# Are temporary work agencies stepping stones into regular employment?

Joakim Hveem

Correspondence:
joakim.hveem@sofi.su.se
Swedish Institute for Social Research
(SOFI), Stockholm University, SE-106
91 Stockholm, Sweden

**Abstract**

This paper estimates the causal effect of temporary work agency (TWA) employment on the subsequent probability of employment in the regular labor market. The main purpose is to estimate the stepping-stone effect separately for natives and immigrants, where the latter group potentially benefits the most from TWA employment. Since no quasi-experiment is available, individual Differences-in-Differences and matching is used to deal with the potential selection bias. The results point at a negative regular employment effect, which slowly fades away over a couple of years. Thus no evidence of a stepping-stone effect is found. When conditioning on immigrants, this negative effect is absent. A long-run significant effect is found on overall employment probability (including TWA employment), there is even a long-run positive effect on annual earnings (mainly driven by women). Unemployment probabilities decreased, however the results in the estimation were less stable over time compared to the employment estimates, suggesting that the TWAs might keep individuals from exiting the labor market.

**Keywords:** Temporary work agencies; Stepping stone; Labor market; Matching

## 1 Introduction

The growth of temporary working agencies (TWAs) in Sweden and Europe has been rapid since the 1990s and the sector is still expanding in Sweden. The 2011 level of penetration[1] was about 1.4% (Bemanningsföretagen 2011) which amounted to an all time high: 62,863 employees (yearly full time equivalents). Thus, it is interesting to investigate whether this development has been of any advantage for the unemployed in terms of increased transition rates into regular employment.

Undoubtedly the deregulation of the market in 1993 was a major contributor to the rapid development since it made TWAs legal, only prohibiting agencies from charging employees for their services and imposing a six months TWA contracting stop if a job position has been terminated. The driving force behind the temporary work industry (TWI) is primarily the demand side of the labor market. Increased competitive pressure has forced employers to change their organizational structure towards the 'lean' production model: the permanent work force is adjusted to the minimum production levels and increased demand is met with atypical employment such as temporary workers or workers hired through a TWA.

The rationale for hiring through a TWA instead of recruiting a regular temporary employee is that there are costs associated with hiring and firing which can be mitigated by the TWAs. Furthermore, the tasks required at a company might not comprise enough to constitute even a part-time position. TWAs have the advantage of bundling together different tasks into one or several employment positions. Since recruiting is the TWAs' main function, it is argued that they have the advantage of economies of scale, which would imply that they are more efficient in both the time elapsed until sealing an employment contract and the quality of the match. This matching efficiency is a theoretical result by (Neugart and Storrie 2006) and also claimed by the TWI itself. Empirically, there is however only inconclusive evidence.

Previous studies that in various degrees might lend support to the stepping stone hypothesis are (Andersson et al. 2007; Ichino et al. 2008; Lane et al. 2003; Summerfield 2009; Garcia-Perez and Munoz-Bullon 2005; Amuedo-Dorantes et al. 2008), while (Kvasnicka 2008) does not find any evidence of such an effect. (Autor and Houseman 2010) takes advantage of a quasi-experimental research design which give a small but negative result. However, heterogeneous effects with respect to country of birth are not explored in these studies.

In the Swedish setting, research is scarce on this question and causal inference is as yet inconclusive. (Andersson and Wadensjö 2004) focus on the TWA's stepping stone role for non-European immigrants. Their findings (based on register data) show that immigrants—in relation to natives—more often leave a TWA for another type of employment. This could be interpreted in favor of the stepping stone hypothesis. The causal inference is however weak and mainly relies on reasoning based on *probit* correlations.

In a recent study, (Jahn and Rosholm 2012) investigate the effect of TWAs on the duration until exit from unemployment into regular employment for immigrants. They find significant and positive results when measuring in-treatment effects but nothing significant when examining the post-treatment effect. However, when dividing the sample into smaller groups, such as non-western immigrants, they found a post-treatment effect but no in-treatment effect[2]. The study was performed in Denmark, implying that the results might be applicable to the Swedish labor market since they are not that different from each other.

There is evidently no real consensus in the field, the results span from positive, none, to negative, underscoring that this is a relatively unexplored research topic.

When client firms hire personnel from a TWA they assume—after accounting for hiring and firing costs—that the worker provided is the best possible match and that their own effort in recruiting would not be able to compete with the TWA's outcome. One hypothesis is that the TWAs increase the probability of a worker's gaining permanent employment in the regular labor market by increasing their human capital, signaling working ambitions, expanding the worker's network, and serving as a cheap screening device. The last two effects are especially vital for immigrants since they are likely to be subject to statistical discrimination. The type of screening service that a TWA offers in effect is likely to be one of the most powerful remedies against statistical discrimination since the client firm will be able to observe the real productivity without hiring the worker. Since employers usually have a hard time adequately assessing an immigrant's abilities, education, and skills acquired in a different environment, TWAs could be a remedy, working as a cheap probation device where the uncertainty and risk has been incorporated into the TWA

itself. It is also reasonable to believe that the immigrant's working network is weaker than natives' and just getting a job helps build up country-specific human capital, such as language skills and a deeper knowledge of how the labor market works[3]. Another reason to further investigate the effect on immigrants is their over-representation in the industry (Andersson Joona and Wadensjö 2010). All these things taken together give reason to believe that immigrants may specially benefit from TWA employment. The opposite sign is of course also possible due to stigmatization. In the present paper I will estimate the causal impact of employment in a TWA on the medium and long-run transition rate from unemployment to regular labor market employment, with extra focus on non-western immigrants[4].

This paper is organized as follows. Section 2 describes the data and the definition of the treatment group. Section 3 presents the estimation framework and outlines the matching estimation. The results from the various estimations are reported in Section 4, both for the matched and the unmatched sample, ending with a brief summary of the robustness check. Section 5 concludes.

## 2   The Data

Even though the TWI is growing rapidly it still does not constitute more than approximately 1.4% of the labor force. Therefore, I will have to use large datasets such as register databases in order to retrieve a sufficiently large sample. I have access to the composite register data used in (Andersson and Wadensjö 2004). The main part of the data comes from the *Longitudinal integration database for health insurance and labor market studies* (LISA) provided by Statistics Sweden. The composite panel database is balanced and gives information about, e.g., age, gender, place of birth, education, place of residence, employment status, annual income, days in unemployment, etc., for about 7.3 million individuals 16–64 years old, covering the years 1997 to 2008, making this study the largest in the TWA field to date[5]. Since the research question at hand is whether a TWA works as a stepping stone out of unemployment into regular employment, identification of the population is based on the unemployment status in November 2001[6]. This means that the sample under study comprises those unemployed or in a labor market program in November 2001. Also, individuals older than 55 in 2001 are pruned out of the data in order to reduce the probability that their subsequent labor market outcome is affected by any early retirement plan[7]. Taking treatment is then defined as being registered at a TWA in November 2002. Subsequent years are recorded as labor market outcomes (see Table 1, "Appendix C. Description of outcome variables" for a description of the outcomes) or effect of treatment. The control group comprises those not joining a TWA in November 2002 (when unemployed in 2001), entry in TWA 2003 and onwards is allowed since I do not want to condition on future outcomes[8].

**Table 1 Definition of labor market outcomes**[a]

| | |
|---|---|
| Outcome 1 | Annual probability of regular employment |
| Outcome 2 | Annual probability of unemployment |
| Outcome 3 | Annual probability of TWA employment |
| Outcome 4 | Annual probability of employment (TWA or regular) |
| Outcome 5 | Annual earnings |

[a]The outcomes are not mutually exclusive.

One drawback of the dataset is that administrative staff is not separately coded with the workers out for hire. The administrative staff is however a small share of TWA employment and this issue should not affect the results in any significant way[9]. Another caveat is that the data frequency is low, relying on annual observations, which makes the treatment definition somewhat imprecise due to the absence of employment status information between the pre-treatment year 2001 and the treatment year 2002. However this will most likely only affect the precision and not bias the estimates. Another drawback with low frequency data is that we cannot observe the in-treatment effect (the contemporary effect of treatment): any treatment effect taking place within a year will not be recorded due to the data structure.

When relying on a selection on observables design, it is crucial to obtain all relevant pre-treatment observables that might be correlated with both the outcome and the selection into the treatment equation. The most vital observable is previous unemployment duration, since this is highly correlated with both the outcome and the selection equation. Due to the long period covered in the data, much credibility is gained since the parallel trends assumption can be tested thoroughly. Moreover, deducing the long-run effect can be done in a more convincing manner than if solely relying on a permanent employment indicator, since we can follow the individuals over several years into post-treatment.

Table 2 presents descriptive statistics for selected variables and regression-based $t$-tests for the treatment and control group (corresponding table for the non-western subsample is found in Table 4 in the Appendix). The two groups are not balanced in most dimensions. The treatment group is younger, has higher education, has on average been unemployed less, resides in Stockholm and Gothenburg to a greater extent, and has a higher share of males and natives. These last two facts run counter to the cross-sectional summary statistics of the population in 2002, where the opposite is true (Andersson and Wadensjö 2004). This is a consequence of the sample selection since I actually identify observations by flow rather than stock in this paper.

Controlling for the observed characteristics observed in Table 2 parametrically by an OLS means that obtaining unbiased results heavily relies on correct model specification.

In Figure 1 we can follow different employment outcomes over different sample groups. The plots show no visible positive effect on the subsequent outcome after treatment by just comparing means. However, when examining income means and medians in Figure 2, we seem to detect a positive income effect, although a slight divergence in earnings is revealed in the early years.

## 3 Methodology

When trying to estimate causal effects, the main issue is to deal with the selection bias caused by omitted variables. Eliminating or at least mitigating selection bias is the key to obtaining unbiased estimates (Angrist and Pischke 2009). Since I do not have access to a randomized experiment (or a quasi-experiment) two approaches will be used: Individual fixed effects and matching, both relying on the *conditional independence assumption* (CIA).

