The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0143-7720.htm

Hope for the best and prepare for the worst. Do short-time work schemes help workers remain in the same firm?

José M. Arranz, Carlos García-Serrano and Virginia Hernanz
Departamento de Economía, Facultad de Ciencias Económicas, Empresariales y Turismo, Universidad de Alcalá, Alcalá de Henares, Spain

Received 28 April 2020
Revised 19 August 2020
Accepted 29 October 2020

Abstract

Purpose – This paper investigates whether short-time work (STW) schemes were successful in their objective of maintaining employment and keeping workers employed within the same firms after the onset of the financial and economic crisis in 2008.

Design/methodology/approach – Spanish longitudinal administrative data has been used, making it possible to identify short-time work (STW) participation not only of workers but also of employers and allowing to know the future labour market status of participants and non-participants. Accordingly, treatment and control groups are defined, and Propensity Score Matching models estimated. The dependent variable is measured as the probability that an individual remained employed with the same employer in the future (one, two and three years) after implementation of a STW arrangement.

Findings – Our results suggest that treated individuals are about 5 percentage points less likely to remain working with the same employer one year later than similar workers, and this negative effect of participation increases over time. Thus, STW schemes would not have the assumed effect of preventing unemployment by keeping the participants employed relative to non-participants.

Research limitations/implications – As our analysis is based on the comparison of the employment trajectories of participant and non-participant workers in firms that have used STW arrangements, our findings cannot be interpreted as the job saving effects of either macro or micro studies carried out previously.

Practical implications – The analysis carried out in the paper is complementary to the country-level and firm-level approaches that have been used in the empirical literature.

Originality/value – We adopt a worker-level approach. This is novel since no previous study has focused attention on the impact of STW participation on the subsequent labour market status of workers.

Keywords Short-time work, Employment stability, Worker-level longitudinal data, Propensity score matching

Paper type Research paper

1. Introduction

The objective of short-time work (STW) schemes is to preserve jobs at firms experiencing temporarily low demand by encouraging work-sharing. In these schemes, the contract of an employee with the firm is maintained during the period of reduced hours or the suspension of work, providing income support to these workers. Nearly all OECD countries currently operate public STW schemes (OECD, 2010).

Academic and political interest in STW usually upsurges in times of economic crisis. This occurred during the early and mid-1990s and during the last recession as well. Facing the

The authors gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (*Programa Estatal de I+D+i Orientada a los Retos de la Sociedad*, ECO2014-57623-R) and from the Institute of Fiscal Studies (Madrid) under the contract "*Evaluación del impacto de políticas públicas: las prestaciones por desempleo y los expedientes de regulación de empleo*". Thanks are extended to professor S. Pi Nette and various seminar and conference participants for their valuable comments and suggestions. The authors also wish to thank the Spanish Social Security for providing the data for this research. Obviously, the opinions and analyses are the responsibility of the authors. The usual disclaimer applies.

International Journal of Manpower

0143-7720
DOI 10.1108/IJM-04-2020-0178

negative financial and economic situation of 2008–2009 with the strongest downturn in decades, fighting unemployment was one of the most important tasks of European countries. Most governments took specific measures in response to the crisis to promote the use of STW schemes, by weakening eligibility and conditionality requirements and increasing generosity, while others established new programs (Eurofound, 2010; Panteia, 2012) [1]. The question is whether these measures contributed to the maintenance of jobs.

The empirical evidence shows that during the last downturn the decline in employment has been small in some countries compared with what would have been expected given the size of the decline in output, due in large part to reductions in average hours having accounted for an unusually high share of the total adjustment in labour input (see, for instance, OECD, 2010). Several analysts and policy-makers have attributed the good performance of some national labour markets (in particular, the German one) to the role played by STW schemes to provide a buffer for downward adjustment during the economic cycle, although their contribution to maintain employment and to avoid dismissals is still not without dispute [2]. Whether that is in fact the case remains uncertain, because there has been little systematic evaluation of the employment preserving effect of work-sharing during the crisis. This has been done by the otherwise limited empirical literature, with contradicting results (see below). This article attempts to contribute to fill this evidence gap by estimating the impact of STW arrangements on the continuation of employment relationships.

Our contribution to this literature is twofold. Firstly, we carry out a worker-level approach. This implies that, using individual data, we examine whether STW schemes achieve their stated goal of keeping workers employed within the same companies. Secondly, we employ longitudinal administrative data. The database used presents at least three advantages for the analysis of STW schemes and its impact on labour market outcomes of workers: the information is available, accurate and detailed on the jobs held by the individuals and on the spells of receipt of unemployment benefits (UB); it makes it possible to identify STW participation not only of workers but also of employers; and it allows to know the subsequent labour market status of participants and non-participants after the implementation of a STW scheme by companies, i.e. remaining working with the same employer, being employed with a different firm or being jobless. Thus, our approach allows us to focus attention on the impact of STW participation on the subsequent trajectories of workers. This is novel since empirical studies usually use either country-level or firm-level data. Examining whether STW programs keep workers employed in the same company is relevant since, when using either country- or firm-level data, one is unable to evaluate the effect of STW on individual labour market biographies. This implies that it may well be that, even when no employment preserving effect of STW on either the country or the establishment level is found, STW might contribute to prevent individual unemployment. Therefore, our analysis fits in with the theoretical debate on whether work-sharing can reduce unemployment. Moreover, our results could be used to improve theoretical models on employment adjustment decisions made by companies, but also models on worker turnover.

Accordingly, in the empirical analysis carried out, treatment and control groups are defined, and the labour market transitions of both groups examined. The dependent variable is measured as the probability that an individual, who was already employed at the time of the implementation of such arrangement, remained employed with the same employer in the future (one, two and three years after implementation). The evaluation method we use is matching based on the propensity score. We examine the sensitivity of estimates of program impacts across alternative algorithms to match participants and nonparticipants and perform several robustness checks to gauge whether unobserved characteristics (the so-called "hidden bias") may play a role (i.e. whether the chance of STW participation is correlated with unobserved factors that determine both eligibility for STW and employment retention, thus inducing a bias on the STW participation effect).

Our case study is Spain, a country that has not been investigated before in the area of work-sharing despite its interest, given the supposedly strict employment protection legislation and the importance of temporary employment [3]. The Spanish STW scheme is the legal instrument used to protect employment in cases of exceptional circumstances, allowing firms either to temporarily reduce employees' working time or to carry out temporary layoffs. The STW scheme is financed by contributions (uniform payroll taxes) paid by employers and workers, like unemployment insurance benefits. In order to benefit from STW schemes, prior consultation with trade unions is required. The company must inform workers' representatives of the reasons for the measures, their extent and expected duration, and the number of employees involved. The procedure implies a process of negotiation on various issues, including the number and types of job affected by the proposed measures, the details of these measures and the criteria used to designate the workers involved (for instance, workers' representatives and employers may negotiate priorities of permanence for specific groups, such as workers with family burdens or older workers). At the same time the firm has to notify the labour authority, providing the necessary information to justify the measures proposed (for more details, see Arranz *et al.*, 2018).

The structure of the article runs as follows. Section 2 provides a review of the empirical literature related to STW schemes and labour market outcomes. Section 3 presents the data and the main variables. Section 4 outlines the econometric methods. Section 5 reports the empirical analysis. Finally, section 6 discusses the findings and concludes.

2. Literature review

Lacking government intervention, there are several reasons by which employers (and even workers) may prefer firings to STW as the way to adjust to declines in firms' demand. These reasons are mostly related to the nature of bargaining, technical restrictions and dynamic effects of the two alternatives, and imply that STW requires the intervention of a "third party" that subsidizes such schemes and penalizes firings (Rosen, 1985; Fitzroy and Hart, 1985; Hall, 1995). The analysis of unemployment insurance in a second-best environment featuring imperfect financial markets suggests that a system combining STW arrangements with UB seems more equitable and efficient than UB only (Abraham and Houseman, 1994; Cahuc and Carcillo, 2011). This system can be more efficient since it reduces excess layoffs encouraged by the implicit subsidies paid out by the public unemployment insurance. It is also more equitable because STW schemes distribute the adjustment burden over a larger number of workers.

