Revisiting the Determinants of Unemployment Duration: Variance Decomposition à la ABS in Spain*

Maia Güell† Cristina Lafuente‡

This version: December 2021

Abstract

Using the methodology developed in Alvarez, Borovičková, and Shimer (2014), we decompose the variance of unemployment duration in Spain into three main components: labour market frictions, duration dependence and heterogeneity. Crucial to our analysis are the comparisons by different demographic groups over the business cycle to shed light in the mechanisms behind duration dependence. We offer a general approach for interpreting administrative data that is closer to unemployment than non-employment and includes different types of unemployment spells, particularly short spells without the right to claim benefits. Labour market frictions account for the bulk of the variance of unemployment duration, and become more prominent during recessions. Duration dependence constitutes 25% of the total variance, followed by heterogeneity, which represents 18%. Women and college graduates have larger shares of duration dependence. Finally, the share of duration dependence does not vary over the business cycle for non-college workers. But it is higher during expansions for college workers, which is consistent with a statistical discrimination mechanism due to dynamic sample selection.

Key words: unemployment duration, administrative social security data, duration dependence.
JEL codes: E24, J64.

---

*This paper is based on material presented by Güell in the Frisch-Tinbergen Lecture at 33th the European Association of Labour Economists Conference in September 2021 in Padova. EALE participants and Michèle Belot provided useful feedback. We are also very grateful to Katarína Borovičková, Carlos Carrillo-Tudela, Mike Elsby, Nacho García-Pérez, Rafael Lopes de Melo, John Moore, Laura Pilossoph, José V. Rodríguez Mora, Rob Shimer, Hélène Turon, and Ludo Visschers for many useful conversations and suggestions. We have also benefited from useful comments from seminar participants at the University of Edinburgh, Universidad Pablo de Olavide, CERGE-EI, Bristol University, CUNEF, EIEF, OIGI (Federal Reserve Bank of Minneapolis) and NYUAB. We are grateful to Joan Gieseke for editorial assistance. This research has been supported by the UK Economic and Social Research Council MacCaLM Grant (ES/L009633/1).

†The University of Edinburgh, CEPR, FEDEA, and IZA. Email: maia.guell@gmail.com

‡Universidad Carlos III de Madrid & European University Institute. Email: Cristina.Lafuente@eui.eu

1 Introduction

The aim of this paper is to better understand the nature of unemployment and duration dependence. We revisit this old question using large administrative social security records for Spain and apply a non-parametric decomposition of the variance of unemployment duration, following the methodology recently developed by Alvarez, Borovičková, and Shimer (2014). This method directly estimates the contributions of the different components without functional form assumptions.

A novelty of our analysis is that, thanks to our large dataset and the volatility of the Spanish labour market, we can relax the assumption that the variance decomposition is the same for all workers at all times. In particular, we let duration dependence differ by demographic groups as well as over the business cycle. This allows us to shed light in the mechanisms behind duration dependence.

Social security administrative data offer many advantages over Labour Force survey data: daily frequency of observations, no attrition, no self-reported bias on spell duration, firm identifier and the complete working history of all workers in the sample. However, social security data were not designed to study unemployment and there is no labour supply information. These data provide an account of spells of employed workers (who contribute to social security) and those of unemployed workers who are entitled to unemployment insurance (who are paid by social security). Therefore any other labour market status will appear as a 'blank' spell in the data. Crucial to unemployment analysis, unemployment spells without benefit entitlement will appear as blank spells in the data. A challenge when using administrative data is how to interpret job search among blank spells. This is important as any analysis using administrative data will be sensitive to the interpretation of the blank spells.

Alvarez, Borovičková, and Shimer (2014) (ABS, hereafter) offer two possible interpretations: on one extreme, ignore such blank spells; on the other extreme, shift the focus to non-employment and consider all blank spells as non-employment. We propose an interpretation of administrative data that expands on the original paper: an intermediate approach which identifies the different types of blank spells, includes only those that can be interpreted as genuine unemployment and excludes those who are clearly not engaged in job search. Following Lafuente (2020), we identify the different types of unemployment spells using the richness of our administrative data as well as labour market institutions, such as the rules of unemployment insurance and temporary contracts.

Traditionally, unemployment duration has been studied through the lens of duration models, which generally assume a parametric structure for the hazard function, distinguishing between duration dependence and unobserved heterogeneity parameters.[1] These methods require long panel data (i.e., large t) as well as some functional form assumptions. The ABS approach requires fewer assumptions and, as long as the data have a large number of observations (i.e., large n), a small number of observations per individual (i.e., small t) will suffice. The ABS method directly estimates the relative contribution of heterogeneity, duration dependence and labour market frictions to the total variance of unemployment duration. In this paper, we use numerical examples to illustrate the method and its capacity to allow the unemployment hazard to depend on duration, as well as to distinguish heterogeneity from duration dependence.

Our main findings are that, as in the ABS application for Austria, the component related to labour market frictions explains more than half of the total variance in unemployment duration. Moreover, in our comparison over the business cycle in Spain, we find the component that is due to frictions becomes even more prominent in recessions. The components related to duration dependence and heterogeneity follow in terms of importance, in this order. We find that duration dependence varies by demographic groups and is higher for women and college graduates. Perhaps surprisingly, the share of the total variance explained by duration dependence does not vary over the business cycle, particularly for non-college workers. For college graduates, the duration dependence component is higher in expansions. These findings are compatible with a statistical discrimination mechanism due to dynamic

[1] See Bentolila, García-Pérez, and Jansen (2017) for a recent paper on the long-term unemployed in Spain using state-of-the-art duration models.

selection into long durations for college graduates.

The rest of the paper is organized as follows. Section 2 explains the ABS method in more detail. Section 3 describes the Spanish social security administrative data. Section 4 discusses possible interpretations of the blank spells. Having established our preferred interpretation of the blank spells, Section 5 describes our results. Finally, Section 6 concludes.

2 The ABS method

In this section, we explain the method proposed by Alvarez, Borovičková, and Shimer (2014) in more detail. The starting point is to assume that there are N individuals with J completed spells of unemployment in settings in which N is large. Our outcome of interest, unemployment duration (y), is a random variable drawn independently for each individual i from a probability distribution function $F_i(y)$, with mean μ_i and variance σ_i^2.

Let y_{ij} be the duration for individual i in spell j. The population mean and variance of unemployment spells are given, respectively, by

$$\mu_y = \frac{1}{N}\sum_{i=1}^{N}\int y_{ij} dF_i(y) \quad \text{and} \quad \sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N}\int (y_{ij}-\mu_y)^2 dF_i(y), \tag{1}$$

The population variance can be decomposed into two components: the *within* and the *between* components:

$$\sigma_y^2 = \sigma_b^2 + \sigma_w^2 = \frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu_y)^2 + \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 \tag{2}$$

That is, the between component is the variance in mean durations across individuals, and the within component is the mean of individual variances across the different spells of individuals.

The between component represents the variance that comes from heterogeneity in the distribution functions $F_i(y)$ across individuals. Consider the following two extreme cases, as in Alvarez, Borovičková, and Shimer (2014). First, all individuals draw from different, individual-specific distributions, but there are no differences within individuals (i.e. $\sigma_i^2 = 0$ for all i). In this case, the total variance in unemployment duration would be attributable to individual heterogeneity captured by σ_b^2 in equation (2). Now consider the opposite case, in which all individuals draw from the same distribution function so all individuals have the same average duration (i.e., $\mu_i = \mu_y$ for all i). In this case, the variance in duration comes from within an individual's draws in their different spells.

We now turn to the interpretation of the within component and how it relates to duration dependence. Assume that the function $F(y)$ is given by a constant hazard rate h of leaving unemployment, as is frequently assumed in search models. Then the population and individual means are $\mu_y = \mu_i = 1/h$ and the standard deviations are $\sigma_y = \sigma_i = 1/h$. In this case, by definition, there is no duration dependence and the total variance is given by a constant for all individuals and all spells. Alvarez, Borovičková, and Shimer (2014) label this component as the "constant hazard within variance." We interpret this as a component coming from the aggregate frictions in the labour market that prevent workers from finding a job immediately. The time that takes workers to find a job, or unemployment duration, is drawn from an exponential distribution with parameter h.

