Temporary paid furlough in the midst of an unprecedented sanitary crisis: evidence from the Spanish reality

E. Congregado, J. Garcia, N. Rubino

December 2021

Abstract

This paper analyzes the average effect of furlough policies in Spain. Ideally divided in two parts, the paper presents an average treatment effect analysis of the ERTE ("Expediente de Regulación Temporal de Empleo" $\approx$ Temporary Paid Furlough) at the national level. Spanning the first three quarters of 2020, we show, everything else equal, that the propensity to be re-employed was significantly higher in the treated (Furlough granted group) than in the control group. In the second part of the paper, we look upon regional heterogeneity to uncover more of the relationship between re-employment probability and furlough participation. Our results show slight differences in the magnitude of the effect related to quarterly aggregation, but overall prove the existence of a positive re-employment effect at the regional level everywhere in the country and in every economic sector.

- JEL: **TO BE ANNOUNCED**
- Keywords: Furlough, short-time work, ERTE, matching techniques, propensity score analysis, Spain.

Resumen

Este trabajo analiza el efecto medio de la política de Expedientes de Regulación Temporal de Empleo (ERTE) en España. Dividido en dos partes, primero se presenta un análisis sobre el efecto medio del tratamiento de los ERTE a nivel nacional. Estudiando los tres primeros trimestres de 2020, observamos que, ceteris paribus, la probabilidad de volver a una situación de empleo fue significativamente más alta en el grupo tratamiento (acogidos a ERTE) que en el grupo control. En la segunda parte del artículo, buscamos la heterogeneidad de este efecto a nivel regional para así entender mejor la relación entre la probabilidad de vuelta al empleo y la adopción del ERTE. Nuestros resultados muestran ligeras diferencias en la magnitud del efecto, sin embargo, prueban la existencia de un efecto positivo sobre la probabilidad de reincorporarse como empleado en todas las regiones del país y sectores económicos.

Key findings

- We found a robust and positive average effect of furlough schemes in the probability of being re-employed at the short-term.
- This positive re-employment effect was widespread at the regional level everywhere in the country and in every economic sector.
- However, although still significant, this effect lessens at the medium-term, thus, when the furlough scheme was held in time for two consecutive quarters.

Resultados clave

- Se ha encontrado un efecto medio significativo y positivo de los esquemas de ERTE sobre la probabilidad de retornar a situaciones de empleo a corto plazo.
- Este efecto positivo habría sido generalizado para todas las regiones y sectores del país.
- Sin embargo, a pesar de seguir siendo significativo, dicho efecto se reduce para el medio plazo, esto es, al considerar esquemas de ERTE sostenidos en el tiempo a lo largo de dos trimestres consecutivos.

Contents

List of Figures

List of Tables

1. Introduction

The Covid-19 outbreak has caused an unprecedented sanitary crisis all over the world, forcing the governments to implement restrictive measures as mandatory lockdown and social distancing. Concerned about a boost in unemployment digits, most countries made a massive use of job retention schemes[1] as a way to temporarily protect the employees' positions meanwhile the labor market were adjusting to the shock. Along this paper, we will study the Spanish case where this scheme is known as "Expediente de Regulación Temporal de Empleo" or just "ERTE", using preferably the furlough terminology when referring to it. This policy, consists in a temporary suspension of the labor relationship between the employer and the employee, or alternatively, a reduction of working hours justified by a major cause. This cause must be related to economic, technical, organizational or production issues, including Covid related consequences from March, 2020. During this period of suspension, the employee is getting a social security allowance while the employer only has to assume a social contribution, which is a minor part of the employee's wage. As a result, it works as a transitory mechanism of flexibility to adjust the labor market, whose cost is essentially assumed by the public administration. This policy main purpose is to maintain the employees' position despite not being working, avoiding the sharp boost of unemployment during the shock.

Although this mechanism was available before the pandemic, the Spanish ERTE were only widely used then, covering around 3 million of workers (more than 20% of the affiliated workers) in the second quarter of 2020. The following quarters it covered around 5% of the affiliated workers, which is still a significantly higher proportion than it was during the previous recession. Moreover, since the Spanish government first approved these Covid-19 related ERTEs in March, 2020, their expiration date has been postponed several times, remaining in the current legislation. Therefore, some evaluation of the impact of this policy in all dimensions is urgently needed to improve the design of these programs for the next future.

To our knowledge, not much literature has managed to establish a direct link between cut off strategies at work and its effect on follow-up outcomes in the labor market. In the literature related to the effect of short-time work schemes (Furloughs) during times of crisis, Hizjen and Martin (2013) point out that timing might be crucial as the positive net effect of Furloughs might be nonlinear with respect to subsequent re-employment and job creation, so much so that the positive causal relationship might turn, in due time, to negative. With some external validity granted by the methodological approach used by the authors, who create a panel of 23 OECD countries ranging from the first quarter of 2004 to the last of 2010, our article tackles the timing issue by considering a robustness check which extends the ERTE period of analysis by aggregating the second and third quarters of 2020.

Osuna and García-Pérez (2015) focused, as we do, on the Spanish case, but find out that the intensity of subsidies to payroll taxes, which ultimately fall on the workers as a direct salary reduction, might amplify or dampen re-employment and job creation/destruction. In line with previous literature findings, the authors confirm the positive causal effect of short-time work on re-employment and document a reduction of the degree of segmentation of labor markets in Spain. As in many other countries where new reforms where introduced to allow for a more flexible labour market, able to deflate wages and adjust working hours at a faster pace, salary losses compensated by law allowed for the causal linkage between Furloughs and Employment to stay positive during the global financial crisis. Arranz et al. (2018) focus on the participation rate extending the analysis to one more GIIPS country other than Spain, Italy. As they establish a comparison between the short-time work scheme take-up rates of the 1990s crises compared to the 2000s crises, the authors look at the workers' profiles to establish which categories are more willing to accept being furloughed. They show that, everything else equal, if the labor markets in the 2000s had the same workforce composition of those same markets in the 1990s, we would have assisted to a higher take-up rate due to the higher weight of older, less

[1] In this context, furlough schemes and short-time work are the most common terms to refer to these job retention schemes. A detailed description about these schemes an their use across Europe can be found in Drahokoupil and Müller (2021).

educated workers, suggesting that policy making should indeed focus on adjusting new short term work policies with a very close eye on the evolution of workforce composition.

Recently, Cahuc et al. (2021) has proved that the impact of Short time work in France has indeed been necessary to save jobs and balance-out strong asymmetric shocks suffered by the firms more exposed to the crisis, furthermore showing that, although hampered by the existence of windfall effects, furloughs prove to be still more cost-efficient at saving jobs than any kind of subsidy. In general, whatever the case of the context, the current general agreement on short terms work schemes and policy intervention is that furloughs, under certain conditions, might have a positive net effect on the economy (Cahuc (2019)) and be of great help in the immediate proximity of the current covid-19 crisis not just in strict market terms but also in terms of potential welfare gains, given the appropriate reforms (Müller and Schulten (2020)).

On the sociological/managerial side, Hufflman et al. (2021) have recently proposed a model of intentions based on the well or ill placed perception of fairness of the employees, affecting turnover choices by the employees at the end of the furlough. On a similar note, Escudero-Castillo and al. (2021) have recently stressed the impact of furlough, smart-working and other contingent measures during the first period covid 19 lockdown in Spain. Their findings based on probability estimates show that furloughed and unemployed persons self-perception of safety and well being was significantly smaller when compared to the share of the population formally employed at the time of the confinement. On the broader side, (unpaid) cut offs during the coronavirus outbreak have been found to affect more firmly women and racial minorities than the average white Caucasian male in the United States (Dias (2021) and Fan (2020)).

As far as we are concerned, most recent contributions in literature appear to have much focused on the health related effects of layoffs and other cut off measures, leaving the matter of deciding what consequential effect such policy tools would have on immediate future market outcomes almost untouched. Furthermore, as layoffs and furloughs are two different policy tools, current literature, has been indeed biased towards observing potential negative consequences of layoffs, but has not offered much yet in furlough studies. Our study fills up this gap by carrying out a national and regional analysis of the effect of ERTEs ("expediente de regulación temporal de empleo", which could be roughly translated with furlough) on re-employment probability in Spain across 2020. In order to do so, Spanish microdata have been filtered to derive a database of salaried workers who have been matched in order to calculate the average effect of being on ERTE on follow-up labor market outcome to prove whether or not being furloughed increases or decreases the likelihood of successful employment. To complete the analysis, we match the propensity score tests with a weighted regional level analysis to look for the marginal effect of the ERTE on employment probability in a logistic model.

