Working Time Reductions at the End of the Career.

Do They Prolong the Time Spent in Employment?

INTERNET APPENDIX

This Internet Appendix contains:

(i) The sample selection mechanism

(ii) The details of the empirical strategy;

(iii) The cost-benefit analysis of the Time Credit (TC) scheme for the treated population entering the treatment in 2003 or 2004;

Other estimation results are available from the authors upon request.

1. Sample selection mechanism

In order to enhance the comparability of treated and control groups, we impose that members of both groups should satisfy criteria slightly stricter than those that determine eligibility to TC (cf. Section 3 of the paper):

1. Being employed in a firm with at least 20 employees at the end of 2002 (2003);
2. Have at least 5 years of tenure in the same firm at the end of 2002 (2003);
3. Have at least 20 years of private sector labour market experience at the end of 2002 (2003);
4. Being full-time employed in all four quarters of 2002 (2003);
5. Being employed in the private sector at the end of 2003Q3 (2004Q3);
6. Not being on sick leave at the end of 2003Q3 (2004Q3).

The fifth selection criterion is imposed at the end of each assignment period into treatment, because it is automatically satisfied for the members of the treatment groups, so that it is natural to impose it on the members of the control groups as well. The other criteria are slightly more restrictive than the TC eligibility conditions, so that a few treated individuals are eliminated from our initial selection. We are slightly more restrictive for the following reasons:

(i) Eligibility conditions 1-4 are uniformly imposed at the end of the year preceding the *contractual* start of the TC so that the same conditions apply to *all* treated and control units;

(ii) Because the data contain only basic information about firm characteristics, we aim at restricting the analysis to sufficiently large firms in which the use of TC does not require the consent of the employer (cf. Section 3 of the paper). If no consent is required, it is less likely that the use of TC is selective in firm characteristics, which therefore enhances the internal validity of the evaluation. However, according to the rules, no consent is required in firms with more than 10 employees. Nevertheless, we include only firms that employ at least 20 workers. This is because the available data on firm size are grouped in intervals that do not allow identifying firms with strictly more than 10 employees;

(iii) While regulations do not impose that the labour market experience should be accumulated in the private sector, we do because we do not have information on *early* experience outside the private sector;

(iv) We impose full-time employment in the last year prior to *contractual* assignment into treatment, while for the 50% TC regime the requirement is only to have worked at least 75% of a full-time job. This is done to consider only individuals who are eligible to both regimes;

(v) Finally, we impose the last condition on sickness in 2003Q3 (2004Q3) because we want to contrast the impact of the benchmark outcome, i.e. survival in employment, to a more restrictive variant that considers survival in employment *without being on sick leave*. If we would not impose this, some of the selected individuals would not be in the risk set of this second outcome at the start of the evaluation period. Imposing this condition only very marginally affects the sample selection. Finally, note that we do not impose the age condition, because our sample only contains individuals than 50.

Our treated sample contains 1,227 men and 762 women, representing 5,124 individuals in the Belgian population (weighted by $W_{cr,i}$). If we retain all individuals employed in 2002q3 (2003q3) and do not impose the eligibility conditions, the comparison group contains 142,154 men and 83,983 women. Adding the eligibility conditions, the size of the control group shrinks to 29,791 men and 9,658 women. Note that the control units in the two years of analysis partly consist of same individuals, while treated units are always different.

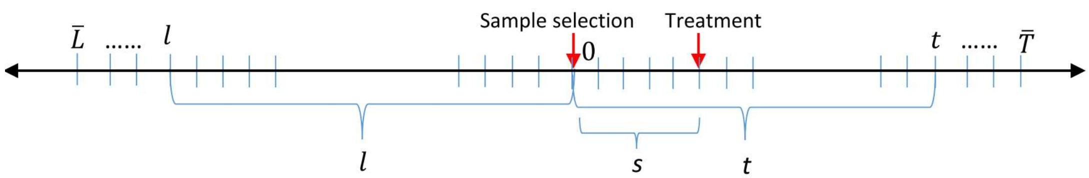
2. The Empirical Strategy

2.1. Notation

Let $t \in \{1,2, \ldots \bar{T}\}$ denote the number of years since sample selection and $l \in \{1, 2, \ldots, \bar{L}\}$ the elapsed number of years in employment at this start. $\bar{T}$ and $\bar{L}$ are the maximum number of years in employment respectively, after and before selection. In the data $l \geq 5$, because this is an eligibility condition for the TC and a sample selection criterion (Section 3). The random time since sample selection until the start of the treatment, i.e. entry in TC, is denoted by S and its realization by s, where $s \in \{0,1, \ldots, \bar{S}\}$ and $\bar{S} \leq \bar{T}$. $Y_{l+t}(s)$ is equal to one in case employment is left in year $l + t$ and treatment started in year $l + s$, and zero otherwise. $Y_{l+t}(\infty)$ denotes the potential outcome in year

$l + t$ if never treated and Y_{l+t} the observed outcome. $\bar{Y}_{l+t}(s) \equiv \{Y_1(s), Y_2(s), \dots, Y_{l+t}(s)\}$ and $\bar{Y}_{l+t} \equiv \{Y_1, Y_2, \dots, Y_{l+t}\}$ denote, respectively, the sequence of potential and of observed outcomes. Figure 1 provides a graphical representation of the introduced notation.

Figure 1: Graphical Representation of the Notation



We aim at identifying the average treatment effect on the treated (ATT) of treatment starting s years after sample selection against the counterfactual of never being treated on the residual survival in employment until year $t > s$, given survival in employment until sample selection:

$$\forall t > s: ATT_t(s) \equiv E_L\{E[\bar{Y}_{L+t}(s) = 0 | S = s, \bar{Y}_{L+s}(s) = 0]$$
$$- E[\bar{Y}_{L+t}(\infty) = 0 | S = s, \bar{Y}_{L+s}(s) = 0] | L \geq 5\} \quad (1)$$

This extends the ATT as parameter of interest to evaluation of a stock sample. Since in a stock sample individuals may have a different elapsed duration, the conditional expectation is taken over these elapsed durations, conditional on being employed for at least 5 years to take into account that one needs at least 5 years of tenure to be eligible for TC. Observe that $E[\bar{Y}_{L+t}(.) = 0 | S = ., \bar{Y}_{L+s}(.) = 0] = Pr[T > L + t | S = ., T > L + s]$ holds, i.e. the conditional probability of surviving $L + t$ years in employment given survival until $L + s$. In case $L = 0$, Equation (1) reduces to the corresponding expression in Vikström (2014) for a flow sample.

