Università degli Studi di Padova
Dipartimento di Scienze Statistiche

Scuola di Dottorato in
Scienze Statistiche
Ciclo XX

An "ex ante" evaluation of the effects of reforms to an Italian labour market policy

Federico Tedeschi

Direttore: Prof. A. Salvan

Supervisore: Prof. E. Rettore

31 Gennaio 2008

Riassunto

L'argomento della mia tesi è la valutazione di politiche *ex ante*, ossia la stima dell'impatto di un intervento pubblico su una variabile risposta prima della sua implementazione.
La politica analizzata è un programma relativo al mercato del lavoro italiano, denominato "Liste di mobilità", caratterizzato dalla presenza contemporanea di una politica "passiva"(sostegno economico ai lavoratori disoccupati) e di una "attiva"(incentivi economici all'impresa che assume lavoratori in movilità). La lunghezza del periodo in mobilità e del diritto a usufruire dei relativi benefici varia in base all'età del lavoratore e alle dimensioni dell'impresa al momento del licenziamento. Tuttavia, i lavoratori indennizzati dal programma ricevono un sussidio proporzionale al loro ultimo reddito prima dell'ingresso in mobilità. I dati utilizzati provengono dall'abbinamento di due diversi archivi amministrativi e sono relativi alla regione Veneto. Le informazioni disponibili sono relative alle caratteristiche sociodemografiche, alla storia lavorativa di tutti gli individui entrati in mobilità fino all'anno 2001, nonché alle caratteristiche delle aziende dalle quali sono stati retribuiti.
Il mio interesse è relativo all'effetto di possibili modifiche alla normativa esistente sulla probabilità di occupazione nei 36 mesi successivi all'ingresso in mobilità, per le quali è necessaria una valutazione di politiche *ex-ante*.
La teoria economica relativa alla determinazione di salario nel mercato del lavoro è rivisitata, al fine di descrivere le ragioni per le quali lavoratori "simili" (ovvero, ugualmente attraenti per le imprese) possano avere retribuzioni differenti.
Tale variabilità nell'ultimo salario prima di entrare in mobilità (e di conseguenza nell'ammontare dell'indennità percepita) è poi utilizzato al fine di identificare l'ef-

fetto di possibili riforme alla normativa attuale. In particolare, due strategie di identificazione differenti sono utilizzate: il “generalized propensity score” e l’approccio “difference in differences”.

Abstract

My thesis is about *ex ante* policy evaluation, i.e. the estimate of the impact of a public intervention on a defined outcome measure prior to its implementation.
The policy regime I analyze is an Italian labour market program targeted to dismissed employees, called "Liste di mobilità" (literally, "Mobility lists", LM hereafter), which includes both a "passive" component (monetary benefits to part of the unemployed workers) and an "active" one (money transfer for the firm hiring them). Length of the period in the LM and entitlement to benefits vary according to the age of the worker and the size of the firm at time of dismissal. However, the amount of the unemployment subsidy (for people entitled to them) is proportional to the last wage earned.
Linked administrative panel data set for the Veneto region (a large region in the Northeastern Italy) are used. Information about people ever entered in the LM (including labour market history, socio-demographic chacteristics, entitlement to receive monetary benefits and characteristics of the firms where they have been employed) is available.
My interest is on the effect of possible changes to the existing policy regime on the probability of re-employment in the 36 months subsequent to enrollment in the LM. Since it deals with potential reforms to the current policy, I need to perform an *ex-ante* policy evaluation.
I will briefly review the job-market economic theory about wage determination, describing why similar workers (from the point of view of attractiveness for firms) may earn different wages.
Basically, I use such random variability in wages, hence in the amount of monetary

benefits received, to mimic a variation in the policy regime. If such policy equivalent variation (PEV) exists and workers are otherwise similar, identification of the effects of *not yet implemented* policy reforms is possible. Specifically, I consider two alternative identification strategies and test their validity. First, I compare individuals basing on the *generalized propensity score* (an extension of the procedure of propensity score *matching* or *subclassification* to non-binary cases; then, I consider the hypothesis that the selection bias pattern is the same in the group of treated and non-treated, using a “difference in differences” approach.

Acknowledgments

First of all, I wish to greet my family and all the people who has supported me with their affection and their psychological support in these years.
I would like to thank my supervisor, Prof. Enrico Rettore, for all the scientific support and for giving me the chance to pursuit this project.
I am also grateful to Prof. Erich Battistin for his precious advise and to Prof. Hidehiko Ichimura for his insight and references on the job-market theory.
I also would like to thank Adriano Paggiaro and Roberto Quaranta for their help in the interpretation of the dataset, and Prof. Giorgio Brunello for his advise about life in Japan.
Thank you also to the Staff of the Department of Statistics of the University of Padua and of the Department of Economics (and of the whole Campus) of the University of Tokyo for their helpfulness.
I want to address a warm thank to my collegues: Francesca, Nedda, Sofia, Antonio, Giuliana, Daniela, Moreno, Simone, Nadia, Nicola, Laura, Vanna, James, Francesco, Daniele, Aldo, Giovanna, Marco for every time they helped me in this effort.
I am especially grateful to all people who supported me in facing economical, bureaucratic and linguistical problems while in Tokyo. I can not mention all of them, but I thank: my parents, Virginia, Jennifer, Yukitoshi-san, Angel, Luzia, Yoko-o San, Niwa-San, Mr. Murai, Tomohiko, Tan-san, Shintaro-San, Koji-san, Takashi-san, Shinya-san, Kengo-san, Yusuke-san, Makoto-san, Yutaka-san, Tsunehiro-san, Satoshi-san, Wataru-san, all my collegues and teachers at the Japanese course in "To-dai".

Contents

Chapter 1

Introduction

My thesis is about *ex ante* policy evaluation, i.e. the estimate of the impact of a public intervention on a defined outcome measure prior to its implementation.
The policy regime I analyze is an Italian labour market program targeted to dismissed employees, called "Liste di mobilita" (literally, "Mobility lists"), which includes both a "passive" component (monetary benefits to part of the unemployed workers) and an "active" one (an unemployment subsidy for the firm hiring them). Length of the period in the LM and entitlement to benefits vary according to the age of the worker and the size of the firm at time of dismissal. However, the amount of the unemployment subsidy (for people entitled to them) is proportional to the last wage earned.
Linked administrative panel data set for the Veneto region[1] are used. Information about people ever entered in the LM (including labour market history, sociodemographic chacteristics, entitlement to receive monetary benefits and characteristics of the firms where they have been employed) up to 2001 is available.
My interest is on the effect of possible changes to the existing policy regime on the probability of re-employment in the 36 months subsequent to enrollment in the LM. Since it deals with potential reforms to the current policy, I need to perform an *ex-ante* policy evaluation.
I will briefly review the job-market economic theory about wage determination, de-

[1] It is a large region in the Northeastern Italy.

scribing why similar workers (from the point of view of attractiveness for firms) may earn different wages.

Basically, I use such random variability in wages, hence in the amount of monetary benefits received, to mimic a variation in the policy regime. If such policy equivalent variation (PEV) exists, identification of the effects of *not yet implemented* policy reforms is possible. Specifically, I consider two alternative identification strategies and test their validity. First, I compare individuals basing on the *generalized propensity score* (an extension of the procedure of propensity score *matching* to non-binary cases -see 2.8.1-) proposed by Hirano, Imbens (2004) and Imai, Van Dyk (2004) to the case of *ex-ante* evaluation (see: Ichimura and Taber, 2000 and 2002; Todd and Wolpin, 2006).

Then, I consider the hypothesis that the selection bias pattern is the same in the group of treated and non-treated, using a "difference in differences" approach (see 2.4). The next chapters are as follows: a review on ex-post and ex-ante evaluation is described in chapter 2; chapter 3 reviews the economic theory on wage dispersion; an explanation on the policy regime under analysis is in chapter 4; chapter 5 shows the theoretical models used, empirical analysis and results.

Chapter 2

From “ ex post” to “ ex ante” evaluation

In this chapter, I start by introducing the notion of causality in Granger (1969) and some following extensions. Then, I relate it to the “Rubin’s model” (Holland, 1986) in order to introduce the policy evaluation setting.

2.1 Causality

Granger (1969) deals with the problem of identification of the direction of causality or the presence of feedback, between two related variables (showing then a generalization for a three-variable case). He adresses the need of extending previous definitions of causality and feedback in order to permit tests for their existence, basing on the stochastic nature of the variables and on the assumption that the future cannot cause the past (he moves in a time-series setting).
He defines U_t as the whole information in the universe accumulated up to time $t-1$, $\overline{A}_t$ as the set of all the past values of the variable A_t and $\overline{\overline{A}}_t$ as the set of all the past and present values of A_t. Then, he defines the notion of *causality*:

$$Y \to X \iff \sigma^2(X|U) < \sigma^2(X|U - \overline{Y}) :$$

"Y causes X" means that we are better able to predict X_t using all available information than if we use all the information but $\overline{Y_t}$. Feedback arises if:

$$\sigma^2(X|U) < \sigma^2(X|U - \overline{Y}), \sigma^2(Y|U) < \sigma^2(Y|U - \overline{X}),$$

i.e.: if X causes Y and Y causes X. Then, the concept of *instantaneous causality* is defined: if

$$\sigma^2(X|U, \overline{\overline{Y}}) < \sigma^2(X|U - \overline{Y}),$$

instantaneous causality $Y \rightarrow X$ is occurring: the current value of X_t is better predicted if the present value of Y_t is included in the prediction than if it is not. He supposes all the relevant information is numerical in nature and belongs to the vector set of time series: $Y_t^D = \{Y_t^i, i \in D\}$, for some integer set D. Then, he explores the situation of omitted variables, introducing the concept of *spurious causality* arising if relevant data are not included in D. For example, if we take $D = (X_t, Y_t)$, but there is a third series Z_t causing both within the enlarged set $D' = (X_t, Y_t, Z_t)$, only spurious causality between X and Y may be found.
Also instant causality may arise spuriously. For example, if: $Y \rightarrow X$ with lag one but the series are sampled every two time units and there is no real instantaneous causality, such causality appears to occurr since we do not use some relevant information (y_{t-1} is not observed, but y_t may inform us about it).
So, predictability is the core of this definition of causality: Y causes X if it contains an information about X that it is contained in no other series. Granger(1969) states that: "In practice it will not usually be possible to use completely optimum predictors, unless all sets of series are assumed to be normally distributed, since such optimum predictors may be nonlinear in complicated ways. It seems natural to use only linear predictors and the above definitions may again be used under this assumption of linearity".
Sims (1972) builds a test (to study the money-income relation in the period '47-'69) for the direction of causality based on Granger's notion. In fact, he admits that data are not able to detect instantaneous causality between Y and X; that the use of time-series is an approximation (time is discrete instead of continuous) and that

more complicated models could be constructed, but concludes that, unless there are sorts of exact relation between the parameters of interest and the sthocastic elements, these issues are not likely to make unidirectional causality wrongly appear, so he mantains that models with a different causality relation could hardly be constructed.
Granger's definition of causality is basically used because of its testability, and the author proves that:

1. When X and Y are two jointly purely linear non-deterministic covariance-stationary stochastic processes, so that we can write: $X(t) = a^*u(t) + b^*v(t)$, $Y(t) = c^*u(t) + d^*v(t)$ (where u and v are mutually incorrelated white noise processes with unit variance, and a, b, c, d vanishes for $t < 0$), Y does not cause X in Granger's definition $\iff$ a or b can be chosen identically 0.

2. *Strict exogeneity*: If (X, Y) has an autoregressive representation, Y can be espressed as a distributed lag function of current and past X with a residual which is not correlated with any values of X, past or future $\iff$ Y does not Granger cause X in Granger's sense.

Angrist and Kuersteiner (2004) will extend the first definition to the case when we condition on a set of covariates. Their generalization is:
$Y_{t+1}, \ldots Y_{t+j}, \ldots \perp D_t | \mathcal{F}_t$, where:
$\mathcal{F}_t = (\overline{X}_t, \overline{Y}_t, \overline{D}_{t-1})$. They show it needs not coincide with Granger's causality by giving a counterexample, where the effect of monetary policy (D_t) on a given output, conditioning on inflation x_t. If the monetary policy affects the outcome only by inflation, Granger tests will fail to detect this causal link , because it conditions on the inflation itself. Sims tests instead will detect this relationship, since it measures the effect on innovations of policy regardless if they occurr because other covariates respond to these policy shocks. They explain it by stating that Granger's and Sims' tests answer to different questions:

1. Granger: What happens to the output if we randomly shock monetary policy, but inflation is fixed?

2. Sims: What happens to the output if we randomly shock monetary policy, and we allow it to affect inflation too?

Hosoya (1977) extends Sim's second theorem to show that identity of Granger's and Sims' causality in the linear predictor case holds even when we introduce an individual latent variable c.
Chamberlain (1982) states that Granger (1969) uses the term "causality" with the meaning of "linear causality", and he also extends Granger's and Sims' (second) definitions by using conditional independence instead of linear predictors. He exits from the time-series and the multivariate normal settings, allowing for the presence of time-invariant variables and exploring situations outside the linear case. Particularly, Granger's noncausality (G) is expressed by: "X is independent of past Y conditional on past X", while Sims' noncausality (S) becomes: "Y is independent of future X conditional on current and past X". Formally, it is:

- (G): x_{t+1} is independent of $y_t, y_{t-1}, \dots$ conditional on $x_t, x_{t-1}, \dots \forall t$.
- (S): y_t is independent of $x_{t+1}, x_{t+2}, \dots$ conditional on $x_t, x_{t-1}, \dots \forall t$.

While (G) implies (S), the opposite does not hold. So, the two definitions of causality are equivalent in the linear case -because they imply identical restrictions on the covariance matrix of $(x_{i1}, \cdots x_{iT}, y_{i1}, \cdots y_{iT})$- but not in the general case. This holds because if a random variable is uncorrelated with two other random variables, then it is uncorrelated with every linear combition of them; but if it is independent with the other random variables, it is not independent of any function of them. Then, he proposes two different versions of Sim's causality, to make Granger's and Sim's definition coincide. In the first case, he extends the conditioning set to past values of the y values and define a regularity condition for the equivalence ($G = S'$) to hold. On the contrary, the second definition is stronger ($S'' \rightarrow S'$) and equivalent to Granger's one: $G = S''$:

- (S'): y_t is independent of $x_{t+1}, x_{t+2}, \dots$ conditional on $x_t, x_{t-1}, \dots, y_{t-1}, y_{t-2} \dots \forall t$.
- (S''): y_t is independent of $x_{t+1}, x_{t+2}, \dots$ conditional on $x_t, x_{t-1}, \dots, Y_t, \quad \forall Y_t, t$, where Y_t is a subset of $y_{t-1}, y_{t-2}, \dots$. For (S'') to be valid, it is sufficient that

the conditional independence holds for Y_t equal to the null set (S) and all sets of the form $(Y_t = [y_{t-1}, \ldots y_{t-k}], k = 1, 2, \ldots)$.

Then, he explores the relation between these definitions, and studies mean conditional independence, finding that even in the regression case Sims' and Granger's noncausality definition do not coincide, despite they do in the linear case, because mean independence is not a simmetric relationship, while correlation is.
Chamberlain (1984) analyses panel data, defining the relationship of y to x as *static* if x is strictly exogenous and y_t is independent of $x_1, \ldots, x_{t-1}$ conditional on x_t (there are no structural lagged dependent variables). He observes that, when there is instant feedback $y_t \leftrightarrow x_t$ and y_t and x_t depends on their past values, we expect that y_2 depends on x_1 even conditional on x_2, so we expect x not to be static to y. Then, he studies the case when we introduce a time invariant latent variable c or an observed one (z), analyzing their effects on the unbiasedness and consistency of the estimators. Particularly, he shows that a latent variable c that makes x strictly exogenous to y conditional on it always exists (this does not hold in the linear predictor case) and that conditional strict exogeneity or a conditional static relationship cannot be tested without restricting the functional form of the distribution of c on X, or of Y conditional on X, c, implying truly nonparametric tests cannot exist (but functional forms can be tested too).