2. Data

Data are drawn from the quarterly flow microdata 2020/1 to 2020/4 of the Spanish Labor Force Survey (henceforth SLFS). The Survey, conducted by the National Statistical Institute (henceforth INE), is a large household sample survey providing results on labor participation of people aged 16 and over as well as persons outside the labor force in which each sampled individual remaining in the survey for a period of six quarters at a time, with no resampling after individuals are rotated out of the sample.

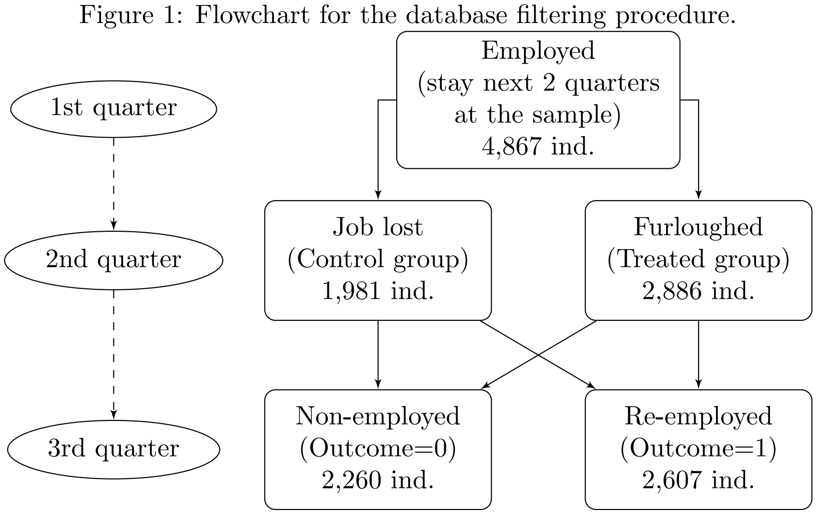
The Survey is targeted at a rotating sample of around 60,000 households throughout the national territory. For every household member, both socioeconomic and labor information is collected in order to summarize the main characteristics of the Spanish workforce each quarter. Individuals in the sample are interviewed for six consecutive quarters, thus we have information on quarterly labor transitions for a maximum period of 18 months for each individual in the sample.[2]

[2]For further methodological information or data access visit INE website.

As we look for a way to rearrange our data for our initial matching analysis, we choose to identify observations and participants by selecting those that would consecutively appear in the first three quarters of 2020, have been getting paid work during the first quarter of 2020, and can be identified and divided for treatment along the second quarter, considering those who had lost their job during such period and those who were full-time furloughed.[3] A binary outcome was finally generated to identify the objective variable for the third quarter of 2020, pointing at those who would have been working during that period non-conditional on the fact that the furlough period might have ended up with the individual reincorporated in its former job or finally getting a new one. The flowchart displayed in figure 1 illustrates the data selection procedure employed for the matching analysis. Therefore, in our final database for this analysis each observation represents an individual who stay at least the 3 initial quarters of 2020 at the sample, was employed in the 1st quarter, was either furloughed (treatment group) or displaced (control group) in the 2nd quarter, and whose employment status was observed in the 3rd quarter (outcome=1 if he/she have been re-employed, outcome=0 otherwise, e.g. no job or furloughed). Then, we have an identifier for each individual, a treatment dummy indicating the furloughed in 2nd quarter, an outcome dummy indicating the re-employment status in the 3rd quarter, and the observable pre-treatment characteristics of each individual, thus, taken their values in the 1st quarter. Note the database keep a cross-section structure because each observation represents a single individual with their 1st quarter characteristics and the time dimension was only used for the treatment assignment and the outcome generation. Hence, once both variables were created, the time dimension was dropped.

Finally, population weights included in the SLFS for the representative interview were not included in the matching analysis, due to obvious limitation in processing power. In the second part of the analysis, as we focus on a single step logistic regression, weights were included at the regional level but did not change the marginal effects of the probability results in any significant way.

Figure 1: Flowchart for the database filtering procedure.



3. Matching analysis

3.1. Methodology

The methodology we employ to look for an average treating effect is propensity score matching. This tech uses as identifying tools, granting the ceteris paribus condition, a number of observable and

[3]Note in the control group we are considering the 2nd quarter displaced workers, no matter whether they were unemployed or inactive that quarter. Since lockdown and containment measures may have significantly hampered the employment active seeking process we made no distinction between both labor states, as suggested by the Spanish Central Bank (BdE) in its latest report (BdE Economic Buletin, 2/2020).

measurable control variables capable of capturing all the relevant differences between groups. Our first step entails outcome creation: as we define:

$$\begin{aligned} y_\alpha &= \text{value of the outcome for the furloughed} \\ y_\omega &= \text{value of the outcome for the non-furloughed} \end{aligned} \tag{1}$$

so that :

$$y = \{.y_\alpha \; if \quad D = 0 y_\omega \; if \quad D = 1 \tag{2}$$

the value of the average effect of the treatment we aim to calculate and make inference on will be $ATT = y_\alpha - y_\omega$.

To talk in terms of expectations, the average treatment effect for the furloughed group will thus be:

$$ATT = E(y_\alpha - y_\omega | D = 1) = e(y|D = 1) - e(y|D = 1) \tag{3}$$

so that the average treatment effect for the control group, in analogy, would thus be:

$$ATNT = E(y_\alpha - y_\omega | D = 0) = e(y|D = 0) - e(y|D = 0) \tag{4}$$

The difference between the ATT and $ATNT$, which measures the changes in magnitude of the effect of furloughs on employment as we jump from the furloughed group to the non-furloughed group, is the measure we are interested in. Our exogeneity pre-assumption will be based on the concept of conditional independence: once a set of observable variables able to capture all possible forms of heterogeneity has been identified, our identifying assumption, for the ATT, will imply that once the control group X is chosen, the results of the outcome for the unfurloughed group (the results for the counterfactual group) will be the same for both the treated and the untreated:

$$E[y_\omega | X, D = 1] = E[y_\omega | X, D = 0] \tag{5}$$

Similarly, and assuming once more the matrix of controls X has been identified correctly, we would want to see that the values of the real outcome variable, once conditioned for the control matrix X, are indeed the same for both the furloughed and unfurloughed workers:

$$E[y_\alpha | X, D = 1] = E[y_\alpha | X, D = 0] \tag{6}$$

One last preliminary condition, necessary for the comparison and average impact estimates to make sense, is the existence of a set of homogeneous counterfactual/individuals couples. Any matching procedure essentially requires that the ceteris paribus condition is respected by associating though different techniques individuals with their almost identical counterfactuals in a given neighborhood of values. An essential way to define the concept of common support has to do with the non existence of a full probability set for any given characteristic inside matrix X. In other words, for the ATT, the common support requirement implies that for any observable X_i the proportion of furloughed individuals with a specific value of that characteristic should always be less than 1. The absence of such condition would imply an empty set for the untreated, and as such the absence of counterfactual for that specific characteristic X_i that would immediately bias the estimated values.

$$ATT : P(D = 1 | X_{i=1,2,3,n}) = P(D = 1 | X) < 1 \tag{7}$$

In a similar fashion, but considering that we need a lower bound for the treated to avoid all workers from being untreated, the common support condition for the average treatment effect for the non furloughed will be:

$$ATNT : P(D = 1|X_{i=1,2,3,n}) = P(D = 0|X) > 0 \tag{8}$$

Joining the two conditions, and considering the average treatment effect, the overall common support condition will thus imply:

$$ATE : 0 < P(D = 1|X) < 1 \tag{9}$$

The last step required is the propensity score matching calculation. The propensity score represents the probability that an individual might be part of the furloughed group given he shows given values of an observable X_i. The values thus represents the proportion of the treated with that given value of the defining characteristic $X_i \implies p(X) = P(D - 1|X = X_i)$. The propensity score is thus calculated though a logit with the full set of X_i as controls:

$$P(D = 1|X) = \phi(\delta' X) \tag{10}$$

Similar values of the propensity score, according to discretional proximity criteria, are then used to match furloughed workers from the treatment group with unfurloughed workers from the control group. The matching finally allow to calculate a comparable result, that is an estimated outcome of those untreated neighbors which have been coupled with the i^{th} treated person:

$$\widehat{y_i} = \sum_{j \in C^0(p_i)} w_{ij} y_j \tag{11}$$

the comparable result $\widehat{y_i}$ is thus defined as the geometric sum of real outcome values of the j^{th} neighbor in the $C^0(p_i)$ set of untreated neighbors of individual i. The set will be bounded, as $\sum_{j \in C^0(p_i)} w_{ij} = 1$ and the different matching methods we have employed offered different definitions of the neighbor set given the way weights have been defined. Once the estimated value of the outcome variable for the counterfactual have been estimated, the only step left to calculate the ATT will thus be the simple average of the difference between actual outcome values and the aforementioned estimated outcome values inside the common support bounds:

$$ATT = \frac{1}{\sum_{k=1}^{n}(D_k = 1 \cap C^0(p_i))} \sum_{i \cap \{D_i = 1 \in C^0(p_i)\}} (y_i - \widehat{y_i}) \tag{12}$$

