Temporary Work Agencies in Italy: A Springboard Toward Permanent Employment?*

Andrea Ichino
EUROPEAN UNIVERSITY INSTITUTE

Fabrizia Mealli
UNIVERSITY OF FLORENCE

Tommaso Nannicini
EUROPEAN UNIVERSITY INSTITUTE

April 7, 2004

Abstract

This paper measures to what extent Temporary Work Agency (TWA) employment represents a "springboard" toward a permanent job or a "trap" of endless precariousness. Applying Propensity Score matching in the presence of choice-based sampling we estimate the causal effect of the treatment "TWA mission" on the outcome "finding a permanent job after 18 months". The data come from Italy, where TWAs were liberalized in 1997, and were specifically collected for this evaluation study. We show that a TWA mission increases the probability of finding a permanent job by 19 percentage points in Tuscany and by 11 percentage points in Sicily, although this second effect is only barely significant. These effects are large given that the observed baseline probabilities in our treated group are respectively 31% and 23% in the two regions. This treatment effect is highly heterogeneous with respect to characteristics such as age, education and firm's sector. We also perform a sensitivity analysis, in order to assess the plausibility of the identifying assumption of "selection on observables". This analysis confirms the robustness of our results.

JEL Classification: C2, C8, J6.

*Andrea Ichino is also affiliated with CEPR, CESifo and IZA. Financial support by the Italian Ministry of Welfare and the Tuscany Region is gratefully acknowledged. We also thank "Manpower Italia Spa" for the help in data collection. We would like to thank seminar participants at Perugia, EUI, and the EC workshop on "Temporary Work in Europe" for insightful comments.

Corresponding author: Andrea Ichino, EUI, Department of Economics, via della Piazzuola 43, 50133 Firenze, Italia, email: andrea.ichino@iue.it.

1 Introduction

Policy makers and labor market analysts are increasingly concerned about the growth of temporary employment in Europe. According to Oecd (1999), during the '90s there has been a considerable continuity in the employment protection legislation of most countries, with one major exception: the deregulation of fixed-term contracts and of temporary work agencies. Particularly in southern European countries, changes of labor market policy consisted mainly of measures aimed at introducing "flexibility at the margin", i.e. making the utilization of non-permanent contracts less strictly regulated leaving the discipline of standard employment unchanged. Precisely in these countries where standard employment is subject to a very rigid legislation, the increasing flexibility at the margin had a stronger effect on the diffusion of temporary contracts.[1]

The growing share of temporary employment in many European countries raised concerns over the risk of labor market "segmentation". Several studies have indicated the existence of a gap in the working conditions of permanent and temporary employees, particularly in terms of wages and working rights.[2] Worried by this gap, the public opinion and policy makers have stressed the importance of searching "an appropriate balance between flexibility and security" (European Commission, 2003). It is the so called "flexicurity" approach, which aims at squaring the circle of ensuring at the same time flexibility, job security and job quality.

While the evidence seems to suggest that "squaring this circle" is not an easy task in a cross-sectional sense, it could be that for most individuals

[1]Similarly in the US, the recent growth of TWA employment appears to be related to the increasing strictness of unjust dismissal doctrine in many States of the Union (Autor, 2000).

[2]See the literature survey in Oecd (2002).

"the circle is squared" in an intertemporal sense. This because a temporary job may represent a costly investment that a young worker undertakes to increase the probability of finding a permanent job. Several theoretical arguments can be constructed to justify this intuition, mostly based around the idea that in the presence of asymmetric information, a temporary contract is a costly signal that allows the worker to inform the market about her ability (Nannicini, 2004b). But it is ultimately an empirical question to measure the extent to which temporary jobs are an effective springboard toward permanent employment or a "trap" of endless precariousness.

This is the question we try to answer in this paper with specific reference to Temporary Work Agency (TWA) employment. By "TWA employment" we mean a triangular contract, in which an agency hires a worker for the purpose of placing her at the disposal of a client firm for a temporary assignment. Our analysis refers to Italy, where this kind of non-standard employment was liberalized in 1997.

Specifically, our goal is to evaluate whether the treatment consisting in a "TWA mission" has a positive and significant causal effect on the outcome "finding a permanent job after 18 months". We use a unique data set, which we have collected precisely to perform this evaluation exercise. The data consist of the universe of TWA workers who went into a mission during the first 6 months of 2001, which we compare to a sample of workers who, at the beginning of this period, were unemployed or employed with a non-permanent contract.

We are interested in the average effect of the treatment on the treated, i.e. in the difference between the outcome for the workers in our treated group with respect to the counterfactual unobservable outcome which would have prevailed for them in the absence of the TWA mission. To do so we use the estimation method of Propensity Score matching in the presence of choice-based sampling. Since this technique relies on the crucial assumption of "selection on observables", we also perform a particular sensitivity analysis to assess the robustness of our estimates with respect to this assumption.

The structure of the paper is as follows. Section 2 describes the take-off of TWA employment in the Italian labor market. Section 3 discusses briefly the possible determinants of the transition from temporary to permanent employment. Section 4 presents our method of evaluation, i.e. the Propensity Score matching estimation in the presence of choice-based sampling. In Section 5.1, the data collection strategy is described. In Section 5.2, sample descriptive statistics are reported and discussed. Section 6 presents the estimation results. In Section 7, the sensitivity analysis is proposed and implemented. Section 8 draws our conclusions.

2 Temporary work agencies in Italy

Italy is a good example of the trend toward flexibility at the margin which has characterized several European countries. Undoubtedly, the major step toward the liberalization of non-standard contracts has been the so-called "Treu law" (law 196/1997), which legalized and regulated the supply of temporary workers by authorized agencies (against the law until then).[3] Afterwards, TWAs have roared and a "hot" policy debate over the effects of this liberalization for firms and workers has begun.

The Treu law (and subsequent modifications) states that TWA employment is allowed except for the following cases: replacement of workers on strike; firms that experienced collective dismissals in the previous 12 months

[3] On the introduction of this kind of non-standard employment in the Italian labor law, see Ichino (2000).

and jobs that require medical vigilance. The law does not set a maximum cumulated duration of missions or legal motivations for using temporary work, leaving the provision of further regulation to collective bargaining. Collective agreements have typically stipulated that temporary workers cannot exceed 8-15% of normal employees (depending on the sector). Moreover they have constrained the allowed motivations for TWAs, which are: peak activity; one-off work and need of skills not available within the firm. Firms cannot extend an individual TWA contract for more than four times and for a cumulated period longer than 24 months.

On the whole, firing costs for standard employment remain high in the Italian labor market, and TWA employment faces less regulatory restrictions than other short-term contracts. In this context, firms might decide to hire temps in situations that generate different kinds of employment relationships in other countries. It should also be noted that, from the firm's point of view, TWA is less associated to a damaging "hire and fire" reputation than other non-standard contracts.

Following the Treu law, implemented in 1998, TWA employment has rapidly expanded, especially in the North of the country and in manufacturing sectors.[4] Nevertheless, in 2002 TWA employment still amounted to just 0.91% of total employment, a figure far below the level observed in countries where TWA developed earlier. Already in year 1999 instead, the overall incidence was 4.5% in the Netherlands, 3.2% in the United Kingdom, 2.5% in France, and 2.5% in the US. The average TWA employment utilization in the European Union was 1.5% in 1999 (Ciett, 2000).

It should be noticed, however, that TWA employment is still at a take-off stage in Italy. According to Ciett (2000), Italy will outmatch the 2% level

[4]For an aggregate overview, see Ministero del lavoro (2001) and Isfol (2001).

by 2010. Moreover, the instantaneous stock measure captures the per capita incidence of this type of work with respect to total employment but not its diffusion among workers: since turnover is high, TWA employment affects a much larger number of workers than those who are actually observed in a mission at any given point in time. Thus it may represent a springboard toward regular employment for a larger part of the labor force. Finally, the intensity of TWA employment utilization varies widely by industry and in 2000 it was already over 3% in some manufacturing sectors such as chemicals, machinery electronics and automotive (Nannicini, 2004a).

3 Springboard or trap?

From a theoretical point of view, there might be two broad reasons why temporary employment might represent a "springboard" to a stable job:

- "signaling", i.e. more-able workers might signal their type by making themselves available to be screened during temporary assignments;
- acquisition of human capital (general or specific), social contacts and information about permanent vacancies.

On the other end, temporary employment might be a "trap" of endless precariousness for the following reasons:

- a TWA experience is a "bad signal" of lack of alternatives (especially under the firm's belief that temps have already been screened by other employers);
- a TWA provides a limited acquisition of human capital because of the high turnover.

Leaving to Nannicini (2004) a detailed discussion of these different mechanisms, it suffices here to say that there is no obvious reason to expect one mechanism to prevail. In the Italian labor market, which is characterized by a high rigidity of standard employment[5], firms appear to be interested in TWA employment not only for screening but also to deal with demand fluctuations. This second motive is typically considered as the factor that transforms TWA employment into a trap. It is however not obvious that this should be the case. For example, a firm might hire a temp worker to face a non-permanent increase in market demand, and decide later to use the same worker (already screened during the short-term assignment) to fill a permanent vacancy. At the end of the day, whether TWA employment is a springboard or a trap is ultimately an empirical question.

Studies in other countries have shown a wide set of different results, depending on the institutional setting and the evaluation strategy. Descriptive evidence for the period 1996-1998 shows a large cross-country variation of the transformation rate of temporary contracts into permanent positions (see Oecd, 2002): from 21% (France) to 56% (Austria) in one year, or from 34% (Spain) to 71% (Austria) in two years. Also evaluation studies in different European countries have found mixed results. Booth, Francesconi and Frank (2002) studies the labor market prospects of temporary workers in UK (where temps represent the 7% of male employees and the 10% of female employees). Their results show that temporary employment is associated to lower wages, less specific training and lower job satisfaction with respect to permanent employment. But it is not associated to negative trajectories. In particular, women that go through a temporary job are able to completely catch up women starting in permanent positions, in terms of wages and job satisfac-

[5]See, for example, Grubb and Wells (1994), OECD (1999), and Nicoletti et al. (2001)

tion. Guell and Petrongolo (2003) studies the transformation of temporary into permanent contracts in Spain. Estimating a duration model, their study shows that temporary contracts might be used by Spanish firms both for flexibility and screening motivations. Malo and Munoz-Bullon (2002) perform an optimal matching analysis for Spain and find that TWA employment characterizes labor market trajectories with a higher probability to end with stable jobs. Other results of this "springboard" literature can be found in Lechner at al. (2000) for Germany and Zijl at al. (2002) for the Netherlands.