2.2 From causality to policy evaluation

Holland(1986) adopts the experiment framework (stating it is the simplest one) for the model of causal inference, basing on Rubin (1974,1977,1978,1980a) and defining it as "Rubin's model", whose he means to describe a "simplified, population level version". He informs "the terms *cause* and *treatment* will be used interchangeably", to underline that "the effect of a cause is *always* relative to another cause". Basically, there is a comparison between two *potential outcomes* (Rubin, 1974) for exposing a unit to a cause or not: each unit is potentially exposable to any of the causes, regardless of whether it can be achieved in pratice or not. The treatment S is assumed to be binary: $S = 1$ for units that are exposed to the treatment, $S = 0$ for

the "control" group. The response variable Y need to be a post-exposure variable, and two potential outcomes must be taken into consideration: Y_1 and Y_0.
This is the *policy evaluation* settings, whose goal is to detect the impact of a program on defined outcome measures as the difference between the potential outcomes. The policy effect on the unit i $(Y_1(i)-Y_0(i))$, but we can not measure both $Y_1(i)$ and $Y_0(i)$ at the same time, so we may not identify the effect of the treatment on each single unit. Holland (1986) defines it as the "fundamental problem of causal inference", and states the statistical solution to this problem is limited to the average causal effect of the treatment $E(Y_1 - Y_0)$ on a population of units.
The author goes on by restating Granger's theory outside the time series setting, applying it to Rubin's model. Basically, the temporal distinction in Holland (1986) is only meant to separate variables, determined before, at or after a given point in time, because "it is the past values of a variable that cause, in Granger's sense, the future values of another variable". Therefore, X may be a cause of Y only if it is determined prior to it. Moreover, he notices later writers (e.g.: Florens and Mauchart, 1985) restated Granger's approach in terms of non-causality, and follows their approach. Furthermore, as Chamberlain (1982), he concentrates on the notion of conditional linear independence rather than on linear prediction.
An application is proposed, with variables defined on a population of units: Y is a variable defined at time s, while X and Z are determined prior to it. In this case, *X is not a Granger cause of Y* (relative to the information in Z) $\iff$

$$X \perp Y|Z. \tag{2.1}$$

Then, he moves to the Rubin model's setting: there is not just one outcome, but as many potential outcomes as points in the support of S (i.e.: treatment status). So, Y becomes Y_s (i.e.: a potential outcome), X becomes S and Z is a set of pre-exposure variables. As a consequence, in the binary case, we have: $Y = Y_1\mathbb{I}(s = 1) + Y_0\mathbb{I}(s = 0) = Y_0 + (Y_1 - Y_0)\mathbb{I}(s = 1)$, implying that condition (2.1) becomes:

$$[Y_0 + (Y_1 - Y_0)\mathbb{I}(s = 1)] \perp S|Z. \tag{2.2}$$

The author states the condition for non-causality to hold as:
$Pr(Y_s = Y|S = 1, Z) = Pr(Y_s = Y|Z)$, that is equivalent to:

$$\{Pr(Y_1 = y|S = 1, Z) - Pr(Y_0 = y|S = 0, Z)\} \times Pr(S = 0|Z) = 0, \tag{2.3}$$

i.e.: treatment status must be independent of potential outcomes for all z values for which it is possible to be part of the control group. Then, he notices that, in a randomized experiment, any pre-treatment Z vector is independent of treatment status (and that both status are possible for every unit):
$0 < Pr(S = 0|Z) = Pr(S = 0) < 1$. So (2.3) becomes:

$$Pr(Y_1 = y|S = 1, Z) - Pr(Y_0 = y|S = 0, Z) = 0.$$

Moreover, treatment status S is also independent of (Y_1, Y_0), so conditional independence between treatment status and potential outcomes would imply the two potential outcomes have the same conditional distribution:
$Pr(Y_1 = y|Z) - Pr(Y_0 = y|Z) = 0$, which implies: $E(Y_1|Z) = E(Y_0|Z)$. So, if S is not a Granger cause of Y_0 relative to Z, the average causal effect for $Z = z$ is always null ($T(z) = E(Y_1 - Y_0|Z = z) = 0 \forall z$). Therefore, in a randomized experiment, Granger noncausality implies zero average causal effect on all subpopulations defined by the values of z. Conversely, if t has a null effect on all units, S will not be a Granger cause relative to any Z that is a pre-exposure variable.
To sum up, Holland (1986) shows the analogy between Rubin's causality and Granger's causality, where the former introduces the concept of potential outcome. Anyway, this analogy holds as long as the condition: $Y_1, Y_0 \perp Z$ holds (as in a randomized experiment).
In case we are outside the randomized experiment setting, a noncausality relation could fail by simply gathering more information (that is, by changing Z): as well as spurious causality (Granger, 1969), also spurious noncausality may arise. This implies the problem of omission of covariates is crucial: it must be taken into consideration before getting to any conclusion regarding causality relations. In the following of the chapter, I describe the main risk omission of covariate may bring

(the selection bias) and the way to eradicate it in order to identify the causal relations of interest.

2.3 Selection bias

If our interest is about the average effects of a treatment in a population, it seems natural to compare the average outcomes of the two groups. The difference may be decomposed in two parts:

$$E[Y_1|T=1]-E[Y_0|T=0] = [E(Y_1|T=1)-E(Y_0|T=1)]+[E(Y_0|T=1)-E(Y_0|T=0)].$$

This implies that the gap between the two groups that we observe is the sum of the average treatment effect on the treated (TT) and the *selection bias*, i.e. the difference between the two groups we would have observed even if the policy had not taken place. Basically, there may be counfounding pre-treatment factors affecting both treatment assignment T (so that these characteristics sistematically differ across treated and non-treated) and the potential outcome Y_0, making a direct comparison of the two observed mean potential outcomes biased. If assignment to the treatment is random (possibly after conditioning on a set of pre-treatment covariates X), we are in a situation as the one described in Neyman, 1923 (rivisited by Rubin, 1990, in terms of causal inference): random sampling from more than one urns (two, in case of binary treatment status) without replacement, and "further suppose that our urns have the property that if one ball is taken from one of them, then balls having the same (plot) label disappear from all the other urns" (Neyman, 1923). The author shows that, in this case, the difference between the two observed outcomes is not biased. This implies:
$E[Y_1|T=1] - E[Y_0|T=0] = E[Y_1|T=1] - E[Y_0|T=1] =$
$E[Y_1|T=0] - E[Y_0|T=0] = E[Y_1] - E[Y_0]$.
In this case, the average policy effect is the same for the treated and for the control group, so we can estimate the average effect of the treatment in the whole population: the intervention entirely accounts for systematic differences between the two groups.

Anyway, policy evaluation is often required also for situations where such an experiment is unfeasible (for practical or ethical reasons): in this case, we have to create an *ex-post quasi experimental setting*, such that: $Y_i \perp T$.

2.4 Difference in differences

Ashenfelter (1978) estimates the effect of training program on earnings, using a two step strategy. First, the difference between outcome measures before and after the program is estimated for both the treated observations and the controls. Then, this differences are compared between the groups. Hence, the name "difference-in-differences" estimator, that requires a stable composition in time (before and after the treatment) of the treated and the control group, and leads to unbiased estimates as long as the extent of selection bias in the pre- and post-program outcomes is equal (Ashenfelter and Card, 1985). This strategy is also called: *natural experiment approach*, because "it considers the policy reform itself as an experiment and tries to find a naturally occurring comparison group that can mimic the properties of the control group in the properly designed experimental context" (Blundell and Costa Dias, 2002). Basically, the assumption made is that unobserved factors affect the treated and the controls in the same way: the average change in the outcomes (the time effect) would have been the same, in absence of the program. Tipically, this assumption is sensible if the time pattern before the treatment has been very similar in the two groups.
A simple regression framework to use for the DD estimator is:

$$y_i = \beta_0 + \beta_1 * treat + \beta_2 * after + \beta_3 * treat * after + \epsilon, \tag{2.4}$$

where *treat* is the dummy for being exposed to the treatment (implying β_1 captures systematic differences between the two groups) and $after$ is the dummy for the outcome to be post-treatment (implying β_2 represents the pre-post difference for the control group). In this way, the parameter β_3 identifies the effect of the treatment. If we believe that some factors may lead to a different time-pattern in the outcomes, they can be added in the regression equation 2.4 (if the treatment may have a different effect on different units depending on these variables, also the interaction terms may be added, as stated by Meyer, 1995). Blundell and Costa Dias (2000) propose the application of the DD technique to DD estimates, with the "difference in difference in differences" procedure. Other possible solutions to the assumption of homogeneous effects of the unobserved factor are proposed by Dee and Fu (2003) and Abadie (2005). Heckmann (1978) and Heckmann and Robb (1985, 1986) propose a generalization of this estimator in case of time-series with autocorrelated stochastic component, and discuss the robustness of "difference-in-differences" estimator for longitudinal or repeated cross-section data. Bertrand, Duflo and Mullainathan (2004) analyze the bias in standard error estimates in case of serially autocorrelated outcomes, proposing techniques to solve this problem.

2.5 Instrumental variables

In case of a regression:

$$Y = \alpha + \beta X + \epsilon, \tag{2.5}$$

a lack of exogeneity (Engle, Hendry and Richard, 1983) of X may bias the parameter estimate β, failing to identify a structural relation between X and Y. Goldberger (1972a) reports that Wright (1928) complied to the need of a generalization of his son's "path coefficients" model (Wright, 1925) by using a variable Z (the price of a substitute, or an index of prosperity) to estimate the parameter of the supply function in a supply-demand model:
$q^D = \alpha p + u$; $q^S = \beta p + v$; $q^S = q^D = q$.
Being the covariance between z and v null ($C(z, v) = 0$), he derives:
$C(z, q) = \beta C(z, p)$, so that $\beta = C(z, q)/C(z, p)$. For the identification of α, he specularly uses a variable x (yield per acre, or lagge price) for which $C(x, u) = 0$, deriving:

$$\alpha = C(x, q)/C(x, p). \tag{2.6}$$

The role played by Z and X is what Reiersol (1945)[1] defines *instrumental variables* [2]. Reiersol (1945) and Geary (1949) independently show that the sample analogue of (2.6) leads to unbiased estimates of α. Basically, being Z the instrumental variable for (2.5), IV leads to identification of the parameter of interest β as long as, at the same time, Z has a non-null correlation with X and a null correlation with ϵ (the magnitude of this correlation is strictly related to the precision of the estimator, as pointed out by Geary, 1949). Durbin (1954) defines the IV estimator as the ratio of sample covariances. In (2.5), IV estimates is: $\frac{\hat{C}(Z,Y)}{\hat{C}(Z,X)}$.

While these models dealt with random disturbances and measurement errors in exact economic relationships, Sargan (1958, 1959) apply IV to more general cases, giving stronger basis to IV methodology and theory (developing significance test and confidence intervals, over-identification and under-identification tests and also studying the case with autoregressive disturbances). Moreover, while Wright (1928) suggested to average between the available IV estimates in case of multiple instruments, more efficient method to use surplus estimator were proposed (see Sargan, 1958).

Theil (1953, 1958) and Basmann(1957) propose the two stage least squares method: in the first stage, the "endogenous" variable X is regressed on all the instruments Z. In the second stage, the predicted values of X from the first-stage regression are either put directly into the equation of interest in place of the endogenous regressor or used as an instrument. In this way, two-stage least squares takes the information in a set of instruments, using it to build a single overall instrument. A further extension to better control for bias is the split-sample two stages least square, from Angrist and Krueger (1995): first stage parameters are estimated in one half of the sample to construct fitted values and second stage parameters for the other half sample.

The policy evaluation setting is explicit in Angrist and Imbens (1991): in case of heterogeneous treatment effect, the average effect of the treatment ($T = 1$) on the outcome Y is identified if there exist values of the instrument D such that proba-

[1]Reiersol attributed the term to his teacher, Ragner Frisch (Morgan, 1990)

[2]For an overlook on the history instrumental variables, see Angrist, Krueger (2001), and references therein.

bility of treatment is null: $P(T = 1|D = d) = 0$. In Angrist and Imbens (1994), a different identification restriction is used: the *monotonicity* assumption. Being z, w two arbitrary values of the instrument Z and i the indicator for individuals, this condition requires that, for all z, w, either $D_i(z) \geq D_i(w) \forall i$, or $D_i(z) \leq D_i(w) \forall i$. In this way, a local average treatment effect ($LATE$) is identified: the one on *compliers* [3] (i.e.: people whose behaviour depend on the instrumental variable). Empirical issues related to LATE estimation are faced in Angrist and Imbens (1995), while conditions for the causal interpretation of this estimand may be found in Angrist, Imbens and Rubin (1996).

2.6 Regression discontinuity design

The first publication on *regression discontinuity design* (RDD) was by Thistlewaite and Campbell (1960). The RDD is described by Trochim (1984) as "one member of a larger class of quasi-experimental methods that can be termed pretest-posttest group designs". It requires the presence of a pre-program continuous measure X based on which people are assigned either to the treatment or to the control group. A crucial assumption needed for the validity of regression discontinuity design is the continuity in X of the average of potential outcomes ($E(Y_1|X = x)$, $E(Y_0|X = x)$) at the threshold $x = x^*$. Imbens and Lemieux (2008) extend this assumption of continuity in X to the distribution of potential outcomes: being $F_{Y(w)|X}(y|x) = Pr(Y(w) \leq y|X = x)$, $F_{Y(0)|X}(y|x)$ and $F_{Y(1)|X}(y|x)$ are continuous in x, for all y. Basically, this means that only treatment status accounts for a possible discontinuity in Y at the cutoff point, because there are not other factors accounting for such a discontinuity.

2.6.1 Sharp regression discontinuity design

Goldberger (1972b,c) shows a first prove of unbiasedness of the sharp RDD. The *sharp* RDD is characterized by the feature that *all* individuals on one side of a cut-

[3] Also Balke and Pearl (1994) deal with the problem of partial compliance in case of random treatment assignment, proposing bounds for the (overall) average treatment effect.

off score x^* are assigned to the program group and, analogously, *all* the ones scoring on the other side belong the control group. The selection process is entirely known: treatment status is a deterministic functions of X. Without losing generality, we may assume treatment is assigned for $X \geq x$. As long as the assumption of continuity of $E[Y_1|x], E[Y_0|x]$ in x^* holds, the difference between the average factual outcomes in the two groups at the threshold x^*:

$$lim_{x \downarrow x^*} E[Y_i|X_i = x] - lim_{x \uparrow x^*} E[Y_i|X_i = x]$$

is the effect of the treatment: $E[Y_1 - Y_0|x^*]$.
Basically, since the two populations do not overlap (with regard to the variable X), identification of the effect of the treatment is possible only at the threshold x^* (i.e., for marginal treated and marginal non-treated), thanks to the continuity assumption. This quasi experimental setting is only local and, if the average effect of the treatment is heterogenous in x, the *RDD* does not inform us about individuals far from the threshold. Since a misspecification of the regression function may bias the estimate of the effect at the threshold, a non-parametric approach is preferable, despite its lower convergence rate (see Hahn, Todd and Var der Klaauw, 2001). Rettore and Battistin (2008) show the sharp *RDD* may be extended to the cases where the population is divided between eligibles and not-eligibles, and the first group self-select into the program.

2.6.2 Fuzzy regression discontinuity design

Campbell (1969) introduced the *fuzzy* RDD, i.e.: the case when the probability of receiving the treatment does not change from 0 to 1 at the threshold: there is still a "jump", but smaller. Basically, there are individuals that do not get the treatment they should receive (given their X value). Trochim (1984) notices that, while the assignment rule in randomization and in the sharp RDD is known, with a fuzzy RDD it is partially unknown. Several solutions to this issue have been proposed (for example: Campbell, 1969; Trochim and Spiegelman, 1980; Trochim, 1984).
The current approach is based on analogy with IV, basically extending to the RDD

the ideas of Angrist and Imbens (1994) and Angrist, Imbens and Rubin (1996) and using treatment assignment at the threeshold as an instrument. If the monotonicity assumption holds, i.e.: there are no *defiers* (people who behave in the opposite way they are expected to: exposed to treatment if they are assigned to the control group, and *viceversa*), the effect of the treatment for compliers (individuals who are treated if and only if they are assigned to the treatment group) is identifiable. This idea is expressed in Hahn, Todd and Van der Klaauw (2001), that also find weaker conditions for identification.

2.7 Adjustment for covariates

Cochran (1965) states that "there is the familiar problem that the response or dependent measurements are usually influenced by many variables other than those under investigation". As seen above (see 2.3), if these factors also affect treatment status T (implying that their distribution in the treatment and the control group is different), the direct comparison of the distributions of the outcome variable in the two groups is generally biased.

He mantains the possible confounding factors have to be listed, and then divided into three groups:

1. The ones implying a small bias, that might be ignored.

2. The ones for which we have to check if we may assume they bring little or no bias. in order to exclude them from analysis.

3. Variables for which some kind of matching or adjustment is required.

In this way, the author underlines the worrying of keeping the number of variables to be matched small, due to the difficiulties that matching over many variables simultaneously brings. Then, he considers three different ways of adjusting for the disturbing variables:

1. *Matching*: Pairs of individuals (one from each population) are selected, such that their values $\mathrm{x_i}$, $\mathrm{x}_{i'}$ are identical, or their distance is minimized (various

kind of matching are possible, depending on the minimization criterion used). This balances the distribution of $\mathbb{X}$ in the two samples, even if pratical problems may arise. Basically, when the counfounding factors are many, and their distribution is very different in the two groups, we are likely to use a very small part of the original sample for matching. Chapin (1947) shows an example where 671 and 523 units are present in the two treatment groups: after matching on six covariates, just 23 pairs are left.

2. *Subclassification*: The population is divided into classes based on the values of $\mathbb{X}$. In each subclass, the tretment effect is estimated, and the overall tretament effect is then calculated as a weighted average of the within-subclass estimates.

3. *Regression adjustment*: Used mainly in the case of continuous $\mathbb{X}$, it consists in regressing y on x within the two groups, and adjusting the difference between treated and controls to remove the bias as estimated by this regression. Linear regression is the most commonly used *covariance adjustment*.

Cochran (1968) studies the performance of subclassification, while Rubin (1973a, 1980b) the one of matching, in bias removal. Cochran and Rubin (1973) and Rubin (1973b, 1979) also compares the effectivness of matching and regression adjustment, showing that the least biased and more robust estimates are generally given by a combination of the two techniques. Rubin (1976a, 1976b) shows how to obtain equal percent bias reduction (among all the confounding factors) in case of multivariate matching method.

2.8 Propensity score

Rosenbaum and Rubin (1983) build a method that aims to control for differences in confounding factors between the treatment groups when the treatment is binary, and has become very popular in social sciences: *propensity score matching.*

A first assumption required is *strong ignorability of treatment assignment*: potential outcomes and treatment status are conditionally independent given X, and both status are possible for any value of X:

$(Y_1, Y_0) \perp z|\mathrm{x}, \forall \mathrm{x} \in \mathrm{X}$, $0 < Pr(z = 1|\mathrm{x}) < 1, \forall \mathrm{x} \in \mathrm{X}$.
A second condition is called the *stable unit treatment value assumption* ($SUTVA$), introduced by Rubin (1980a). This assumption requires that the representation $Y_i(t)$ is "adeguate": for each treatment status t of unit i, the potential outcome $Y_i(t)$ is univocally determined . The main implication of this condition ($SUTVA$) is the independence between potential outcomes of a unit and treatment status of another unit: $(Y_0(i), Y_1(i)) \perp T(j), \forall \mathrm{x} \in \mathrm{X}$. This means "interference between units" (Cox, 1958) makes this assumption fail[4]. Omission of covariates can make these assumptions fail, seriously biasing the estimates.
To get unbiased estimate of the treatment effect, we need to condition on a *balancing score*, $b(x)$, conditioning on which X is independent from treatment status: $\mathrm{X} \perp T|b(\mathrm{x})$. In this way, strong ignorability holds even if we condition on $b(x)$: $(Y_1, Y_0) \perp z|b(x), \forall \mathrm{x} \in \mathrm{X}$. The vector of the covariates trivially satisfies this condition, being the finest balancing score. This implies that we can match or subclassify individuals based on their value of x. To make matching or subclassification easier (reducing the dimensionality), it is enough to do it on the coarsest balancing score, i.e. the *propensity score*, i.e. the probability of being treated given the set of pre-treatment covariates: $e(\mathrm{x}) = P(T|\mathrm{x})$. Therefore, the propensity score allows to condition just on a scalar. Even if the *true* propensity score is known (that is: when we know how treatment assignment has been decided), conditioning on the *estimated* one may be adventageous. In fact, it accounts not only for the systematic relationship between the distribution of X and T, but also for the random differences in this conditional distribution in the observed sample. These sample specific differences average to 0 over all the possible samples (stochastic independence), but in single samples they lead to an increase in variance.

