Paper prepared for the 122nd EAAE Seminar
"EVIDENCE-BASED AGRICULTURAL AND RURAL POLICY MAKING: METHODOLOGICAL AND EMPIRICAL CHALLENGES OF POLICY EVALUATION"
Ancona, February 17-18, 2011

Evaluating the CAP Reform as a multiple treatment effect: evidence from Italian farms

Roberto Esposti[1]

[1] Department of Economics, Università Politecnica delle Marche, Ancona (Italy)

r.esposti@univpm.it

Evaluating the CAP Reform as a multiple-treatment effect: evidence from Italian farms

Roberto Esposti

Abstract

This paper aims at evaluating the impact of the 2003 CAP reform on farm production choices. The 2003 Reform of the Pillar I and the Pillar II measures are considered as two distinct but interacting treatments eventually generating the expected outcome, that is, market (ri)orientation of farmers. The outcome of "market orientation" is measured by considering both the short-term production choices and the long-term investment decisions. The Average Treatment effect on the Treated (ATT) is estimated through alternative versions of the Propensity Score Matching (PSM) estimator. Results show that the 2003 Reform of Pillar I actually had a role in (ri)orienting short-term farm production decision and this effect is significantly reinforced, especially in investment decisions, when Pillar II measures are also taken into account. Pillar I reform seems to prevalently affect short-run production decisions while Pillar II support, when present, influences long-run choices (investments).

Keywords: Common Agricultural Policy, Farm Production, Treatment Effects, Propensity Score Matching,

JEL classification: Q18, Q12, C21, O13

1. INTRODUCTION

In 2003 the CAP underwent what can be considered the most radical reform of its half-century history. The so-called Fischler Reform (Henke et al., 2011) substantially changed the way support is delivered to farms conditional on their production activities (Pillar I). The decoupling of Pillar I support was the hinge of that reform. Among the different policy objectives of decoupling, one of the most relevant declared purposes was (ri)orientation to market, that is, to eliminate the distorting effect that the precedent Pillar I support had on production decisions and leave the farmers free to produce what they consider be more profitable for them in the market (European Commission, 2011).

This reform was actually enforced in 2005 (at least in Italy, the country under study here) when the Pillar II of the CAP (Rural Development Policy; RDP) was reformed, in turn. Though much less radical, even this latter reform was intended, among other purposes, to support farmers in taking investment decisions to improve their economic performances in terms of efficiency, productivity and capacity to generate income (European Commission, 2005). In both the 2000-2006 and 2007-2013 programming periods, some of the Pillar II measures (particularly under Axis 1) specifically support farm investments favouring market orientation.

In practice, even though we may agree on the fact that the 2003 CAP reform had clear-cut objectives, an appropriate empirical *ex-post* evaluation of the impact of this reform encounters several serious problems. First of all, the analyst faces the typical problem of programme evaluation with observational (non-experimental) data, that is, the issue of *selection bias* or *selection on unobservables* (Nichols, 2007). It arises because, in any non-experimental setting, we can not exclude that policy treatment and outcome are correlated *ex-ante* through unobservable characteristics that affect, at once, the probability of a farm to be treated and its observed outcome. A second empirical issue concern the fact that the policy outcome itself may be neither obvious nor easily measurable. In particular, the policy impact can be actually revealed by multiple outcomes each of them expressing a different implication of the policy on farmer behaviour. A third complication regards the nature itself of the CAP, since it is actually delivered at the farm level not as a single treatment but as a whole menu of measures whose effects may either compensate or interact. Analysing the farm-level impact of the CAP reform, therefore, takes the form of estimating the treatment effect in a *non-experimental*, *multitreatment* and *multioutcome* setting.

The main purpose of the present paper is to evaluate the impact of the 2003 reform of Pillar I on farmers' production choices. The research question the paper aims at answering, therefore, is whether and to what extent the decoupling of Pillar I support really oriented farmers to market. This evaluation is performed by trying to recreate a quasi-experimental situation where a group of treated farms are compared to a group of non-treated (or control) farms. Such comparison, however, has to explicitly consider, on the one hand, that the treatment effect may be represented by different outcome variables all expressing different aspects of production decisions (*multiple outcomes*) and, on the other hand, that such policy treatment is often accompanied by the presence of other policy measures and, in particular, of Pillar II measures that may have an impact on production choices as well (*multiple treatments*).

2. THE SAMPLE

To create a quasi-experimental setting the first critical issue is the selection of the sample to be considered in the empirical analysis. As the objective here is to assess the impact of the 2003 Pillar I reform on Italian farm's choices, this sample has to satisfy some specific requisites. It has to be a representative random sample. Sample farms have to be observed over the pre and post-treatment periods, therefore the sample must be a balanced panel not just a cross-sectional sample. It must contain all the needed information about the outcomes, the treatments and the so-called confounding variables (see below).

These conditions can be met by extracting a constant sample of farms yearly observed over the pre and post-2005 period. This balanced panel is extracted from the FADN (RICA) database. Though FADN database also covers years prior to 2003, the sampling and data collection procedures and criteria do not allow reconstructing a balanced panel backward. Moreover, 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). Thus, a 2001-2007 comparison, for instance, would overlap different CAP reforms and would mix-up different policy treatments. Year 2008 could also be added but some significant changes in FADN data collection would make year-by-year comparison more difficult. Moreover, the huge price turbulence observed in agricultural markets in 2008 (Esposti and Listorti, 2010) suggests particular caution in adding this year to the post-treatment period. Farmers' behaviour, as well as farms' performance, might be strongly affected by this price bubble and this year could confound permanent responses due to policy treatment with those temporarily induced by peculiar market conditions.

To preserve to the maximum possible extent a quasi-experimental setting some considerations on FADN sample are required. First of all, 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).[1] 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). A second aspect to be considered here is that the Italian RICA sample is not entirely obtained drawing a random sample from this reference population. A small part of the sample is actually constituted by voluntary participation of farms. Nonetheless, it is possible to extract from the larger FADN constant sample a restricted sample that can be actually considered as a random extraction from the underlying FADN reference population according to the ISTAT (the Italian Institute of Statistics) criteria (Cagliero et al., 2010). This sub-sample contains 6542 farms observed over years 2003-2007. This balanced panel constitutes the sample on which the present analysis is performed.

Figure 1 displays the geographical distribution of these sample farms over the Italian provinces (NUTS III level). It may be noticed that the 6542 farms are quite homogeneously distributed across the national territory. Though the sample may tend to concentrate in some specific provinces and, thus, the across-province distribution may be biased and not representative, 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.

3. THE CAP REFORM: MULTIPLE OUTCOMES AND MULTIPLE TREATMENTS

3.1. Multiple Outcomes

As the objective of the present evaluation exercise is to assess the impact of Pillar I 2003 reform on farm production choices, a major problem arises in the appropriate definition of a synthetic variable expressing such choices. Roughly speaking, farmers' production decisions are expressed by how farm resources (land, labour, capital and investments) are allocated and farm performances (revenues, value added, profits) are distributed over the vector of (potential)

---

[1] According to 2000 Census data, more than 82% of Italian agricultural holdings has an economic size smaller than 8 ESU and they account for more than 27% of total Italian agricultural area (Sotte, 2006).

production activities (output vector). By definition, therefore, production decisions can be hardly expressed by a scalar variable. Moreover, for any element of this vector the production decision itself can take a different form. The decision can just be to produce or not that 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 standards and so on. Moreover, whatever this decision eventually is, its timing may be different. Decision about the introduction of a new perennial crop in the farm output vector (for instance wine production) implies a long-term horizon; what we may appreciate in the short-term, in such case, 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 input expenditure or higher revenue related to that specific crop.

As the present analysis regards the impact of 2003 Pillar I reform in (ri)orienting farmers' to market, such policy treatment is expected to affect production decisions. As mentioned above, the hypothesis is that decoupling leaves farmers free to adjust their production decisions, therefore their output vector, to the market conditions to achieve better economic performances (Moro and Sckokai, 2011). This is explicitly mentioned in many EU Commission document and, therefore, it can be considered as the main declared objective of the reform itself: "the next movement towards market orientation for the European agricultural sector came in 2003, when a major overhaul of the CAP was undertaken. […] The current decoupled direct payment […] ensures that farmers respond to market signals while providing income support" (European Commission, 2011, p. 6).

In more technical terms, we may reasonably assume that one of the most significant impact expected from the policy treatment is to improve farm's allocative efficiency. While the impact on technical efficiency (and technological change) may be more controversial (Bartolini et al., 2011) and indisputably takes more time farms are expected to immediately react to decoupling by reorienting their production decisions given their individual characteristics and market conditions (i.e., prices). This search for allocative efficiency may even imply extensivisation, that is, selection of production activities that imply a lower use of inputs (labour and capital, in particular) and, therefore, lower production costs.

