(10) $e_j(\mu_j; \mathbf{X}_i) = E\{Y_j(j) - \mu_j | \mathbf{X}_i\} = E\{Y_j(j) - \mu_j | \mathbf{X}_i, T_i = j\}$

The EIF estimation is then obtained by imposing the following set of moment conditions:

(11) $E[\psi_{EIF,j}\{\mathbf{z}_i; \mu_j, p_j(\mathbf{X}_i), e_j(\mu_j; \mathbf{X}_i)\}] = 0$

with $\psi_{EIF,j}\{\mathbf{z}_i; \mu_j, p_j(\mathbf{X}_i), e_j(\mu_j; \mathbf{X}_i)\} = \frac{d_i(j)(Y_i - \mu_j)}{p_j(\mathbf{X}_i)} - \frac{e_j(\mu_j; \mathbf{X}_i)}{p_j(\mathbf{X}_i)}\{d_i(j) - p_j(\mathbf{X}_i)\}$.

As in the case of the IPW estimator, the EIF estimator uses these moment conditions, replacing expectations by sample averages and unknown functions by appropriate (parametric or nonparametric) estimators. Beside the GPS estimation, $p_j(\mathbf{X}_i)$, (11) also requires the estimation of $e_j(\mu_j; \mathbf{X}_i)$. This leads to the following estimates:

(12) $\hat{\mu}_{EIF,j}$ s.th. $\frac{1}{N}\sum_{i=1}^{N} \psi_{EIF,j}\{\mathbf{z}_i; \hat{\mu}_{EIF,j}, \hat{p}_j(\mathbf{X}_i), \hat{e}_j(\hat{\mu}_{EIF,j}, \mathbf{X}_i)\} = 0$

that can be expressed in closed form as follows:

(13) $\hat{\mu}_{EIF,j} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{d_i(j) Y_i}{\hat{p}_j(\mathbf{X}_i)} - \left\{\frac{d_i(j)}{\hat{p}_j(\mathbf{X}_i)} - 1\right\}\hat{Y}_i(j)\right]$

where $\hat{Y}_i(j)$ is the outcome variable predicted value obtained regressing Y_i on $\mathbf{X}_i$ for those observations with $d_i(j) = 1$.

Though asymptotically equivalent, there may be reasons to prefer one of the two estimators. The EIF is expected to be more robust in finite samples as it enjoys the so-called double-robust property when viewed from a (flexible) parametric implementation perspective, while the IPW estimator does not have this property. On the contrary, however, the IPW estimation could be preferred to the EIF case because of its simplicity though, in fact, if also the $\hat{\mu}_j$ variance-covariance matrix has to be estimated (and this is evidently needed for inference purposes) the IPW estimator requires implementing the same ingredients of the EIF estimator (Cattaneo, 2010; Cattaneo et al., 2013).

Therefore, in practice, the implementation of both estimators is very similar and requires the same three steps. First of all, $p_j(\mathbf{X}_i)$ must be estimated. As mentioned, this corresponds to the GPS estimation of Hirano and Imbens (2004), the only difference being that we have here a categorical dependent variable. Therefore, $p_j(\mathbf{X}_i)$ is here estimated with a nonlinear Multinomial Logit (ML) estimation

[15] In statistics, the *influence function* is a measure of the dependence of the estimator on the value of one sample point. The EIF (Efficient Influence Function) is a concept adopted in the estimation of semiparametric models, that is, models combining a parametric form for some component of the data generating process with weak nonparametric restrictions on the remainder of the model.

approach. Nonlinearity comes from the fact that as independent variables of the ML model we use $\overline{\mathbf{X}}_i$, that is, a flexible fully interacted polynomial of the covariates. This ML model is estimated via MLE. The second step requires the estimation $e_j(\mu_j, \mathbf{X}_i)$. Even in this case, estimation is obtained by specifying a flexible fully interacted polynomial (i.e., a parametric function) then estimated with OLS. In both cases, $p_j(\mathbf{X}_i)$ and $e_j(\mu_j, \mathbf{X}_i)$, the actual adopted specification is selected among all possible forms following the usual AIC.[16]

The final step consists in computing the ATE that can be easily obtained as:

(14) $ATE_{IPW,j} = (\hat{\mu}_{IPW,j} - \hat{\mu}_{IPW,j-1})$ and $ATE_{EIF,j} = (\hat{\mu}_{EIF,j} - \hat{\mu}_{EIF,j-1})$, $\forall j = 1,...,J$

To perform inference on these estimated ATEs the estimated standard errors of $\hat{\mu}_{IPW,j}$ and $\hat{\mu}_{EIF,j}$ are obtained through bootstrapping of the whole estimation process.

4. THE SAMPLE, THE TREATMENT AND THE OUTCOME VARIABLE

A suitable sample to apply the estimation methods discussed in the previous section must be observed over a period including both pre and post-treatment observations (years). In other words, the sample must be a balanced panel and must contain all the needed information about the outcome variables, the treatments and the confounding variables (covariates $\mathbf{X}$). As the objective here is to assess the impact of the FPR on (Italian) farm's production choices, these conditions are met by extracting a constant sample of Italian Farm Accountancy Data Network (FADN/RICA) farms yearly observed over a period including the pre and post-FPR years. The numerosity of the FADN database allows for a quite large balanced panel.

For the selection of the time period covered by this panel, the choice is here made to avoid years that are progressively far from the moment of the treatment (2005). Moreover, it seems appropriate to select a period of analysis that is symmetric with respect to the treatment (FPR) and contains most, if not all, of its effects while excluding other possibly overlapping effects due to other policy treatments (or other confounding factors) and that could occur before 2003 and after 2007. For instance, adding years 2000-2003 in the pre-treatment period can be troublesome as they may still incorporate some effects of the previous CAP reform (Agenda 2000) (Esposti, 2007). At the same time, the post-2007 years could raise the same kind of problem due the implementation of the so-called Health Check Reform (HC) (Esposti, 2011). In addition, the considerable price turbulence observed in agricultural markets in years 2008, 2009 and 2010 (Esposti and Listorti, 2013), then accompanied by the negative effects on agriculture of the general economic crisis (De Filippis and Romano, 2010), suggests particular caution in adding these years to the post-treatment period.

For this reason, the balanced sample is here limited to the 2003-2007 interval. 6542 farms are observed over these five years. This balanced panel constitutes the sample on which the present analysis is performed. It is worth noticing that the FADN sample is not fully representative of the whole national agriculture. The reference population from which the FADN sample is ideally drawn, in fact, excludes a significant (at least in terms of numerosity) amount of Italian farms (those with Economic Size<4 ESU, that is, less than 4800 Euro of Standard Goss Margin).[17] In this respect, the FADN sample is only representative of a sub-population of Italian farms, those farms that can be here refereed as *professional or commercial farms* (Cagliero et al., 2010; Sotte, 2006). Nonetheless, these 6542 farms are quite homogeneously distributed across the national territory, and the scattering of farms across the Italian macro-regions (North-West, North-East, Centre, South and Islands) well represents the pretty diverse agricultural conditions and structures of these different parts of the country.

---

[16] It may seem now more clear why the IPW and EIF estimations of the ATE are considered a semiparametric estimation compared to the fully parametric estimation of Hirano-Imbens. While the underlying estimation theory is grounded in nonparametric estimation, in practice, they still imply (though not strictly necessary) some parametric specifications.

[17] According to 2000 Census data, more than 82% of Italian agricultural holdings had an economic size smaller than 8 ESU but they accounted for just 27% of total Italian agricultural area (Sotte, 2006). According to 2010 Census data, about 67% of Italian agricultural holdings has an economic size smaller than 18 ESU but they account for just 17% of total Italian agricultural area (Sotte and Arzeni, 2013).

The treatment is the FPR, that is, the change of first pillar support from coupled to decoupled payments. Evidently, the participation to the treatment is not voluntary. It depends, in Italy, on the history of the individual farm and on the respective support it received in the 2000-2002 period; farms can not decide to remain in the old regime, as this is not admitted. Therefore, the treated units ($T = 1$) are those farms that received the first pillar CAP support in the form of coupled payment before the reform and then in the form of SFP. According to such interpretation, the treatment consists in the change of the form of support and not in its amount as in Italy the conversion from coupled to decoupled payments has been defined on a purely historical basis. It is still possible to find farms that did not receive any CAP support in the old regime (for whatever reason mostly due to peculiar production and managerial choices). For them, the change in regime did not occur simply because they remain in a no-policy situation both before and after 2005, and no CAP first pillar support has been received over the entire 2003-2007 period. Therefore, they are not treated (control) units ($T = 0$) simply because they experience no change in the form of support.[18]

Among treated farms, the treatment level is here measured in terms of *treatment intensity* (*TI*). To take into account the different economic size of farms, *TI* is computed as the amount of first pillar support received by a given farm divided by its Gross Production Value (GPV). For both values, the yearly average over the whole 2003-2007 period is computed. Here, the multivalued treatment case only considers the treated units while non-treated units are excluded from the analysis under the assumption that, as already discussed in the case of the FPR, the selection-on-unobservables bias can not be ruled out. Therefore, *TI* is here always greater than 0.[19]