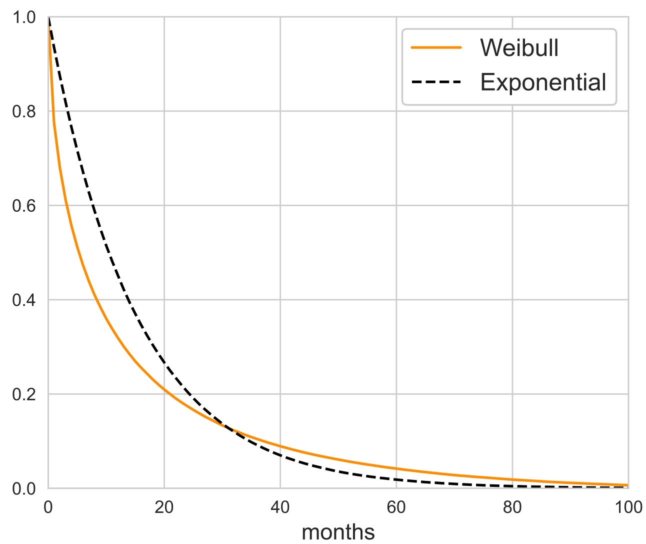
However, to the extent that the hazard rate varies with elapsed duration in unemployment (e.g., the longer in unemployment, the longer it takes to leave unemployment), then the variance will not be fully captured by the "constant hazard within variance." Duration dependence manifests itself by increasing the variance relative to the mean of individual spells (Stoyan and Daley 1983).

We can illustrate this point with the following numerical example. We simulate the labour market with 100,000 individuals.[2] We consider two hazards of leaving unemployment: (i) an exponential with constant hazard of leaving unemployment in any given month given by λ, and (ii) a Weibull distribution with CDF $F(t) = 1 - e^{-(\lambda t)^k}$, where duration dependence

[2]Our simulations show than at least 100,000 observations are needed for separating the different components accurately, based on the parametrizations in Table 2.

is captured by k. If $k = 1$ there is no duration dependence and for $k < 1$ there is negative duration dependence. Figure 1 shows the implied survival functions keeping constant the average duration of unemployment. As can be seen, the Weibull distribution with $k = 0.6$ looks like a stretched exponential distribution featuring negative duration dependence: at short durations, workers leave unemployment faster, while at long durations they take much longer.

Figure 1: Simulated survival functions



Notes: Simulated data with 100,000 individuals with 2 spells each. The average duration 15 months. The exponential has $\lambda = 0.066$ and the Weibull has $\lambda = 0.10$ and $k = 0.6$.

Given that duration dependence can be thought of as the distance between the variance and the squared mean, the within component from equation (2) can be expressed as:

$$\sigma_w^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 = \frac{1}{N}\sum_{i=1}^{N}(\sigma_i^2 - \mu_i^2) + \frac{1}{N}\sum_{i=1}^{N}\mu_i^2 \tag{3}$$

The last term of equation (3) is the squared mean of individual spells which is equal to the total variance of the exponential distribution. Therefore this term represents the share of the variance that would correspond to the constant hazard benchmark.

This allows us to interpret the first term (the difference between the variance and the (squared) mean of the individual spells) as the *excess* variance or the remaining variance that is not explained by the constant hazard model. That is, under the constant hazard, this term would be zero. But if there is duration dependence it would be different from zero: positive if there is negative duration dependence while negative if there is positive duration dependence. Alvarez, Borovičková, and Shimer (2014) label this term as the "excess within variance."

In sum, we have that the total variance is given by the following three components: the between variance, the excess within variance and the constant within variance.

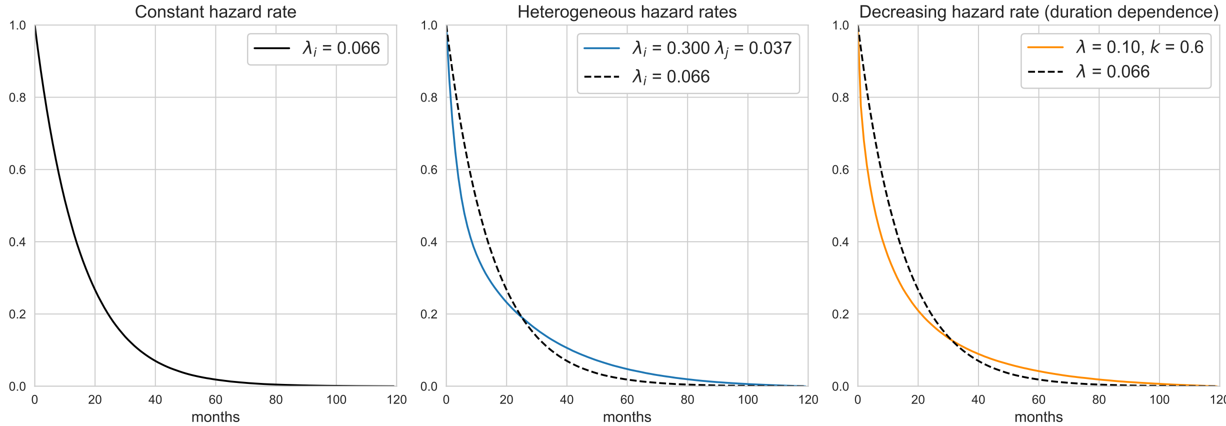
$$\sigma_y^2 = \sigma_b^2 + \sigma_w^2 = \sigma_b^2 + \sigma_e^2 + \sigma_c^2 \tag{4}$$

2.1 Estimation of the variance components

ABS show that using only two completed spells per individual one obtains unbiased estimators of the variance components with a large number of individuals. In this case, the average and the variance of the two spells are unbiased estimators of the individual means and variances. And the sample mean and sample variance are unbiased estimators of the population mean and variance, where the sample is composed by two spells from n individuals. These unbiased estimators of μ_i, σ_i^2, μ_y and σ_y^2 can be used to obtain unbiased estimators of the three components of the variance as explained in ABS. Denoting with a circumflex the estimators of the different components, the estimators in equation (2) are given by $\hat{\sigma}_b^2$, $\hat{\sigma}_e^2$ and $\hat{\sigma}_c^2$, respectively.

Let us numerically illustrate how the proposed estimators are able to capture duration dependence and separate it from heterogeneity without parametric assumptions. Figure 2 shows the survival plot of our simulated duration data generated by drawing 100,000 individuals from 3 different distributions. The first panel shows the survival plot when the underlying distribution is an exponential, implying a constant hazard rate of leaving unemployment. The second panel illustrates heterogeneity and shows the survival plot when

Figure 2: Simulated survival functions



Notes: Simulated data with 100,000 individuals with 2 spells each. The average duration is 15 months. The exponential has $\lambda = 0.066$. The mix of 2 exponential distributions has $\lambda = 0.3$ for half of the observations and $\lambda = 0.037$ for the other half. The Weibull has $\lambda = 0.10$ and $k = 0.6$.

half of the observations come from an exponential with a very high hazard rate and the other half of the observations come from an exponential with a very low hazard rate, keeping constant the average duration of the first panel. In this figure, the survival plot from the single exponential case is superimposed (dashed line) to this mix of two exponentials. We can observe how the mix of two exponential distributions *stretches* the survival function compared to the single exponential case. This reflects that now there are more very short spells (coming from the exponential with higher λ) and more very long spells (coming from the the exponential with lower λ). The third panel is the same as in Figure 1 above, and shows that the stretching of the distribution with respect to the single exponential if the data come from a Weibull distribution. In this case, workers find jobs faster at the beginning (resulting in more short spells) but take longer to leave at higher durations (resulting in more longer spells). Comparing the second and third panel shows why it is challenging to assess if the extra variance with respect to a constant hazard comes from heterogeneity in hazard rates (second panel) or duration dependence (third panel).