Obviously, design features are important in affecting both their success among employers and workers and its efficiency properties. There is considerable variation in the institutional design of STW schemes across countries. This is likely to reflect different strategies for balancing concerns about assuring adequate take-up while maintaining cost-effectiveness, thus limiting potential deadweight and displacement effects. Dimensions such as the strictness of eligibility and entitlement requirements, the conditionality requirements, and the generosity, may significantly affect the take-up rates and the overall results of the STW schemes (see Hijzen and Venn, 2011) [4]. Estimates of the impact of design features of STW programs on take-up rates using cross-country time-series observations (Boeri and Bruecker, 2011; Hijzen and Martin, 2013; Lydon *et al.*, 2019) confirm that more restrictive eligibility requirements and higher costs for employers are associated with lower take-up rates and lower responsiveness of take-up rates to output shocks. Restrictions on the permissible range of working-time reductions lower the take-up rates and limit their responsiveness to changes of output. Moreover, the inclusion of variables capturing labour market institutions reveals that lower take-up rates are associated with less stringent employment protection legislation, more generous UB systems and more decentralized wage bargaining structures.

Most studies on STW have focused on the analysis of the potential employment preserving effect of STW schemes. Both country-level and firm-level approaches have been used in the empirical literature. On the one hand, macro estimates are designed to exploit the country and time variation in take-up rates to analyze the quantitative impacts of STW schemes on labour market outcomes (Abraham and Houseman, 1994; Van Audenrode, 1994; Arpaia *et al.*, 2010; Hijzen and Venn, 2011). On the other hand, micro estimates are based on establishment-or firm-level data and exploit the variation between participating and non-participating employers within countries (Calavrezo *et al.*, 2009, 2010, and Duhautois *et al.*, 2009, for France; Dietz *et al.*, 2011, and Crimmann *et al.*, 2012, for Germany).

Although previous studies do not usually take account of the potential endogeneity of STW schemes, subsequent works do. On the macro side, they do this by instrumenting the STW take-up rate (Cahuc and Carcillo, 2011; Boeri and Bruecker, 2011; Hijzen and Martin, 2013). On the micro side, authors either use instruments based on the STW experience of firms before the crisis (Boeri and Bruecker, 2011; Bellman *et al.*, 2015) or apply propensity score matching (PSM) techniques (Kato and Kodama, 2019). Macro studies tend to find short-run positive impacts of STW on employment during the 2008–2009 recession. However, deadweight costs seem to be sizable as well, especially in the medium run. The empirical evidence from micro studies which control for STW endogeneity are less clear-cut, showing either small positive effects or no effects at all on stabilizing employment or avoiding dismissals, although some studies find more positive results (see Siegenthaler and Kopp, 2019, who find strong evidence that STW in Switzerland increased establishment survival and prevented rather than postponed dismissals in the period 2009–2014).

Almost all studies on STW are conducted at country or company level. The analysis of the effect of STW at the worker level is rare. Speckesser (2010) focuses on Germany during the recession of 1993–1994 and, applying PSM techniques, finds very short-lived employment effects of STW: the employment rate of participants is higher than non-participants during the first three months after the beginning of the program but virtually the same afterwards. His analysis also shows negative and significant wage effects for STW participants as compared with non-participants in the long run (six years). Calavrezo and Lodin (2012) examine the French program in 2008–2010 and, using quarterly data, find that the probability of unemployment is twice as high for participants than non-participants within three months after being in STW. However, their analysis is basically descriptive and does not take account of firm effects, selection issues and the potential endogeneity of the STW variable.

Finally, Pavlopoulos and Chkalova (2019) investigates the effects of STW in surviving firms in the Netherlands in 2009–2011. They apply a discrete-time survival model using a dataset with monthly register data. Participants in the STW program are compared with two control groups: non-participant workers from firms that used STW and workers from firms that did not make use of the program. Their findings indicate that STW had a positive effect: the risk of unemployment and job separation is lower for STW participants than non-participant, although, after 25–30 months, STW participants exhibit the same hazard of unemployment (job separation) as workers from non-participant firms (non-participants from firms that use STW).

Based on the results of the empirical literature and on theoretical considerations, we may hypothesize that, if STW is an internal arrangement that, *ceteris paribus*, helps firms improve their response to a demand shock, it would reduce the probability of unemployment for all workers of the participant company (without differences among workers). However, it also may be the case that STW can be used to create different perspectives for participant and non-participant workers within an employer. For instance, firms could use STW arrangements to protect their core staff during crises and to avoid brain drains (Crimmann *et al.*, 2012), so the likelihood of becoming unemployed would be different for participants than non-participants. In the empirical section of the article, we try to test these predictions on the impact of STW on the permanence of workers within firms.

3. Econometric methods

Estimating the causal effect of STW participation on the permanence of workers in their jobs implies being able to appraise the counterfactual situation in which those same workers are instead not participating in STW. This is called the fundamental evaluation problem. Another problem is the presence of selection bias, that is the possibility that STW workers and non-STW workers could have significant differences even in the absence of treatment. The econometric methodology we use is propensity score matching (PSM) [5]. Assumptions along with this technique imply treatment estimates can be interpreted as causal.

The outcome variable is measured as the probability that an individual, who was already employed at the time of the implementation of such arrangement, remains employed with the same employer in the future (one, two and three years after implementation), while the measure of exposure to a STW scheme is a variable indicating whether or not the worker was involved in an arrangement at a given period. Therefore, two groups are identified: workers affected by STW schemes ($T_i = 1$) and workers not affected ($T_i = 0$). PSM methods allow to construct a statistical comparison group (C) based on a model of the probability of participating in the treatment (T) conditional on observed characteristics, X, the so-called propensity score:

$$P(X_i) \quad = \quad \Pr(T_i = 1|X_i) \tag{1}$$

with ($0 < P(X_i) < 1$). The average treatment effect of the program is calculated as the mean difference in outcomes across these two groups. Under certain assumptions, Rosenbaum and Rubin (1983) show that matching on $P(X)$ is as good as matching on X.

The necessary assumptions for identification of the program effect are (1) conditional independence and (2) presence of a common support. Conditional independence (or unconfoundedness) states that, given a set of covariates, X, that are not affected by treatment, potential outcomes, Y, are independent of treatment assignment, T: $(Y_i^T, Y_i^C) \perp T_i|X_i$. It implies that the uptake of the program is based on observed characteristics. As noted by Imbens (2004), if the previous condition holds, conditioning on the propensity score removes all biases due to observable attributes. Conditional independence is a strong assumption. Nevertheless, due to the features of the process of negotiation between employers and workers' representatives (explained in the introduction section) and the set of variables at our disposal in the dataset used in the empirical analysis (see below), we consider that assuming selection on observables is justified in this case. Regarding the common support (i.e. $0 < P(T = 1|X_i) < 1$), the basic intuition is that there has to be at least one similar individual in the counterfactual state for each treated one. This condition ensures that treatment observations have comparison observations "nearby" in the propensity score distribution (Heckman *et al.*, 1999).

The treatment effect of the program using these methods can be represented as either the Average Treatment Effect (ATE) or the Average Treatment Effect on the Treated (ATET). Typically, researchers can ensure internal as opposed to external validity of the sample, so only the ATET (the difference between the expected outcome of treated individuals who have been treated and the expected outcome of treated individuals had not they been treated) is estimated. To do so, weaker assumptions of conditional independence ($(Y_i^C) \perp T_i|X_i$) as well as common support ($P(T = 1|X_i) < 1$) are needed. In our estimations, we will impose a common support by dropping treatment observations whose *pscore* is either higher than the maximum or lower than the minimum *pscore* of the controls. We will also test that there is not hidden bias and, thus, that the conditional independence assumption (CIA) holds.

In practice, the PSM must be estimated. The predicted values from standard logit or probit models are normally used to estimate the propensity score for each observation in the treatment and the control group samples. Using the propensity score, $\widehat{P}(X)$, matched pairs are constructed on the basis of how close the scores lie across the two samples. Different matching methods are used to do so and obtain estimates of the outcomes of those assigned to treatment and those assigned to the control group (see Mueser *et al.*, 2007, in the context of an application in a workforce training setting, and Caliendo and Kopeinig, 2008) [6].

4. Data

The data we use to investigate the impact of STW on the retention of workers within firms come from the Continuous Sample of Working Life (*Muestra Continua de VidasLaborales*; hereinafter, MCVL), an administrative dataset built upon the computerized records of the Spanish Social Security. The reference population includes both workers registered with the Social Security as working, as well as recipients of contributory and non-contributory pensions and UB. This administrative dataset provides information on individual, job and employer attributes as well as on the UB received by each worker in the event that they were separated from their job and eligible to receive them (in particular, whether each individual was receiving UB when out of work, the type of benefits, and the number of days of benefit). This also applies to spells of benefit receipt while in STW because of either suspension of contract or reduction of working time. For more details on the characteristics of the dataset, see Arranz *et al.* (2013, 2014) [7].