4 Methodology

4.1 Framework and notation

The aim of our analysis is to assess whether a TWA experience has a causal effect on the probability of finding a permanent job at a certain time in the future. Such a problem of causal inference involves "what if" statements and thus counterfactual outcomes. Hence, it can be "translated" into a treatment-control situation typical of the experimental framework. The fact that the treatment might be considered "endogenous" reflects the idea that the outcomes are jointly determined with the treatment status or, that there are unobservable variables related to both treatment status and outcomes. Thus "endogeneity" prevents the possibility of comparing "treated" and "untreated" individuals, as such comparison is very unlikely to have a causal interpretation because the two groups are different irrespective of their treatment status. A growing list of papers in the economic literature have tried to identify causal effects of interventions from observational, i.e., non experimental, studies, using the conceptual framework of randomized experiments and the so-called "potential outcomes approach", that allows to translate causal questions into a statistical model (Rubin, 1974).

This perspective was called "Rubin's Causal Model" by Holland (1986), because it views causal inference as a problem of missing data with explicit mathematical modeling of the assignment mechanism as a process for revealing the observed data. The essential feature of this approach is to define a causal effect as the comparison of the potential outcomes for the same unit measured at the same time: Y_0 = (the value of the outcome variable Y if the unit is exposed to treatment $T = 0$), and Y_1 = (the value of Y if exposed to treatment $T = 1$). Only one of these two potential outcomes can be observed, i.e., the one corresponding to the treatment the unit received, but the causal effect is defined by their comparison, i.e., $Y_1 - Y_0$. Thus, causal inference becomes a problem of inference with missing data. The focus of the analysis is usually that of estimating some features of the distribution of $Y_1 - Y_0$, e.g.,

$$E\{Y_1 - Y_0\} = E\{Y_1\} - E\{Y_0\}, \tag{1}$$

that is usually called the *Average Treatment Effect* (ATE), or the average treatment effect for subpopulations of individuals defined by the value of some variable, most notably the subpopulation of the treated individuals (*Average effect of Treatment on the Treated*, ATT):

$$E\{Y_1 - Y_0 | T = 1\}. \tag{2}$$

The assignment mechanism is a stochastic rule for assigning treatments to units and thereby for revealing Y_0 or Y_1 for each unit. This assignment mechanism can depend on other measurements, i.e., $P(T = 1|Y_0, Y_1, X)$; if these other measurements are observed values, then the assignment mechanism is ignorable; if given observed values it involves missing values, possibly even missing Y's, then it is non ignorable.

In order to make explicit the identifying assumptions underlying the estimators of the causal effects in our case, let us introduce further notation. Consider a set of I individuals, and denote each of them by subscript i: $i \in \{1, ..., I\}$. At time t_0, some of these individuals are "treated", i.e., they have an experience of TWA employment, whereas the others, usually named "controls", do not have such an experience at t_0. The treatment indicator is $T \in \{0, 1\}$. We are interested in the binary outcome variable indicating permanent employment at time $t_1 > t_0$. The two potential outcomes are thus: $Y_1 \in \{0, 1\}$ and $Y_0 \in \{0, 1\}$.

The decision to have a TWA experience can be represented, without loss of generality, as a process of utility maximization V:

$$V = f(Z, U_v) \quad T = I(V > 0) \tag{3}$$

where Z and U_v are observed and unobserved characteristics determining the choice, respectively. These sets of variables may contain both characteristics that are individual specific, which represent the individual life history up to time t_0, and characteristics of the area and the labor market where the individual resides.

Analogously, the two potential outcomes can be written as functions of observed (X) and unobserved (U) pre-treatment variables:

$$Y_1 = g_1(X, U) \tag{4}$$

$$Y_0 = g_0(X, U). \tag{5}$$

Also for these variables the previous comments hold, that is, they may include both characteristics that are individual specific and characteristics of the area and the labor market where the individual resides, that determine the occupational status in t_1. The two sets of variables X and Z may coincide or overlap to a certain extent.

Our objective is to identify and consistently estimate the ATT. Problems may arise because of the potential association between some of the unobservables U and the treatment indicator T. Our identification and estimation strategy is presented in the following sections.

4.2 Identification strategy

One of the assumptions that allow identification of the ATT is "unconfoundedness" (Rosenbaum e Rubin, 1983a), which is a special case of ignorable missing mechanism and the rationale behind common estimation strategies such as *matching* and regression modeling. This assumption does not distinguish between X and Z, but considers the whole *conditioning set* of pretreatment variables $W = (X, Z)$ and assumes that

$$(Y_1, Y_0) \perp T|W \tag{6}$$

and

$$0 < Pr(T = 1|W) < 1. \tag{7}$$

That is, conditioning on observed covariates W, treatment assignment is independent of potential outcomes. Unconfoundedness says that treatment assignment is independent of potential outcomes after accounting for a set of observable characteristics W. In other words, exposure to treatment is random within cells defined by the variables W. Although very strong, the plausibility of this assumptions heavily relies on the amount and on the quality of the information contained in W.

In our study, unconfoundedness might be violated both from the labor supply and labor demand sides. Some of the characteristics of the area where the individual resides (e.g., the presence of a high labor demand) may have attracted temporary work agencies, making it easier for a worker to have a temporary job experience. These same characteristics of the area may also ease the subsequent search of a permanent job. This is the reason why in Section 6, we used the distance of each single individual from her residence to the nearest TWA, in order to capture local labor market features not directly observed by the econometrician. Analogously, some individual unobserved characteristics may affect the propensity to have such an experience, and, at the same time, facilitate the access to a permanent job. These remarks notwithstanding, the quality of the data and the results of the sensitivity analysis that we will implement in Section 7 lead us to find defendable the unconfoundedness assumption in our case.

Under unconfoundedness, one can identify the average treatment effect within subpopulations defined by the values of W:

$$E\{Y(1) - Y(0)|W\} = E\{Y(1)|W\} - E\{Y(0)|W\} \tag{8}$$

$$= E\{Y(1)|T = 1, W\} - E\{Y(0)|T = 0, W\} \tag{9}$$

and also the overall ATT as :

$$E\{Y_1 - Y_0|T = 1\} = E\{E\{Y_1 - Y_0|T = 1, W\}\} = \tag{10}$$

$$= E\{E\{Y_1|T = 1, W\} - E\{Y_0|T = 0, W\}|T = 1\} \tag{11}$$

where the outer expectation is over the distribution of W in the subpopulation of treated individual. An implication of the above result is that if we could just divide the sample into subsamples depending on the exact value of the covariates W, then we could just take the average of the within subsample estimates of the average treatment effects. Often the covariates are more or less continuous, so that some smoothing techniques are in order: under unconfoundedness several estimation strategy can serve this purpose. One of those is regression modeling: using regression models to "adjust" or "control for" pre-intervention covariates is in principle a good strategy, although it has some pitfalls. For example, if there are many covariates it can be difficult to find an appropriate specification; in addition regression modeling obscures information on the distribution of covariates in the two treatment groups. In principle, in fact, one would like to compare individuals that have the same values of all covariates; unless there is a substantial overlap on the two covariates distributions, with a regression model one relies heavily on model specification, i.e. on extrapolation, for the estimation of treatment effects.

It is thus crucial to check how much the two distributions overlap, and which is the "region of common support" for the two distributions. When the number of covariates is large, this task is not an easy one. An approach that can be followed is to reduce the problem to one dimension by using the "Propensity Score", that is, the individual probability of receiving the treatment given the observed covariates: $p(W) = P(T = 1|W)$. In fact, under unconfoundedness, the following results hold (Rosenbaum and Rubin, 1983a): T is independent of W given the Propensity Score $p(W)$ and Y_0 and Y_1 are independent of T given the Propensity Score.

Note that the Propensity Score satisfies the so-called "balancing property", i.e., observations with the same value of the Propensity Score have the same distribution of observable (and possibly unobservable) characteristics independently of the treatment status; also exposure to treatment and control status is random for a given value of the Propensity Score. These two properties allow to use the Propensity Score as a univariate summary of all the W to check the overlap of the distributions of W. This because it is enough the check the distribution of the Propensity Score in the two groups and use the Score in the ATT estimation procedure as the single covariate that needs to be adjusted for. Adjusting for the Propensity Score automatically controls for all observed covariates, at least in big samples. As a result, given a population of units, if the Propensity Score $p(W_i)$ is known, the ATT can be estimated as follows:

$$\begin{aligned}\tau &\equiv E\{Y_{1i} - Y_{0i}|T_i = 1\} \qquad (12)\\ &= E\{E\{Y_{1i} - Y_{0i}|T_i = 1, p(W_i)\}\}\\ &= E\{E\{Y_{1i}|T_i = 1, p(W_i)\} - E\{Y_{0i}|T_i = 0, p(W_i)\}|T_i = 1\}\end{aligned}$$

where the outer expectation is over the distribution of $(p(W_i)|T_i = 1)$.