We are looking, in other words, for an average calculated on a difference embodying the treatment effect. To define it disjointedly between the treatment and control groups:

$$ATT = E(Y|\textit{treated in common support}) - E(Y|\textit{non-matched treated/weighted}) \tag{13}$$

The $ATNT$ will instead be calculated as:

$$ATNT = \frac{1}{\sum_{k=1}^{n}(D_k = 0 \cap C^0(p_i))} \sum_{i \cap \{D_i = 0 \in C^0(p_i)\}} (y_i - \widehat{y_i}) \tag{14}$$

The difference between the values in Equation 12 and Equation 14 will ultimately give us the desired differential effect. In the following Sections, we shall offer a complete overview of our results, together with the necessary post-estimation checks and inference.

The matching procedure has been executed based on the individual pre-treatment characteristics, hence using the 1st quarter values. The set of observable controls selected for the propensity score calculation can be classified in two categories: social-demographics and labor conditions. For the social-demographic dimension we considered a set of standard controls as sex, age, region (Spanish Autonomous Community), educational stage, marital status, and nationality. On the other hand, the labor economic dimension comprise the industry or economic activity where the individual is employed (by the Spanish National Classification of Economic Activities), their professional situation (i.e. public servant, private employee, etc), the contract duration (either temporary or permanent) and the type of working day (either part or full-time).

4. National propensity score matching results

4.1. Preliminary results for the National level outcome analysis, single quarter analysis

In this section, results from the matching procedure will be shown to illustrate how the average treatment effect stands out as being statistically significant, favoring ERTE as a mean of re-employment. In this exercise, we evaluate the transition from a state of furlough/unemployment to a state of employment on the basis of the first three quarters of 2020. Our first preliminary look at the data starts with a simple smoothing baseline approach as we choose to pair each treated individual with up to his/her ten nearest neighbors with a caliper of 0.01. We also allow for replacement. Table 1 and Figure 2 show the difference between the unmatched sample and the matched individuals in the common support. Figure 3 highlights the difference between the propensity score distribution across all possible values for both the Furloughed and the Unfurloughed groups. Figure 4 and Figure 5 shows the estimates of the densities of the propensity score for the furloughed and the unfurloughed before and after the matching procedure (whole sample vs. common support). Figure 6, equivalently, shows a box and whisker plot of the distribution of the propensity score after completion of the matching procedure, highlighting quartiles. Finally, Table 2 shows the estimates for the average effect on the furloughed, the average effect on the unfurloughed, and most importantly the total differential effect (with both single estimates and bootstrapped ones).

As for these preliminary estimates, two facts can be clearly stated: first of all, the average treatment effect on the difference between Furloughed and Unfurloughed Personnel is a positive quantity, where the average effect on the treated of the furlough almost doubles its effect in terms of reemployability with respect to the matched Unfurloughed group. To add to the robustness of the result, Table 2 the estimated for the bootstrapped differential are even sharper, showing standard errors a tenfold lower than the naive calculations (from 0,250 to 0,021). The magnitude of the differential, when compared to the unmatched statistic, is perhaps slightly higher but never the less significant both after and before matching. This both suggests the goodness of fit of this initial matching exercise and the general validity of the covariates. To infer on the robustness of the matching and the (joint) validity of the instruments, Figures 5 and 4 show the score kernels before and after matching. Most strikingly, the two distributions appear to be almost identical after the procedure. This last result is further enhanced by the bias reduction plot in Figure 2: after matching, the average bias is clearly reduced and less dispersed around an average more or less equal to 0.[4]

[4]The cautious reader might have understood that an average bias reduction does not necessarily imply that every single control in every single cross section led to a bias reduction. As a further indication of this, the Likelihood Ratio test in Table 1 for the matched analysis did not reject the joint null any better than in the umatched analysis. This, however, a calculated risk entailed by having to resort to a high number of covariates to check for any sort of observable heterogeneity, exposing the analysis to the risk of "overcontrolling " it.

Table 1: Propensity score matching, quality check.

Sample	*Pseudo* R^2	*LR* χ^2	$p_{¿}\chi^2$	*Mean bias*	*Median Bias*	*B*	*R*
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VII)
Umatched	0,295	1828,08	0,000	13,8	9,3	142,500*	0,460*
Matched	0,009	71,92	0,003	2,8	2,6	22,700	1,080

Values of the pseudo correlation coefficient and related tests on the joint hypothesis of non-significance of the control variables. Mean and median bias changes in column (V) and (V). Rubin's B and R in Columns (VII) and (VIII). Suggested values for B and R: B less than 25, R between 0.5 and 2.

Figure 2: Standardized Bias with k=10 neighbors, caliper=0.1, replacement.

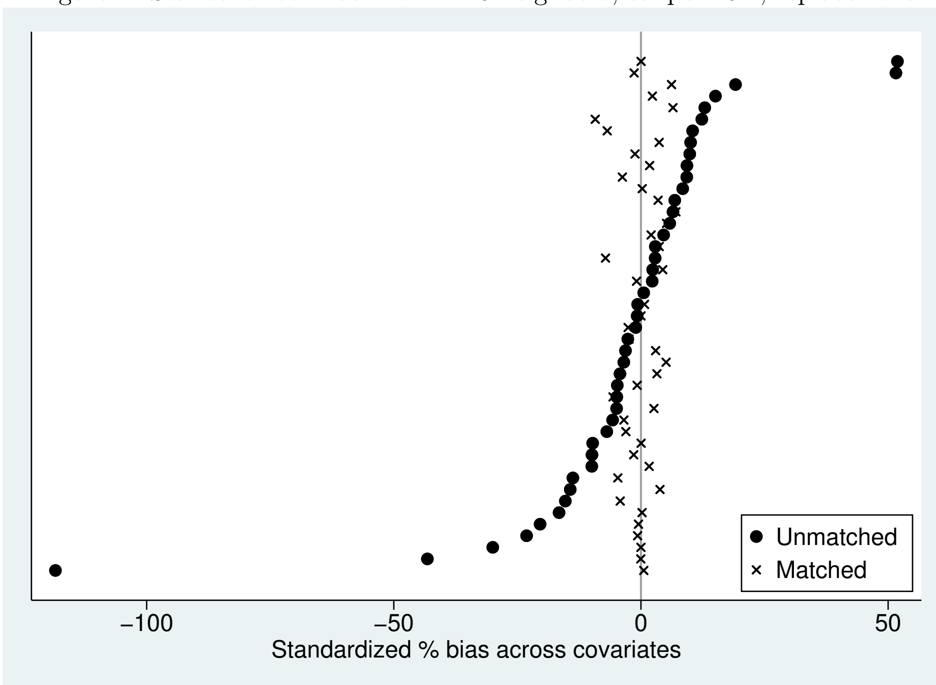


Figure 3: Bar plot of the overall distribution of the propensity score for the treatment and the control groups. Replacement.