Table 1 shows the results of applying the variance decomposition à la ABS as outlined above. The method is able to correctly distinguish each case. In the first scenario (exponential), the share of the *constant* component is virtually 100 per cent of the variance, while the

Table 1: Variance decomposition with simulated data

	Exponential	Exponential Mix	Weibull
$\hat{\mu}_y$	15.037	15.035	15.038
$\hat{\sigma}_y$	226.057	496.085	707.687
$\hat{\sigma}_c$	225.828	363.132	226.759
	(0.999)	(0.732)	(0.320)
$\hat{\sigma}_b$	-0.291	137.083	0.629
	(-0.001)	(0.276)	(0.001)
$\hat{\sigma}_e$	0.520	-4.13	480.298
	(0.002)	(-0.008)	(0.679)
N (individuals)	100,000	100,000	100,000

Notes: Distribution parameter values as in Figure 2. Shares of each component in parenthesis.

components representing duration dependence and heterogeneity are estimated to be very close to zero.[3] In the second column, corresponding to the case of a mix of two exponential distributions, as expected, the share of the variance explained by the *excess* variance is very close to zero again (if slightly negative). The share of the *between* component is around 28 per cent of the total variance.[4] The share of the *constant* component, which comes from the fact that there is variance of the two individual spells for both type of workers, explains the rest of the total variance. Finally, in the case where the duration of unemployment comes from a Weibull distribution, the share of the *between* component is very close to zero, while the share of the *excess* variance is now around 68 per cent of the total variance. These extreme cases illustrate how the ABS variance decomposition successfully distinguishes between heterogeneity and pure duration dependence.

We have assumed so far that the two observations per individual are from the same

[3] The unbiased estimator can be negative when the true between variance is zero, as explained in Alvarez, Borovičková, and Shimer (2014).

[4] Note that in order for the between variance to be 100 percent of the variance every individual would have to have the same spell duration twice, and therefore all individual variances to be zero. This is a very unrealistic scenario.

individual distribution F_i. This may not necessarily be the case for all applications, and $F_{i,1}$ may differ from $F_{i,2}$. ABS show that the proposed estimate of the individual variance is unbiased if and only if the two distributions have the same mean. ABS argue that it is reasonable to assume this is the case if the mean of all first spells across individuals is the same as the mean of all second spells across individuals. In practice, to achieve this one has to first regress the duration variable controlling for being the first or second spell and then do the decomposition on the residuals from the regression. This is represents the variance that is not explained by the spell number. The same procedure can be applied for any other observable covariate.

To sum up, we have:

$$\hat{\sigma}_y^2 = \hat{\sigma}_s^2 + \hat{\sigma}_b^2 + \hat{\sigma}_w^2 = \hat{\sigma}_s^2 + \hat{\sigma}_b^2 + \hat{\sigma}_e^2 + \hat{\sigma}_c^2 \tag{5}$$

where $\hat{\sigma}_s^2$ is the estimated variance explained by the spell number. In practice, in our application, as in ABS, the share of $\hat{\sigma}_s^2$ out the total variance is close to zero.

3 Data

We use Spanish social security administrative data. The "Spanish Continuous Working Life Sample" (MCVL, the Spanish acronym, hereafter) comprises the complete working histories of a 4% sample of the working population (including self-employment) in 2005, that is increased yearly to keep representative of the working population. Working histories go back to when individuals had their first employment spell up until 2013. Relevant to this paper is the large number of observations in terms of both individuals at any point in time and the large number of years available, which is suitable for the ABS method.

Administrative social security data offer daily frequency of observations, which means we can capture shorter spells that get lost in quarterly surveys; they have no self-reporting bias; they have firm identifiers which for our analysis will become crucial to identify workers who are called back to work by their former employer; and finally they contain the entire working history of workers, which allows us to identify workers who are not eligible for unemployment benefits under unemployment insurance rules. This accuracy and the lack of attrition means we can perform the decomposition in different periods of the business cycle, which we explore in section 5. Moreover, similarity in data structure and sampling allows cross-country comparability, although one needs to take into account that the coverage and data classifications might differ.[5] In sections 4.1 and 4.2 we contrast our results for Spain with Alvarez, Borovičková, and Shimer (2014) results for Austria, providing useful qualitative comparisons even where the samples are not strictly comparable.

A disadvantage of these data for labour market study is that because they were designed for keeping count of the number of employed workers (who contribute to social security) and of registered unemployed workers (who are receiving unemployment insurance, UI, hereafter, paid by social security), they lack information on labour supply. Therefore, there are three possible spells in the data: employed, registered unemployed or *blank*. The last of these are spells of workers that do not fall into the first two categories. For instance, these can be workers who have run out of unemployment insurance or workers without the right to claim unemployment benefits.

A challenge with using social security data is how to interpret job search among the blank spells. This is important because otherwise spell duration and the number of unemployment spells could be under-reported. However, the fact that we observe the complete history of individual spells will allows us to identify them and propose in section 4 a strategy to account for them. Our estimates will be very sensitive to the interpretation of blank spells. Because our assumptions about blank spells can have a large effect on the results, this motivates us to think carefully about them and to look for the best approach. Finally, note that this would also be the case for other estimation methods using these data and not only the ABS method.

[5] See Güell, Lafuente, Sánchez, and Turon (*forthcoming*) for a study comparing the social security data of Germany and Spain in the context of unstable spells.

The MCVL records daily one person-spells, information on the reason why the contract has ended, firm identifier and, from 1996, type of employment contract. It also records some demographic variables such as gender, age and education level.[6] As in Alvarez, Borovičková, and Shimer (2014), we use the first two consecutive spells of unemployment observed in the data for all individuals for workers aged 25-50. We use spells from 1996 up to 2013.

As mentioned earlier, in the ABS method, first and second spells need to be comparable – come from the same individual distribution. Appendix A shows that first and second spells are not too different in our data. In any case, as in ABS, we will first control for spell number before doing the variance decomposition.

Given the two spell requirement for the ABS method, it is worth checking the representativeness of the sample used. In Appendix F we compare the histograms of the spells in our sample with those without a two spell requirement. We also check that our sample represents around 70% of all registered unemployed. And even for college workers, it constitutes 66% of those registered unemployed. One possible reason for these high shares is that temporary contracts in Spain generate at least one short spell of unemployment for a large share of the workforce. This provides reassurance with regard to concerns about selection bias due to the two spell requirement.

4 Interpretation of blanks

In this section, we explain how we can interpret the blank spells. Following Lafuente (2020), we exploit the richness of information in MCVL together with institutional information (e.g., unemployment insurance rules) and use information from the Spanish Labor Force Survey to guide our interpretations.[7]

There are three main approaches: one extreme is to ignore the blank spells; another

[6]Please note that in the Spanish social security data, maternity leave spells appear as part of employment spells.

[7]As explained in Lafuente (2020) the overlap in observable characteristics between administrative and survey data is very limited. This precludes the use of machine learning techniques to address under-reporting.

extreme is to shift the focus from unemployment to non-employment and interpret all blank spells as non-employment. These two approaches can be found in Alvarez, Borovičková, and Shimer (2014).[8] We will implement these two approaches and compare our findings with those of Austria.[9] Moreover, in this paper, we also offer a more nuanced approach, identifying the different types of blank spells and include only those that can be interpreted as genuine unemployment. We will show the advantage of our approach in a context in which the data allow us to identify workers who are clearly not engaged in job search.

4.1 Ignoring the blank spells: RU interpretation

We start by not giving any interpretation to the blank spells. In other words, ignoring such spells. This means that only registered unemployment spells (i.e., those receiving UI) are considered. We label this interpretation as *registered unemployed* (RU, hereafter). This interpretation is also considered in ABS for Austria. Figure 3 displays the duration distribution for these unemployment spells. As can be seen, striking spikes appear at various durations due to administrative reasons and/or end of UI eligibility. The spike at 4 weeks comes from the fact that UI payments are made in monthly installments. The spikes at 16, 24 and 32 weeks relate to UI eligibility.[10] However, all these spikes represent a mixture of a genuine end-of-benefit effect[11] and the nature of the data (i.e. some workers do not leave unemployment at the end of their UI but the administrative data stops recording their unemployment spell becoming a blank spell).