The MCVL has a longitudinal design: an individual who is present in an edition of the sample and subsequently remains registered with the Social Security stays as a sample member. Its longitudinal nature makes it possible to know the subsequent labor market status of a given individual after a STW scheme has been adopted in a company or a job separation has taken place.

The production of unbiased estimates of a treatment effect through PSM methods depends heavily on the quality of the data used. As pointed out by Mueser *et al.* (2007), the use of administrative data to obtain PSM estimates of the average treatment effect can be a very effective tool. One reason for this is the fact that data on the outcome for both treated and untreated individuals come from the same source. Another important characteristic is the availability of large datasets. In our case, the use of longitudinal administrative data allows us to have a representative sample of workers and a relatively rich set of variables to employ in our analyses.

Our sample is limited to wage and salary individuals aged 16–59 who work in the non-agriculture private economy (the individuals are registered with the General System of Social Security in their current job), hold an open-ended contract, work full-time and have a job tenure longer than one year. The purpose is to exclude those workers with a marginal attachment to firms who cannot be potentially (and legally) eligible in the event that an employer would decide to run a STW arrangement.

The outcome variable in our analysis is the labor market status of the individual several (one, two and three) years after the participation on STW has taken place. We take advantage of the characteristics of the dataset to construct treatment and control groups. An observation window (or period of entry) is defined which refers to the second quarter of 2009 (once the financial and economic crisis was on its height). The individuals starting a spell of receipt of STW in this period conforms the *treatment group*. In fact, we select individuals who begin STW spells anytime during the second quarter of 2009. Individuals can start such a spell between the 1st of April 2009 and the 30th of June 2009, while their successive entries into either the unemployment compensation system or other jobs may occur until the 31st of December 2013 (the last day available in our dataset). The *comparison group* is comprised of employed individuals who are not involved in STW schemes during the afore-mentioned period (i.e. April 2009–December 2013). This implies that the comparison group is not contaminated by either previous or future participation in STW arrangements. In fact, we consider only non-participating workers in firms that have made use of STW schemes. We do so because this is the group of workers that can be considered the most appropriate comparison group to carry out the evaluation exercise proposed.

For each group defined according to their treatment status (treated and non-treated), period one covers the quarter in which treatment (participation in STW) takes place, while period two corresponds to one, two and three years later. In period two, the outcome is measured by the individual's permanence in employment within the same firm.

Table A1 of the Appendix provides summary statistics, separately for the sample of treated and non-treated workers, for a comprehensive set of explanatory variables. These are individual, job and employer attributes (gender, age, nationality, region, job category, industry affiliation and size). When compared with non-participating workers in firms which have run STW arrangements, the picture is that certain categories of workers and jobs are over-represented among STW participants. These are male and older workers, long-tenured employees, manual occupations, medium-sized and large employers (over 50 employees), manufacturing industries, and the regions of Catalonia and the Basque Country.

5. Results

5.1 Simple means estimates

Table 1 provides descriptive evidence on labor market outcomes of the samples of the treatment and control groups after one, two and three years. The top panel offers the detailed information distinguishing all possible statuses, while the bottom panel provides a summary: the first two rows show the mean proportion of employees who are still working with the same employer distinguishing between treatment and control groups, and the final row offers the differences in that indicator across groups of workers.

Detailed information		STW same firm	Employed same firm	Employed different firm	UI	Other benefits	Non-employment	Total
1 year later (2010Q2)								
Control	Obs	0	10,029	774	3	41	194	11,676
	%	0.0	85.9	6.6	0.0	0.4	1.7	100
Treatment	Obs	804	1,081	114	2	4	22	2,277
	%	35.3	47.5	5.0	0.1	0.2	1.0	100
2 years later (2011Q2)								
Control	Obs	0	8,967	1261	22	111	489	11,676
	%	0.0	76.8	10.8	0.2	1.0	4.2	100
Treatment	Obs	352	1,294	263	3	10	67	2,277
	%	15.5	56.8	11.6	0.1	0.4	2.9	100
3 years later (2011Q2)								
Control	Obs	0	8,101	1597	59	219	837	11,676
	%	0.0	69.4	13.7	0.5	1.9	7.2	100
Treatment	Obs	356	1,106	337	31	82	138	2,277
	%	15.6	48.6	14.8	1.4	3.6	6.1	100

Summary	2010Q2 1 year later	2011Q2 2 years later	2012Q2 3 years later
Control	85.9	76.8	69.4
Treatment	82.8	72.3	64.2
Difference T-C	−3.1	−4.5	−5.2

Table 1. Labor market performance of individuals by treatment status and control groups: proportion of individuals who remain employed with the same firm several years after treatment. Spain (MCVL, 2009–2013)

The raw data presented in the table show that employment stability is lower in the treatment group than in the control group and that the differential between both groups increases over time (from 3.1 pp after one year to 4.5 pp after two years and 5.2 pp after three years). This behavior is due to the fact that, although the proportion of workers remaining in the same firm diminishes for both groups, it falls more intensively in the case of the treatment group. Although this *prima facie* result would point out to a relatively unfavourable impact of STW schemes on the stability of employment which is not only evident in the short run but also in the medium run, the negative correlation between STW participation and employment stability might not be reflecting causation. To answer formally the question on whether STW schemes keep jobs in the same firm and therefore they have a causal effect on the preservation of employment, we use PSM methods.

5.2 Matching estimates

Finding perfect matches between STW participants and nonparticipants on all relevant observable pre-treatment characteristics can be a rather complicated task. PSM methods involve matching individuals with similar likelihoods of exposure to STW schemes. The theory assures us, in principle, that the distribution of independent variables will be the same across cases with a given propensity score, even when values differ for a particular matched pair. The matching process models the probability of participation and matches individuals with similar propensity scores. Our estimate of $P(X)$ to participate in STW is based on a probit model, where the dependent variable is the treatment (T) regressed against a set of observable personal, job and firm characteristics which are not influenced by the employer decision to put an employee in STW work. The variables we have chosen include gender, nationality, age, job category, firm size, industry affiliation and region [8].

The estimated marginal effects of the probit model on the probability of STW participation (see Table 2) suggest that male, Spanish, older workers are more likely to be chosen to participate and that the job category, the industry affiliation and the employer size seem to play an important role in determining that probability. Thus, being employed in blue-collar medium-skilled jobs, in medium and large firms (at least 50 employees) and in heavy manufacturing increases that likelihood. The result on the effect of industry is substantial and reflects the fact that STW tends to be concentrated in the goods-producing sector and that incentives to hoard labor may be stronger there due to the greater importance of firm-specific skills [9].

One of the key for the validity of the PSM (estimated with a probit model) is the overlap or common support condition. The range of control variables included in the probit regression makes it likely that the outcome of the treated and control group, given the propensity score, differs only due to treatment and, hence, the CIA holds. To ensure that this probability is met, we have proceeded with a graphical analysis, plotting the propensity score distribution for the treated and control groups to look at the overlap that we have achieved. Results (not shown but available upon request) suggests that the overlap is quite good, although differences in the density of the propensity score can be observed in the tails of the distribution.

Another method to confirm the previous results is the min-max method to discard from our analysis the treated observations that are outside the common support region. This method involves finding the minima and maxima of the propensity score distribution for both the treated (0.0003306, 0.9445475) and control group (4.84e-07, 0.9579591) and defining the support region by selecting the highest of the two minima and the lowest of the two maxima (0.0003306,0.9445475). As can be seen, the region of the common support corresponds to the interval showing the distribution of the propensity score for the treated individuals. This means that we have a perfect overlap for the estimation of the ATET effect and that our estimates would be representative.

	dF/dX	S.E.	Sign
Gender (men)	0.030	0.004	***
Nationality (Spanish)	0.026	0.006	***
Age groups			
<30 years old	−0.107	0.005	***
30–39 years old	−0.063	0.005	***
40–49 years old	−0.040	0.004	***
≥50 years old	–	–	–
Job category			
WCHS and WCMS	–	–	–
WCLS	0.005	0.006	
BCHS	0.042	0.007	***
BCMS	0.069	0.010	***
BCLS	0.008	0.008	
Industry			
Manufacture of food products, textiles, wood and paper	−0.047	0.003	***
Extraction, energy, chemicals and manufacture of metals	−0.013	0.004	***
Manufacture of machinery, electrical and electronic or transport equipment	–	–	–
Construction (and rest of services)	−0.115	0.005	***
Trade and tourism	−0.097	0.005	***
Transport	−0.062	0.003	***
Region			***
Balearic Is., Canary Is., Andalusia, Ceuta, Melilla and Murcia	−0.035	0.004	***
Aragon, Navarre, La Rioja, Basque Country and Catalonia	–	–	–
Asturias, Galicia and Cantabria	−0.003	0.006	
Extremadura, Castile-La Mancha and Castile- Leon	0.022	0.009	***
Valencia			***
Madrid	−0.018	0.004	***
Firm size	−0.076	0.004	***
1–19 workers	−0.092	0.004	***
20–49 workers	−0.065	0.003	***
+50 workers	–	–	–
Log-likelihood		−3445.167	
Sample		13,953	
LR chi2(21)		5525.690***	
Pseudo R2		0.445	