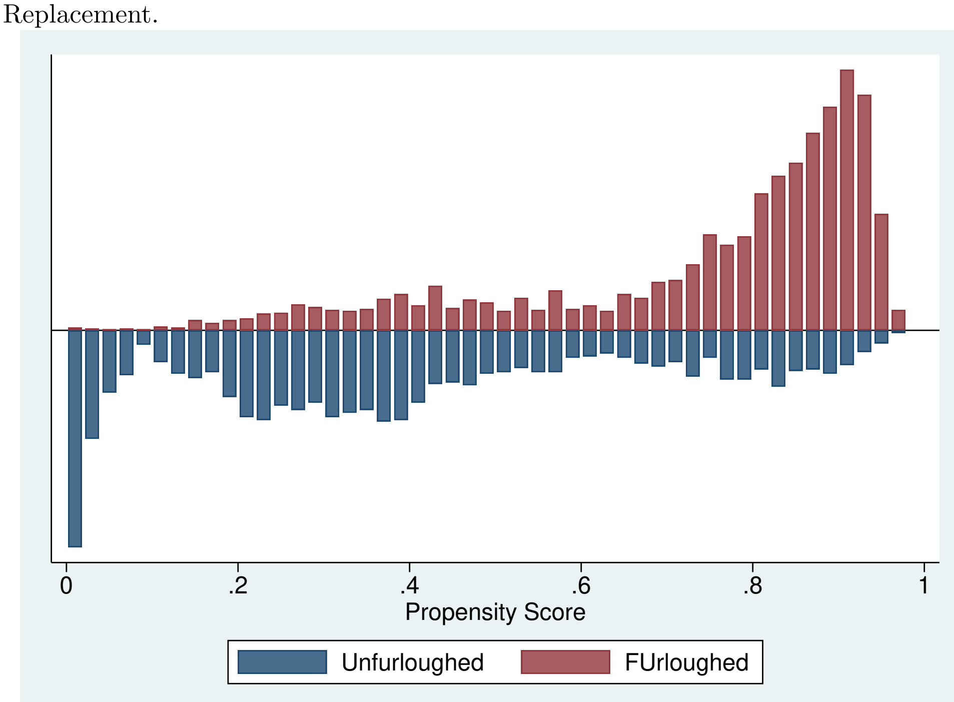


Figure 4: Densities of the treatment and control groups before matching (whole sample).

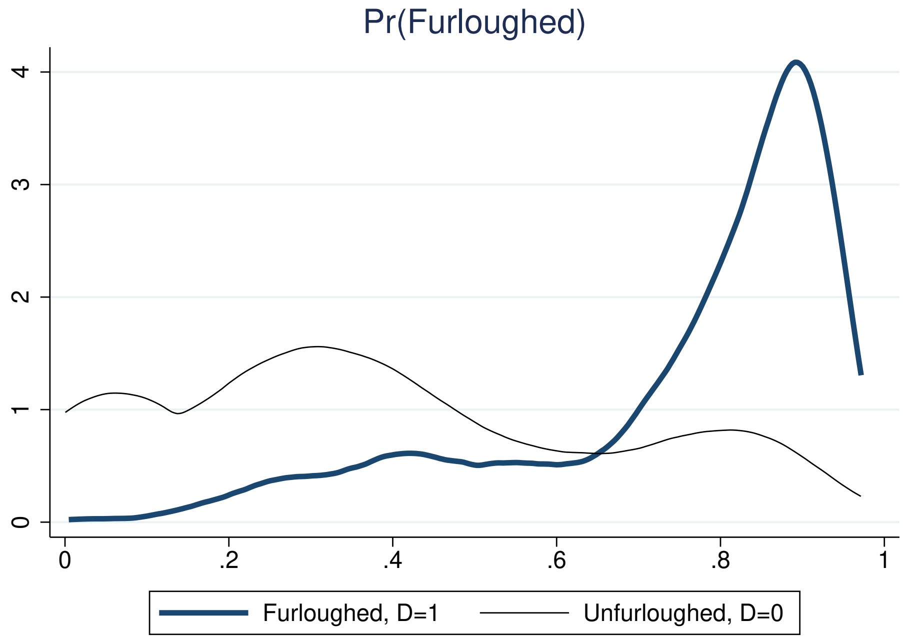


Figure 5: Densites of the treated and untreated inside the common support (k=10, caliper=0.01).

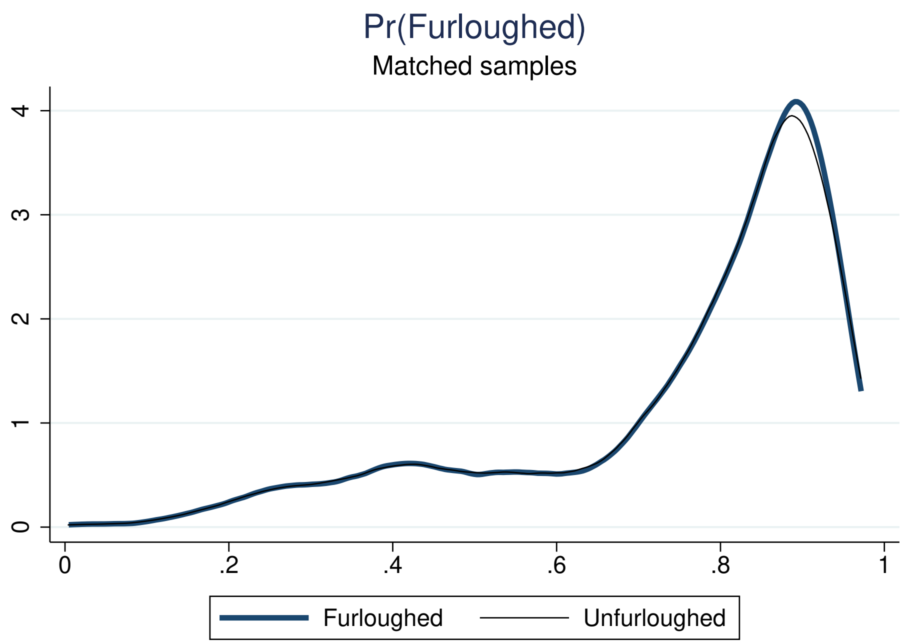
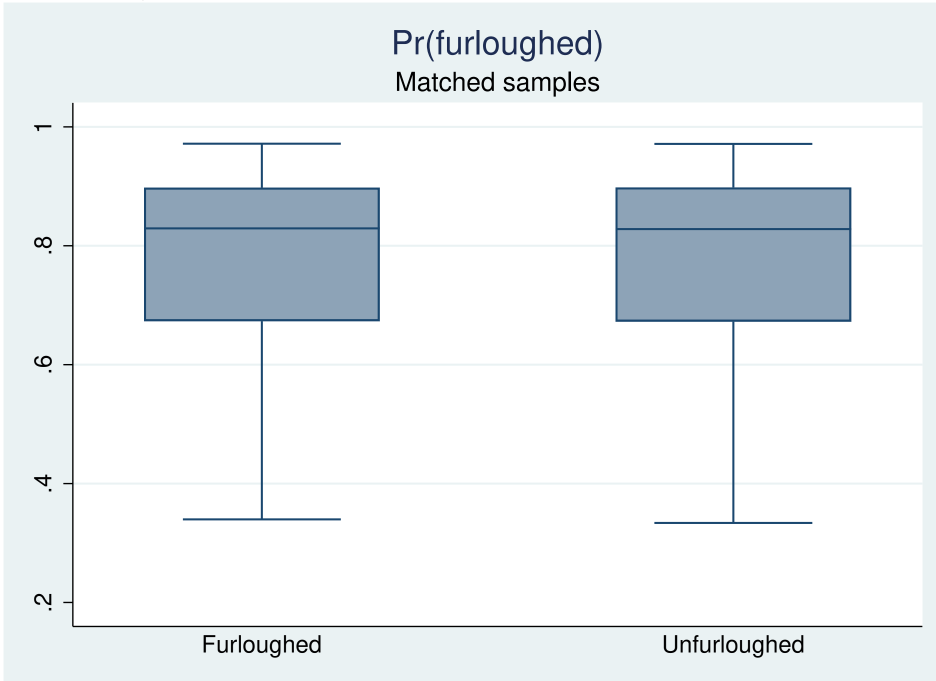


Table 2: Propensity score matching, estimates.

Matching results	*Furloughed*	*Unfurloughed*	*Difference*	*S.E.*	*T-stat (BS Z)*
(I)	(II)	(III)	(IV)	(V)	(VI)
Unmatched	0,645	0,368	0,277	0,015	19,16*
ATT-ATNT-ATE	0,645	0,358	0,287	0,025	11,47*
Bootstrapped ATE			0,287	0,021	13,72*

Estimates of the average treatment effects. Bootstrapped values for the ATE have been replicated fifty times

Figure 6: Box and Whiskers plot, distribution of the propensity scores across treatment and control groups after matching).