The distribution of the *TI* within the sample is displayed in Figure 1. If we exclude the 1124 (17%) non-treated units (for which $T = 0$), the *TI* across the 5430 treated units (for which $T = 1$) tends to concentrate in the [0-10%[interval (52% of treated units); the mode is, in fact, around 5%. For only 8% of treated units *TI* is equal or greater than 30%, while for only 1% *TI* is equal or greater than 50%. The maximum *TI* observed within the sample amounts to about 72%. Therefore, the continuous variable *TI* has a truncated distribution starting at 0, finding a peak at about 5% and then regularly declining up the maximum value. It is clearly a non-normal distribution. To restore normality, therefore, the *TI* variable must be properly transformed.[20]

At this micro-level, the FPR is expected to affect production decisions by (ri)orienting farmers' to market. The hypothesis is that decoupling leaves farmers free to adjust and reorient their production decisions given their individual characteristics and market conditions (i.e., prices).[21] Therefore, a proper outcome variable should express the degree of change in production orientation or mix. Finding a synthetic variable expressing such change in farm production choices, however, is not trivial. Within typically multioutput activities, production decisions are expressed by an output vector rather than by a scalar variable. For any element of this vector, the change in production choices can take a different form: to start producing a new (for the farm) agricultural product but also (in the case of a product that is already part of the farm's supply) to increase or reduce the amount of production of that particular good, to improve or not its quality level and so on. Moreover, whatever this change eventually is, its timing may be different. The introduction of a new perennial crop in the farm output vector (for instance, wine production) implies a long-term horizon; in such case, what we observe in the short-term, is just an investment decision. On the contrary, the introduction (or a larger production) of an annual crop (for instance, durum wheat) operates in a short-term horizon and can be directly observed in terms of higher cultivated area or higher revenue or higher input expenditure related to that specific crop.

To take this multiple nature of the farm-level production response to FPR into account, different outcome variables are considered. We can divide them in two typologies. The first type of outcome variable is a synthetic (scalar) measure of the change in the supply vector (that is, in the shorter-term

[18] This panel excludes farms that are strongly specialized in crops whose CMOs have been reformed only starting in 2008. Therefore, farms of typologies with an high revenue share of vegetable or wine production higher are excluded from the sample (see also Esposti, 2014).

[19] See Bia and Mattei (2008, p. 362) for more details.

[20] When needed, the Box-Cox transformation is here adopted to restore a distribution that passes the statistical test of normality. See Bia and Mattei (2008, p. 362) for more details.

[21] In more technical terms, the most significant impact expected from the FPR is to improve farm's allocative efficiency. See Moro and Sckokai (2011) for an exhaustive theoretical background on this aspect.

production decisions) between two years or periods. The second type considers the investment behaviour, that is, production decisions oriented towards a longer-term programming horizon.
In the first typology, the outcome variable expresses the distance between two output vectors. This distance is computed using two different metrics:

$$(15) \qquad y_i^1 = \sqrt{\sum_{k=1}^{K} \left(s_{ik,B} - s_{ik,A}\right)^2}$$

$$(16) \qquad y_i^2 = \sum_{k=1}^{K} \frac{\left|d_{ik,B} - d_{ik,A}\right|}{N}$$

where $k=1,....,K$ indexes the k-th product within the vector of potential production activities, s_{ik} expresses the share of the k-th commodity on the total revenue of the i-th farm, d_{ik} is a dummy variable taking value 1(0) if the k-th product is (is not) produced by i-th farm. Finally, indexes A and B express the two points in time when these variables are observed. Typically, A = pre-treatment year, B = post-treatment year. y_i^1 and y_i^2 are just distance variables: the former is an Euclidean distance, the latter is a variant of a conventional similarity index.
$0 \leq s_{ik} \leq 1$, y_i^1 varies between 0 and 2, with the lower value taken by farms whose revenue distribution across potential products remains the same between the two years/periods. In such case, no change in production decision is observed over time. The maximum value, on the contrary, is taken by those farms that concentrate all revenue in only one product and this unique product changed between the two periods. Therefore, this outcome variable not only accounts for the change in production decisions between the two years/periods but also for the degree of specialization of the given farm.
As d_{ik} is a dummy variable, it is $0 \leq y_i^2 \leq 1$. Even in this case, the outcome variable increases the more the output vector changes. The 0 value is taken by farms for which all productions observed in A are confirmed in B and no other activity is added. In this case, however, specialization does not tend to increase the value of the outcome variable as, on the contrary, an higher value is observed for those farms that change their production activities over a large range of products. It must be also noticed that this second outcome variable does not take into account the different relevance (share) of a given k-th production in the i-th farm revenue. Therefore, it is not able to take into account changes in production decisions that take the form of an extension (reduction) of an activity over a continuous domain.
Apparently, therefore, y_i^2 is a less accurate measure of the treatment outcome than y_i^1. This latter, however, may encounter a major drawback because revenue shares s_{ik} does not only depend on farmer's production decisions but also on market prices. Prices may not only be independent on the treatment but may be even unpredicted by producers. Under remarkable price volatility, therefore, the former outcome variable may overestimate the response of farmers to treatment by attributing to it an exogenous movement of prices.
Finally, we may also argue that, under market reorientation, the response of a farm's production choices measured with y_i^1 is higher than the response of the same farm measured with y_i^2. First of all, as mentioned, the former can take into account also a change in the production decisions that give more relevance to activities with higher market convenience even if no new activities is really added (therefore, y_i^1 can be >0 even when y_i^2=0). Secondly, the latter simply counts the addition or substitution of production activities. Under market reorientation these new activities are expected to show more market convenience (for instance due to a temporary positive price dynamics) than those they replace or the preexisting ones, therefore this addition/replacement impacts more on y_i^1 than on y_i^2. According to this interpretation, a positive TE of the FPR is expected to be revealed, in treated farms, not only by a response of both y_i^1 and y_i^2, but also by an higher response of y_i^1 than y_i^2.
The second typology of outcome variable consists of a scalar measure expressing the investment decisions taken in response to the FPR. The idea simply is that the treatment may induce extra (more than "business-as-usual") investments allowing the farm to activate (extend) new (existing) activities in the longer-term. Therefore, the outcome measuring such effect is simply the change in investment

expenditures of the i-th farm (I_i) between years/periods A and B. This change is here measured in two different ways:

(17) $$y_i^3 = \left(I_{i,B} - I_{i,A}\right)$$

(18) $$y_i^4 = \left(\frac{I_{i,B}}{VA_{i,B}} - \frac{I_{i,A}}{VA_{i,A}}\right)$$

y_i^3 merely is the difference between the yearly total investment expenditure. y_i^4 expresses this difference not in absolute terms but in relative terms, that is, as investment rate given by the ratio between total investment expenditure and the respective farm value added.[22] This latter outcome variable may better capture the real investment effort of the i-th farm and get rid of the wide size heterogeneity among farms both in physical and economic terms. In doing this, however, y_i^4 may partially sterilize the effect of the treatment on investment decisions over the whole sample. In fact, a real increase in investments concentrated in larger farms may be entirely compensated by a decline in investment rates in smaller farms.

In practice, all these four outcome variables present pros and cons. The first typology only partially captures the farms' production response to the treatment. At the same time, as already mentioned, one possible problem in using investment decisions as outcome variable in the present case is that both anticipation and lagged effects may occur. For instance, a pre-treatment year/period could already contain some anticipated investments response of the farmers to the FPR and this makes the identification and estimation of the TE more complex and, consequently, results less reliable and robust. Therefore, all outcome variables (15)-(18) represent a relevant but incomplete dimension of the production response to the FPR. Actually, they are more complements than alternatives in providing a comprehensive picture of the reorientation to market. For these reasons, all the four outcome variables will be used throughout the present empirical study.

Table 1 reports some descriptive statistics by treatment group for the four outcome variables. It may be easily appreciated that, comparing the two extreme years 2003-2007, for all outcome variables the average values tend to increase moving from the non-treated (or control) group to the treatment group. Moreover, within the treatment group, farmers also receiving second pillar payments show an additional positive impulse. Nonetheless, if we look at comparisons between other couples of years, the picture becomes less clear. While y_i^1 and y_i^2 tend to confirm higher values, more mixed evidence emerges for y_i^3 and y_i^4 especially when years considered are those around the treatment year (2005). This can be explained by the presence of anticipation and/or lagged effects, but another explanation could simply be that the observed differences in outcome variables are not caused by the FPR.

In fact, the most relevant evidence emerging in Table 1 is the high variability of all outcome variables in the whole sample, as well as in treatment groups. In practice, if we constructed a conventional 95% confidence interval around the sample averages we would notice that these intervals are largely overlapping across the groups for all outcome variables. More generally, though these differences in outcome variables are mostly consistent with the expectations in terms of policy TE, these simple statistics can not be conclusive on the fact that such differences across groups can be indisputably attributable to FPR.