Note(s): See the "Job category" classification in Table A1 of the Appendix. Significance levels: * 10%, ** 5%, *** 1%

Table 2. Propensity score estimation (probit model). Marginal effects. Spain (MCVL, 2009–2013)

In performing the PSM it is crucial to evaluate the quality of the matching. This means that we have to know if the balancing property is satisfied. If a good balance is achieved, the marginal distribution of each covariate is similar in the treated and control groups. There exist some methods to check the balancing assumption. One of the methods frequently used is the "standardized per cent bias" proposed by Rosembaun and Rubin (1985). This is the percentage difference of the sample means in the treated and control group (full or matched) subsamples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups. A reduction of the standardized bias after matching could prove that the covariates balance is improved by the matching. Rosembaun and Rubin (1985) consider a standardized difference in means with an absolute value lower than 20 as acceptable. Another method proposed by Sianesi (2004) consists in comparing the pseudo R2 before and after the matching. If the covariates after the matching are well balanced, the pseudo R2 should be low. For the calculus of both methods, we use the command *pstest* in

Stata. As a matter of example, we report the results after the estimation of a PSM radius caliper (0.01) in Table A2 of the Appendix. Results are quite similar for the alternative matching algorithms we use below.

Table A2 reports unmatched (U) and matched (M) means of the covariates for treated and control groups, the percentage of bias, and the *t*-test on the hypothesis that the mean value of each variable is the same in the treatment and control groups. The values of the pseudo R2 before and after the matching are displayed at the bottom of the table. The levels of standardized bias are very low, always lower than 20 (look at the % bias for *M*). The *t*-test statistic tests the hypothesis that the mean value of each variable is the same in the treatment group and the control group. It has been calculated before and after the matching. The result is that the null hypothesis cannot be rejected at the 5% of significance level after the matching for each variable. Finally, the pseudo R2 declines from a value of 0.445 for the unmatched sample to a value of 0.003 for the matched sample. In sum, all this information leads to conclude that the balance of the covariates and the quality of matching we have achieved are quite good [10].

Table 3 provides PSM estimates of STW participation effects using a variety of methods for creating comparison samples based on $P(X)$. The ATET has been estimated using a number of alternative algorithms to match participants and nonparticipants: nearest neighbor without replacement without caliper; nearest neighbor without replacement within caliper (0.01, 0.05, 0.1); nearest neighbor with replacement within caliper (0.1, 0.05, 0.1); one to one matching; kernel matching; multiple neighbours (5, 10, 15) and radius caliper (0.01, 0.05, 0.1). The outcome variable is the probability of remaining in the same firm in the future after the "treatment" (STW in 2009/Q2) has taken place. As we described above, this was done separately for each individual one, two and three years after participating in a STW scheme.

If STW had the assumed effect of preventing unemployment by keeping the participants employed relative to non-participants, there should be a positive (or null) difference between the two groups, so employment would fall significantly faster (or at the same rate) in the control group of workers than in the treated one. This result would show up as a positive (or null) ATET. On the contrary, what we observe with all matching estimators is a significantly negative difference in the employment level between the two groups.

The nearest neighbor estimator with no replacement gives a significant effect equal to a decrease of the job retention for participants in STW schemes relative to non-participants of nearly 6 pp lower one year later, about 10 pp lower two years later and 12 pp three years later. Kernel, radius and multiple neighbours matching provide quite similar results. When using the multiple neighbor algorithm, we obtain slightly different estimates depending on the number of neighbours used. Decreasing this number from 15 to 5 increases the estimated ATET for one and three years later between one and three pp. The latter estimates are closer to the larger ones obtained with the nearest neighbor algorithms without and with replacement: the first ones give the highest values for the effects two and three years later, while the second ones for the effects one year later. The estimates do not change when we vary the caliper from 0.1 to 0.01 over the nearest neighbor with or without replacement. Although the nearest neighbor estimations with replacement give higher values for standardized bias (maybe reflecting that this type of matching trades a reduced bias with an increase in variance), the estimates of the ATET continue being significant.

In sum, the estimate results indicate that the probability of remaining employed in the same firm for treated individuals involved in STW is significantly lower relative to other similar non-treated workers and decreases with the course of time.

5.3 Robustness checks: the role of unobserved factors and IV estimates

Our strategy is based on the assumption that outcomes are independent of the treatment once we control for measured characteristics and depends on the observable attributes available. This implies that all existing selection bias is assumed to be determined by the observable

	Sample	Treated	Controls	Difference	S.E.	*T*-stat
Nearest neighbor without replacement without caliper						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.884	−0.056	0.010	5.37
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.822	−0.099	0.012	8.04
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.766	−0.124	0.013	9.27
Note(s): On support: observations treated, 2277; observations control, 11676						
Nearest neighbor without replacement within caliper (0.01)						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.819	0.892	−0.074	0.013	5.68
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.706	0.831	−0.126	0.016	8.10
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.624	0.785	−0.161	0.017	9.63
Note(s): On support: observations treated, 1450; observations control, 11676						
Nearest neighbor without replacement within caliper (0.05)						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.818	0.892	−0.074	0.013	5.71
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.704	0.832	−0.127	0.015	8.24
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.623	0.786	−0.162	0.017	9.76
Note(s): On support: observations treated, 1460; observations control, 11676						
Nearest neighbor without replacement within caliper (0.1)						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.821	0.892	−0.072	0.013	5.57
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.701	0.832	−0.131	0.015	8.45
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.621	0.786	−0.164	0.017	9.88
Note(s): On support: observations treated: 1460; observations control, 11676						
Nearest neighbor with replacement within caliper (0.01)						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.935	−0.107	0.027	4.00
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.833	−0.111	0.034	3.28
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.778	−0.136	0.037	3.690
Note(s): On support: observations treated, 2276; observations control, 11676						
Nearest neighbor with replacement within caliper (0.05)						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.935	−0.107	0.027	4.00
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.834	−0.111	0.034	3.28
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.778	−0.136	0.037	3.67
Note(s): On support: observations treated, 2277; observations control, 11676						

(*continued*)

Table 3.
Propensity Score Matching: ATET of STW. Different algorithms. Spain (MCVL, 2009–2013)

	Sample	Treated	Controls	Difference	S.E.	T-stat
Nearest neighbor with replacement within caliper (0.1)						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.935	−0.107	0.027	4.00
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.834	−0.111	0.034	3.28
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.778	−0.136	0.037	3.67
Note(s): On support: observations treated, 2277; observations control, 11676						
1 to 1 matching						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.935	−0.107	0.027	4.00
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.834	−0.111	0.034	3.28
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.778	−0.136	0.037	3.67
Note(s): On support: observations treated, 2277; observations control, 11676						
Kernel matching						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.878	−0.050	0.013	3.72
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.809	−0.086	0.016	5.33
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.755	−0.113	0.018	6.44
Note(s): On support: observations treated, 2277; observations control, 11676						
Multiple neighbors N = 15						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.895	−0.067	0.017	3.83
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.820	−0.097	0.021	4.67
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.779	−0.137	0.023	6.03
Note(s): On support: observations treated, 2277; observations control, 11676						
Multiple neighbors N = 10						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
	ATT	0.828	0.883	−0.055	0.016	3.51
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
	ATT	0.723	0.807	−0.084	0.019	4.45
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
	ATT	0.642	0.753	−0.111	0.021	5.34
Note(s): On support: observations treated 2277; observations control, 11676						
Multiple neighbors N = 15						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
ATT	0.828	0.880	−0.052	0.015	3.44	
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
ATT	0.723	0.802	−0.079	0.018	4.34	
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
ATT	0.642	0.751	−0.109	0.020	5.48	
Note(s): On support: observations treated, 2277; observations control, 11676						

Table 3. (*continued*)