4.2. Robustness I: no replacement

In order to prove the robustness of the results previously seen, this section presents and compares an alternative matching choice to the benchmark we already established with the k=10, 1% caliper approach with replacement. We need to point out that, in the benchmark case, the loss in terms of comparable individuals was basically 0. However, a trade off exists when matching techniques with replacement are compared, everything else equal (width of the caliper, numerosity of the common support), to their no-replacement counterpart. Although replacement ensures lower bias and higher matching quality, since the distance between propensity scores is minimal as the best optimal choices from the control group are not systematically ruled out of the matching procedure, variances (and thus standard error) end up being comparatively higher with respect to the no-replacement option. That naturally happens since the information set is smaller (as ”far away individuals ” are never selected given the replacement mechanism), leading us to smaller precision and higher uncertainty for close to zero results in the estimates of the average effects. On the other side of the coin, avoiding replacement reduces variances as all the available information is being employed, but naturally worsen the quality of the matching as less likely individuals are paired by their less similar propensity scores. All in all, we expect the loss in terms of unmatched individuals to be non-negligible when no-replacement is present, but hope results stay robust, if not in terms in magnitude, at least in terms of sign when compared to the benchmark case. The results reported in this sections thus refer to the k=1, 1% caliper, no replacement case.[5]

A first glimpse at the outcome of the matching procedure can be seen in Figure 7. When compared to Figure 3, we observe how the no replacement option, together with the 1% caliper, leaves out of the pairing procedure around 1800 Furloughed workers from upper regions of the propensity score, too many when compared to their comparable equivalent in the Unfurloughed group. Besides the significant loss (1800 unavailable treated individuals over a total of 2814 treated individuals), we expect the ratio we are left with (1014 treated compared to 1811 untreated individuals) to be enough to reduce the variance of the estimates (even if slightly) while preserving the magnitude of the estimates.

[5]As the simplest way to implement no replacement is in the case where matching is done one to one.

Figure 7: Bar plot of the overall distribution of the propensity score for the treatment and the control groups. No replacement.

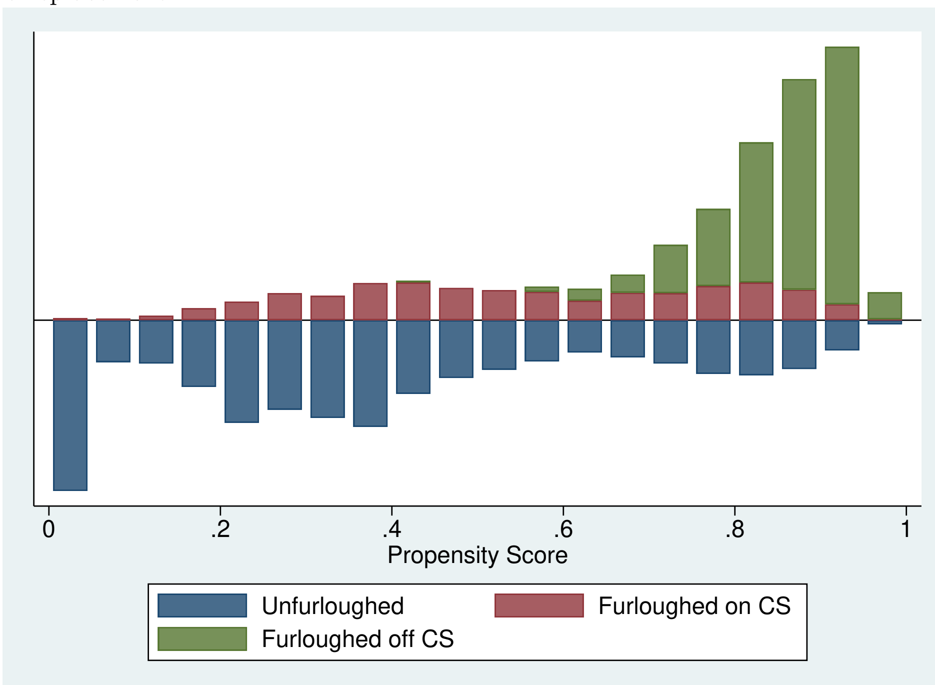


Table 3: Propensity score matching, quality check, no replacement.

Sample	*Pseudo* R^2	*LR* χ^2	$p¿\chi^2$	*Mean bias*	*Median Bias*	*B*	*R*
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VII)
Umatched	0.295	1828.080	0.000	14.0	8.5	142.5*	0.46*
Matched	0.007	18.360	0.999	2.4	2.0	19.1	1.00

Values of the pseudo correlation coefifcient and related tests on the joint hypothesis of non-significance of the control varaibles. Mean and median bias changes in column (V) and (V). Rubin's B and R in Columns (VII) and (VIII). Suggested values for B and R: B less than 25, R between 0.5 and 2.

Table 4: Propensity score matching, no replacement.

Matching results	*Furloughed*	*Unfurloughed*	*Difference*	*S.E.*	*T-stat (BS Z)*
(I)	(II)	(III)	(IV)	(V)	(VI)
Unmatched	0.645	0.368	0.277	0.014	19.16*
ATT-ATNT-ATE	0.627	0.385	0.243	0.022	11.26*
Bootstrapped ATE	0.243			0.017	14.54*

Estimates of the average treatment effects. Bootstrapped values for the ATE have been replicated fifty times

Table 3 reports the post-estimation checks. All values related to the goodness of the matching appear to suggest the stability of the procedure, with a very low matched R^2, a strong non-rejection of the contemporaneous null of the coefficients of the covariates in the Log-likelihood ratio test, reduced bias and within-limits values for the parameters B and R. Estimates, reported in Table 4, stay true to the value found n the benchmark case: the bias reduction is present although minimal (Standard errors decrease by around 0.01) but this just strengthens results which appeared already precise to begin with (both in the simple and the bootstrapped estimates). All in all, the positive differential implied by the ATE, favoring the idea that Furloughs might have a positive impact on re-employment opportunities, still stands.

4.3. Robustness II: higher caliper and kernel estimates

Following a reasoning not far away from the one we introduced in the last section, we now produce results based on a different discretional choice of the caliper. We remind that the reason why a caliper can be imposed on this kind of discretional procedures is that some treated individuals could be very far away from the closest treated individual. That would imply a reduction in the matching the more untreated individuals are paired with the same treated one. Having a caliper is a way to ensure the existence of a common support interval, but the lower the value of it, the higher the chances to leave some individuals outside of the estimates. Although our k=10 neighbors and 1% caliper did extraordinarily well in junction with replacement (no Furloughed individuals were left out), the discretional choice of the width of the analysis might cast a shadow on the (internal) validity for the results. That is why we tested This section presents an alternative robustness exercise where a higher caliper of 10% is used alongside the usual k=10 and k=1 choice and replacement and no replacement alternatively. Table 5 reports the numerosity of the treatment and effect groups conditional on decreasing values of the caliper. The number of the treated individuals inside the commons support and eligible for matching stayed relatively higher than the number of untreated individuals up until a equivalent to 0.05% of the overall sample size. Only when the value reached 0.01 the proportion of Furloughed to Unfurloughed inside the common support collapsed (767 om 1811). Estimates of the average effect, however, remained close to the benchmark case in every single attempted estimation.[6]

Table 5: Common support, absoulte terms

Group	*Off Support*	*On Support*	*Total*	*Caliper*
(I)	(II)	(III)	(IV)	(V)
Unfurloughed		1811	1811	1%
Furloughed		2814	2814	1%
Unfurloughed		1811	1811	0.5%
Furloughed	8	2806	2814	0.5%
Unfurloughed		1811	1811	0.1%
Furloughed	266	2548	2814	0.1%
Unfurloughed	0	1811	1811	0.05%
Furloughed	768	2046	2814	0.05%
Unfurloughed	0	1811	1811	0.01%
Furloughed	2047	767	2814	0.01%

Common Support Numerosity according to caliper variation.