A further issue in the proper definition of the outcome variable, concerns the selection of observations/years A and B. In principle, several couples of years/periods could be compared to compute the outcome variables in (15)-(18). Years 2003 and 2004 unquestionably represent before-reform (thus, before-treatment) years as the implementation of the reform started in 2005 in Italy. At the same time, years 2005, 2006 and 2007 can be considered as after-reform (after-treatment) years. As a consequence, the following pairs of years can be candidate for a before and after-treatment comparison: 2004-2005, 2004-2006, 2004-2007, 2003-2005, 2003-2006 and 2003-2007. However, the choice here is to consider years that are symmetric with respect to the treatment year (2005) and are far enough from it

[22] The value added rather the value of production is here considered because the former can be more properly considered a proxy of farm profits, that is, of the capacity to generate surplus from which further investments can be undertaken.

to exclude (or minimize) anticipated or lagged responses. For this reasons, the outcome variables are here computed assuming A=2003 and B=2007.[23]
Finally, Table 1 also reports the (sub)sample averages of the covariates considered in the present analysis. The covariates (or confounding factors) here considered are those pre-treatment variables (covariates **X**)[24] expected to incorporate all the relevant aspects that may affect the production choices before the treatment (thus, affecting the outcome regardless the treatment itself) as well as the treatment assignment. We selected these variables to capture, with the minimum redundancy, three different types of factors (Table 3).
First of all, we consider the relevant *individual characteristics* of the farmer (AGE) and of the farm (Altitude - ALT). Secondly, the economic (ES, FC)[25] and physical (AWU, HP, UAA and, at least partially, LU) *size* of the farm clearly matters. All these variables evidently affect the outcomes but presumably are not directly correlated with the treatment assignment. Still, they are definitely linked to production choices and, since pre-treatment production choices are unquestionably correlated with the treatment assignment, this correlation indirectly occurs even with respect to these first two categories of covariates.
The third typology of confounding factors, in fact, consists of those variables (TF and, in part, LU) that directly express the *production specialization* of the farm. The linkage between these covariates and the treatment assignment is evident as this actually concerns those farms that were interested by specific OCM measures while, on the contrary, farms not involved in first pillar are those whose production specialization was less (or not at all) targeted by specific policy measures. To express farm production specialization, the 4-digit "Type of Farm" (TF) FADN classification is adopted (2000 classification).[26]
Even for these covariates the dominating evidence concerns the large variability observed in both the whole sample and in treatment groups, and this prevents from clear-cut statements about structural differences across the groups. Only for few variables a difference between treatment groups' clearly emerges. In particular, non-treated units tend to show some peculiarities compared to the treated ones while the difference between the treated units with and without second pillar payments seems mostly negligible. Non-treated units show a smaller physical size (UAA) but this is not necessarily true if we consider the economic size (ES). Moreover, as expected, the production specialization of the non-treated group is evidently less dependent on first pillar support (TF_R), it practically excludes livestock activities (LU) while it favours activities mostly run in flat areas (ALT). The immediate interpretation is that most of these non-treated units are farms with small area and high output values strongly specialized in a certain kind of production (e.g., horticulture) largely disregarded by coupled first pillar payments. This reinforces the idea that these farms might not be reliable as control units in identifying and estimating the TE of the FPR.

[23] As stressed by Esposti (2014), the choice of the years A and B to compute the outcome variable may raise significant measurement issues. Not only the presence of possible anticipated and lagged response of farmers can create problems. Other highly time-varying external (e.g., price dynamics) and internal (e.g., rotation practices) factors can make the measurement of the outcome variable sensitive to the chosen years. The choice here made to consider years (2003 and 2007) that are symmetric with respect to the treatment year (2005) and far enough from it to exclude (or minimize) anticipated or lagged responses is expected to minimize these measurement problems. Nonetheless, other years (or averages of years) has been alternatively considered and estimations consequently performed. Results are qualitatively not different from those here presented though, in general terms, of lower statistical quality. They are available upon request.
[24] Pre-treatment variables have been observed in 2003, the only exceptions being FC, for which the 2003-2004 average has been considered since this variable may largely vary on a yearly base.
[25] The relative (with respect to net value added) amount of fixed costs expresses the importance of fixed factors (especially labour and physical capital) within the farm and, therefore, it is a proxy of the scale of the farm business itself.
[26] By itself, however, this qualitative variable is not suitable in this empirical exercise as it has not a monotonous linkage with the treatment assignment. For instance, class 4210 (beef production) is more dependent on first pillar support than classes 2022 (flowers' production) and 6010 (horticulture); therefore, farms belonging to the former class are more likely to be assigned to the treatment than farms of the latter classes. To overcome this problem, the official TF classification has been reclassified by assigning to any 4-digit class a number (ranging from 1 to 7) expressing its dependency on first pillar CAP support. This number expresses a qualitative monotonous variable (TF_R) that increases as the dependency on CAP support declines.

Table 1. Descriptive statistics over treatment groups: sample averages of the outcome and pre-treatment variables (standard deviation in parenthesis)

Treatment group:	*Not treated*	*Treated*	*Whole Sample*
Outcome variables:			
y^1 (distance index)			
2007 vs. 2003	0.169 (0.435)	0.594 (0.739)	0.522 (0.688)
2004 vs. 2003	0.084 (0.322)	0.562 (0.778)	0.481 (0.700)
2007 vs. 2006	0.054 (0.249)	0.119 (0.304)	0.108 (0.294)
y^2 (distance index)			
2007 vs. 2003	0.006 (0.014)	0.017 (0.020)	0.015 (0.019)
2004 vs. 2003	0.003 (0.009)	0.011 (0.017)	0.010 (0.016)
2007 vs. 2006	0.007 (0.012)	0.020 (0.017)	0.018 (0.016)
y^3 (in €)			
2007 vs. 2003	-20477 (179122)	-5596 (136912)	-8126 (171088)
2004 vs. 2003	-8862 (163765)	2008 (104365)	160 (141463)
2007 vs. 2006	-4048 (35240)	-4019 (110414)	-4024 (103035)
y^4 (in €)			
2007 vs. 2003	-0.401 (1.533)	-0.272 (1.870)	-0.294 (2.083)
2004 vs. 2003	-0.101 (1.193)	-0.0250 (1.768)	-0.038 (1.883)
2007 vs. 2006	-0.128 (0.958)	-0.213 (2.651)	-0.199 (2.903)
Pre-treatment variables:			
AGE (of the holder) (years)	51.85 (13.84)	52.68 (14.89)	52.54 (14.71)
Altitude (ALT) (m)	154.28 (192.11)	290.38 (290.52)	267.24 (273.79)
Annual Working Units (AWU)	2.79 (6.22)	2.06 (3.06)	2.18 (3.60)
Economic Size (ES) (classes)	6.41 (2.19)	6.12 (2.31)	6.17 (2.29)
Fixed Costs (on Net Value Added) (FC)	2.79 (36.41)	1.81 (16.64)	1.98 (20.00)
Horse Power (HP)	93.23 (129.58)	187.52 (235.09)	171.49 (217.15)
Livestock Units (LU)	5.73 (50.53)	46.69 (255.24)	39.73 (220.44)
Utilized Agricultural Area (UAA) (ha)	7.50 (24.34)	37.89 (76.86)	32.72 (67.93)
Type of Farm (TF) (4-digits)*	(fruits)	(arable crops)	(arable crops)
Type of Farm (reclassified) (TF_R)	5.07 (1.67)	3.37 (1.65)	3.66 (1.65)

* In this case the Table reports the higher frequency class

5. Estimation Results

5.1. GPS estimates

The initial stage that is common to the two multivalued estimation methodologies here considered consists in the GPS estimation. Nonetheless, the specification of the GPS function, $GPS_i = r(T_i, \mathbf{X}_i)$ is necessarily different in the two cases. In the Hirano-Imbens approach, T (i.e., the *TI*) is a continuous variable and, therefore, $r(T_i, \mathbf{X}_i)$ is specified as a linear flexible functional forms, i.e., a fully interacted polynomial in both arguments, T_i and $\mathbf{X}_i$. In the Cattaneo approach, on the contrary, T (*TI*) is a categorical variable thus $r(T_i, \mathbf{X}_i)$ necessarily takes a multinomial specification. A Multinomial Logit Model (MLM) is here adopted. Even for this MLM a fully interacted polynomial specification in both arguments of the MLM is adopted. In both cases, however, the functional forms actually estimated are the best fitting specifications selected according to the AIC starting from a fully interacted second-order polynomial (quadratic specification).

Table 2 reports the GPS function estimates in the continuous (linear regression) case. It emerges that most estimated parameters are significantly different from 0, the only exceptions being those concerning variables ALT and FC. A higher generalized propensity (that is, a higher probability to receive a higher *TI*) is associated to larger size (in both economic and physical terms) farms and older farmers while, on the contrary, a lower propensity is associated to farms with higher labour intensity and specialization in less supported activity included a higher presence of livestock activities.

