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

May 23, 2004

Abstract

This paper measures to what extent Temporary Work Agency (TWA) employment creates a "springboard" toward permanent jobs, 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 TWA employment was liberalized in 1997, and they 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 the 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. A sensitivity analysis is performed, in order to assess the plausibility of the identifying assumption of "selection on observables". This analysis confirms the robustness of the 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 Venezia, Perugia, EUI, and the EC workshop on "Temporary Work in Europe" for insightful comments. Address correspondence to: andrea.ichino@iue.it; tommaso.nannicini@iue.it.

1 Introduction

Policy makers and labor market analysts are becoming increasingly concerned about the growth of temporary employment in Europe. According to OECD (1999), during the '90s there was a considerable continuity in the employment protection legislation of most countries, with one major exception: the deregulation of fixed-term contracts and 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 more loosely regulated while leaving the discipline of standard employment unchanged. In those 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] Triggered by this gap, 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 flexibility, job security and job quality, all at the same time.

While the evidence seems to suggest that "squaring the 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 (Autor, 2000).

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

"the circle is squared" in an intertemporal sense. This is 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). Ultimately, it is an empirical question of measuring the extent to which temporary jobs are an effective springboard toward permanent employment or a "trap" of endless precariousness.

This paper will attempt to answer this question with specific reference to Temporary Work Agency (TWA) employment. "TWA employment" signifies 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. The analysis refers to Italy, where this kind of non-standard employment was liberalized in 1997. Specifically, the goal of our study 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 will use a unique data set, which was collected precisely to perform this evaluation exercise. The data consist of the universe of TWA workers who went into a mission during the first six months of 2001, which is then compared to a sample of workers who, at the beginning of this period, were unemployed or employed with a non-permanent contract.

Interest lies mainly in the average effect of the treatment on the treated, i.e., in the difference between the outcome for the workers in the treated group with respect to the counterfactual unobservable outcome which would have prevailed for them in the absence of the TWA mission. The estimation method of Propensity Score matching in the presence of choice-based sampling will be used. Since this technique relies on the crucial assumption of "selection on observables", a particular sensitivity analysis to assess the robustness of estimates with respect to this assumption will also be performed.

The structure of the paper is as follows. Section 2 describes the take-off of TWA employment in the Italian labor market. Section 3 briefly discusses the possible determinants of the transition from temporary to permanent employment. Section 4 presents the 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 some 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 in all but the following cases: replacement of workers on strike, firms that experienced collective dismissals in the previous 12 months, and jobs that require medical vigilance. The law does not set a maximum cu-

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

mulated 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 more than four times or a cumulation period longer than 24 months.

On the whole, firing costs for standard employment remain high in the Italian labor market[4] 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, using TWA employment as a tool of personnel screening and selection is less associated with a negative "hire and fire" reputation than the utilization of other non-permanent contracts.

Following the Treu law, implemented in 1998, TWA employment has rapidly expanded, especially in the north of the country and in manufacturing sectors.[5] Nevertheless, in 2002 TWA employment still amounted to only 0.91% of total employment, far below the level observed in countries where it developed earlier. In 1999, in fact, the overall incidence was 4.5% in the Netherlands, 3.2% in the UK, 2.5% in France, and 2.5% in the US (Ciett, 2000). The average TWA utilization in the European Union was 1.5%.

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

[5]For an aggregate overview, see Ministero del lavoro (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 manufacturing sectors such as chemicals, machinery and electronics, and transportation manufacturing (Nannicini, 2004a).

3 Springboard or trap?

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

- signaling, i.e. more-able workers might signal their type by making themselves available for screening 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 might be a "bad signal" as to the lack of alternatives (especially under the firm's belief that temps have already been screened by other employers);
- a TWA experience is associated with a limited acquisition of human capital because of the high turnover.

Leaving a detailed discussion of these different mechanisms to Nannicini (2004b), 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, firms appear to be interested in TWA employment not only for screening workers 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 is 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 varying results, depending on institutional setting and evaluation strategy. Descriptive evidence for the period 1996-1998 shows a large cross-country variation in 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 analysis in different European countries leads to mixed conclusions.

Booth, Francesconi and Frank (2002) studied the labor market prospects of temporary workers in the UK (where temps represent 7% of male employees and 10% of female employees). Their results show that temporary employment is associated with lower wages, less specific training and lower job satisfaction in respect to permanent employment. But, it is not associated with negative trajectories. In particular, women that go through a temporary job are able to completely catch up to women starting in permanent positions, in terms of wage and job satisfaction. Guell and Petrongolo

(2003) studied the transformation from 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) performed an optimal matching analysis for Spain and found that TWA employment characterized 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 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. Such a comparison is unlikely to have a causal interpretation because the two groups are different irrespective of their treatment status. A growing number 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", which allows the translation of 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 analysis usually aims at estimating some features of the distribution of $Y_1 - Y_0$, e.g.,

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

which 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):

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

In order to make explicit the identifying assumptions underlying the estimators of the causal effects in our case, there is need to 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 $T \in \{0, 1\}$. Interest lies 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:

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

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

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, which determine the occupational status in t_1, the previous comments hold, that is, they may include both characteristics that are specific to the individual and characteristics of the area or local labor market where the individual lives. The two sets of variables X and Z may coincide or overlap to a certain extent. Our aim is to identify and consistently estimate the ATT. Problems may arise because of the potential association between some of the U and the treatment indicator T, as determined by the observable and unobservable variables expressed in equation (3). The identification strategy is presented in the following section.