One last step is perhaps missing in our robustness check: the possibility to allow for heterogeneous weights in the matching procedure. That is, instead of averaging out the k propensity scores from the matched untreated, we want to resort to a method that creates a comparable result as a weighted average based on some function. Kernel densities can thus be used so that the comparable result (the propensity score associated with the i^{th} treated) is the weighted averaged of the propensities of the untreated neighbors with weights negatively proportional to the distance between propensity score of the i^{th} treated and the j^{th} matched untreated individual. The more far away the untreated, the less

[6]Results for estimates based on lower than 0.01% calipers are available upon request.

its contribution to the result. In mathematical terms, the calculated comparable result $\widehat{y_i}$, given p_i, p_j and h, (the propensity score of the ith treated, the propensity score of the jth treated, and the bandwidth of the kernel respectively) will thus be:

$$\widehat{y_i} = \frac{\sum_{j\in\{d=0\}} K(\frac{p_i-p_j}{h})y_j}{\sum_{j\in\{d=0\}} K(\frac{p_i-p_j}{h})} \tag{15}$$

with $\widehat{w}_{ij} = \frac{K(\frac{p_i-p_j}{h})y_j}{\sum_{j\in\{d=0\}} K(\frac{p_i-p_j}{h})}$ representing the contribution of the weight of the jth untreated on the comparable result associated to the ith treated. We anticipate to the reader that the outcome of the matching remained basically unchanged, when compared to the benchmark case regardless of the approximated shape of our kernel density (normal, biweight, tricubic, epanechnikov), given a fixed chosen bandwidth of 0.1, the only difference being the fact that the off support treatment group could not be considered an empty set anymore, with just a single individual out of the common support (!). The last discretional choice we did put to test was the bandwidth length. As larger values of h imply a more platicurtic distribution while lower values imply a narrower one, we tried to experiment on the results given by the normal kernel density on a similar way we did with the simple smoothing: that is increasingly halving its value. Our results stayed absolutely in line with the baseline case, with just a slight decrease in the precision of the estimates (the standard error of the ATE differential went from 0.024 to 0.030 when we went from h=0.01 to h=0.0001) and a handful more of unmatched treated individuals (from 1 to 11 in total).[7] Results did not change radically when we dropped those treated individuals whose score was lower than the lowest and higher than the highest score from the untreated group (widest common support). Estimates based on a weighted matching with a normal kernel, bandwidth equal to 0.0001 and outlier exclusion (*Wide CS=No*) are available in Table 6.

Table 6: Propensity score matching, normal kernel.

Wide CS	*Matching results*	*Furloughed*	*Unfurloughed*	*Difference*	*S.E.*	*T-stat (BS Z)*
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Yes	*Unmatched*	0.645	0.368	0.277	0.014	19.16*
Yes	*ATT-ATNT-ATE*	0.645	0.384	0.261	0.030	8.66*
No	*Unmatched*	0.645	0.368	0.277	0.014	19.16*
No	*ATT-ATNT-ATE*	0.644	0.384	0.261	0.030	8.66*

Estimates of the average treatment effects. Wide CS represents outlier exclusion from above the maximum propensity of the untreated and below the minimum propenstiy of the untreated.

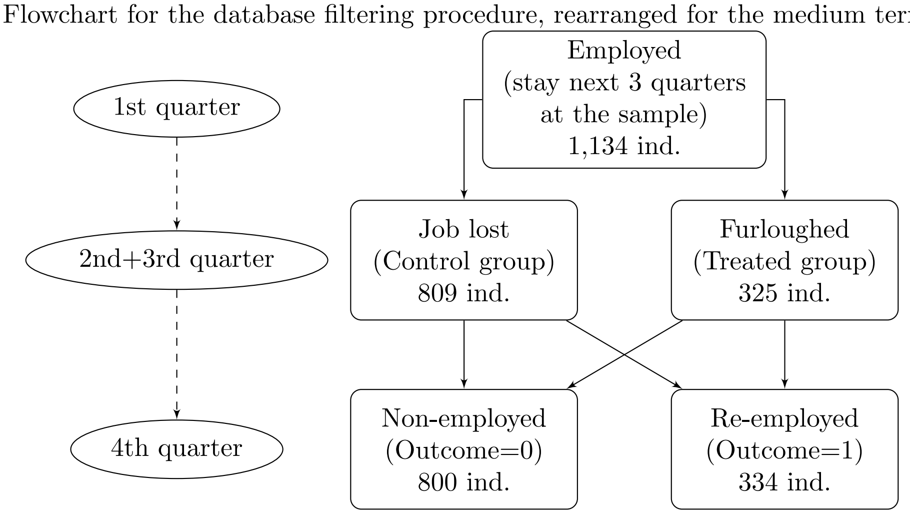
4.4. Medium term impact: average effect of the second and third quarter ERTE on the fourth quarter of 2020

We now ask what would happen if we were to aggregate those workers who have been Furloughed during both the second and third quarter of 2020 in order to verify the causal consistency of the impact of the Spanish ERTE on employment in the last quarter of 2020. In this definition of medium term, we expect the positive causal effect to persist as the ERTE gives more time to both the worker and the entrepreneurs to adjust to new market conditions. This time, the filtering procedure for the data is analogous. However, now the individuals considered for both control and treatment group must necessarily stay in the same situation during the second and third quarter as illustrates figure 8. Despite reducing the sample size, this medium term analysis still preserves significance as we will see afterward.

The point of this section is to compare the average treatment effect in the previous quarter to quarter analysis with an estimate coming from a treatment group that has spent relatively more time

[7] For the sake of simplicity results are available on request.

Figure 8: Flowchart for the database filtering procedure, rearranged for the medium term analysis.



being furloughed. We are thus mainly interested in: 1) confirming causality in the medium term; 2) establishing a relative comparison in terms of magnitude between the previous "short term" exercise and the current "medium term" one. For these purposes, we will stick to the one to one and one to many modelling, showing the estimates of a k=10, 0.1 caliper, a k=1, 0.1 caliper with replacement, a k=1, 0.1 caliper with no replacement and two kernel density estimates (based on a normal and on a on the rearranged dataset. Estimates and Stability tests are visible in Tables 7 and 8. The estimates of the Average treatment effect on the treated (Column (IV), Table 7,) as with the previous exercise, stay consistent across all matching models. The ATE differential, in column (V) is also precisely estimated with values ranging from 0.199 to 0.203. The value of the ATE is comparable to what we found in the previous exercise, where the estimated range across all models was comprised between 0.175 and 0.287. Though the effect on re-employment is perhaps smaller in magnitude when compared to the previous exercise, all estimates appear to be stable as visible in the post-estimation checks in Table 8.[8]. All in all, the positive causal effect of Furloughs on employment appears to have helped workers in Spain across the whole year, regardless of the duration of the Furlough.

Table 7: Propensity score matching, second and third quarter 2020 aggregated.

Model	*Matching results*	*Furloughed*	*Unfurloughed*	*Diff*	*S.E.*	*T-stat(BS Z)*
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
$k = 10,\ c = 0.1$	ATT-ATNT-ATE	0.374	0.172	0.202	0.048	4.21*
$k = 1, c = 0.1,\ r.$	ATT-ATNT-ATE	0.374	0.175	0.199	0.052	3.81*
$k = 1, c = 0.1,\ n.\ r.$	ATT-ATNT-ATE	0.386	0.211	0.175	0.049	3.60*
normal, $h = 0.01$	ATT-ATNT-ATE	0.364	0.162	0.203	0.051	3.95*
epanech., $h = 0.01$	ATT-ATNT-ATE	0.375	0.174	0.202	0.049	4.09*

Estimates of the average treatment effects. Column (I) shows the model of choice. r. means replacement, n.r. no replacement.

[8] Rubin's B, is perhaps above the suggested 25% upper limit. However, this result, together with the lower magnitude of the average treatment effect estimates, is perhaps justified by the reduced sample size of this exercise. In fact, once the second and third quarters of 2020 are aggregated the sample size decreases from around 4000 individuals to around 1000. That obviously affect matching quality, as it is evident comparing replacement with no replacement common support matching in Figures A1 and A2.

Table 8: Propensity score matching, quality check, medium term.

Sample	*Pseudo* R^2	*LR* χ^2	$p_{¿}\chi^2$	*Mean bias*	*Median Bias*	*B*	*R*
(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
$k = 10,\ c = 0.1$	0,020	16.810	1.000	3.2	2.6	33.3*	1.41
$k = 1, c = 0.1,\ r.$	0.034	27.780	0.835	5.4	3.9	43.5*	1.53
$k = 1, c = 0.1,\ n.\ r.$	0.038	17.950	0.995	5.8	4.9	46.3*	1,28
$normal, h = 0.01$	0.023	20.650	0.995	3.6	2.9	35.9*	1,.39
$epanechnikov, h = 0.01$	0.022	18.240	0.999	3.4	2.7	34.9*	1.44

Values of the pseudo correlation coefifcient and related tests on the joint hypothesis of non-significance of the control varaibles. Mean and median bias changes in column (V) and (V). Rubin's B and R in Columns (VII) and (VIII). Suggested values for B and R: B less than 25, R between 0.5 and 2.