Any standard probability model can be used to estimate the Propensity Score. For example, $Pr\{T_i = 1|W_i\} = F(h(W_i))$, where $F(.)$ is the normal or the logistic cumulative distribution and $h(W_i)$ is a function of covariates with linear and higher order terms. Inasmuch as the specification of $h(W_i)$ which satisfies the balancing hypothesis is more parsimonious than the full set of interactions needed to match cases and controls on the basis of observables, the Propensity Score reduces the dimensionality problem of matching treated and control units on the basis of the multidimensional vector X.[6]

4.3 Matching estimators of the ATT based on the Propensity Score

The estimation of the Propensity Score is not enough to estimate the ATT of interest using equation (12). In fact, the probability of observing two units with exactly the same value of the Propensity Score is in principle zero, since $p(W)$ is a continuous variable. Various methods have been proposed

[6]It is important to note that the outcome plays no role in the algorithm for the estimation of the Propensity Score. This is equivalent, in this context, to what happens in controlled experiments in which the design of the experiment has to be specified independently of the outcome.

in the literature to overcome this problem.[7] In this paper we adopt two of them, *Nearest Neighbor Matching* and *Kernel Matching*, which we now present formally.

Let D be the set of treated units and C the set of control units, and Y_i^D and Y_j^C be the observed outcomes of the treated and control units, respectively. Denote by $C(i)$ the set of control units j matched to the treated unit i with an estimated value of the propensity score of p_i. Nearest Neighbor matching sets

$$C(i) = \{j \mid j = \arg\min_j \parallel p_i - p_j \parallel\} \tag{13}$$

which is a singleton set unless there are multiple nearest neighbors. In practice, the case of multiple nearest neighbors should be very rare, in particular if the set of characteristics W contains continuous variables.

Denote the number of controls matched with observation $i \in D$ by N_i^C and define the weights $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$ and $w_{ij} = 0$ otherwise. Then, the formula for the Nearest Neighbor Propensity Score matching estimator can be written as follows:

$$\begin{aligned}
\tau^M &= \frac{1}{N^D}\sum_{i \in T}\left[Y_i^D - \sum_{j \in C(i)} w_{ij} Y_j^C\right] \qquad (14)\\
&= \frac{1}{N^D}\left[\sum_{i \in D} Y_i^D - \sum_{i \in D}\sum_{j \in C(i)} w_{ij} Y_j^C\right]\\
&= \frac{1}{N^D}\sum_{i \in D} Y_i^D - \frac{1}{N^D}\sum_{j \in C} w_j Y_j^C
\end{aligned}$$

where the weights w_j are defined by $w_j = \Sigma_i w_{ij}$.

To derive the variances of these estimators the weights are assumed to be fixed and the outcomes are assumed to be independent across units.

$$Var(\tau^M) = \frac{1}{(N^D)^2}\left[\sum_{i \in D} Var(Y_i^D) + \sum_{j \in C}(w_j)^2 Var(Y_j^C)\right] \tag{15}$$

[7] See Becker and Ichino (2003) for further discussion.

$$= \frac{1}{(N^D)^2}\left[N^D Var(Y_i^D) + \sum_{j\in C}(w_j)^2 Var(Y_j^C)\right]$$
$$= \frac{1}{N^D}Var(Y_i^D) + \frac{1}{(N^D)^2}\sum_{j\in C}(w_j)^2 Var(Y_j^C).$$

Standard errors can also be obtained by bootstrapping.

The Kernel Propensity Score matching estimator is instead given by

$$\tau^K = \frac{1}{N^D}\sum_{i\in D}\left\{Y_i^D - \frac{\sum_{j\in C} Y_j^C G(\frac{p_j-p_i}{h_n})}{\sum_{k\in C} G(\frac{p_k-p_i}{h_n})}\right\} \tag{16}$$

where $G()$ is a kernel function and h_n is a bandwidth parameter. Under standard conditions on the bandwidth and kernel,

$$\frac{\sum_{j\in C} Y_j^C G(\frac{p_j-p_i}{h_n})}{\sum_{k\in C} G(\frac{p_k-p_i}{h_n})} \tag{17}$$

is a consistent estimator of the counterfactual outcome Y_{0i}. Standard errors can be obtained by bootstrapping.

4.4 Choice-based sampling

The estimators presented in the previous section can be straightforwardly applied if data are obtained with a simple random sampling design or a stratified sampling design, with known and observed stratification variables included in the vector W. Our data collection scheme (see Section 5.1) is a stratified sampling design, where one of the two stratifying variables is the province of residence, which is included in the pre-treatment variables' set W, while the other is the treatment indicator T. One of the stratifying variables, T, is thus an *endogenous* variable with respect to the specification of a model for the Propensity Score, i.e., $Pr(T = 1|W)$. This type of sampling scheme is usually called "choice based sampling" (Manski e Lerman, 1977) or, in general, "endogenous stratification".

In our case, we opted for this sampling scheme, in order to obtain information on an adequate number of treated individuals (i.e., temporary workers). With a simple random sampling, this would have required a sample size above our budget, because of the relatively small size of the treated group in the population. In addition, because we intended to use (even though not exclusively) an estimation strategy based on matching of treated and control units, and because variables describing the geographical and economical context are, a priori, particularly relevant, the stratification by province of residence allowed to select a number of controls that could guarantee an appropriate number of potential controls for each treated individual.

Under unconfoundedness, if a regression model is used to estimate the average causal effect, regression analysis is robust with respect to such an endogenous sampling scheme. Denoting with A the variable that identify the province of residence, in our sampling scheme a certain number of observations are sampled at random from each of the strata defined by $A \times T$. So every observation is characterized by the probability distribution $Pr(Y, W|A, T)$, with $Y = Y_1T + Y_0(1 - T)$. Assuming that the conditional distribution of Y can be modeled with a linear regression, the log-likelihood function for each observation is equal to $\log P(y|w, t, \beta) + \log P(w|a, t)$, and so factorizes into two terms, only one of which depends on the regression parameters β. Stratification in this case can only affect efficiency.

The application of estimation strategies based on the preliminary estimation of the Propensity Score seems more problematic. Sample data allow in fact to estimate the following distributions $Pr(W|A, T = 0)$ and $Pr(W|A, T = 1)$, whereas the Propensity Score is the conditional distribution $Pr(T = 1|W, A)$. Nevertheless, these distributions are linked one another, via Bayes theorem, in the following way:

$$Pr(W|A,T=j)Pr(T=j|A)Pr(A) = Pr(T=j|W,A)Pr(W|A)Pr(A), j=0,1 \tag{18}$$

so that

$$\frac{Pr(W|A,T=1)Pr(T=1|A)}{Pr(W|A,T=0)Pr(T=0|A)} = \frac{Pr(T=1|W,A)}{Pr(T=0|W,A)} \tag{19}$$

and

$$\frac{\tilde{P}(T=1|W,A)}{\tilde{P}(T=0|W,A)} = \frac{P(T=1|W,A)}{P(T=0|W,A)}\frac{\tilde{P}(T=1|A)}{\tilde{P}(T=0|A)}\frac{P(T=0|A)}{P(T=1|A)} \tag{20}$$

where $\tilde{P}(T=1|W,A)$ disregards choice-based sampling and $\tilde{P}(T=1|A)$ is conditioned on the province in the choice-based sample. Hence the odd of the misspecified (i.e., choice-based) Propensity Score can be used to implement matching within each province, because it is equal, up to a constant, to the odd of the *true* Propensity Score, which is itself a monotonic transformation of the Propensity Score (see Heckman and Todd, 1999).

5 Data

5.1 Data collection

The data collection implemented for this evaluation project had the following characteristics and stages. The analysis focused on a region of the Center (Tuscany) and one of the South (Sicily), which were among the areas with incomplete penetration of TWAs in 2000. We selected 5 provinces with agency (Livorno, Pisa, Lucca, Catania, Palermo) and 4 provinces without agency (Grosseto, Massa, Messina, Trapani). This strategy gave us two opportunities: 1) to identify observations very similar with respect to all individual characteristics but the access to TWA employment; 2) to consider the distance from an agency as a proxy of local labor demand, and use it as a matching variable in order to control for area-specific characteristics. In the empirical analysis, this second opportunity will be exploited, under the assumption that, within every province, TWAs locate themselves in the areas with higher labor demand.

We obtained access to the dataset of workers hired by "Manpower Italia Spa", a major company operating in the TWA sector with a national market share of about 25%. From this dataset, we extracted the workers who were in a TWA mission in one of the 9 provinces mentioned above during the first semester of 2001 and tried to interview them all. Hence, the first semester of 2001 was chosen as the "treatment" period, i.e. the period in which treated individuals went through their TWA experience. Data collection developed along the following two steps: i) phone interviews to all temps who were resident in the 9 provinces and were in a TWA mission during the first semester of 2001; ii) phone interviews to a random sample of "controls" drawn from the population of the 9 provinces, in order to match them with the treated units. Controls were chosen so as to have two characteristics: to be aged between 18 and 40 and not to have a permanent job (open-ended contract or self-employment) in January 2001. In a sense, also this first screening might be interpreted as part of the matching strategy, aimed at identifying a common support for the treated and the controls with respect to observable characteristics.

In order to get a sufficient number of controls in each area, we stratified the sample according to the province of residence. Hence, our data collection strategy leads to both choice-based sampling and geographical stratification. These two elements will be properly taken into account when deriving the empirical results, using the methodology described in Section 4.4.

Following this sampling strategy, we tried to get a lot of comparable information on treated and control units. For the treated, the reference point in time is the date of the TWA mission in 2001. For controls, it is January 2001. Information on the period before these reference points gave us "pre-treatment" variables, while information on the date of the interview (November 2002) gave us "outcome", i.e. post-treatment, variables. For both the treated and the control units, interviews followed an identical path, asking: a) demographic characteristics; b) family background; c) educational attainments; d) work experience before the treatment period; e) job characteristics during the treatment period; f) work experience from the treatment period to the end of 2002; g) job characteristics at the end of 2002.

After a first analysis of the data, we decided to drop from the sample control individuals who were out of the labor force in the treatment period (e.g. students). In fact, these subjects showed characteristics that made them not easily comparable with the treatment units. Notice that this is a conservative choice with respect to the estimated treatment effects, since all these individuals had a very low probability of having a permanent job at the end of 2002. Dropping these observations is another step of the search for a common support of treated and control units. The final data set used for the empirical evaluation is already a data set containing control units who could be more meaningfully "matched" with treated units.