2.2. Identification

In order to identify $ATT_t(s)$ we use two identifying assumptions: CIA and no anticipation (NAA). These assumptions can be formalized as follows:

$$\boldsymbol{CIA}\ \forall l > 5, \forall s, \forall t > s: S \perp Y_{l+t}(s) \mid X \quad (2)$$

and

$$\boldsymbol{NAA}\ \forall l > 5, \forall t < min(s', s''): Pr(Y_{l+t}(s') = 1) = Pr(Y_{l+t}(s'') = 1), \quad (3)$$

The latter condition means that individuals do not alter their behaviour in response to a future assignment to the treatment. Based on these assumptions Fredriksson and Johansson (2008), Crépon et al. (2009) and Vikström (2014) prove that for $l = 0$ $ATT_t(s)$ can be identified by successively using the not yet treated at $l + t$ to estimate the exit rate under no treatment at $l + t$ for those treated at $l + s$. Vikström (2014) generalizes by explicitly allowing for selectivity on observables in subsequent assignments into treatment. We follow his approach. Because the identification proof is not affected for different values of l, we refer the reader to Vikström (2014).

2.3. Estimation and Inference

Vikström derives the Inverse Probability Weighting (IPW) estimator, introduced by Horvitz & Thompson (1952) and Hirano et al. (2003), to estimate the $ATT_t(s)$ defined in (2). We follow this approach for the following reasons: (i) Busso et al. (2014) find in their Monte Carlo simulation that the normalized IPW estimator is one of the best performing matching estimators in the presence of good overlap. Other Monte Carlo simulations of Huber et al. (2013) and Frölich et al. (2015) confirm the good performance of the IPW estimator, although it does not outperform other Propensity Score-based and non-parametric estimators; (ii) It is easy to integrate the endogenous sampling weights. This merely requires to include an additional weight in the estimation; (iii) Compared to other matching estimators, the IPW estimator is simple and computationally fast.

We provide the most general estimator that does not only allow to take into account selective (on observables) right censoring as a consequence of not yet treated individuals getting treated (Vikström, 2014), but also more general forms of selective right censoring that may involve both treated and not yet treated individuals. For instance, we consider a competing risk framework, with estimations of the treatment effect on different exit destinations when terminating employment. We distinguish between exits to bridge pensions, statutory *early* retirement and "other" exit routes.

To write down the estimator, we denote the random censoring duration since sample selection for individual i by C_i. Generalising Vikström's formula (see his Appendix A.3) for the endogenous sampling weights $W_{cr,i}$ and taking the elapsed employment duration l_i into account, we obtain:

$$\widehat{\mathbf{ATT}}_{\mathbf{t}}(\mathbf{s}) = \prod_{\mathbf{k=s+1}}^{\mathbf{t}} \left[\mathbf{1} - \frac{\sum_{\mathbf{i}} \mathbf{W}_{\mathbf{cr,i}} * \mathbf{W}^{\mathbf{C}}_{\mathbf{l,k(s),i}}(\mathbf{s})\, \mathbf{Y}_{\mathbf{k,i}}\, \mathbf{1}\left(\overline{\mathbf{Y}}_{\mathbf{l+k-1,i}} = \mathbf{0}\right)\mathbf{1}(\mathbf{S_i} = \mathbf{s})\mathbf{1}(\mathbf{C_i} > \mathbf{s})}{\sum_{\mathbf{i}} \mathbf{W}_{\mathbf{cr,i}} * \mathbf{W}^{\mathbf{C}}_{\mathbf{l,k(s),i}}(\mathbf{s})\, \mathbf{1}\left(\overline{\mathbf{Y}}_{\mathbf{l+k-1,i}} = \mathbf{0}\right)\mathbf{1}(\mathbf{S_i} = \mathbf{s})\mathbf{1}(\mathbf{C_i} > \mathbf{s})} \right]$$
$$- \prod_{\mathbf{k=s+1}}^{\mathbf{t}} \left[\mathbf{1} - \frac{\sum_{\mathbf{i}} \mathbf{W}_{\mathbf{cr,i}} * \mathbf{W}^{\mathbf{C}}_{\mathbf{l,k(\infty),i}}(\mathbf{s})\, \mathbf{Y}_{\mathbf{k,i}}\, \mathbf{1}\left(\overline{\mathbf{Y}}_{\mathbf{l+k-1,i}} = \mathbf{0}\right)\mathbf{1}(\mathbf{S_i} \geq \mathbf{k})\mathbf{1}(\mathbf{C_i} \geq \mathbf{k})}{\sum_{\mathbf{i}} \mathbf{W}_{\mathbf{cr,i}} * \mathbf{W}^{\mathbf{C}}_{\mathbf{l,k(\infty),i}}(\mathbf{s})\, \mathbf{1}\left(\overline{\mathbf{Y}}_{\mathbf{l+k-1,i}} = \mathbf{0}\right)\mathbf{1}(\mathbf{S_i} \geq \mathbf{k})\mathbf{1}(\mathbf{C_i} \geq \mathbf{k})} \right] \qquad (4)$$

where

$$W^C_{l,k(s),i}(s) = \frac{1}{\prod_{m=s+1}^{k}[1 - c_m(X_i, l_i)]}$$
$$W^C_{l,k(\infty),i}(s) = \frac{p_s(X_i, l_i)}{1 - p_s(X_i, l_i)} \frac{1}{\prod_{m=s+1}^{k}[1 - p_m(X_i, l_i)][1 - c_m(X_i, l_i)]}$$
$$p_t(X_i, l_i) = \Pr(S_i = t | X_i, S_i \geq l_i + t, \bar{Y}_{l+t-1,i} = 0)$$

$$c_t(X_i, l_i) = \Pr(C_i = t | X_i, S_i \geq l_i + t, \bar{Y}_{l+t-1,i} = 0)$$

where X_i denotes the vector of predetermined explanatory variables, $W^C_{l,k(s),i}(s)$ and $W^C_{l,k(\infty),i}(s)$ are the IPW weights in year $l_i + k$ for individual i treated in year $l_i + s$ and not yet treated in year $l_i + k$, respectively. $p_t(X_i, l_i)$ and $c_t(X_i, l_i)$ denote the conditional probability of entering the treatment, respectively censoring state after $l_i + t$ years conditional on still being employed in $l_i + t - 1$. In other words, they represent the discrete hazard of entering treatment, respectively censoring in year $l_i + t$.