	Sample	Treated	Controls	Difference	S.E.	*T*-stat
Radius caliper = 0.01						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
ATT	0.828	0.887	−0.059	0.016	3.61	
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
ATT	0.723	0.799	−0.077	0.020	3.87	
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
ATT	0.642	0.744	−0.102	0.022	4.72	
Note(s): On support: observations treated, 2276; observations control, 11676						
Radius caliper = 0.05						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
ATT	0.828	0.877	−0.049	0.015	3.33	
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
ATT	0.723	0.802	−0.079	0.018	4.43	
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
ATT	0.642	0.747	−0.105	0.019	5.44	
Note(s): On support: observations treated, 2277; observations control, 11676						
Radius caliper = 0.1						
1 year later	Unmatched	0.828	0.859	−0.031	0.008	3.84
ATT	0.828	0.879	−0.051	0.014	3.75	
2 years later	Unmatched	0.723	0.768	−0.045	0.010	4.62
ATT	0.723	0.809	−0.087	0.016	5.28	
3 years later	Unmatched	0.642	0.694	−0.052	0.011	4.87
ATT	0.642	0.755	−0.113	0.018	6.37	
Note(s): On support: observations treated, 2277; observations control, 11676						

Table 3.

attributes used as covariates in the propensity score estimation. However, non-observable characteristics can play a role. Any characteristic that is associated with both program participation and the outcome measure, after conditioning on measured attributes, can induce bias (the so-called hidden bias). In matter of STW, unobservable characteristics might be potentially relevant. Since it is not possible to test directly that PSM estimates are free of hidden bias, a sensitivity analysis will be carried out in order to assess the robustness of our results to its failure. We do so by following two strategies. Firstly, the so-called Rosenbaum Bounds are used. Secondly, we perform estimates based on instrumental variables (IV) methods.

5.3.1 Using the Rosenbaum bounds. Here the idea is that the probability of being treated is a function of both observed and unobserved factors (Rosenbaum, 1987). If the uncounfoundedness assumption holds and there is no hidden bias, it implies that either the effect of the unobserved factors (γ) takes a value of zero, such that they have no effect on the probability of participation, or that the unobserved factors are the same across treated and untreated individuals. Assuming that the probability of being treated follows a logistic distribution, the only case where the treated and untreated individuals have the same probability is when $e^{\gamma} = 1$. In this case, the hidden bias is not present and the CIA holds. Higher values of e^{γ} will mean that there is hidden bias. Becker and Caliendo (2007) have devised a Stata routine which allows to implement this sensitivity analysis exploiting the Mantel-Haenszel test statistics (Q_{MH}). Rosenbaum (2002) shows that for values of $\gamma > 1$, the Q_{MH} test-statistics is bounded by two known distribution Q_{MH}^{+} and Q_{MH}^{-}. They represent the case in which the ATET has been overestimated and underestimated, respectively.

Table 4 reports the Q_{MH} bounds and their significance level for different values of e^{γ} and with three types of PSM estimators. The Q_{MH} bounds for $e^{\gamma} = 1$ show a scenario where the

•Nearest neighbor radius caliper 0.01				
Γ	Q_mh+	Q_mh–	p_mh+	p_mh–
1 year later				
1	3.829	3.829	0.000	0.000
1.25	7.497	0.198	0.000	0.421
1.5	10.551	2.705	0.000	0.003
1.75	13.192	5.224	0.000	0.000
2	15.535	7.423	0.000	0.000
2 years later				
1	4.606	4.606	0.000	0.000
1.25	8.969	0.284	0.000	0.388
1.5	12.594	3.194	0.000	0.001
1.75	15.722	6.189	0.000	0.000
2	18.492	8.802	0.000	0.000
3 years later				
1	4.859	4.859	0.000	0.000
1.25	9.543	0.214	0.000	0.415
1.5	13.429	3.530	0.000	0.000
1.75	16.777	6.749	0.000	0.000
2	19.736	9.555	0.000	0.000
•Nearest neighbor radius caliper 0.05				
γ	Q_mh+	Q_mh–	p_mh+	p_mh–
1 year later				
1	3.810	3.810	0.000	0.000
1.25	7.479	0.180	0.000	0.429
1.5	10.533	2.724	0.000	0.003
1.75	13.173	5.243	0.000	0.000
2	15.516	7.442	0.000	0.000
2 years later				
1	4.587	4.587	0.000	0.000
1.25	8.950	0.265	0.000	0.396
1.5	12.575	3.213	0.000	0.001
1.75	15.703	6.209	0.000	0.000
2	18.473	8.822	0.000	0.000
3 years later				
1	4.839	4.839	0.000	0.000
1.25	9.524	0.194	0.000	0.423
1.5	13.410	3.551	0.000	0.000
1.75	16.757	6.770	0.000	0.000
2	19.717	9.577	0.000	0.000
•Nearest neighbor radius caliper 0.1				
γ	Q_mh+	Q_mh–	p_mh+	p_mh–
1 year later				
1	3.810	3.810	0.000	0.000
1.25	7.479	0.180	0.000	0.429
1.5	10.533	2.724	0.000	0.003
1.75	13.173	5.243	0.000	0.000
2	15.516	7.442	0.000	0.000

(*continued*)

Table 4.
Sensitivity to the presence of hidden bias

•Nearest neighbor radius caliper 0.1				
γ	Q_mh+	Q_mh–	p_mh+	p_mh–
2 years later				
1	4.587	4.587	0.000	0.000
1.25	8.950	0.265	0.000	0.396
1.5	12.575	3.213	0.000	0.001
1.75	15.703	6.209	0.000	0.000
2	18.473	8.822	0.000	0.000
3 years later				
1	4.839	4.839	0.000	0.000
1.25	9.524	0.194	0.000	0.423
1.5	13.410	3.551	0.000	0.000
1.75	16.757	6.770	0.000	0.000
2	19.717	9.577	0.000	0.000

Table 4.

estimated ATET is free of hidden bias: they are both statistically significant with all the algorithms used. Higher values of e^{γ} represent the effect an unobserved factor would have to produce on the odds of being treated in order to explain away the estimated ATET. ${Q_{MH}}^{+}$ is always significant and in the case of ${Q_{MH}}^{-}$ our results are only sensitive to the existence of an observed factor which would increase the probability of being treated by 25% but not 50% or 100%. Therefore, we can conclude safely that our estimates are quite robust to the existence of hidden bias.

5.3.2 IV methods. If it is more likely that individuals with certain observable or unobservable characteristics are chosen for participation, the estimation of a model that does not take into account this endogeneity would bias the effects of STW participation on subsequent employment. In order to avoid this bias, we employ an IV approach, setting out a model which involves estimating an equation for STW program participation in addition to an equation that indicates the labor market outcomes of individuals.

Both equations can be estimated simultaneously using a bivariate probit model for two binary outcomes. The aim of forming a consistent estimator for the impact of STW on the probability of remaining within the same employer in the future relies on the ability to construct instruments for the variable indicating exposure to a STW scheme. In the estimation of this model, participation is instrumented using two variables: whether the employer was involved in a STW scheme prior to the starting of the observation period in 2009/Q2, and whether the worker received unemployment benefits prior to the participation in STW. The first variable tries to take account of the fact that the first important factor explaining the individual probability of being in STW is the employer decision of implementing such a programme. Here we follow the approach used by Boeri and Bruecker (2011). The second variable may be proxying relevant aspects that firms and workers' representatives take into account when they have to choose which workers will participate in a given STW arrangement: labor costs, expectations and labor hoarding, and fairness (see Scholz, 2012).

The IV estimate results (shown in Table 5) are in line with the PSM estimates obtained with the algorithm of the nearest neighbor without replacement, suggesting that the job retention would be 10 pp lower one year later, about 16 pp lower two years later and 18 pp three years later for participants in STW schemes relative to non-participants. Therefore, they confirm our previous findings.

6. Conclusions and discussion

This article has investigated whether STW schemes are successful in their objective of maintaining employment. For that, Spanish longitudinal administrative data has been used and a worker-level approach adopted, examining whether STW programs achieved to keep workers employed within their firms. Applying PSM methods, we have estimated the treatment effect of STW as the ATET, the outcome variable being the individual probability of remaining working with the same employer.

Our main findings are the following. First, STW schemes do not bring about the expected effect of preserving jobs (retaining workers) in the short run (within one year after implementation), although the effect is relatively small, since treated individuals are about 5 pp less likely to remain working with the same employer one year later than similar workers. This means that there is a significantly negative difference in the job retention between the two groups, that is a negative treatment effect, so that STW programs would not have the assumed effect of preventing unemployment by keeping the participants employed relative to non-participants. Second, workers participating in STW arrangements are significantly less likely to remain in employment with the same employer two and three years after their participation when compared with similar non-participating workers. Thus, the negative effect of participation increases over time.