2.8.1 Generalized propensity score

Some extensions of the propensity score outside the binary case have been proposed in literature: Imbens (2000), for categorical treatment variables, proposes to adjust

[4]For a discussion of other possible violations of this condition, see Rubin (1986), Holland (1987), Rosenbaum (1987).

for several PS; Joffe and Rosenbaum (1999) (that propose the possibility of adjusting for a low-dimensional linear PS) and Lu et al (2001) for ordinal ones; Hirano, Imbens (2004) for continuous variables, and Imai, Van Dyk (2004) for arbitrary treatment regimes. I describe the generalized propensity score as defined by Imai, Van Dyk (2004).

First, they define T^A as the random variable for the treatment received, $\mathcal{T}$ as the set of potential treatment status and t^p as a particular potential treatment. To estimate the distribution of potential outcomes ($Y(\mathrm{t}^p)$), the two standard assumptions of the binary PS must be generalized:

1. *Stable unit treatment value assumption*: The distribution of potential outcomes for one unit is assumed to be independent of potential treatment status of another unit given the observed covariates X:

$$Y_i(\mathrm{t}^p) \perp T(j), \forall \mathrm{x} \in \mathrm{X}. \tag{2.7}$$

2. *Strong ignorability of Treatment Assignment*: The distribution of the actual treatment does not depend on potential outcomes given the observed covariates:

$$p(\mathrm{T}^A|Y(\mathrm{t}^p), \mathrm{X}) = p(\mathrm{T}^A|\mathrm{X}), \forall \mathrm{t}^p \in \mathcal{T}. \tag{2.8}$$

   Moreover, all the set of status ($A \subset T$) with positive measure are possible for any value of X:

$$0 < p(\mathrm{T}^A|\mathrm{X}) \forall \mathrm{x} \in \mathrm{X} \tag{2.9}$$

These two assumptions have the same role as in the binary case: they may fail in case of omission of covariates, leading to biased estimates but, if they hold, we can obtain valid inference.

The generalization of the propensity score is the *propensity function* ($e(\cdot|X)$), that has the standard characteristics and goals: it must be a balancing score ($p(\mathrm{T}^A|\mathrm{X}) = p\{\mathrm{T}^A|\mathrm{X}, e(\cdot|\mathrm{X})\} = p\{\mathrm{T}^A|e(\cdot|\mathrm{X})\}$) for strong ignorability to hold ($p\{Y(\mathrm{t}^P)|\mathrm{T}^A, e(\cdot|\mathrm{X})\} = p\{Y(\mathrm{t}^P)|e(\cdot|\mathrm{X})\}, \forall \mathrm{t}^P \in \mathcal{T}$) and it must reduce the dimensionality of the conditioning set in order to make the matching or subcalssification procedure easier. When it is

unknown, misspecification of the model is possible, which may bias causal inference, so care must be taken in selecting the model form of the PF and in examining the balance and computing the treatment effect after conditioning on it.
To simplify the representation of the propensity function and to facilitate subclassification and matching, authors make the assumption of a *uniquely parametrized* PF: $\forall \mathrm{X} \in \mathcal{X}$, there exists a unique finite-dimensional parameter, $\boldsymbol{\theta} \in \Theta$, such that $e(\cdot|\mathrm{X})$ depends on X only through $\boldsymbol{\theta}$. So, $\boldsymbol{\theta}$ uniquely represents $e\{\cdot|\boldsymbol{\theta}, \mathrm{X}\}$: the PF is characterized by $\boldsymbol{\theta}$, which is tipically of much lower dimension than X. Matching or subclassification can be implemented on $\boldsymbol{\theta}$. For example, if we assume the distribution of treatment status to be normal with a fixed variance, $\boldsymbol{\theta}$ is univariate (and it is the mean of this normal, given by a function of X); if also the variance is not fixed but depends on covariates, $\boldsymbol{\theta}$ is bivariate. In the binary case, $\boldsymbol{\theta}$ is the parameter of the binomial distribution (PS is a scalar). We need the vector $\boldsymbol{\theta}$ characterizing the propensity function to be of lower dimension than X. Thanks to these results, we can estimates $p\{Y(\mathrm{t}^P)\}$. It is:

$$p\{Y(\mathrm{t}^P)\} = \int p\{Y(\mathrm{t}^P)|\mathrm{T}^A = \mathrm{t}^P, \boldsymbol{\theta}\}p(\boldsymbol{\theta})d(\boldsymbol{\theta}). \tag{2.10}$$

This integration can be approximated by subclassifying similar values of $\boldsymbol{\theta}$ into J subclasses of roughly equal size. Within each subclass, $P_\phi\{Y(\mathrm{t}^P)|\mathrm{T}^A = \mathrm{t}^P\}$ can be modeled parametrically, where $\boldsymbol{\phi}$ is an unknown parameter that is allowed to vary across subclasses (being the model the same in each subclass). Then, a weighted average of the within-subclass distributions is computed, with weights (W_j) equal to the relative size of the subclasses. Formally, (2.10) is approximated by:

$$\sum_{j=1}^{J} p_{\hat{\phi}_j}\{y(\mathrm{t}^P)|\mathrm{T}^A = \mathrm{t}^P\}W_j,$$

where $\hat{\phi}_j$ is the estimate of ϕ in subclass j and W_j is the relative weight of subclass j . We can often summarize distributions with one relevant causal effect (generally, a function of $\boldsymbol{\phi}$). Authors average between $\hat{\boldsymbol{\phi}}_j$ to obtain $\hat{\boldsymbol{\phi}}$, the overall average causal effect. To further reduce bias, they adjust for covariates in the within-subclass model:

$$\hat{\boldsymbol{\phi}} = \sum_{j=1}^{J} \hat{\boldsymbol{\phi}}_j \{Y(\mathrm{t}^P)|\mathrm{T}^A = \mathrm{t}^P, \mathrm{X}\} W_j.$$

If W_j is known and the estimate of the causal effect is unbiased within each subclass, then this procedure results in an unbiased estimate of the causal effect. In practice, authors estimate W_j by the relative proportion of the observations that fall into subclass j. Results may be sensitive to the number of subclasses and the choice of subclassification on $\boldsymbol{\theta}$, so a sensitivity analysis is suggested. Usually, the single subclass has a too large standard error to distinguish the effect among the subclasses. Alternatively, it is possible to allow the causal effect $\boldsymbol{\phi}$ to vary as a smooth function of θ, via *smooth coefficient models* (Di Nardo and Tobias, 2001; Li,Huang and Fu, 2002; Yatchew, 1998). This model allows to study the hypothesis of constance of $\boldsymbol{\phi}$.

2.9 Ex-ante evaluation

2.9.1 Importance of evaluating programs prior to their implementation

Marschak (1953) describes the study of the effects of policy changes prior to their implementation as one of the most challenging problems facing empirical economists. He adresses this topic in the decision-making setting, wondering which kind of knowledge is important to make the best decisions, and how to use it. Especially, he deals with the knowledge needed by a monopolistic firm in its choice of the most profitable output level, and by the government (who knows firm behaviour) in its choice of the rate of excise tax on firm product. He discusses the conditions under which the best decision can be made just with the consideration of past observations (exploiting their variability), showing the cases when, contrarywise, knowledge of the structure of the model (in this case, the demand curve as a function of taxes and the profit as a function of production) is necessary. The econometric literature has then typically used structural models or historical variation corresponding to the policy under consideration to make predictions.

Ichimura and Taber (2000, 2002) propose to exploit other types of variation (defining it "policy equivalent variation") in the data to mimic the effects of a policy change, in order to derive (weak) conditions under which the impacts of a policy can be identified using data generated under a different policy regime. In this way, they can use a non-parametrical approach, avoiding possible misspecification of the model and focusing on the parameter of interest.
Todd and Wolpin (2006) describe the importance of evaluating social programs before they take place. This allows to simulate the impact of potentially many hypothetical programs in order to choose the optimal one (from the point of view of costs and impacts), instead of implementing all of them: the cost of implementing ineffective programs may be avoided. Moreover, in case of existing programs, it would be interesting to predict what would happen if we altered some parameters of them. Finally, even an *ex-post* evaluation may be better implemented when there is an idea of the expected impacts.

2.9.2 Conditions for ex ante evaluation

The main feature of *ex ante* evaluation as proposed by Ichimura, Taber (2000, 2002) and Todd, Wolpin (2006) is the fact that all the factual outcomes are about non-treated individuals, i.e. none of them has been exposed yet to the policy the analyst is willing to evaluate. The matching procedure is between an individual i whose we observe (or estimate) the outcome as non-treated, and an individual j that mimicks the outcome individual i would have under the new policy. Todd and Wolpin (2006) underline a broad difference with the conventional matching approach: individuals must not be similar in all respect, but different in a given way. It must be: $Y_j(0) = Y_i(1)$, i.e.: the factual outcome for individual j under the *status quo* policy regime must be equal the one of individual i under the new policy.
The basic model from Ichimura, Taber (2000) includes a choice variable D, the policy under consideration Π, the outcome variables (Y_1, Y_0) and a vector of observable random variables Z.
A first assumption is:

$$Y \perp \pi | D, \tag{2.11}$$

meaning that the policy affects the outcome only through the choice variable. This means the distribution of potential outcomes is not altered by the introduction of the new policy: this assumption is sensible only for local programs, otherwise general equilibrium effects should be taken into consideration.
Using the notation $\Delta(z, \pi', \pi)$ and $\Delta_c(z, \pi', \pi)$ to point, respectively, for people with $Z = z$, the *mean policy effect* $(E[Y(z, \pi') - Y(z, \pi)])$ and the one on the treated: $\Delta_c(z, \pi', \pi) = E[Y(z, \pi') - Y(z, \pi)|D(z, \pi') \neq D(z, \pi)]$ (that they define *conditional policy effect*), authors also explore the case when Π assumes $n-$*dimensional* real values to define two marginal treatment effects (being π the current policy parameter value, the impact of an infinitesimal change). If $\lambda > 0$ is a real number and $\pi' = \pi + \lambda\widetilde{\pi}$, letting $\lambda \downarrow 0$, we have the *marginal conditional mean impact*:

$$\Delta_c^m(z, \widetilde{\pi}, \pi) = \lim_{\lambda \downarrow 0} \Delta_c(z, \pi', \pi)$$

and the *marginal mean impact*:

$$\Delta^m(z, \widetilde{\pi}, \pi) = \lim_{\lambda \downarrow 0} \frac{\Delta(z, \pi', \pi)}{\lambda},$$

implying that:

$$\Delta^m(z, \widetilde{\pi}, \pi) = \Delta_c^m(z, \widetilde{\pi}, \pi) \times$$
$$\lim_{\lambda \downarrow 0} \frac{Pr\{D(z, \pi') \neq D(z, \pi)\}}{\lambda},$$

(conditional mean impact crucially depends on the concept of directional limit, since standardization using λ is needed). Marginal effects can be approximated by $\frac{\Delta(z,\pi',\pi)}{\lambda}$ and $\Delta_c(z, \pi', \pi)$, for small values of λ. Anyway, the identification conditions of the marginal effects are weaker than those of $\Delta(z, \pi', \pi)$ and $\Delta_c(z, \pi', \pi)$.
Identification of policy impacts starts from their effects on the decision D. Defining ω as the stochastic element that drives the participation decision, D is influenced by ω, the policy Π and the covariates $\mathcal{Z}$, that authors divide in two separate vectors $(Z = (\widetilde{Z}, Z_\pi))$ depending on whether a given covariate is required to be equal or different across the two groups, as we shall see later. To identify the choice behaviour, the set:

$\mathcal{D}(\widetilde{z}, z_\pi, \pi', \pi) = \{(\widetilde{z}, z^*_\pi) \in \mathcal{Z} | Pr\{D(\widetilde{z}, z_\pi, \pi'; \omega) = D(\widetilde{z}, z^*_\pi, \pi; \omega)\} = 1\}$ is defined. This means we have to find the values z^* in $\mathcal{Z}$ such that $D(z^*, \pi)$ mimics the choice behaviour under the new policy, $D(z, \pi')$. The covariates with the same values between z and z^* are included in the vector $\widetilde{Z}$, while the others set Z_π, that is defined as *policy equivalent variation.* To identify z^*, structural assumptions are needed (understanding of the relationship between z and π); to guarantee it is not an empty set, data variability is required (at least in Z_π). The assumptions that authors enumerate are:

1. A PEV and the unobserved variation in choice variable are independent given some conditioning variables under two policies π and π':

$$Z_\pi \perp \omega | \widetilde{Z} \tag{2.12}$$

2. Being $\mathcal{Z}_0(z, \pi) = \{z^* \in \mathcal{Z} | Pr\{Y_0(z) = Y_0(z^*) | D(z^*, \pi) = 0\} = 1\}$, and: $\mathcal{Z}_1(z, \pi) = \{z^* \in \mathcal{Z} | Pr\{Y_1(z) = Y_1(z^*) | D(z^*, \pi) = 1\} = 1\}$, $\mathcal{Z}_0(z, \pi)$ and $\mathcal{Z}_1(z, \pi)$ are known, and their intersection with $\mathcal{D}(z, \pi', \pi)$ is nonempty for $z \in \mathcal{Z}$.

3. $\forall z \in Z$, either
   $Pr\{D(z, \pi') \geq D(z, \pi)\} = 1$ or
   $Pr\{D(z, \pi') \leq D(z, \pi)\} = 1$.

4. There exists $\lambda^*(z, \widetilde{\pi}, \pi) > 0$ such that Assumption 2 holds for all π' that correspond to λ such that $0 < \lambda < \lambda^*(z, \widetilde{\pi}, \pi)$.

Assumption 1 guarantees that the concept of probability is well defined for two different points z, z^*, and gives sense to the expression "policy equivalent variation". Assumption 2 states that $Y_1(z^*)$ and $Y_0(z^*)$ need to match $Y_1(z)$ and $Y_0(z)$, respectively. This is basically an exclusion restrictions: z_π must not influence Y, given D. Therefore, the policy equivalent variation must have an impact only on choice behaviour, not on policy outcome. Moreover, it tells us that there is a set $z^* \in \mathcal{D}(z, \pi', \pi) \cap \mathcal{Z}_0(z, \pi) \cap \mathcal{Z}_1(z, \pi)$ that allows us to identify the distribution of $Y(z, \pi')$, since we have:

$Pr(Y(z,\pi') < y) = Pr(Y(z^*,\pi) < y)$. Also identification of the distribution of potential outcomes is possible: under assumption 1 and 2, the distribution of $Y_0|D(z;\pi) = 0$ is identified if $Pr\{D(z,\pi') = 0\} > 0$, while identification of the distribution of $Y_1|D(z;\pi) = 1$ requires: $Pr\{D(z,\pi') = 1\} > 0$. If $E\{D(z,\pi')Y_1(z)\}$ and $E[\{1 - D(z,\pi')\}Y_0(z)]$ are finite, $\Delta(z,\pi',\pi)$ is identified.
Under Assumption 3, $Pr\{D(z,\pi') \neq D(z,\pi)\}$ is identified, and equal to:
$|Pr\{D(z,\pi') = 1\} - Pr\{D(z,\pi) = 1\}|$ (otherwise, only $Pr\{D(z,\pi') = 1\}$ and $Pr\{D(z,\pi) = 1\}$ could be identified, not their joint distribution).
Assumption 1,2 and 3 guarantee identification of the conditional mean impact $\Delta_c(z,\pi',\pi)$; if marginal impacts exist (continuous case), Assumption 1 and 4 allow us to identify the mean one (and also the conditional one, together with assumption 3). Since Assumption 4 is a weaker support condition than Assumption 2 (that is required to hold only in a neighborhood of the current policy value π), *ex-ante* evaluation is fit to study local effects. Its main features, anyway, are the fact that (differently from the standard matching procedure) two groups of non-treated are compared, and that variation in the data is required (the concept of PEV here is crucial).