### 3.1 Differences-in-Differences

An ideal econometric approach would have been to identify an exogenous variation into taking treatment or not that would have ensured a causal inference. Since the focus is

**Table 2 Summary statistics for 2001, full sample**

| | Group | | Differences in means | |
|---|---|---|---|---|
| | Control | Treatment | Coefficient | *t*-statistic |
| **Gender** | | | | |
| Male | 57.0 | 65.1 | 0.082 | (17.03)*** |
| Female | 43.0 | 34.9 | -0.082 | (-17.03)*** |
| **Aggregate days in unemployment 1998–2001** | **403** | **340** | **-62.43** | **(-16.27)*** |
| 0 | 0.8 | 0.6 | | |
| 1–30 | 2.2 | 1.4 | | |
| 31–90 | 8.0 | 11.4 | | |
| 91–182 | 14.4 | 17.2 | | |
| 183–274 | 12.3 | 15.1 | | |
| 275–364 | 11.2 | 14.6 | | |
| 365–730 | 37.6 | 32.2 | | |
| 731–1094 | 12.8 | 7.2 | | |
| 1095–1457 | 0.4 | 0.2 | | |
| **Age groups** | **35** | **31** | **-3.430** | **(-10.75)*** |
| 16–20 | 8.7 | 12.3 | 0.022 | (8.42)*** |
| 21–25 | 15.4 | 23.6 | 0.055 | (14.50)*** |
| 26–30 | 15.3 | 18.8 | 0.057 | (14.05)*** |
| 31–35 | 14.5 | 15.0 | 0.017 | (4.65)*** |
| 36–40 | 14.9 | 10.8 | -0.018 | (-5.56)*** |
| 41–45 | 12.0 | 8.8 | -0.036 | (-12.37)*** |
| 46–50 | 10.6 | 6.0 | -0.041 | (-15.77)*** |
| 51–55 | 8.6 | 4.7 | -0.040 | (-18.31)*** |
| **Country of birth** | | | | |
| Sweden | 73.9 | 75.0 | 0.012 | (2.69)*** |
| Nordic countries (except Sweden) | 3.2 | 2.8 | -0.004 | (-2.34)*** |
| EU12 | 0.9 | 0.6 | -0.003 | (-3.79)*** |
| Other European countries | 7.0 | 6.7 | -0.003 | (-1.23) |
| Africa | 2.2 | 2.2 | -0.000 | (-0.23) |
| North America | 0.4 | 0.4 | 0.000 | (0.72) |
| South America | 1.5 | 1.7 | 0.002 | (1.44) |
| Asia | 10.7 | 10.3 | -0.004 | (-1.25) |
| Oceania | 0.0 | 0.0 | -0.000 | (-9.95)*** |
| Soviet Union | 0.2 | 0.2 | 0.001 | (1.15) |
| **Highest education level** | | | | |
| Primary school less than 9 years | 3.9 | 1.5 | -0.024 | (-17.80)*** |
| Primary school 9 (10) years | 17.9 | 14.4 | -0.019 | (-4.87)*** |
| Upper secondary 2 years or less | 34.0 | 27.5 | -0.065 | (-14.49)*** |
| Upper secondary 2 years or more | 27.2 | 35.5 | 0.072 | (15.75)*** |
| Higher education less than 3 years | 8.6 | 10.5 | 0.016 | (4.98)*** |
| Higher education 3 years or more | 7.3 | 9.9 | 0.023 | (7.80)*** |
| Post-graduate education | 0.2 | 0.2 | -0.001 | (-4.59)*** |
| Unknown | 1.0 | 0.6 | -0.003 | (-3.89)*** |
| **Resides in a large city** | | | | |
| Stockholm | 12.3 | 19.4 | 0.076 | (18.98)*** |
| Gothenburg | 17.0 | 23.7 | 0.060 | (14.37)*** |
| Malmoe | 15.4 | 10.9 | -0.045 | (-14.60)*** |
| Other | 55.4 | 46.0 | -0.091 | (-18.26)*** |
| **Number of observations** | 25,158 | 953 | 26,111 | |

Robust standard errors.
*** *p* < 0.01, ** *p* < 0.05, * *p* < 0.1.
*Note: Taking treatment is defined as being registered at a TWA in November 2002.*
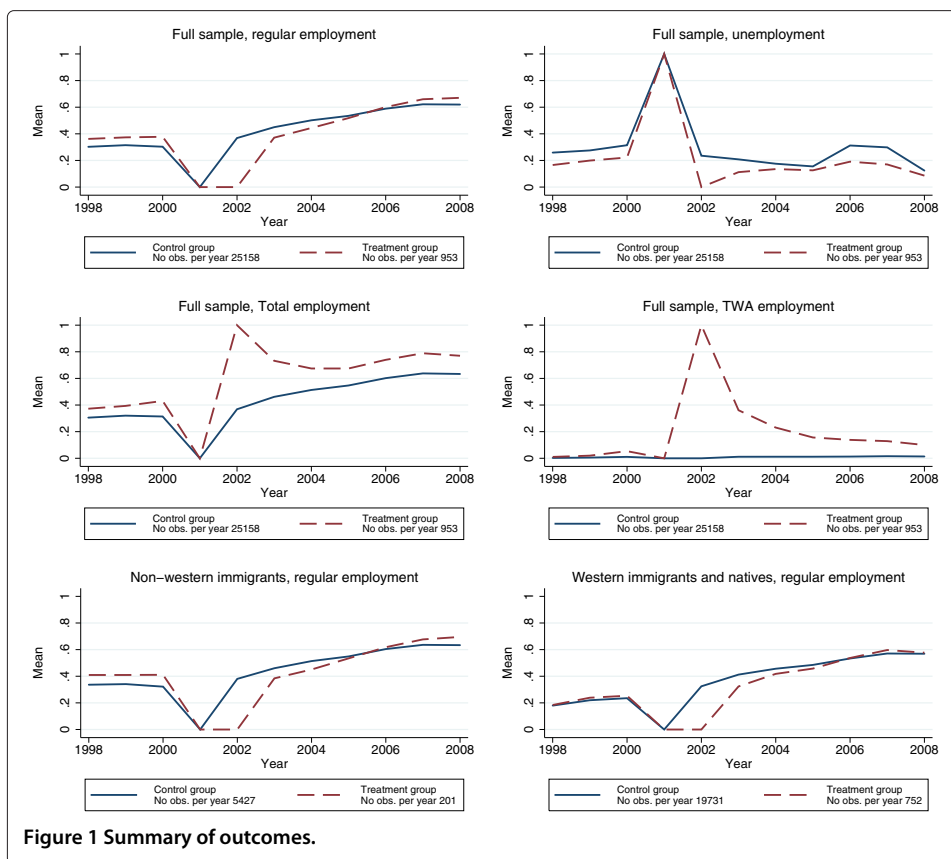
**Figure 1 Summary of outcomes.**

on the causal effect of TWA employment (taking treatment) on the subsequent regular employment probability, I will turn to the difference-in-difference model. When estimating this model it is crucial that: (i) both groups exhibit parallel trends, (ii) selection into treatment is conditionally random or at least not correlated with the outcome, and (iii) nothing else that affects the outcome variable occurs at the same time as the treatment timing. Since I have not been able to identify an exogenous instrument that selects individuals into treatment, I instead control for individual effects (exploiting the panel data structure) which might be correlated with both selection into treatment and the outcome. However, individual trends cannot be captured by this way of modeling. Another way to deal with this endogeneity is to employ matching techniques which relies on the CIA. It states that conditional on observable characteristics, treatment is as good as randomly assigned, more formally:

$$T|X \quad \perp \quad Y_{0i} \tag{1}$$

Here, $T$ is getting treatment, $X$ is a set of confounders, and $Y_{0i}$ the potential outcome if not taking treatment. Having a rich set of observables is, as previously stated, crucial for being able to claim that the CIA holds and thus that the conditional differences-in-differences (cDiD) is valid. More specifically, observables such as previous labor market performance prior to treatment might control for the unobservable characteristics that cause a selection bias. One way of applying the matching approach is to weight the most crucial variables in the estimation in order to balance the two groups so that they look very similar along observable dimensions. This can be done by *coarsened exact matching*
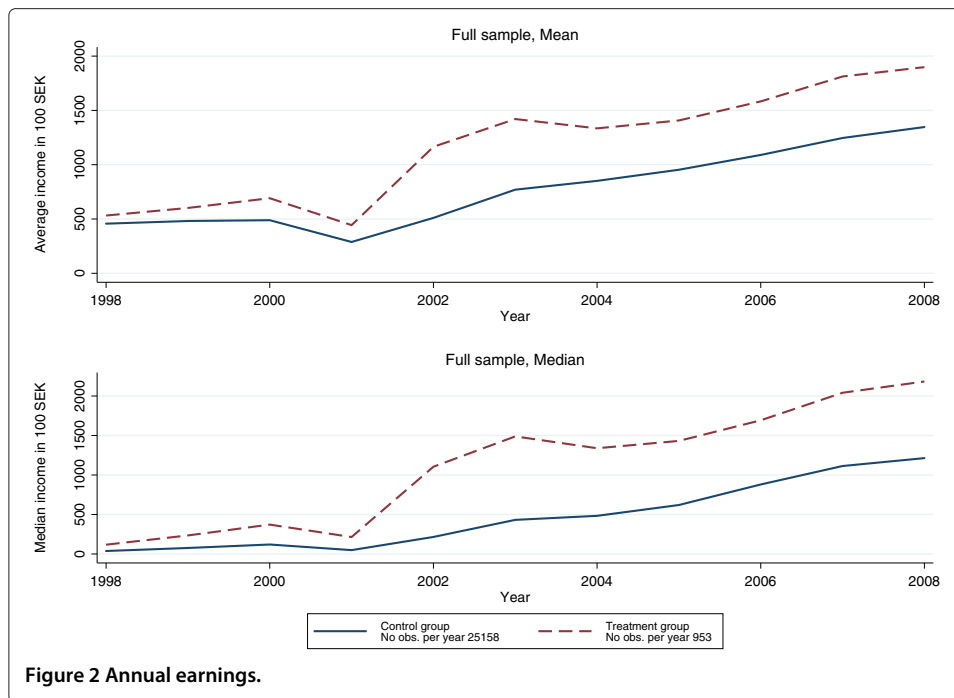
**Figure 2 Annual earnings.**

(CEM) where the reweighing is done individually on the different confounders depending on relevance[10].

The outcome variable of main interest is the long-run probability of getting employed in the regular sector, relying on employment status in several periods afterwards, which will basically capture this long-run performance in the labor market[11]. The outcome will be defined as $P(Y = 1)$, the probability of getting employed in the regular sector. An initial problem is that $P(Y = 1)$ is not observed in the pre-treatment period of 2001 since the criterion in this year is that the individuals under study should be unemployed. However we do not need outcome data for 2001 as long as we have outcomes for earlier years such as 2000 and 1999. 1998 will be used as the reference year, the parallel trends assumption can be tested using data for 1999 and 2000[12]. The main iDiD and cDiD[13] models estimated in this paper will be specified as

$$Y_{i,t} = \gamma_t + a_i + \sum_{\rho=1999}^{2000} \delta_\rho T_{i,\rho} + \sum_{\tau=2002}^{2008} \delta_\tau T_{i,\tau} + \beta' \mathbf{X}_{i,t} + v_{i,t} \qquad (2)$$

and

$$Y_{i,t} = \gamma_t + \sum_{\rho=1999}^{2000} \delta_\rho T_{i,\rho} + \sum_{\tau=2002}^{2008} \delta_\tau T_{i,\tau} + \beta' \mathbf{X}_{i,t} + v_{i,t}. \qquad (3)$$

Here, $\alpha$ is the intercept, the $a_i$ are the individual dummies, $\gamma_t$ is a set of time dummies, $\mathbf{X}_{i,t}$ is a set of confounders, $v_{i,t}$ is the error term, $T_{i,\rho} = 1$[if will be getting treatment], and $T_{i,\tau} = 1$[if treated]. Given that there are no anticipation effects (by construction impossible in this setting) the coefficient of the leads ($\delta_\rho$) should be zero (i.e., parallel trends) strengthening our causal link hypothesis. Since the matching is performed before

estimating the equation, we omit any treatment group dummy in Equation (3) since its coefficient will be zero if the matching was successful. Including a group dummy would, in that case, only inflate the standard errors while not contributing to the model. The timing of the treatment is chosen to be 2002 to ensure a long follow-up.