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 regression modeling and *matching*. This assumption does not distinguish between X and Z, but considers the whole *conditioning set* of pre-treatment variables $W = (X, Z)$ and assumes that

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

and

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

This means that, conditioning on observed covariates W, treatment assignment is independent of potential outcomes. In other words, exposure to treatment is random within cells defined by the variables W. Although very strong, the plausibility of this assumption heavily relies on the amount and quality of the information contained in W.

In this study, unconfoundedness might be violated both from the labor supply and demand sides. Some of the characteristics of the area where the individual lives (e.g., the presence of high-pressure labor demand) might have attracted TWAs, making it easier for a worker to get a temporary job. These same area-specific characteristics might also ease the subsequent search for a permanent job. This is the reason why Section 6 will use the distance of each individual's home from the nearest agency to capture local labor market features not directly observed by the econometrician. Analogously, some individual unobserved characteristics might affect the propensity to get a temporary job and, at the same time, facilitate the jump into a permanent job. These remarks notwithstanding, the quality of the data and the results of the sensitivity analysis implemented in Section 7 lead us to find defendable the unconfoundedness assumption in this case.

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

$$\begin{aligned} E(Y_1 - Y_0|W) = E(Y_1|W) - E(Y_0|W) = \\ = E(Y_1|T = 1, W) - E(Y_0|T = 0, W) \end{aligned} \tag{8}$$

and also the overall ATT as

$$\begin{aligned} E(Y_1 - Y_0|T = 1) = E(E(Y_1 - Y_0|T = 1, W)) = \\ = E(E(Y_1|T = 1, W) - E(Y_0|T = 0, W)|T = 1), \end{aligned} \tag{9}$$

where the outer expectation is over the distribution of W in the subpopulation of treated individuals. An implication of the above result is that, if it is possible to 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 W are more or less continuous, so that some smoothing techniques are in order: under unconfoundedness, several estimation strategies 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 instance, if there are many covariates, it can be difficult to find an appropriate specification. Moreover, regression modeling obscures information on the distribution of covariates in the two treatment groups. In principle, 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 the treatment effects.

It is thus crucial to check how much the two distributions overlap, and what the "region of common support" is for the distributions. When the number of covariates is large, this task is not an easy one. One approach is to reduce the problem to a single 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 Score have the same distribution of observable (and possibly unobservable) characteristics independently of the treatment status; also, the exposure to treatment or control status is random for a given value of the Score. These two properties allow use of the Propensity Score as a univariate summary of all W. It is enough to check the distribution of the Score in the two groups, and use the Score in the ATT estimation procedure as the single covariate that needs to be adjusted for. In fact, adjusting for the Propensity Score automatically controls for all observed covariates, at least in big samples. As a result, if $p(W_i)$ is known, the ATT can be estimated as follows:

$$\begin{aligned}
\tau &\equiv E(Y_{1i} - Y_{0i}|T_i = 1) = \\
&= E(E(Y_{1i} - Y_{0i}|p(W_i), T_i = 1)) = \\
&= E(E(Y_{1i}|p(W_i), T_i = 1) - E(Y_{0i}|p(W_i), T_i = 0)|T_i = 1)
\end{aligned} \tag{10}$$

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 the covariates with linear and higher order terms. Inasmuch as the specification of $h(W_i)$ which satisfies the balancing property is more parsimonious than the full set of interactions needed to match treated and control units according to observable characteristics, the Propensity Score reduces the dimensionality problem of matching procedures based on the multidimensional vector W.[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 (10). In fact, the probability of observing two units with exactly the same value of the Score is in principle zero, since $p(W)$ is a continuous variable. Various methods have been proposed in the literature to overcome this problem.[7] This study adopts two of them, *Nearest Neighbor Matching* and *Kernel Matching*, presented here 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

$$C(i) = \{j \mid j = \arg\min_j \parallel p_i - p_j \parallel\}, \tag{11}$$

[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.