To clarify the intuition of the estimator defined in Equation (4) consider first the case without right censoring, i.e. $C_i = \infty$, $W^C_{l,k(s),i}(s) = 1$ and $W^C_{l,k(\infty),i}(s) = \frac{p_s(X_i,l_i)}{1-p_s(X_i,l_i)} \frac{1}{\prod_{m=s+1}^{k}[1-p_m(X_i,l_i)]}$. Apart from the weights, the first sequence of products in (5) is the standard Kaplan Meier survivor estimator for the treated group. This represents the conditional survival rate in employment until year $l_i + t$, conditional on treatment and survival in employment until year $l_i + s$, i.e. the product of one minus the discrete hazards from employment between $l_i + s + 1$ and $l_i + t$. The second sequence of products is a similar Kaplan Meier estimator for the control group (or not yet treated individuals), which estimates the survival rate of the treated in the counterfactual of no treatment. In order to make these control units comparable to the treated they are reweighted using the standard IPW weights $\frac{p_s(X_i,l_i)}{1-p_s(X_i,l_i)}$ in a static evaluation approach, where $p_s(X_i, l_i)$ is the estimated Propensity Score (PS) for an individual treated in year $l_i + s$. However, to take into account that not yet treated individuals gradually become treated, we must consider that this may change the composition of the control group over time. Hence, Vikström (2014) shows that we must in addition weigh the control units by $\frac{1}{\prod_{m=s+1}^{k}[1-p_m(X_i,l_i)]}$, i.e. by the inverse of the probability of *not yet* being treated in each period between $l_i + s + 1$ and $l_i + t$.

If individuals are right censored before exiting to the destination of interest and this is selective (i.e. depends on X), then this may similarly gradually change the composition of now not only the control group, but also of the treatment group over time. We therefore need to weigh both treated and control samples by $\frac{1}{\prod_{m=s+1}^{k}[1-c_m(X_i,l_i)]}$, i.e. the probability of not yet being right censored in each period between $l_i + s + 1$ and $l_i + t$.

In contrast to Vikström (2014), the discrete hazards to treatment and censoring depend on the elapsed employment duration l_i at sample selection. Observe that we can only proxy for this elapsed employment duration, because prior to 1998 we only have annual (instead of quarterly) information on private sector employment and no information on self-employment, neither on employment as civil servant. Given that we selected individuals with at least 5 years of tenure and 20 years of experience in the private sector, we believe that the bias induced by using this proxy is negligible.

We estimate separate ATTs for individuals entering treatment in 2003 and 2004. Subsequently, we pool, as Vikström, these analyses to have more precise estimates. This is done by averaging the estimated ATTs in each survival year, using the endogenous sampling weights to take into account the size of the two different treated groups in the population.

$$\widehat{ATT_t} = \sum_s \frac{n_s}{\sum_s n_s} \widehat{ATT_{s+t}}(s) \quad (5)$$

where $n_s \equiv \sum_i W_{cr,i} * 1(S_i = s)$.

As a lack of overlap of the PS can bias the estimator and increase the variance (Lechner and Strittmatter, 2014), we trim treated units who, due to their very high PS, do not have a correspondent control unit. In particular, we remove treated units with a PS above the 99.9 percentile of the control units. After trimming, we remain with about 99% of the treated units, counting 1,212 and 755 men and women. Huber et al. (2013) propose to remove the control units with a weight higher than 4% of the total. However, because the sample of control units is large, this additional trimming is not required. In the four analyses (2003 and 2004, men and women) the highest relative weight is only 0.17% of the total sample.

To take into account that the PS in the weights $W^C_{l,k(.),i}(s)$ are estimated, we bootstrap the standard errors. As our data come from an endogenously stratified sample, we augment the standard bootstrap method and implement a *stratified bootstrap* by randomly drawing for each replication n_{cr} individuals within each cohort-stratum cr. This is valid because the bootstrap randomly samples individuals within each cohort-stratum (for a review on bootstrap and stratified data see e.g. Shao, 2003). To take individual serial correlation into account we re-sample within each replication the same individuals (i.e. clusters) in the two analyses (2003 and 2004 sample).

In general, once conditioned on the eligibility conditions, the selectivity into treatment on the observables is low in 2003. The Pseudo R-squared of a standard logit model is 0.068 and 0.026 for men and women. The selectivity is slightly higher in 2004. The corresponding Pseudo R-squared are 0.127 and 0.084. This indicates that, once the eligibility conditions are imposed, the sample becomes relatively homogenous. The IPW estimator performs well in balancing the distribution of the covariates (see Table 3 below). In our worst scenario (women selected in 2004), after reweighting the control units by $W^C_{l,1(\infty),i}(s)$, the median Standardized Bias (SB) is as low as 1.2%, the highest SB is 2.9%, the Pseudo R-squared of the reweighted sample is 0 and the Wald test for the joint significance of the variables after the reweighting has a p-value of 1.

3. Cost-Benefit Analysis

To obtain an order of magnitude of the costs and benefits of the TC for the government budget and for society, we perform a cost-benefit analysis (CBA) along the lines proposed by Staubli and Zweimüller (2013). To that purpose we make use of the information available in the administrative dataset on the benefits and gross wages that are paid out to participants and non-participants in the TC. This information is then used to calculate for each individual and in each of the 8 years of analysis after the year of (counterfactual) entry in TC the real costs (or gains) in constant 2004 Euros for the government budget and for society. We weigh the control group by the appropriate IPW to make them comparable to the treated group and calculate for each of these years the average difference of these costs (gains). This provides an estimate of the average net cost (gain) per participant in TC for the government and society in each of these eight years (Staubli and Zweimüller, 2013; Albanese and Cockx, 2015). Differently from the ATT on the survival in employment, these have to be interpreted as the *instantaneous* net costs (gains) during those years.

We consider two scenarios in the CBA. One assumes that the part-time work is *not* replaced by another part-time worker, the other assumes that this part-time work *is* replaced by another who earns an equivalent wage and is equally productive as the part-time worker. In this second scenario we do not take into account, however, the gains in terms of UB payments that would no longer have to be paid if the replacing worker came from unemployment.

There are a number of reasons why we cannot perform a full-fledged CBA. First, as we will explain more in detail below, not all required information at the individual level is available in the administrative database. In these cases we substitute the individual level information by aggregate information obtained from other sources or, if it refers to a very small share of individuals, we ignore the information by setting it to zero. Second, the analysis ignores some important dimensions. For instance, in Section 5.2.3 we found some limited evidence that participation in the TC may have some small positive health effects. However, as we lack information on health costs, we cannot take this dimension into account. Moreover, we ignore the impact of participation in TC on the distribution of welfare or on poverty. All this means that the CBA should be taken as a crude approximation.