Joining a TWA after the treatment year by either the control or treatment group is permitted and thus selection is orthogonal to future outcomes. Violating this and in effect conditioning on future outcomes would lead to a bias of the estimated effect (Fredriksson and Johansson 2003).

Both groups are identified by being unemployed in 2001, thus a form of matching on pre-treatment labor market outcomes is already performed here. The control group is defined as not joining a TWA in 2002 and instead engaging in something else or staying unemployed. The counterfactual path is then, e.g., taking up studies, dropping out of the labor force, taking a regular job, etc. No restrictions were put on any outcomes from 2003 to 2008.

The reason for using both iDiD and cDiD (to be described in the following section) is that we can expect iDiD to give more precise estimates by construction, compared to a regular DiD by controlling for individual effects rather than two group effects. Also, controlling for unobserved and observed heterogeneity with fixed effects does not prune out observations like *coarsened exact matching* (CEM)[14] does: CEM might result in very few observations. On the other hand, cDiD can, by balancing the two groups, mitigate the bias occurring when, for instance, the two groups have different age compositions, something which can give rise to diverging income progressions (steeper for younger people). Using well-balanced groups it is also more convincing to point at the control groups' outcome as the actual counterfactual outcomes since they are the same in all observed aspects. The drawback then is of course the low number of observations that arise due to tight matching criteria. Contrasting these two methods with each other will also give the reader a feel for how big the self-selection bias might be in this application.

### 3.2   Matching

Matching is a technique to overcome the selection bias threatening causal inference. The approach is, however, not uncontroversial. Evidence pointing in favor of the technique comes from, e.g., (Dehejia and Wahba 1999), who report a successful non-experimental analysis on the data in (LaLonde 1986): using matching, they replicate the experimental impact estimates. (Smith and Todd 2005) criticizes (Dehejia and Wahba 1999), but conclude that the matching technique is best put in a DiD-design, which is what is done in the present paper (*conditional DiD*). A principle conceptual difference between regular regression estimations and matching estimation is that the latter gives the researcher greater flexibility in choosing how to aggregate heterogeneous effects, especially when using the specific technique *coarsened exact matching*. Since previous work has shown that the impact TWAs have on individuals differ greatly among groups, this is of great importance. Due to the explicit and easily manipulated weighting procedure, which is in the hands of the researcher instead of implicitly in the estimator (as in OLS), matching makes it easier to estimate the interesting parameters such as the ATT in a stratified way (Cobb-Clark and Crossley 2003).

The basic idea with matching estimators is that we try to find a 'twin' for each individual taking treatment. This is done by matching on observable characteristics. The idea is that

if the individuals are very similar in the observables that are related to the outcome and selection process, the risk of their being different in unobservables that are correlated with outcome and selection is reduced or even eliminated. In practice, we explicitly try to calculate the counterfactual untreated outcome $E[Y_{0i}]$.

$$\delta_i^* = Y_{1i} - E[Y_{0i}]. \tag{4}$$

Matching estimations rely on the CIA as discussed earlier. Furthermore, (Rosenbaum and Rubin 1983) noted that an additional condition was needed, *common support*: If we define $P(x)$ as being the probability of getting treatment ($T$) for an individual with characteristics $x$, then the common support condition requires $0 < P(x) < 1, \quad \forall x$. This is also called the *overlap condition* and it rules out the perfect predictability of $T$ given $x$; without this assumption, we have no information to construct our counterfactuals. CEM takes care of this by construction.

*Coarsened exact matching* is a member of the Monotonic Imbalance Bounding (MIB) class of matching methods (further described in (Iacus et al. 2008)). It is a method of pre-processing data which deals with the 'curse of dimensionality'[15] by coarsening continuous data into bins where the researcher by in-depth knowledge of the variables at hand can determine the size of the bins to preserve information and maximize the number of matches. When the continuous variable is coarsened into bins, matching will take place on the respective strata and then observations are finally re-weighted according to the size of their strata. The bin width can be constant ($\epsilon_j$) within the variable $j$ or it can vary within each variable, $\epsilon_j^v$, where the $v$ are the cut-off points. Then basically any type of regression can be performed while including the new weights on the uncoarsened data. If the matching is exact in a variable—which is done for, e.g., the educational level—then this confounder is not needed in the regression since the balancing is perfect, unless the variable is time varying. If the matching is exact only on the coarsened values and/or is time varying, then the confounder should be included in the regression to control for the within-bin correlation which most likely will be very small if the bin width ($\epsilon_j$) is tightly defined.

The rationale for using CEM instead of, e.g., *propensity score matching*, is because this technique is more transparent, straightforward, by construction deals with the *common support*, gives priority to balancing (thus reducing bias and model dependence) to variance (high precision), meets the congruence principle, is computationally efficient, and reduces the sensitivity to measurement error (which would lead to biased estimates of the ATT, see (Iacus et al. 2008)). In Figures 5 and 6 in the Appendix, two kernel density plots over aggregated days in unemployment from 1998 to 2001 and two histograms over age are graphed before and after matching to give a visual representation of what is going on in the matching process. The treatment group has a higher density over the left region of *days in unemployment* and vice versa over the right region, this skewness is adjusted through the matching. A similar adjustment takes place in the variable age, where the treatment group has a lower density in the left region and vice versa in the right. Notably the sample under study becomes a quite young sample compared with the population. Since balancing the two groups to each other changes the average sample characteristics compared to the population, we in effect measure the *sample average treatment effect on the treated* (SATT) when applying CEM.

Matching was performed exactly in 2001—unless otherwise is specified—on gender, level of education, and marital status; coarsened exact matching on aggregate days in unemployment from 1998 to 2001[16] ($\epsilon_{unemp.}$ = 2 days), age ($\epsilon_{age}$ = 5 years), annual earnings in 2000 ($\epsilon^{v}_{earnings}$ where $v$ = [0, 5000, 10,000, 15,000, 20,000, 25,000, 30,000, 35,000, 40,000, 60,000, 100,000, 200,000]). For non-western immigrants, the income distribution was completely different and the following break points were chosen to get a sufficient number of matches: $v$= [0, 3000, 10,000, 50,000]. The number of children over age groups ($\epsilon^{v}_{children_{age}}$ where $v$=[0.5, 1.5, 2.5] and $age$=[0–3, 4–6, 7–10, 11–15, 16–17, 18+]). When matching the non-western sample, *marital status* was not included since it reduced the number of matches and did not help to establish parallel trends. *Region of birth* was not included since it reduced the number of matches severely: if *region of birth* were included in the regression, the *F*-test would not be able to reject the hypothesis that the coefficients are zero (at the 5% significance level)[17].

## 4   Results

This section will begin with multiple cross-sectional regressions measuring how the subsequent labor market outcome varies over time from the year 2002, followed by brief results from the iDiD. Lastly, the cDiD design will be contrasted with the iDID results.

### 4.1   Unmatched results

I first use repeated OLS over the years 2003 to 2008:

$$Y_{t,i} = \alpha_t + \beta_t (TWA)_{2002,i} + \gamma'_t \mathbf{X}_{t,i} + u_{t,i} \quad \text{for} \quad t = 2003 : 2008. \tag{5}$$

Here, $Y$ is either regularly employed, unemployed or employed. Thus each line of Table 7 in the Appendix is a separate regression measuring the effect of joining a TWA in year 2002 on subsequent labor market outcomes.

The estimates in Table 7 in the Appendix hint at a locking-in effect during the first years, which fades away and becomes insignificant in 2006 and 2007 though still negative. For unemployment, we find the same mechanical decrease of unemployment in the first year. Those who joined a TWA have a risk of unemployment in 2003 that is 6.3 percentage points lower, but already in 2004 the estimate becomes insignificant. In 2006 and 2007 the estimates once again turn significant. It would, however, be quite a stretch to draw any inference about that, since the earlier insignificant estimates suggest that we are most likely picking up something unobserved systematic, e.g., the business cycle. The overall probability of employment rises by 0.22 and then hovers around 0.10, implying that treatment raises the participants' overall employment rate. The non-western immigrants section of the table show similar results but with more unfavorable figures compared to the full sample in all aspects. These results will be contrasted against the more causally robust cDiD and iDiD estimates.

Making use of the individual DiD design, we estimate the equation for the full sample and the subsample *non-western immigrants*[18]. The estimated coefficients, i.e., the *average treatment effects on the treated* (ATT), are seen in Table 8 in the Appendix[19]. The estimates exhibit a similar pattern as in Table 7 in the Appendix but overall they are more unfavorable for the stepping stone hypothesis. The results hardly vary when conditioning

on *non-western immigrants*, though one could argue that they do not suffer any TWA stigma[20](or, at least, suffer less of a TWA stigma) since the post-treatment estimates on regular employment are lower than for the full sample and revert back to insignificance more quickly. When performing the individual DiD estimation on income, the parallel trends assumption is in the danger zone. Figure 2 suggests a small diverging trend from 1998 to 2001; the individual FE:s do not control for the differences in the pre-existing trend. This is picked up by the iDiD estimator in Figure 8 in the Appendix and suggests that the estimates are unreliable. In the following section, matching will prove itself useful compared to fixed effects in mitigating these sorts of problems.

### 4.2 Matched results

To measure the overall imbalance in the unmatched dataset, which is visible in Table 2, we can use the $\mathcal{L}_1$ statistic introduced by (Iacus et al. 2008) as a comprehensive measure of global imbalance. Starting by discretizing the continuous variables by using a pre-defined binning algorithm[21] and binning the variables by the researcher's choice. A comparison will be made between the two approaches using the following statistic

$$\mathcal{L}_1(f,g) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

where $k$ is the number of dimensions (or variables), $f$ is the treated, $g$ is the control, and $\ell$ is the variable imbalance. $f_{\ell_1 \cdots \ell_k}$ is then the $k$-dimensional relative frequency for the treated. $\mathcal{L}_1 = 0$ means perfect balance and $\mathcal{L}_1 = 1$ means perfect imbalance. The measure by itself is not that informative but computing the pre-matching $\mathcal{L}_1$ and comparing it to the post-matching $\mathcal{L}_1$ will show if the matching was successful.