The empirical strategy followed and the results obtained in this article are not fully comparable to previous studies that use either country-level or firm-level data. This literature examines the effect of either the aggregate STW take-up rate or the companies' participation in STW schemes on (the change of) employment at either the nation-wide or firm level. The small positive or null effects encountered in general on either unemployment or employment do not rule out the possibility that a portion of STW workers are subsequently laid off, that is they would be compatible with negative impacts of STW schemes on workers' retention within firms. Therefore, as our analysis is based on the comparison of the employment trajectories of participant and non-participant workers in firms that have used STW arrangements, our findings are more related to the almost non-existent literature investigating the impact of STW on individual workers. While some authors (Calavrezo and Lodin, 2012) have found similar negative effects as ours (although restricted to a very short period of time: one-quarter), others report either null effects (Speckesser, 2010) or significantly positive effects (Pavlopoulos and Chkalova, 2019) for STW participants as compared to similar non-participants.

Table 5. Estimate results of IV recursive bivariate probit models: the impact of STW on the probability of remaining within the same employer one, two and three years later. Marginal effects. Spain (MCVL, 2009–2013)

	1 year later	2 years later	3 years later
STW participation (*D*)	−0.080***	−0.154***	−0.178***
	(0.014)	0.017	(0.020)
Wald chi2(46)	7532.20	8963.38	10061.74
Prob > chi2	0.000	0.000	0.000
Individuals	11,676	11,676	11,676

Note(s): The estimated marginal effects (standard errors) in each column come from a different regression on two equations: one equation for STW program participation and another for the labor market outcomes of individuals. For each regression, the estimates of an indicator variable for exposure to a STW scheme on the equation for labour market outcomes are reported. Additional explanatory variables include personal, job and employer attributes (gender, nationality, age, industry, region, job category, job tenure and firm size). Worker's participation on STW is instrumented using two variables: whether the employer was involved in a STW scheme prior to the starting of the observation period in 2009/Q2, and whether the worker received unemployment benefits prior to the participation in STW. Standard errors are shown in brackets. Significance levels: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

These differential results may be due to various reasons. One relevant issue to be considered is that the assessment of STW schemes refers to their effects during a recession that differed in intensity and duration among countries. Several studies (Hijzen and Martin, 2013; Kruppe and Scholz, 2014) stress that STW arrangements are helpful and have an economic impact on preserving jobs during a downturn if lack of work is transitory. Nevertheless, if the downturn continues much longer, the flexibility buffers exhaust, with unavoidable layoffs consequently. In the case of the 2008 crisis, as it turned out to be V-shaped for some countries and U-shaped for others, the fast recovery may have favoured the effectiveness of some STW schemes (such as the German or the Swiss one), while the prolonged recession may have affected negatively the impact of others (such as the Spanish one). Thus, the result we obtain might be related to the duration of the economic downturn.

Another factor to take into account is how companies use this work-sharing device with their workforce. Firms could use STW arrangements to protect their core jobs, thus inducing differences among workers on the likelihood of remaining within an employer. However, it could also happen that, as certain jobs and workers are less likely to survive when a recession hits the economy, companies used STW programs to try to preserve jobs in the short-run that are not economically viable and leave workers that hold less productive jobs in STW, something that would help them stabilize employment and protect their core employees during crises, bringing about "displacement effects". Further research should shed more light on the potential relationship between usage of STW schemes and the skill content of jobs.

Despite our findings, it cannot be assumed that the policy recommendation is not to use STW arrangements as a way to reduce unemployment. There are at least three reasons for that. One is methodological. Even though the results seem to be quite robust to various sensitivity analyses, such as the use of different matching algorithms and estimation strategies, the endogeneity and unobserved heterogeneity that plagues this sort of studies may also be present in our analyses. This would imply, for instance, that workers (and jobs) who were chosen for STW may have some unobservables (not captured by our methods) that make them more prone to be laid off in the future; or firms with greater financial difficulties are more likely to take advantage of the STW incentives, so the financial and economic problems (rather than the STW program) lead to more layoffs in the future.

The second reason is related to the institutional framework. Labour market institutions and employment policies (employment protection legislation, systems of collective bargaining, etc.) vary across countries, so similar STW schemes may bring about not only different STW take-up rates, but also different employment and unemployment outcomes. While the first issue has been previously analyzed (see Boeri and Bruecker, 2011; Hijzen and Martin, 2013; Lydon *et al.*, 2019), the second remains unexplored and merits investigation.

And the third reason has to do with the welfare effects and efficiency outcomes associated with STW programs. In this sense, policymakers should be aware of all the costs associated with the use of STW schemes (not only fiscal costs but also deadweight and displacement costs). As far as we know, only Siegenthaler and Kopp (2019) have conducted a cost–benefit analysis of a STW program (the Swiss one) and concluded that the savings from reduced unemployment benefit payments was large enough to fully compensate the payments related to STW. This limited empirical evidence calls for an analysis which provides further evidence on the efficiency of STW schemes and the resources devoted to them during the recessions, so further research should be addressed.

Notes

1. Take-up rates increased substantially at the beginning of the recession; they rose from near zero in the pre-crisis period to over 1% of dependent employment (more than 4.5 million workers involved) in 2009, at the peak of the downturn, amounting to 7% in Belgium, 4–5% in Germany and Japan,

1–2% in Austria, France, Italy, Czech Republic, the Netherlands and Slovakia, and less than 1% in other countries in 2009 (Hijzen and Martin, 2013). At the same time the resources devoted to them were substantial; in 2009 expenditure amounted to between 0.1% and 0.3% of GDP in Germany, Italy and Japan (see Boeri and Bruecker, 2011).

2. Herzog-Stein *et al.* (2017) attribute the German employment miracle mainly to strong temporary working-time reductions (not only short-time work but also working-time accounts and discretionary reductions of regular working hours) by cushioning about 40% of the output shock. For some nuances on this perspective, see Brenke *et al.* (2013) and Balleer *et al.* (2016, 2017). Antosiewicz and Lewandowski (2017) perform contrafactual simulations to study the factors behind cyclical fluctuations and differences in adjustments to economic shocks in Greece, Italy, Portugal and Spain, taking the German economy as the yardstick. They find that those countries would have fared better if their economies responded to shocks more similarly to the German economy.

3. In a companion paper (Arranz *et al.*, 2018), we focus our attention on a reform on STW incentives, examining whether this change had an effect on the probability of remaining in employment, so we used treatment and control groups before and after the reform and estimated difference-in-differences regressions and instrumental variable bivariate probit models with endogenous covariates. Quite differently, the current paper focuses on the difference that participating in STW may have on the stability of employment, so treatment and control groups are used and Propensity Score Matching (PSM) models estimated. Therefore, here we try to evaluate the programme itself and not the effect of a specific reform.

4. Hijzen and Venn (2011) consider the following dimensions: (1) work-sharing requirements: they specify the range of permissible reductions in weekly hours for short-time workers; (2) eligibility requirements: they set conditions that employers or workers must meet in order to participate in STW programs; (3) conditionality requirements: they set behavioural requirements for both employers and workers; (4) generosity: it determines the cost of participation for both firms and workers and the maximum length of participation (for firms this depends on the extent to which they are required to share in the cost of hours not worked, while for workers this depends on the extent to which they are compensated for hours not worked). Subsequent studies follow this characterisation.

5. We follow the empirical strategy developed by Rubin (1974) and Rosenbaum and Rubin (1983).

6. The estimation of the propensity score and the matching procedure has been conducted using the Stata module *psmatch2* developed by Leuven and Sianesi (2003). This strategy has also been used by Kruppe and Scholz (2014) in the context of STW.

7. Arranz *et al.* (2013) describe a process through which the original information of the different files of the administrative data source can be organised in such a way as to permit the accurate study of work histories, while Arranz *et al.* (2014) use the MCVL to document the importance of recalls in labour market transitions and estimate a duration model with competing risks of exits in order to investigate the individual, job and firm attributes that influence the probabilities of leaving unemployment to return to the same employer or to find a new job. In both studies, the authors explain in detail the characteristics (and pros and cons) of the MCVL database.

8. We have included several interaction terms (between the job category and the age groups) as well in order to help eliminate differences between treatment and control groups (see Smith and Todd, 2005; Caliendo and Kopeinig, 2008). The effects of these interactions have never been significant.

9. In this vein, it is a well-known fact that large companies in manufacturing (especially in sectors such as the automotive sector, the metalworking industry, and the textile, clothing and leather industry) are the heaviest users of STW schemes in an economy and their employees are overrepresented in the group of workers participating in STW (Eurofound, 2010; Boeri and Bruecker, 2011; Calavrezo and Lodin, 2012).