In practice, ignoring the blank spells means that unemployment duration will be under-reported in the data. So much so, that we will find positive duration dependence when

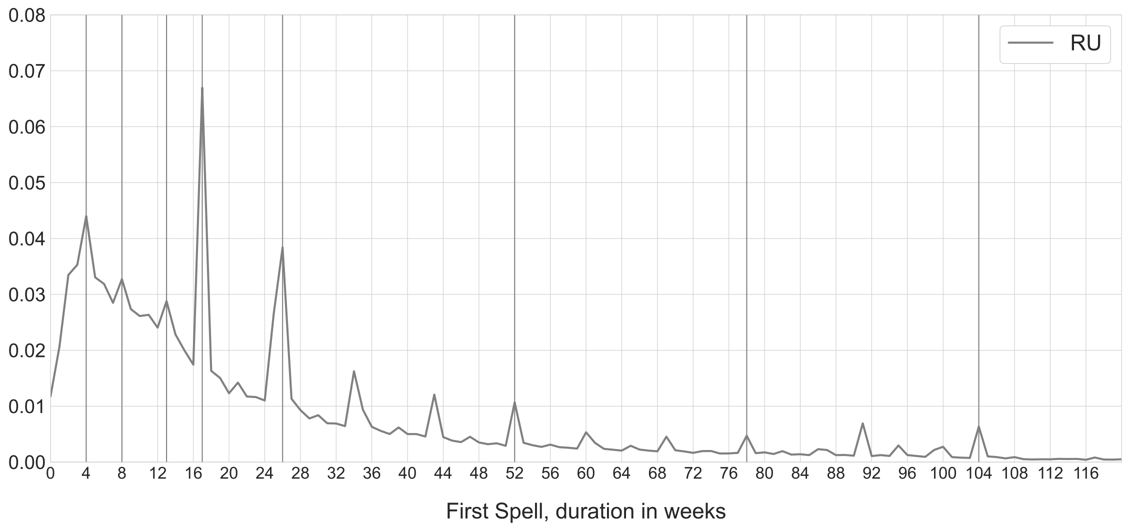
[8]Fitzenberger and Wilke (2009) also discuss the possible interpretations of blank spells using social security administrative data for Germany in a context in which spells can be right censored.

[9]In Appendix C we offer a comparison of the labour markets in Austria and Spain.

[10]In Spain, workers receive UI during a period that corresponds to a third of their previous employment spell. The minimum employment spells for eligibility is one year, which gives 4 months (16 weeks) of UI. Spikes at around 24 and 32 weeks reflect contracts of 1.5 and 2 years, respectively.

[11]See Tatsiramos and van Ours (2014) for a recent overview and Rebollo-Sanz and García-Pérez (2015) for recent evidence for Spain.

Figure 3: Duration of unemployment spells, registered unemployed spells



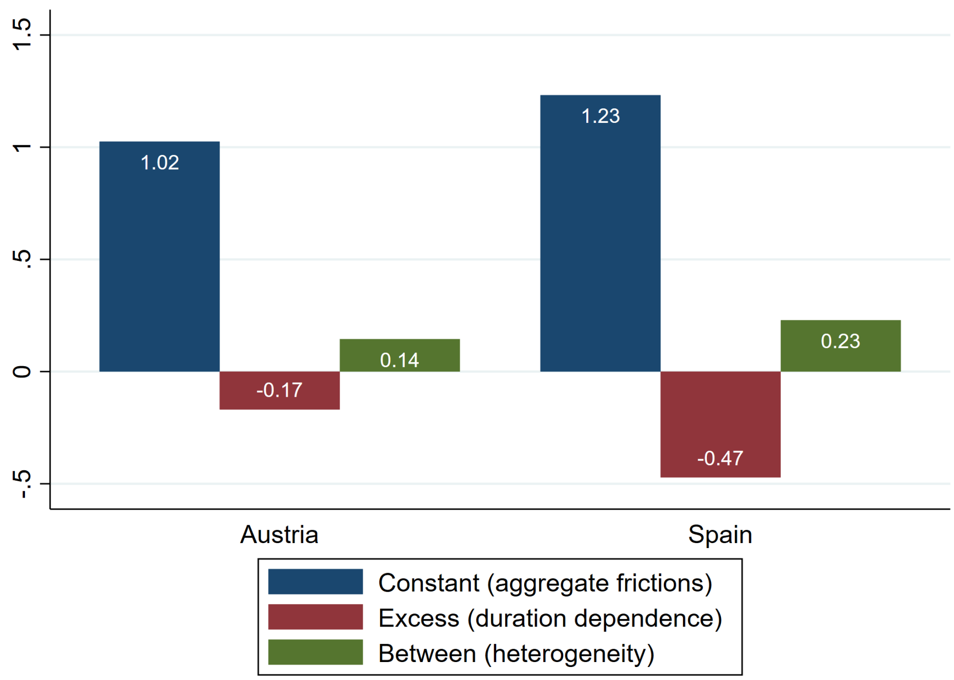
Notes: Histogram of registered unemployment (RU) spells. This plot shows the first spell of all individuals in the sample (210, 825). Data source: MCVL.

performing the decomposition of the variance. This comes from the fact that a longer duration means being closer to the end of UI and thus leaving the registered unemployment spell (but not necessarily unemployment).

In what follows we present the decomposition of the variance of unemployment duration into the 3 components described in section 2, see equation (5). In Figure 4, we display the decomposition for Spain and also report the Austrian one from ABS to highlight that the positive duration dependence is a common finding for these two countries when following this interpretation of blank spells.

Several aspects from Figure 4 are worth noting. First, for both countries, the most important and relevant component is the constant. Second, for both countries, the duration dependence component is actually a negative number (i.e, there is positive duration dependence). This is an artefact of the data under this interpretation of the blanks. On the one hand, very long spells of unemployed whose UI has expired are missing. On the other hand, very short spells without UI with are also missing. These two types of spells would stretch the hazard rate in the two extremes contributing to a monotonic decreasing function and thus negative duration dependence. In Austria this issue is less prominent than in Spain.

Figure 4: Unemployment duration variance decomposition, RU interpretation



Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). The share of $\hat{\sigma}_s^2$ out the total variance is 0.01. Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$. Data source: MCVL (Spain); Results for Austria taken from Alvarez, Borovičková, and Shimer (2014), where the share of $\hat{\sigma}_s^2$ out the total variance is 0.

This is for two reasons.[12] Firstly, the incidence of long-term unemployment is lower than in Spain and thus fewer blank spells due to UI expiration. Secondly, regarding short spells, Austria has fewer temporary contracts which imply fewer short blank spells without UI.[13]

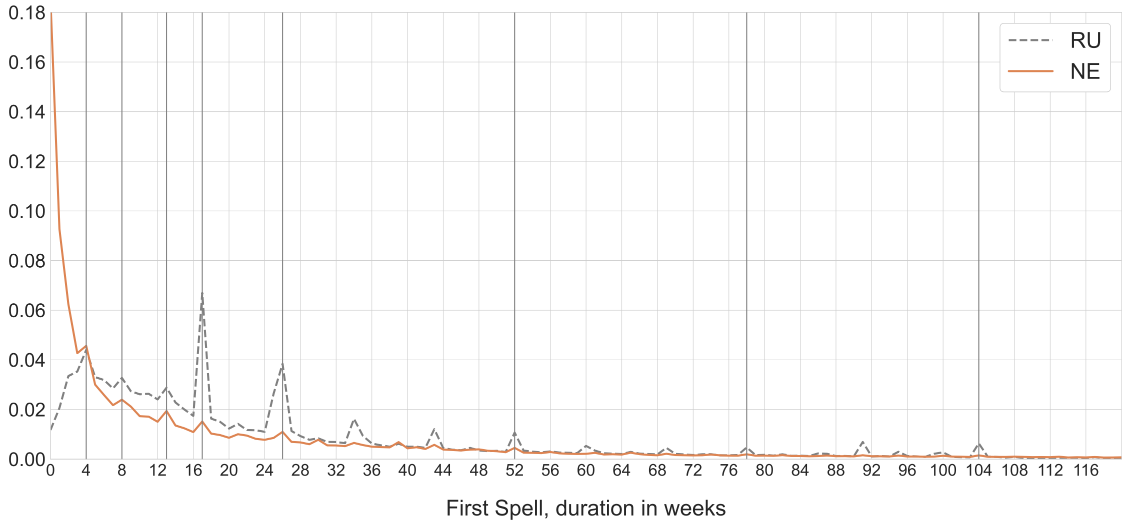
4.2 Blank spells as non-employment: NE interpretation

We now turn to the opposite interpretation of blank spells. That is, we interpret all blank spells that are not employment as non-employment. Equally, we interpret all registered unemployment spells as non-employment. We label this interpretation as *non-employment* (NE, hereafter). Figure 5 displays the duration distribution for treating unemployment spells

[12]Please see Appendix C for a comparison of labour institutions in the two countries.