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

which is a singleton set unless there are multiple nearest neighbors. In practice, the case of multiple nearest neighbors should be rare, particularly if the set of observable 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:

$$\begin{aligned} \tau^M &= \frac{1}{N^D} \sum_{i \in T} \left[Y_i^D - \sum_{j \in C(i)} w_{ij} Y_j^C \right] \\ &= \frac{1}{N^D} \left[\sum_{i \in D} Y_i^D - \sum_{i \in D} \sum_{j \in C(i)} w_{ij} Y_j^C \right] \\ &= \frac{1}{N^D} \sum_{i \in D} Y_i^D - \frac{1}{N^D} \sum_{j \in C} w_j Y_j^C \end{aligned} \tag{12}$$

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:

$$\begin{aligned} Var(\tau^M) &= \frac{1}{(N^D)^2} \left[\sum_{i \in D} Var(Y_i^D) + \sum_{j \in C} (w_j)^2 Var(Y_j^C) \right] \\ &= \frac{1}{(N^D)^2} \left[N^D Var(Y_i^D) + \sum_{j \in C} (w_j)^2 Var(Y_j^C) \right] \\ &= \frac{1}{N^D} Var(Y_i^D) + \frac{1}{(N^D)^2} \sum_{j \in C} (w_j)^2 Var(Y_j^C). \end{aligned} \tag{13}$$

Standard errors can also be obtained by bootstrapping.

The Kernel Propensity Score matching estimator is instead given by:

$$\tau^K = \frac{1}{N^D} \sum_{i \in D} \left[Y_i^D - \frac{\sum_{j \in C} Y_j^C G(\frac{p_j - p_i}{h_n})}{\sum_{k \in C} G(\frac{p_k - p_i}{h_n})} \right] \tag{14}$$

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

$$\frac{\sum_{j \in C} Y_j^C G(\frac{p_j - p_i}{h_n})}{\sum_{k \in C} G(\frac{p_k - p_i}{h_n})} \tag{15}$$

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

4.4 Choice-based sampling

The matching estimators presented in the previous section can be straightforwardly applied if data are obtained with a simple random or stratified sampling design, with known and observed stratification variables included in the vector W. The data collection scheme (see Section 5.1) here is a stratified sampling design, where one of the two stratifying variables is the province of residence, which is included in the pre-treatment 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 the model for the Propensity Score, i.e., $Pr(T = 1|W)$. This type of sampling scheme is usually called "choice based sampling" (Manski and Lerman, 1977) or, in general, "endogenous stratification".

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

Under unconfoundedness, regression analysis is robust with respect to such an endogenous sampling scheme. With regression modeling, endogenous stratification can only affect efficiency. On the contrary, the application of estimation strategies based on the preliminary estimation of the Propensity Score is more problematic in the presence of choice-based sampling. Denoting with A the variables that identify the province of residence, our sampling scheme allows a certain number of observations to be sampled at random from each of the strata defined by $A \times T$. Hence, every observation is characterized by the probability distribution $Pr(Y, W|A, T)$, with $Y = Y_1T + Y_0(1 - T)$. Sample data allow estimation of the 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 to each other, via Bayes theorem, in the following way:

$$Pr(W|A, T = j)Pr(T = j|A)Pr(A) = Pr(T = j|W, A)Pr(W|A)Pr(A) \tag{16}$$

where $j = 0, 1$, so that

$$\frac{Pr(W|A, T = 1)Pr(T = 1|A)}{Pr(W|A, T = 0)Pr(T = 0|A)} = \frac{Pr(T = 1|W, A)}{Pr(T = 0|W, A)} \tag{17}$$

and

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

where $\tilde{Pr}(T = 1|W, A)$ disregards choice-based sampling and $\tilde{Pr}(T = 1|A)$ 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 strategy implemented in the evaluation project had the following stages and characteristics. The analysis focused on a region at the center of Italy (Tuscany) and one in the south (Sicily), which were among the areas with incomplete penetration of TWAs in 2000. Five provinces with an agency (Livorno, Pisa, Lucca, Catania, Palermo) and four provinces without any agency (Grosseto, Massa, Messina, Trapani) were selected. This first step allowed for two opportunities: 1) to identify observations very similar in respect to all individual characteristics except 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 econometric analysis performed in Section 6, this second opportunity will be exploited, under the assumption that -within every province- TWAs locate themselves in the areas with higher labor demand, making it easier to meet potential client firms.

"Manpower Italia Spa", a major company operating in the TWA sector with a national market share of about 25%, provided the dataset of workers they hired. From this dataset, workers who were on a TWA mission in one of the nine provinces mentioned above during the first semester of 2001 were extracted and interviewed. 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: 1) phone interviews to all temps who were resident in the nine provinces and were in a TWA mission during the first semester of 2001; 2) phone interviews to a random sample of "controls" drawn from the population of the nine 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 stable job (either an open-ended contract or self-employment) on January 1, 2001.

In a sense, this first screening of potential control observations might be interpreted as part of the matching strategy, aimed at identifying a common support for the treated and the control units 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, the 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, according to the methodology described in Section 4.4. It should be taken into account, moreover, that this data collection strategy combines *flow sampling* for the treated group and *stock sampling* for the control group. As long as the transition probability from permanent employment to unemployment or non-permanent employment is very low, there is no loss in relevant information from the control group.