10. The Rubin's R and Rubin's B are reported at the bottom of Table A2. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the matched samples to be considered sufficiently balanced. In our case they lie within these ranges (look at M in Table A2).

References

Abraham, K. and Houseman, S. (1994), "Does employment protection inhibit labor market flexibility? Lessons from Germany, France, and Belgium", in Blank, R. (Ed.), *Social Protection versus Economic Flexibility: Is There a Trade-off?*, University of Chicago Press, Chicago, pp. 59-94.

Antosiewicz, M. and Lewandowski, P. (2017), "Labour market fluctuations in GIPS–shocks vs adjustments", *International Journal of Manpower*, Vol. 38 No. 7, pp. 913-939.

Arpaia, A., Curci, N., Mejermans, E., Peschner, J. and Pierini, F. (2010), "Short time working arrangements as response to cyclical fluctuations", *European Economy Occasional Paper*, Nº 64, European Commission.

Arranz, J.M., García-Serrano, C. and Hernanz, V. (2013), "How do we pursue 'labourmetrics'? An application using the MCVL", *Revista Estadística Española*, Vol. 55 No. 181, pp. 231-254.

Arranz, J.M. and García-Serrano, C. (2014), "The interplay of the unemployment compensation system, fixed-term contracts and rehirings: The case of Spain", *International Journal of Manpower*, Vol. 35 No. 8, pp. 1236-1259.

Arranz, J.M., García-Serrano, C. and Hernanz, V. (2018), "Short-time work and employment stability: evidence from a policy change", *British Journal of Industrial Relations*, Vol. 56 No. 1, pp. 189-222.

Balleer, A., Gehrke, B., Lechthaler, W. and Merkl, Ch. (2016), "Does short-time work save jobs? A business cycle analysis", *European Economic Review*, Vol. 84, pp. 99-122.

Balleer, A., Gehrke, B. and Merkl, Ch. (2017), "Some surprising facts about working time accounts and the business cycle in Germany", *International Journal of Manpower*, Vol. 38 No. 7, pp. 940-953.

Becker, S.O. and Caliendo, M. (2007), "Sensitivity analysis for average treatment effect", *Stata Journal*, Vol. 7 No. 1, pp. 71-83.

Bellman, L., Gerner, H. and Upward, R. (2015), "The response of German establishments to the 2008-2009 economic crisis", in Commendatore, P., Kayam, S. and Kubin, I. (Eds), *Complexity and Geographical Economics: Topics and Tools*, Springer, Vol. 19, pp. 165-207, Dynamic Modelling and Econometrics in Economics and Finance.

Boeri, T. and Bruecker, H. (2011), "Short-time work benefits revisited: some lessons from the Great Recession", *Economic Policy*, Vol. 26 No. 68, pp. 697-765.

Brenke, K., Rinne, U. and Zimmermann, K. (2013), "Short-time work: the German answer to the great recession", *International Labour Review*, Vol. 152 No. 2, pp. 287-305.

Cahuc, P. and Carcillo, S. (2011), "Is short-time work a good method to keep unemployment down?", *Nordic Employment Policy Review*, Vol. 1 No. 1, pp. 133-169.

Calavrezo, O. and Lodin, F. (2012), "Short-time working arrangements in France during the crisis: an empirical analysis of firms and employees", *Comparative Economic Studies*, Vol. 54, pp. 299-320.

Calavrezo, O., Duhautois, R. and Walkoviak, E. (2009), *The Sort-Time Compensation Program in France: An Efficient Measure against Redundancies?*, CEE, Working Paper No. 114.

Calavrezo, O., Duhautois, R. and Walkoviak, E. (2010), *Short-time Compensation and Establishment Survival: An Empirical Analysis with French Data*, IZA, Discussion Paper No. 4989.

Caliendo, M. and Kopeinig, S. (2008), "Some practical guidance for the implementation of propensity score matching", *Journal of Economic Surveys*, Vol. 22 No. 1, pp. 31-72.

Crimmann, A., Wiesner, F. and Bellman, L. (2012), "Resisting the crisis: short-time work in Germany", *International Journal of Manpower*, Vol. 33 No. 8, pp. 877-900.

Dietz, M., Stops, M. and Walwei, U. (2011), *Safeguarding Jobs in Times of Crisis: Lessons from the German Experience*, International Labour Office, Discussion Paper no. 207.

Duhautois, R., Walkoviak, E. and Calavrezo, O. (2009), "The substitution of worksharing and short-time compensation in France: a difference-in-differences approach", *Economics Bulletin*, Vol. 29 No. 2, pp. 820-833.

Eurofound (2010), *ERM Report 2010. Extending Flexicurity – the Potential of Short-Time Working Schemes*, European Foundation for the Improvement of Living and Working Conditions, Luxembourg, Publications Office of the European Union.

Fitzroy, F.R. and Hart, R.A. (1985), "Hours, layoffs and unemployment insurance funding: theory and practice in an international perspective", *Economic Journal*, Vol. 95 No. 379, pp. 700-713.

Hall, R.E. (1995), "Lost jobs", *Brookings Papers on Economic Activity*, Vol. 26 No. 1, pp. 221-273.

Heckman, J.J., LaLonde, R. and Smith, J. (1999), ""The economics and econometrics of active labor market programs", in Ashenfelter, O. and Card, D. (Eds), *Handbook of Labor Economics*, North-Holland, Amsterdam, Vol. 3, pp. 1865-2097.

Herzog-Stein, A., Lindner, F. and Sturn, S. (2017), "The German employment miracle in the Great Recession: the significance and institutional foundations of temporary working-time reductions", *Oxford Economic Papers*, Vol. 70 No. 1, pp. 206-224.

Hijzen, A. and Martin, S. (2013), "The role of short-time work schemes during the global financial crisis and early recovery: a cross-country analysis", *IZA Journal of Labor Policy*, Vol. 2 No. 5, pp. 1-31.

Hijzen, A. and Venn, D. (2011), *The Role of Short-Time Work Schemes during the 2008-09 Recession*, OECD Publishing, OECD Social, Employment and Migration Working Papers, Nº 115.

Imbens, G. (2004), "Nonparametric estimation of average treatment effects under exogeneity: a review", *The Review of Economics and Statistics*, Vol. 86 No. 1, pp. 4-29.

Kato, T. and Kodama, N. (2019), *The Consequences of Short-Time Compensation: Evidence from Japan*, IZA, Discussion Paper No. 12596.

Kruppe, T. and Scholz, T. (2014), *Labour Hoarding in Germany. Employment Effects of Short-Time Work during the Crises*, IAB, Discussion Paper, 17/2014.

Leuven, E. and Sianesi, B. (2003), *"PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing", Statistical Software Components S432001*, Boston College Department of Economics, (accessed 02 May 2017), Boston, Massachusetts.

Lydon, R., Mathä, T.Y. and Millard, S. (2019), "Short-time work in the great recession: firm-level evidence from 20 EU countries", *IZA Journal of Labor Policy*, Vol. 8 No. 2, pp. 1-29.

Mueser, P.R., Troske, K.R. and Gorislavsky, A. (2007), "Using State administrative data to measure program performance", *The Review of Economics and Statistics*, Vol. 89 No. 4, pp. 761-783.

OECD (2010), *Moving beyond the Jobs Crisis*, OECD Publishing, Paris, Chapter 1, Employment Outlook.

Panteia (2012), *Short-time Working Arrangements during the Crisis and Lessons to Learn*, Report for the European Commission DG Employment.

Pavlopoulos, D. and Chkalova, K. (2019), "Short-time work: a bridge to employment security or a springboard to unemployment?", First Published December 11, *Economic and Industrial Democracy*. doi: 10.1177/0143831X19890674.

Rosen, S. (1985), "Implicit contracts: a survey", *Journal of Economic Literature*, Vol. 23 No. 3, pp. 1144-1175.

Rosenbaum, P.R. (1987), "The role of a second control group in an observational study", *Statistical Science*, Vol. 2, pp. 292-316.

Rosenbaum, P.R. (2002), *Observational Studies*, NY: Springer, New York.

Rosenbaum, P.R. and Rubin, D.B. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, Vol. 70, pp. 41-55.

Rosenbaum, P.R. and Rubin, D.B. (1985), "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score", *The American Statistician*, Vol. 39 No. 1, pp. 33-38.

Rubin, D.B. (1974), "Estimating causal effects of treatments in randomised and non-randomised studies", *Journal of Educational Psychology*, Vol. 66, pp. 688-701.

Rubin, D.B. (2001), "Using propensity scores to help design observational studies: application to the tobacco litigation", *Health Services and Outcomes Research Methodology*, Vol. 2, pp. 169-188.