[13]Interestingly, the decomposition in *levels* does not display a negative share, neither for Austria or Spain, see Appendix B. The decomposition in logs puts relatively more weight to short spells, emphasising the more the early exits out of unemployment which translate into positive duration dependence.

Figure 5: Duration of unemployment spells, non-employment spells



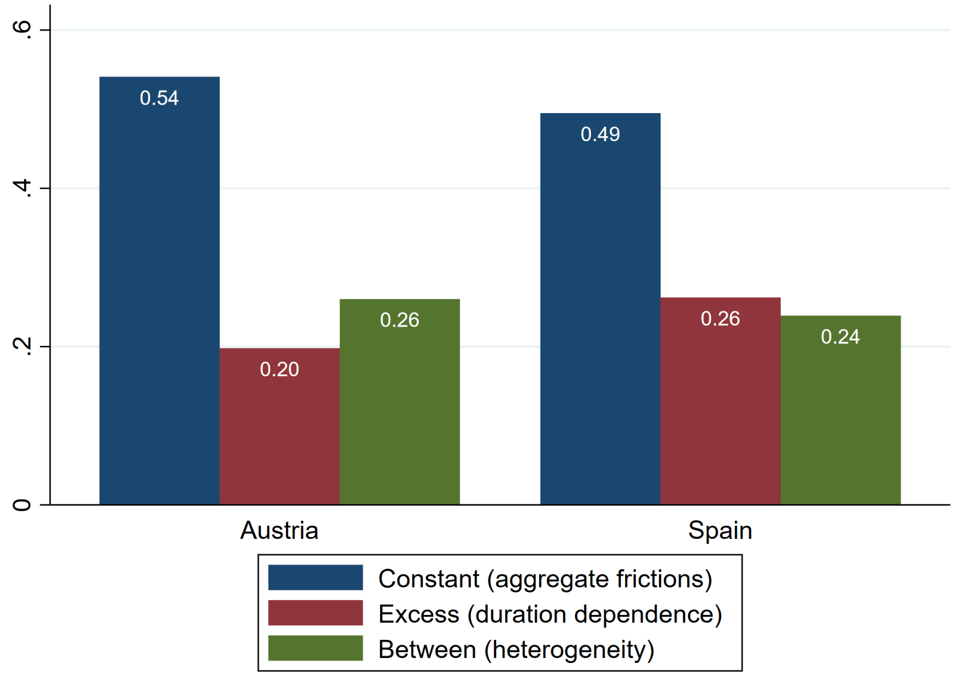
Notes: Histogram of non-employment (NE) and RU spells. This plot shows the first spell of all individuals: 436,634 in the NE sample and 220,825 in the RU sample. Data source: MCVL.

in this way. As can be seen, under this interpretation, the original spikes at the end of UI are smoothed out. The reason is that non-employment spells end when the worker finds a job and not when UI ends. This means that the duration of the non-employment spell is the genuine duration and therefore we will find negative duration dependence because the histogram displays a decaying shape.

Another change in Figure 5 is that for very short spells, less than 4 weeks, the histogram is decreasing while it was increasing under the previous interpretation. This makes the overall histogram decreasing in duration. Under this interpretation, the number of spells increases substantially because unemployment spells without UI were missing under the previous interpretation. Workers without UI have no payment at the end of the month and therefore do not have any reason not to exit unemployment before then. Interestingly, 83 percent of these new very short spells are recalls. Section 4.3 discusses recalls further.

We next present the decomposition of the variance of duration under this non-employment interpretation. In Figure 6, we display the decomposition for Spain and also report the Austrian one from ABS to highlight that now *negative* duration dependence emerges in these two countries. It is higher in Spain than in Austria.

Figure 6: Unemployment duration variance decomposition, NE interpretation



Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). The share of $\hat{\sigma}_s^2$ out the total variance is 0.004. Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$. Data source: MCVL (Spain); Results for Austria taken from Alvarez, Borovičková, and Shimer (2014), where the share of $\hat{\sigma}_s^2$ out the total variance is 0.000.

4.3 Accounting for non-registered unemployment

We propose a more nuanced approach, identifying the different types of blank spells and include in our analysis those that can be interpreted as unemployment. After excluding recalls, we can distinguish in the data three types of blank spells in between two different employment spells: (i) Registered unemployed workers whose spell runs over the duration of UI, (ii) Non-employed workers that do not have the right of UI, (iii) Non-employed workers who have the right to claim UI but do not do so. Since the ABS method only considers completed spells of unemployment in between employment spells, it is highly unlikely that these constitute episodes out of the labour force. Table 2 reports summary statistics of these three different types of blank spells. Blank spells are much shorter than registered unemployment spells.

Table 2: Summary statistics of blank spells under each interpretation

	RU	RU with expired UI	NE without right to UI	NE who do not claim UI
Mean	188.47	171.19	163.11	134.27
Standard deviation	256.92	357.79	369.37	336.85
Median	117.00	50.00	35.00	25.00
N (spells)	441,650	132,922	364,939	626,341

Notes: Duration in days. The number of observations refers to the number of spells *added* under each interpretation. Data source: MCVL.

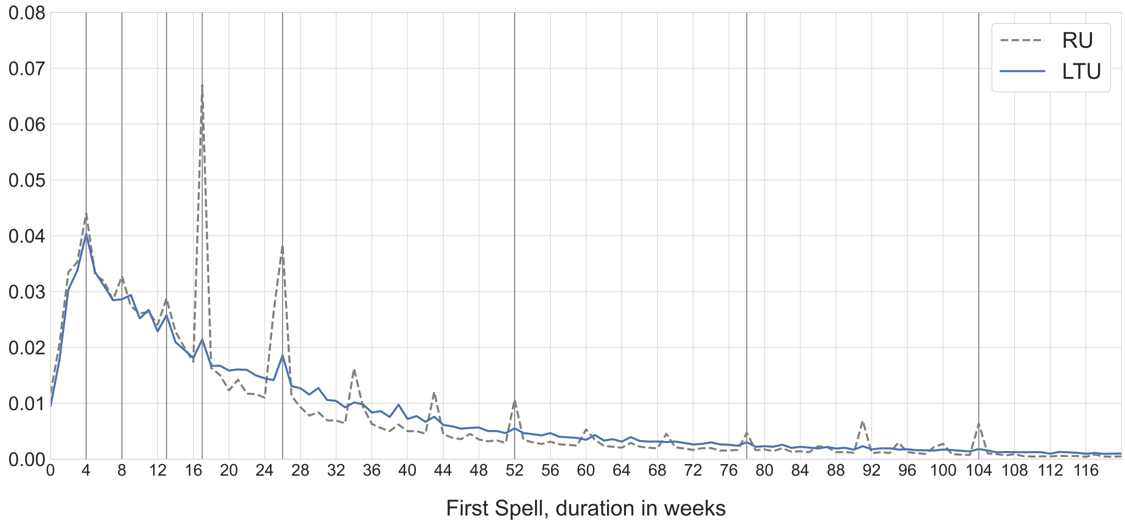
Accounting for UI expiration: LTU interpretation

The first main source of blank spells are registered unemployed workers who run out of UI before finding a job. This implies that unemployment duration in the data will be under-reported. Also, the number of unemployment spells would be overstated as it would include recalls. We interpret these blank period as unemployment and therefore the entire spell from employment to employment is considered as unemployment. Since this affects mostly workers with long durations we label this interpretation as *long-term unemployment* (LTU, hereafter).