These results are consistent with the expectation given the well-known distribution of first pillar's coupled payments in Italy (Povellato and Velazquez, 2005) and with what observed in the binary treatment case. The reliability of these results can be assessed by testing the balancing condition over the space of the estimated GPS and distinguishing the *TI* in the abovementioned 7 intervals. The balancing tests on the common support accept the balancing condition at the 95% confidence level: for no regressor and in no treatment level we observe a mean difference between the units belonging to that treatment group and the respective average that is statistically different from 0.[27]

Table 2. GPS estimation: linear regression of the continuous treatment on the covariates (standard errors in parenthesis)[a]

AGE (of the holder) (years)	0.0132* (0.0019)	Horse Power (HP)	0.0004* (0.0002)
Altitude (ALT) (m)	-0.0000 (0.0001)	Livestock Units (LU)	-0.0003* (0.0001)
Annual Working Units (AWU)	-0.1882* (0.0123)	Utilized Agricultural Area (UAA) (ha)	0.0067* (0.0005)
Economic Size (ES) (classes)	0.0405* (0.0130)	Type of Farm (reclassified) (TF_R)	-0.5639* (0.0188)
Fixed Costs (on Net Value Added) (FC)	-0.0018 (0.0013)	Constant term	4.365* (0.1620)

[a]The BoxCox transformation of the treatment variable is used; the assumption of Normality is statistically satisfied at .05 level
*Statistically significant at 0.05 level

Moreover, indirectly, the robustness of these results is also confirmed by the other GPS function, i.e., the MLE estimation of the MLM implied by the Cattaneo (2010) approach.[28] In this latter case, a further difference with respect to the continuous case consists in the exclusion of the categorical independent variables (ES, TF_R) as they would prevent the MLE estimation of the MLM to reach convergence. Parameter estimation of the MLM actually refer to the different *TI* levels and, as well known (Cattaneo et al., 2013.), can not be directly interpreted as marginal effects passing from one treatment level to the subsequent.[29] Nonetheless, also these results indicate that a higher *TI* is associated with larger farms (in physical and economic terms), lower labour intensity and older farmers, while on the contrary lower *TI* is associated with flat-areas agriculture and more labour-intense activities included livestock productions.

As a matter of fact, these estimates of the GPS function's parameters do not provide any particular novel evidence compared to expectations and to what already emerges from a simple descriptive analysis. Nonetheless, this is just the first necessary step to achieve the multivalued ATE estimation pursued here and its reliability and robustness is needed to make ATE estimation itself reliable.

[27] Due to space limitations the balancing tests are not reported here; they are available upon request. More details on these aspects can be found in Bia and Mattei (2008, p. 368-369; 2012) and Flores and Mitnik (2008). Bia and Mattei (2012) and Bia et al. (2013) also put forward some first attempts to perform a sensitivity analysis assessing for the validity of the weak unconfoundedness under multivalued treatments.
[28] Due to space limitations, these estimates are reported in the Annex.
[29] Estimated marginal effects are available upon request.

5.2. The Hirano-Imbens approach: DRF estimates

Table 3 reports the results of the second estimation step in the Hirano-Imbens approach for the four outcome variables under consideration here. It consists of the estimation of *uDRF*, $g(G\hat{P}S_i, T_i) = E\left[Y_i \middle| G\hat{P}S_i, T\right]$. Again, the empirical parametric specification of the estimand function started from a fully-interacted quadratic form in both arguments $G\hat{P}S_i$ and *T* (see (4)), where $G\hat{P}S_i$ is the GPS estimated in the first stage (Table 2). Then, the best fitting specification, according to the AIC, has been adopted. As evident in Table 3, this adopted specification includes both the quadratic terms and the interaction term of the two arguments. It is well emphasized in this empirical literature (Hirano and Imbens, 2004; Bia and Mattei, 2008) that these estimates can not be given any direct economic interpretation and are just intermediate results. However, they are needed to achieve the final estimation step providing the estimation of $aDRF(T) = E[(\hat{g}(T)]$ in (5) and, consequently, of the ATE we are mainly interested in.

Estimation results are of good quality for the first outcome variables, y^1, where parameters of most terms are statistically significant expressing a complex relationship among $G\hat{P}S_i$, T and the *uDRF*. Starting with the second outcome variable (y^2), however, results become statistically weaker and the treatment level significantly affect the *uDRF* only interacting with the estimated GPS: the DRF comes more from the propensity to receive a given *TI* rather than by the *TI* itself. This is even more clear in y^3 and y^4, where few parameters (just one for y^3) are statistically significant and they all involve the estimated GPS, alone or in interaction with the *TI.* It remains true, however, that results seem statistically weaker passing from the first two outcome variables, dealing with the production mix, to those related to investment choices. Generally speaking, the fact that the impact on the DRF mostly comes more from the propensity to receive a given *TI* is a further indirect evidence of how, even in the multivalued case, the selection-on-observables bias may be relevant.

Figures 2-3 display both the estimated average DRF and ATE (point estimates and confidence intervals) for the four outcome variables over the observed continuous domain of TI (therefore, we can also call the latter *average TE function*). Again, we may appreciate a significant different behavior of the first two outcomes compared to y^3 and y^4. For y^1 and y^2 we observe similar results in several respect. First of all, in both cases DRF and TE estimates appear to be statistically robust as standard errors are small and, consequently, the confidence intervals thin. The response increases in the initial TI levels and reaches its maximum at about *TI* = 10% and *TI* = 30%, respectively. However, the corresponding TE, though positive, is declining. Once the maximum response is reached, it starts declining regularly and this evidently implies a negative TE. This behaviour is particular evident and relevant in the case of y^1 where the TE rapidly vanishes to 0, while, on the contrary, in the case of y^2 a positive TE is observed until TI levels that encompass the large part of the observed farms.

The interpretation of this evidence seems particularly interesting. The FPR induced the farms' response in terms of production mix/market reorientation, but this mostly regards those farms receiving a relatively limited CAP payment with respect to its GVP (i.e., a low TI). Moreover, this TE is particularly evident in terms of composition of the production mix; on the contrary, it is quite limited if this change of production mix is measured in terms of revenue composition. Evidently, the most responding farms are those that were already market-oriented before the FPR as demonstrated by the limited incidence of the CAP support on their GPV. Whenever we move towards farms strongly dependent on the CAP their response is weaker. Moreover, this response may be more apparent than substantial. It implies that new productions are activated while others are quitted and this affects y^2. At the same time, this change might not provide any significant change in revenue composition, that is, in y^1. In practice, many farms, especially those strongly dependent on first pillar support, remain locked-in their original rent-seeking production choices and only timidly attempt some minor changes that actually result in marginal adjustments in revenue composition.

As could expected, the results emerging for the other two outcome variables, y^3 and y^4, are much less robust and interpretable. Expectations come from the fact that investment decisions may depend on many strategic long-term aspects, not considered here, beside the decoupling of CAP support. Moreover, as discussed in previous sections, an investment response to the policy change may have a timing (anticipation and lagged effects) that makes a proper identification of the ATE in the present circumstance much more difficult. It is still worth noticing that the response, though with an increasing statistical variability, linearly increases over the observed TI domain. This occurs in both cases thus suggesting that farm size (that is, what really makes the difference between y^3 and y^4) does not qualitatively influence the response.

Table 3. DRF coefficient estimates (GPS=estimated Generalised Propensity Score; T=treatment level) [a]

	y^1	y^2	y^3	y^4
T	0.0182*	0.0001	-729.8	-0.0103
	(0.0037)	(0.0001)	(903.9)	(0.0118)
T^2	-0.0003*	0.0000	9.437	0.0002
	(0.0001)	(0.0000)	(16.616)	(0.0002)
GPS	4.921*	-0.0575*	107956	-4.092*
	(0.9162)	(0.0261)	(221262)	(1.879)
GPS^2	-9.239*	0.2389*	-703988	17.82*
	(3.603)	(0.1024)	(870085)	(8.320)
GPS*T	-0.1378*	0.0023*	7411*	0.0467
	(0.0168)	(0.0005)	(3065)	(0.0529)
Constant term	0.2094*	0.0165*	-11528	-0.1136
	(0.0549)	(0.0016)	(13262)	(0.1725)

[a]The BoxCox transformation of the treatment variable is used
*Statistically significant at 0.05 level

The slight but regular and statistically significant increase of the TE across the *TI* can be interpreted in two directions. On the one hand, we can conclude that the TE is constant over the *TI* domain, so the ATE corresponds to the TE observed in any different treatment level. This seems particularly reasonable in the case of y^4, as it implies that an increase of *TI* has a constant positive impact on investment rates. On the other hand, we can conclude that, though slightly, the TE is increasing with the TI and this seems consistent with the decoupling of support: the higher the *TI*, the higher the amount, relative to the farm's GPV, of free financial resources provided by the FPR to make investments. The fact that these resources are free, i.e. not associated to specific production activities, may indirectly suggest that these investments are dedicated to other/new activities. However, this is not granted given the definition of the two outcome variables under question (y^3 and y^4). Even if we acknowledge that a positive and maybe increasing TE occurs, this does exclude that investments are made on the same activities carried out by the farm even before the FPR.

5.3. EIF and IPW estimates

The robustness of the DRF and ATE estimation obtained with the Hirano-Imbens approach can be hardly assessed. Once the balancing condition is confirmed in the GPS estimation stage, the remaining steps strongly depend on arbitrary (parametric) specification assumptions. To indirectly evaluate to what extent these assumptions may influence results, the Cattaneo (2010) approach can be helpful as it is based on a substantially different sequence of estimation steps.
The second step of this approach consists in the estimation of $e_j(\mu_j, \mathbf{X}_i)$, that is, the mean of the potential outcome conditional on covariates $\mathbf{X}$. Again, a flexible form is initially specified and the best fitting specification, selected with AIC, is then adopted. Here, the selected specification is the fully interacted quadratic form that implies the estimation of an high number of parameters. Therefore, due to space limitations, these estimates are neither reported nor commented here.[30] After all, again, this is an intermediary estimation step, without particular economic content, to achieve the final stage, that is, the

[30] These estimates are available upon request.

estimation of the average potential outcome for any given treatment level under the two alternative estimation procedures ($\mu_{EIF,j}$ and $\mu_{IPW,j}$).