Nonetheless, as mentioned, this response may be of different entity and take different forms. On the one hand, the extent to which such reorientation to market is accomplished is questionable. Several studies (see Renwick and Revoredo-Giha, 2008, for instance) show that, though decoupled, Pillar I support still acts as a cross-subsidisation of pre-existent farm activities. On the other hand, reorientation may imply different choices: introducing a "new" (for the farm) production, skipping an "old" one, reallocating resources across existing productions, investing to have a new product in the future.

To take into account this multiple nature of policy outcomes, different outcome variables are taken into account here. 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 production decisions) between the period before (2003-2004) and after (2005-2007) the treatment. The second type considers the investment behaviour (thus, production decisions oriented towards a longer-term programming horizon). In the first case we consider two outcome variables both expressed in terms of distance between the two (before and after the treatment) output vectors. This distance is computed using two different metrics:

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

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

where: *i=1,…,N* indexes the i–th farm within the sample, *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: A indicates the pre-treatment period (2003-2004 average) and B the post-treatment period (2006-2007 average).[2] 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 firms whose revenue distribution across potential products remains the same between pre and post-treatment periods. In such case, the treatment did not induce any effect on production decision. The maximum value, on the contrary, is taken by those firms 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 but also for the degree of specialization of the given farm, the intuition being that the maximum treatment effect is observed in those hyper-specialized farms that decide to completely change their specialization over the treatment period.

As d_{ik} is a dummy variable, it is $0 \leq y_i^2 \leq 1$. Even in this case, the outcome variable increases as the change in the output vector increases. The 0 value is taken by farms for which all productions observed in the pre-treatment period are confirmed in the post-treatment period and no other activity is added. In this case, however, specialization does not tend to increase the 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

---

[2] 2005 being considered an year of transition, it is excluded from the calculation of the *ex-post* outcomes.

changes in production decisions induced by the treatment that take the form of an extension (reduction) of an activity over a continuous domain.

Apparently, y_i^2 is a less accurate measure of the treatment outcome than y_i^1. This latter, however, may encounter a major drawback in the fact that 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 a treatment effect what actually is an exogenous movement of prices. For this reason, in the present exercise, both outcome variables expressing distance in production decisions are maintained.

The outcome variable expressing the treatment effect on investment decisions may take two forms, as well. In both cases, the idea simply is that the treatment may induce extra (that is, more than "business-as-usual") investments that allow the farm to activate (or extend) new (or existing) activities in the longer-term. Therefore, the outcome measuring the treatment effect could simply be the change in investment expenditures of the i-th farm (I_i) before and after the treatment. The amount of this extra-investment can be measured in two different ways:

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

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

The former outcome variable simply is the difference between the yearly average total investment expenditure before (2003-2004 yearly average) and after (2006-2007 yearly average) the treatment. The latter variable does not calculate this difference in absolute terms but in relative terms, that is, as investment rates given by the ratio between total investment expenditure and the respective farm value added.[3] This latter outcome 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 (see Table 2). In doing this, however, it may partially sterilize the effect of the treatment on investment decisions over the whole sample as a real increase in investment levels concentrated in larger farms may be entirely compensated by a decline in investment rates in smaller farms. Even for this second category of outcome variables, therefore, both alternatives are maintained.

3.2. Multiple Treatments[4]

Recreating a quasi-experimental setting within an observational sample requires particular caution in how the sample is extracted and how it is articulated between treated and non-treated

[3] The value added rather the value of production is here considered as it can be more properly considered a proxy of farm profits, that is, of the capacity to generate surplus from which further investments can be made.

[4] Another application of the treatment-effect concepts to CAP impact analysis can be found in Esposti (2007).

(or control or counterfactual) farms. Sample extraction has been already discussed in section 2. Here we want to clarify how the sample has been sorted out among treatment groups. There are two basic problems in achieving the appropriate definition of the treatment groups. The first problem is how a control group can be defined. The second problem is how to deal with multiple treatments.

With respect to the first concern, it is worth reminding that, according to the declared objectives of the present paper, the treated farms are those that experienced the change in Pillar I regime (the introduction of the decoupled Single Farm Payment instead of coupled *per ha* or *per head* payments) starting in 2005. The control group, therefore, should be made of farms that did not underwent that change in regime. This situation would mimic in observational data the purely hypothetical experimental setting where the same farms are alternatively treated and non-treated. Such circumstance could be recreated by randomly assigning the treatment to randomly drawn farms. Unfortunately, even though we may accept that the sample is really a random one (see above), the treatment is not randomly assigned. Participation to the treatment it is not event voluntary as it depends, in Italy, on the history of the individual farm and on the respective support it received in the 2001-2003 period. Farms can not decide to do not move to the new regime, as remaining in the old regime is not admitted. However, 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. The control group, therefore, is the no-policy group, as it includes those farms for which no CAP support has been observed over the entire 2003-2007 period.

On the contrary, the treatment group is made of farms that experienced a change in support in 2005. This, however, takes us to the second problem as, in fact, there is no such a "Pillar I treatment" in real observations. Pillar I support is actually made of many different measures. This is evidently true in the pre-treatment period when support was mostly delivered through almost 30 Common Market Organisations (CMOs). But even in the post-treatment period, though the SFP clearly became dominant, Pillar I support at the farm level is still made by a set of different coupled and decoupled possible measures. Moreover, additionally to Pillar I measures support may be delivered through Pillar II measures. Not only there are many possible Pillar II measures, but any given farm may receive, at the same time, Pillar I and Pillar II measures. In other words, and beside the two pillars, the CAP (both in 2003-2004 and 2005-2007 periods) is a multiple-measure policy. Therefore, an appropriate treatment effect assessment should explicit consider this complex multiple-treatment nature of the CAP.

Figure 2 reports, for the 5 years under consideration, the percentage distribution of the total support within the sample among the most significant (those with more than 1% on total support) measures.[5] Even if we exclude the negligible measures, the fragmentation of the

[5] See Annex 1 for a description of the measures reported in Figure 2.

support remains both in Pillar I and in Pillar II even after the introduction of the simplifying SFP scheme with the 2003 reform. Moreover, though OCM and SFP payments are clearly dominant, some Pillar II measures show a considerable overall expenditure. A so complex multi-treatment setting is hardly empirically affordable (Frölich, 2004). To simplify the approach, however, we can treat Pillar I measures as an unique aggregate, though articulated in different measures with the common objective, after the 2005 reform, of market orientation.

An issue arises, however, with respect to Pillar II measures. It must be acknowledged that farms' production decisions are affected by Pillar II measures, as well, in two possible directions. Axis 1 measures, by supporting competiveness and structural adjustment, may directly imply investments or production choices that are themselves oriented toward allocative efficiency thus overlapping with the impact of Pillar I reform.[6] Other measures (especially in Axis 2), on the contrary, may actually represent constraints to a rapid adjustment towards a more efficient output vector. Both pillars, therefore, may impact on production decisions and, then, on the observed treatment outcomes. In addition, they may even reciprocally interact (positively or negatively, that is, reinforcing or reciprocally offsetting) with respect to the expected outcome such that a simultaneous treatment can generate a remarkably different effect compared to the two separate treatments.

As Pillar II measures may clearly interact with production decision it would be of particular interest to assess which kind of contribution Pillar II measures actually provide to achieve the Pillar I objective of market orientation. The timing and the nature of Pillar II reform, however, was substantially different. The RDP was reformed in 2005 and implemented only in 2007. On the one hand, it did not radically change the way the support is delivered and its fundamental objectives. Axes, measures and actions themselves are partially analogous between the two programming periods (2000-2006 vs 2007-2013) (Esposti, 2011). On the other, using farm-level 2003-2007 data we do not really observe the new (2007-2013) RDP in action. Even in 2007 farm data mostly report Pillar II funding that still refers to measures of the former period (2000-2006). The treatment under evaluation here is the change in policy regime but not policy regime change is actually observed for Pillar II support. Therefore, the two policy treatments must remain separate.

On the basis of this consideration, the observed farms are eventually distinguished in three groups:

- *the control group (C farms)*: farms that did not receive any kind of CAP support (either Pillar I or Pillar II) over the whole period of investigation (2003-2007)
- *the "Pillar I regime change" treatment group (T^I farms)*: farms that experienced the shift in Pillar I support starting from 2005 but that received no Pillar II support over the whole period of investigation

[6] "Measures relating to structural adjustment of farming […] enhancing the economic viability of agriculture through investment and modernisation" (European Commission, 2011, p. 8).