2.4. Methodology

We calculate, for each year *t* of the period of analysis above, the effect of the policy on two indicators, the Net Budgetary Cost for the Government (NBC) and the Net Welfare Cost for Society (NWC). Both indicators are expressed per treated individual and in monthly terms (2004 Euros).

i. **Net budgetary cost (gain) for the Government (NBC)**: This is the average cost (gain) of the policy for the state, net of savings for the public budget:

$$NBC_t = allow_t - tax_t * \ remu_t - SSC_t, \tag{7}$$

with:

- $allow_t$: expenditures of the Government on allowances of the Social Security scheme, such as unemployment benefits and the TC allowance. Because the database lacks information on statutory pension, sickness and disability allowances, we impute these allowances as follows:
  - We assume that the worker has worked his entire career in the private sector and assume that the individual is paid the average pension in the private sector according to the age bracket to whom he belongs: 50-54, 55-59, 60-64, 65-69.
  - For sickness and disability benefits we set the allowance equal to the theoretical level of entitlement, i.e. to 60% of the individuals' average monthly remuneration over the last six quarters.
- tax_t: The average personal income tax rate on the gross remuneration in every year (OECD, 2015a).
- $remu_t$: the gross wage earnings. This is observed in the data for employees in the private and public sector, but not for the self-employed, for whom we impute a zero value for both treated and control units. Since the share of self-employed individuals in the control group is larger (4.6%/2.0% of men / women) than in the treated group (2.0%/0.7% of men / women), this slightly biases our cost estimate downwards.
- SSC_t: employer and employee contributions to Social Security.

ii. **Net Welfare cost (gain) for Society (NWC):** the efficiency cost ("excess burden" or "deadweight loss") of the net budgetary expenditures mentioned in (i) plus the opportunity cost of working minus the production value of employment (PV):

$$\begin{aligned} NWC_t &= (MCF - 1) * NBC_t + LEIS_t - PV_t \\ &= (MCF - 1) * NBC_t + LEIS_t - (1 - PG_t) * LC_t \end{aligned} \tag{8}$$

with:

- MCF: the Marginal Cost of Public Funds.[1] For Belgium a MCF equal to two is considered to be appropriate (Kleven and Kreiner, 2006; Barrios Cobos et al., 2013).

[1] The net budgetary cost is *in se* not a cost to Society as it just involves transfers between individuals.

- $LEIS_t$: the opportunity cost of working, which has to be between zero and the net wage plus the SSC). In the latter we include both employee's and employer's Social Security contributions as they can be seen as an insurance premium to entitlements to Social Security benefits. Similarly to Greenberg and Robins (2008), we use the mid-point between the two bounds as our benchmark estimate.
- PG: the age-related pay-productivity gap takes into account that the wage cost of older workers exceeds their productivity. An estimate of the production value of labour (PV) is obtained by downward adjusting the labour costs (LC) by this gap. We use estimates of this pay-productivity gap provided by Vandenberghe et al. (2013) for the Belgian case.

We consider a sensitivity analysis in two directions. First we consider two scenarios: one in which the effect of reduced working time on labour costs is taken into account and one in which it is not (i.e. full replacement of the reduced working hours for both treated and controls). Second, we check the robustness of our results by varying three key parameters of our model:

1. Two personal income tax rates: for persons earning 100% (benchmark) or 133% of the average wage.
2. Three marginal costs of public funds: 1.41, 2.14 (benchmark) and 3.23 (Kleven and Kreiner, 2006).
3. Three values for the opportunity cost of working: the aforementioned lower and upper bound, as well as the midpoint (benchmark).

In order to measure the impact of TC on these indicators, we proceed in the following way. We run by gender a pooled weighted regression on all 8 years of analysis ($t \in \{1,2, \dots 8\}$) separately for $s \in \{0,1\}$, i.e. the two samples of analysis:

$$\sqrt{W_{it}^R} Y_{it} = \sum_{t=s+1}^{8} (\alpha_t + \beta_t 1(S_i = s)) \sqrt{W_{it}^R}\, D_t + \sqrt{W_{it}^R} u_{it}, \tag{9}$$

with

$$W_{it}^R \equiv W_{cr,i} \left[1 + \frac{p_s(X_i,l_i)}{1-p_s(X_i,l_i)} \frac{1}{\prod_{m=s+1}^{t} [1-p_m(X_i,l_i)]} 1(S_i > t) \right] \tag{10}$$

where Y_{it} measures the outcome of interest, i.e. *NBC* or *NWC*, for each individual i in year t after treatment assignment, D_t is a year indicator equal to one in year t and zero otherwise, u_{it} is the error term of the regression. In this regression α_t measures the average outcome for the control units that have not yet been treated in year t, i.e. $S_i > t > s$, so that $1(S_i = s) = 0$, while β_t measures the average treatment effect on the treated in year t (*ATT*$_t$), i.e. $S_i = s < t$, so that $1(S_i = s) = 0$. The weights W_{it}^R ensure that (i) the endogenous sampling is taken into account by weighing all individual observations by $W_{cr,i}$, (ii) the control units that have not yet been treated are made, respectively kept comparable to the treated units by weighing them by the standard IPW $\frac{p_s(X_i,l_i)}{1-p_s(X_i,l_i)}$ and by $\frac{1}{\prod_{m=s+1}^{t}[1-p_m(X_i,l_i)]}$ to take the selective assignment into treatment over time into account. Note, in contrast to the analysis on the survival rate in employment, individuals who leave employment are not dropped from the analysis. Only individuals in the control group who become treated are dropped, but the weights avoid that this induces selectivity in the comparison between treated and control units. For each time period $t \in \{1,2, \dots 8\}$ we can then estimate the ATT_t on *NBC* or on *NWC* for each treatment group $s \in \{0,1\}$ by $\hat{\beta}_{t-s}$, where the hat denotes the estimate of the weighted regression in (10). Subsequently, we average over treatment groups in a similar way as in (6).

2.5. Results

The results of the benchmark cost-benefit analysis (CBA) can be found in Figure 1, which shows the monthly cost per treated for the government budget and the welfare cost to society. Both the scenarios with and without replacement of part-time workers are discussed.