Scholz, T. (2012), *Employers' Selection Behavior during Short-Time Work*, IAB-Discussion Paper 18/2012.

Sianesi, B. (2004), "An evaluation of the Swedish system of active labour market programmes in the 1990s", *Review of Economics and Statistics*, Vol. 86 No. 1, pp. 133-155.

Siegenthaler, M. and Kopp, D. (2019), *Short-time Work and Unemployment in and after the Great Recession*, KOF Swiss Economic Institute, ETH Zurich, KOF Working papers 19-462.

Smith, J. and Todd, P.E. (2005), "Does matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of Econometrics*, Vol. 125 Nos 1-2, pp. 305-353.

Speckesser, S. (2010), *Employment Retention in the Recession: Microeconomic Effects of the Short-Time Work Programme in Germany*, Westminster Business School, University of Westminster, London.

Van Audenrode, M. (1994), "Short-time compensation, job security, and employment contracts: evidence from selected OECD countries", *Journal of Political Economy*, Vol. 102 No. 1, pp. 76-102.

Corresponding author
Carlos García-Serrano can be contacted at: carlos.garcia@uah.es

Appendix

	Control group		Treatment group	
	Mean	Std. Dev	Mean	Std. Dev
Gender (men)	0.602	0.489	0.798	0.401
Nationality (Spanish)	0.947	0.223	0.974	0.160
Age groups				
<30 years old	0.349	0.477	0.126	0.332
30–39 years old	0.371	0.483	0.357	0.479
40–49 years old	0.204	0.403	0.294	0.456
≥50 years old	0.076	0.266	0.222	0.416
Job category				
WCHS and WCMS	0.209	0.407	0.152	0.359
WCLS	0.371	0.483	0.148	0.356
BCHS	0.214	0.410	0.342	0.474
BCMS	0.109	0.312	0.303	0.459
BCLS	0.097	0.296	0.055	0.229
Industry				
Manufacture of food products, textiles, wood and paper	0.081	0.273	0.108	0.311
Extraction, energy, chemicals and manufacture of metals	0.080	0.272	0.381	0.486
Manufacture of machinery, electrical and electronic or transport equipment	0.062	0.241	0.333	0.471
Construction (and rest of services)	0.310	0.463	0.055	0.229
Trade and tourism	0.332	0.471	0.080	0.271
Transport	0.134	0.340	0.042	0.201
Region				
Balearic Is., Canary Is., Andalusia, Ceuta, Melilla and Murcia	0.152	0.359	0.067	0.250
Aragon, Navarre, La Rioja, Catalonia and Basque Country	0.358	0.479	0.609	0.488
Asturias, Galicia and Cantabria	0.053	0.224	0.088	0.284
Extremadura, Castilla-La Mancha and Castilla-Leon	0.045	0.207	0.078	0.269
Com. Valenciana	0.109	0.311	0.119	0.323
Madrid	0.283	0.450	0.039	0.194
Firm size				
1–19 workers	0.387	0.487	0.087	0.282
20–49 workers	0.159	0.365	0.044	0.205
+50 workers	0.454	0.498	0.869	0.337

(*continued*)

Table A1. Summary statistics (means): treatment and control groups. Spain (MCVL, 2009–2013)

	Control group		Treatment group	
	Mean	Std. Dev	Mean	Std. Dev
Tenure				
1-<3 years	0.113	0.317	0.058	0.233
3-<6 years	0.202	0.402	0.148	0.355
6-<10 years	0.386	0.487	0.342	0.474
10-<20 years	0.257	0.437	0.387	0.487
>20 years	0.041	0.197	0.066	0.248
Individuals (sample)	11,676		2,277	

Note(s): "Job category" is classified as White-collar high-skilled occupations, WCHS (managers, workers with university degree, technical engineers and qualified assistants); White-collar medium-skilled occupations, WCMS (clerical and workshop heads and assistants); White-collar low-skilled occupations, WCLS (administrative officials and other clerical workers); Blue-collar high-skilled occupations, BCHS (first and second class officials); Blue-collar medium-skilled occupations, BCMS (third class officials and specialists) and Blue-collar low-skilled occupations, BCLS (laborers)

Table A1.

	Sample	Mean Treated	Mean Control	% Bias	% Reduction |Bias|	t-test t	t-test $p > t$
Gender (men)	U	0.79842	0.60218	43.8	–	17.99	0
	M	0.79833	0.79269	1.3	97.1	0.47	0.637
Nationality (Spanish)	U	0.97365	0.9475	13.5	–	5.33	0
	M	0.97364	0.97147	1.1	91.7	0.45	0.654
Age groups							
<30 years old	U	0.12648	0.34858	−54.1	–	−21.25	0
	M	0.12654	0.13645	−2.4	95.5	−0.99	0.322
30–39 years old	U	0.35749	0.37093	−2.8	–	−1.22	0.224
	M	0.35764	0.36088	−0.7	75.9	−0.23	0.82
40–49 years old	U	0.29425	0.20418	20.9	–	9.54	0
	M	0.29438	0.27583	4.3	79.4	1.39	0.166
Industry							
Manufacture of food products,	U	0.10848	0.08128	9.3	–	4.24	0
textiles, wood and paper	M	0.10852	0.10251	2.1	77.9	0.66	0.509
Extraction, energy, chemicals and	U	0.3812	0.08025	76.5	–	41.49	0
manufacture of metals	M	0.38093	0.39127	−2.6	96.6	−0.72	0.474
Construction and rest of services	U	0.05534	0.31029	−69.9	–	−25.69	0
	M	0.05536	0.05769	−0.6	99.1	−0.34	0.734
Trade and tourism	U	0.07993	0.33222	−65.6	–	−24.77	0
	M	0.07996	0.07246	2	97	0.95	0.34
Transport	U	0.04216	0.13378	−32.8	–	−12.43	0
	M	0.04218	0.04916	−2.5	92.4	−1.13	0.26
Region							
Balearic Is., Canary Is., Andalusia,	U	0.06675	0.15245	−27.7	–	−10.88	0
Ceuta, Melilla and Murcia	M	0.06678	0.07394	−2.3	91.7	−0.94	0.346
Asturias, Galicia and Cantabria	U	0.08827	0.0531	13.8	–	6.53	0
	M	0.08831	0.09127	−1.2	91.6	−0.35	0.727
Extremadura, Castilla-La Mancha	U	0.07817	0.04505	13.8	–	6.62	0
and Castilla-Leon	M	0.07777	0.08254	−2	85.6	−0.59	0.554
Com. Valenciana	U	0.11858	0.1086	3.1	–	1.39	0.164
	M	0.11863	0.10504	4.3	−36.2	1.45	0.146
Madrid	U	0.03909	0.28289	−70.3	–	−25.37	0
	M	0.0391	0.03976	−0.2	99.7	−0.11	0.91
Job category							
WCLS	U	0.14844	0.37059	−52.4	–	−20.87	0
	M	0.14851	0.13276	3.7	92.9	1.53	0.127
BCHS	U	0.34168	0.21394	28.8	–	13.24	0
	M	0.34183	0.33497	1.5	94.6	0.49	0.625
BCMS	U	0.30259	0.10928	49.2	–	24.78	0
	M	0.30272	0.31249	−2.5	94.9	−0.71	0.475

(*continued*)

Table A2.
Unmatched (U) and matched (M) means of the covariates for treated and control groups, percentage of bias, and t-test on the hypothesis that the mean value of each variable is the same in the treatment and control groups. Spain (MCVL, 2009–2013)

	Sample	Mean Treated	Mean Control	% Bias	% Reduction |Bias|	t-test t	t-test $p > t$
BCLS	U	0.05534	0.09704	−15.8	–	−6.36	0
	M	0.05492	0.04254	4.7	70.3	1.94	0.052
Firm size							
1–19 workers	U	0.08696	0.38703	−75.4	–	−28.48	0
	M	0.08699	0.08663	0.1	99.9	0.04	0.965
20–49 workers	U	0.04392	0.15853	−38.7	–	−14.53	0
	M	0.04394	0.04177	0.7	98.1	0.36	0.718

Sample	Pseudo R2	LR chi2	p > chi2	Mean Bias	Med Bias	B	R
Unmatched	0.445	5525.69	0.00	37.1	32.8	209.7*	0.68
Matched	0.003	20.87	0.467	2	2	13.6	1.06

Note(s): B is Rubin's B (the absolute standardized difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group) and R is Rubin's R (the ratio of treated to (matched) non-treated variances of the propensity score index)

Table A2.