2.9.3 The estimation procedure

The mean policy effect $\Delta(z,\pi',\pi)$ is given by the difference:
$E[Y(z,\pi') - E[Y(z,\pi)]$.
While $E[Y(Z,\pi)]$ can be estimated using the standard non-parametric regression methods, estimation of $E[Y(Z,\pi')]$ requires some discussion (since it is the average counterfactual outcome: noone is currently exposed to the policy π').
The counterfactual set $\mathcal{Z}^*(z,\pi',\pi) = \mathcal{D}(z,\pi',\pi) \cap \mathcal{Z}_0(z,\pi) \cap \mathcal{Z}_1(z,\pi)$ is sometimes known *ex ante*; in other cases, it must be estimated. After the definition of $\mathcal{Z}^*(z,\pi',\pi)$, the equality: $E[Y(z,\pi')] = E[Y(z^*,\pi)]$ allows the use of standard non-parametric regression methods as well, if $\mathcal{Z}^*$ is a singleton. If, contrariwise, $\mathcal{Z}^*$ has got multiple elements, $E[Y(Z^*,\pi)|Z^* \in \mathcal{Z}^*(z,\pi',\pi)]$ must be estimated across the different

points. Since:

$$\Delta_c(z, \pi', \pi) = \frac{\Delta(z, \pi', \pi)}{|Pr\{D(z, \pi') = 1\} - Pr\{D(z, \pi) = 1\}|},$$

estimation of $Pr\{D(z, \pi) = 1\}$ and $Pr\{D(z, \pi')\} = 1$ is required. The situation is analogous to the case of conditional outcomes above: $Pr\{D(z, \pi) = 1\}$ can be estimated directly, while estimation of $Pr\{D(z, \pi') = 1\}$ needs the use of the counterfactual set again: $Pr\{D(z, \pi') = 1\} = Pr\{D(z^*, \pi)|z^* \in \mathcal{Z}^*(z, \pi', \pi)\}$.
An estimation problem may arise from *curse of dimensionality*, i.e.: if the linear space including $\mathcal{Z}^*(z, \pi', \pi)$ is high-dimension.
A first idea is to give up on estimating the effects on people with $\mathcal{Z} = z$ ($\Delta(z, \pi', \pi)$ and $\Delta_c(z, \pi', \pi)$) and instead to condition on a larger set of observables than $\mathcal{Z}^*$, which can be estimated with smaller variance. Basically, the pointwise policy effects considered above are generalized, conditioning not on a point z, but on a set $S \subset \mathcal{Z}$.
So:

$$\Delta(S, \pi', \pi) = E(Y(z, \pi') - Y(z, \pi)|z \in S)\Delta_c(S, \pi', \pi) = E(Y(z, \pi') - Y(z, \pi)|z \in S),$$

$$d(z, \pi') \neq d(z, \pi)) = \frac{\Delta(S, \pi', \pi)}{Pr(d(z, \pi') \neq d(z, \pi)|Z \in S}.$$

The choice of set S obviously depends on the subgroup whose impact is estimable, but it may be also dictated by the group one is interested in studying. In fact, program impacts may be estimated only where the supports of $\mathcal{Z}$ and $\mathcal{Z}'$ overlap: the policy equivalent variation must be there in the available data. Then, averages across people within a subgroup of interest and the one of matches provide the average policy effect for that subgroup. The estimated effects are averages of the pointwise effects:
$E\{\Delta(z, \pi', \pi)|Z \in S\} = E\{E[\Delta(z, \pi', \pi)|Z = z]|Z \in S\} =$
$= E\{E[Y(z^*, \pi)|z^* \in \mathcal{Z}^*(z, \pi', \pi)]|Z \in S\} - E\{Y(z, \pi)|Z \in S\}$.
In this case, the curse of dimensionality may arise again, because we need to estimate either the conditional mean of $Y(z^*, \pi)$, or the density $g(z^*)$, that could be high-dimensional objects (even when $\mathcal{Z}$ is a singleton). Nevertheless, averaging should

yield higher convergence rates.
Todd and Wolpin (2006) propose a one-by-one match, building the *ex-ante* matching estimator:
$$\frac{1}{n}\sum_{i=1}^{n}(Y_{0j}(Z=z')-Y_{0i}(Z=z)).$$

An alternative approach is given by the use of parametric restrictions, typically on $E(Y|D,Z)$ and $E(D|Z,\pi)$.

2.9.4 An empirical example

An application of *ex ante* evaluation may be found in Todd and Wolpin (2006), using the PROGRESA experiment. It is a program implemented in Mexico, that provides cash transfers to parents conditional on their children attending school. In this setting, an *ex post* evaluation of the effect of cash transfer on household's decision is possible, because the program was implemented as a randomized experiment, in which 320 villages were assigned to treatment group and 186 to the control group. Authors exploit this experimental framework to study the performance of *ex-ante* evaluation methods and estimators, using the randomized-out control group, within which there is not any direct variation in the policy program. The PEV here is the variation in child wage offers w. The assumptions made are that a child who does not attend school is working in the labor market at wage w (equal to minimum wage paid to daily laborers in each village) and that household's decision is driven by utility maximization, depending on consumption (c) and an indicator for child's attendance to school (s):
$$max_s U(c,s),$$
with:$c=y+w(1-s)$ (being y household's income net of child's earnings). The optimal choice is: $S^*=\varphi(y,w)$. A conditional subsidy to attend school changes the budget constraint:
$$c=y+w(1-s)+\tau s.$$

So, the constraint may be rewritten as:
$c = (y+\tau)+(w-\tau)(1-s)$, which shows that the optimal choice in the presence of the subsidy is: $s^{**} = \varphi(\widetilde{y}, \widetilde{w})$, where: $\widetilde{y} = y + \tau$, $\widetilde{w} = w - \tau$. Under these assumptions, the schooling choice for a family with (y, w) that receives the subsidy is the same as the choice for an otherwise equivalent family with $(\widetilde{y}, \widetilde{w})$ who does not.

Chapter 3

Wage dispersion

Mincer (1974), using data from the 1950 and 1960 Censuses, related income distribution in America to human capital (schooling and work experience) among workers. The Mincer equations estimated in the 70's and the 80's showed large differences in wages across schooling and experience groups, which could be considered as productivity differences. Panel data sets on wages, that begun to be widely available to labor economists in the 80's, permitted a thorough analysis of the residuals of these equations, further showing that a large part of wage dispersion resulted from unobserved heterogeneity in individual ability and complex accumulation of idiosyncratic shocks. At that point, the competitive view of wages reflecting individual productivity still held (see Heckmann and Honoré, 1990).
On the contrary, with the advent of matched employer-employee data at the end of the 90's, systematic wage differentials both across individuals and across employers became apparent: the current models could not fully account for wage dispersion, due to perfect information assumption.

3.1 A model with individual fixed effects

Abowd, Kramarz and Margolis (1999, AKM hereafter) propose a standard error-component model with individual fixed effects:

$$w_{it} = x_{it}\beta + \sum_{j=1}^{J} \psi_j d_{it}^j + u_{it},$$

where i is an index for the worker, j for the firm, and d_{it}^j are indicator variables of worker i working at firm j at date t. The authors propose to estimate β, α and ψ by OLS, but the parameter space would often be too large. A necessary condition for the unbiasedness of the parameter estimates is:
$(d_{it}^j)_{\forall t,j} \perp (u_{it})_{\forall t}, \forall i \in \{1, \dots, I\}$. So, a basic requirement is u_{it} not to have impact on job mobility decisions: workers decide whether or not to change employers based on relative values of firm fixed effects ψ_j.
With exogeneity of mobility, the OLS estimator of β is consistent when I tends to infinity for fixed T; while the OLS estimators of α and ψ are consistent while T tends to infinity for fixed I and J. For an arbitrary variable z_{it}, defining
$z_{i.} = \frac{1}{T} \sum_{j=1}^{J} z_{it}$, we get:

$$w_{it} - w_{i.} = (x_{it} - x_{i.})\beta + \sum_{j=1}^{J} (d_{it}^j - d_{i.}^j) + u_{it} - u_{i.}$$

If $d_{it}^j = d_{i.}^j$ for all j (workers never change employer), we can not distinguish between firm and worker fixed effects. In practice, often $T < 10$, and workers are typically matched with two or three different employers. Therefore, the OLS estimates are expected to be very imprecise. OLS estimates of person effects α are given by:
$\hat{\alpha} = w_{i.} - x_{i.}\hat{\beta} - \sum_{j=1}^{J} \hat{\psi}_j d_i^j$. So, any statistical error affecting firm effects is transmitted to worker effects with a sign reversal: there is a negative spurious correlation between $\hat{\alpha}_i$ and $\hat{\psi}_{j(i,t)}$ in every year t. Lack of precision of the estimates arises when there is not enough worker's job mobility. Moreover, pointwise estimation of worker and firm effects is not very interesting: only the joint cross-sectional distribution

of $\hat{\alpha}_i$ and $\hat{\psi}_{j(i,t)}$ provides useful parameters to interpret. Data from France and US labor market (Abowd et. al, 2003) show that this model leaves a significant fraction of wage dispersion unexplained: there are systematic differences across workers and across firms that classical individual and market attributes are not able to account for. Although AKM model considers worker's and firm heterogeneity, there is a residual component of variance, that may reflect not only wage indeterminacy, but also productivity shocks and measurement errors. Computational burdens and noisy information may be avoided by allowing firm and worker's effect to be casual rather than fixed, or by considering time-varying firm covariates. If we use employer's mean log productivity ($\overline{y}_j$) instead of firm fixed effects ψ_j, we can evaluate the extent to which firm effects reflect differences in labor productivity. Considering productivity as a measure of firm effect ($\overline{y}_j = \psi_j + \eta_{jt}$), we can evaluate the explanatory power of productivity on wages. Since correlation between worker and firm effects is weak, there does not seem to be sorting of workers by firms. This model underlined the importance of quantitative firm-specific effects in wage determination, showing that the Law of One Price does not hold in the labor market, making the departure from the competitive paradigm necessary.

3.2 Equilibrium search models

Outside the competitive market, wages are no more equal to marginal productivity, so the problem of the determination of the way in which rents are shared arises. Mortensen (2003), in a review of this starnd of literature, argues that equilibrium search models are both realistic and empirically implementable enough These models rely on competition between employers, but they also limit such a competition, allowing search frictions to reflect information imperfection. By varying the intensity of search frictions, we get a broad array of equilibrium patterns, ranging between the *competitive wage equilibrium* (where workers force employers into complete competition, getting paid their marginal productivity) and *monopsony wage equilibrium* (where employed job search is infinitely costly and firms offer unemployed workers their reservation wages, equalizing utility of working and utility of unemployment: $W = U$).

3.2.1 The basic framework

Burdett and Mortensen (1998, *BM* hereafter) review previous developments in search equilibrium with frictions, unifying the literature about the level of unemployment with the one on wage dispersion. They emphasize mobility rather than focusing on stocks, following a "flow approach" to market analysis. Moreover, they describe the equilibrium search approach as a useful setting to study the effects of alternative labor market policy regimes on unemployment and its costs. Their goal is also the identification of the policy effects on unemployment duration and incidence (as a consequence of decisions made in both sides of the market) and the possible comparison of the effectiveness of different interventions.
BM model admits endogeneity of the wage offers distribution, assuming homogeneity of workers and firms. Time is continuous, and workers are infinitely lived. Unemployed workers sample job offers sequentially at some finite Poisson rate $\lambda_0 > 0$, while employed workers are allowed to search on the job, and face a rate of job offers $\lambda_1 > 0$. Matches dissolve at rate $\delta > 0$, then the workers become unemployed. λ_0, λ_1 and δ are exogenous. Each firm offers the same wage both to unemployed (that accept any job offer, prefering any wage draw to unemployment) and employed workers (that move from job to job only if this corresponds to a wage increase, because jobs are otherwise identical for the worker). When a match is created, the wage remains constant until job-destruction. If all the firms offered the same wage $R \leq w < p$ (where p is match productivity and R the common reservation wage), an employer offering a slightly higher wage would hire all the worker s/he contacts. If all the firms offered p, employer's profit would be 0, so for the single firm it would be better to decrease its wage offer, because they would make a positive profit, still hiring all the unemployed workers they contact. This implies heterogeneity of wage offers: there is no possible equilibrium where all firms offer the same wage. Firm profit per worker contacted is given by $\pi = (p - w)P(w|R, F)$, and wages are set in

order to maximize profit. If there were a mass point in a given value of $w < p$, a firm could increase $P(w|R, F)$ (by offering slightly more), reducing $(p - w)$ only by an amount of ϵ, as small as possible. No firm would offer $w \geq p$, in order to make a positive profit. So, the only possible equilibrium is given by a mixed strategy: there is a continuous nondegenerate distribution F of wage offers, implying wage dispersion among identical firms and workers. Anyway, the predicted wage density is upward sloping, which is at odds with empirical evidence.
If we assume a discrete distribution of firm productivity, we would still observe wage dispersion within the same productivity level, having that higher productivity firms offer higher wages. In fact, they have an higher probability of hiring contacted workers:

$$\pi_2 = (p_2 - w_2)P(w_2) \geq (p_2 - w_1)P(w_1) > (p_1 - w_1)P(w_1) = \pi_1 \geq (p_1 - w_2)P(w_2).$$

Then, authors extend their model to change the distribution of wage densities. The distribution of productivity types of firm p is assumed to be continuous . Wage offer maximizes firm profit, so w is a function of p ($w = w(p)$), and the sample distribution of wages and firm types are equal: $F[w(p)] = \Gamma(p) \forall p \in [b, \overline{p}]$. In equilibrium, the firm with the smallest productivity, $\underline{p}$, offers unemployed workers their reservation wage ϕ, and hires workers only from the unemployment pool. Free entry will ensure: $\underline{p} = \phi$. This model rationalizes any observed wage offer distribution as resulting from a properly chosen underlying productivity distribution Γ, provided that $w(p)$ is an increasing function. Anyway, the estimated distribution of firm productivity Γ exhibits a too long right tail, because the market power of firms is overestimated.

3.2.2 A sequential auction model

To temper this market power, Postel-Vinay and Robin (2002, *PR* hereafter) allow employers to counter the outside offers made to their employees. Moreover, they allow workers to differ in ϵ, their ability parameter, assuming that employers know ability and employment status of the worker they contact and (in case s/he is employed) also the productivity of the incumbent one, that, in turn, knows the type of the poaching firm. Wage contracts can be renegotiated by mutual consent only, and the marginal productivity of a match between an ϵ-type worker and a p-type firm is ϵp. A firm contacting an unemployed worker will offer him/her $\phi_0(\epsilon, p)$, i.e. just enough to make him/her prefer employment to unemployment. If a firm with (high enough) productivity p' meets a worker employed in a firm with productivity p, the two employers compete to hire the same worker. The more productive firm bids the worker away from the less productive one and pays the value-equivalent of the best wage the latter firm can offer, which equals match productivity. This implies that, if $\phi(\epsilon, p', p) < w$, the incumbent firm does not need to offer an higher wage to its employer. Otherwise, workers have an increased utility, due to the firms competition. If $p > p'$, they remain in the incumbent firm, but with an higher wage. If $p' > p$, they change employers, and they get all the surplus of the previous match. PR model is characterized by two main features: during job spells, wages may increase, due to outside-offers arrival (so, within-job wage dynamics are possible), but job-to-job changes may be associated with wage-cuts. Basically, workers only move up the *productivity* ladder, even if such move may correspond to a decrease in wage. This may happen because, in this model, mobility wage is less than the match productivity ϵp : if the worker was getting the full surplus in the previous firm, s/he will give up some income today, for the prospect of future wage raises ($p' > p$). Anyway, authors show that the mean wage paid by a firm to its employers is an increasing function $y(p)$, so job-mobility occurs only if the mean wage is higher in the poaching firm than in the incumbent one. To sum up, wage dispersion here is due not only to worker ability (ϵ) and firm type (p), but also to outside job offers (q). The third element (arising because of the presence of search-frictions) may explain why identical workers employed at identical firms can earn different wages: it depends on luck in drawing outside job offers. Applying this model to French register data, authors find this component may explain about 50% of wage variability, while person effect explains 40% of variability for managers, but it is negligible for unskilled categories. Even in PR model, there is no sorting: worker's ability ϵ is equally distributed in all firms. Formally, $(q, p) \perp \epsilon$, where q is the productivity of the employer from which the worker was last able to extract the full surplus in negotiation. Anyway, the authors also notice the predicted distribution of wage changes among job-movers first-order stochastically dominates the observed one, and the predicted distribution is more skewed than the observed one (putting into question the use of classical measurement errors). Moreover, PR model is not strong enough to fully explain neither wage cuts, nor wage raises for job stayers: in particular, data show job cuts for the latter, ruled out by the model.

3.2.3 Recent contributions

Postel-Vinay, Turon (2005, PT hereafter) add to PR model simple i.i.d. match-level productivity shocks. Due to on-the-job search and wage renegotiation by mutual consent, transitory productivity shocks are translated into persistent wage shocks. They assume that: $w_{it} = \alpha_i + \nu_{it}$, where the shocks $\nu_{i,t}$ are equal to $\nu_{i,t-1}$ with probability $(\overline{F}(\nu_{it}) - \frac{\lambda_1}{1-\delta}\overline{F}(\nu_{it}^2)$, and have got a continuous distribution conditional on $\nu_{i,t} \neq \nu_{i,t-1}$. Therefore, since wage shocks may also be negative, within-job wage cuts are allowed too. Conditional on p, wages follow a first order linear process, with specific acceptance-rejection scheme of i.i.d. wage innovations. Using data on high-educated workers of both genders along twelve years from the British Household Panel Study, PT fit the covariance structure of data amazingly well. Variance due to permanent earning shocks is significantly positive, and a difficulty of distinguishing between a linear process exhibiting a first root and other types of highly persistent processes arises (Baker, '97). PT model is similar to the ones describing wages as designed to allocate risk between a risk-neutral employer and a risk-adverse employee faced with uncertainty about match productivity and market opportunities.

Cahuc, Postel-Vinay and Robin (2006, CPR hereafter) merge administrative data with firm accounting data in order to obtain direct estimates of the firms productivity levels, instead of inferring them from observed wages and the structure of the model. The comparison of the so estimated productivity distribution with the ones determined by previous models may be used as a test for their validity. In CPR model, workers are allowed to have bargaining power. They negotiate with a single firm when they are unemployed: in case of outside job-offers, a three-player bargaining process is started. Bargaining power raises the value of the match got by the worker: $V(\phi;\epsilon,p') = (1-\beta)V(\epsilon p;\epsilon,p) + \beta V(\epsilon p';\epsilon,p')$, where $\beta \in [0,1]$ measures worker's bargaining power. If $\beta = 0$, the PR model holds: CPR model is a way to test it. This seems to hold for low skill categories, while it lies between 0 and 0.3 for high skill workers (data about France, 1993-2000). Further research is needed to explain bargaining power. CPR model fits the relation between wage and productivity very well; wage paid by the lowest-p firms in all sample and categories of workers is very close to match productivity, and profit rates increase with productivity. Thus, authors improve on previous models, but they remain focused on "cross-sectional" aspects of the data, failing to describe the process followed by individual wages over time properly.