The first balancing is done on the full sample where pre-matching $\mathcal{L}_1 = 0.99$ and post-matching $\mathcal{L}_1 = 0.86$. The summary statistics reported in Table 5 in the Appendix exhibit a clear improvement relative to Table 2. Since matching on *aggregate days in unemployment* was performed at two-day precision—and not in the intervals specified in the table—the balanced result cannot be entirely visible in the table (though it can be viewed in Figure 5 in the Appendix. *Country of birth* is not perfectly balanced since no matching was made on that variable. Still, improvement has been made; only the *Nordic countries* are unbalanced in the treatment group's favor (2 percentage points more in the treatment group). The fact that balancing on some variables gives rise to balancing on other observable variables adds to the plausibility of the CIA assumption, since it is not unreasonable to believe that unobserved characteristics also might become balanced. The post balancing results for the subsample *non-western immigrants* are shown in Table 6 in the Appendix. The balance is not perfect in the mean *Aggregate days in unemployment* but the difference is insignificant. Balancing had not been performed on *country of birth* or *year of arrival*, yet they still balanced after matching. The $\mathcal{L}_1$ statistic equals 1.0 for pre-matching and 0.8 post-matching, an even better improvement than for the full sample matching. If we compare the $\mathcal{L}_1$ statistics and the matched tables to the unmatched, it is clear that improvements have been made. However, the common support condition implies a substantial reduction in the number of observations, which affects the precision of the estimates.

Figure 3 compares the iDiD and the cDiD estimates of the probability of being regularly employed for the non-western immigrants and the full sample. In Figure 3 the different *sample average effects on the treated* (SATT) for the outcome employment are plotted over time, the estimates for all labor market outcomes are displayed in Table 3. For the full sample, the probability of regular employment is not significantly different from zero in 2007 while the iDiD estimates were always significantly negative[22]. Stratification on gender shows that women endure negative regular employment estimates until 2006 while men's estimates only remain significant until 2004, see Figure 7 in the Appendix. Another change that has taken place—compared to the iDiD estimates—is in the regular employment probability for non-western immigrants, where there is no evidence of a negative significant effect, which is in line with the theory of a more favorable subsequent labor market outcome for immigrants than for natives. However, this is rather an absence of adverse effects than a prevailing positive effect. Notably, the standard errors are large and the estimated insignificant effect for 2003 might just be due to imprecisions caused by the reduced number of observations.

The estimates for unemployment (Table 3) are in agreement with the overall employment results, by never getting anywhere near significantly positive. The cyclical pattern of unemployment exhibited in the iDiD estimates has been reduced by the matching process, thus showing a favorable image for TWA's effect on the transition out of unemployment. Apart from 2005, until 2007 all estimates are significantly negative, hovering around 0.04 to 0.08.

For instance, joining a TWA in 2002 reduced the probability of getting unemployed by 4 percentage points in 2004. Compared to the iDiD estimates where the unemployment significantly fluctuated around zero, the cDiD estimates do not. If negative self-selection was the reason for the cyclical pattern then it seems likely that the matching has mitigated the problem, making these estimates more reliable. The unemployment estimates
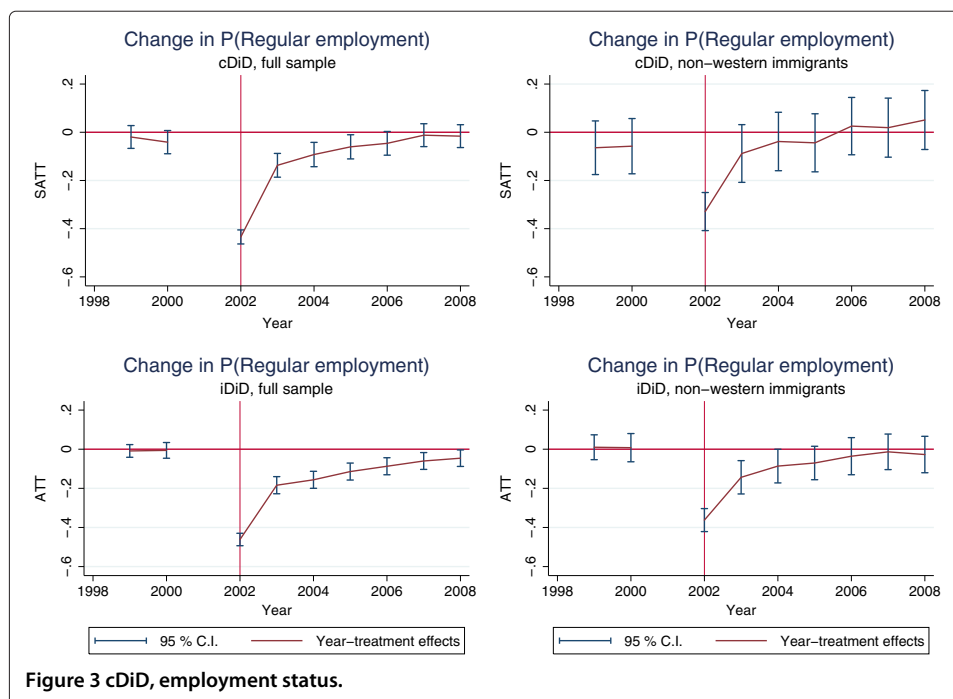


**Figure 3 cDiD, employment status.**

**Table 3 cDiD coefficient estimates**

| VARIABLES | (1) Regular employment Full sample | (2) Unemployment Full sample | (3) Employment Full sample | (4) Regular employment Non-western immigrants | (5) Unemployment Non-western immigrants | (6) Employed Non-western immigrants |
|---|---|---|---|---|---|---|
| Treatment × 1999 | -0.022 | 0.007 | -0.004 | -0.071 | 0.018 | -0.069 |
| | (0.024) | (0.019) | (0.024) | (0.056) | (0.051) | (0.057) |
| Treatment × 2000 | -0.043* | 0.003 | -0.008 | -0.059 | 0.094 | -0.012 |
| | (0.025) | (0.020) | (0.025) | (0.058) | (0.058) | (0.061) |
| Treatment × 2002 | -0.439*** | -0.172*** | 0.561*** | -0.331*** | -0.250*** | 0.667*** |
| | (0.015) | (0.011) | (0.015) | (0.041) | (0.037) | (0.041) |
| Treatment × 2003 | -0.143*** | -0.053*** | 0.203*** | -0.091 | -0.043 | 0.198*** |
| | (0.025) | (0.017) | (0.024) | (0.060) | (0.051) | (0.058) |
| Treatment × 2004 | -0.097*** | -0.038** | 0.122*** | -0.038 | -0.054 | 0.139** |
| | (0.025) | (0.017) | (0.024) | (0.061) | (0.051) | (0.058) |
| Treatment × 2005 | -0.063** | -0.011 | 0.073*** | -0.036 | -0.044 | 0.119** |
| | (0.025) | (0.016) | (0.024) | (0.061) | (0.048) | (0.058) |
| Treatment × 2006 | -0.050** | -0.048** | 0.068*** | 0.028 | -0.097* | 0.117** |
| | (0.025) | (0.020) | (0.023) | (0.061) | (0.057) | (0.058) |
| Treatment × 2007 | -0.017 | -0.060*** | 0.087*** | 0.017 | -0.154*** | 0.146** |
| | (0.024) | (0.020) | (0.022) | (0.062) | (0.056) | (0.057) |
| Treatment × 2008 | -0.020 | 0.004 | 0.052** | 0.044 | -0.103** | 0.153*** |
| | (0.024) | (0.014) | (0.022) | (0.062) | (0.045) | (0.058) |
| | | | | | | |
| Observations | 40,400 | 40,400 | 40,400 | 3,730 | 3,730 | 3,730 |
| R-squared | 0.126 | 0.056 | 0.140 | 0.182 | 0.101 | 0.230 |
| Control variables | YES | YES | YES | YES | YES | YES |

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

*Control variables* ($X_{i,t}$): Age, number of kids over age groups, and dummies for: region of birth, year of arrival, educational orientation and level and social welfare benefit received current year. For the non-western immigrant sample, region of birth was not included, but dummies for marital status were added.

Standard errors were clustered on individual level to account for possible serial correlation.

for non-western immigrants exhibit an insignificant positive trend before taking treatment, the in-treatment effect is then significantly negative, followed by four consecutive years of negative insignificant estimates and then two years of significantly negative estimates. The pattern displayed in the unemployment column for *non-western immigrants* is similar to the full sample unemployment column even though—maybe due to imprecision—more estimates are insignificant and also the impact is larger. For instance: in 2007 taking treatment lead to a 15.4 percentage points drop in the probability of being unemployed for non-western immigrants and 8.7 percentage points for the full sample.

The overall employment has also changed significantly from the unmatched estimates. The iDiD estimates where significant at the 5%-level for only two years whereas the cDiD estimates never go insignificant. The employment columns in Table 3 shows clear cut evidence of a long-run change in the probability of employment for both the full sample and the immigrant subset. Given that the matching successfully eliminated the self-selection bias, we can causally interpret this as joining a TWA increases one's probability of employment in the long run by approximately 7 to 9 percent points in general and 12 to 15 percent points for non-western immigrants.

Given the effect on employment probability it might also be interesting to take a closer look at the income progression that one would expect from joining a TWA. Previous studies usually show that the working conditions in TWAs are worse and that the salaries are lower than those in regular employment contracts (Jahn 2008). The coarsened exact matching technique showed itself useful in the income equation where the process purged out individuals with diverging pre-treatment income trends and reduced the standard errors. A time trend has also been included in the earnings regression apart from the other confounders included in the employment status cDiD. Figure 4 and Table 10 in the Appendix report the estimates. There are no subsequent adverse earnings effects from joining a TWA. In fact the income progression seem to benefit from TWA employment, supporting the descriptive results in Figure 2. When stratifying on gender (Figure 4 and Table 10 in the Appendix), it seems as though the effect is mostly driven by women. Since the parallel trends assumption cannot be empirically supported in the non-western immigrants sample, the estimates are unreliable.