Table 4 reports the $\mu_{EIF,j}$ and $\mu_{IPW,j}$ estimates for all outcome variables and respective treatment levels, while Figures 4-5 display the respective variations across the treatment levels, that is, the estimated ATE according to (14). In general terms, results confirm what emerged within the Hirano-Imbens approach. For both y^1 and y^2 the potential outcome increases with the treatment for the first *TI* levels, then we have the inversion in this relationship with potential outcome declining with the *TI*. This inversion comes quite soon in the case of y^1, i.e. when *TI* exceeds 10%, while it comes later for y^2. This implies that the estimated TE is positive in the first levels of treatment but with higher TI it declines to zero and then becomes negative. Compared to the DRF estimation of the Hirano-Imbens approach, the most evident difference consists in the fact that confidence intervals are here larger, thus indicating a less efficient estimate of the ATE. Nonetheless, the tests of "zero TE" reported in Table 4 clearly indicate that such hypothesis can be excluded for both y^1 and y^2.

Table 4. Average potential outcome for each treatment level: EIF and IPW estimates (standard errors in parenthesis)

Outcome variable:	y^1	y^2	y^3	y^4
Treatment level:				
	EIF estimates			
1	0.6342*	0.0141*	-2799	-0.3877*
	(0.0163)	(0.0004)	(3502)	(0.0471)
2	0.8540*	0.0175*	-21658*	-0.3551*
	(0.0203)	(0.0006)	(4360)	(0.0480)
3	0.7885*	0.0195*	-8364	-0.3540*
	(0.0240)	(0.0008)	(2945)	(0.0617)
4	0.7008*	0.0195*	-11230*	-0.3681*
	(0.0294)	(0.0009)	(9281)	(0.0495)
5	0.6609*	0.0190*	6892	-0.3234*
	(0.0287)	(0.0011)	(11470)	(0.0729)
6	0.6883*	0.0186*	-20934*	-0.2265*
	(0.0289)	(0.0013)	(9764)	(0.1130)
7	0.6404*	0.0146*	-29781*	-0.1909*
	(0.0263)	(0.0014)	(6480)	(0.0684)
Test of zero treatment effect[a]	126.9*	47.18*	12.23*	12.34*
	IPW estimates			
1	0.6155*	0.0142*	-5581	-0.4045*
	(0.0187)	(0.0010)	(6925)	(0.0910)
2	0.8289*	0.0177*	-27916*	-0.4010*
	(0.0219)	(0.0007)	(3946)	(0.0966)
3	0.6458*	0.0194*	-13203	-0.1351
	(0.0814)	(0.0076)	(20588)	(0.1036)
4	0.5324*	0.0194*	-11968*	-0.3166*
	(0.0550)	(0.0037)	(5767)	(0. 0901)
5	0.4921*	0.0196*	5393	-0.2712
	(0.1728)	(0.0043)	(12822)	(0.1576)
6	0.4849	0.0178	-7816	-0.2381
	(0.2797)	(0.0172)	(50054)	(0.6418)
7	0.4094*	0.0171	-8617	-0.0109
	(0.4674)	(0.0126)	(76096)	(0.9844)
Test of zero treatment effect[a]	90.1*	41.1*	16.5*	7.5

[a] The test distributes as a χ^2 with 6 d.o.f. under the H_0 that the potential outcome is equal for any treatment level

*Statistically significant at 0.05 level

In the case of y^3 and y^4 it seems more difficult to conclude that results tend to be qualitatively concordant with what obtained with the Hirano-Imbens approach. In particular, confidence intervals around the estimated potential outcome are large such that, for some treatment levels, the it is not statistically different from zero. This is particularly true for y^3 and in the case of the IPW estimation. Therefore, also the consequent estimated ATE are often not statistically significant and do not necessarily show a regular pattern across the treatment levels. In the case of y^3, though we can reject the hypothesis of no TE, the only robustly significant ATE is that observed passing from the first to the second treatment level, that is, around 5% of *TI*. However, this effect is negative. For all other TI the ATE is statistically zero. Better evidence is obtained for y^4, as expected due to the lower statistical noise caused by heterogeneous size. The ATE is positive passing from the first to the second treatment levels than it goes to 0, but it comes back into positive territory and with an apparent increasing pattern after the forth treatment level. In this respect, these results confirm what obtained in the case of Hirano-Imbens approach. The investment response to the treatment is much less clear than the production choice response. Nonetheless, when the investment rate, rather than the investment absolute value, is considered the response seems to be positive and, to a certain extent, slightly increasing with the increase of the *TI*. Higher *TI*, as mentioned, can be associated to an higher amount of free financial resources to make brand new investments.
A further general evidence that is worth noticing concerns the comparison between EIF and IPW estimates. Though results are qualitatively similar, IPW estimates are clearly less efficient. Therefore, despite the asymptotical equivalence of these two estimators, in the finite sample and in this specific application, the EIF estimation appears to be generally superior.

6. CONCLUDING REMARKS

This paper primarily aims at showing how the TE toolkit can be successfully applied also to a peculiar kind of treatment, the 2005 CAP first pillar reform. With a long and large enough balanced panel and an appropriate definition of the outcome variable, the impact of the reform on farms' market reorientation can be estimated by taking advantage of the fact that this peculiar case can be considered a multivalued treatment. Recent estimation approaches suitable for such circumstances have been applied.
Results show that, as expected, the farms' response in terms of market reorientation tends to be limited to short-run choices and has been declining with the treatment intensity. More surprisingly, it emerges that this response is significant and positive just for the lower levels of treatment intensity, that is, for those farms that are less dependent on first pillar support. In more dependent farms the response rapidly declines to 0. Moreover, this response is more evident in terms of introduction of new productions or, more generally, of change in the production mix. It significantly reduces when the response is measured in terms of change in revenue composition thus indicating that, even when present, this response remains fairly conservative.
While results provide quite robust evidence about the effect of the reform, however, some steps forward could be proposed with respect to the present approach. The adopted estimation strategy is interesting and promising but also raises several practical issues. The main problems are represented by its complexity and by the need of many arbitrary assumptions. As a consequence, the role of possible misspecification of the models estimated in the intermediate steps needs to be explored. Nonetheless, the methodological toolkit seems rich enough not only to check for robustness of results by comparing different estimation approaches, but also to push the investigation further by refining these estimation approaches, as well as the definition of the treatment and outcome variables (as well as of the relevant covariates), and the construction of the balanced panel. For instance, it could be argued that some second pillar measures substantially interfere with production choices, thus with the FPR TE itself. In this respect, a more sophisticated articulation of treatment groups could be attempted. Actually, the analysis of multiple continuous treatments is at the forefront of the current TE econometrics literature (Frölich, 2004; Imbens and Wooldridge, 2009) and is well beyond the scope of the present paper. Nonetheless, some methodological solutions accompanied by appropriate estimation techniques could be proposed and attempted in future research.

REFERENCES

Abadie, A., Drukker, D., Herr, J.L. and Imbens, G.W. (2004). Implementing Matching Estimators for Average Treatment Effects in Stata. *The Stata Journal*, 4(3), 290-311.

Becker, S. O. and Ichino, A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal*, 2(4), 358–77.

Bia, M., Mattei, A. (2007). *Application of the Generalized Propensity Score. Evaluation of public contributions to Piedmont enterprises*. POLIS Working Paper 80, Department of Public Policy and Public Choice, Università del Piemonte Orientale, Alessandria.

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

Bia, M., Mattei, A. (2012). Assessing the Effect of the Amount of Financial Aids to Piedmont Firms Using the Generalized Propensity Score. *Statistical Methods and Applications*, 21 (4), 485-516.

Cagliero, R., Cisilino, F. and Scardera, A. (2010). *L'utilizzo della RICA per la valutazione di programmi di sviluppo rurale*. Roma: Rete Rurale Nazionale.

Cattaneo, M.D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155, 138–154.

Cattaneo, M.D., Drukker, D.M. and Holland, A.D. (2013). Estimation of multivalued treatment effects under conditional independence. *The Stata Journal*, 13(3), 407–450.

Chabé-Ferret, S. (2010). To Control or Not to Control? Bias of Simple Matching vs Difference-In-Difference Matching in a Dynamic Framework. Paper presented at the 10th World Congress of the Econometric Society, Shanghai, 17-21 August.

Chabé-Ferret, S. and Subervie, J. (2013). How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching. *Journal of Environmental Economics and Management*, 65, 12–27.

De Filippis, F., Romano, D. (a cura di) (2010). *Crisi economica e agricoltura*. Gruppo 2013-Coldiretti, Roma.

Duflo, E., Glennerster, R., Kremer, M. (2006). *Using Randomization in Development Economics Research: A Toolkit*. NBER Technical Working Papers 0333, Cambridge: National Bureau of Economic Research.

EENRD (2010). *Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors*. European Evaluation Network for Rural Development (EENRD), European Commission, Brussels.

Esposti, R. (2007). Regional Growth and Policies in the European Union: Does the Common Agricultural Policy Have a Countertreatment Effect? *American Journal of Agricultural Economics,* 89, 116–134.