This interpretation has been implemented by other authors (see, e.g., García-Pérez 2008). Lafuente (2020) offers an additional empirical justification based on the Spanish Labour Force Survey: 88% of registered unemployed workers with unemployment benefits in a given quarter remain unemployed after losing their benefits in the following quarter (as opposed to being out of the labour force). This interpretation results in the aggregate unemployment rate being more consistent with that of the LFS (for more detail, see Lafuente 2020).

Figure 7 displays the duration distribution of unemployment spells. As expected, under this interpretation, the original spikes related to the end-of-benefit are smoothed out, and just capturing the genuine end-of-benefit effect.

Figure 7: Duration of unemployment spells, long-term unemployment spells



Notes: Histogram of long-term unemployment spells and RU spells. This plot shows the first spell of all individuals: 200,357 in the LTU sample and 220,825 in the RU sample. Data source: MCVL.

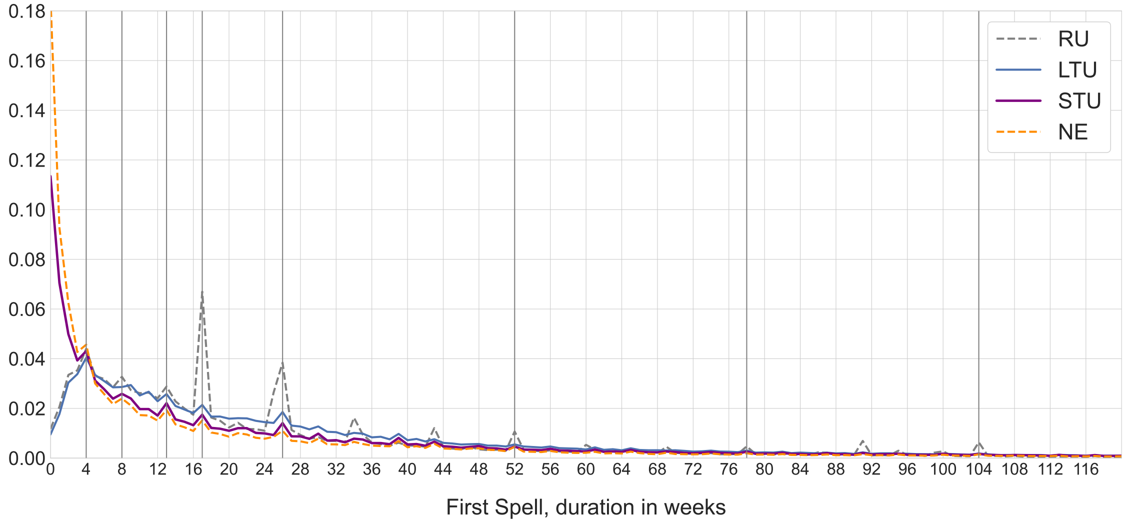
Accounting for lack of UI entitlement: STU interpretation

The next source of blank spells are unemployed workers who do no have the right of UI. Complete work histories and the UI institutional rules allow us to detect such spells.[14] Since these are clearly not registered unemployed, this implies that the number of unemployed workers is under-reported.

We find three possible cases among these blank spells: workers whose previous employment spell was not long enough (56% of cases), workers who quit their previous job (33% of cases), or those coming from self-employment (11% of cases). Thus the majority are workers coming from temporary employment whose past tenure is too low to accumulate the minimum contribution period that entitles them to receive UI. This group is comprised of workers who cycle between short unemployment and employment spells. Half of these blanks are less than 35 days and the mean duration is around 163 days (see Table 2). We interpret these blank spells as unemployment and therefore add them to the spells considered in the

[14]In particular, in any of the following cases the worker has not right to claim UI: if the worker has not had a total of 360 days accumulated in registered employment since her last registered unemployment spell, or if "voluntary" is listed as the reason for the end of her last employment spell (quit) or if her previous spell was registered as self-employed. There are not any other benefits relating to housing, sustenance or other benefits for unemployed workers in our data. We do record unemployment assistance, which is is the name of UI after it runs out in some particular cases.

Figure 8: Duration of unemployment spells, short-term unemployment spells



Notes: Histogram of short-term unemployment spells (STU), LTU and RU spells. This plot shows the first spell of all individuals: 436,634 in the NE sample, 336,228 in the STU sample, 200,357 in the LTU sample and 220,825 in the RU sample. Data source: MCVL.

LTU interpretation.[15] Given that this mostly affects workers with short durations and we label this interpretation as *short-term unemployment* (STU, hereafter).[16]

The main advantage of this approach is that, unlike analysis with quarterly labour force survey data, it includes short durations. Figure 8 displays the duration distribution of unemployment spells under differing interpretations of blank spells. This interpretation adds new short unemployment spells, some of which leave unemployment before the end of one month. Because there are enough of these observations, the histogram is now also decreasing prior to the spike due to first-month UI payments.

This STU interpretation has been used by Lafuente (2020) who finds that for workers who do not have the right to claim UI (such as many women and young workers) it results in quarterly unemployment rates that closely track the LFS.

[15]Very short spells without UI could be a job-to-job transition. We consider this in Appendix D.

[16]The ABS method requires the *first* two completed unemployment spells observed for every individual. This means that in our STU interpretation of the blanks the first and second spells of individuals may change with respect to that in which blanks are ignored. That is, a worker with two registered spells may had had a previous non-registered spell. In the STU interpretation, this becomes her first spell in the sample.

Accounting for non-claimants

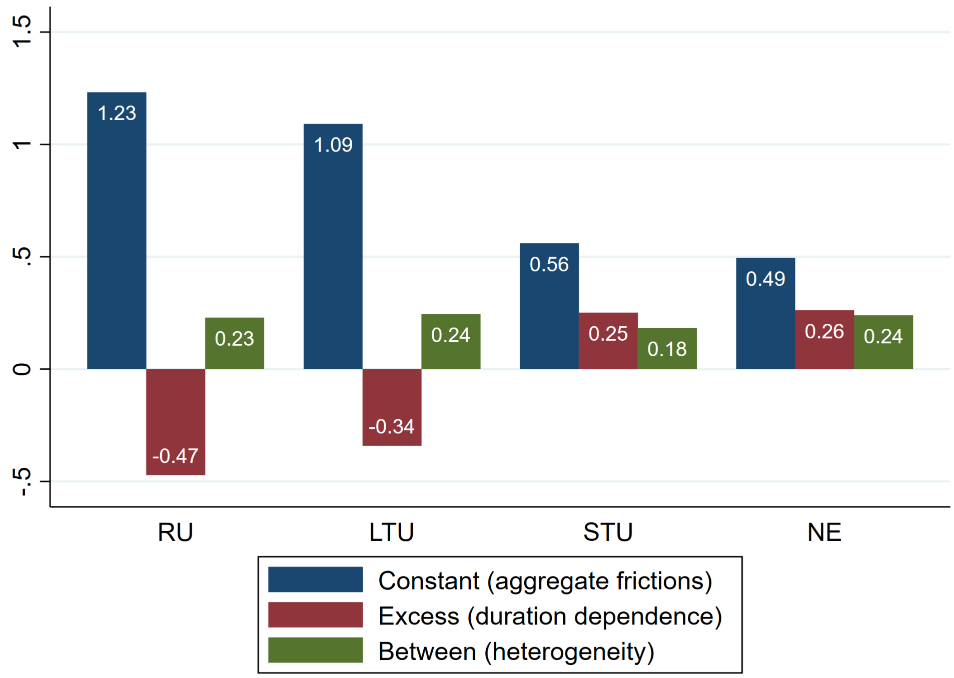
The rest of the blank spells are non-claimants of UI. These are workers who are eligible but for some reason do not claim their benefits. Some studies show that in the US there are cases of unemployed workers who do not claim UI due to stigma or other reasons (see Blank and Card 1991 and Anderson and Meyer 1997, for example). To our knowledge there are no studies showing this is an important factor in Spain.

Further inspection of these non-claimants spells in our data reveals that the majority are recalls, that is, workers who return to the same firm. In particular, 70% of all non-claimant blank spells are recalls.