At the end, the treated sample contains all residents in the 9 provinces who were on a TWA mission through "Manpower" during the first semester of 2001; while the control sample contains residents in the 9 provinces, aged 18-40, who belonged to the labor force but were not permanent employees as of January 1, 2001. This choice of the control sample is driven by the counterfactual question: what would have been the outcome of temporary workers at the end of 2002, if they had chosen to keep looking for a stable job or accept another kind of non-standard contract at the beginning of 2001?

The final dataset contains 2030 individuals: 511 treated (temporary workers); 1519 controls (other atypical workers or unemployed). The next section discusses some descriptive statistics of this data set.[8]

5.2 Descriptive statistics

Table 1 reports the distribution of the observations across the 9 provinces. The weighted proportion of each group (treated and controls) in the reference population (composed by unemployed and atypical workers aged between 18 and 40) is estimated by using "Manpower" and Istat data.[9] "Manpower" temps are 0.58% of this population in Tuscany and 0.15% in Sicily.[10] These small figures notwithstanding, it should be noted that in Tuscany 32% of the reference population declared to have contacted a TWA at least once, and 15% in Sicily.[11]

Table 2 summarizes the relevant information available for every individual in the sample. In this table as well as in the following ones, we present also the average characteristics of an important sub-sample of controls which we call the "matched controls". These are the control units used as "nearest neighbors" of at least one treated unit in the Nearest Neighbor Propensity Score matching.[12] Inasmuch as our treated units are more similar to the "matched controls" than to "all controls", our matching strategy has succeeded in improving the quality of the comparison used to estimate the causal effect of

[8]See Ichino, Mealli and Nannicini (2003) for further details on data collection.

[9]We know the exact number of "Manpower" temps in each province in the first semester of 2001. We estimate the population of unemployed and atypical workers aged between 18 and 40 in each province, by combining Istat statistics and the answer rate of our phone interviews. The ratio of the second and the first term is the province-specific weight.

[10]Notice that "Manpower" declared a market share of 32% in Tuscany and 45% in Sicily.

[11]See Ichino, Mealli and Nannicini (2003) for further data details.

[12]See Section 4.3 for a description of this estimator.

interest. We first give the descriptive statistics for the all the treated and control units and then comment on the sub-sample of "matched controls".

Treated individuals are prevalently young, male, single and without children. As far as education is concerned, there are not significant differences in years of schooling or educational attainment between treated and controls. Before the treatment period, a greater fraction of the treated was out of the labor force. In 2001, obviously, all treated are employed. Among controls, in Tuscany (Sicily) 36% (25%) had an atypical contract, while 64% (75%) were looking for a job. In 2002 -the "outcome" period- 31% of the treated have a permanent position in Tuscany, compared with 17% of controls. In Sicily, the same comparison gives 23% versus 13%.[13] Of course, these are simple correlations that need to be cleansed from observable and unobservable influences. This is exactly what our evaluation strategy aims to do.

Table 3 reports additional characteristics on the treated and controls who were employed in the pre-treatment period. Among the treated, there is a greater fraction of individuals previously employed with an atypical contract and as blue-collar workers in manufacturing. We can also notice that the pre-treatment wage of the treated was lower on average than the wage of controls, while hours of work were greater (due to a lower utilization of part-time arrangements).[14]

Table 4 reports additional characteristics on the treated (all) and the controls who were employed in the treatment period. The more relevant

[13]Incidentally, notice that employers mention to 51% of TWA workers the possibility of hiring them on a permanent basis at the end of the mission. Among these temps, 32% are effectively hired by the firm. But also among the others the percentage of direct-hiring is high: 20%. See Ichino, Mealli and Nannicini (2003) for further data details.

[14]Another interesting element concerns wage mobility (even though the small sample size prevents us to use this information as an alternative outcome): 36.9% of the treated with wage below the median in 2000 had a wage above the median in 2002, compared with 15.1% of controls. See Ichino, Mealli and Nannicini (2003) for further data details.

difference is in the firm's sector: TWA workers are mostly used in the manufacturing sector (60% in Tuscany and 53% in Sicily), while the other atypical workers are prevalently employed in the service sector (68% in Tuscany and 74% in Sicily). The motivations for the choice of atypical work are quite similar. For instance, in Tuscany 59% of temps could not find permanent job (against 59% of the other atypical workers); 22% did it to make up their mind on what they wanted to do (against 18%); 16% did it for personal flexibility needs (against 18%). Table 5 reports additional characteristics on the treated and controls who were employed at the end of 2002, i.e., in the outcome period. The "manufacturing gap" persists also in this period.

The previous descriptive tables provide information also on matched controls, i.e., control units used in the Nearest Neighbor Propensity Score matching estimation. It is particularly informative to check whether (and to what extent) the treated-control gap in observable pre-treatment characteristics is reduced when considering only matched controls (see again Tables 2 and 3). Figure 1 does so in a graphic way by reporting the relative reduction of such a gap for Tuscany. For each variable, the difference between the averages of the treated and the averages of all controls is set equal to 100 and displayed as such. The figure also displays the difference between the average of the treated and the average of matched controls as a fraction of the analogous difference between treated and all controls. Inasmuch as this relative difference is smaller than 100 our matching strategy has improved the quality of the comparison used for the estimation of the treatment effect. Figure 3 does the same for Sicily.

Figure 2 reports instead a similar relative reduction in the "pre-treatment gap" for those variables that are available only for individuals who were employed in the pre-treatment period, i.e., the period of unemployment as a fraction of the transition from school to work, and the job characteristics in 2000. Figure 4 does the same for Sicily.

It is evident that the Nearest Neighbor algorithm for the choice of the control units to be compared with treated units reduces considerably the "pre-treatment gap". This reduction is large both in Tuscany and in Sicily, even though it encounters some problems in the case of employment variables in Sicily. As we will see in the next section, this might be due to the specific characteristics of this regional labor market.

6 Estimated causal effects

Tables 6, 7, 8 and 9 contain the estimated ATTs for Tuscany and Sicily separately. Each regional ATT is obtained as the weighted average of the province-specific ATTs, in order to control for geographical stratification. The province-specific ATTs are obtained by using the regional estimates of the odd of the Propensity Score, in order to control for choice-based sampling (see Section 4.4). Standard errors are calculated as: $SE = (\sum \frac{N_i^2}{N^2} SE_i^2)^{1/2}$, where i indicates the province of residence. Matching variables include: gender, age, place of birth, nationality, marital status, number of children; years of schooling and prevalent job of the father, living father; educational level, grade in the last degree, post-school training; share of time without any occupation from school to the pre-treatment period; occupational status in the pre-treatment period, as well as type of contract, sector, profession, wage, working hours; province of residence and distance from the nearest temporary agency in the pre-treatment period.

Table 6 reports the results of Nearest Neighbor Propensity Score matching in Tuscany. TWA employment has a significant and positive effect of 19 percentage points on the probability to be in a stable positions 18 months after the treatment. As a reference, note that in Tuscany the observed probability of finding a permanent job for controls is 17%, while the observed probability for the treated is 31%. Hence, for the treated, the estimated "counterfactual" probability to get a permanent job in the case of non-treatment is 12% (even lower than the observed outcome of controls). Table 8 shows the results of Kernel Propensity Score matching, with a similar effect equal to 18 percentage points.

Tables 7 and 9 report the results of Nearest Neighbor and Kernel Propensity Score matching, respectively, for Sicily. Both estimators find a lower and not very significant effect of TWA employment: 11 and 10 percentage points, respectively. As a reference, note that in Sicily the observed probability of finding a permanent job for controls is 13%, while the observed probability for the treated is 23%. Hence, for the treated, the estimated "counterfactual" probability to get a permanent job in the case of non-treatment is 12-13% (exactly the same of the observed outcome of controls). The result that in Sicily TWAs do not seem an effective springboard to permanent employment might be linked to the fact that, in this region, the public sector is the primary source of stable positions, and in this sector the recruitment channels are different from TWA employment. In Tuscany, on the contrary, the private sector is creating a relevant number of stable positions that may be reached through the TWA channel.

In Table 10, some sources of heterogeneity in the treatment effect are investigated. In all these cases, the ATT is estimated by means of Weighted Nearest Neighbor Propensity Score Matching. In order to control for both geographical stratification and choice-based sampling, the ATT is estimated at the regional level by using appropriate weights.[15] In Table 10, analytical

[15]See footnote 9.

standard errors are reported. Bootstrapped standard errors have been calculated too, but the analytical ones lead to more conservative estimates. This heterogeneity analysis confirms non-significant results for Sicily. Only for a marginal minority of workers with university degree TWA employment has a strong and significant effect. In Tuscany, on the contrary, the heterogeneity analysis finds interesting results.

In the first row of Table 10, the ATT is estimated by dropping the unemployed from the control group. In this case, the ATT looses much of its significance also in Tuscany. TWAs are a springboard to permanent employment, but such a springboard does not seem more effective than the one offered by other forms of temporary employment. Hence, the aggregate effect of the liberalization of TWAs on permanent employment depends on the magnitude of the possible "crowding-out" of other non-permanent contracts.

Again in Table 10, the ATTs for individuals under 30 or over 30, and the ATTs for individuals with university degree and with or without high school degree are computed separately in these sub-samples. The ATTs in manufacturing or service sectors are computed by interacting the TWA experience with the sector of the using firm. These estimations show that the effect of the TWA treatment on the treated is greater for individuals over 30 years, for individuals with a university degree (even though they are a small minority) and in the service sector. The most surprising result is the one about age, which shows that young workers in the Italian labor market have to wait for a quite long period before finding a stable job.

7 Sensitivity analysis

Our analysis of the effects of temporary work on permanent employment is based on the critical assumption of unconfoundedness. As in all observational studies, our results might be criticized since this assumption rules out the role of unobservables. Parametric selection models, that in principle allow to relax the unconfoundedness assumption, are formally identified thanks to other types of non-necessarily preferable hypotheses, as pointed out in several papers (e.g., Little, 1985; Copas and Li, 1997). However, parametric models can be used as the basis for a sensitivity analysis of the robustness of non-parametric estimates. This is the spirit of the method proposed by Rosenbaum and Rubin (1983b), which allows to assess the sensitivity of the estimated causal effects with respect to assumptions about an unobserved binary covariate that is associated with both the treatment and the response.