Esposti, R. (2011), Reforming the CAP: an agenda for regional growth? In: Sorrentino, S., Henke, R., Severini, S. (eds.), *The Common Agricultural Policy after the Fischler Reform. National Implementations, Impact Assessment and the Agenda for Future Reforms,* Farnham: Ashgate, 29-52.

Esposti, R. (2014). To match, not to match, how to match: Estimating the farm-level impact of the CAP-first pillar reform (or: How to Apply Treatment-Effect Econometrics when the Real World is a Mess). Quaderno di Ricerca n. 400, Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche.

Esposti, R., Listorti, G. (2013). Agricultural Price Transmission across Space and Commodities during Price Bubbles. *Agricultural Economics*, 44 (1), 125.

European Commission (2006). *Rural Development 2007-2013. Handbook on common monitoring and evaluation framework. Guidance document*. DG Agriculture and Rural Development, European Commission, Brussels.

European Commission (2011). The CAP in perspective: from market intervention to policy innovation. *Agricultural Policy Perspectives Briefs*, Brief n° 1, European Commission, Directorate-General for Agriculture and Rural Development, Brussels.

Flores, C.A., Mitnik. O.A. (2009). Evaluating Nonexperimental Estimators for Multiple Treatments: Evidence from a Randomized Experiment. Paper presented at the 2009 ASSA (Allied Social Science Association) meeting, San Francisco, 3-5 January.

Frascarelli A. (2008). L'Ocm unica e la semplificazione della Pac. Gruppo 2013-Coldiretti, Roma.

Frölich, M. (2004). Programme evaluation with multiple treatments. *Journal of Economic Surveys*, 18, 181-224.

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

Imbens, G.W. and Wooldridge, J.M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47 (1), 5–86.

Kluve, J., Schneider, H., Uhlendorff, A., Zhao, Z. (2012). Evaluating continuous training programs using the generalized propensity score. *Journal of the Royal Statistical Society*, 175, 587-617.

Lukesch R., Schuh B. (eds.) (2010). Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors. Working Paper, March 2010, Findings of a Thematic Working Group established and coordinated by The European Evaluation Network for Rural Development, Brussels.

Magrini, E., Montalbano, P., Nenci, S., Salvatici, L. (2013). Agricultural trade distortions during recent International price spikes: what implications for food security? Paper presented at the 54th Annual Conference of the Italian Economic Association (SIE), Bologna, 24-26 October.

Michalek, J. (2012). Counterfactual impact evaluation of EU rural development programmes – Propensity score matching methodology applied to selected EU member states: vol. 1, a micro-level approach. Joint Research Center (JRC), Sevilla.

Moro, D., Sckokai, P. (2011). The impact of pillar I support on farm choices: conceptual and methodological challenges. Paper presented at the 122nd EAAE Seminar "Evidence-Based Agricultural and Rural Policy Making: Methodological and Empirical Challenges of Policy Evaluation", Ancona (Italy), February 17-18.

OECD (2011). Evaluation of Agricultural Policy Reforms in the European Union. Paris: OECD Publishing.

Povellato A., Velazquez B.E. (2005). La riforma Fischler e l'agricoltura italiana. Roma: INEA.

Pufahl, A., Weiss, C. (2009). Evaluating the effects of farm programmes: results from propensity score matching. *European Review of Agricultural Economics*, 36 (1), 79-101.

Renwick, A., Revoredo-Giha, C. (2008). Measuring Cross-Subsidisation of the Single Payment Scheme in England. Paper presented at the 109th EAAE Seminar "The CAP after the Fischler Reform: National Implementations, Impact Assessment and the Agenda for Future Reforms", Viterbo (Italy), November 20-21.

Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. *Journal of Educational Psychology*, 66, 688-701.

Salvioni, C., Sciulli, D. (2011). Impact of Rural Development Policy and Less Favoured Areas Scheme: A Difference in Difference Matching Approach. Paper presented at the EAAE 2011 Congress, Change and Uncertainty Challenges for Agriculture, Food and Natural Resources, August 30-September 2, Zurich.

Smith, J., Todd, P. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators. *Journal of Econometrics*, 125 (1–2), 305–353.

Sorrentino, S., Henke, R., Severini, S. (eds.) (2011). *The Common Agricultural Policy after the Fischler Reform. National Implementations, Impact Assessment and the Agenda for Future Reforms,* Farnham: Ashgate.

Sotte, F. (2006). Imprese e non-imprese nell'agricoltura Italiana. *Politica Agricola Internazionale*, 1/2006, 13-30.

Sotte, F., Arzeni, A. (2013). Imprese e non-imprese nell'agricoltura italiana. *AgriRegioniEuropa* (*ARE*), 9(32), 65-70.

Figure 1. Distribution of the continuous the treatment, First Pillar support on farm's GPV (in %): Kernel density (**K**) and frequency histogram (avg. over 2003-2007 period)