Jolivet, Postel-Vinay and Robin (2005, JPR hereafter) use european and american data to study worker turnover and wage dispersion, using a simple search model whose assumptions are common to the one they review. They state one should account for:

1. workers' transitions (both from job to job, and in and out of employment), especially for the wage cuts that are sometimes associated with them;
2. the negative duration dependence of job separation hazard;
3. wage dispersion, and particularly the fact that the distribution of wages statistically dominates the distribution of entry wages, and is less positively skewed.

JPR model differs from the BM one because the distribution of wage offers F exogenous. In steady-state equilibrium, stocks are constant: workers' inflows and outflows from any given stock balance each other. For it to hold true for u (proportion of unemployed workers), it is: $u = \frac{\delta}{\delta+\lambda_0}$. Steady-state equilibrium of workers paid less than w gives: $G(w) = \frac{F(w)}{1+\kappa\overline{F}(w)}$, or:
$F(w) = \frac{(1+\kappa)G(w)}{1+\kappa G(w)}$, where $\kappa = \frac{\lambda_1}{\delta}$ is an index of search frictions (the average number of job offers received by a worker before a job destruction), $\overline{F} = 1-F$, and G is the distribution of wages. So, estimation of f can be constructed using g and κ. JPR estimates f from the sample of wages of employed workers who were just hired from unemployment, g from the sample of all employees' wages, and κ from job mobility

data. Wage offer densities are to the left of earnings densities, are less dispersed and more positively skewed. Discrepancy between f and g is near to the one predicted by JPR model. Data show that employed workers accept take-up jobs associated with better wages than unemployed workers, and this selection process is related to the process of job mobility (κ). Since wages are assumed constant over job spells, the model predicts negative duration dependence through wage heterogeneity: job spells with longer elapsed durations tend to be associated with higher wages, which in turn makes them more likely to last longer in the future. JPR model assumes constancy of wages over a given job spell, which implies that workers only move up the "wage ladder". So, within job dynamics and wage cuts are not allowed (as in the BM model). Moreover, negative duration dependence of job spell hazard rates is not fully explained by the model, and constancy of the rate of job offers is not consistent with observed duration profiles. So, constant exogenous search intensity seems not to be realistic.

Rogerson, Shimer and Wright (2005) survey the literature on search-theoretic models, considering various alternative sets of assumptions, analyzing analogies and differences among them and discussing their implications on predictions and efficiency. Another review on the literature on job-search models comes from Postel-Vinay and Robin (2006). Authors state that last twenty years of empirical literature on wage dynamics show that earnings shocks are highly persistent over time, and that a rich mix of random processes is needed to explain the intricate autocovariance structure of earnings. They also suggest some further possible future extensions of the theory, to better explain the economic mechanism at the root of these properties, to evaluate which set of assumptions is more suitable to describe a given labour market and to abandon som strong assumptions as perfect substituability among workers and constant returns to scale.

Summing up, to my purpose the key point from these theories is that wage dispersion arises even among workers who are equally acctractive from the point of view of firms.

Chapter 4

The policy regime under analysis

4.1 The institutional context

The Italian labour market is characterized by a strong protection of workers ”on the job”, both with the employment protection regulations[1] and with a generous pension system. In particular, the OECD’s Employment Protection Legislation (EPL) summary indicator for Italy at the end of the 1990s is one of the highest (3.1), mainly because of the value of the index regarding ”collective dismissal” (4.9, the highest among all OECD’s country[2]). Moreover, this protection varies depending on the size of the firm the worker is employed by: at the threshold of 15 employees there are several discontinuities in the Italian labour market regulations. In particular, workers of *large* (with more than 15 employees) firms have much more protection in case of unjustified dismissal[3]. The situation is utterly different if we look at welfare measures. In particular, the OECD index of net replacement rates (given by unemployment insurance and welfare benefits) is around 45% for Italy, while the OECD average (Martin and Grubb, 2001, Figure 3) is about 65%. Specifically, over the period I analyze, for workers not enrolled in the LM but having been previously

[1]see Ichino (2004) for protection against ”unjustified dismissal”.

[2]OECD (2004), p.105 and p.108. Data for 2003, p.70, show a strong decline of EPL strictness (from 3.1 to 2.4), entirely due to a reduction in the index about temporary employment (from 3.6 to 2.1). Previous estimates for Italy were upward biased; see also Del Conte, Devillanova, Morelli (2004).

[3]Law no.300,1970. See Schivardi and Torrini (2004).

employed for at least 24 months (with payment of social security contributions), the standard UI (30% of his/her last wage up to six months) is granted. People without this eligibility requirement (included unemployed workers looking for a job for the first time) do not receive any unemployment benefit. Also the public spending on active and passive labour market policies is modest, if compared to the European Union average[4], and it is characterized by traditional measures (training and recruitment incentives) and not by *welfare to work* and *mutual obligations* policies, that are meant not only to help people, but also to strongly encourage them to find a job, combining activation strategies with monitoring of job-seekers and enforcement of work tests (OECD, 2005, Ch.3).

4.2 The "Liste di mobilità " program

The policy regime under analysis is *"Liste di mobilità "*. It is an italian program introduced in 1991, mainly regulated by the laws No.233/1991 and 236/1993[5], and meant to support, for a predefined period of time, employees permanently dismissed by their employers .

Permanence in the program ends with the conclusion of the eligibility period, or when the worker is hired permanently. During their staying in the LM, they are allowed to engage in temporary jobs mantaining their LM status. When a LM worker is hired on a temporary basis, the clock measuring time in the LM stops, restarting as the job-spell ends. This implies that workers entitled to monetary benefits do not receive them until their temporary contract expires, but then they come back to the same situation as before. The duration of a single temporary employment spell has to be no longer than one year, and the total duration of temporary unemployment spells of an LM worker may not be longer than the maximum duration in the LM to which s/he is allowed.

Permanence in the LM requires, in principle, fulfilment of some obligations with re-

[4] see Martin and Grubb, 2001, Table 1; OECD, 2005, Statistical Appendix, Table H.

[5] provisions vary according to industry, worker's occupation and geographic area, and have been frequently modified over time. For details, see Anastasia *et al.*(2004), pp.49-64, Caruso and Pisauro (2005) and references therein.

spect to training and job offers: people who refuse an appropriate[6] job offer by the local public labour exchange should be dropped from the program. Nevertheless, these rules are not enfoced: actually, enrolled workers may refuse any job offer they receive without losing any benefit they are entitled to.
In the following, I use the notation from Mortensen, Pissarides (1999). Different values are taken by the policy parameters, depending on age and the kind of firm the worker was employed by when fired, as we will see below. Even the eligibility period depends on age. For the Northern and Central Italy[7], enrolled workers younger than 40 years, aged 40 to 49 and older than 49 are eligible for the LM benefits over one, two and three years respectively.
Extension of the eligibility period for older workers is based on the idea that they face more difficulties in finding a new job once they are dismissed by the previous one. Nevertheless, as pointed out by Paggiaro, Rettore, Trivellato (2007), the empirical evidence of a lower re-employment probability of women compared to men (at any age) is much more overwhelming, but the policy does not differentiate basing on gender.
Anyway, I drop the group above 50 from analysis, since their extended eligibility period is even longer if they are close to being eligible for retirement provisions (some of them are eligible to "long mobility", that allow them to draw monetary benefits up to retirement age): basically, the program is meant to bring them to early retirement, not to help them in finding a new job.
For people fired by firms with less than 15 employees (let us say: *small firms*), registration in the LM is voluntary. Anyway, evidence indicates that most of eligible workers dismissed by small firm do register in the LM. On the contrary, people collectively[8] fired by large firms due to firm restructuring or plant closing are automatically enrolled in the LM.
The program consists in two different part: a *passive* component (an unemployment compensation for the fired worker) and an *active* one (incentives for the firm hiring him-her), that I describe below.

[6]with respect to distance from the place of residence and closeness to previous job and wage.
[7]in Southern Italy, the eligibility period is one year longer, for each age group.
[8]From firms with at least five redundancies in a period of four months.

4.2.1 The passive component

People fired by large firms after being at least one year on a permanent contract[9] with the dismissing firm are entitled to receive monetary benefits. In fact, the program includes an unemployment compensation for the fired worker (ρw, where ρ is the replacement ratio, and w is the last wage earned by the worker before firing). During the first year, the replacement ratio is equal for everyone, with ρ equal to 0.8. People who are in their forties enjoy it also during the second year of unemployment, but with $\rho = 0.64$. Since they also enjoy a substantial reduction in the tax rate during the first year of permanence and, if they are eligible for two years, during the second year they are even exempted from social security contributions, the real take-home benefits are around 83% and 70% of the last take-home pay. Anyway, for higher wages the replacement ratio may be lower, because there are ceilings that vary over time. In particular, between 1995 and 1998, the contribution rates varies between 5.54% and 6.04%. In 1995, the maximum amount of gross benefits was £1.287.306 (664.84 €) for people earning less than £2.784.990 (1438.33 €) , £1.547.217 (799.07 €) for the others. Then, in each year, the gross value two ceilings and the threeshold are raised by the same proportionality factor. This implies the monetary benefit pattern is as shown in Figure 4.1[10]: the gross amount of unemployment compensation is 80% of the last pay earned for the lower wages; otherwise, workers receive an amount of monetary benefits equal to one to the two ceilings set for a given year.

4.2.2 The active component

The program also influences the employer's point of view. In fact, firms hiring workers fired by large firms on an open-ended basis receive 50% of the residual benefits the worker would have received if s/he had remained unemployed, but with a ceiling of one year. Moreover, employers hiring people in the LM (regardless of the size of the last firing firm, worker's age and the time spent in the LM) face a substantial reduction in τ (the rate of social security contributions): from the

[9] with at least six months of actual work.

[10] Unemployment insurance is set to one for workers in the first ceiling

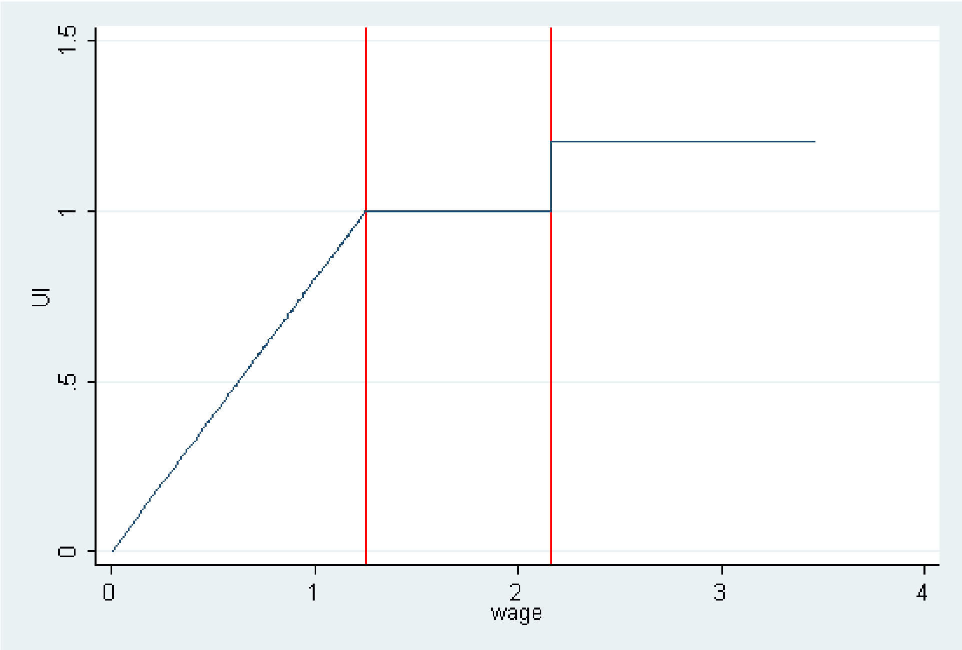

Figure 4.1: Wage vs gross unemployment benefits

standard rate (35% in 1998) to a fixed amount for apprentices (about 3% of the standard one), for a maximum of 18 months. If they hire him-her with a temporary contract, they just benefit from the reduction of τ until the contract expires (so, up to 1 year). They can cumulate rebates by hiring a worker on a temporary contract, and then switching to a permanent one as the former expires: in this way, the reduction of SSC for a given worker can last 2 years.

These incentives are substantial, so firms might be induced to fraudulently dismiss workers and enroll them in the LM in order to re-hire them, directly or indirectly (by newly created firms or affiliates), with the goal of drawing benefits. Law No.451/1994 has drastically reduced this possibility[11].

[11] re-hiring was formally forbidden during the first six months of enrollment in the LM, and social security agency has contrasted practices meant to evade the law.

4.2.3 Lack of "activation" strategies

Hazard to employment could be raised by employment bonuses or cash payments to UI recipients who find a job quickly and keep it for a given time period (Meyer, 1995). In many OECD countries, starting from the second part of the nineties, labour market policies has changed their eligibility rules towards greater "activation" (Martin and Grubb, 2001; OECD, 2005, Chapter 3-5), while the LM programs is basically consistent with previous italian public spending measures, neglecting activation principles (for example, failing to enforce eligibility requirements, as seen above).

4.3 The likely effect of the program on re-employment

Even if the duration of the eligibility period depends only on age at the moment of entering the LM, the main difference in the characteristics of the benefits set by the program is the one regarding the size of the last firing firm. In fact, while workers dismissed by large firms enjoy both the passive (monetary benefits) and the active (benefit transfer and reduction in the firm's social security contribution rate τ) component of the program, people fired by small firms are entitled only to a reduction in SSC for the hiring firm, so they just enjoy an active component. A further difference is relative to age: people above 40 are allowed to two years of permanence, the others to just one. A longer eligibility period and, above all, entitlement to monetary benefits may lead to be more willing to wait for a better job-offer, while approaching to the conclusion of the period of entitlement should make acceptance of a job-offer easier.
From the point of view of employers, hiring a worker on a two-step basis (first with a temporary contract and, after one year, with a permanent one) is always more profitable than offering the worker a permanent contract directly. In fact, this is the only way they can enjoy a reduction in SSC for two years (without losing the benefit transfer in case of people fired by large firms). The hiring incentive is obviously stronger if the worker has been dismissed by large firms (unless they are in their last day of eligibility in the LM, in which case they do not bring any monetary benefit

to the hiring firm). In this case, firms have an incentive to hire workers are soon as possible (the benefit transfer reduces as long as permanence in the LM of the worker goes by), and they prefer to hire individuals older than 40 to comparable younger workers (unless we compare individuals in their first day in the LM), because the benefit transfer decreases more sharply with permanence in the LM for people below 40.

This implies probability of re-employment should be affected by two contrasting incentives: the one for firms to increase job-offers to workers because of the subsequent benefits, and the one for workers to refuse the job-offers they receive (and both incentives are stronger for people entitled to receive monetary benefits, above 40 and at the beginning of their permanence in the LM), because their reservation reservation wage (the wage making people indifferent between working or not) is raised by participation in the LM. While the first effect (driven by the active component of the policy) raises re-employment probability, the second effect decreases it. Anyway, for people fired by small firms, only an active component holds, so that the dominant effect must be the one increasing re-employment probability.

For workers fired by large firms, on the contrary, the direction of the change in probability of finding a new job is uncertain. In fact, which component (the active or the passive one) of the program dominates is not *a priori* predictable. Moreover, the change of the amount of monetary benefits and benefit transfer in time make the net results of these incentives time-dependent.

Chapter 5

Empirical analysis

5.1 The data

The analysis is done in the Veneto region, using a linkage between data from a *Netlabor* (public labour exchange archives) and administrative data files, coming from *INPS* (National Institute of Social Security). Veneto is a well developed region in the NorthEastern Italy, with a tight labour market and a *per capita* GDP above the national average. In the second part of the nineties, the difference between regional and national employment rate has always ranged between 6% and 7%, and the Veneto unemployment rate has constantly been much lower than the italian one (in 1998, the employment rate was 59.6% compared to 52.9% in Italy, while unemployment rate was 6.1% *vs.* 11.4% in the whole country[1]).

The list of workers comes from the Netlabor file. Observation of individuals entered in the LM until 2005 is available, but for people entered in the program after 1998 we can reconstruct job history following enrollment in the LM for less than 3 years, while years until 1994 show serious problems of data quality. For this reason, attention is focused on workers enrolled in the LM in the period 1995-1998. Labour market histories (from 1975 to 2001) of workers are reconstructed from the INPS files. Only job spell in the private sector with a temporary or permanent contract are recorded:

[1]For working ages, 15-64. These data are revised official estimates from the Labour Force Survey.

self-employment, public sector employment or non-regular jobs are not observed (we can not distinguish between unemployment and these job experiences). The matching has been found for 27,565 workers (93.3% of the total). Then, individuals with more than one episode of mobility in the LM (about 3% of workers and 6% of episodes) are removed, since their past and/or future history is different from the one of workers enrolled just once. Moreover, only people whose enrollment date in the LM is consistent with an event of firing noticed in INPS are kept, and individuals with more than one firing event in the same month are dropped (since identification of the entry event in the LM is not straightforward for them). Resulting matches are 24,178 (81.8% from the original Netlabor file), from which 534 (about 2%) further workers are dropped, due to problems of missing data or inconsistencies in working histories, leading to a sample of 23,644 enrolled workers (80% of the original population, 86% of workers found also in Inps archives).
A pay episode for year is available: a sequence of 12 ”dummy” variables for whether the worker is paid or not for that month, the number of paid days and weeks, global yearly pay. As a consequence, a multi-year employment episode is divided into many yearly records with their relative pay, while during a year several records for each individual are possibly listed. This may happen in case of:

1. job change (episodes separated in time);

2. overlapping episodes (for example, in case of part-time jobs);

3. accounting particularities leading to duplication or partition of the information concerning a single episode, much more frequent in cases of firing.

The episode length seems to be better identified by paid weeks than by paid days: their value is more realistic.