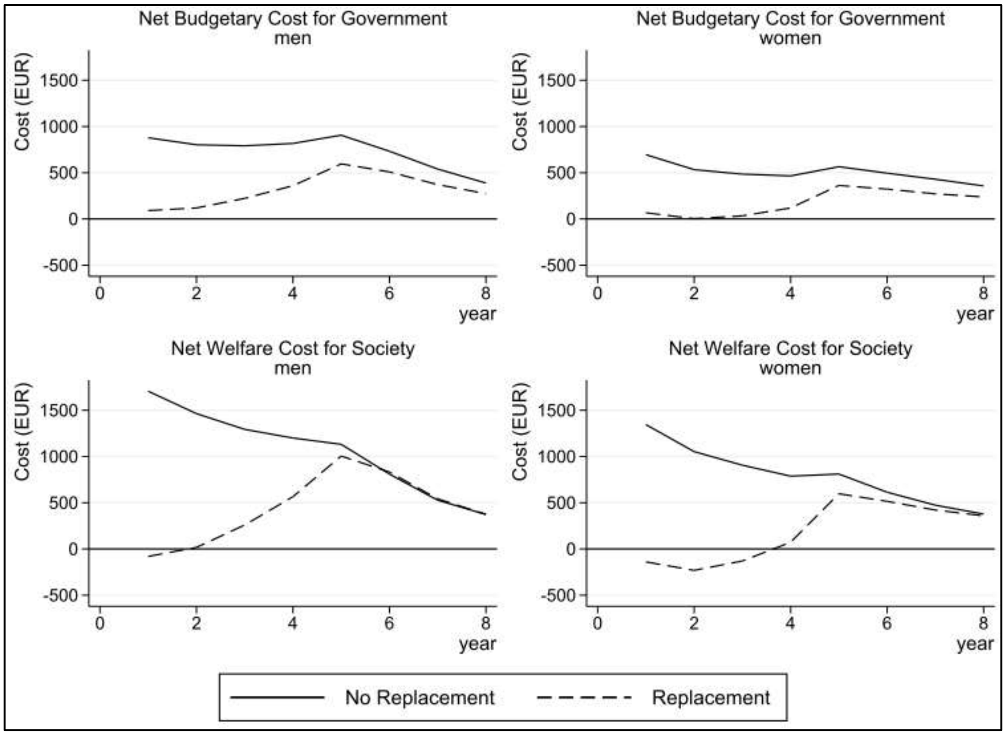
Without the replacement of part-time workers, the results indicate that the TC scheme is an expensive policy that fails the cost-benefit test. Although in the first years we have estimated a positive employment effect, the costs of the policy immediately dominate the benefits. In the first year of the analysis the monthly costs to the government budget in the baseline case are €878 (€696) per treated person for men (women). That these budgetary costs for the government already exist from the first years indicate that the positive effects on employment at the extensive margin are immediately dominated by the reduction of working hours and the allowance paid to the TC participants. In particular, (i) treated people reduce their working hours, resulting in less income for the government in the form of taxes; (ii) TC participants receive the TC benefit, resulting in higher expenditures for the government. As a reduction in working hours also implies lower production, the monthly total welfare costs per treated individual are even higher: €1,706 (€1,345) for men (women). The costs to society show a decreasing pattern, while near the end of the analysis we observe a stronger decrease for both the costs to the government and the society. As already mentioned, this convergence is induced by the fact that all individuals eventually retire, *irrespectively of treatment*, so that costs converge to zero.

If full replacement of the reduced working hours is assumed, a small short-run positive welfare gain for society emerges, but only for women. This is because in this scenario no working hours are lost when a participant enters the TC scheme. To the extent that there are positive employment effects as discussed in Section 5.2, this reduces the negative budgetary impact, while the net gain for society also comes from the positive impact of the TC scheme on the value of leisure. The largest potential benefits are obtained at the start of the analysis and are larger and longer-lasting for women, as they have more favourable short-run employment effects.2 The welfare cost to Society is about zero in year two and four, respectively for men and women. At those points the ATTs on the survival rate in employment was estimated to be about 3.5 pp. This suggests that a positive ATT of about 3.5 percentage points is required to break-even in terms of social welfare. However, this is measured for the most optimistic scenario. In reality it is unlikely that employers could replace all reduced working time. We included two summary tables (Table 1 and Table 2) containing the effects for our sensitivity analyses in which we consider all possible scenarios. Though the magnitude of the estimates in the sensitivity analysis changes, the qualitative results are in line with the findings of the baseline scenario. Our CBA ignores distributional effects and effects on other channels such as health. However, it is clear that without full replacement, the policy fails the cost-benefit test.

In line with our employment analysis we considered the heterogeneous effects of splitting the sample by age, using the same cut-off as in the employment analysis (see Figure 2). This analysis shows that while the policy is more costly for younger workers when we do not consider replacement, it also has the greatest scope for positive budgetary effects once the replacement is taken into account. The potential benefits are greater and longer-lasting for (especially female) younger workers, as they also have longer lasting positive employment effects (cf. Section 5.2.1).