For the treated units, the reference point in time is the date of the TWA mission during the first semester of 2001. For the control units, it is January 2001. Information on the period before these reference points provided "pre-treatment" variables, while information on the date of the interview (November 2002) provided "outcome" or "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 achievements; 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, control individuals who were out of the labor force in the treatment period (e.g. students) were dropped from the sample. In fact, these subjects showed characteristics that made them not easily comparable with the treatment units. Notice that this was 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 for treated and control units. The final data set used for the empirical evaluation already contains control units who could be more meaningfully *matched* with the treated units.

At the end, the treated sample contains individuals who lived in the nine provinces and were on a TWA mission through "Manpower" during the first semester of 2001; while the control sample contains residents in the nine provinces, aged 18-40, who belonged to the labor force but were not permanent workers 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? Notice that the control sample might include subjects who went through a TWA experience in a period different from the first semester of 2001. This is because the treatment coincides with "a TWA mission during the first semester of 2001". If the outcome of some control units were affected by a TWA experience in another period, our exercise would result in conservative estimates.

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

5.2 Descriptive statistics

Table 1 reports the distribution of the observations across the nine 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% did the same in Sicily.[11]

Table 2 summarizes the relevant information available for all individuals in the sample. This table, as well as the following ones, presents the average characteristics of an important sub-sample of controls, dubbed the "matched controls". These control units are used as "nearest neighbors" of at least one treated unit in the Nearest Neighbor Propensity Score matching estimation.[12] Inasmuch as the treated units are more similar to the "matched controls" than to "all controls", the matching strategy has succeeded in improving the quality of the comparison used to estimate the causal effect of interest.

[8] See Ichino, Mealli and Nannicini (2004) for further details on data collection.

[9] The exact number of "Manpower" temps in each province in the first semester of 2001 is known. The population of unemployed and atypical workers aged between 18 and 40 in each province was estimated by combining Istat statistics and the answer rate of the first screening question in phone interviews. The ratio of the second to the first term is the province-specific weight.

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

[11] See Ichino, Mealli and Nannicini (2004) for further data details.

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

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 36% (Sicily 25%) had an atypical contract, while 64% (75%) were looking for a job. In 2002 -the "outcome" period- 31% of the treated had a permanent position in Tuscany, compared with 17% of the controls. In Sicily, the same comparison is 23% versus 13%.[13] Of course, these are simple correlations that need to be cleansed from observable and unobservable influences, just as 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. One 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 most relevant difference concerns the firm's sector: TWA workers are mainly used in the manufacturing sector (60% in Tuscany and 53% in Sicily), while the

[13]Incidentally, note 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%. Among the treated who are employed in the outcome period, 38% (34% in Tuscany and 43% in Sicily) are working in the same firm of the TWA mission. See Ichino, Mealli and Nannicini (2004) 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 (2004) for further data details.

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 jobs (against 59% of the other atypical workers); 22% became temps 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 also provide information 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 (again see 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 matched controls as a fraction of the analogous difference between treated and all controls. Inasmuch as this relative difference is smaller than 100, the 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 the treated units considerably reduces the "pre-treatment gap". This reduction is large, both in Tuscany and Sicily, even though it encounters some problems in the case of employment variables in the latter. As explored in the next section, this might be due to the specific characteristics of this regional labor market.

6 Estimated causal effects

Tables 6 through 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, whether the father is living; 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 barely 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 able to create 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 standard errors are reported. Bootstrapped standard errors have been calculated as well, but the analytical ones lead to more conservative estimates.

[15]On the procedure to calculate these weights, see footnote 9.

This heterogeneity analysis confirms non-significant results for Sicily. Only for a marginal minority of workers with university degrees does TWA employment have a strong and significant effect. In Tuscany, on the contrary, the heterogeneity analysis shows interesting results.

In the first row of Table 10, the ATT is estimated by dropping the unemployed from the control group. In Tuscany, only the 228 controls who were employed with an atypical contract in the treatment period were considered.[16] 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 ones 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, over 30, with a university degree, and with or without a 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"). These estimations show that the effect of the treatment on the treated is greater for individuals over 30 years, with a university degree (even though they are a small minority) or in the service sector. The most surprising result is the one regarding age, which shows that young workers in the Italian labor market generally wait for quite a long period of time before finding a stable job.

[16] In Tuscany, these workers had the following contracts: 53% fixed-term, 14% co.co.co. (a particular Italian arrangement), 6% training contract, 18% irregular employment, 9% other occasional work. In Sicily: 44% fixed-term, 11% co.co.co., 2% training contract, 26% irregular employment, 17% other occasional work.

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, the results might be criticized since this assumption rules out the role of unobservables. Parametric selection models, which in principle allow a relaxation in 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, and should, 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 one 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. Here, the analysis by Rosenbaum and Rubin (1983b) is adapted to the Propensity Score matching estimation.