The unobservables are assumed to be summarized by a binary variable in order to simplify the analysis, although similar techniques could be used assuming some other distribution for the unobservables. Note, however, that a Bernoulli distribution can be thought of as a discrete approximation of any distribution, and thus we believe our analysis results in no particular loss of generality.

The central assumption of this sensitivity analysis is that the assignment to treatment is not unconfounded given a set of observable variables X, i.e.,

$$Pr(T = 1|Y(0), Y(1), X) \neq Pr(T = 1|X) \tag{21}$$

but unconfoundedness holds given X and an unobserved binary covariate U, that is

$$Pr(T = 1|Y(0), Y(1), X, U) = Pr(T = 1|X, U). \tag{22}$$

Given this assumption, Rosenbaum and Rubin (1983b) suggest to derive the full-likelihood and maximize it, holding the sensitivity parameters as fixed known values. It is then possible to judge the sensitivity of conclusions to certain plausible variations in assumptions about the association of U with

T, $Y(0)$, $Y(1)$ and X. If conclusions are relatively insensitive over a range of plausible assumptions about U, then causal inference is more defensible. Since $Y(0)$, $Y(1)$ and T are conditionally independent given X and U, the joint distribution of $(Y(t), T, X, U)$ for $t = 0, 1$ is

$$Pr(Y(t), T, X, U) = Pr(Y(t)|X, U)Pr(T|X, U)Pr(U|X)Pr(X) \qquad (23)$$

We could further assume that

$$Pr(U = 0|X) = Pr(U = 0) = \pi \qquad (24)$$

and

$$Pr(T = 0|X, U) = (1 + \exp(\gamma' X + \alpha U))^{-1} \qquad (25)$$

and

$$Pr(Y(t) = 1|X, U) = \exp(\beta' X + \tau T + \delta_t U)(1 + exp(\beta' X + \tau T + \delta^t U))^{-1} \quad (26)$$

where π represents the proportion of individuals with $U = 0$ in the population, and the distribution of U is assumed to be independent of X. This should render the sensitivity analysis even more unfavorable to causal conclusions, since, if U were associated with X, controlling for X should capture at least some effects of the unobservables.

The sensitivity parameter α captures the effect of U on treatment receipt while the δ_t's are the effects of U on the potential outcomes. Given plausible but arbitrary values of the parameters π, α, and δ_t's, we estimated the parameters γ and β by maximum likelihood. It can be shown that for given values of the sensitivity parameters the conditional maximum likelihood estimates $\widehat{\gamma}(\pi, \alpha, \delta_0, \delta_1)$ and $\widehat{\beta}(\pi, \alpha, \delta_0, \delta_1)$ are uniquely defined. This enables to define the profile log-likelihood

$$\begin{aligned} L^*(\pi, \alpha, \delta_0, \delta_1) &= \max_{\gamma, \beta|\pi, \alpha, \delta_0, \delta_1} L(\gamma, \beta, \pi, \alpha, \delta_0, \delta_1) & (27) \\ &= L(\widehat{\gamma}(\pi, \alpha, \delta_0, \delta_1), \widehat{\beta}(\pi, \alpha, \delta_0, \delta_1), \pi, \alpha, \delta_0, \delta_1). & (28) \end{aligned}$$

In our case, in order to account for choice-base sampling, each individual contribution to the log likelihood should be multiplied by the sampling weights. Once the parameters have been estimated, the ATT estimates can be derived as follows:

$$A\hat{T}T = \frac{1}{N^T}\sum_{i \in T}[\hat{Y}_i^1 - \hat{Y}_i^0] \tag{29}$$

where

$$\hat{Y}_i^t = \pi \hat{P}r(Y(t) = 1|X, U = 0) + (1 - \pi)\hat{P}r(Y(t) = 1|X, U = 1) \tag{30}$$

To further adjust this methodology to our needs, we can condition on the treated and matched control samples, that have equal distribution of all the covariates. The idea is to treat the two samples as an imperfect randomized experiment, where instead of assigning the treatment with probability $P(T = 1)$, assignment is based on $P(T = 1|U)$, where U is an unobserved stratifying variable. This approach allows to bypass the problem of choice-based sampling so that the likelihood based method defined above can be applied without the use of weights and assuming that

$$Pr(T = 0|X, U) = (1 + \exp(\alpha U))^{-1}. \tag{31}$$

In addition to simplifying the problem of choice based sampling, this further adjustment makes the maximum likelihood estimates of the ATT even more comparable to the ones based on the Propensity Score based matching procedure.

Tables 11 and 12 show the results obtained with our version of the sensitivity analysis proposed by Rosenbaum and Rubin (1983b). Standard errors are computed by bootstrapping, based on 500 equal-sized replications of the original samples of treated and matched controls. Hence, only observations used in the Nearest Neighbor algorithm enter the maximum-likelihood estimations.

For each sample, L^* is maximized numerically. The value of the estimated ATT in the first row is derived under the assumption of unconfoundedness, i.e. with all the sensitivity parameters set to zero. It is the reference point of the sensitivity analysis, and it is only slightly lower than the ATT estimated by Propensity Score matching: 17 percentage points against 18-19 points in Tuscany; 8 points against 10-11 points in Sicily. In our case, one might interpret the binary unobservable variable U as individual ability, affecting both the (self-)selection into treatment and the potential outcomes. From this point of view, π is the probability of having low ability; α is the effect of ability on the selection into treatment; and δ_t (with $t = 0, 1$) are the effects of ability on the potential outcomes. In tables 11 and 12, "ATT" is calculated using the estimated average outcome for both the treated and controls. "ATT OBS", differently, uses the observed average outcome for the treated as $Y(1|T)$.

First of all, notice that in both cases the likelihood function is relatively flat, indicating that the data provide little information about the selectivity parameters. If anything, the profile likelihood attributes a greater plausibility to configurations with relatively low values of the parameters. The ATT estimates appear rather robust with respect to the removal of the unconfoundedness assumption. In Sicily, the effects remain positive and non significant, except for the case when the association between U and the potential outcomes and treatment assignment is (unbelievably) big, with a coefficient equal to 2: this would mean that, after conditioning on all the pre-treatment variables included in the model, individuals with $U = 1$ would have an odd of having a permanent job more than 7 ($=e^2$) times larger than individuals with $U = 0$. Such an association would be bigger than that of any other observed pre-treatment variable.[16] Similar robust results are obtained in Tuscany, with the estimated ATT remaining positive and significant in nearly all cases. Only for implausibly high values of the sensitivity parameters the effect disappears if calculated as "ATT OBS". On the whole, the sensitivity analysis strongly confirms the robustness of the Propensity Score matching estimates.

8 Conclusions

This paper investigates whether (and to what extent) TWA employment represents a "springboard" to a permanent job or it is a "trap" of endless precariousness. Applying Propensity Score matching in the presence of choice-based sampling, we estimated the causal effect of the treatment "TWA mission" on the outcome "finding a permanent job after 18 months". The analysis referred to Italy, where TWAs were liberalized in 1997 and we had the opportunity to get data appropriately collected for this evaluation exercise. Our estimates find a positive effect of a TWA mission on the probability to find a permanent job in Tuscany (19 percentage points) and a less significant effect (of about 11 percentage points) in Sicily. These effects are large given that the observed baseline probabilities in our treated group are respectively 31% and 23% in the two regions.

Relevant heterogeneity in the treatment effect along observable characteristics such as age, education and firm's sector is also detected. The estimated ATT is greater for individuals over 30 years, for individuals with an university degree (even though they are a small minority of temps) and in the service sector (rather than in manufacturing). We also performed a sensitiv-

[16]Notice that no variable has a coefficient higher than 1 in the previous estimations of the Propensity Score and the outcome equations.

ity analysis, in order to assess the plausibility of the identifying assumption of "selection on observables". This analysis confirms the robustness of our results.

From a policy perspective, our study finds that TWA employment has not been a "trap" of endless precariousness in Italy, but has been an effective springboard toward permanent employment. A similar springboard, however, is offered by other types of non-permanent labor contracts and it is not equally effective everywhere (e.g. it is in Tuscany, but not in Sicily) or for all workers (e.g. for workers in services, but not for workers in manufacturing sectors).

It should be noted however, that precisely because TWA employment allows workers to signal their (unobservable) ability to employers, it facilitates the emergence of a separating equilibrium in the labor market. Such a separating equilibrium benefits the workers who are better equipped to compete, while worsening the employment prospects of the weakest workers. The commendable attention that the Italian society (and unions in particular) devote to these weak workers, may appear to justify an opposition to TWA employment. However, banning the signaling possibilities offered by TWA employment would not help the weakest much, and would typically result in a less efficient outcome, not to mention the cost for the strongest workers. The correct way to help the weakest workers is to offer them the tools (e.g. training and better information) to compete effectively and send the right signals in the labor market.

Finally, from a methodological perspective, our study suggests that labor market programs in Europe, and in Italy in particular, should be increasingly evaluated with econometric methods aimed, as much as possible, at the identification of *causal effects*. Only in this way the political debate has a chance to become more productive, being based on relevant empirical findings instead of ideological prejudices.

References

Autor D.H. (2000), *Outsourcing at Will: Unjust Dismissal Doctrine and the Growth of Temporary Help Employment*, WP 7557, NBER.

Becker S. and Ichino A. (2002), "Estimation of average treatment effects based on Propensity Scores", *The Stata Journal*, Vol.2, 4, 358-377.

Booth A.L., Francesconi M. and Frank J. (2002), "Temporary Jobs: Stepping Stones or Dead Ends?", *Economic Journal*, 112, 480, 189-213.

Ciett (2000), *Orchestrating the Evolution of Private Employment Agencies towards a Stronger Society*, Brussels.

Copas J.B. and Li H.G. (1997), "Inference from Non-random Samples", in *Journal of the Royal Statistical Society*, B, 59, 1, 55-95.

Dehejia R.H. and Wahba S. (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*.