- *the "Pillar I regime change with Pillar II" treatment group (T^{II} farms)*: farms that experienced the shift in Pillar I support starting from 2005 and also received at least one Pillar II payment over the whole period of investigation

It is a multiple (double) dichotomous treatment setting where the research questions are the following: which is the impact of the 2005 reform on T^{I} farms? Which is the impact of the reform combined with Pillar II measures on farms T^{II}? In the quasi-experimental setting here adopted the answer to these questions can be provided by estimating the *Average Treatment Effect on the Treated* (ATT) in the sequence of comparisons C - T^{I} , T^{I} - T^{II} and C - T^{II}. The latter two comparisons provide alternative answers to the second question depending on the control group. Thus, C always operates as control group, T^{II} always as treatment group, while T^{I} alternatively operates as control and treatment group.

Table 1 illustrates how the sample farms distribute across the three groups above and the amount of the CAP support (both Pillars) delivered to them. The largest group is T^{I} that counts for about 58% of the total number of farms, while C and T^{II} accounts for about 14% and 28%, respectively. Therefore, all the abovementioned group comparisons are unbalanced, with the number of control units being different than the number of treated units. An appropriate methodology to estimate the ATT, therefore, must admit comparison with replacement (see below). The treatment intensity (though not explicitly considered here) strongly varies within the treated sample. However, if we measure it in terms of support per unit of value of production rather than of absolute support, this heterogeneity evidently falls.[7] The overall support per farm increases by about 17% from 2003 and 2007 in nominal terms, though this increase is lower than 10% in real terms. The dispersion of support, on the contrary, is reducing though remains remarkably high. This reduction of variability can be only partially attributed to the decoupling of Pillar I support that, in fact, stabilized the pre-reform differences among farms. The group T^{II} support shows a more significant reduction of dispersion and, in any case, variability is sensibly lower for group T^{II} payments.

Overall, group T^{II} receives a larger support, on average. This difference increases over years and might seem obvious given group definition (these farms receive an additional treatment). It can be partially explained, however, also by the fact that farms of group T^{II} show a larger size on average (see Table 2) and, therefore, Pillar I support is itself expected to be larger in group T^{II}. The average increase of support from 2003 to 2007 is mostly due to increase in T^{II} support, that is, to increase in Pillar II payments. At the same time, however, even in 2007 we observe this counterintuitive evidence that payments designed to support investments (those additionally observed in T^{II}) are, in fact, less variable than payments designed to support income (those exclusively observed in T^{I}). As emerges in Table 2, this larger variability of support in group T^{I}

[7] If we divide the support by the farm's value of production, the coefficient of variation becomes 1.9 in 2003 and 2.4 in 2007.

can not be attributed to larger variability in farms' size but should be almost exclusively attributed to how Pillar I payments are allocated.

4. THE AVERAGE TREATMENT EFFECT ON THE TREATED (ATT): PROPENSITY SCORE MATCHING (PSM)

4.1. Identification issues: Unconfoundness and ATT

Let $y_{i,T}^{n}$ indicate the n-th (*n=1,…,4*) outcome variable (see section 3.1) observed in the i-th farm and *T=0*, *I* and *II* the three policy treatments outlined above (i.e., comparisons *C* - T^{I} , T^{I} - T^{II} and *C* - T^{II}), respectively. Let's assume that the attribution of a given treatment to the i-th unit (farm) does not affect the treatment effect on the j-th farm. This assumption, called *stable-unit-treatment-value assumption,* is always maintained in conventional treatment effect analysis and seems largely plausible here and in all other cases when micro data are used and treatment assignment to single units may hardly have aggregate (or macro) effects.
By Average Treatment Effects (ATE) we intend the following expected values:[8]

(5) $ATE_0 = E\left(y_{i,I}^{n} - y_{i,0}^{n}\right)$, $ATE_I = E\left(y_{i,II}^{n} - y_{i,I}^{n}\right)$ and $ATE_{II} = E\left(y_{i,II}^{n} - y_{i,0}^{n}\right)$

These ATE actually express the difference that would be observed in the n-th outcome in a purely experimental (or randomization) situation, that is, as the i-th farm were observed, in sequence, under the non-treatment and treatment situations. 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; Imbens and Wooldridge, 2009). Moreover, with observational data, identifying and estimating the *ATE* could be difficult and non necessarily meaningful. In these cases, the actual research question is: which is the impact of the treatment on the outcome observed in treated units? The answer to this question is rather provided by the ATT:

(6) $ATT_0 = E\left(y_{i,I}^{n} - y_{i,0}^{n} \middle| T = I\right)$, $ATTE_I = E\left(y_{i,II}^{n} - y_{i,I}^{n} \middle| T = II\right)$, and $ATT_{II} = E\left(y_{i,II}^{n} - y_{i,0}^{n} \middle| T = II\right)$

where the answer about the effect of the treatment only concerns the units that were actually treated and does not apply to units that were not treated. Frölich (2004) and Imbens and Wooldridge (2009) provide a clear explanation on why in non-experimental settings $ATE_T \neq ATT_T$. As we can write $ATE_0 = E\left(y_{i,I}^{n} - y_{i,0}^{n}\right) = E\left(y_{i,I}^{n}\right) - E\left(y_{i,0}^{n}\right)$ and

[8] In this form, this is also called Population Average Treatment Effect (PATE) as it refers to the expected value within the whole population. For this reason, in (5) we can write ATT_0 rather than ATT_{i0}: as *i* indexes the randomly drawn unit in the population, it is $ATT_0 = ATT_{i0}$. In the sample, however, we may have heterogeneous treatment effects and, therefore, $ATT_0 \neq ATT_{i0}$. In such case ATT_0 is then calculated averaging across the sample units. When ATE is calculated as a sample average it is also called Sample Average Treatment Effects (SATE) (see Abadie et al., 2004, for details). In the present application, the ATE has to be always intended as SATE though, for simplicity, in the notation we do not explicit the *i* index .

$ATT_0 = E\left(y_{i,I}^n - y_{i,0}^n \middle| T = I\right) = E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = I\right)$, *ATE* and *ATT* are equal only if $E\left(y_{i,0}^n\right) = E\left(y_{i,0}^n \middle| T = I\right)$ and $E\left(y_{i,I}^n\right) = E\left(y_{i,I}^n \middle| T = I\right)$, that is, only if the expected value of a given treatment is independent on the subsample (treated units or counterfactuals) on which we are measuring it. But, in non-experimental settings, this is granted only under specific (and strong) assumptions. Even more important is that, unless these assumptions are made, in non-experimental settings both the *ATE* and the *ATT* remain unidentified (Frölich, 2004; Imbens and Wooldridge, 2009).

If we focus on $ATT_0 = E\left(y_{i,I}^n - y_{i,0}^n \middle| T = I\right) = E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = I\right)$ we may easily understand where this identification issue comes from. The problem is that, in observational data, we can not observe the counterfactual or potential outcome $E\left(y_{i,0}^n \middle| T = I\right)$, that is, the outcome that would be observed if the treated units were not treated. What we really observe is only $E\left(y_{i,0}^n \middle| T = 0\right)$. In practice, with our observations we can compute the difference $E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right)$ but this difference does not necessarily correspond (i.e, does not identify) the *ATT* as it is:

(7)

$$E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right) = \left[E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = I\right)\right] - \left[E\left(y_{i,0}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right)\right] = ATT_0 - \left[E\left(y_{i,0}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right)\right]$$

The latter term of the right-hand side is the so-called *selection bias* as it corresponds to the difference between what we can observe ($E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right)$) and what we want to estimate (*ATT*). This bias occurs whenever a difference in the outcome would be observed between the treated units and the control units regardless the treatment itself. There is some unobserved difference between the two groups that, at the same, conditions the participation to the treatment and the outcome regardless of the treatment. As the selection bias depends on the presence of some unobserved characteristics, it is also called *selection-on-unobservable bias* or *omitted variable bias*.