The estimated effects can at first sight seem a bit large, but there is more than a wage effect induced by TWA employment driving these estimates. To disentangle the wage effect of TWA, we can decompose the earnings in the following simple equation: *earnings* = *P*(*employment*) × *wage* × (*annual*)*hours*. Taking the natural logarithm and differencing gives

$$\Delta \ln earnings = \Delta \ln employment + \Delta \ln wage + \Delta \ln hours$$

where

$$\Delta \ln employment \approx \frac{\Delta employment}{employment}.$$

Using the full sample estimates from 2008 to exemplify: $\Delta \widehat{employment} = 0.056$ (Table 3), $\Delta \ln \widehat{earnings} = 0.177$ (Table 10 in the Appendix) and the average employment rate $\overline{(employment_{2007}} = 0.643)$.

$$0.177 = \frac{0.056}{0.643} + \Delta \ln wage + \Delta \ln hours$$

**Figure 4 cDiD, percentage change in earnings.**
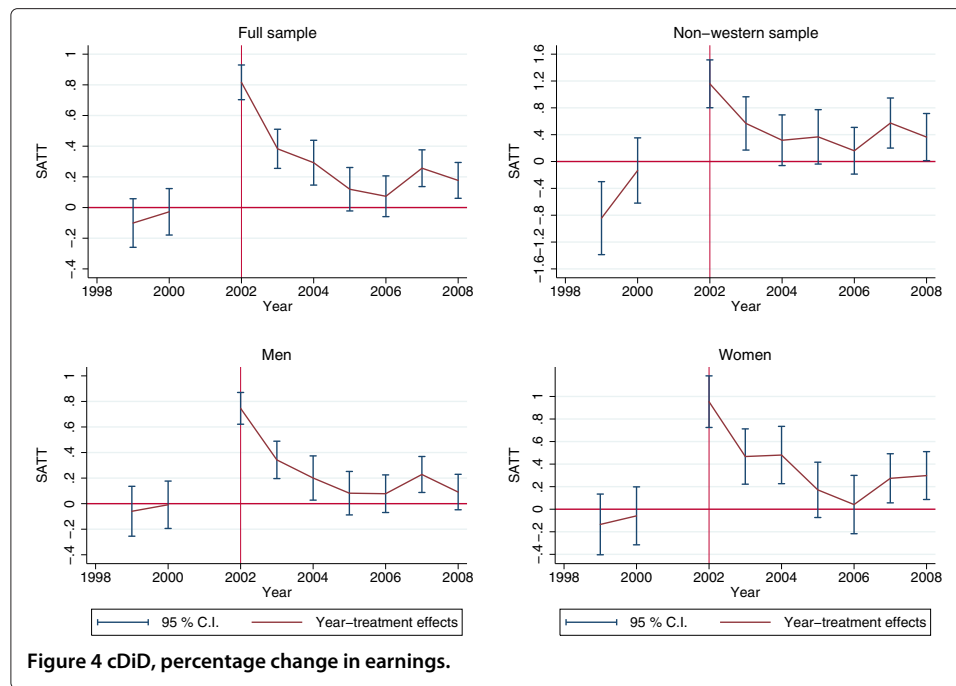
Then subtracting the increased earnings effect stemming from the increased probability of being employed,

$$0.177 - 0.087 = \Delta \ln wage + \Delta \ln hours$$
$$0.09 = \Delta \ln wage + \Delta \ln hours.$$

This means that the treatment group on average increased their annual earnings by 9% in 2008 compared to those not taking treatment when accounting for employment effects. How much of this that is an effect of increased working hours rather than a wage increase is unfortunately not possible to determine with the available data, but presumably the hours effect dominates. One has to keep this equation in mind when interpreting the plots.

### 4.3 Robustness check

To make sure the results do not hinge on the specific timing of treatment (2002) a type of robustness check was performed: All equations were re-estimated with 2004 as treatment year instead of 2002. Neither the iDiD nor the cDiD estimator showed a sensitivity to the treatment timing. The overall pattern was unchanged in all estimations. Two changes are worth noting though: the unemployment estimates for the cDiD shifted down a few points, making the estimates significant at all post-treatment years. Secondly, the earnings estimates shifted upwards, resulting in significantly positive estimates for all panels in the post-treatment years.

By this I conclude that the reported results most likely are robust since the pattern and the estimates barely varied when switching treatment year.

## 5  Summary

This paper investigates how TWAs affect the subsequent labor market outcomes for the unemployed in terms of employment status and income. The outcomes have been re-estimated on non-western immigrants and stratified by gender to control for heterogeneous effects. The selection bias associated with TWA studies has been tackled by individual Differences-in-Differences (iDiD) and conditional Differences-in-Differences (cDiD). The study was concluded with a robustness check where the sensitivity to the treatment year was tested. Both estimators were found to be robust.

The most solid result that can be drawn from the estimations is that joining a TWA (taking treatment) decreases the probability of getting a regular job (TWAs excluded) for years to come in general but not for non-western immigrants. Conversely, the effect on overall employment (TWAs included) has a long-run positive effect when using the matched estimator (only three years with the iDiD). The estimates for unemployment are only marginally significant at times and fluctuate around zero when estimating an iDiD, but when matching the groups we found evidence of a positive transition rate out of unemployment. When stratifying by gender, women showed stronger and more persistent negative regular employment effects even though the other outcomes did not diverge much, suggesting that women tend to stay for longer periods in TWA employment. When turning to the income estimations, the treatment group seems to have gained a bit from the TWA in the long-run. Stratification by gender showed that the result is mainly driven by women.

The evidence provided in this article does not support the stepping stone hypothesis, since regular employment is negatively affected or not affected at all in the medium and long run. It might on the other hand work as a way to escape unemployment that, especially if you are a woman, might benefit your future income (though it is not clear if this is an effect of increased working hours or wages). The TWAs also had a clear long-run effect on employment probabilities. Compared to the unemployment estimates, the employment estimates were larger and more stable over time, suggesting that the TWAs keep individuals from exiting the labor market. This effect, together with the increased long-run earnings effect, puts the TWI in a quite favorable light. The biggest difference between the full sample and the non-western immigrants sample is in the regular employment outcome, where the latter group does not seem to 'get stuck' in the TWA to the same extent as the full sample. However both in the iDiD and in the cDiD the standard errors are quite large and the lack of a negative regular employment effect might be just out of imprecision.

It should here be emphasized that the results are valid for people in unemployment; it is still an open question whether they are valid for a weaker subset (e.g. social assistance recipients) or a stronger subset (e.g. students). External validity is also affected by the matching process, which distorts the sample's characteristics to some extent. In this case, for instance, the results are first and foremost applicable to a younger subset of the population.

### Endnotes

[1] TWA employment as a share of total employment. There is reason to believe that the figure is an understatement since it is survey based and non-repliers are more common among the agencies.

[2] The definition of *non-western immigrants* in (Jahn and Rosholm 2012) is narrower than in the present paper, thus the results are not completely comparable.

[3] (Bennmarker et al. 2009) find significant positive effects for immigrants on time in employment after participating in a private job placement agency compared to the public employment service, indicating that immigrants might benefit more from private options, arguably due to the increased access to norms and networks on the Swedish labor market.

[4] Non-western immigrants: Born in Africa, South America, Asia, the Soviet Union, or other European countries (i.e., excluding the Nordic countries and EU15).

[5] The entire database presently holds annual registers since 1990 and includes all individuals 16 years of age or older that were registered in Sweden as of December 31 for each year.

[6] To increase the computational efficiency, the control group was a 20% random sample, drawn from the population (excluding individuals in TWAs 2001) before any other restriction was put on the group.

[7] Meaning, in effect, that they are at most 63 in the last year, 2008.

[8] Examples of counterfactual outcomes in 2002 could be: taking up studies, getting a regular job, or continuing their unemployment spell, etc.

[9] This data problem is common to all register-based TWA research.

[10] The technique is outlined in (Iacus et al. 2008) and (Blackwell et al. 2009).

[11] Estimations will also be performed on unemployment, overall employment, and annual earnings. But the focus will be on regular employment for simplicity.

[12] This time span might be considered a bit short, but I have also been running a probit on the entire working population and predicting values for 2001: they all support the parallel trends assumption together with the years 1999 and 2000, however since that approach is unorthodox it has not been included and the model does not rely on these results at all.

[13] *Conditional Differences-in-Differences*, i.e., DiD performed on a matched data set.

[14] This matching technique is described in Section 3.2.

[15] Matching on a continuous variable will in effect rule out any matches.

[16] Serves as a measure of labor market attachment.

[17] Four regions of birth dummies in a joint hypothesis test give `P-value > 0.05` and inclusion of them actually increases the standard errors on the treatment variables while leaving the estimates unaffected.

[18] Stratification is also done by gender but no large diverging results were found, the estimates can be found in Table 9 in the Appendix.

[19] Standard errors were clustered on the individual level to account for possible serial correlation.

[20] Having worked at a TWA might signal low ability and regular employers might then shy away from these employees

[21] The default coarsening algorithm by the matching software is the *Scott break method*: $\epsilon_{scott} = 3.5\sqrt{\bar{s}_n^2}n^{-1/3}$, where $n$ denotes the sample size and $\sqrt{\bar{s}_n^2}$ the sample standard deviation (Scott 1992).

[22] Standard errors were clustered on the individual level to account for possible serial correlation.

[23] Register based labor market statistics.

[24] SNI 2007 Standard for Swedish industrial classification 2007.

## Appendix A. Tables

Table 4, 5, 6 show descriptive statistics. Table 7 shows parameter estimates of the multiple OLS regressions. Table 8, 9 shows parameter estimates of the full iDiD regressions

and stratified on gender. Table 10 shows the cDiD parameter estimates from the wage equation.

**Table 4 Summary statistics for 2001, non-western immigrants**

| | Group | | Differences in mean | |
|---|---|---|---|---|
| | Control | Treatment | Coefficient | *t*-statistic |
| **Gender** | | | | |
| Male | 58.5 | 76.1 | 0.177 | (5.73)*** |
| Female | 41.5 | 23.9 | -0.177 | (-5.73)*** |
| **Aggregate days in unemployment 1998–2001** | **447** | **396** | **-50.898** | **(-2.87)*** |
| 0 | 0.8 | 0.5 | | |
| 1–30 | 1.6 | 0.5 | | |
| 31–90 | 6.6 | 9.0 | | |
| 91–182 | 11.5 | 13.4 | | |
| 183–274 | 11.2 | 13.9 | | |
| 275–364 | 10.0 | 13.4 | | |
| 365–730 | 41.4 | 37.8 | | |
| 731–1094 | 16.3 | 11.4 | | |
| 1095–1457 | 0.6 | 0.0 | | |
| **Age groups** | **36** | **32** | **-3.905** | **(-6.04)*** |
| 16–20 | 4.1 | 10.9 | 0.069 | (3.11)*** |
| 21–25 | 10.9 | 19.4 | 0.085 | (3.01)*** |
| 26–30 | 13.2 | 12.4 | -0.008 | (-0.32) |
| 31–35 | 17.3 | 18.9 | 0.016 | (0.56) |
| 36–40 | 19.2 | 15.9 | -0.033 | (-1.26) |
| 41–45 | 17.4 | 14.4 | -0.030 | (-1.18) |
| 46–50 | 11.6 | 5.0 | -0.066 | (-4.14)*** |
| 51–55 | 6.3 | 3.0 | -0.033 | (-2.66)*** |
| **Country of birth** | | | | |
| Other European countries | 32.5 | 31.8 | -0.007 | (-0.21) |
| Africa | 10.4 | 10.4 | 0.001 | (0.03) |
| South America | 6.9 | 8.0 | 0.010 | (0.53) |
| Asia | 49.4 | 48.8 | -0.007 | (-0.18) |
| Soviet Union | 0.7 | 1.0 | 0.003 | (0.36) |
| **Highest education level** | | | | |
| Primary school less than 9 years | 10.7 | 5.5 | -0.052 | (-3.14)*** |
| Primary school 9 (10) years | 15.8 | 15.4 | 0.004 | (-0.16) |
| Upper secondary 2 years or less | 25.6 | 22.9 | -0.027 | (-0.90) |
| Upper secondary 2 years or more | 23.7 | 30.8 | 0.071 | (2.15)** |
| Higher education less than 3 years | 9.4 | 10.4 | 0.010 | (0.46) |
| Higher education 3 years or more | 11.3 | 10.9 | -0.003 | (-0.14) |
| Post-graduate education | 0.6 | 1.0 | 0.004 | (0.60) |
| Unknown | 2.9 | 3.0 | 0.001 | (0.11) |
| **Year of arrival** | | | | |
| 1936–1965 | 0.4 | 0.0 | -0.004 | (-4.59)*** |
| 1966–1972 | 2.8 | 0.5 | -0.023 | (-4.20)*** |
| 1973–1985 | 19.9 | 14.9 | -0.050 | (-1.95)* |
| 1986–1993 | 48.5 | 52.2 | 0.038 | (1.06) |
| 1994–2001 | 28.4 | 32.3 | 0.039 | (1.17) |
| **Resides in a large city** | | | | |
| Stockholm | 20.3 | 32.3 | 0.121 | (3.61)*** |
| Gothenburg | 22.4 | 28.9 | 0.065 | (1.99)** |
| Malmoe | 21.0 | 10.0 | -0.111 | (-5.06)*** |
| Other | 36.3 | 28.9 | -0.075 | (-2.29)** |
| **Number of observations** | 5,427 | 201 | 5,628 | |