5.2 Which policy changes can be ex-ante evaluated?

The goal of my analysis is to predict the impact of a change in the monetary benefits the LM workers are entitled to (so, the amount of monetary benefits has to be considered as the treatment T) on the probability of re-employment. The average yearly flow of workers enrolled in the LM in Veneto is less than $8,000$, while the average stock of unemployed workers in the region is around $100,000$. So, the effect of the current program (and of potential reforms to it) on the whole labour market can be assumed to be negligible. Using the notation from Ichimura, Taber (2000), this implies that the assumption: $(Y_0, Y_1) \perp \pi | D$ is sensible, i.e. we can assume a change in the current policy addressed only to already eligible people does not alter potential outcomes. A non-parametric approach requires, for a given individual i whose we mean to identify the behaviour under a counterfactual policy regime, the possibility of finding an individual j that mimics it with his/her own current situation.

Workers in the dataset may be divided into eight groups, based on gender, eligibility period (two years vs one year), entitlement to benefits.

In the group with only one year of eligibility, people below 30 years old are dropped, because their past history is likely to be too short (previous job history for 6 years before enrollment in the LM is used in the set of covariates in the identification strategy based on the propensity function).

Let me define the policy parameters $\rho(t)$ (the pattern of replacement ratio in time, being t the time from entrance in the LM, using years as time units), h, H (the fraction and the total amount of the potential future unemployment benefits that is transfered to the hiring firm) and τ (social security contribution rate for firms hiring people in the LM).

The effect of a change in τ may not be evaluated with the available dataset, since this parameter is constant for all workers in the program, regardless of their characteristics. This also implies that we can not measure the impact of the policy in its whole (that is, being the counterfactual: non-partecipation in the LM).

For people fired by small firms, the policy parameters ρ, h, H are always 0: unemployed workers just receive the standard unemployment compensation, and firms hiring them do not receive any direct subsidy by hiring them. Moreover, people fired by large firms can be considered as “different” from them (see above). Therefore, in the group of people not receiving any monetary benefit from the program,

identification is possible for the sole changes in the eligibility period (especially, 1 year vs 2 year).
For the problems arising towards identification of a change in h, a comparison between people with different current time-spells of permanence in the LM, considerations about the extension of the current policy to the whole population of unemployed workers (or a non-negligible fraction of them) or arbitrary changes to the policy parameter time patterns, see Appendix A.
The most promising comparison seems the one among individuals eligible to receive monetary benefits in the same group (for gender and eligibility period) to detect the effect of a change in the replacement ratio $\rho(t)$, keeping h constant. For individuals to be comparable (same transfers and same reservation wage, conditioning on X), we need the policy we mean to evaluate to have the same pattern of the factual one: $\rho(t) = \rho_1$ for the first year, and, during all the second year: $\rho(t) = \rho_2 = 0.8\rho_1$ for people above 40 ($\rho(t) = 0$ for people below). In this case, to simulate a change in the replacement ratio for the first year from $\rho(t) = 0.8$ to $\rho(t) = \rho^*$, the counterfactual for individual i would be an (otherwise identical) individual j such that: $0.8w_j = \rho^* w_i$.
Therefore, for people below 40, identification is possible for a policy setting a replacement ratio lasting one year and constant over time: $\rho(t) = k$ for $t \leq 1$, $\rho(t) = 0$ for $t > 1$. On the contrary, for people above 40, we can identify a replacement ratio with the same pattern of the existing one (in the second year the subsidy is just 80% of the first one):
$\rho(t) = \rho_1$ for $0 \leq t \leq 1$; $\rho(t) = \rho_2 = 0.8 * \rho_1$ for $1 < t \leq 2$; $\rho(t) = 0$ for $t > 2$.

5.3 The model setting

To make individuals comparable, we need to eliminate the selection bias deriving from the fact that workers with different wages at the moment they enter in the LM may be different with regard to probability of re-employment even in absence of the monetary benefits.
The *propensity function* setting from Imai, Van Dyk (2004) could be suitable to

evaluate the effect of changes in the amount of monetary benefits b at the individual level. Evaluation of the effects of a policy change (from the current policy regime π to a new one, with a new value of replacement ratio ρ') needs a generalization of the *ex ante* evaluation setting from Ichimura, Taber (2000) to the case of an arbitrary distribution of the treatment (here, the amount of monetary benefits received).

5.3.1 Assumptions required

For any possible new policy π' differing from the current one only in the amount of replacement ratio, the following assumptions are required to *ex ante* identify the average effect of the possible reforms relying on the "selection on observables" restriction:

- SUTVA: (2.7): independence between the distribution of potential outcomes (probability of re-employment under a given benefit regime) for one worker and treatment status (the benefit pattern received) of another worker, given age, month of entrance in the LM and past job history included in X($work_t$, $wage_t$).

- Strong ignorability (2.8): the distribution of potential outcomes (probability of re-employment under a given benefit regime) for a worker and his/her treatment status (the benefit pattern received) are independent, given age and past job history included in X($work_t$, $wage_t$). (2.9): any set of benefits with positive measure has a positive probability for any value of the covariates.

- (2.11): the policy regime affects the outcome only through the received benefit scheme. This assumption can be made as long as we study the effects of changes in a small fraction of the labour force.

- Here, there is not unobserved variation in the benefit transfer received, as long as (for people in the same group) it is a deterministic function of the policy regime π, year of entrance in the LM and last wage earned, so condition (2.12) is trivially satisfied.

- Assumption 2 from Ichimura and Taber (2000) [2] needs an extension to the continuous case. Being:
  $\mathcal{Z}(z, \pi) = \{z^* \in \mathcal{Z} | Pr\{Y_b(z) = Y_b(z^*) | D(z^*, \pi) = b\} = 1\}$, $\forall b$ in D, $\mathcal{Z}(z, \pi)$ is known, and its intersection with $\mathcal{D}(z, \pi', \pi)$ is nonempty for $z \in \mathcal{Z}$. Here, the assumption needed is: the difference in the last wage corresponding to a change in ρ do not to alter potential outcomes (conditioning on covariates).

- Assumption 3 is satisfied: $\forall z \in Z$, either
  $Pr\{D(z, \pi') \geq D(z, \pi)\} = 1$ or
  $Pr\{D(z, \pi') \leq D(z, \pi)\} = 1$ holds, because, if the replacement ratio is increased, the UI gotten by the worker may not decrease, and *viceversa*: $\rho' > \rho \rightarrow D' \geq D$, $\rho' < \rho \rightarrow D' \leq D$.

- Assumption 4 requires Assumption 2 to hold at least in a neighbourhood of π.

5.4 Evaluation of a reform changing the replacement ratio

The first step of the analysis is the estimation of the amount of benefits received in the first year, that is calculated as a function of the last wage before firing, year of entrance in the LM, the ceilings, threshold values and social security contributions for each year. Then, this subsidy is transformed, in order to standardize for price level: year 2003 is used as benchmark. With regard to the group of people entitled to monetary benefits, since workers who receive higher subsidies have an higher wage at the moment they are fired, they are plausibly more skillful than the others, so a selection bias may arise. This implies the need to control for ability of workers, but we obviously do not observe such a variable. Therefore, to reduce bias controlling for covariates, we can use a set of covariates X (age, past history, month of entrance in the LM). Since the dimensionality of this vector may be quite large, appeal to generalized (given the amount of monetary benefits is a continuous variable) propensity score is more suitable.

[2] I refer to their paper also for Assumption 3 and 4.

Finally, we need to study the distribution of the last wage earned in terms of past history. I assume that the distribution of logwage depends on X only through its mean and its variance. So, the propensity function is characterized by no more than two parameters.
I use individual job histories starting from 6 years before entering in the LM up to 3 years afterwards, and use the notation t and t^*, depending on whether periods are before or after entrance in the LM. I use previous job history, together with age, to build variables that may bring to less biased estimates.
Therefore, interest is on job history during 10 years. I estimate the mean weekly wage and the number of worked weeks for each of these years. For each year, I build a binary variable ($work_t$) regarding whether the person has ever received a wage in the year ($work_t = 1$ if the worker has received a wage in year t, $work_t = 0$ otherwise). Moreover, I define a continuous variable ($wage_t$) as the average weekly wage in year t (setting this quantity to 0 if the individual has never worked in the period), and $week_t$ as the number of paid weeks in the year. Of course, $wage_t$, $work_t$ and $week_t$ are either all equal to 0, or all positive. In this way, the year of entrance in the LM is the 7th. I include information related to the first 5 years observed in the set of covariates (wage in the 6th year is closely related to the one of entrance in LM, since, at least for a period, the firm where workers are employed is the same that will fire them the following year).
Therefore, at the moment of entrance in the LM, 17 variables are available for the set X: 15 relative to previous job history ($work_t$, $wage_t$ and $week_t$), *age* and *month* of enrollment in the LM (from 1 to 48, since the period goes from January 1995 to December 1998).

5.4.1 Testing the validity of the propensity function approach

The propensity function setting by Imai-Van Dyk (2004) is meant to lead to unbiased estimates. A possible validation strategy is feasible using the four groups of individuals not receiving any monetary benefit from the program. In particular, if the propensity function eradicates all the bias arising from the confounding factors, the estimated effect for these groups of the last wage before firing on probability of re-employment must turn out statistically equal to zero (since it does not correspond to any monetary benefit coming from the current policy regime[3]). In case the estimates are significantly different from 0, this would imply that selection bias arises for people who do not get any benefit from the program, thus making estimates of the impact for the corresponding (same gender and age group) groups receiving monetary benefits less trustworthy.

Balance

The first step is the application of a Box-cox regression of $log(W)$ on the set of covariates. The balance is checked by studying the correlation between the inverse of the residuals of the regression and each dummy variable ($work_t$, $1 \leq t \leq 5$), then considering the partial correlation between the residuals and two transformations of the other covariates (the logarithm and the inverse)[4]. In case of bad balance, interaction and square and cubic terms are added one by one. If needed, the square residuals from this estimation procedure are then used to estimate the log wage variance in the same way.

5.4.2 Difference in differences approach

An alternative identification strategy may be based on the hypothesis that the selection bias pattern does non depend on entitlement to receive monetary benefits. To make it clear, let us define D as the dummy variable for receiving monetary benefits ($D = 1$ for people entitled to have them, $D = 0$ otherwise) and $B = B(w, year)$ as the amount of monetary benefits received by a worker entitled to receive the subsidy who enters in the LM in a given year with a given last wage earned. Basically, if re-employment probability at time period t is additive in the last wage earned and the amount of monetary benefit received, so that:$EY|D = 1, w = f(w) + g(B)$ for

[3]There is not any benefit transfer to firms, and I assume here that the standard unemployment compensation is so poor that its variation due to a different last pay does not alter the reservation wage.

[4]Balance for variables regarding the number of worked weeks and the weekly pay is considered conditional on working in the given year.

people with $D = 1$ and $EY|D = 0, w = h(w)$ for workers with $D = 0$, I have to evaluate if:$f(w) = h(w)$, meaning that the dependence on the last pay is the same amongthe two groups ($D = 1$ and $D = 0$). This may allow us to estimate the effect of unemployment benefits on re-employment probability including also people not entitled to receive them, in a model similar to (2.4). Here, defining D as the dummy variable for receiving monetary benefits ($D = 1$ for people entitled to have them, $D = 0$ otherwise), w as the last wage earned and UI as the amount of monetary benefits gotten by a worker, we can replace the variable $treat$ with D (here, treated people are the one receiving monetary benefits) and the variable $after$ with w (here, bias is not determined by changes in time, but by the last wage earned), and consider the further difference that we do not mean to measure the effect of the last wage earned, but of the benefits received.
So, the final model would look like:

$$y_i = \beta_0 + \beta_1 * D + \beta_2 * w + \beta_3 * UI + \epsilon. \tag{5.1}$$

The particular pattern of monetary benefits implies the possibility to check for the presence of bias even within the groups of people receiving the subsidy. For example, for workers entered in the LM in 1998, we have a group of individuals receiving an amount of monetary benefits of £1.325.749, and another one getting £1.593.422. Let us define: $B = B_i$ ($i = 1, 2$) if a worker is in $i^t h$ ceiling (or would be, since we consider also workers not entitled to receive them). For each year, the variable B depends only on last wage earned: $B = B(w)$. Being Y_t our outcome variable (probability of finding a job in the t^{th} semester after dismissal), we can evaluate if, $\forall (w_1, w_2) \in \mathrm{W}|B(w_1) = B(w_2) = B(1)$ and $\forall t$,

$$E\{Y_t|w = w_1, D = 1\} - E\{Y_t|w = w_2, D = 1\} =$$

$$= E\{Y_t|w = w_1, D = 0\} - E\{Y_t|w = w_2, D = 0\};$$

and, analogously, $\forall w_3, w_4 \in \mathrm{W}|B(w_3) = B(w_4) = B_2$,

$$E\{Y_t|w = w_3, D = 1\} - E\{Y_t|w = w_4, D = 1\} =$$

$$= \{Y_t|w = w_3, D = 0\} - E\{Y_t|w = w_4, D = 0\},$$

i.e.: whether selection bias in the two groups is the same, both among people receiving the first ceiling and among people getting the highest amount of benefits for that year.
The same hypothesis of the same pattern of selection bias between $D = 0$ and $D = 1$ may also be evaluated using probability of being employed in time periods t not affected by the policy regime, i.e.: before entrance in the LM. In this case, the last wage does not correspond to any monetary benefit. We have to test if: $\forall(A, B)$, $\forall t^*$,

$$E\{Y_{t^*}|w = w_1, D = 1\} - E\{Y_{t^*}|w = w_2, D = 1\} = \tag{5.2}$$

$$= E\{Y_{t^*}|w = w_1, D = 0\} - E\{Y_{t^*}|w = w_2, D = 0\},$$

Finding a different pattern for the groups with $D = 0$ and $D = 1$ would make the "diff-in-diffs" estimate of the effect of the monetary benefits lest trustworthy.

5.5 Analysis in the group of "young" women

I show the results for the group of women below 30, that is the only one for which I found both a valide test (in the sense explained above) and some significative estimated effect of the unemployment subsidy on re-employment probability.
After getting a good balance (no correlation is significative at 5% level), I generate the variable $var(W)$ (given by square residuals of the final Box-Cox regression) and study the dependence of the estimated variance on covariates. Regressing this new variable on X, the observed level of significance of the set of covariates is 0.2039, while the one on the estimated logwage prediction is 0.0451. Moreover, the partial correlation between covariates and variance conditional to the predicted logwage is never significant at the 10% level, apart for the variable related to month of entrance in the LM. Therefore, defining the prediction of logwage using covariates as *prop*, dependence between the actual logwage and covariates conditional on *prop* seems to be non-negligible for *month*, while a good balance is gotten by the other covariates. So, even after controlling for the propensity function, the distribution of time of

entrance in the LM may still be a cause of bias.

5.5.1 Validation test for the propensity function

To test if the propensity function approach may lead to unbiased estimates, the group of workers not entitled to receive monetary benefits is used. I control for the variable *prop* together with time of entrance in the LM, building dummy variables for each year (generating the variables $year_i$,
$i = 1995, 1996, 1997, 1998$, such that $year_i = 1$ if the year of entrance is i, $year_i = 0$ otherwise).
In case this approach eradicates the bias arising from confounding factor, after controlling for *prop* and $year_i$, the effect of the last wage earned on probability of re-employment should turn out to be null (see Section 5.4.1). Re-employment probabilities are considered in the six semesters after entering the LM (generating the dummy variables:
$post_t$,$1 \leq t \leq 6$, where $post_t = 1$ if the individual receives a wage in the t^{th} semester, $post_t = 0$ otherwise).
A model including the estimated propensity function up to the third degree together with year of entrance seems to capture the relation between re-employment probability and the propensity function itself (see B.1-B.6):

$$Y_t = \alpha + \beta prop + \beta_1 prop2 + \beta_2 prop3 + \beta_4 year1996 + \beta_5 year1997 + \beta_6 year1998 \quad (5.3)$$

Defining the logarithm of the last wage earned as *logpay*, including in the logistic regression 5.3 the last wage and its squared term ($logpay, logpaysq$), they are always jointly significant at 5% level (see B.1). This implies that, in the group of young women not receiving any monetary benefit from the program, we are not able to control for the bias arising from omission of covariates with the use of the propensity function.

5.5.2 Testing the non-interacted model

In a logistic regression of the probability of re-employment in given months (24, 36, 48, 60, 72) before entering in the LM, I evaluate the joint significance of the interaction of the polynomials of logarithm of the last wage earned *logpay* (respectively of first, second, third and fourth degree) with the dummy for being eligible to monetary benefits D (to validate the "diff-in-diffs" hypothesis made in Section 5.4.2). In the linear case, it is:

$$Y_{t^*} = \alpha + \beta_1 D + \beta_2 logpay + \beta_3 logpay * D \tag{5.4}$$

Significance at the 5% level is found only in the linear and quadratic function of the prediction of employment 72 months before enrollment in the LM (see Appendix C.1), implying the bias due to last wage earned is statistically the same in the two groups (at least for periods previous to entrance in the LM).

5.5.3 Estimation of the impact of a policy change on probability of re-employment

Defining as $logUI$ the logarithm of the received amount of monetary benefits (net of social security contributions), a second degree polynomial in this variable is fitted, in the logistic regression of probability of re-employment on last wage earned (entering the regression with a fourth degree polynomial):

$$Y_{t^*} = \alpha + \beta_1 D + \beta_2 logpay + \beta_3 logpay2 + \beta_2 logpay3 + \beta_3 logpay4 + \beta_4 logpUI + \beta_5 logUI2 \tag{5.5}$$

I find significance in the first two semesters (see Appendix D.1). The estimated parameter values may be used to simulate the effect of an increase of 1% in the unemployment compensation received for young women fired by large firms. The average overall probability would approximately change by 0.3% both in the first and in the second semester (the estimated average re-employment probability raises from 0.468661 to 0.4717946 in the first semester, from 0.5633903 to 0.5662506 in the second one).