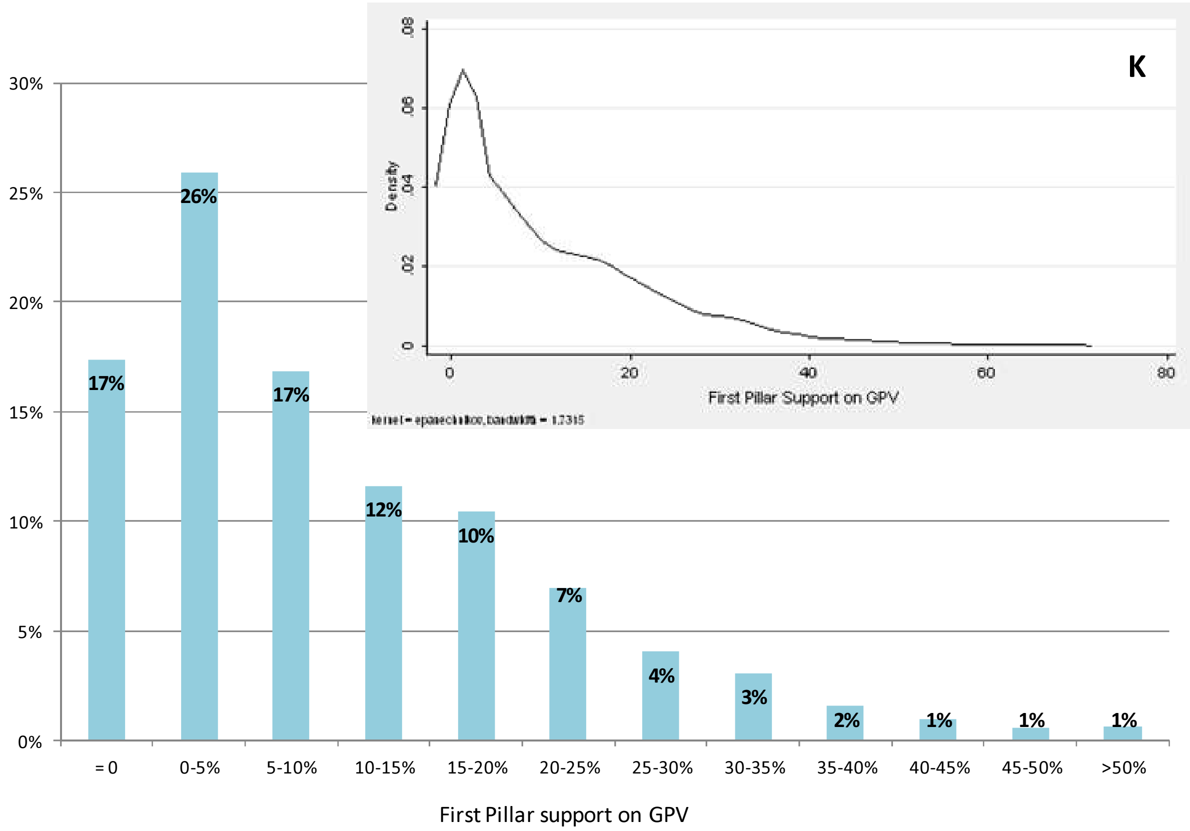
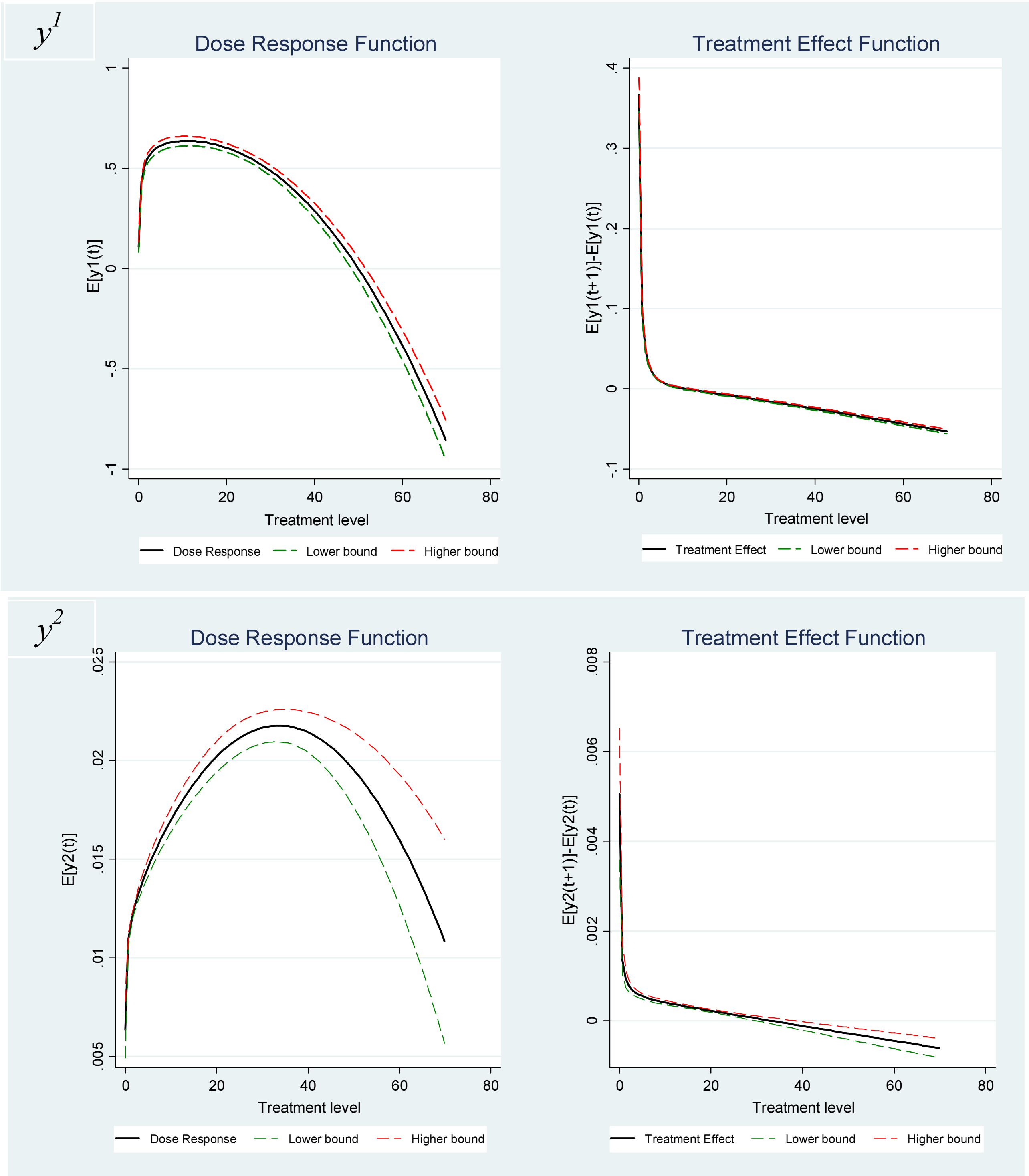
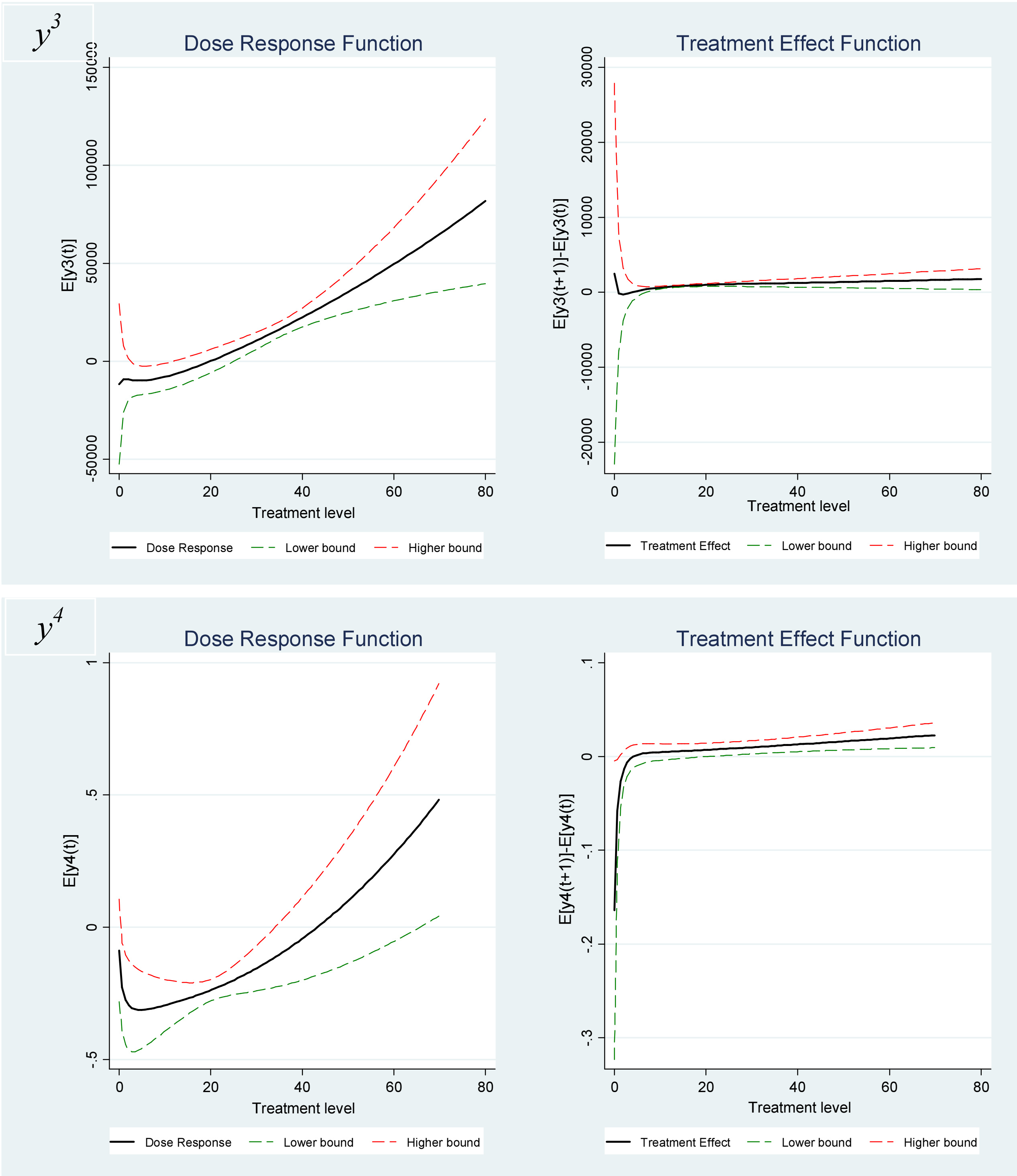


Figure 2. Estimated average DRF and TE for outcome variables y^1 and $y^{2\,a,b}$



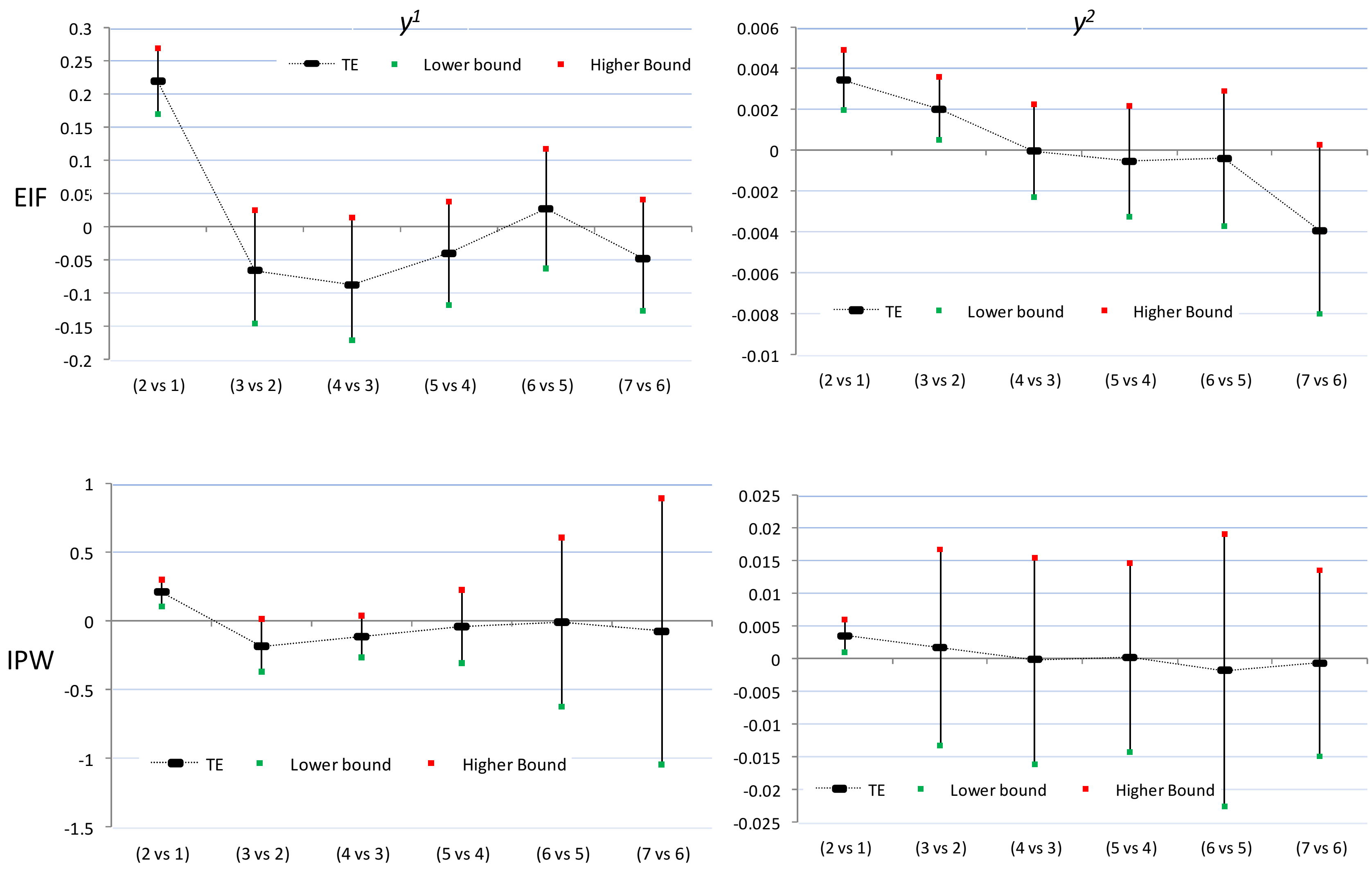
[a] Bootstrap standard errors (100 replications)
[b] Confidence bounds at .95 % level