Given that workers on recalls do not engage in job search, and it is not straightforward to think about duration dependence for such cases, our preferred interpretation of the blanks in our data is the *STU* interpretation, which excludes blank spells of non-claimants. That is, the key difference between the STU and NE interpretations is the absence of these non-claimant spells in the former. This results in fewer very short spells, as evidenced by Figure 8 where there are fewer spells of a week or less in the STU compared to the NE histogram.

The importance of choosing an interpretation for the blank spells becomes apparent when we apply the decomposition to each case separately. Figure 9 shows these results for the RU, LTU, STU and NE interpretations. The excess variance is still negative in the LTU case because, although most spikes at UI expiration are smoothed out (as Figure 7 showed), the histogram is increasing up to 4 weeks. This increase is analogous to an increasing hazard, which results in overall positive duration dependence. Excess variance becomes positive under the STU interpretation, as the naturally shorter durations of those without the right to claim UI are enough to make the hazard monotonically declining. Notice how the between component, related to heterogeneity, is also smaller under the STU interpretation (18% of the total variance) compared to the NE (24% of total variance). As explained above, the additional observations that are considered under the NE interpretation are mostly recalls.

Figure 9: Unemployment duration variance decomposition, different interpretations



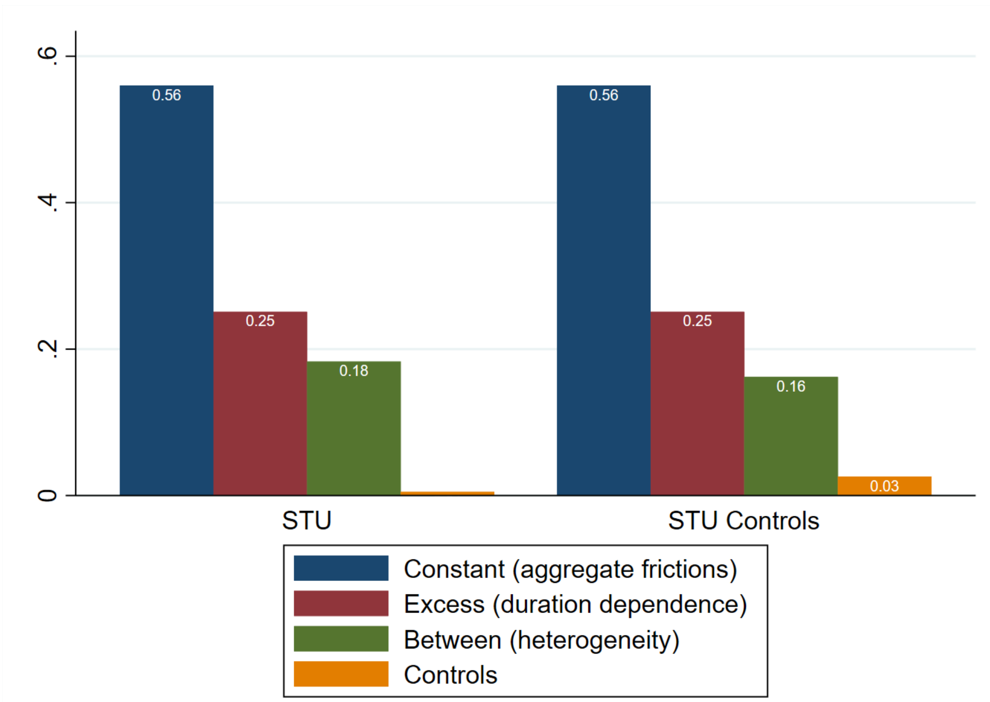
Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$. Data source: MCVL.

The weight of duration dependence is a bit larger for NE, which is consistent with Fujita and Moscarini (2017), who argue that duration dependence emerges mostly for those who are eventually recalled to the same firm. Since recalls are of a different nature than other unemployment spells, in what follows we present the results under the STU interpretation. The STU interpretation offers a balance between capturing the unemployment experience of those without the right to claim benefits and including spells that likely related to highly unstable jobs. Under this interpretation, the *constant* component (capturing the aggregate frictions) is the largest, followed by the *excess* component (duration dependence) and lastly the *between* component (duration dependence).

5 Results

Section 4 established our reasoning to prefer the STU interpretation of blank spells as unemployment. In this section, we present the results of the variance decomposition using this approach.

Figure 10: Unemployment duration variance decomposition, demographic controls



Notes: Unemployment duration in logs. Decomposition after controlling for spell number (STU), and after controlling for spell number, gender and education (STU controls). Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$ and *Controls* refers to the share of $\hat{\sigma}_s^2$ in equation (5). Data source: MCVL.

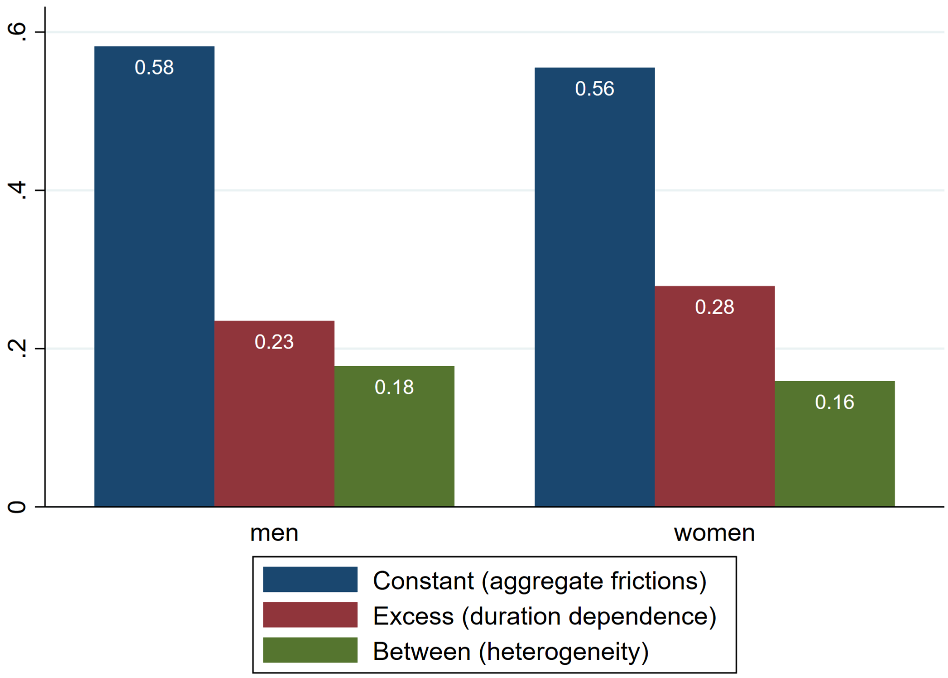
As pointed out in Section 2, the ABS method can accommodate differences in spells by observable characteristics in a similar fashion as differences between the first and second spells of individuals. A regression of duration against such characteristics is run to remove the share of the variance due to them. The ABS method is then applied to the remaining unexplained variance. Figure 10 reproduces the main decomposition and one with controls for gender and education.[17] As one would expect, the share of the heterogeneity component is now smaller

[17]Notice that performing the decomposition in logs gives relatively more weight to short spells, as the

as some of it is captured by the controls. But the share of total heterogeneity remains unchanged. Likewise, the total within component does not vary after the introduction of observable controls.

Given our large data set, we next allow all the components to differ by demographic groups as well as over the business cycle. This would allow us to shed some light on the possible mechanisms behind duration dependence. We focus on gender and education.[18]

Figure 11: Unemployment duration variance decomposition by gender



Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$. Data source: MCVL.

We start with the decomposition by gender. In our sample, around 45 percent of all spells correspond to women.[19] As Figure 11 shows, the share of the different components is similar

log transformation compresses the duration of very long spells. This is useful in that it limits the impact of outliers of very long durations in the data, which are not uncommon in Spain, while also keeping these observations. For results in levels, see Table B.2.