Robust standard errors.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 5 Summary statistics for 2001, matched sample**

| | Group | | Differences in mean | |
| --- | --- | --- | --- | --- |
| | Control | Treatment | Coefficient | *t*-statistic |
| **Gender** | | | | |
| Male | 69.9 | 69.9 | 0.000 | (0.00) |
| Female | 30.1 | 30.1 | 0.000 | (0.00) |
| **Aggregate days in unemployment 1998–2001** | **332** | **331** | **-0.790** | **(-0.07)** |
| 0 | 0.4 | 0.4 | 0.4 | |
| 1–30 | 2.9 | 1.6 | 2.7 | |
| 31–90 | 10.1 | 12.3 | 10.4 | |
| 91–182 | 19.1 | 18.1 | 18.9 | |
| 183–274 | 16.5 | 15.4 | 16.4 | |
| 275–364 | 14.0 | 13.9 | 14.0 | |
| 365–730 | 29.1 | 31.3 | 29.4 | |
| 731–1094 | 7.5 | 6.5 | 7.4 | |
| 1095–1457 | 0.4 | 0.4 | 0.4 | |
| **Age groups** | **30** | **30** | **0.043** | **(0.09)** |
| 16–20 | 12.9 | 12.9 | -0.000 | (-0.00) |
| 21–25 | 30.1 | 30.1 | 0.000 | (0.00) |
| 26–30 | 21.0 | 21.0 | -0.000 | (-0.00) |
| 31–35 | 13.6 | 13.6 | -0.000 | (-0.00) |
| 36–40 | 7.2 | 7.2 | -0.000 | (-0.00) |
| 41–45 | 7.4 | 7.4 | -0.000 | (-0.00) |
| 46–50 | 3.8 | 3.8 | -0.000 | (-0.00) |
| 51–55 | 4.0 | 4.0 | -0.000 | (-0.00) |
| **Country of birth** | | | | |
| Sweden | 82.6 | 82.1 | -0.006 | (-0.29) |
| Nordic countries (except Sweden) | 1.1 | 3.3 | 0.0214 | (2.75)*** |
| EU12 | 0.4 | 0.5 | 0.002 | (0.40) |
| Other European countries | 4.9 | 4.0 | -0.009 | (-0.79) |
| Africa | 2.4 | 1.4 | -0.009 | (-1.31) |
| North America | 0.4 | 0.5 | 0.001 | (0.34) |
| South America | 1.1 | 1.1 | -0.000 | (-0.00) |
| Asia | 7.0 | 7.1 | 0.001 | (0.07) |
| Oceania | 0.0 | 0.0 | -0.000 | (-1.00) |
| Soviet Union | 0.1 | 0.0 | -0.001 | (-1.41) |
| **Highest education level** | | | | |
| Primary school less than 9 years | 0.5 | 0.5 | 0.000 | (0.00) |
| Primary school 9 (10) years | 12.7 | 12.7 | 0.000 | (0.00) |
| Upper secondary 2 years or less | 28.6 | 28.6 | 0.000 | (0.00) |
| Upper secondary 2 years or more | 42.0 | 42.0 | 0.000 | (0.00) |
| Higher education less than 3 years | 7.8 | 7.8 | 0.000 | (0.00) |
| Higher education 3 years or more | 8.2 | 8.2 | 0.000 | (0.00) |
| Unknown | 0.2 | 0.2 | 0.000 | (0.00) |
| **Resides in a large city** | | | | |
| Stockholm | 15.0 | 15.0 | 0.000 | (0.00) |
| Gothenburg | 22.1 | 22.1 | 0.000 | (0.00) |
| Malmoe | 10.5 | 10.5 | -0.000 | (-0.00) |
| Other | 52.4 | 52.4 | -0.000 | (-0.00) |
| **Number of observations** | 3,488 | 552 | 4,040 | |

Robust standard errors.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 6 Summary statistics for 2001, non-western immigrants matched sample**

| | Group | | Differences in mean | |
| --- | --- | --- | --- | --- |
| | Control | Treatment | Coefficient | *t*-statistic |
| **Gender** | | | | |
| Male | 82.7 | 82.7 | -0.000 | (-0.00) |
| Female | 17.3 | 17.3 | 0.000 | (0.00) |
| **Aggregate days in unemployment 1998–2001** | **405** | **385** | **-19.30** | **(-0.60)** |
| 0 | 0.2 | 1.0 | | |
| 1–30 | 2.2 | 0.0 | | |
| 31–90 | 6.2 | 8.2 | | |
| 91–182 | 16.9 | 12.2 | | |
| 183–274 | 14.1 | 18.4 | | |
| 275–364 | 10.0 | 11.2 | | |
| 365–730 | 38.2 | 40.8 | | |
| 731–1094 | 11.3 | 8.2 | | |
| 1095–1457 | 1.0 | 0.0 | | |
| **Age groups** | **31** | **31** | **0.270** | **(0.27)** |
| 16–20 | 9.2 | 9.2 | -0.000 | (-0.00) |
| 21–25 | 24.5 | 24.5 | 0.000 | (0.00) |
| 26–30 | 13.3 | 13.3 | -0.000 | (-0.00) |
| 31–35 | 25.5 | 25.5 | 0.000 | (0.00) |
| 36–40 | 11.2 | 11.2 | 0.000 | (0.00) |
| 41–45 | 14.3 | 14.3 | 0.000 | (0.00) |
| 46–50 | 2.0 | 2.0 | 0.000 | (0.00) |
| **Country of birth** | | | | |
| Other European countries | 27.2 | 32.7 | 0.055 | (0.90) |
| Africa | 18.1 | 13.3 | -0.048 | (-1.23) |
| South America | 6.1 | 7.1 | 0.011 | (0.001) |
| Asia | 46.9 | 46.9 | 0.001 | (0.01) |
| Soviet Union | 1.8 | 0.0 | -0.018 | (-1.54) |
| **Highest education level** | | | | |
| Primary school less than 9 years | 4.1 | 4.1 | -0.000 | (-0.00) |
| Primary school 9 (10) years | 16.3 | 16.3 | 0.000 | (0.00) |
| Upper secondary 2 years or less | 22.4 | 22.4 | -0.000 | (-0.00) |
| Upper secondary 2 years or more | 36.7 | 36.7 | -0.000 | (-0.00) |
| Higher education less than 3 years | 8.2 | 8.2 | 0.000 | (0.00) |
| Higher education 3 years or more | 9.2 | 9.2 | 0.000 | (0.00) |
| Unknown | 3.1 | 3.1 | 0.000 | (0.00) |
| **Year of arrival** | | | | |
| 1936–1965 | 0.4 | 0.0 | -0.004 | (-1.20) |
| 1966–1972 | 1.6 | 1.0 | -0.005 | (-0.37) |
| 1973–1985 | 18.5 | 17.3 | -0.012 | (-0.24) |
| 1986–1993 | 51.0 | 42.9 | -0.082 | (-1.27) |
| 1994–2001 | 28.4 | 38.8 | 0.103 | (1.66)* |
| **Resides in a large city** | | | | |
| Stockholm | 34.7 | 34.7 | 0.00 | (0.00) |
| Gothenburg | 27.8 | 27.8 | 0.00 | (0.00) |
| Malmoe | 6.9 | 6.9 | -0.00 | (-0.00) |
| Other | 30.6 | 30.6 | 0.00 | (0.00) |
| **Number of observations** | 275 | 98 | 373 | |

Robust standard errors.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 7 Multiple OLS regressions**

| | Independent variable: TWA | | | | | |
| | Full sample | | | Non-western immigrants | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| OUTCOME YEAR | P(Regular emp.) | P(Unemployment) | P(Employment) | P(Regular emp.) | P(Unemployment) | P(Employment) |
|---|---|---|---|---|---|---|
| 2003 | -0.128*** | -0.063*** | 0.217*** | -0.155*** | -0.013 | 0.148*** |
| | (0.016) | (0.011) | (0.015) | (0.035) | (0.027) | (0.034) |
| 2004 | -0.094*** | -0.018 | 0.121*** | -0.105*** | 0.026 | 0.073** |
| | (0.016) | (0.011) | (0.015) | (0.035) | (0.028) | (0.034) |
| 2005 | -0.051*** | -0.010 | 0.090*** | -0.076** | 0.004 | 0.061* |
| | (0.016) | (0.011) | (0.015) | (0.036) | (0.027) | (0.034) |
| 2006 | -0.026* | -0.072*** | 0.095*** | -0.035 | -0.052* | 0.062* |
| | (0.016) | (0.013) | (0.014) | (0.036) | (0.031) | (0.033) |
| 2007 | -0.005 | -0.078*** | 0.104*** | -0.021 | -0.080*** | 0.083*** |
| | (0.015) | (0.012) | (0.013) | (0.035) | (0.030) | (0.032) |
| 2008 | 0.007 | -0.013 | 0.089*** | -0.039 | 0.010 | 0.051 |
| | (0.015) | (0.009) | (0.013) | (0.036) | (0.024) | (0.033) |
| Observations | 26,111 | 26,111 | 26,111 | 5,628 | 5,628 | 5,628 |
| Control variables | YES | YES | YES | YES | YES | YES |

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

*Control variables:* Number of children over six different age groups, and dummies for: gender, world region of birth, parents born abroad, labor market region, educational level, marital status, year of arrival, age groups, social welfare benefit received current year. For non-western immigrants the dummies for parents born abroad and some of the world region of birth dummies were dropped due to collinearity.