Figure 3. Estimated average DRF and TE for outcome variables y^3 and y^4 [a,b]



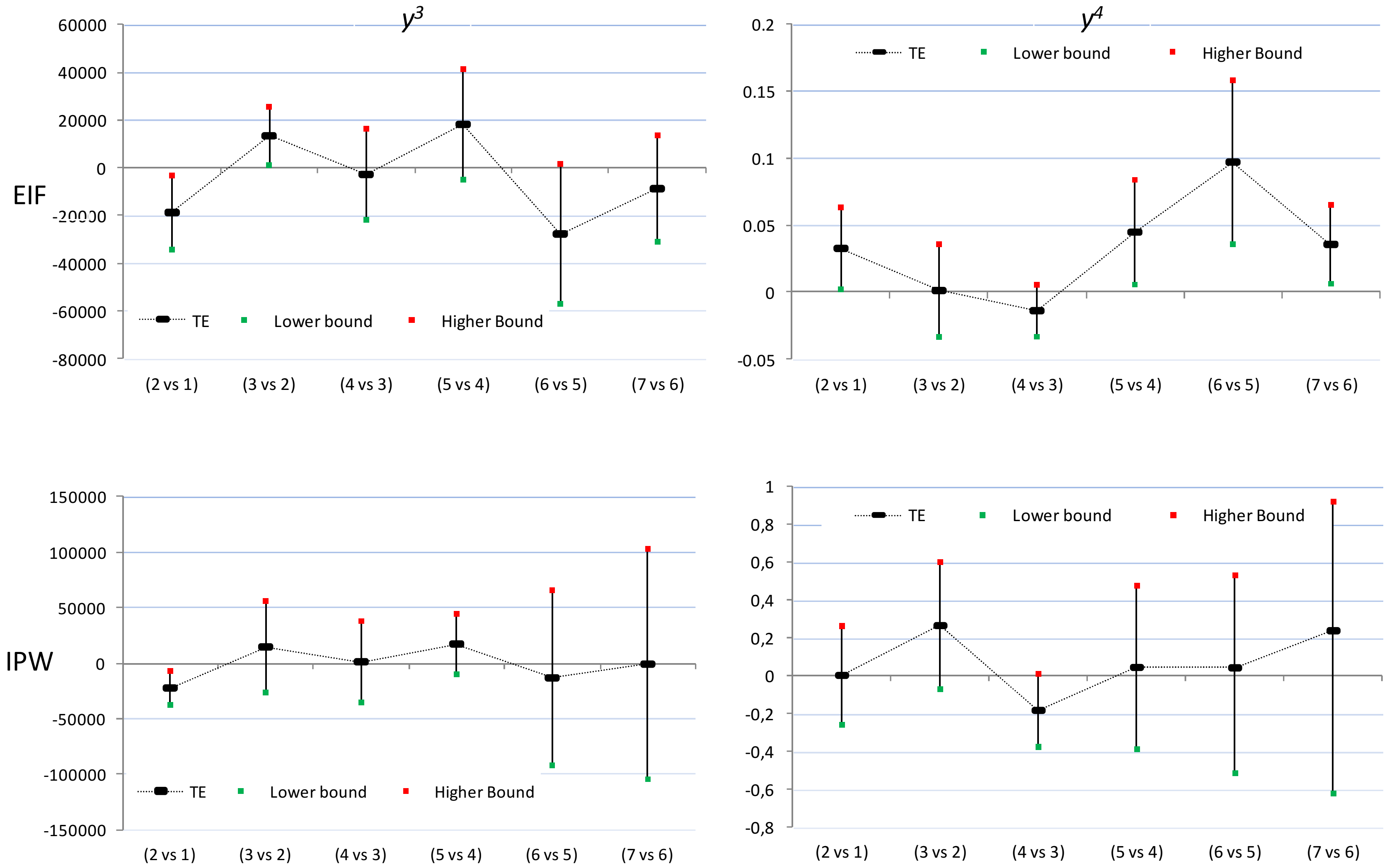
[a] Bootstrap standard errors (100 replications)
[b] Confidence bounds at .95 % level

Figure 4. Potential outcome variation across the treatment levels: EIF and IPW semi-parametric estimations of the TE for outcome variables y^1 and y^{2} [a, b]



[a] Bootstrap standard errors (100 replications)
[b] Confidence bounds at .95 % level

Figure 5. Potential outcome variation across the treatment levels: EIF and IPW semi-parametric estimations of the TE for outcome variables y^3 and y^4 [a, b]



[a] Bootstrap standard errors (100 replications)
[b] Confidence bounds at .95 % level

ANNEX

GPS function: MLM (Multinomial Logit Model) estimation of the discrete treatment on the covariates (standard errors in parenthesis)

Treatment level = 1	(base outcome)
Treatment level = 2	
Constant term	-0.5545 (0.1771)*
Altitude (ALT) (m)	0.0004 (0.0001)*
Utilized Agricultural Area (UAA) (ha)	0.0099 (0.0013)*
Livestock Units (LU)	-0.0002 (0.0002)
Horse Power (HP)	0.0020 (0.0003)*
Fixed Costs (on Net Value Added) (FC)	-0.0002 (0.0015)
AGE (of the holder) (years)	0.0003 (0.0028)
Annual Working Units (AWU)	-0.1678 (0.0239)*
Treatment level = 3	
Constant term	-1.179 (0.2036)*
Altitude (ALT) (m)	0.0000 (0.0002)
Utilized Agricultural Area (UAA) (ha)	0.0162 (0.0014)*
Livestock Units (LU)	-0.0009 (0.0003)*
Horse Power (HP)	0.0019 (0.0003)*
Fixed Costs (on Net Value Added) (FC)	-0.0138 (0.0107)
AGE (of the holder) (years)	0.0094 (0.0032)*
Annual Working Units (AWU)	-0.2221 (0.0308)*
Treatment level = 4	
Constant term	-1.491 (0.2144)*
Altitude (ALT) (m)	0.0000 (0.0002)*
Utilized Agricultural Area (UAA) (ha)	0.0189 (0.0014)*
Livestock Units (LU)	-0.0031 (0.0006)*
Horse Power (HP)	0.0027 (0.0003)*
Fixed Costs (on Net Value Added) (FC)	-0.0043 (0.0050)
AGE (of the holder) (years)	0.0132 (0.0033)*
Annual Working Units (AWU)	-0.2856 (0.0346)*
Treatment level = 5	
Constant term	-2.159 (0.2623)*
Altitude (ALT) (m)	0.0002 (0.0002)
Utilized Agricultural Area (UAA) (ha)	0.0217 (0.0015)*
Livestock Units (LU)	-0.0057 (0.0012)*
Horse Power (HP)	0.0034 (0.0004)*
Fixed Costs (on Net Value Added) (FC)	0.0007 (0.0026)
AGE (of the holder) (years)	0.0192 (0.0038)*
Annual Working Units (AWU)	-0.4235 (0.0533)*
Treatment level = 6	
Constant term	-2.793 (0.3204)
Altitude (ALT) (m)	0.0003 (0.0002)
Utilized Agricultural Area (UAA) (ha)	0.0235 (0.0015)*
Livestock Units (LU)	-0.0077 (0.0018)*
Horse Power (HP)	0.0036 (0.0004)*
Fixed Costs (on Net Value Added) (FC)	-0.0010 (0.0045)
AGE (of the holder) (years)	0.0191 (0.0047)*
Annual Working Units (AWU)	-0.3867 (0.0568)*
Treatment level = 7	
Constant term	-1.405 (0.2651)*
Altitude (ALT) (m)	0.0004 (0.0002)*
Utilized Agricultural Area (UAA) (ha)	0.0247 (0.0015)*
Livestock Units (LU)	-0.0168 (0.0024)*
Horse Power (HP)	0.0049 (0.0003)*
Fixed Costs (on Net Value Added) (FC)	-0.0127 (0.0130)
AGE (of the holder) (years)	0.0050 (0.0039)
Annual Working Units (AWU)	-0.5845 (0.0584)*

*Statistically significant at 0.05 level