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 the analysis results in no particular loss of generality. The central assumption of this analysis is that the assignment to treatment is not unconfounded given the set of observable variables W, i.e.,

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

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

$$Pr(T = 1|Y(0), Y(1), W, U) = Pr(T = 1|W, U). \tag{20}$$

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 W. 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 W and U, the joint distribution of $(Y(t), T, W, U)$ for $t = 0, 1$ is

$$Pr(Y(t), T, W, U) = Pr(Y(t)|W, U)Pr(T|W, U)Pr(U|W)Pr(W). \quad (21)$$

It is further assumed that:

$$Pr(U = 0|W) = Pr(U = 0) = \pi \quad (22)$$

and

$$Pr(T = 0|W, U) = (1 + \exp(\gamma' W + \alpha U))^{-1} \quad (23)$$

and

$$Pr(Y(t) = 1|W, U) = \exp(\beta' W + \tau T + \delta_t U)(1 + exp(\beta' W + \tau T + \delta^t U))^{-1} \quad (24)$$

where π represents the proportion of individuals with $U = 0$ in the population, and the distribution of U is assumed to be independent of W. This does not mean, of course, that there is need to assume that the unobservables whose effect we are trying to capture (e.g. ability) have to be independent of all the observable characteristics, but that we are testing the sensitivity of the estimates to the part of the unobservables that is orthogonal to the variables W. This renders the sensitivity analysis even more unfavorable to causal conclusions, since, if U were associated with W, controlling for W 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, one can estimate 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 the definition of the profile log-likelihood

$$\begin{aligned} L^*(\pi, \alpha, \delta_0, \delta_1) &= \max_{\gamma, \beta | \pi, \alpha, \delta_0, \delta_1} L(\gamma, \beta, \pi, \alpha, \delta_0, \delta_1) \\ &= L(\widehat{\gamma}(\pi, \alpha, \delta_0, \delta_1), \widehat{\beta}(\pi, \alpha, \delta_0, \delta_1), \pi, \alpha, \delta_0, \delta_1). \end{aligned} \tag{25}$$

In our case, 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:[17]

$$A\hat{T}T = \frac{1}{N^T} \sum_{i \in T} [\hat{Y}_i^1 - \hat{Y}_i^0] \tag{26}$$

where

$$\hat{Y}_i^t = \pi \hat{P}r(Y(t) = 1 | W, U = 0) + (1 - \pi) \hat{P}r(Y(t) = 1 | W, U = 1). \tag{27}$$

To further adjust this methodology to the needs of our project, one can condition on the treated and matched control samples, that have equal distribution of all the covariates W. 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

[17] Note that one could have used, intuitively, $\frac{1}{N^T} \sum_{i \in T} [Y_i^1 - \hat{Y}_i^0]$, as an estimate of the ATT, i.e., the observed rather than the estimated Y_1 for the treated. However, if assumption $Pr(T = 1|Y(0), Y(1), W, U) = Pr(T = 1|W, U)$ holds, then averaging over the observed values of Y_1 for the group of treated individuals implicitly implies averaging over the distribution $Pr(W, U|T = 1)$, whereas in $\hat{Y}_i^0$ the average is over the distribution $Pr(U)Pr(W|T = 1)$.

is an unobserved stratifying variable. This approach allows one 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

$$Pr(T = 0|W, U) = (1 + \exp(\alpha U))^{-1}. \tag{28}$$

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 Nearest Neighbor Propensity Score matching procedure, since one is using only the control units selected by the Nearest Neighbor algorithm.

Tables 11 and 12 show the results obtained with the modified version of the sensitivity analysis 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. 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. This is the reference point of the sensitivity analysis, and it is only slightly lower than the ATT estimated by Propensity Score matching: 16 percentage points against 19 in Tuscany; 8 points against 11 in Sicily. One might interpret the binary unobservable variable U as individual ability, affecting both (self-)selection into treatment and 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. "ATT" is calculated using the estimated outcome for both the treated and control units, for the reasons explained in footnote 17.

Note 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 Tuscany, the effects remain positive and significant, except for the extreme 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 the odds of 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. Note, in fact, that no variable has a coefficient higher than 1 in the previous estimations of the Propensity Score and outcome equations. As expected, in the case of negative selection (e.g. less able workers self-selecting into treatment), the estimated ATTs are higher than those at the baseline. Similar robust results are obtained in Sicily, with the estimated ATT remaining positive (but not significant) in nearly all cases. On the whole, the sensitivity analysis strongly confirms the robustness of the Propensity Score matching estimates.

8 Conclusions

This paper has investigated 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, the causal effect of the treatment "TWA mission" on the outcome "finding a permanent job after 18 months" was estimated.