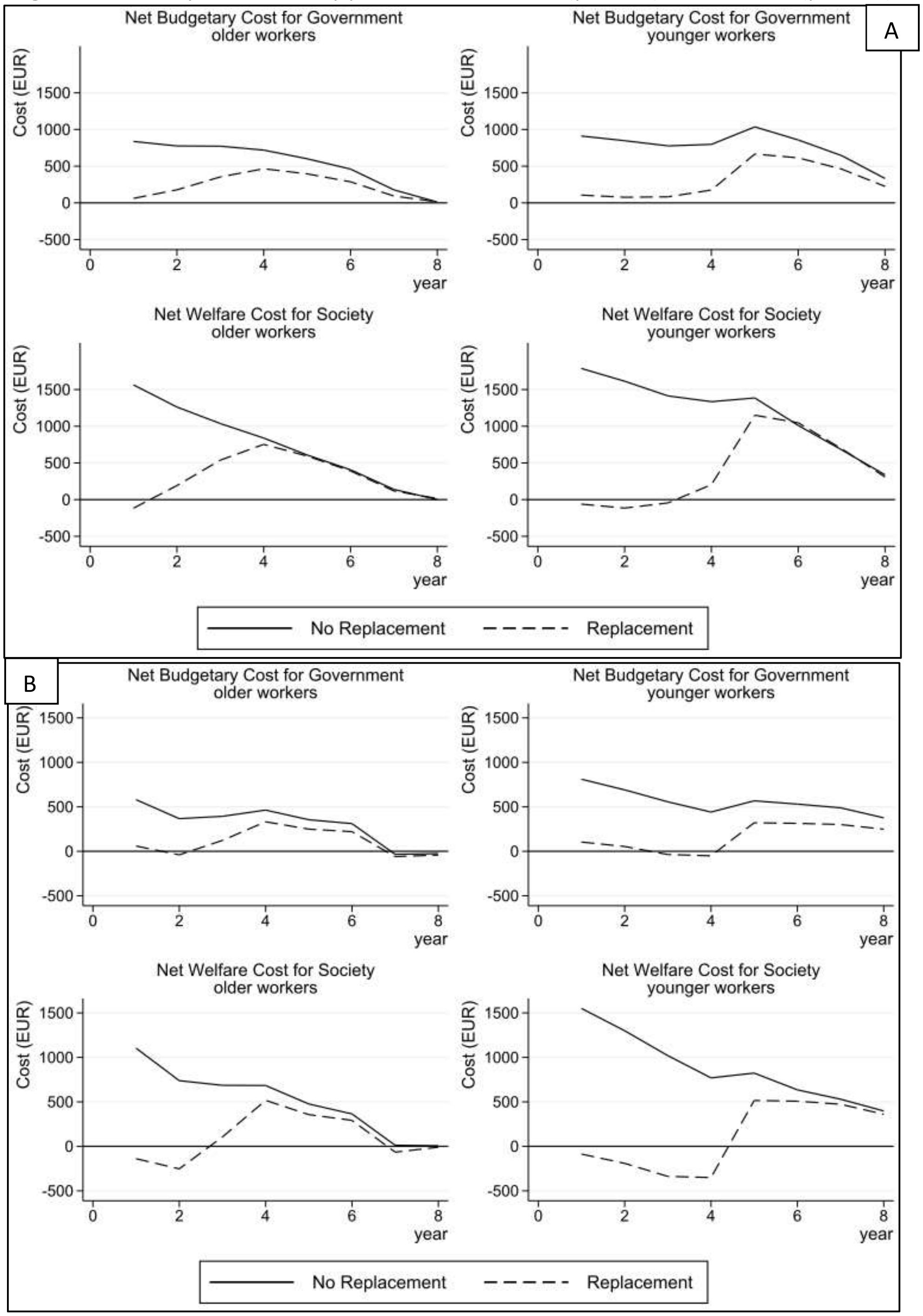
[2] There are almost no benefits for the government budget, while the monthly welfare benefits for society are larger in year one (two) for men (women), when they amount to €80 (231) per treated man (woman). Similar to the ATT on survival in employment, the period with a welfare gain lasts only one year for men, while three years for women.

Figure 1: Monthly Cost of the TC per Treated of 2003 (2004)



Cost-benefit analysis (CBA) on the pooled sample of participants in TC of 2003 and 2004. CBA in monthly costs (benefits if negative) in 2004 euros per treated individual (the size of the treated sample as defined in 2003/2004). The Net Budgetary Cost (NBC) for the government is the average cost (gain) of the policy for the state, net of savings for the public budget. The Net Welfare Cost (NWC) for society is the efficiency cost of the NBC minus the production value of employment (PV). Baseline scenario: CBA without replacement of the part-time workers. Replacement scenario: baseline scenario with the additional assumption that all hours reduced by part-time workers (treated and controls) are recovered by hiring extra workers with similar characteristics. The CBA ignores potential substitution and anticipation effects. The costs to society ignore the value of leisure and potential distributional and health impacts of the measure. The CBA spans all eight years of the ATT analysis (Section 5.2). Year 1 is the first year for which we calculate the ATT. Year 8 only contains information from the 2003 sample.

Figure 2: Monthly Cost of Policy per Treated for Men (panel A) and Women (panel B)



Cost-benefit analysis (CBA) on the pooled sample of participants in TC of 2003 and 2004 (panel A for men, panel B for women). CBA in monthly costs (benefits if negative) in 2004 euros per treated individual (the size of the treated sample as defined in 2003 / 2004). The Net Budgetary Cost (NBC) for the government is the average cost (gain) of the policy for the state, net of savings for the public budget. The Net Welfare Cost (NWC) for society is the efficiency cost of the NBC minus the production value of employment (PV). Baseline scenario: CBA without replacement of the part-time workers. Replacement scenario: baseline scenario with additional assumption that all hours reduced by part-time workers (treated and controls) are recovered by hiring extra workers with similar characteristics. Younger workers are aged strictly below age 56.5 at the moment of sample selection (year 0), older workers are aged 56.5 and above at that moment. The CBA ignores potential substitution and anticipation effects. The costs to society ignore the value of leisure and potential distributional and health impacts of the measure. The CBA spans all eight years of the ATT analysis from Section 5.2. Year one is the first year for which we calculate the ATT. Year eight only contains information from the 2003 sample.