Guell M. and Petrongolo B. (2003), How Binding Are Legal Limits? Transition from Temporary to Permanent Work in Spain, IZA DP N.782.

Grubbs D. and W. Wells (1994), "Employment Regulations and Patterns of Work in EC Countries", in OECD *"Economic Studies"* 21, 7–58.

Heckman J. and Todd P., *Adapting Propensity Score Matching and Selection Models to Choice-based Samples*, mimeo, University of Chicago

Holland P. (1986), "Statistics and Causal Inference (with discussion)", *Journal of the American Statistical Association*, 81, 396, 945-970.

Ichino P. (2000), *Il contratto di lavoro - I*, Trattato di diritto civile e commerciale, volume XXVII, t.2, Giuffr Editore, Milano.

Isfol (2001), *Il lavoro interinale. Prima ricerca nazionale sui dati dei Centri per l'impiego*, Roma.

Lechner M., Pfeiffer F., Spengler H. and Almus M. (2000), *The impact of non-profit temping agencies on individual labour market success*, ZEW Discussion Paper 00-02.

Little R.J.A (1985), "A note about models for selectivity bias", in *Econometrica*, 53, 1469-1474.

Malo M.A. and Munoz-Bullon F. (2002), *Temporary Help Agencies and the Labour Market Biography: A Sequence-Oriented Approach*, EEE 132, FEDEA.

Manski C.F. and Lerman S.R.(1977), "The Estimation of Probabilities from Choice Based Samples", *Econometrica*, 45, 8.

Ministero del Lavoro e delle Politiche Sociali (2001), *Rapporto di monitoraggio sulle politiche occupazionali e del lavoro*, N.1/2001, Roma.

Nannicini T. (2004a), *The Take-Off of Temporary Help Employment in the Italian Labor Market*, EUI-ECO Working Paper N.09/04.

Nannicini T. (2004b), *Temporary Workers: How Temporary Are They?*, EUI, mimeo.

Nicoletti G., A. Bassanini, E. Ekkerhard, J. Sebastien, P. Santiago and P. Swaim (2001), *Product Market and Labour Market Regulation in OECD Countries*, OECD Economics Department WP, n. 312.

Oecd (1999), *Employment Protection and Labour Market Performance*, in *Employment Outlook*, Paris.

Oecd (2002), *Taking the measure of temporary employment*, in *Employment Outlook*, Paris.

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

Rosenbaum P. and Rubin D. (1983b), "Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome", *Journal of the Royal Statistical Society*, Series B, 45, 212-218

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

Zijl M., Heyma A. and van der Berg G. (2002), *Stepping stones for the unemployed? Effects of temporary jobs on job search duration of the unemployed*, mimeo, IZA.

Tables and Figures

Table 1: Province of residence before the treatment

	Agency	Distance	Treated	Controls	Tot.
Pisa	Yes	11.0	126	130	256
			(1.09)	(98.91)	(100)
Lucca	Yes	8.9	69	99	168
			(0.76)	(99.24)	(100)
Livorno	Yes	18.8	63	156	219
			(0.46)	(99.54)	(100)
Massa	No	39.9	10	130	140
			(0.15)	(99.85)	(100)
Grosseto	No	40.6	13	113	126
			(0.20)	(99.80)	(100)
TOSCANA	-	21.0	281	628	909
			(0.58)	(99.42)	(100)
Palermo	Yes	13.6	76	276	352
			(0.15)	(99.85)	(100)
Catania	Yes	15.4	112	195	307
			(0.22)	(99.78)	(100)
Messina	No	74.8	27	206	233
			(0.10)	(99.90)	(100)
Trapani	No	68.5	15	214	229
			(0.09)	(99.91)	(100)
SICILIA	-	38.0	230	891	1121
			(0.15)	(99.85)	100

The variable "distance" measures the average distance from the nearest agency (in km), computed by means of the Postal Code. In brackets, the weighted proportion of each group (controls and treated) on the reference population. The weighted proportion of the treated refers to "Manpower" temps only.

Table 2: Characteristics of the whole sample

	TUSCANY			SICILY		
	Treated	Matched Controls	All Controls	Treated	Matched Controls	All Controls
Age	26.5	27.5	29.1	26.8	27.8	30.0
Male	0.56	0.41	0.29	0.67	0.57	0.29
Single	0.90	0.87	0.66	0.83	0.81	0.49
Children	0.09	0.16	0.45	0.20	0.23	0.86
Father school	9.3	9.2	8.6	8.7	9.2	7.6
Father blue	0.33	0.39	0.43	0.30	0.31	0.39
Father active	0.53	0.46	0.37	0.46	0.45	0.29
School	12.5	12.7	12.3	12.0	12.4	11.6
Grade	75.9	77.1	76.9	74.7	74.6	76.5
Training	0.32	0.30	0.28	0.42	0.42	0.34
Unemployment	0.38	0. 42	0.48	0.42	0.44	0.62
Employed 2000	0.35	0.36	0.42	0.34	0.35	0.30
Unemployed 2000	0.52	0.53	0.52	0.60	0.60	0.67
Out l.force 2000	0.13	0.10	0.05	0.06	0.05	0.03
Employed 2001	1.00	0.36	0.36	1.00	0.30	0.25
Unemployed 2001	0.00	0.64	0.64	0.00	0.70	0.75
Permanent 2002	0.31	0.16	0.17	0.23	0.14	0.13
Atypical 2002	0.42	0.36	0.31	0.39	0.17	0.18
Unemployed 2002	0.16	0.44	0.45	0.30	0.59	0.63
Out l.force 2002	0.11	0.04	0.07	0.07	0.09	0.07
N.individ.	281	135	628	230	128	891

All variables except age, number of children, father's years of schooling, grade (expressed as a fraction of the highest mark), years of schooling and unemployment period (expressed as a fraction of the transition from school to work) are dummies. "Matched controls" are individuals who belong to the control sample and are used in the propensity-score matching estimation.

Table 3: Characteristics of the employed before the treatment

	TUSCANY			SICILY		
	Treated	Matched Controls	All Controls	Treated	Matched Controls	All Controls
Permanent	0.16	0.22	0.26	0.14	0.16	0.36
Atypical	0.84	0.78	0.74	0.86	0.84	0.64
Blue-collar	0.62	0.59	0.39	0.44	0.24	0.22
White-collar	0.36	0.41	0.54	0.54	0.71	0.67
Self-empl.	0.02	0.00	0.07	0.01	0.04	0.10
Manufact.	0.53	0.41	0.23	0.39	0.20	0.15
Service	0.39	0.45	0.67	0.49	0.67	0.70
Other	0.08	0.14	0.11	0.11	0.13	0.15
Wage	5.2	5.6	6.8	5.6	7.6	7.0
Hours	38.0	36.3	33.3	34.5	32.1	31.1
N.individuals	98	49	266	79	45	267

All variables except the hourly wage (expressed in Euros) and the weekly hours of work are dummies. "Matched controls" are individuals who belong to the control sample and are used in the propensity-score matching estimation.

Table 4: Characteristics of the employed in the treatment period

	TUSCANY			SICILY		
	Treated	Matched Controls	All Controls	Treated	Matched Controls	All Controls
Manufact.	0.60	0.35	0.22	0.53	0.13	0.15
Service	0.36	0.56	0.68	0.42	0.79	0.74
Other	0.04	0.08	0.10	0.05	0.08	0.12
Wage	7.1	7.5	7.8	8.8	10.7	8.8
Hours	40.5	31.0	31.5	39.0	28.4	30.5
No stable job	0.59	0.69	0.59	0.70	0.61	0.55
Preferences	0.22	0.17	0.18	0.13	0.18	0.22
Flexibility	0.16	0.13	0.18	0.15	0.13	0.13
N.individuals	281	48	228	230	38	224

All variables except the hourly wage (expressed in Euros) and the weekly hours of work are dummies. The last three dummies refer to the motivation why the workers chose an atypical contract in the treatment period: 1) because they could not find a stable job; 2) because they wanted to clear up their preferences; 3) because of flexibility needs. "Matched controls" are individuals who belong to the control sample and are used in the propensity-score matching estimation.

Table 5: Characteristics of the employed after the treatment

	TUSCANY			SICILY		
	Treated	Matched Controls	All Controls	Treated	Matched Controls	All Controls
Permanent	0.43	0.31	0.35	0.38	0.45	0.42
Atypical	0.57	0.69	0.65	0.63	0.55	0.58
Manufact.	0.47	0.37	0.26	0.42	0.07	0.14
Service	0.45	0.49	0.63	0.49	0.82	0.71
Other	0.09	0.14	0.11	0.08	0.10	0.16
Wage	6.2	7.2	7.3	6.6	7.9	7.3
Hours	37.4	34.8	32.8	36.3	29.1	30.5
N.individuals	206	70	299	144	40	268

All variables except the hourly wage (expressed in Euros) and the weekly hours of work are dummies. "Matched controls" are individuals who belong to the control sample and are used in the propensity-score matching estimation.

Table 6: Effect of a temporary mission on the probability to find a permanent job in Tuscany - *Weighted Nearest Neighbor Propensity Score Matching*

	ATT	N.treated	N.controls
Grosseto	0.31	13	11
	(0.19)		
Livorno	0.17	63	43
	(0.07)		
Lucca	0.16	69	28
	(0.07)		
Massa-Carrara	0.10	10	8
	(0.26)		
Pisa	0.21	126	45
	(0.08)		
TUSCANY	0.19	281	135
	(0.06)		

The ATT for Tuscany is obtained as the weighted average of the province-specific ATTs, in order to control for geographical stratification. Standard errors are calculated as: $SE = (\sum \frac{N_i^2}{N^2} SE_i^2)^{1/2}$, where i = pi, lu, li, gr, ms. The province-specific ATTs are obtained by using the regional estimates of the Odd of the Propensity Score, in order to control for choice-based sampling. Standard errors are reported in brackets. As a reference, note that in Tuscany the observed probability of finding a permanent job for controls is 17%, while the observed probability for the treated is 31%.