The analysis referred to Italy, where TWAs were liberalized in 1997 and we had the opportunity to gather data appropriately collected for this evaluation exercise. 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, with an university degree (even though they are a small minority of temps) or in the service sector (rather than in manufacturing). A particular sensitivity analysis was also performed, in order to assess the plausibility of the identifying assumption of "selection on observables". This analysis confirms the robustness of the results.

From a policy perspective, this 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, this study suggests that labor market programs in Europe, Italy in particular, should be increasingly evaluated with econometric methods specifically aimed at the identification of *causal effects*. Only in this way does the political debate have 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 A., Mealli F. and Nannicini T. (2004), *Il lavoro interinale in Italia. Trappola del precariato o trampolino verso un impiego stabile?*, Regione Toscana, Edizioni Plus - Universit di Pisa

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

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 postal codes. 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.individuals	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 by workers to choose 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 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 available controls and not only to matched controls.

Table 11: Sensitivity analysis for Tuscany

π	α	δ_0	δ_1	ATT	L^*
0	0	0	0	0.16	-480.42
				(0.04)	
0.75	0.25	0.25	0.25	0.16	-480.44
				(0.04)	
0.5	0.25	0.25	0.25	0.16	-480.84
				(0.04)	
0.75	0.5	0.5	0.5	0.16	-481.01
				(0.04)	
0.5	0.5	0.5	0.5	0.15	-481.07
				(0.04)	
0.75	1	1	1	0.13	-480.59
				(0.05)	
0.5	1	1	1	0.13	-481.07
				(0.04)	
0.75	2	2	2	0.06	-480.61
				(0.05)	
0.5	2	2	2	0.03	-480.47
				(0.05)	
0.75	-1	1	1	0.19	-482.98
				(0.05)	
0.5	-1	1	1	0.19	-480.54
				(0.04)	
0.75	1	0	1	0.15	-481.15
				(0.05)	
0.5	1	0	1	0.15	-480.51
				(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 $t = 0, 1$. 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	L^*
0	0	0	0	0.08	-388.41
				(0.04)	
0.75	0.25	0.25	0.25	0.07	-388.46
				(0.05)	
0.5	0.25	0.25	0.25	0.07	-388.44
				(0.05)	
0.75	0.5	0.5	0.5	0.07	-388.63
				(0.05)	
0.5	0.5	0.5	0.5	0.07	-388.53
				(0.04)	
0.75	1	1	1	0.05	-388.56
				(0.04)	
0.5	1	1	1	0.04	-388.51
				(0.04)	
0.75	2	2	2	-0.02	-388.31
				(0.05)	
0.5	2	2	2	-0.03	-390.35
				(0.06)	
0.75	-1	1	1	0.10	-391.13
				(0.05)	
0.5	-1	1	1	0.10	-388.54
				(0.04)	
0.75	1	0	1	0.06	-388.54
				(0.04)	
0.5	1	0	1	0.07	-388.43
				(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 $t = 0, 1$. 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