**Table 8 iDiD coefficient estimates**

| Variables | (1) Regular employment Full sample | (2) Unemployment Full sample | (3) Employment Full sample | (4) Regular employment non-western immigrants | (5) Unemployment non-western immigrants | (6) Employment non-western immigrants |
|---|---|---|---|---|---|---|
| Treatment × 1999 | -0.009 | 0.017 | -0.002 | 0.010 | 0.050 | 0.007 |
|  | (0.017) | (0.016) | (0.016) | (0.032) | (0.040) | (0.032) |
| Treatment × 2000 | -0.006 | -0.003 | 0.028 | 0.008 | 0.057 | 0.062 |
|  | (0.021) | (0.018) | (0.021) | (0.037) | (0.045) | (0.040) |
| Treatment × 2002 | -0.461*** | -0.153*** | 0.529*** | -0.362*** | -0.169*** | 0.623*** |
|  | (0.016) | (0.013) | (0.016) | (0.030) | (0.030) | (0.031) |
| Treatment × 2003 | -0.184*** | -0.008 | 0.156*** | -0.144*** | 0.033 | 0.152*** |
|  | (0.022) | (0.016) | (0.022) | (0.043) | (0.038) | (0.046) |
| Treatment × 2004 | -0.156*** | 0.045*** | 0.053** | -0.086** | 0.077** | 0.082* |
|  | (0.022) | (0.017) | (0.022) | (0.044) | (0.039) | (0.045) |
| Treatment × 2005 | -0.114*** | 0.055*** | 0.021 | -0.070 | 0.063* | 0.057 |
|  | (0.022) | (0.016) | (0.022) | (0.044) | (0.037) | (0.044) |
| Treatment × 2006 | -0.087*** | -0.037** | 0.028 | -0.036 | -0.016 | 0.054 |
|  | (0.022) | (0.017) | (0.021) | (0.048) | (0.042) | (0.048) |
| Treatment × 2007 | -0.060*** | -0.047*** | 0.043** | -0.013 | -0.053 | 0.084* |
|  | (0.022) | (0.018) | (0.020) | (0.046) | (0.040) | (0.045) |
| Treatment × 2008 | -0.046** | 0.042*** | 0.030 | -0.027 | 0.052 | 0.053 |
|  | (0.022) | (0.015) | (0.021) | (0.048) | (0.037) | (0.047) |
| Observations | 261,110 | 261,110 | 261,110 | 56,280 | 56,280 | 56,280 |
| R-squared | 0.424 | 0.264 | 0.431 | 0.422 | 0.248 | 0.428 |
| Control variables | YES | YES | YES | YES | YES | YES |

Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
*Control variables ($X_{i,t}$):* Dummies for labor market region, education level, educational orientation, marital status, age groups, social welfare benefit received current year.
Standard errors were clustered on individual level to account for possible serial correlation.

**Table 9 iDiD, stratified on gender**

| Variables | (1) Regular employment Women | (2) Unemployment Women | (3) Employment Women | (4) Regular employment Men | (5) Unemployment Men | (6) Employment Men |
|---|---|---|---|---|---|---|
| Treatment × 1999 | -0.031 | 0.031 | -0.024 | 0.004 | 0.010 | 0.010 |
| | (0.028) | (0.026) | (0.028) | (0.020) | (0.021) | (0.020) |
| Treatment × 2000 | -0.033 | -0.005 | -0.002 | 0.006 | -0.002 | 0.042 |
| | (0.034) | (0.027) | (0.033) | (0.025) | (0.024) | (0.026) |
| Treatment × 2002 | -0.502*** | -0.113*** | 0.488*** | -0.437*** | -0.174*** | 0.554*** |
| | (0.028) | (0.019) | (0.028) | (0.020) | (0.017) | (0.020) |
| Treatment × 2003 | -0.200*** | -0.002 | 0.166*** | -0.176*** | -0.010 | 0.148*** |
| | (0.038) | (0.024) | (0.035) | (0.027) | (0.021) | (0.027) |
| Treatment × 2004 | -0.177*** | 0.060** | 0.060* | -0.145*** | 0.039* | 0.048* |
| | (0.037) | (0.025) | (0.036) | (0.027) | (0.022) | (0.027) |
| Treatment × 2005 | -0.124*** | 0.059** | 0.022 | -0.109*** | 0.056*** | 0.019 |
| | (0.039) | (0.026) | (0.037) | (0.027) | (0.020) | (0.027) |
| Treatment × 2006 | -0.110*** | -0.055** | 0.030 | -0.075*** | -0.023 | 0.027 |
| | (0.038) | (0.027) | (0.035) | (0.027) | (0.022) | (0.026) |
| Treatment × 2007 | -0.068* | -0.065** | 0.059* | -0.056** | -0.032 | 0.034 |
| | (0.037) | (0.027) | (0.033) | (0.027) | (0.023) | (0.026) |
| Treatment × 2008 | -0.021 | 0.016 | 0.061* | -0.058** | 0.060*** | 0.015 |
| | (0.036) | (0.022) | (0.033) | (0.027) | (0.020) | (0.026) |
| Observations | 111,591 | 111,591 | 111,591 | 149,519 | 149,519 | 149,519 |
| R-squared | 0.426 | 0.260 | 0.433 | 0.426 | 0.267 | 0.433 |
| Control variables | YES | YES | YES | YES | YES | YES |

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

*Control variables (X$_{i,t}$):* Dummies for labor market region, education orientation level, marital status, age groups and social welfare benefit received current year.

Standard errors were clustered on individual level to account for possible serial correlation.

**Table 10 cDiD coefficient estimates corresponding to Figure 4**

| | Percentage change in earnings | | | |
| | (1) | (2) | (3) | (4) |
| Variables | Full sample | Women | Men | Non-western immigrants |
|---|---|---|---|---|
| Treatment × 1999 | -0.101 | -0.135 | -0.059 | -0.844*** |
| | (0.081) | (0.137) | (0.100) | (0.277) |
| Treatment × 2000 | -0.028 | -0.059 | -0.009 | -0.133 |
| | (0.077) | (0.131) | (0.095) | (0.248) |
| Treatment × 2002 | 0.816*** | 0.954*** | 0.746*** | 1.158*** |
| | (0.058) | (0.117) | (0.063) | (0.182) |
| Treatment × 2003 | 0.383*** | 0.467*** | 0.342*** | 0.568*** |
| | (0.065) | (0.125) | (0.075) | (0.203) |
| Treatment × 2004 | 0.292*** | 0.481*** | 0.201** | 0.317 |
| | (0.075) | (0.130) | (0.088) | (0.193) |
| Treatment × 2005 | 0.119* | 0.172 | 0.082 | 0.368* |
| | (0.072) | (0.125) | (0.087) | (0.207) |
| Treatment × 2006 | 0.074 | 0.041 | 0.078 | 0.161 |
| | (0.068) | (0.132) | (0.075) | (0.177) |
| Treatment × 2007 | 0.256*** | 0.274** | 0.228*** | 0.574*** |
| | (0.061) | (0.111) | (0.072) | (0.190) |
| Treatment × 2008 | 0.177*** | 0.298*** | 0.091 | 0.364** |
| | (0.060) | (0.108) | (0.071) | (0.179) |
| Observations | 28,891 | 7,431 | 21,460 | 2,443 |
| R-squared | 0.219 | 0.193 | 0.256 | 0.274 |
| Control variables | YES | YES | YES | YES |

Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
*Control variables ($X_{i,t}$):* Age, number of kids over age groups and dummies for: region of birth, year of arrival, educational orientation, social welfare benefit received current year and level, and a linear time trend.
Standard errors were clustered on individual level to account for possible serial correlation.