Table 7: Effect of a temporary mission on the probability to find a permanent job in Sicily - *Weighted Nearest Neighbor Propensity Score Matching*

	ATT	N.treated	N.controls
Catania	-0.02	112	51
	(0.09)		
Messina	0.15	27	18
	(0.12)		
Palermo	0.09	76	49
	(0.07)		
Trapani	0.26	15	10
	(0.17)		
SICILY	0.11	230	128
	(0.06)		

The ATT for Sicily is obtained as the weighted average of the province-specific ATTs, in order to control for geographical stratification. Standard errors are calculated as: $SE = (\sum \frac{N_i^2}{N^2} SE_i^2)^{1/2}$, where i = ct, pa, me, tp. The province-specific ATTs are obtained by using the regional estimates of the Odd of the Propensity Score, in order to control for choice-based sampling. Standard errors are reported in brackets. As a reference, note that in Sicily the observed probability of finding a permanent job for controls is 13%, while the observed probability for the treated is 23%.

Table 8: Effect of a temporary mission on the probability to find a permanent job in Tuscany - *Weighted Kernel Propensity Score Matching*

	ATT	N.treated	N.controls
Grosseto	0.23	13	85
	(0.18)		
Livorno	0.16	63	130
	(0.07)		
Lucca	0.14	69	78
	(0.07)		
Massa-Carrara	0.18	10	105
	(0.16)		
Pisa	0.19	126	104
	(0.08)		
TUSCANY	0.18	281	502
	(0.05)		

The ATT for Tuscany is obtained as the weighted average of the province-specific ATTs, in order to control for geographical stratification. Standard errors are calculated as: $SE = (\sum \frac{N_i^2}{N^2} SE_i^2)^{1/2}$, where i = pi, lu, li, gr, ms. The province-specific ATTs are obtained by using the regional estimates of the Odd of the Propensity Score, in order to control for choice-based sampling. Standard errors are reported in brackets. As a reference, note that in Tuscany the observed probability of finding a permanent job for controls is 17%, while the observed probability for the treated is 31%.

Table 9: Effect of a temporary mission on the probability to find a permanent job in Sicily - *Weighted Kernel Propensity Score Matching*

	ATT	N.treated	N.controls
Catania	0.01	112	137
	(0.08)		
Messina	0.11	27	176
	(0.13)		
Palermo	0.08	76	255
	(0.05)		
Trapani	0.27	15	175
	(0.15)		
SICILY	0.10	230	743
	(0.05)		

The ATT for Sicily is obtained as the weighted average of the province-specific ATTs, in order to control for geographical stratification. Standard errors are calculated as: $SE = (\sum \frac{N_i^2}{N^2} SE_i^2)^{1/2}$, where i = ct, pa, me, tp. The province-specific ATTs are obtained by using the regional estimates of the Odd of the Propensity Score, in order to control for choice-based sampling. Standard errors are reported in brackets. As a reference, note that in Sicily the observed probability of finding a permanent job for controls is 13%, while the observed probability for the treated is 23%.

Table 10: Heterogeneity of the treatment effect

	TUSCANY			SICILY		
	ATT	Treated	Controls	ATT	Treated	Controls
Only atypical	0.14	281	228	-0.33	230	224
	(0.16)			(0.18)		
Under 30	0.12	199	326	0.00	170	410
	(0.11)			(0.06)		
Over 30	0.37	82	302	-0.23	60	481
	(0.12)			(0.13)		
University	0.34	35	113	0.35	17	112
	(0.08)			(0.12)		
High school	0.20	174	332	-0.09	149	460
	(0.09)			(0.08)		
No high school	0.24	72	183	0.14	64	319
	(0.16)			(0.11)		
Manufacturing	0.04	169	740	0.02	123	998
	(0.06)			(0.06)		
Services	0.17	100	809	-0.01	96	1025
	(0.08)			(0.06)		

All ATTs are estimated by means of Weighted Nearest Neighbor Propensity Score Matching. They are estimated at the regional level by using appropriate weights, in order to control for both geographical stratification and choice-based sampling. Analytical standard errors are reported in brackets (also bootstrapped standard errors have been calculated, but the analytical ones lead to more conservative estimates). The first-row ATT is estimated by dropping the unemployed from the control group. The ATTs for individuals under 30, over 30, with university degree, with or without high school degree, are computed separately in these sub-samples (treatment-effect heterogeneity). The ATTs in manufacturing or service sectors are computed by interacting the TWA experience with the sector of the using firm (treatment heterogeneity). The number of controls refers to all controls and not only to matched controls.

Table 11: Sensitivity analysis for Tuscany

π	α	δ_0	δ_1	ATT	ATT OBS	L^*
0	0	0	0	0.17	0.17	-477,01
				(0.04)	(0.04)	
0.75	0.25	0.25	0.25	0.17	0.16	-478,03
				(0.04)	(0.04)	
0.5	0.25	0.25	0.25	0.17	0.15	-477,09
				(0.04)	(0.04)	
0.75	0.5	0.5	0.5	0.17	0.15	-477,37
				(0.05)	(0.04)	
0.5	0.5	0.5	0.5	0.18	0.13	-477,09
				(0.05)	(0.05)	
0.75	1	1	1	0.18	0.13	-477,79
				(0.05)	(0.05)	
0.5	1	1	1	0.19	0.09	-477,81
				(0.05)	(0.05)	
0.75	2	2	2	0.18	0.08	-477,41
				(0.05)	(0.05)	
0.5	2	2	2	0.19	-0.02	-477,43
				(0.06)	(0.06)	
0.75	-0.5	0.5	0.5	0.17	0.15	-477,48
				(0.05)	(0.04)	
0.5	-0.5	0.5	0.5	0.18	0.13	-477,32
				(0.05)	(0.05)	
0.75	-1	1	1	0.18	0.13	-479,05
				(0.05)	(0.05)	
0.5	-1	1	1	0.19	0.09	-478,01
				(0.05)	(0.05)	

This sensitivity analysis explicitely models a potential binary confounding factor, U. The full-likelihood is estimated and maximized by calibrating the sensitivity parameters. One might interpret U as unobservable ability, π as the probability of low ability, α as the effect of ability on the selection into treatment, and δ_t as the effects of ability on the potential outcomes $t = 0, 1$. "ATT" uses the estimated average outcome for both the treated and controls. "ATT OBS" uses the observed average outcome for the treated. Bootstrapped standard errors are reported in brackets. The last column reports the profile likelihood L^*.

Table 12: Sensitivity analysis for Sicily

π	α	δ_1	δ_2	ATT	ATT OBS	L^*
0	0	0	0	0.08	0.08	-388,41
				(0.04)	(0.04)	
0.75	0.25	0.25	0.25	0.07	0.07	-388,46
				(0.04)	(0.04)	
0.5	0.25	0.25	0.25	0.07	0.07	-388,44
				(0.04)	(0.04)	
0.75	0.5	0.5	0.5	0.07	0.07	-388,63
				(0.05)	(0.05)	
0.5	0.5	0.5	0.5	0.07	0.07	-388,53
				(0.04)	(0.04)	
0.75	1	1	1	0.05	0.06	-388,56
				(0.04)	(0.04)	
0.5	1	1	1	0.04	0.06	-388,51
				(0.04)	(0.04)	
0.75	2	2	2	-0.02	0.01	-388,31
				(0.05)	(0.05)	
0.5	2	2	2	-0.03	0.01	-390,35
				(0.06)	(0.06)	
0.75	-0.5	0.5	0.5	0.08	0.08	-389,02
				(0.05)	(0.05)	
0.5	-0.5	0.5	0.5	0.08	0.08	-389,59
				(0.04)	(0.04)	
0.75	-1	1	1	0.10	0.09	-391,13
				(0.05)	(0.05)	
0.5	-1	1	1	0.10	0.09	-388,54
				(0.04)	(0.04)	

This sensitivity analysis explicitely models a potential binary confounding factor, U. The full-likelihood is estimated and maximized by calibrating the sensitivity parameters. One might interpret U as unobservable ability, π as the probability of low ability, α as the effect of ability on the selection into treatment, and δ_t as the effects of ability on the potential outcomes $t = 0, 1$. "ATT" uses the estimated average outcome for both the treated and controls. "ATT OBS" uses the observed average outcome for the treated. Bootstrapped standard errors are reported in brackets. The last column reports the profile likelihood L^*.