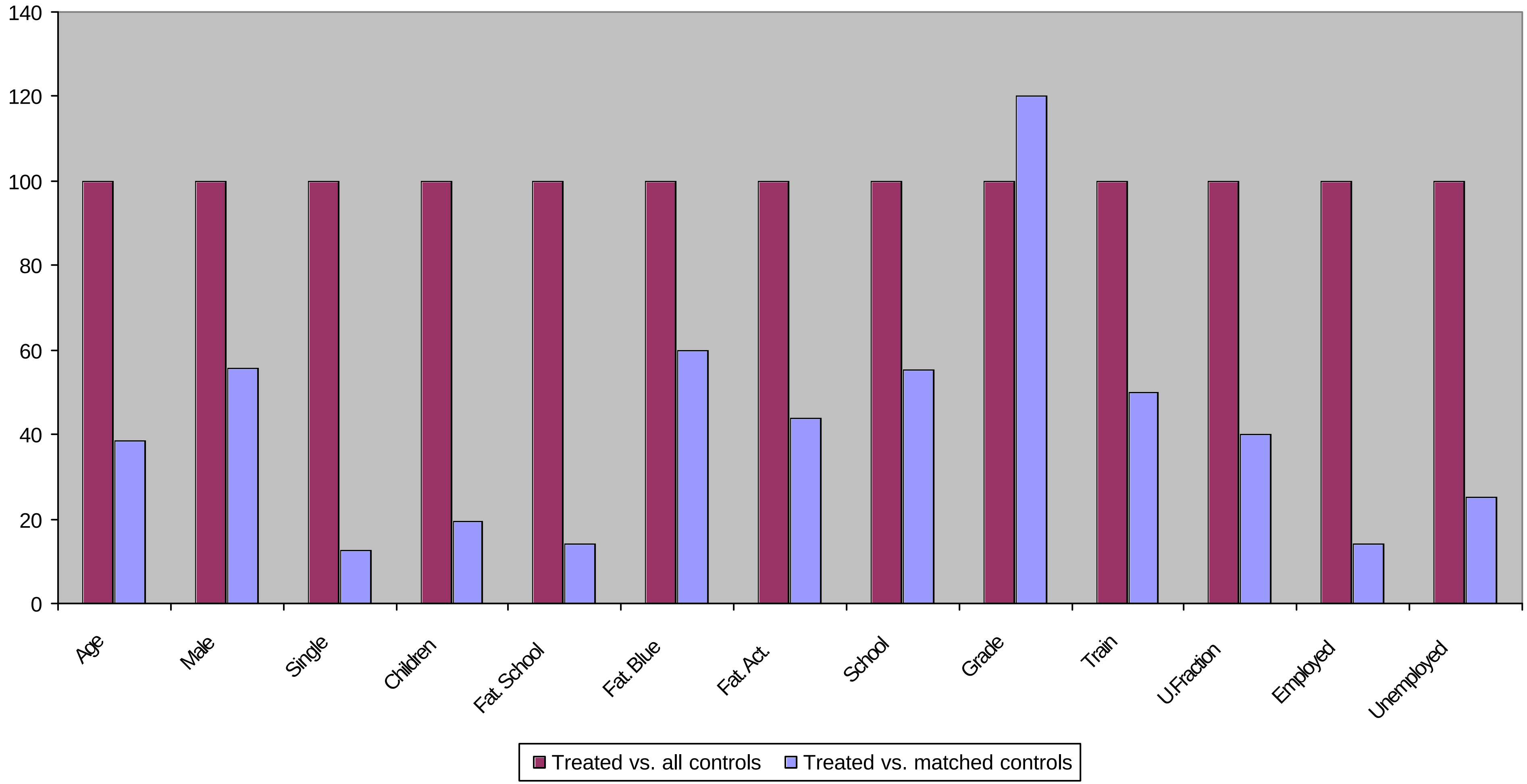

Fig.2) Pre-treatment "gap" in Tuscany: employed controls vs. matched employed controls