4.5. Regional and sectoral effects, single quarter analysis

The last section proved how the analysis of the Furlough measures at the national aggregated level led us to conclude, ceteris paribus, that people Furloughed showed a higher probability of reemployment with respect to unfurloughed workers during the most relentless, initial phase of the coronavirus outbreak. We might now wonder whether or not any residual heterogeneity is left at a more disaggregated territorial level. Since the SLFS is currently available at different aggregation levels, this section present a logit regression of the binary outcome variable, employment state in the third quarter of 2020, on the treatment dummy and the controls we have already made use of in the matching exercise, that is considering the first three quarters of 2020 separately. The output of the logistic regression is available in Table A1. Marginal effects to analyze the direct impact of Furlough in each autonomous region and sector have also been calculated and are presented in Table 9. To give a more solid idea of the precision and consistency of the results (distance from the null of the coefficients and similarities in magnitude) we also report marginal effects and their confidence intervals in Figures A3 and A4. By a raw and direct inference on the logit, a few important hints on the impact of the main control variables on re-employment can help us trust the estimates. First of all, the negative coefficient in the dummy and age variables are consistent with the empirical occurrence that being a woman and older than a given threshold decreases your chances to get once more employed. Secondly but perhaps more importantly, the geographical position at the autonomous level does condition propensity to re-employability: the mere fact of being a worker based in Cantabria, Castilla and Leon, Valencia, Galicia or la Rioja is by itself a characteristic that could lead to higher likelihood of future re-employment.

However, as we switch our attention to the marginal effects of the Furloughs on Re-employment, conditional on the region and the economic activity, any possible noise induced by the results on the controls mentioned before disappear, and we are left with a very statistically precise and homogeneous sequence of coefficients very close in magnitude. In the case of the autonomous communities, a unitary increase in the ERTE variable would increase the probability of reemployment in the third quarter of 2020 in a range of values comprised between 22,3 and 25,4%. On a similar note, conditional on each sector of activity, the same increase would lead to an increase in probability within a range of 23,8 and 25,4%. Results for the marginal effects are visible in Table 9. Every single coefficient falls comfortably into a confidence interval well above 0.

Table 9: Logit estimates, marginal effects.

Variables	$\delta y/\delta x$	*S.E.*	z	*P(Z)*
(I)	(II)	(III)	(III)	(IV)
Andalucía	0,250	0,008	29,680	0,000
Aragón	0,250	0,009	29,400	0,000
Asturias	0,251	0,009	29,420	0,000
Baleares	0,249	0,009	28,370	0,000
Canarias	0,238	0,009	27,830	0,000
Cantabria	0,223	0,014	16,230	0,000
Castilla y León	0,243	0,009	27,930	0,000
Castilla-La Mancha	0,247	0,009	27,930	0,000
Cataluña	0,251	0,008	29,600	0,000
C. Valenciana	0,242	0,008	28,860	0,000
Extremadura	0,249	0,009	28,770	0,000
Galicia	0,241	0,009	29,630	0,000
C. de Madrid	0,251	0,008	28,470	0,000
R. de Murcia	0,247	0,009	28,470	0,000
C. F. Navarra	0,240	0,010	24,240	0,000
País Vasco	0,250	0,09	29,390	0,000
La Rioja	0,226	0,16	13,840	0,000
General manufacturing	0,254	0,009	29,750	0,000
Extract, supply and other industr	0,252	0,009	29,290	0,000
Machinery/ install and repair	0,252	0,009	29,220	0,000
Construction	0,238	0,009	27,650	0,000
Trade and hospitality	0,251	0,009	29,38	0,000
Transp and storage/info and comm	0,254	0,009	29,670	0,000
FandI/r.state/prof,sci,tech	0,254	0,009	29,670	0,000
Public admin/educ and health	0,250	0,009	28,920	0,000
Other services	0,249	0,008	29,290	0,000

Estimates of the marginal effects according to the autonomous community and by economic activity in Spain.

5. Final remarks

Our work has proven that the propensity to be re-employed was significantly higher in the treated (Furlough granted group) than in the control group (displaced workers), leading to a positive net effect on after-ERTE re-employability. The analysis, carried out through propensity score matching techniques, has led us to results which appears to be robust to a variety of alternative specifications, may that be a time-wise different data arrangement or a series of tweaks to the selection procedure related to the matching method. To add more to this, we have shown the existence of a positive re-employment effect at a lower level of aggregation (autonomous regions) everywhere in the country. Every marginal effect, calculated for each of the Spanish autonomous regions, appeared statistically significant and close in magnitude to the other.

As a suggestion to the Spanish Policy makers, we precise that short term work public schemes appear to be once more a very relevant policy tool both at the national and the regional level when labor market stability is the target, but any public choice related to this kind of tool should be considered keeping in mind that the duration and timing of the maneuver is essential for it to reduce social costs and achieve the highest possible effect on re-employability given the conditions of the labor market in the Covid era.

6. References

[1] Drahokoupil, J., Müller, T. Job retention schemes in Europe: a lifeline during the Covid-19 pandemic. ETUI Research Paper-Working Paper, DOI: 10.2139/ssrn.3931230, 2021.

[2] A. H. Huffmann, M. D. Albritton, R. A. Matthews, L A. Muse, and S. S. Howes, "Managing furloughs: how furlough policy and perceptions of fairness impact turnover intentions over time ", The International Journal of Human Resource and Management, DOI: 10.1080/09585192.2021.1879207, 2021.

[3] I. Escudero-Castillo, F. J. Mato-Diaz, and A. Rodriguez-Alvarez, "Furloughs, Teleworking and Other Work Situations during the COVID-19 Lockdown: Impact on Mental Well-Being ", International Journal of Environmental Research and Public Health, 18(6), 2898, 2021.

[4] F. A. Dias, "The Racial Gap in Employment and Layoffs during COVID-19 in the United States: A Visualization " Socius: Sociological Research for a Dynamic World, 7, pp. 1-3, 2021.

[5] H. Fan, and X. Nie, "Impacts of Layoffs and Government Assistance on Mental Health during COVID-19. AN Evidence-Based Study of the United States " Sustainability, 12(18), 7763, 2021.

[6] Hijzen, A., & Martin, S., "The role of short-time work schemes during the global financial crisis and early recovery: a cross-country analysis ", IZA Journal of Labor Policy, 2(1), 1-31, 2013.

[7] Osuna, V., & García-Pérez, J. I., "On the effectiveness of short-time work schemes in dual labor markets ", De Economist, 163(3), 323-351, 2015

[8] Arranz, J. M., García-Serrano, C., & Hernanz, V., "The changing use of short-time work schemes: Evidence from two recessions ", European Journal of Industrial Relations, 25(1), 5-22, 2019.

[9] Cahuc, P., "Short-time work compensation schemes and employment ", IZA World of Labor, 2019

[10] Cahuc, P., Kramarz, F., & Nevoux, S., "The Heterogeneous Impact of Short-Time Work: From Saved Jobs to Windfall Effects ", IZA Discussion Paper No. 14381, 2021

[11] Müller, T., & Schulten, T., "Ensuring fair short-time work-a European overview ", ETUI Research Paper-Policy Brief, 7, 2020.

Appendix

Figure A1: Cumulative 2 quarters Furlough: Bar plot of the overall distribution of the propensity score for the treatment and the control groups. One to one matching, replacement.