Fig.1) Pre-treatment "gap" in Tuscany: controls vs. matched controls

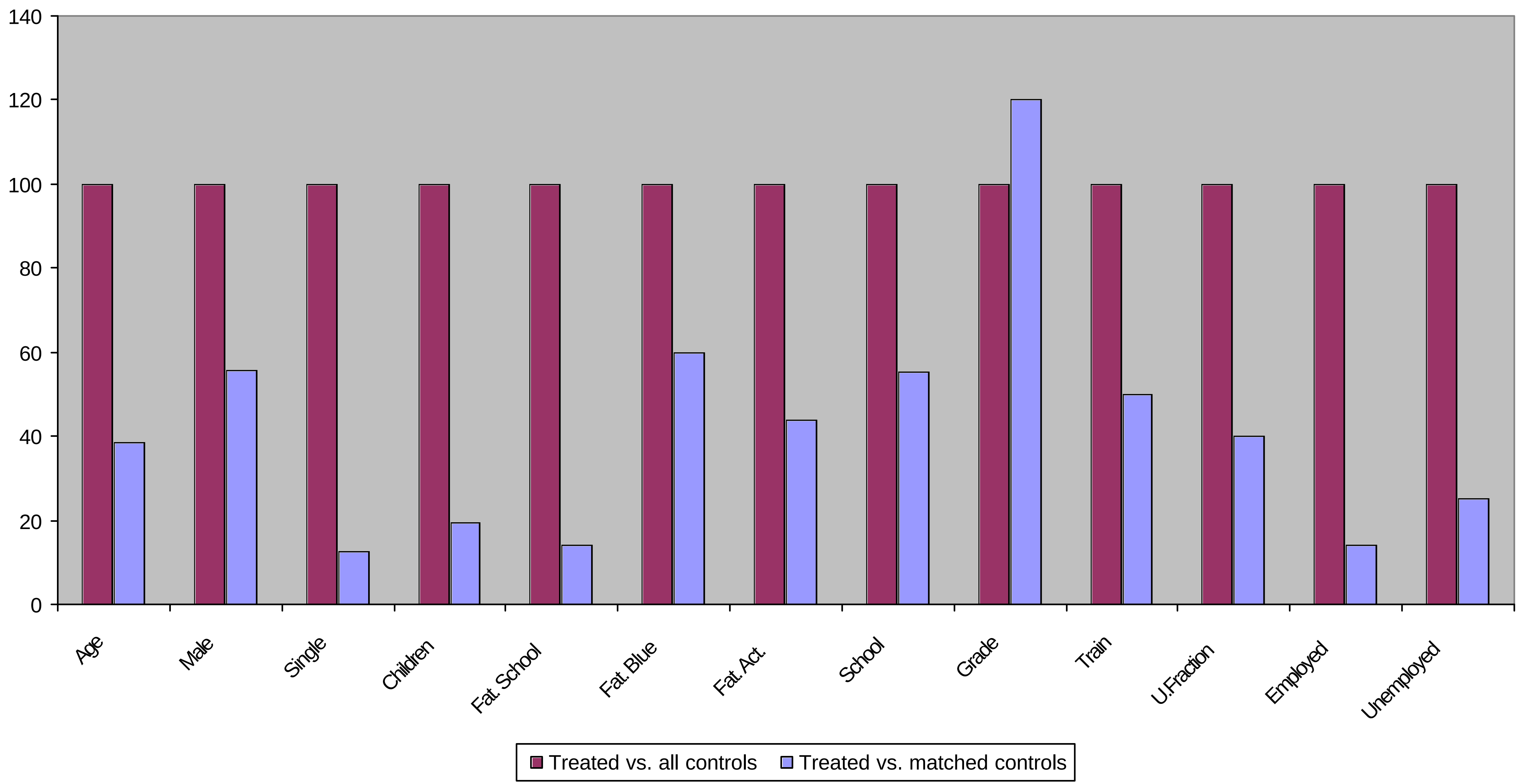

Fig.2) Pre-treatment "gap" in Tuscany: employed controls vs. matched employed controls

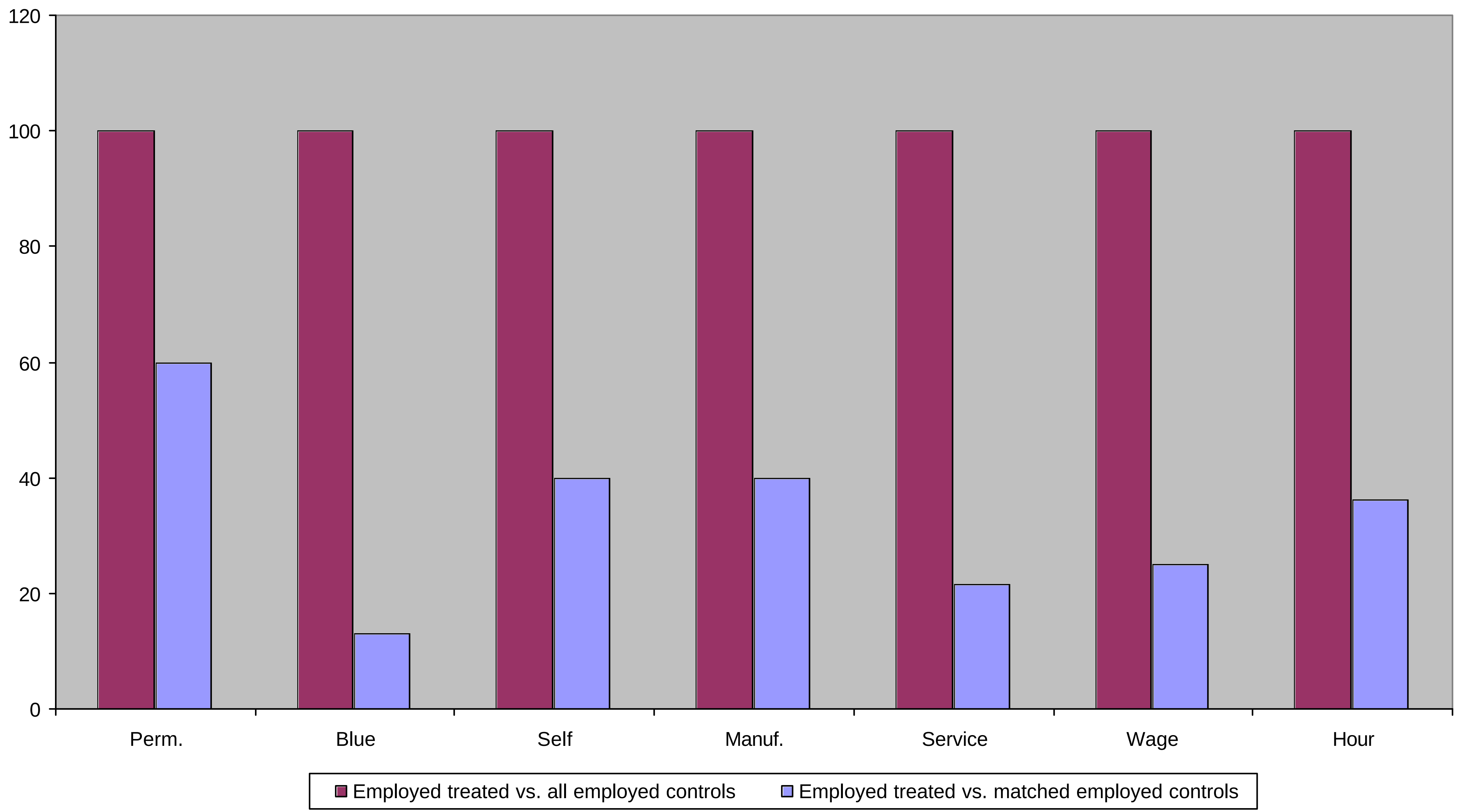

Fig.3) Pre-treatment "gap" in Sicily: controls vs. matched controls

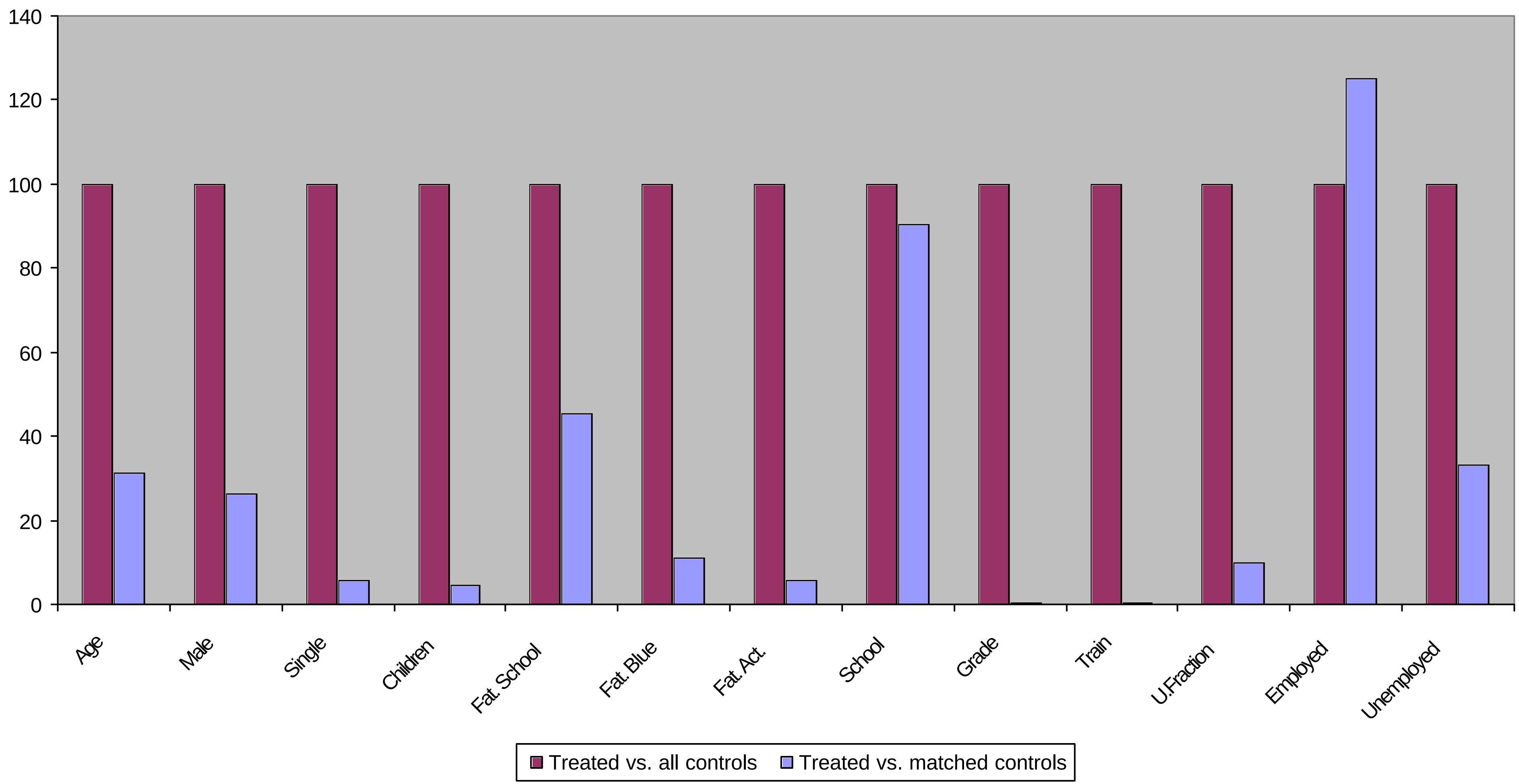

Fig.4) Pre-treatment "gap" in Sicily: employed controls vs. matched employed controls