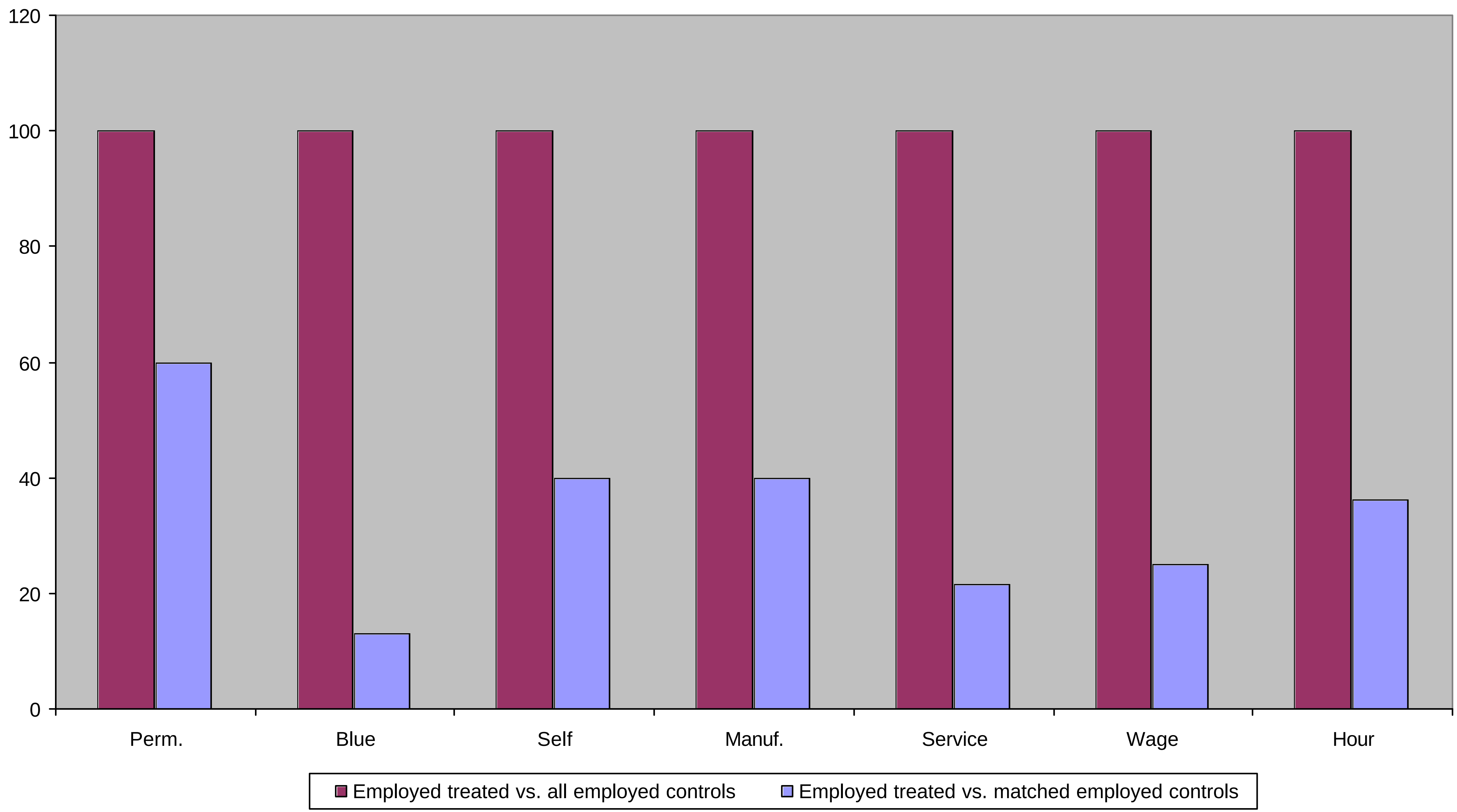

Fig.3) Pre-treatment "gap" in Sicily: controls vs. matched controls

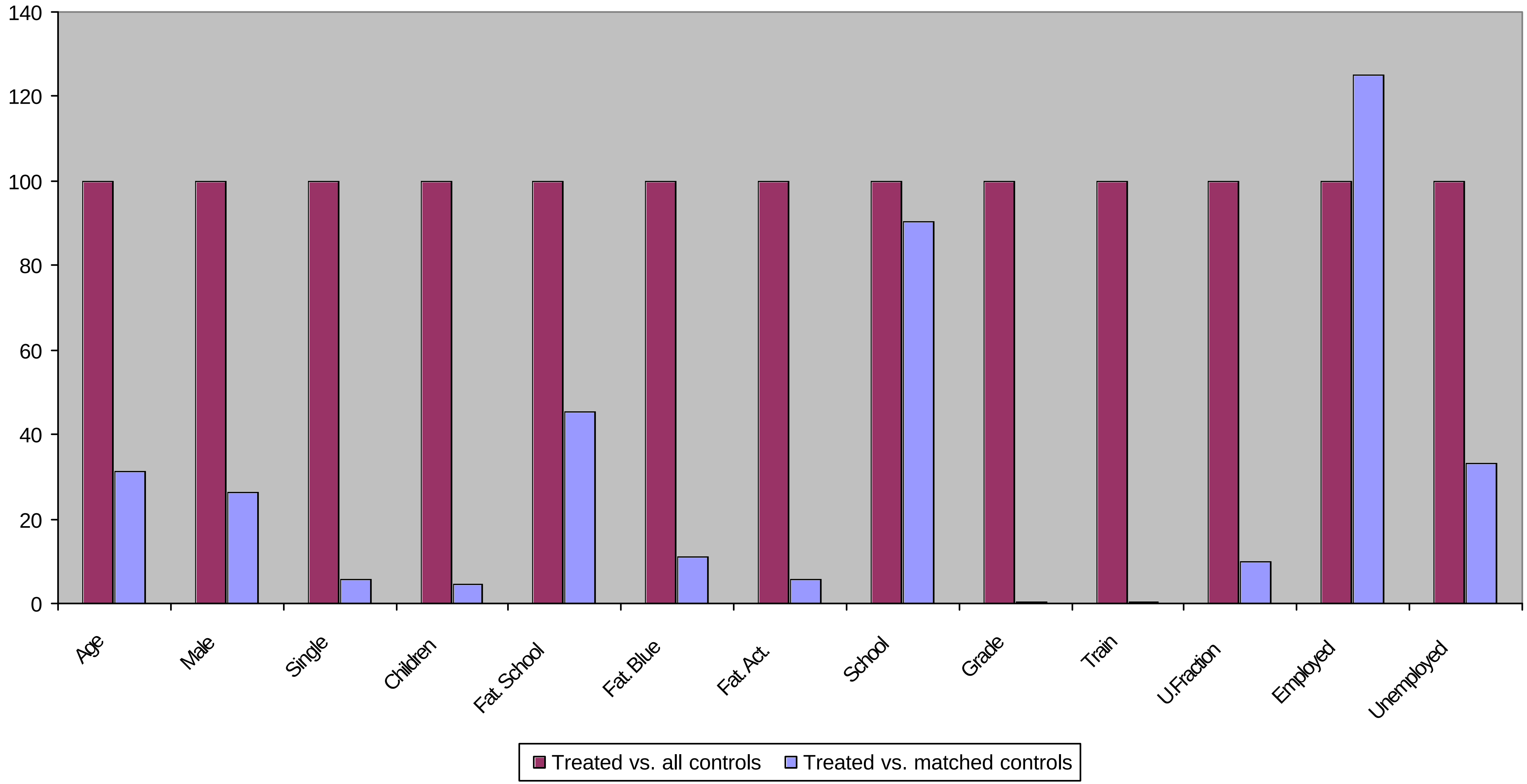

Fig.4) Pre-treatment "gap" in Sicily: employed controls vs. matched employed controls