Appendix A

Possible evaluation of different policy effects

A.1 Reforms to the rate of the transfer to firm

Using the notation of Section 5.2 and being X our set of covariates and Y probability of re-employment in given time-periods after enrollment in the LM, let me start the analysis with the possibility to evaluate the effects of a change in the hiring subsidy parameter h for people receiving the subsidy leaving the other parameters unchanged, calling h^* the potential new value of h. In the factual policy regime, it is: $h = 0.5$ (the hiring firm receives half of the residual benefit transfers the worker is entitled to), but there is a ceiling of one year. So, the amount of benefit-transfer to firms is: $H = 0.5w[0.64 \times t + 0.8 \times (1-t)], t \leq 1; H = 0.5w[0.64 \times (2-t)], 1 < t < 2$. For people below 40, it is:$H = 0.5w[0.8 \times (1-t)], t \leq 1$; 0 for $t > 1$. So, a change in h would raise both worker's subsidy and the benefit transfer to the hiring firm. We need to detect an observed individual j that must simulate a policy equivalent variation for individual i. This means worker j is required to bring to the potential hiring firm the same transfer individual i would do under the new policy regime: $h^* w_i = 0.5 w_j$, but that this difference must not influence potential outcomes, conditioning to our set of covariates X: $Y_i(H) = Y_j(H)|X_i = X_j$. Nevertheless, even if these individuals were identical with respect to X (to guarantee the different value of their last wage does not mean different ability), they receive a different level of unemployment compensation, so they have different reservation wages: their decision about job-acceptance would not be the same. We could try to extend the comparison to workers already entered in the LM, and with different current time-spells of permanence, but even when: $0.5w_j \sum_{k=t_j}^{t_j+1} \rho_k = h^* w_i \sum_{k=t_i}^{t_i+12} \rho_k$ (j, under the current policy regime, brings the same transfers of i in the new one), the value of their reservation wage must be the same, for them to be comparable (they need to have the same decision rule about whether accepting a job-offer or not). Anyway, these people are in different moments of their entitlement in the LM: to evaluate their comparability, we would need to model the reservation wage, as a function of future potential unemployment benefits and transfers to firms: $w_r = f(\rho(t), H)$.

A.2 Reduction of the duration of entitlement period

Another possibility would be to compare people fired by large firms with the same pattern of future $\rho(t)$, assuming the vector of covariates X is such that the probability of receiving offers and the acceptation rule do not depend on history in the LM so far, given X. For example, we could identify the impact of such a policy: reducing to six months the period in which workers in the LM (above 40, fired by large firms) enjoy: $\rho = 0.8$. We could identify their behaviour via their factual observable equivalent, that is: people who did not get a job after six months in the LM. Anyway, since people who are not able to get a job in a given time period are likely to be less skillful than workers just entered (or with a shorter permanence) in the LM, problems in finding a region where the two supports of X overlap are likely to arise.

Evaluation of the effects of other policy changes about $\rho(t)$ (possibly using counterfactual individuals also for single unemployment periods, not only for the whole spell in the LM) would require again the expression of the time-varying reservation wage as a function of future potential benefits and incentives to firms.

A.3 Structural models

Basically, the characteristics that the counterfactual j worker should have to mimic a specific change in such pattern for a given i individual is: same attractiveness for firms (given by possible future benefits and productivity) and same reservation wage i would have under the new policy regime.
To extend the group of policies whose impact is to be evaluated, we should express attractiveness of individual i at time t with an unidimensional value q_{it}.
Modeling re-employment probability would require the consideration of two opposite effects: being more attractive from the employer's point of view raises the probability of being hired, while higher reservation wages decrease that probability. Moreover, the reservation wage itself also depends on worker attractiveness.
Basically, the consideration of the effects of the policy in a structural scheme (so, in case it involved the whole population of unemployed workers, or a non-negligible fraction of them) seems even more troublesome. In fact, as underlined by Mortensen (2000), to analyze equilibrium effects of policy interventions, endogenization of contact and job destruction rates would be required, but, since endogenous variables are distributions, i.e. infinite-dimensional parameters, it is very hard to analyze their dynamics (Shimer, 2003).
Even assuming constant marginal productivity and contact rate and supposing information on worker's heterogeneity and dynamics (so, human capital accumulation) and on firm characteristics is available, the problem of estimation of firm productivity and of job destruction and job-to-job transition in the period with tax reduction would arise. Finally, as Mortensen (1977) and Burdett (1979) point out, an higher amount (or duration) of unemployment compensations not only have the well-known direct effect of increasing reservation wage, but also the reversal indirect effect of decreasing it for unemployed workers who know they can receive the subsidy after a dismissal.

Appendix B

Results on the validation of the propensity function approach

B.1 Fractional polynomial regression

To evaluate if equation (5.3) properly describe the relation between the estimated propensity function and the probability of re-employment in each of the six semesters after enrollment in the LM, a fractional polynomial (Royston and Altman, 1994) with 4 degree of freedom is used.

In this approach, power terms are restricted to a small predefined set of integer and non-integer values, selected so that conventional polynomials are a subset of the family. The set of fractional polynomial powers from which I select the model leading to the best fit (in terms of likelihood) is: (-2, -1, -0.5, 0, 0.5, 1, 2, 3)[1], as suggested by Royston and Altman (1994). As outcome variables, the Pearson's residuals (adjusted for number sharing covariate pattern) of the logistic regressions of $post_t$ ($1 \leq t \leq 6$) on the third degree polynomial on *prop* and on the dummy variables $year_i$ ($i = 1996, 1997, 1998$) are used. Confidence interval at 95% level are also drawn.

[1] 0 means the logaritmic transformation

B.2 Results

Figure B.1-B.6 show there is not a statically significant relation between the estimated propensity function and the residuals of (5.3) for the group of young women not receiving any monetary benefit.

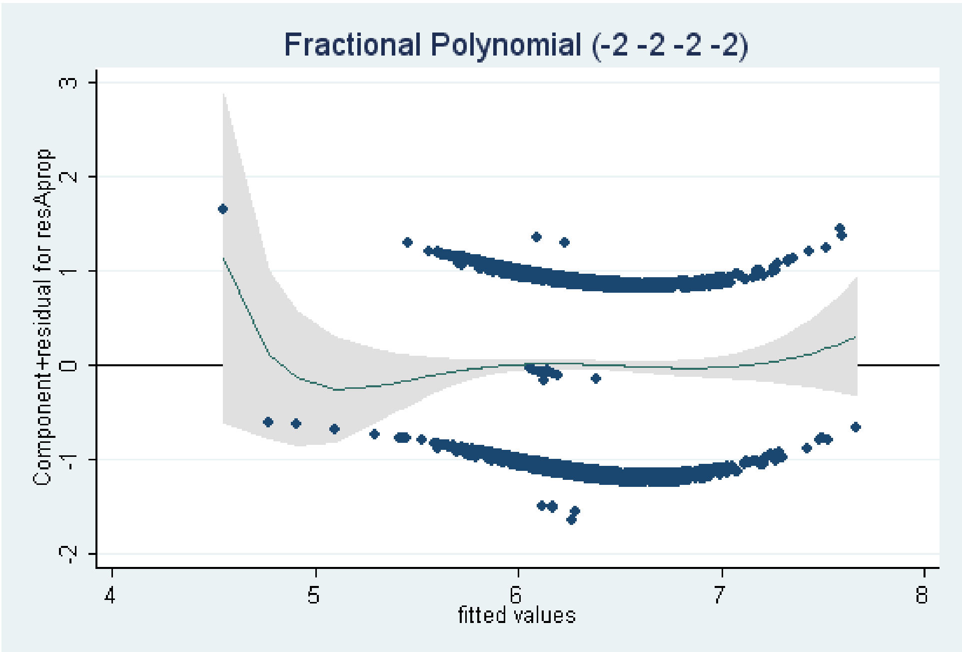

Figure B.1: Residuals from equation (5.3) vs *prop*, first post-program semester, young women

B.2.1 Results on the residual bias

The model where a two degree polynomial on *logpay* is added to (5.3) shows the control variables used are not enough to eradicate the bias, since the quadratic polynomial on the last wage earned is always significant.

Semester after enrollment in the LM					
1	2	3	4	5	6
0.0021	0.0024	0.0279	0.0016	0.0007	0.0006

Table B.1: Observed significance level of the quadratic polynomial of *logpay* on re-employment probability after controlling for covariates

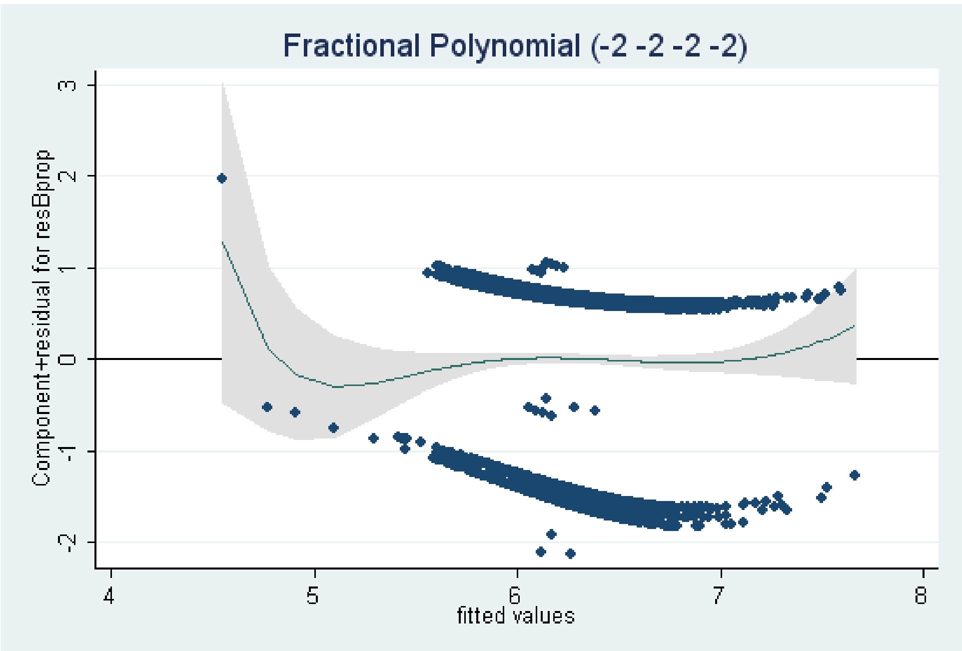

Figure B.2: Residuals from equation (5.3) vs *prop*, second post-program semester, young women

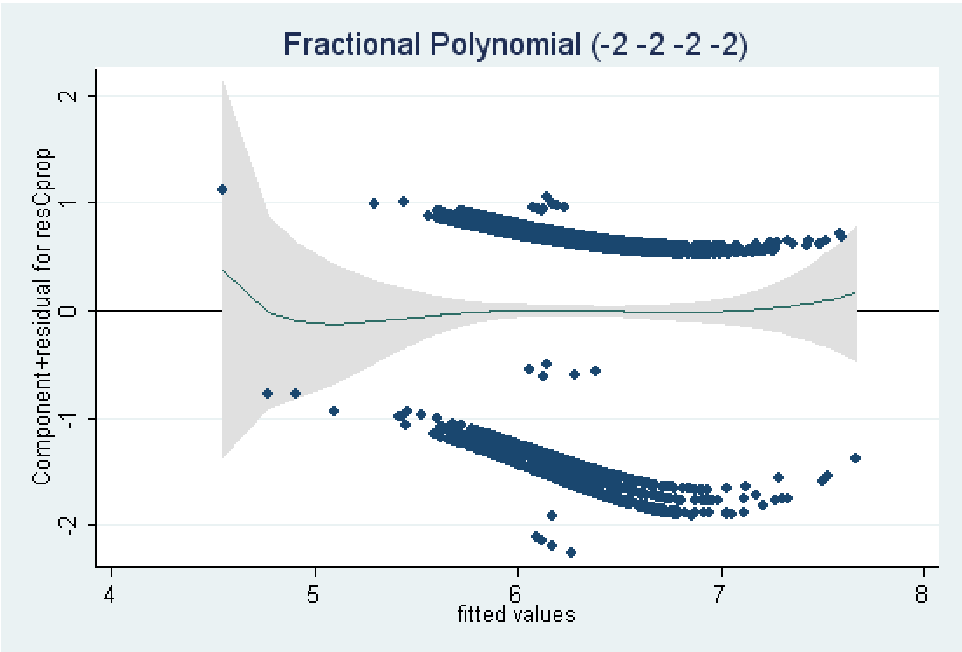

Figure B.3: Residuals from equation (5.3) vs *prop*, third post-program semester, young women

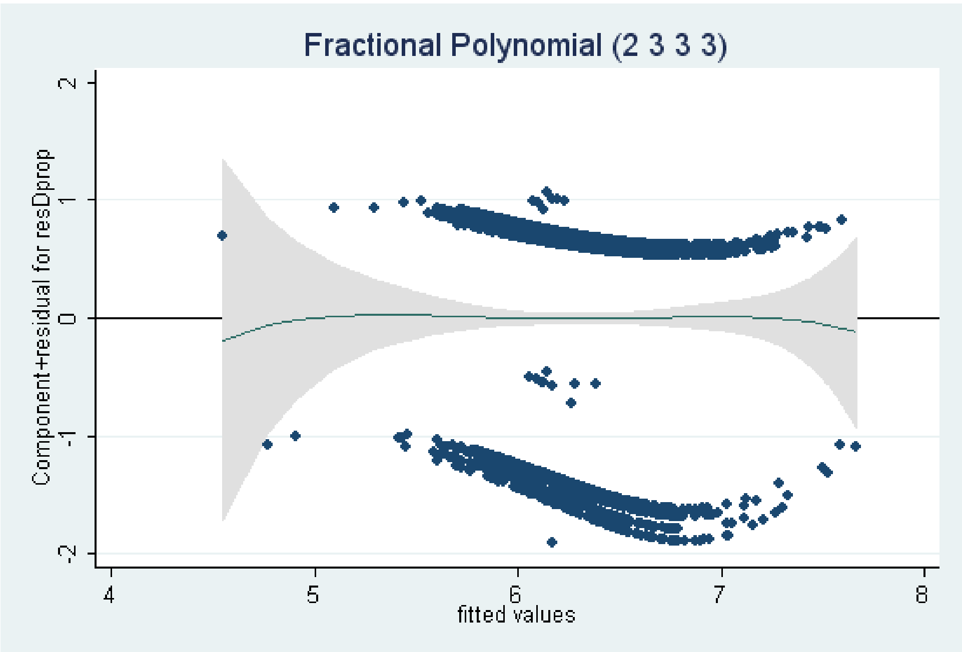

Figure B.4: Residuals from equation (5.3) vs *prop*, fourth post-program semester, young women

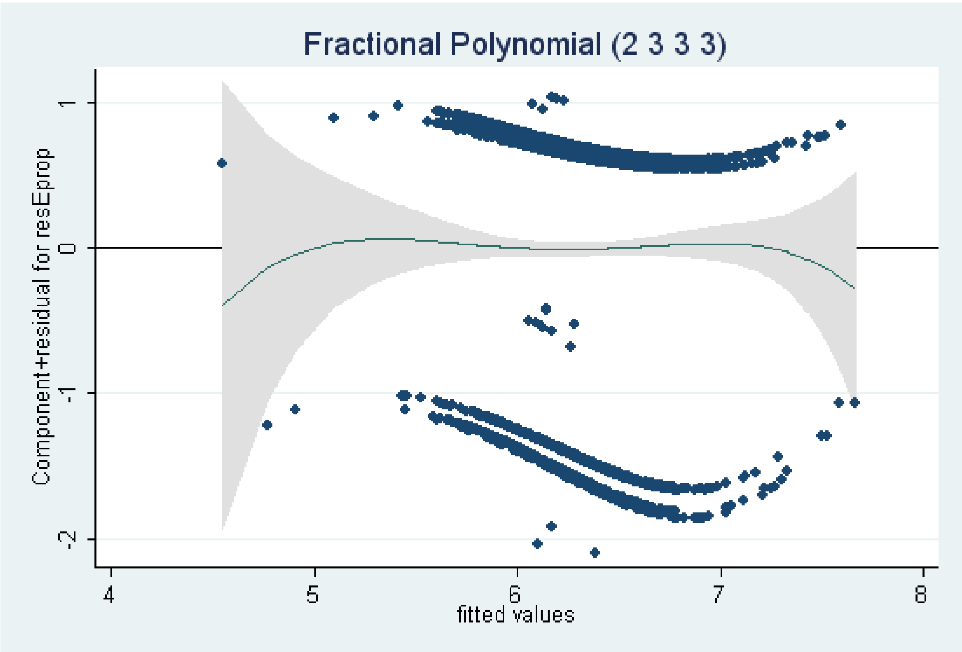

Figure B.5: Residuals from equation (5.3) vs *prop*, fifth post-program semester, young women

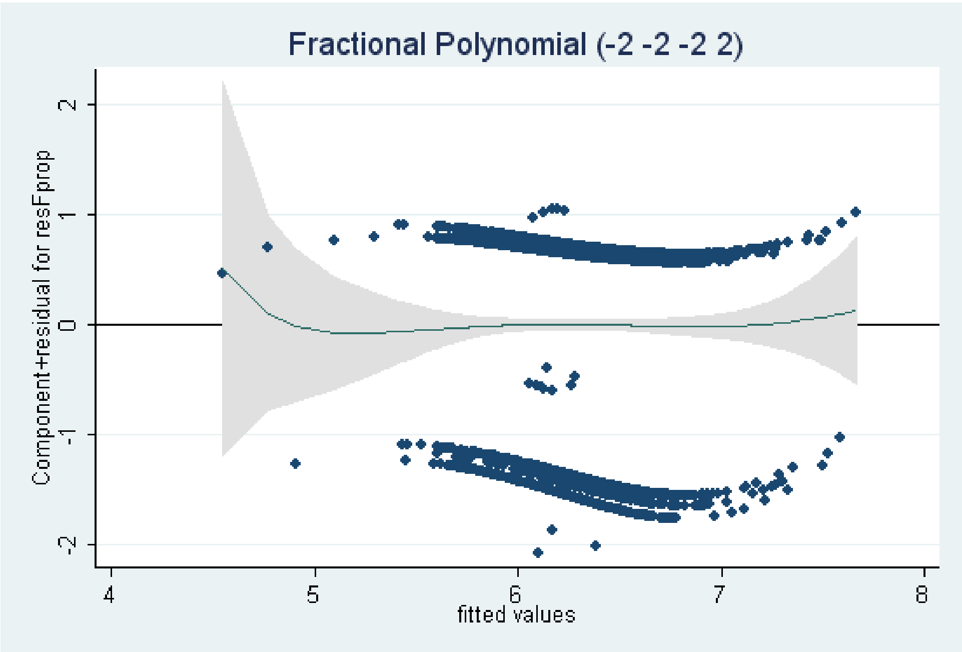

Figure B.6: Residuals from equation (5.3) vs *prop*, sixth post-program semester, young women

Appendix C

Results of the test on the validity of the "diff-in-diffs" approach

Table C.1 show the joint significance of the interacted terms (D with a polynomial on *logpay*) for 24, 36, 48, 60, 72 months before enrolment in the LM. Numbers on the left indicates the polynomial degree used. When it is 1, the model is:(5.4), and the significance level is relative to the hypothesis: $\beta_3 = 0$. In general, this is a test for the validity of the "diff-in-diffs" approach.