Getting rid of this bias is the key issue of the estimation of the *ATT* (or *ATE*). This is the so-called identification issue (Frölich, 2004; Imbens and Wooldridge, 2009) which consists, in practice, in finding ways to make the term $\left[E\left(y_{i,0}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right)\right] = 0$ and, consequently, $E\left(y_{i,I}^n \middle| T = I\right) - E\left(y_{i,0}^n \middle| T = 0\right) = ATT_{i0}$. Imbens and Wooldridge (2009) and Nichols (2007) provide excellent reviews of the alternative methodological solutions put forward over time by treatment-effect econometrics.[9]

---

[9] An extensive analysis of these alternatives is beyond the scope of this paper. It is worth reminding, however, that a widely adopted approach in programme evaluation is the so-called Difference-In-Difference (DID) identification and estimation of the ATT. Such approach is affordable whenever the outcome is repeatedly observed over time in both

Given the specific characteristics of available data and of the policy under evaluation here, the methodological solution that seems more suitable is the approach that more directly tackle the issue of selection on unobservables. This approach can be generally (and generically) referred to as *selection-on-observables* approach. The idea is that, though a selection bias may be observed, the analyst is in the condition to detect and observe all the pre-treatment variables or characteristics **X** that generate it. In practice, the identification of the ATT is achieved by assuming that:

(8) $ATT_0 = E\left(y^n_{i,I} \middle| \mathbf{X}, T = I\right) - E\left(y^n_{i,0} \middle| \mathbf{X}, T = 0\right)$ as $\left[E\left(y^n_{i,0} \middle| \mathbf{X}, T = I\right) - E\left(y^n_{i,0} \middle| \mathbf{X}, T = 0\right)\right] = 0$

This is the so-called *Conditional Independence Assumption* (CIA) or *Unconfoundness Assumption* as we are assuming that, once we control for all relevant pre-treatment covariates **X**, we recreate the condition of a randomized experiment and the ATT can be estimated by directly computing the difference between the observed outcome of the treated and of the control units. Vector **X** is expected to contain all the pre-treatment variables that are, at the same time, correlated to the treatment assignment and to the outcome variable: once we control for them the difference in the outcome can be exclusively attributed to the treatment. This identification assumption can be somehow tested but still remains the critical point of this approach as in no case we can definitely exclude that an unobserved confounding variable (i.e., correlated with both the treatment assignment and the outcome variable) still exists.

If we are willing to accept unconfoundness, however, identification is achieved and the problem becomes how to estimate the conditional expected values in (8), that is, $E\left(y^n_{i,I} \middle| \mathbf{X}, T = I\right)$ and $E\left(y^n_{i,0} \middle| \mathbf{X}, T = 0\right)$. A natural way could be to specify a regression model, that is, estimate $E\left(y^n_{i,I} \middle| \mathbf{X}, T = I\right)$ and $E\left(y^n_{i,I} \middle| \mathbf{X}, T = I\right)$ by regressing $y^n_{i,I}$ on **X** and the treatment-assignment variable (a dummy variable). Nonetheless, this parametric approach finds two major drawbacks. First of all, it has to assume a parametric linear specification of the relation between $y^n_{i,I}$ on **X** while this relation may be more complex and may vary over the sample. Secondly, such approach estimates the ATT by using all the observed variation of covariates **X** while, in fact, it may be the case that only a portion of this range of variation is common to both treated and control units. This common portion is also called *common support* and it is intuitively (see

---

the treated and the control units, that is, in practice, when a panel data set is available. Moreover, it is based on the assumption that the spontaneous (that is, independent on the treatment) dynamics of the outcome variable is the same for both groups (for this reason this assumption is also called *parallel trends assumption*). This assumption seems particularly strong in the present application as, in fact, the structural differences between the control and the treated groups do imply different dynamics over the observed period (for instance due to different market conditions). Moreover, the DID approach requires appropriate time series of the outcome variables as these must be repeatedly observed over time. In the present case, however, this condition is not met. The available dataset is, in fact, a panel but the treatment outcome is not repeatedly observed. The treatment being a policy regime change occurring in 2005, all the proposed observed outcomes (see section 3) are variations observed only once. The way the analysis is constructed makes the DID approach, as well as all other identification solutions based on pre and post-treatment comparisons, not suitable in the present application.

Imbens and Wooldridge, 2009, for more technical details) more appropriate to limit the estimation of the conditional expectations in (8), thus of the ATT, to this common support.
For these two reasons, the recent empirical treatment-effect literature tends to prefer nonparametric approaches to the estimation of (8). This is achieved through matching estimators. Among this, the Propensity Score Matching (PSM) estimator has become very popular.

4.2. PSM: Alternative Estimators

Generally speaking, matching is a statistical procedure aiming at bringing observational (non-experimental) data back to an hypothetical experimental setting. The idea consists in pairing each treated observation with one or more control units on the basis of the observed covariates. Matching looks for pairs (or groups) of treated and control units showing the closest **X** in such a way we can assume that treatment assigned to these pairs (groups) is randomly attributed (i.e., units are equal for all relevant and observable characteristics except for the treatment).
In practice, matching raises two serious empirical issues. The first problem consists in finding a metric (a scalar variable) measuring the distance among observations across the elements of vector **X** (Nichols, 2007). Once this metric has been established, the second problem consists in finding appropriate rules to match treated and non-treated units (or groups) on the basis of this metric. Both steps are computationally more complex the greater is the dimension of vector **X**. At the same time, accounting for a higher number of observed confounding characteristics (thus a larger **X**) guarantees about the validity of the unconfoundness assumption. Therefore, the key empirical problem in matching (also know as the *curse of dimensionality*) is the trade-off between the validity of the identifying assumption and the often unaffordable computational burden to achieve estimation of the ATT.
In their seminal contribution, Rosenbaum and Rubin (1983) proposed a very appealing solution to this empirical problem. They proposed the use of a propensity score as the metric on which matching can be then based. The Propensity Score (PS) of the i-th unit ($p_i^T(\mathbf{X})$) is its probability of being treated with treatment *T* conditional on covariates **X**:

(9) $$p_i^T(\mathbf{X}) \equiv \Pr(T|\mathbf{X}) = E(T|\mathbf{X})$$

The key intuition of Rosenbaum and Rubin (1983) consists in demonstrating that if treatment assignment is random conditional on **X** then, under the CIA, it is also random conditional on $p_i^T(\mathbf{X})$. As $p_i^T(\mathbf{X})$ is a scalar, matching based on $p_i^T(\mathbf{X})$ is empirically much more affordable than matching based on **X**.[10] The ATT can be thus more easily computed as:

(10) $$ATT_0 = E\left(y_{i,I}^n \middle| p_i^I(\mathbf{X}), T = I\right) - E\left(y_{i,0}^n \middle| p_i^I(\mathbf{X}), T = 0\right)$$

[10] For more details and discussion on the asymptotic properties of these (PS) matching estimators see Hahn, 1998, and Imbens and Wooldridge, 2009.

The identification of the ATT on the basis of the PS still depends on the CIA assumption. In this case, however, the assumption must hold on **X** as well as on $p_i^T(\mathbf{X})$. In other words, if unconfoundness on **X** holds true, then the PSM implies that $E\left(y_{i,I}^n \middle| p_i^T(\mathbf{X})\right)$ is independent on **X** as the PS already contains all the information about how **X** conditions the treatment assignment. Unconfoundness on $p_i^T(\mathbf{X})$ is also called the *Balancing Hypothesis* (Becker and Ichino, 2002). The validity of this hypothesis can be tested within the sample by checking whether observations showing a very close PS really have a statistically equal distribution of **X** independently on the treatment status, that is, both in the treatment and control groups.

A further empirical practice that makes the balancing condition more easily met and improves the quality of matching consists in imposing the common support. It means that balancing is assessed only over the range of $p_i^T(\mathbf{X})$ for which we have observation in both the treated and control groups. Then, matching is made only considering units belonging to this common support. It means that in balancing and matching, treated (control) units whose $p_i^T(\mathbf{X})$ does not find a corresponding value in at least one control (treated) unit, are excluded from the analysis. In most econometric packages, balancing is tested by dividing the common support in blocks containing an often different number of treated and control units.

PSM identifies and estimates the ATT in a three-step procedure. Firstly, a parametric binary choice model (also called the Propensity Score Equation, usually taking the form of a conventional binomial probit or logit model) is estimated to obtain estimated $p_i^T(\mathbf{X})$. Secondly, usually imposing common support and once the balancing condition has been assessed,[11] matching on the basis of $p_i^T(\mathbf{X})$ is performed and pair-wise (or group-wise) ATTs are computed as in (10). Finally, the average non-parametric ATT is computed over the whole sample (the common support) as the weighted average (where weights depend on the number of treated units) of the pair or group-wise ATTs.

The second step (matching) can be achieved following two different strategies. The first is matching *strictu sensu*, that is, for any treated unit (or for blocks of units) it consists in looking for the closest control unit(s) in terms of $p_i^T(\mathbf{X})$ (*best matches*). This is the case of *Stratification Matching* (where matching is made on groups or block of units), of the *Radius Matching* (where any treated unit is matched with all control units following within a predetermined distance, or radius *r*, from its own $p_i^T(\mathbf{X})$) and of the *Nearest Neighbour Matching* (where matching is made one-by-one). In this latter case, as the number of treated and control units may be different, replacement is allowed, that is, the same control unit can be the best match for more than one treated unit.