Table 1: Sensitivity analysis on Cost-Benefit Analysis - men

(A) MCF	(B) Reservation Wage	(C) Income taxes	(1) No replacement of part-time workers								(2) Replacement of part-time workers							
			Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
(α) NET BUDGETARY COST (NBC) FOR THE GOVERNMENT																		
-	**-**	**Medium**	**878.3**	**803.0**	**791.2**	**817.0**	**906.1**	**732.6**	**540.0**	**388.6**	**90.7**	**119.2**	**222.7**	**359.6**	**596.2**	**508.9**	**371.5**	**273.8**
-	-	High	914.1	835.7	823.1	850.5	941.4	760.2	559.5	402.6	78.6	111.5	224.3	373.8	627.5	537.0	391.3	288.0
(β) NET WELFARE COST (NWC) FOR SOCIETY																		
1.41	Low	Medium	1,505.2	1,306.4	1,185.0	1,142.1	1,123.5	817.1	530.7	371.0	-306.8	-170.3	126.0	494.0	992.0	839.2	544.8	373.1
1.41	Low	High	1,519.8	1,319.8	1,198.1	1,155.9	1,138.0	828.4	538.7	376.7	-311.8	-173.5	126.7	499.8	1,004.8	850.7	552.9	378.9
1.41	Medium	Medium	1,022.1	864.3	755.4	698.0	653.1	450.9	271.5	182.8	-143.7	-67.9	100.0	305.7	575.4	464.8	279.5	182.1
1.41	Medium	High	1,054.7	894.1	784.4	728.4	685.3	475.9	289.2	195.5	-154.7	-74.9	101.5	318.7	603.9	490.4	297.6	195.0
1.41	High	Medium	539.1	422.1	325.8	253.8	182.6	84.6	12.4	-5.5	19.3	34.6	74.0	117.4	158.9	90.5	14.2	-8.9
1.41	High	High	589.6	468.3	370.7	300.9	232.5	123.4	39.7	14.3	2.3	23.6	76.3	137.5	203.0	130.1	42.2	11.0
2.14	Low	Medium	2,189.4	1,906.3	1,723.1	1,644.4	1,602.2	1,176.1	786.7	560.6	-243.5	-85.9	285.4	753.1	1,421.1	1,206.5	807.4	566.6
2.14	Low	High	2,230.2	1,943.6	1,759.5	1,682.5	1,642.6	1,207.6	808.8	576.5	-257.3	-94.7	287.3	769.4	1,456.8	1,238.6	830.0	582.8
2.14	**Medium**	**Medium**	**1,706.4**	**1,464.1**	**1,293.5**	**1,200.2**	**1,131.8**	**809.9**	**527.5**	**372.3**	**-80.4**	**16.5**	**259.4**	**564.9**	**1,004.6**	**832.2**	**542.1**	**375.6**
2.14	Medium	High	1,765.1	1,517.8	1,345.7	1,255.1	1,189.8	855.1	559.3	395.3	-100.3	3.8	262.1	588.3	1,055.9	878.3	574.6	398.8
2.14	High	Medium	1,223.3	1,022.0	863.9	756.0	661.4	443.6	268.3	184.1	82.6	119.0	233.4	376.6	588.0	457.9	276.8	184.7
2.14	High	High	1,299.9	1,092.1	932.0	827.6	737.1	502.6	309.9	214.1	56.8	102.4	236.9	407.1	654.9	518.1	319.3	214.9
3.23	Low	Medium	3,211.1	2,801.9	2,526.6	2,394.4	2,317.1	1,712.2	1,168.9	843.7	-149.0	40.2	523.5	1,140.1	2,061.9	1,755.1	1,199.4	855.7
3.23	Low	High	3,290.9	2,874.9	2,597.6	2,469.0	2,396.0	1,773.6	1,212.2	874.9	-176.0	22.9	527.0	1,171.9	2,131.6	1,817.8	1,243.7	887.2
3.23	Medium	Medium	2,728.0	2,359.8	2,097.0	1,950.2	1,846.6	1,345.9	909.7	655.4	14.1	142.6	497.5	951.9	1,645.3	1,380.8	934.1	664.7
3.23	Medium	High	2,825.8	2,449.2	2,183.9	2,041.5	1,943.2	1,421.1	962.7	693.6	-18.9	121.4	501.8	990.8	1,730.7	1,457.5	988.4	703.3
3.23	High	Medium	2,245.0	1,917.6	1,667.3	1,506.0	1,376.2	979.7	650.5	467.2	177.2	245.0	471.5	763.6	1,228.8	1,006.4	668.8	473.7
3.23	High	High	2,360.6	2,023.4	1,770.2	1,614.1	1,490.5	1,068.7	713.2	512.4	138.1	220.0	476.7	809.7	1,329.8	1,097.3	733.0	519.4

Cost-benefit analysis (CBA) on the pooled sample of male participants in TC of 2003 and 2004. Treated sample size defined in 2003-2004. CBA in monthly costs (benefits if negative) in 2004 euros per treated individual under different scenarios. Scenario in **bold** denotes baseline scenario. (1) No replacement of part-time workers scenario; and (2) replacement scenario: baseline scenario with additional assumption that all hours reduced by part-time workers (treated and controls) are recovered by hiring extra workers with similar characteristics. Additionally (A) Marginal Cost of public Funds (MCF) equal to 1.41, 2.14, or 3.23 (Kleven and Kreiner, 2006), (B) opportunity cost of working (Reservation Wage) with a lower, medium and upper bound and (C) Income Tax Rate, variable over time: medium (average income) which is on average 28.26%, higher (133% average income) which is on average 32.36% (OECD stat extract, 2003-2011). The first outcome variable is the Net Budgetary Cost (NBC) for the government (α), i.e. the average cost (gain) of the policy for the state, net of savings for the public budget. The second is the Net Welfare Cost (NWC) for society (β), i.e. the efficiency cost of the NBC minus the production value of employment. The CBA ignores potential substitution and anticipation effects. The costs to society ignore potential distributional and health impacts of the measure. The CBA spans all eight years of the ATT analysis from Section 5.2. Year one is the first year for which we calculate the ATT. Year eight only contains information from the 2003 sample

Table 2: Sensitivity analysis on Cost-Benefit Analysis - women

(A) MCF	(B) Reservation Wage	(C) Income taxes	(1) No replacement of part-time workers								(2) Replacement of part-time workers							
			Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8
(α) NET BUDGETARY COST (NBC) FOR THE GOVERNMENT																		
-	**-**	**Medium**	**695.5**	**533.5**	**485.0**	**466.0**	**565.1**	**495.5**	**429.6**	**357.6**	**67.0**	**3.8**	**33.5**	**119.3**	**362.8**	**322.8**	**271.7**	**237.8**
-	-	High	721.9	554.2	504.2	484.8	589.2	515.5	446.2	371.4	53.9	-11.5	22.3	115.9	379.7	339.2	286.0	250.7
(β) NET WELFARE COST (NWC) FOR SOCIETY																		
1.41	Low	Medium	1,139.7	872.2	761.1	684.7	777.2	598.3	467.6	374.5	-366.8	-442.7	-302.2	-50.5	558.5	496.6	414.6	355.8
1.41	Low	High	1,150.6	880.8	769.0	692.4	787.1	606.5	474.4	380.2	-372.2	-449.0	-306.8	-51.9	565.4	503.3	420.5	361.1
1.41	Medium	Medium	790.1	601.2	511.0	439.7	461.9	335.8	248.9	192.9	-185.7	-228.9	-145.7	-2.8	336.2	281.3	224.9	186.6
1.41	Medium	High	814.2	620.1	528.5	456.8	483.8	354.1	264.0	205.5	-197.6	-242.8	-155.8	-5.9	351.5	296.3	238.0	198.3
1.41	High	Medium	440.5	330.1	261.0	194.7	146.6	73.3	30.1	11.3	-4.6	-15.0	10.9	44.9	113.9	66.1	35.1	17.3
1.41	High	High	477.8	359.4	288.0	221.2	180.5	101.6	53.5	30.8	-23.1	-36.6	-4.9	40.1	137.7	89.2	55.4	35.5
2.14	Low	Medium	1,694.9	1,324.0	1,156.3	1,032.4	1,125.8	876.3	692.4	559.0	-320.9	-445.0	-286.3	25.1	819.6	730.0	611.2	527.5
2.14	Low	High	1,725.0	1,347.7	1,178.2	1,053.8	1,153.2	899.2	711.3	574.7	-335.8	-462.5	-299.1	21.3	838.9	748.7	627.6	542.2
2.14	**Medium**	**Medium**	**1,345.3**	**1,052.9**	**906.2**	**787.5**	**810.5**	**613.8**	**473.6**	**377.4**	**-139.8**	**-231.1**	**-129.8**	**72.9**	**597.3**	**514.7**	**421.4**	**358.3**
2.14	Medium	High	1,388.6	1,087.0	937.7	818.2	849.9	646.7	500.8	400.0	-161.3	-256.3	-148.2	67.3	625.0	541.6	445.0	379.4
2.14	High	Medium	995.7	781.8	656.2	542.5	495.1	351.4	254.8	195.8	41.3	-17.3	26.7	120.6	375.0	299.5	231.6	189.1
2.14	High	High	1,052.3	826.4	697.3	582.7	546.6	394.3	290.4	225.3	13.3	-50.1	2.8	113.3	411.2	334.6	262.4	216.7
3.23	Low	Medium	2,523.8	1,998.5	1,746.4	1,551.7	1,646.3	1,291.5	1,027.9	834.5	-252.3	-448.4	-262.7	138.1	1,209.5	1,078.5	904.6	784.0
3.23	Low	High	2,582.7	2,044.9	1,789.2	1,593.6	1,699.9	1,336.2	1,065.0	865.2	-281.6	-482.6	-287.6	130.6	1,247.2	1,115.0	936.7	812.7
3.23	Medium	Medium	2,174.2	1,727.4	1,496.3	1,306.8	1,330.9	1,029.0	809.2	652.9	-71.2	-234.5	-106.1	185.8	987.2	863.2	714.8	614.8
3.23	Medium	High	2,246.3	1,784.2	1,548.8	1,358.0	1,396.6	1,083.7	854.5	690.5	-107.0	-276.4	-136.7	176.6	1,033.3	908.0	754.1	649.9
3.23	High	Medium	1,824.6	1,456.4	1,246.3	1,061.8	1,015.6	766.5	590.4	471.3	109.9	-20.7	50.4	233.5	765.0	648.0	525.1	445.6
3.23	High	High	1,910.0	1,523.5	1,308.3	1,122.4	1,093.3	831.3	644.1	515.8	67.5	-70.2	14.3	222.7	819.5	700.9	571.5	487.2