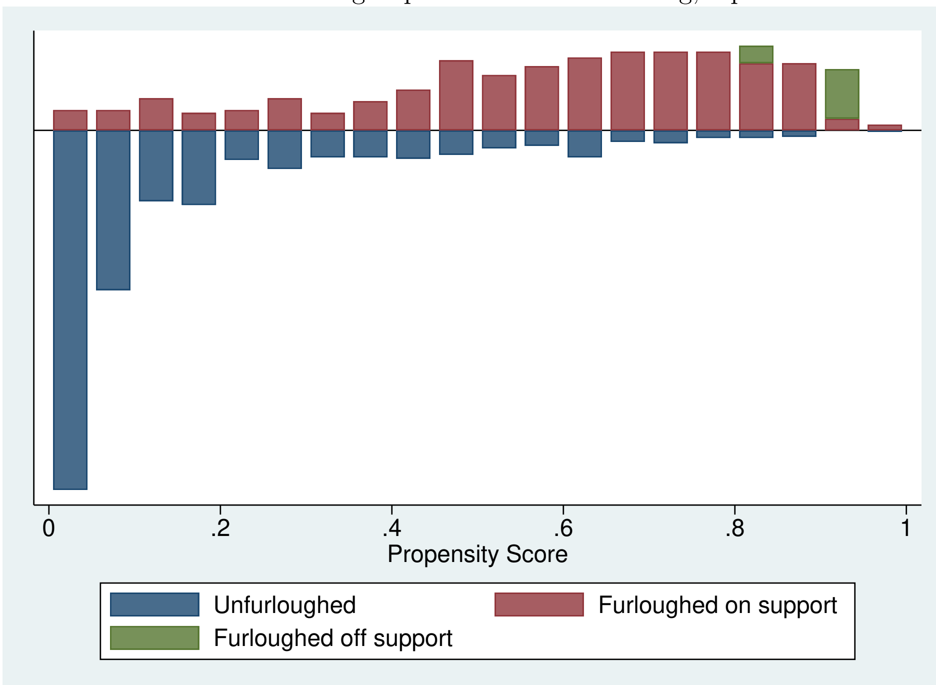


Table A1: Logit, outcome variable: employment condition 3rd quarter 2020

Variables	*Reemployment*	*S.E.*	*z*	*P(Z)*
(I)	(II)	(III)	(IV)	(V)
Furloughed	1.122***	(0.0418)	26,810	0,000
Female	-0.0878**	(0.0375)	-2,340	0,019
Age	-0.0175***	(0.00186)	-9,440	0,000
Aragón	-0.0483	(0.117)	-0,410	0,680
Asturias	-0.0962	(0.141)	-0,680	0,497
Baleares	0.0696	(0.121)	0,580	0,564
Canarias	-0.691***	(0.0775)	-8,920	0,000
Cantabria	0.653***	(0.188)	3,460	0,001
Castilla y León	0.265***	(0.0888)	2,980	0,003
Castilla-La Mancha	0.127	(0.102)	1,240	0,214
Cataluña	-0.112*	(0.0593)	-1,900	0,058
C. Valenciana	0.287***	(0.0671)	4,280	0,000
Extremadura	0.0309	(0.142)	0,220	0,827
Galicia	0.308***	(0.0850)	3,620	0,000
C. de Madrid	-0.141**	(0.0624)	-2,250	0,024
R. de Murcia	0.124	(0.113)	1,100	0,272
C. F. Navarra	0.350**	(0.140)	2,510	0,012
País Vasco	-0.301***	(0.0965)	-3,120	0,002
La Rioja	0.616**	(0.242)	2,540	0,011
Ceuta	0.889	(0.719)	1,240	0,216
Melilla	-0.0229	(0.830)	-0,030	0,978
Married	0.206***	(0.0435)	4,740	0,000
Widowed	0.318*	(0.170)	1,870	0,062
Divorced	0.296***	(0.0782)	3,790	0,000
Foreign	0.127***	(0.0465)	2,730	0,006
Incomplete primary education	-0.307	(0.376)	-0,820	0,414
Primary education	-0.865***	(0.330)	2,620	0,009
Lower secundary education	-0.474	(0.325)	-1,460	0,144
Upper secondary education	-0.553*	(0.327)	1,690	0,090
Post-second non-tertiary education	-0.284	(0.328)	-0,870	0,387
Higher education	-0.535	(0.326)	-1,640	0,101
General manufacturing	0.595***	(0.134)	4,430	0,000
Extract, supply and other industr	0.739***	(0.135)	5,450	0,000
Machinery/i install and repair	0.345***	(0.130)	2,660	0,008
Construction	1.134***	(0.124)	9,180	0,000
Trade and hospitality	0.786***	(0.109)	7,210	0,000
Transport and storage/ inf and comm	0.525***	(0.122)	4,300	0,000
FandI/r.state/prof,sci,tech	0.523***	(0.120)	4,340	0,000
Public admin/educ and health	0.225*	(0.125)	1,810	0,071
Other services	0.865***	(0.119)	7,290	0,000
Private employee	-0.134	(0.104)	-1,290	0,197
Temporary Contract	0.0565	(0.0421)	1,340	0,180
Part-time	-0.482***	(0.0423)	-11,41	0,000
Constant	0.102	(0.357)	0,290	0,775

Logit estimates. Observations: 15773. *** p<0.01, ** p<0.05, * p<0.1

Figure A2: Cumulative 2 quarters Furlough: Bar plot of the overall distribution of the propensity score for the treatment and the control groups. One to one matching, no replacement.

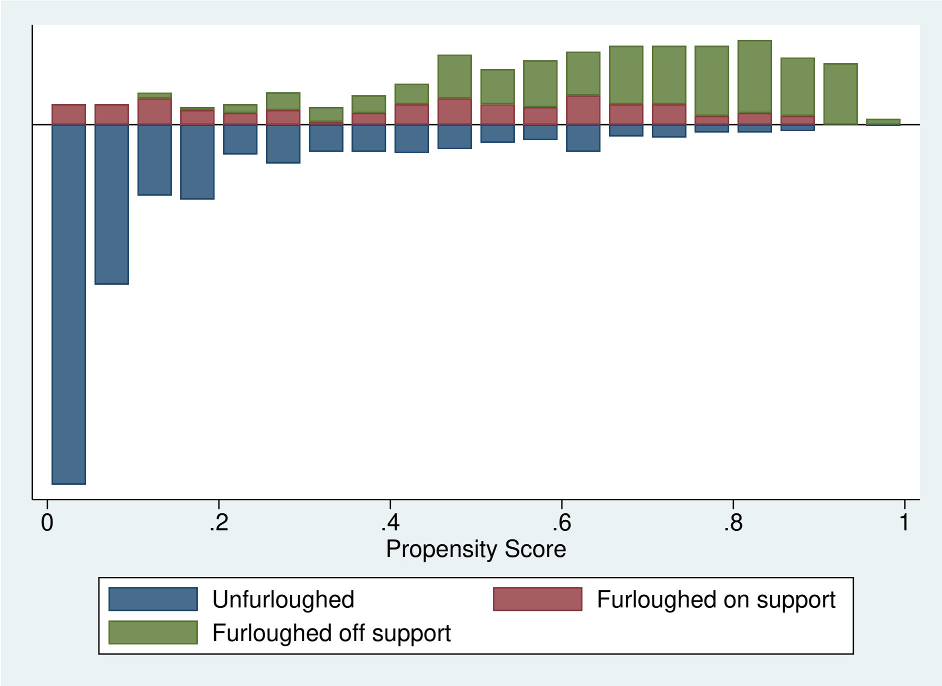


Figure A3: Marginal Effects of Furloughs on re-employment opportunity conditional on autonomous community with 95% C.I.s

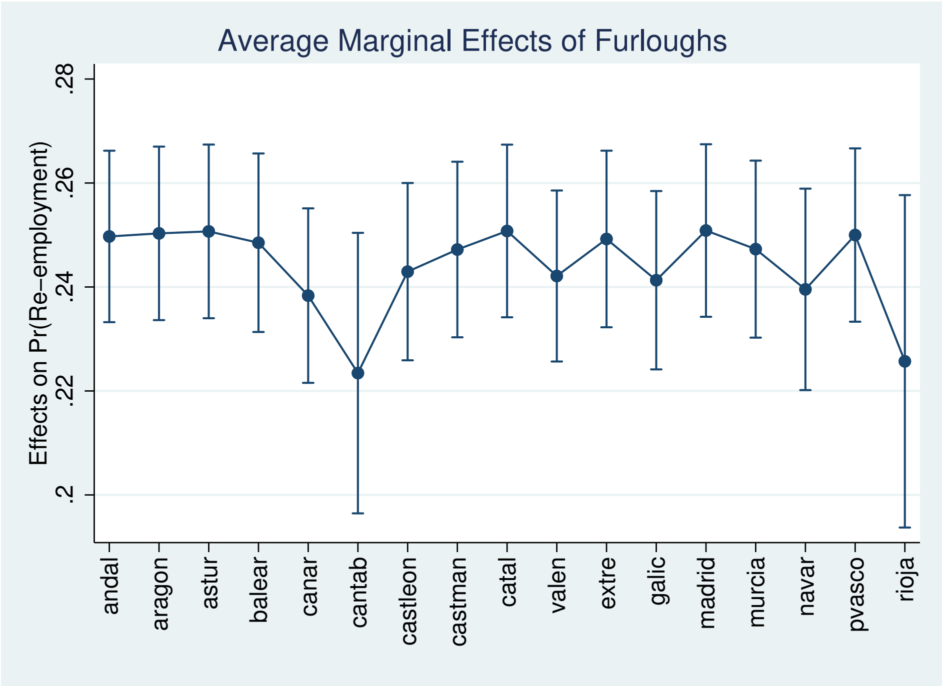


Figure A4: Marginal Effects of Furloughs on re-employment opportunity conditional on economic sectors with 95% C.I.s