[11] See Becker and Ichino (2002) for more details on this first stage of the PSM estimation.

The second strategy is, in fact, a weighting procedure. Any i-th treated unit is matched with all control units (within the common support) but each is weighted by the inverse of its distance from the $p_i^T(\mathbf{X})$. This is the case of *Kernel Matching*. In practice, however, all these matching procedures actually are kinds of weighting as in three of them matching implies giving weight = 0 to some control units, that is, excluding them from comparison with the given treated unit.
There is no clear-cut and univocal indication on which of the matching approaches should be preferred, though they may be asymptotically, and under unconfoundness, equivalent In practice, in finite sample the Stratification Matching performance is usually poorer compared to the other solutions while Kernel Matching is often preferable. At the same time, given the specific conditions on which matching is performed a trade-off between bias and variance (accuracy) is often observed. For these reasons, presenting results of all these matching procedures may serve as sensitivity analysis to assess the robustness of the ATT estimates.

5. THE EMPIRICAL APPLICATION

5.1. The treated and the control groups: descriptive evidence

Table 2 reports some descriptive statistics by treatment group for the four outcome variables and the pre-treatment covariates (or confounding variables).[12] Covariates are those pre-treatment variables 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. These covariates can be grouped in three categories. First of all, we consider the relevant individual characteristics of the farmer (AGE) and of the farm (Altitude - ALT). Secondly, the economic (ES, FC)[13] and physical (AWU, HP, UAA and, at least partially, LU) size of the farm clearly matters, as well. 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 category of confounding variables, in fact, are those variables that directly express the production specialization of the farm (TF and, in part, LU). The linkage between these covariates and the treatment assignment is evident at least in the case of Pillar I support as this actually concerns those farms that were interested by specific OCM measures while, on the contrary, farms not involved in Pillar I are those whose production specialization was less (or not at all) targeted by specific policy measures. To express farm production specialization, the

[12] Pre-treatment variables have been observed in 2003, the only exclusion being FC for which the 2003-2004 average has been considered since this variable may largely vary on a yearly base.

[13] 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.

4-digit "Type of Farm" (TF) FADN classification is adopted (2000 classification). This qualitative variable, however, 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 Pillar I support than classes 2022 (flowers' production) and 6010 (horticulture); therefore, farms belonging to the former class are more likely to be assigned to group T^I (or T^{II}) 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 Pillar I CAP support. This number expresses a qualitative monotonous variable (TF_R) that increases as the dependency on CAP support declines. Annex 2 details this reclassification of the TF variable.

From table 2 it may be easily appreciated that for the first three outcome variables, the average values tend to increase by moving from the control group (*C*) to the treatment groups (T^I and T^{II}). The forth outcome variable actually increases passing from the control to the first treatment group but then it declines in the second treatment group. Nonetheless, though these differences in outcome variables may be consistent with expectation in terms of policy treatment effect, the most significance evidence is their very high variability in the whole sample, as well as in any single treatment group. In practice, if we constructed a conventional 95% confidence interval around the sample average in any treatment group we would notice that these intervals are largely overlapping across the groups for all outcome variables. Therefore, looking at these simple statistics, there is no clear evidence of a significant difference in any outcome variable across the treatment groups.

In the case of some covariates, treatment groups' averages move monotonically, either increasing (variables ALT, HP, LU, UAA) or decreasing (FC), from group *C* to group T^{II}. In other cases (AGE, AWU, ES, TF_R), this monotonicity is not observed. Generally speaking, groups T^I and T^{II} show closer characteristics between them compared to group *C*. In this latter case, the physical size of farms tends to be smaller but this is not necessarily true if we consider the economic size. Moreover, production specialization of group *C* is evidently less dependent on Pillar I CAP support (see TF_R), as could be expected by the fact that this group is excluded from CAP treatments, and it practically excludes livestock activities. The difference between T^I and T^{II}, on the contrary, is negligible.

Nevertheless, even for most 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. The only exceptions are variables AGE, ES and TF_R. In the former two variables, however, the difference between treatment groups is relatively small and not particularly informative. Difference observed in TF_R, on the contrary, is explicative of the fact that the treatment assignment strongly depends on the production specialization of the farms. This confirms that any treatment-effect assessment must be conditioned on such specialization, that is, should be performed comparing units (farms) with similar specialization.

5.2. Results[14]

5.2.1. The estimated Propensity Score (PS)

Table 3 reports the estimates of the PS equation. The specification adopted is a conventional binomial probit with the covariates as explanatory variables and the treatment assignment as the dependent binary variable. The estimation is repeated three times, each for any treatment under consideration here (*T=0*, *I* and *II*). For any estimation the balancing property has been checked on the common support. Whenever balancing was not found, the number of blocks was adjusted until balancing was satisfied. Limiting the balancing check to the common support is justified by the fact that the following matching and ATT estimation excludes the units that are outside the common support.

Parameter estimates of the probit equation are in most cases statistical significant. Only variable FC is not significant in any of the three equations. This can be reasonably interpreted with the fact that fixed costs are strongly collinear with other explanatory variables, in particular those expressing the amount of labour and capital. Older farmers seem to have an higher probability to be treated with exclusively Pillar I support while younger farmers, on the contrary, show higher propensity in either being not treated at all by CAP support or receiving support from both Pillars. The altitude operates univocally as the propensity to receive the Pillar I support increases moving from plain areas to hilly and mountainous farms; the same holds true in the case of the double treatment (both Pillars). Evidently, farms that self-excluded themselves from Pillar I support, through their production decisions, are positioned in relatively more favourable areas while, on the contrary, accessing to Pillar II support prevails in farms located at higher altitude. Former farms tend to be more intensive as the propensity to receive *T=0* declines with a larger amount of working units. On the contrary, the latter ones seem to be oriented towards extensiveness in terms of labour use. The increasing substitution of agricultural labour with physical capital (machinery) is evident moving from group *0* to groups *I* and *II* as indicated by the parameter associated to HP.

The propensity score is only marginally affected by the physical size of farms (UAA).[15] Larger farms tend to have an higher propensity for all CAP treatments but this effect is actually negligible and partially misleading. As a matter of fact, whenever the economic size (ES) is taken into account, the propensity score to receive the Pillar I treatment actually decreases but, at the same time, a larger economic size increases the propensity to receive the Pillar II treatment combined with the Pillar I. We can argue that farms with larger economic size (therefore more professional farms) show a dualistic attitude towards policy treatments. On the one hand, we have labour intensive farms of relevant economic size and prevalently positioned

[14] The whole PSM procedure has been performed using software STATA10 and, in particular, modules `pscore, atts, attnd, attnw, attk, attr` (Becker and Ichino, 2002).

[15] It must be noticed, however, that the UAA variable actually enters the PS equation as square transformation (UAA^2). This transformation was needed to satisfy the balancing property.

in plain and well-endowed areas that tend to exclude themselves from Pillar I support due to their specific production specialization. On the other hand, we find capital intensive farms with large economic size that tend to be positioned in less favoured hilly and mountainous areas for which we observe an higher probability to receive Pillar II support together with Pillar payment. In the middle, we have smaller farms that only receive Pillar I payments.

Variables expressing production specialization (LU and, above all, TF_R) confirm that the prevailing production decisions eventually bring about self-selection in terms of treatment assignment. Moving towards TF with a lower dependency on the CAP is associated to a lower propensity score to receive all the treatments. This evidence suggests, on the one hand, that the way TF variable has been reclassified is appropriate. On the other hand, however, it also suggests that farms receiving both policy treatments show a significantly different production orientation compared to farms that are exclusively treated with Pillar I support. This is also confirmed by LU variable suggesting that livestock activities tend to be more present in group T^{II}. In any case, more than any other structural or idiosyncratic characteristic of the farm, production orientation and specialization is the factor that primarily induces self-selection of farms in terms of treatment assignment.

5.2.2. Matching and the estimated Average Treatment effect on the Treated (ATT)

Tables 4 to 6 report the PSM estimation of the *ATT* in the treatments under investigation here. The four alternative matching procedures presented in section 4 are performed and displayed to assess robustness of results with respect to how matching is achieved. Stratification Matching has been performed using the blocks identified in PS equation estimation. Nearest Neighbour Matching is obtained sorting all units by the respective estimated propensity score, and then searching forward and backward for the closest control unit(s). In the case of multiple nearest neighbours, either the forward or backward matches are randomly drawn.[16] Radius Matching is performed by taking a 0.05 radius of the PS, while for Kernel Matching the conventional Gaussian Kernel function is adopted.[17] As illustrated by Becker and Ichino (2002), analytical standard errors can be obtained only for some of these *ATT* estimates. For Stratification and Kernel Matching estimators, in particular, standard errors are estimated through bootstrapping with 1000 replications.