Cost-benefit analysis (CBA) on the pooled sample of female participants in TC of 2003 and 2004. Treated sample size defined in 2003-2004. CBA in monthly costs (benefits if negative) in 2004 euros per treated individual under different scenarios. Scenario in **bold** denotes baseline scenario. (1) No replacement of part-time workers scenario; and (2) replacement scenario: baseline scenario with additional assumption that all hours reduced by part-time workers (treated and controls) are recovered by hiring extra workers with similar characteristics. Additionally (A) Marginal Cost of public Funds (MCF) equal to 1.41, 2.14, or 3.23 (Kleven and Kreiner, 2006) (B) opportunity cost of working (Reservation Wage) with a lower, medium and upper bound and (C) Income Tax Rate, variable over time: medium (income tax on the average income) which is on average 28.26%, higher (income tax on 133% of the average income) which is on average 32.36% (OECD stat extract, 2003-2011). The first outcome variable is the Net Budgetary Cost (NBC) for the government (α), i.e. the average cost (gain) of the policy for the state, net of savings for the public budget. The second is the Net Welfare Cost (NWC) for society (β), i.e. the efficiency cost of the NBC minus the production value of employment. The CBA ignores potential substitution and anticipation effects. The costs to society ignore potential distributional and health impacts of the measure. The CBA spans all eight years of the ATT analysis from Section 5.2. Year one is the first year for which we calculate the ATT. Year eight only contains information from the 2003 sample.

Bibliography

Albanese, A., Cockx, B., 2015. Permanent Wage Cost Subsidies for Older Workers: An Effective Tool for Increasing Working Time and Postponing Early Retirement? . IZA Discussion Paper No. 8988, IZA, Bonn.

Barrios Cobos, S., Saveyn, B., Pycroft, J., 2013. The Marginal Cost of Public Funds in the Eu: The Case of Labour Versus Green Taxes. Taxation papers 35, Publications Office of the European Union.

Busso, M., DiNardo, J., McCrary, J., 2014. New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators. The Review of Economics and Statistics 96(5), 885–897.

Crépon, B., Ferracci, M., Jolivet, G., van den Berg, G., 2009. Active Labor Market Policy Effects in a Dynamic Setting. Journal of the European Economic Association 7(2–3), 595–605.

Fredriksson, P., Johansson, P., 2008. Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data. Journal of Business & Economic Statistics 26(4), 435–445.

Frölich, M., Huber, M., Wiesenfarth, M., 2015. The Finite Sample Performance of Semi- and Nonparametric Estimators for Treatment Effects and Policy Evaluation . IZA Discussion Papers No. 8756, IZA, Bonn.

Greenberg, D., Robins, P., 2008. Incorporating nonmarket time into benefit-cost analyses of social programs: An application to the self-sufficiency project. Journal of Public Economics 92(3–4), 766–794.

Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. Econometrica 71(4), 1161–1189.

Horvitz, D.G., Thompson, D.J., 1952. A Generalization of Sampling Without Replacement from a Finite Universe. Journal of the American Statistical Association 47(260), 663–685.

Huber, M., Lechner, M., Wunsch, C., 2013. The Performance of Estimators Based on the Propensity Score. Journal of Econometrics 175(1), 1–21.

Huber, M., Lechner, M., Wunsch, C., 2016. The Effect of Firms' Phased Retirement Policies on the Labor Market Outcomes of Their Employees. ILR Review 19793916644755.

Kleven, H.J., Kreiner, C., 2006. The Marginal Cost of Public Funds: Hours of Work Versus Labor Force Participation. Journal of Public Economics 90(10–11), 1955–1973.

Lechner, M., Strittmatter, A., 2014. Practical Procedures to Deal with Common Support Problems in Matching Estimation . Economics Working Paper Series No. 1410, University of St. Gallen, St. Gallen.

Shao, J., 2003. Impact of the Bootstrap on Sample Surveys. Statistical Science 18(2), 191–198.

Staubli, S., Zweimüller, J., 2013. Does raising the early retirement age increase employment of older workers? Journal of Public Economics 108(C), 17–32.

Vandenberghe, V., Waltenberg, F., Rigo, M., 2013. Ageing and Employability. Evidence from Belgian Firm Level Data. Journal of Productivity Analysis 40, 111–136.

Vikström, J., 2014. IPW estimation and related estimators for evaluation of active labor market policies in a dynamic setting. Working Paper Series from IFAU - Institute for Evaluation of Labour Market and Education Policy, No 2014:1, Uppsala.