		Months before enrollment in the LM				
		24	36	48	60	72
Polynomial degree	1	0.1432	0.1379	0.5896	0.5987	0.0135
	2	0.3285	0.0589	0.6672	0.7539	0.0416
	3	0.4402	0.1266	0.7967	0.7700	0.0988
	4	0.3093	0.1068	0.8176	0.8294	0.1615

Table C.1: Observed significance level of the interacted polynomials on probability of employment in the t^{th} month before enrolment in the LM

Appendix D

Results of the test on the significativity of unemployment benefit on re-employment probability

A second degree polynomial of the logarithm of the monthly subsidy received $\log UI$ on re-employment probability in equation (5.5) turns out to be significant only for the first two semesters after enrolment in the LM.

Semester after enrollment in the LM					
1	2	3	4	5	6
0.0060	0.0097	0.2840	0.1822	0.2807	0.1944

Table D.1: Observed significance level of the quadratic polynomial of $\log UI$ on re-employment probability in the "difference in differences" model.

The bibliography

- Abadie A. (2005), "Semiparametric Difference-in-Differences Estimators",*Review of Economic Studies*, **72(1)**, 1-19.

- Abowd J. M., Kramarz F., Lengermann P., Roux S.(2003), "Interindustry and Firm-Size Wage Differentials in the United States and France", *Cornell University* working paper.

- Abowd J. M., Kramarz F., Margolis D.N.(1999), "High Wage Workers and High Wage Firms", *Econometrica*, **67**, 251-333.

- Anastasia B. et al. (2004), "Interazione fra sussidi passivi e incentivi al reimpiego: provenienze ed esiti di lavoratori iscritti nelle Liste di Mobilità", Rapporto finale, Venezia: Agenzia Veneto Lavoro (mimeo).

- Angrist J. D., Imbens G. W. (1991), "Sources of Identifying Information in Evaluation Models", Technical Working Paper 117, National Bureau of Economic Research.

- Angrist J. D., Imbens G.W. (1994), "Identification and Estimation of Local Average Treatment Effects",*Econometrica*, **62(2)**, 467-476.

- Angrist J. D., Imbens G.W. (1995), "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity", *Journal of the American Statistical Association*, **90:430**, 431-42.

- Angrist J. D., Imbens G.W. (1996),"Identification of Causal effects Using Instrumental Variables",*Journal of Econometrics*, **71(1-2)**, 145-160.

- Angrist J. D., Krueger A.B. (1995), "Split-Sample Instrumental Variables Estimates of the Returns to Schooling", *Journal of Business and Economic Statistics*, **132**, 225-35.

- Angrist J.D., Krueger A.B. (2001), "Instrumental variables and the search for identification: From supply and demand to natural experiments", *Journal of Economic Perspectives* **15**, 69-85.

- Angrist J.D., Kuersteiner G.M. (2004), "Semiparametric Causality Tests Using the Policy Propensity Score",*NBER Working Paper No. 10975.*

- Ashenfelter O. (1978), "Estimating the effect of training programs on earnings",*Review of Economics and Statistics*, **60**, 47-57. Ashenfelter O., Card, D. (1985): "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs",*Review of Economics and Statistics*, **67**, 648-660.

- Balke A., Pearl J. (1994), "Counterfactual probabilities: Computational methods, bounds, and applications", in R. Lopez de Mantaras and D. Poole (Eds.), *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI-94)*, Morgan Kaufmann, San Mateo, CA, 46-54, July 29-31, 1994.

- Bassmann (1957), "A generalized classical method of linear estimation of coefficients in a structural equation", *Econometrica* **25**, 77-83.

- Bertrand M., Duflo E., Mullainathan S. (2004), "How Much Should We Trust Differences-in-Differences Estimates?", *The Quarterly Journal of Economics*, MIT Press, **119(1)**, 249-275.

- Blundell R., Costa Dias M. (2000), "Evaluation methods for non-experimental data", *Fiscal Studies***21(4)**, 427-468.

- Blundell R., Costa Dias M. (2002), "Alternative approaches to evaluation in empirical microeconomics", *Portuguese Economic Journal*, **1**, 91-115.

- Burdett K.(1979), "Unemployment Insurance Payments as a Search Subsidy: A Theoretical Analysis", *Economic Inquiry*, **17:3**, 333-343.

- Burdett K., Mortensen D.T.(1998), "Wage Differentials, Employer Size, and Unemployment', *International Economic Review*, **39 (2)**, 257-273.
- Cahuc P., Postel-Vinay F., Robin J.M., 2006, "Wage Bargaining with On-the-job search: Theory and Evidence",*Econometrica*,**74(2)**, 323-64.
- Campbell D. T. (1969), "Reforms as Experiments",*American Psychologist*, **24**, 409-429.
- Caruso E., Pisauro G. (2005), "Licenziamenti definitivi o temporanei? Durata della disoccupazione nelle Liste di mobilità tra nuovi e vecchi datori di lavoro",*Politica Economica*,**21(3)**, 361-399.
- Chamberlain G.(1982), "The General Equivalence of Granger and Sims Causality",*Econometrica*, **50(3)**, 569-582.
- Chamberlain, G. (1984), "Panel Data", in *Handbook of Econometrics*, Z. Griliches and M. Intriligator, Volume 2, Amsterdam: Elsevier Science.
- Chapin F. S. (1947), "Experimental Designs in Sociological Research", Harper, New York.
- Cochran W. G.(1965), "The planning of observational studies in human populations",*Journal of The Royal Statistical Society*, Series A **128**, 234-266.
- Cochran W. G. (1968),"The effectiveness of subclassification in removing bias in observational studies", *Biometrics* **24**, 295-313.
- Cochran W.G., Rubin D.B. (1973), "Controlling Bias in Observational Studies: A Review", *Sankhya*, Series A **35**, 417-446.
- Cox D.R.(1958), "Planning of Experiments", New York: John Wiley.
- Dee T. S., Fu H. (2003), "Do Charter Schools Skim Students or Drain Resources?", *Economics of Education Review* **23**, 259-271.
- Del Conte M., Devillanova C., Morelli S. (2004), "L'indice OECD di rigidità nel mercato del lavoro: una nota", *Politica Economica*,**20**, 335-356.

- DiNardo J., Tobias J. L. (2001), "Nonparametric Density and Regression Estimation",*Journal of Economic Perspectives*,**15**, 11-28.

- Durbin, J. (1954), "Errors in Variables", *Review of the International Statistical Institute*, **22**, 23-32.

- Engle R.F., Hendry D.F. and Richard J.F. (1983), "Exogeneity", *Econometrica*, **51(2)**, 277-304.

- Florens J.P., Mouchart, M., 1985: "A Linear Theory for Noncausality", *Econometrica*, **53(1)**, 157-75.

- Geary R.C. (1949), "Determination of Linear Relations Between Systematic Parts of Variables with Errors of Observations, the Variances of Which are Unknown", *Econometrica*,**17(1)**, 30-58.

- Goldberger A. S. (1972a), "Structural Equation Methods in the Social Sciences", *Econometrica*, **40**, 979-1001.

- Goldberger, A.S. (1972b), "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations", Madison, Wisconsin: University of Wisconsin Press.

- Goldberger, A.S. (1972c), "Selection Bias in Evaluating Treatment Effects: The case of interaction", Madison, Wisconsin: Institute for Research on Poverty.

- Granger, C. W. J. (1969), "Investigating causal relations by econometric models and cross-spectral methods", *Econometrica* **37**, 424-438.

- Hahn J., Todd P., Van der Klaauw W. (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica*, **69**, 201-209.

- Heckmann J.J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, *Econometric Society*, **46(4)**, 931-59.

- Heckmann J. J., Honoré B. E.(1990), “The Empirical Content of the Roy Model”, *Econometrica* **58(5)**, 1121-1149.

- Heckman J.J., Robb R. (1985), “Alternative methods for evaluating the impact of interventions”, in *Longitudinal analysis of labour market data*, Wiley, New York.

- Heckman J.J., Robb R. (1986), “Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes”, in: Wainer H, (ed) *Drawing inferences from self-selected samples*, Springer, Berlin Heidelberg New York.

- Hirano K., Imbens G.(2004), “The Propensity Score with Continuous Treatments”, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, A. Gelman and X.-L. Meng, New York: Wiley.

- Holland P.W. (1986),“Statistics and Causal Inference”, *Journal of the American Statistical Association* **81**, 945-960.

- Holland P.W.(1987),“The Role of a Second Control Group in an Observational Study: Comment”,*Statistical Science*,**2(3)**, 306-308.

- Hosoya, Y. (1977), “On the Granger condition for non-causality”, *Econometrica 45* **7**, 1735-1736.

- Ichimura H., Taber C.(2000), “Direct Estimation of Policy Impacts”, *NBER Technical working paper No. 254.*

- Ichimura H., Taber C.(2002), “Semiparametric Reduced Form Estimation of Tuition Subsidies”, *American Economic Review*, **92(2)**, 286-292.

- Ichino P. (2004), “Job security and the value of equality. A tentative law & economics approach to the problem of effectiveness of the workers protection against dismissal for economic reasons”, paper presented at the University of Vienna, January 31, 2004 (mimeo).

- Imai K., Van Dyk D.A.(2004), "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score", *Journal of the American Statistical Association*, **99(467)**, 854-866.

- Imbens G. W. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*,**87**, 706-710.

- Imbens G.W., Lemieux T. (2008), "Regression discontinuity designs: A guide to practice", *Journal of Econometrics*, **142(2)**, 615-635.

- Joffe M. M., Rosenbaum P. R. (1999), "Propensity Scores", *American Journal of Epidemiology*,**150**, 327-333.

- Li Q., Huang C. J., Li D., Fu T.T. (2002), "Semiparametric Smooth Coefficient Models", *Journal of Business Economic Statistics*, **20**, 412-422.

- Lu B., Zanutto E., Hornik R., Rosenbaum P. R. (2001), "Matching With Doses in an Observational Study of aMedia Campaign Against Drug Abuse",*Journal of the American Statistical Association*,**96**, 1245-1253.

- Marschak J.(1953), "Econometric Measurements for Policy and Prediction", *Studies in Econometric Method*, edito da W.Hood and T.Koopmans (New York, John Wiley, 1-26).

- Martin J.P., Grubb D.(2001), "What works and for whom: A review of OECD countries experience with active labour market policies", Swedish Economic Policy Review, 8: 9-56.

- Meyer B.(1995), "Natural and Quasi-Natural Experiments in Economics", *Journal of Business and Economic Statistics*, **13**, 151-162.

- Mincer J. (1974), "Schooling, Experience and Earnings", Columbia University Press: New York.

- Mortensen D.T. (2000), "Equilibrium unemployment with wage posting: Burdett-Mortensen meet Pissarides", *Panel data and structural labor market models*, Bunzel H., B.J. Christensen, P. Jensen, N.M. Kiefer and D.T. Mortensen (eds.), Elsevier, North-Holland.

- Mortensen D.T.(2003), "Wage Dispersion- Why Are Similar Workers Paid Differently?", MIT Press.

- Mortensen D.T., Pissarides C.A. (1999), "New developments in models of search in the labor market", Handbook of Labor Economics, Ashenfelter and Card ed., edition 1, volume 3, chapter 39, pages 2567-2627.

- Morgan, M.S. (1990), "The History of Econometric Ideas", Cambridge: Cambridge University Press.

- Mortensen D.T.(1977), "Unemployment Insurance and Job Search Decisions", *Industrial and Labor Relations Review*,**30(4)**, 505-517.

- Neyman, J.(1923), "On the Application of Probability Theory to Agricultural Experiments", *Essay on Principles*, Section 9, translated in *Statistical Science*, (with discussion), **5(4)**, 465-480, 1990.

- OECD (2004), OECD Employment Outlook 2004, Paris: OECD Publishing.

- OECD (2005), OECD Employment Outlook 2005, Paris: OECD Publishing

- Paggiaro A., Rettore E., Trivellato U. (2007),"The effect of extended duration of eligibility in an Italian labour market for dismissed workers", paper presented at the Conference "Labor Market Flows, Productivity and Wage Dynamics: Ideas and Results from Empirical Research on Employer-Employee Linked Longitudinal Databases", LABORatorio R. Revelli, Fondazione Carlo Alberto, Moncalieri (Torino), September 2007.

- Postel-Vinay, F. Robin J.M.(2002), "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity",*Econometrica*,**70(6)**, 2295-350.

- Postel-Vinay F., Robin J.M. (2006), "Microeconometric Search-Matching Models and Matched Employer-Employee Data" in R. Blundell, W. Newey and T.

Persson, editors, *Advances in Economics and Econometrics, Theory and Applications*, Ninth World Congress, Volume 2, Cambridge: Cambridge University Press.

- Postel-Vinay F., Turon H. (2005), "On-the-job Search, Productivity Shocks, and the Individual Earnings Process", *CEPR* Discussion Paper No. 5593.

- Reiersol O. (1945), "Confluence Analysis by Means of Instrumental Sets of Variables", *Arkiv for Matematik, Astronomi och Fysik*, **32a:4**, 1-119.

- Rettore E., Battistin E. (2008), "Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs", *Journal of Econometrics*, **142(2)**, 715-730.

- Rogerson R., Shimer R., Wright R., "Search-Theoretic Models of the Labor Market: A Survey",*Journal of Economic Literature*, **63(4)**, 959-988.

- Rosenbaum P.R.(1987),"The Role of a Second Control Group in an Observational Study",*Statistical Science*,**2(3)**, 292-306.

- Rosenbaum P.R., Rubin D.B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects" *Biometrika*, **70(1)**, 41-55.

- Royston P., Altman D.G.(1994):"Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with disc.)", *Applied Statistics*,**43**, 429-467.

- Rubin, D.B. (1973a),"Matching to remove bias in observational studies",*Biometrics*,**29**,159-183.

- Rubin, D.B. (1973b),"The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies", *Biometrics*,**29(1)**, 185-203.

- Rubin, D.B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*,**66(5)**, 688-701.

- Rubin, D.B. (1976a),“Multivariate Matching Methods That are Equal Percent Bias Reducing, I: Some Examples”, *Biometrics*,**32(1)**, 109-120.

- Rubin, D.B. (1976b),“Multivariate Matching Methods That are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes”, *Biometrics*,**32(1)**, 121-132.

- Rubin, D.B. (1977), “Assignment to a Treatment Group on the Basis of a Covariate”,*Journal of Educational Statistics*,**2**, 1-26.

- Rubin, D.B. (1978), “Bayesian Inference for Causal Effects: The Role of Randomization”,*Annals of Statistics*,**6(1)**, 34-58.

- Rubin, D.B.(1979), “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias In Observational Studies”,*Journal of the American Statisrical Association*, **74(3)**, 18-328.

- Rubin, D.B. (1980a), “Randomization Analysis of Experimental Data: The Fisher Randomization Test. Comment”,*Journal of the American Statistical Association*,**75**,**371(3)**, 591-593.

- Rubin, D.B. (1980b), “Bias Reduction Using Mahalanobis-Metric Matching”,*Biometrics*,**36(2)**, 293-298.

- Rubin, D.B. (1986), “Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers”, *Journal of the American Statistical Association*, **81(396)**, 961-962.

- Rubin, D.B. (1990), “Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies”,*Statistical Science*,**5(4)**, 472-480.

- Sargan J. (1958), “The estimation of economic relationships using instrumental variables”,*Econometrica* **26**, 393-415.

- Sargan, J. (1959), “The estimation of relationships with autocorrelated residuals by the use of the instrumental variables”, *Journal of the Royal Statistical Society*, Series B **21**, 91-105.

- Schivardi F., Torrini R. (2004), "Firm size distribution and employment protection legislation in Italy", *Temi di discussione* No. 504, Roma: Banca d'Italia.
- Shimer R. (2003), "The Cyclical Behavior of Equilibrium Unemployment and Vacancies: Evidence and Theory", *NBER* Working Papers 9536, National Bureau of Economic Research, Inc.
- Sims C.A. (1972),"Money, Income, and Causality", *American Economic Review*, **62(4)**, 540-552.
- Theil H. (1953), "Repeated Least Squares Applied to Complete Equation Systems", The Hague: Central Planning Bureau.
- Theil H.(1958), "Economic Forecasting and Policy", North Holland, Amsterdam (1958).
- Thistlewaite D.L., Campbell D.T. (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment", *Journal of Educational Psychology*, **51**, 309-317.
- Todd P.E., Wolpin K.I.(2005), "Ex Ante Evaluation of Social Programs", *PIER* Working Paper 06-022.
- Trochim W. M. K. (1984), "Research Design for Program Evaluation", Beverly Hills, CA: Sage Publications.
- Trochim W. M. K., Spiegelman C. (1980), "The Relative Assignment Variable Approach to Selection Bias in Pretest-Posttest Designs",*American Statistical Association.*
- Wright, P.G. (1925), "Corn and Hog Correlations", US Department of Agriculture Bulletin 1300, January 1925, Washington, DC.

- Wright, P.G. (1928), "The Tariff on Animal and Vegetable Oils", New York: MacMillan.
- Yatchew A. (1998), "Nonparametric Regression Techniques in Economics", *Journal of Economic Literature*, **36**, 669-721.