Some regularities among the three estimated *ATT*s emerge. First of all, in the case of the output distance variables y^1 and y^2, results are quite robust across the four matching procedures. Moreover, for these outcome variables *ATT* estimates are in most cases statistically different

[16] In STATA10, this is the `attnd` procedure. An alternative is the `attnw` procedure where forward and backward matches are both taken into account and equally weighted. These alternative estimates do not substantially differ from what presented here and are available upon request. More technical details on these PSM procedures can be found in Becker and Ichino (2002).

[17] The alternative Epanechnikov Kernel function has been also tested by imposing a bandwidth of 0.05. These alternative estimates do not substantially differ from what presented here and are available upon request.

from 0. On the contrary, estimates are more variable and often not statistically significant in the case of the outcome variables related to investment decisions, that is, y_i^3 and y_i^4. Secondly, estimates obtained with Kernel and, above all, Radius matching are systematically higher than *ATT* estimated with the other two matching procedures. Radius matching estimates are also dependent on the definition of the radius as a lower radius may significantly change the results.[18] Limiting the attention to statistically significant *ATT* estimates, all the results are concordant in sign and similar in magnitude (except Radius matching). Therefore, common conclusions can be drawn from the four different matching approaches.

Treatment 0 (i.e., ATT_0) always has a positive impact on inducing a change in the output vector and on inducing investment decisions; in this latter case, however, it is barely statistical significant. If we consider the average value of y_i^1 over the treated group T^I we can also appreciate that this treatment effect is quite strong as it would imply a >50% impact. In the case of treatment *I* (i.e., ATT_I), the impact on the output vector is significant in the case of y_i^1 but it is almost an half compared to treatment *0*. In such case, however, the estimated effect on investment decisions is not only larger but also more clearly significant in the case of variable y_i^3.

Finally, treatment *II* (i.e., ATT_{II}) somehow summarizes the previous two effects, as expected, since group T^{II} undergoes a sort of double treatment compared to group *C*. The treatment effect measured with respect to the distance of the output vector before and after the treatment (i.e., y^1 and y^2) is always statistically significant and exceeds 70%. Moreover, the estimated *ATT* is very robust across the different matching approaches. The same holds true for at least one of the two outcome variable expressing change in investment decisions (y_i^3) with significant, robust and large *ATT* estimates.

The policy interpretation of these results brings us back to the original research questions of the present paper. On the one hand, we can state that results support the idea that Pillar I reform did oriented farms to market as induced treated farm to a stronger change in their output composition. Nonetheless, the impact on investment decisions is much less evident and may be the case that the reform convinced the farms to change their short-run production decisions but did not affect as much the long-run production choices. This may somehow depend on the fact that most farms reduced their investment level over the periods under study, regardless the policy treatment (see Table 2). In a period of persistent market crises or difficulties, many Italian farms suspended long-run investment decisions regardless the change in the policy regime. At the same time, however, it must be noticed that when Pillar II measures are added not only they reinforce the impact of Pillar I reform on short-run decisions; they also, and above all, extend the impact also to investment decisions, that is, to the long-run horizon. The

[18] Estimates under alternative values of the radius are available upon request.

conclusion could be that the capacity of Pillar I reform to orient farms to market is substantially reinforced when it occurs in combination with Pillar II support and also have a more persistent nature since it also affects the long-term farmers' choices. As confirmed by the comparison between groups 0 and T^{II}, Pillar II measures seem to play a strategic role in conditioning the impact of the Pillar I reform, not only because they reinforce the impact on shorter-term production decisions but also because they show complementarity in combining short-run and long-run impacts.

6. CONCLUDING REMARKS

The empirical approach here proposed aims at assessing the effect of the Pillar I 2003 reform on farms' production choice. While results provide quite robust evidence about the effect of the reform and, even more significantly, on the role of Pillar II measures, some steps forward could be proposed with respect to the proposed methodology. First of all, multiple treatment groups could be defined with a more detailed distinction across policy measures. For instance, different measures of Pillar I and, above all, of Pillar II may induce substantially different production choices. As a consequence, a more sophisticated articulation of treatment groups could be attempted. Secondly, and more importantly, the policy treatment could enter the analysis not just as a binary variable (on/off) but as a continuous treatment, that is, explicitly considering the different amount of CAP delivered to the respective farms.
The analysis of multiple continuous treatments is at the forefront of the current treatment-effect 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 matching and estimation techniques could be proposed and attempted in future research.

ACKNOWLEDGMENT

The author is grateful to the National Institute of Agricultural Economics (INEA) for having provided him with the FADN Italian database.

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.

Bartolini1, F., Latruffe, L. and Viaggi D. (2011). Assessing the effect of the CAP on farm innovation adoption. An analysis in two French regions. 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.

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

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

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 Henke, R., Severini, S. and Sorrentino, A. (eds), *The Cap After The Fischler Reform: National Implementations, Impact Assessment And The Agenda For Future Reforms*. Aldershot, UK: Ashgate (forthcoming).

Esposti, R., Listorti, G. (2010). Agricultural Price Transmission Across Space and Commodities. The Case of the 2007-2008 Price Bubble. Paper presented at the XLVII SIDEA Conference "L'agricoltura oltre le crisi", Campobasso (Italy), September 22-25.

European Commission (2005). *Putting rural development to work for jobs and growth*. Special Edition Newsletter, Directorate-General for Agriculture and Rural Development, Brussels.

European Commission (2011). The CAP in perspective: from market intervention to policy innovation. *Agricultural Policy Perspectives Briefs*, Brief nº 1, European Commission, Directorate-General for Agriculture and Rural Development, Brussels.

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

Hahn, J. (1998). On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica*, 66(2), 315–31.

Henke, R., Severini, S. and Sorrentino, A. (eds) (2011). *The Cap After The Fischler Reform: National Implementations, Impact Assessment And The Agenda For Future Reforms*. Aldershot, UK: Ashgate (forthcoming).

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

Moro, D. and 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.

Nichols, A. (2007). Causal inference with observational data. *The Stata Journal*, 7(4), 507-541.

Renwick, A. and 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.

Rosenbaum, P. R. and. Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

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

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

Table 1: Distribution of sample farms across the three treatments and the respective amount of CAP support (both Pillars)

Treatment group:	*C*	*T^I*	*T^{II}*	*Whole Sample*
Number of farms (%)	938 (14%)	3775 (58%)	1829 (28%)	6542 (100%)
CAP support (2003)				
Avg. amount of support (€)	0	12715	19047	12662
Standard Deviation	0	64698	57122	57978
Coefficient of Variation (CV)	0	5.1	3.0	4.6
Minimum	0	0	0	0
Maximum	0	2205000	2004153	2205000
CAP support (2007)				
Avg. amount of support (€)	0	14273	25513	15376
Standard Deviation	0	69936	63783	60735
Coefficient of Variation (CV)	0	4.9	2.5	4.0
Minimum	0	0	0	0
Maximum	0	1635650	622351	1635650

Source: own elaboration

Table 2: Descriptive statistics over treatment groups: sample averages (standard deviation in parenthesis)

Treatment group:	*C*	*T^I*	*T^{II}*	*Whole Sample*
Outcome variables (y_i^n):				
y_i^1 (distance index)	0.169 (0.435)	0.505 (0.653)	0.738 (0.736)	0.522 (0.688)
y_i^2 (distance index)	0.006 (0.014)	0.016 (0.019)	0.018 (0.021)	0.015 (0.019)
y_i^3 (in €)	-12421 (94964)	-3887 (91365)	-1316 (166063)	-4392 (117589)
y_i^4 (in €)	-0.322 (1.382)	-0.165 (3.338)	-0.239 (2.424)	-0.208 (2.890)
Pre-treatment variables (X):				
AGE (of the holder) (years)	51.85 (13.84)	53.95 (15.05)	49.99 (14.07)	52.54 (14.71)
Altitude (ALT) (m)	154.28 (192.11)	234.95 (215.60)	391.81 (359.34)	267.24 (273.79)
Annual Working Units (AWU)	2.79 (6.22)	1.90 (2.71)	2.41 (3.30)	2.18 (3.60)
Economic Size (ES) (classes)	6.41 (2.19)	5.85 (2.42)	6.71 (1.90)	6.17 (2.29)
Fixed Costs (on Net Value Added) (FC)	2.79 (36.41)	2.12 (17.95)	1.30 (9.27)	1.98 (20.00)
Horse Power (HP)	93.23 (129.58)	173.94 (206.31)	206.55 (260.50)	171.49 (217.15)
Livestock Units (LU)	5.73 (50.53)	41.20 (214.38)	54.12 (277.26)	39.73 (220.44)
Utilized Agricultural Area (UAA) (ha)	7.50 (24.34)	30.52 (63.46)	50.17 (85.15)	32.72 (67.93)
Type of Farm (TF) (4-digits)*	3211 (fruits)	1310 (arable crops)	4110 (dairy)	1310 (arable crops)
Type of Farm (reclassified) (TF_R)	5.07 (1.67)	3.40 (1.59)	3.54 (1.40)	3.66 (1.65)