## Appendix B. Figures

Figure 5, 6 show the effect of matching on the selected variables *aggregate days in unemployment* and *age*. Figure 7, 8 plots the cDiD estimated coefficients of the employment status stratified by gender, and the iDiD estimates of the wage equation. The section ends with a description of outcome variables.
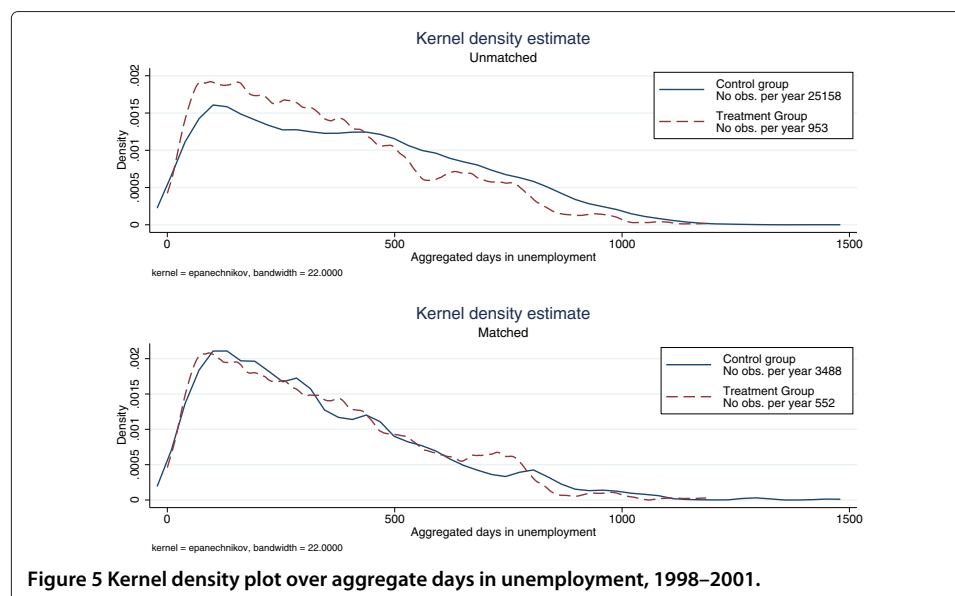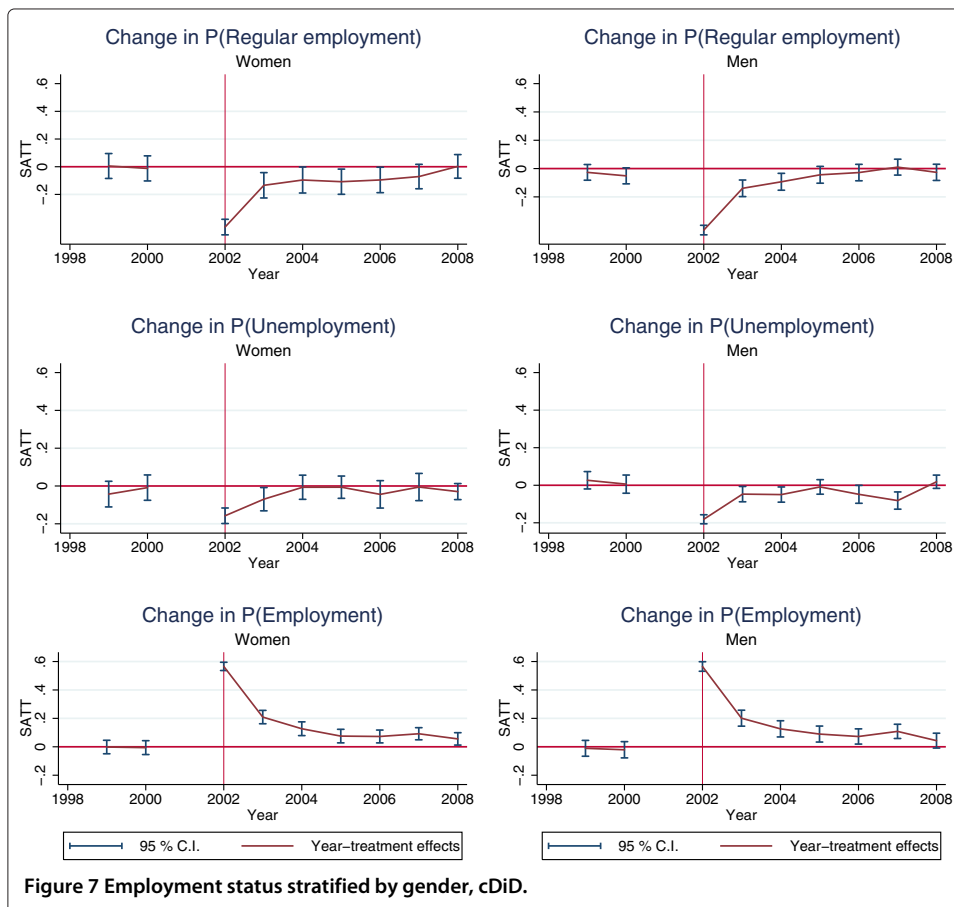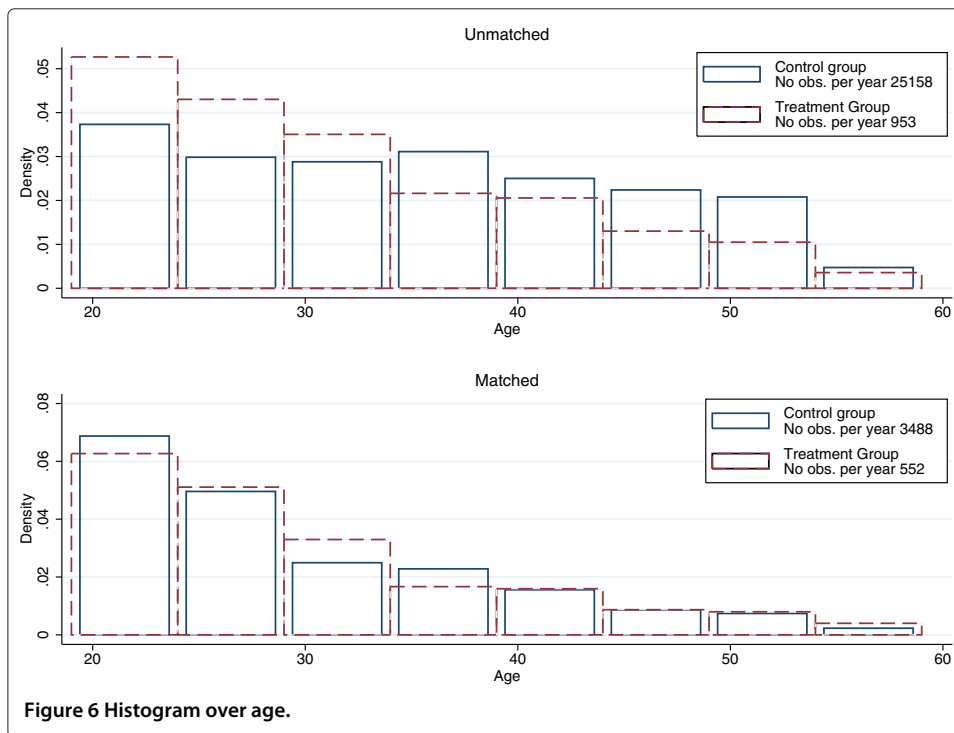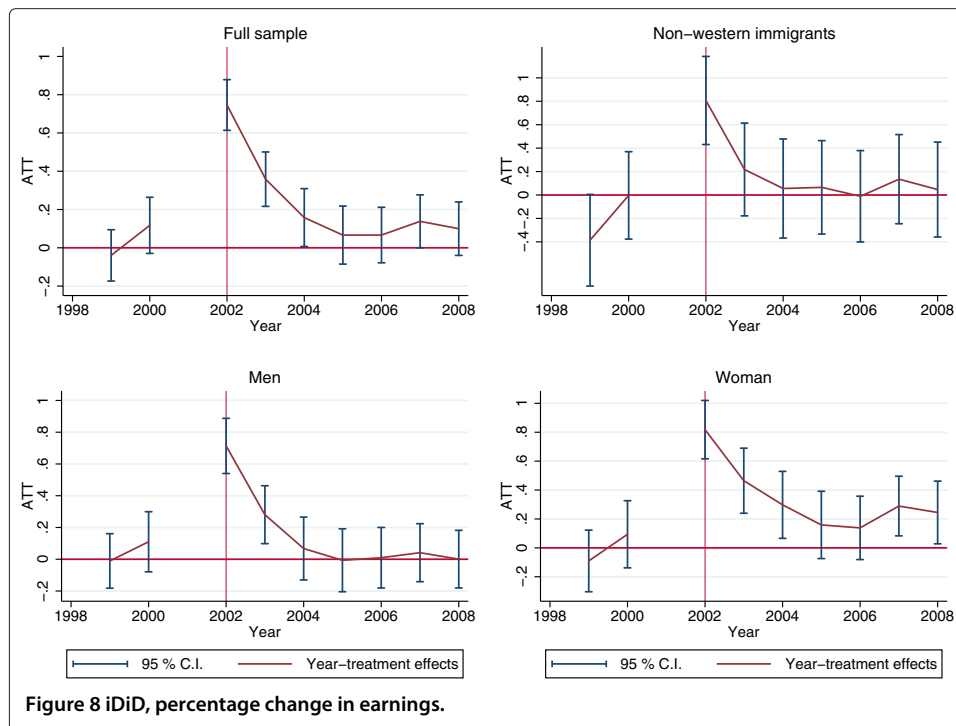


**Figure 5 Kernel density plot over aggregate days in unemployment, 1998–2001.**

**Figure 6 Histogram over age.**



**Figure 7 Employment status stratified by gender, cDiD.**

**Figure 8 iDiD, percentage change in earnings.**

## Appendix C. Description of outcome variables

**Regular employment** defined as all working in November in the employment register. The official definition of being employed in RAMS[23] closely tries to follow the ILO definition, meaning that if any income-generating labor has been performed during the week of measurement that is regarded as being employed (includes income from own business). In addition to this, all TWA workers are excluded to define *regular* employment.

**Unemployment** defined as searching for a job at the unemployment offices at the end of November including those registered in a labor market program.

**TWA employment** defined as working at a TWA in November. The TWA definition is number 78 in the SNI 2007[24].

**Total employment** defined as regular employment but not excluding TWA employment.

**Earnings** defined as total gross annual reported income from work, recorded by the tax office.

**References**
Amuedo-Dorantes C, Malo MA, Munoz-Bullon F (2008) The role of temporary help agency employment on temp-to-perm transitions. J Labor Res 29(2): 138–161

Andersson F, Holzer HJ, Lane J (2007) Temporary help agencies and the advancement prospects of low earners. NBER Working Papers 13434, National Bureau of Economic Research, inc. http://www.nber.org/papers/w13434

Andersson P, Wadensjö E (2004) Temporary employment agencies: A route for immigrants to enter the labour market? IZA Discussion Papers 1090, Institute for the Study of Labor (IZA), Bonn, Germany

Andersson Joona P, Wadensjö E (2010) Bemanningsbranschen 1998-2005: En bransch i förändring? Working Paper Series 6/2010, Swedish Institute for Social Research (SOFI), Stockholm University, Stockholm, Sweden

Angrist J, Pischke J (2009) Mostly Harmless Econometrics. Princeton Univ. Press, Princeton, NJ

Autor DH, Houseman SN (2010) Do temporary-help jobs improve labor market outcomes for low-skilled workers? Evidence from 'Work First'. Am Econ J: Appl Econ 2(3): 96–128

Bemanningsföretagen (2011) Antal anställda och penetrationsgrad i bemanningsbranschen 2011. Retrieved from http://www.bemanningsforetagen.se/MediaBinaryLoader.axd?MediaArchive_FileID=2d854ee3-a01e-49fb-8aad-3f9127a4927b&FileName=Anst_och_penetrationsgrad_2011_A.pdf

Bennmarker H, Grönqvist E, Öckert B (2009) Effects of outsourcing employment services: Evidence from a randomized experiment. Working Paper Series 2009:23, The Institute for Labour Market Policy Evalutation. IFAU, Uppsala, Sweden

Blackwell M, Iacus S, King G, Porro G (2009) Cem: Coarsened exact matching in stata. Stata J 9(4): 524–546

Cobb-Clark DA, Crossley T (2003) Econometrics for evaluations: An introduction to recent developments. Econ Rec 79(247): 491–511

Dehejia RH, Wahba S (1999) Causal effects in non-experimental studies: Reevaluating the evaluation of training programs. J Am Stat Assoc 94(448): 1053–1062

Fredriksson P, Johansson P (2003) Program evaluation and random program starts. CESifo Working Paper Series 844, CESifo Group Munich

Garcia-Perez JI, Munoz-Bullon F (2005) Temporary help agencies and occupational mobility. Oxford Bull Econ Stat 67(2): 163–180

Iacus S, King G, Porro G (2008) Matching for causal inference without balance checking. UNIMI–Research Papers in Economics, Business, and Statistics UNIMI-1073, Universitá degli Studi di Milano

Ichino A, Mealli F, Nannicini T (2008) From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? J Appl Econ 23(3): 305–327

Jahn EJ (2008) Reassessing the wage penalty for temps in germany. IZA Discussion Papers 3663, Institute for the Study of Labor (IZA), Bonn, Germany

Jahn EJ, Rosholm M (2012) Is temporary agency employment a stepping stone for immigrants? IZA Discussion Papers 6405, Institute for the Study of Labor (IZA)

Kvasnicka M (2008) Does temporary help work provide a stepping stone to regular employment? NBER Working Papers 13843, National Bureau of Economic Research Inc.

LaLonde RJ (1986) Evaluating the econometric evaluations of training programs with experimental data. Am Econ Rev 76(4): 604–620

Lane J, Mikelson KS, Sharkey P, Wissoker D (2003) Pathways to work for low-income workers: The effect of work in the temporary help industry. J Policy Anal. Manag 22(4): 581–598

Neugart M, Storrie D (2006) The emergence of temporary work agencies. Oxford Econ Papers 58(1): 137–156

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70(1): 41–55

Scott DW (1992) Multivariate Density Estimation: Theory, Practice, and Visualization. Vol. 8. John Wiley, New York

Smith JA, Todd PE (2005) Does matching overcome lalonde's critique of nonexperimental estimators? J Econ 125(1–2): 305–353

Summerfield F (2009) Help or hindrance: Temporary help agencies and the united states transitory workforce. Working Papers 0911, University of Guelph, Department of Economics.