* In this case the Table reports the higher frequency class
Source: own elaboration

Table 3: Estimates of the Propensity Score equation for the three treatments[a] (standard error in parenthesis)

Treatment group:	*T=0*	*T=I*	*T=II*
Common support	[0.107, 1]	[0.064, 0.980]	[0.017, 1]
Avg. estimated PS:			
Total	0.80	0.33	0.66
Group *C*	0.60	-	0.44
Group T^I	0.85	0.29	-
Group T^{II}	-	0.40	0.77
Probit parameter estimates:			
Constant term	1.7409* (0.1408)	-1.4045* (0.1140)	0.5917* (0.1848)
AGE (of the holder) (years)	0.0048* (0.0017)	-0.0068* (0.0013)	-0.0009 0.0021
Altitude (ALT) (m)	0.0009* (0.0001)	0.0013* (0.0001)	0.0018* (0.0001)
Annual Working Units (AWU)	0.0631* (0.0097)	0.0041 (0.0071)	-0.0420* (0.0081)
Economic Size (ES) (classes)	-0.0589* (0.0119)	0.1017* (0.0096)	0.0394* (0.0166)
Fixed Costs (on Net Value Added) (FC)	-0.0017 (0.0010)	-0.0003 (0.0013)	-0.0012 (0.0016)
Horse Power (HP)	0.0024* (0.0002)	0.0002* (0.0001)	0.0021* (0.0003)
Livestock Units (LU)	0.0035* (0.0005)	-0.0001 (0.0001)	0.0026* (0.0006)
Utilized Agricultural Area (UAA) (ha)	0.0001* (0.0000)	0.0000 (0.0000)	0.0001* (0.0000)
Type of Farm (reclassified) (TF_R)	-0.2965* (0.0160)	0.0629* (0.0125)	-0.2674* (0.0199)

[a] The balancing property of the propensity score is satisfied on the common support in all estimates at significance level of 0.005
*Statistically significant at 0.05 level
Source: own elaboration

Table 4: ATT_0 estimates for the four outcome variables (standard errors in parenthesis)[a]

	Stratification Matching	*Nearest Neighbour Matching*	*Radius Matching* [c]	*Kernel Matching*
$y_{i,0}^1$	0.283*[b] (0.061)	0.244* (0.039)	0.337* (0.022)	0.281*[b] (0.056)
$y_{i,0}^2$	0.010*[b] (0.001)	0.009* (0.001)	0.010* (0.001)	0.009*[b] (0.001)
$y_{i,0}^3$	1661 [b] (2068)	1814 (1400)	4165* (1144)	2785 [b] (1555)
$y_{i,0}^4$	0.138 [b] (0.089)	0.129 (0.091)	0.228* (0.082)	0.157*[b] (0.068)

[a] All estimates are performed on the common support
[b] Bootstrap standard errors obtained with 1000 replications
[c] Radius = 0.05
*Statistically significant at 0.05 level
Source: own elaboration

Table 5: ATT_I estimates for the four outcome variables (standard errors in parenthesis)[a]

	Stratification Matching	*Nearest Neighbour Matching*	*Radius Matching* [c]	*Kernel Matching*
$y^1_{i,I}$	0.131*[b]	0.156*	0.241*	0.136*[b]
	(0.024)	(0.031)	(0.022)	(0.032)
$y^2_{i,I}$	0.001[b]	0.001*	0.002*	0.000[b]
	(0.001)	(0.000)	(0.001)	(0.001)
$y^3_{i,I}$	4442*[b]	4723	8030*	4970*[b]
	(1846)	(2680)	(1588)	(1873)
$y^4_{i,I}$	-0.088[b]	-0.176	-0.051	-0.110*[b]
	(0.061)	(0.224)	(0.067)	(0.048)

[a] All estimates are performed on the common support
[b] Bootstrap standard errors obtained with 1000 replications
[c] Radius = 0.05
*Statistically significant at 0.05 level
Source: own elaboration

Table 6: ATT_{II} estimates for the four outcome variables (standard errors in parenthesis)

	Stratification Matching	*Nearest Neighbour Matching*	*Radius Matching* [c]	*Kernel Matching*
$y^1_{i,II}$	0.547*[b]	0.589*	0.561*	0.546*[b]
	(0.042)	(0.053)	(0.026)	(0.044)
$y^2_{i,II}$	0.012*[b]	0.012*	0.011*	0.012*[b]
	(0.001)	(0.002)	(0.001)	(0.001)
$y^3_{i,II}$	10100*[b]	13284*	11539*	11206*[b]
	(3001)	(3342)	(1731)	(2498)
$y^4_{i,II}$	0.056 [b]	0.072	0.130	0.056
	(0.072)	(0.201)	(0.082)	(0.091)

[a] All estimates are performed on the common support
[b] Bootstrap standard errors obtained with 1000 replications
[c] Radius = 0.05
*Statistically significant at 0.05 level
Source: own elaboration

Figure 1 – Distribution of the sample farms across Italian provinces (NUTS III level)

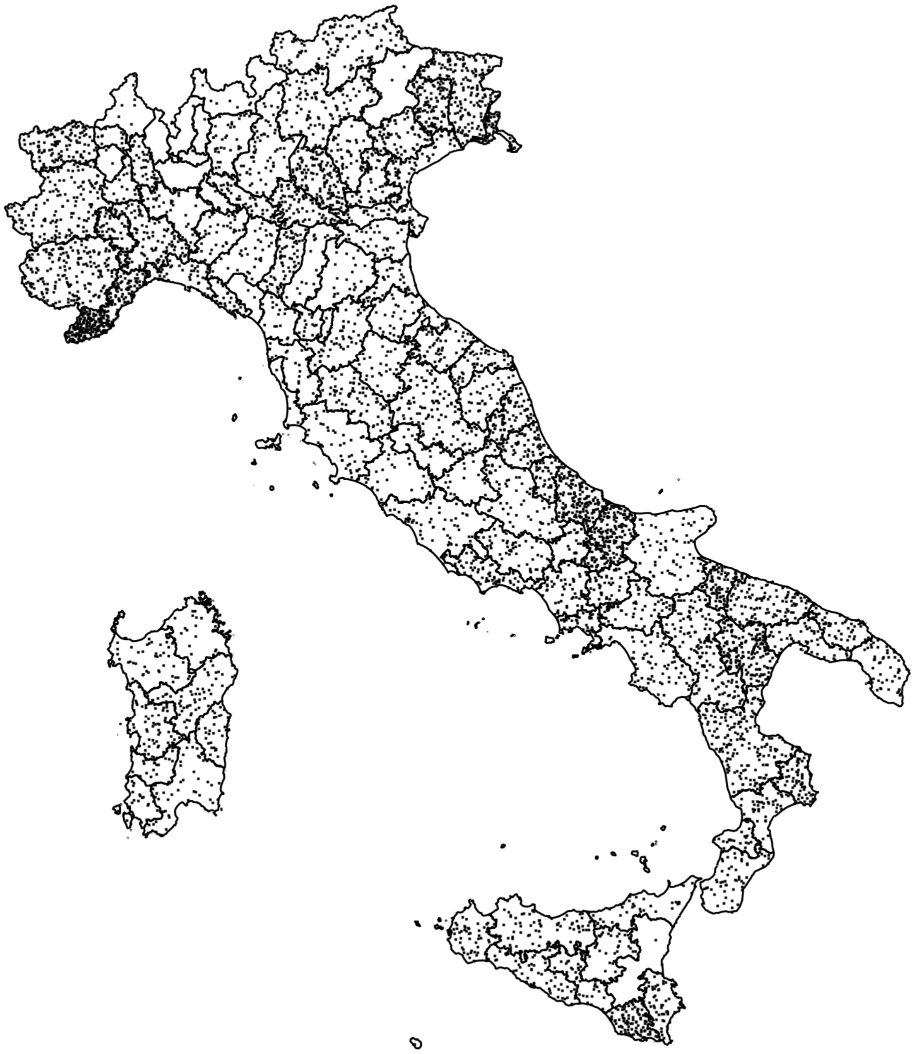
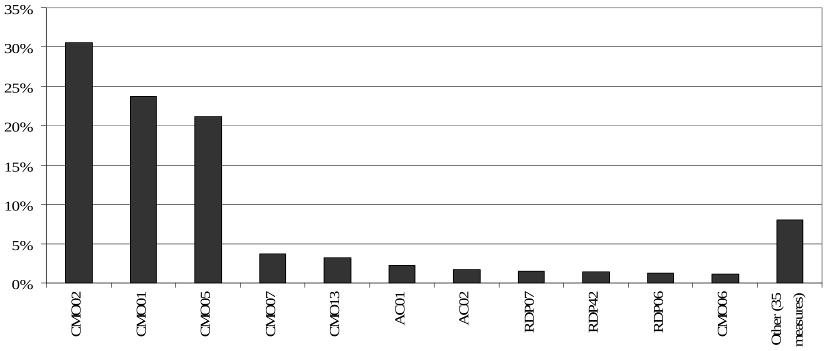
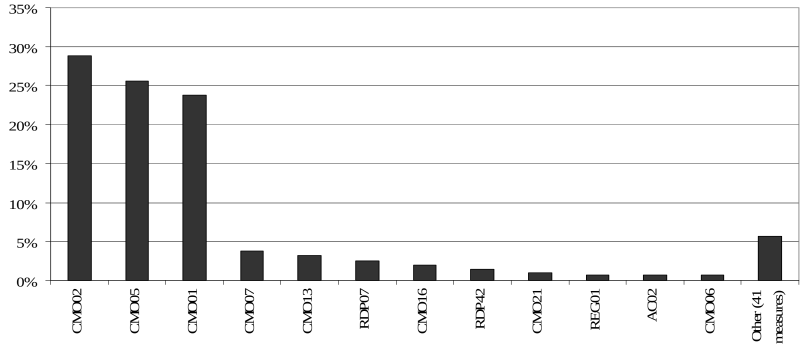


Figure 2. Distribution (%) of the total support across measures (most significant measures: >1% on total support) within the sample (see Annex 1)

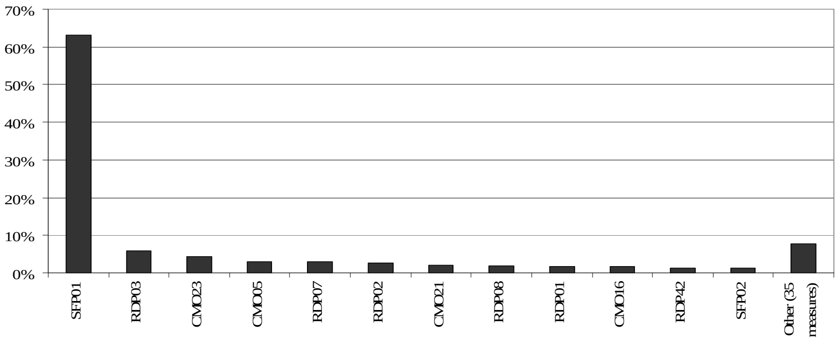
2003



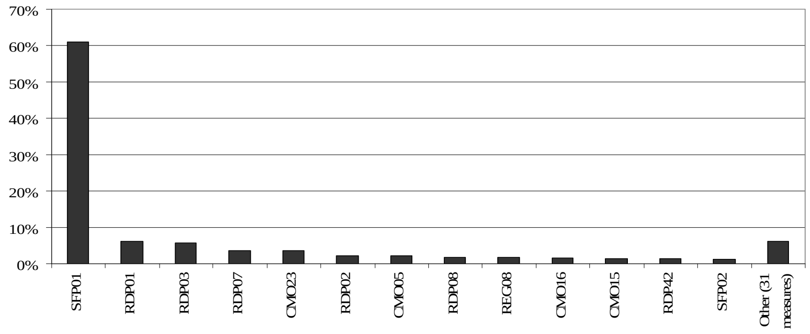
2004



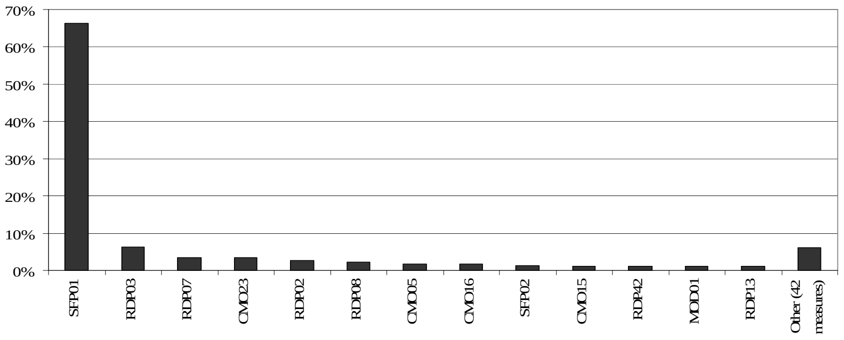
2005



2006



2007



Total number of treatments:

2003 = 16892 (2,6 per farm on avg.). Total support = 82,8 mln € (12600 € per farm on avg.)

2004 = 18864 (2,9 per farm on avg.). Total support = 92,4 mln € (14300 € per farm on avg.)

2005 = 20203 (3,1 per farm on avg.). Total support = 97,2 mln € (14900 € per farm on avg.)

2006 = 19629 (3,0 per farm on avg.). Total support = 105,7 mln € (16200 € per farm on avg.)

2007 = 19721 (3,0 per farm on avg.). Total support = 93,3 mln € (14700 € per farm on avg.)

Source: own elaboration

ANNEX 1 – Abbreviations of policy measures in Figure 2

AC01	Accompanying measure: integrated agriculture (2078/92)
AC02	Accompanying measure: organic agriculture (2078/92)
CMO01	Common Market Organization: arable crops' compensatory payment (1251/99 and others)
CMO02	Common Market Organization: arable crops' supplementary payment (1251/99)
CMO05	Common Market Organization: other crops
CMO06	Common Market Organization: bovine special premium (1254/99)
CMO07	Common Market Organization: suckler cow premium (1254/99)
CMO13	Common Market Organization: sheep premium (3013/89)
CMO15	Common Market Organization: fruits - investments
CMO16	Common Market Organization: durum wheat special quality premium (1782/03)
CMO21	Common Market Organization: dairy premium (1782/03)
CMO23	Common Market Organization: supplementary quality aid for arable crops (art.69, 1782/03)
MOD01	Modulation: supplementary payment (art.12 1782/03)
RDP01	Rural Development Plan: investments in agricultural holdings
RDP02	Rural Development Plan: settlement of young farmers
RDP03	Rural Development Plan: training
RDP06	Rural Development Plan: low environmental impact
RDP07	Rural Development Plan: organic farming
RDP08	Rural Development Plan: breeds in danger of being lost to farming
RDP13	Rural Development Plan: afforestation of agricultural land
RDP42	Rural Development Plan: compensatory payment for less favour areas
REG01	Regional measure: other payments for livestock activities
REG08	Regional measure: rehabilitation and prevention for livestock activities
SFP01	Single Farm Payment (1782/03)
SFP02	Mandatory set-aside

ANNEX 2 – Type of Farm (TF) reclassification (TF_R)

TF 2000 4-digit Classification (TF)	TF reclassified (TF_R)	TF 2000 4-digit Classification (TF) *(continues)*	TF reclassified (TF_R) *(continues)*
1310	1	5011	6
1320	1	5012	6
1330	1	5013	6
1410	2	5021	6
1420	2	5022	6
1430	2	5023	6
1441	2	5031	6
1442	2	5032	6
1443	2	6010	5
2011	7	6020	5
2012	7	6030	5
2013	7	6040	5
2021	7	6050	5
2022	7	6061	5
2023	7	6062	5
2031	7	7110	6
2032	7	7120	6
2033	7	7210	6
2034	7	7220	6
3110	4	7230	6
3120	4	8110	5
3130	4	8120	5
3141	4	8130	5
3143	4	8140	5
3211	4	8210	5
3212	4	8220	5
3213	4	8231	5
3220	4	8232	5
3230	4		
3300	4		
3400	4		
4110	3		
4120	3		
4210	3		
4220	3		
4310	3		
4320	3		
4410	3		
4420	3		
4430	3		
4440	3		