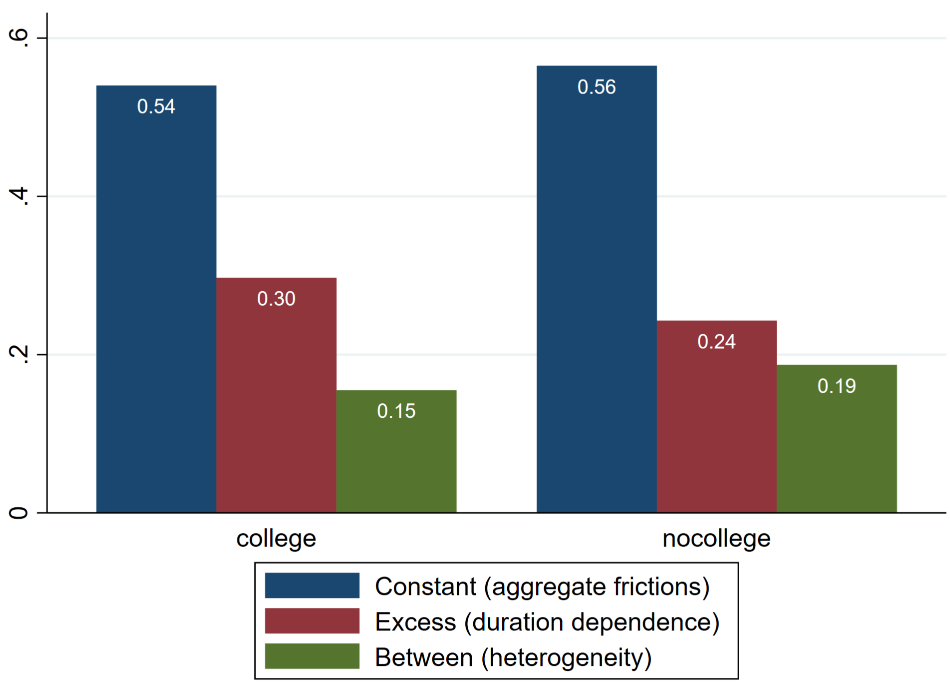
[18] Bover, Arellano, and Bentolila (2002) and, more recently, Rebollo-Sanz and García-Pérez (2015) and Rebollo-Sanz and Rodríguez-Planas (2020) explore the effects of receiving UI on the exit rate of unemployment and unemployment duration.

[19] Table E.1 reports descriptive statistics of our STU sample.

between men and women, with the constant component accounting for more than half of the total variance, followed by duration dependence taking slightly more than a quarter with heterogeneity explaining the rest. The main difference between the sexes is that duration dependence is more prominent among women.

We perform an analogous exercise by education. We split the sample into individuals who have attained at least a college degree and the rest. Around 15 percent of individuals in our sample are college graduates, a rather small number due to the two spell requirement. However they still represent 66% of all unemployed with a college degree (see Table F.2). As Figure 12 shows, the decomposition is similar for the two groups, with a more prominent role for heterogeneity among the non-college educated and duration dependence being more pronounced among college graduates. The constant component accounts for around 55% of the total variance for both groups.

Figure 12: Unemployment duration variance decomposition by education



Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}^2_c$, *Excess* refers to the share of $\hat{\sigma}^2_e$ and *Between* refers to the share of $\hat{\sigma}^2_b$. Data source: MCVL.

As shown in Appendix B, when decomposing the variance of the *logarithm* of unemployment duration, the excess within variance component only depends on the average of individual variances between spells (i.e., the mean of σ_i^2). We can therefore relate our findings of duration dependence to differences of this moment. Table 3 displays the mean of individual variances as well as its standard deviation by gender and education. The two groups with higher duration dependence are also characterized by a higher mean of individual variances. This implies that their two individual unemployment experiences are more different than for their comparison group (e.g., female/college workers have consistently higher variances than men/non-college workers). Moreover, these two groups are also characterized by a higher dispersion of individual variances. That is, female and college educated workers are more different in terms of their unemployment spells *among themselves* — more so than men and non-college, respectively. The variance decomposition shows that this greater dispersion translates into higher duration dependence for these groups.

Table 3: Individual variances, demographic groups

	Men	Women	College	Non college
Mean of σ_i^2	2.33	2.51	2.59	2.38
Standard deviation of σ_i^2	3.29	3.69	3.58	3.45
N (individuals)	185,105	151,123	50,501	285,727

Notes: Duration in log days. Data source: MCVL.

In other words, for women and college graduates the distribution of unemployment duration is riskier: they are more likely to receive extreme values (low and high), which implies a higher slope of the duration distribution and thus higher duration dependence. This higher risk is shared among all individuals in the group. That is, it is not that some women have inherently higher risk than others – that would be captured by the heterogeneity component.

Business cycle and duration dependence

An important question in the duration literature is on the nature of duration dependence. Different mechanisms have been proposed (Machin and Manning 1999): skill depreciation along the unemployment spell (e.g., Acemoglu 1995), stock-flow matching (e.g., Coles and Smith 1998), decrease in search intensity (e.g., Schmitt and Wadsworth 1993, or Krueger and Mueller 2011), or statistical discrimination (e.g., Lockwood 1991).[20]

In general, it is hard to disentangle these explanations. We can attempt to shed some light on the nature of unemployment by performing the decomposition in two very different periods of the business cycle. The key idea being that mechanisms related to structural forces, such as human capital depreciation, imply that longer duration makes workers less employable and thus duration dependence will be higher during recessions. While the opposite is true if the mechanism is related to statistical discrimination and employers consider long duration as a bad signal of worker's ability. This is because during economic downturns unemployment duration as a signal becomes noisier as there are fewer vacancies and more workers have longer durations. In other words, the dynamic sample selection will be more acute when unemployment is low (see, for instance, Blanchard and Diamond 1994 and Lockwood 1991).

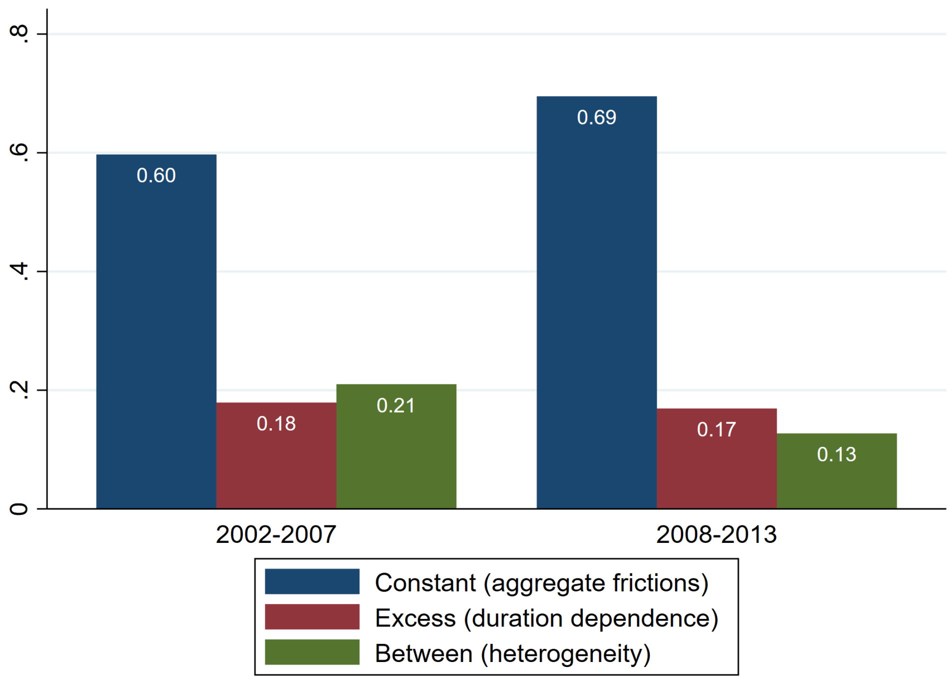
In particular, we can explore the Great Recession of 2008 which was very severe in Spain.[21] We split our sample into an expansion period and a recession period, namely, 2002-2007 and 2008-2013, respectively. We focus on a balanced sample with 5 years of recession and 5 years of economic expansion. These two periods contrast extremely. The expansion period was characterized by years of record low unemployment and a record low share of long-term unemployment (around 10% and 30%, respectively, see Figure C.1). The recession hit Spain hard, with record highs in unemployment and the share of long term unemployment (around

[20] More recent approaches have also considered the role of workers' beliefs in their perceived job finding probability and its effects of unemployment duration (see Mueller, Spinnewijn, and Topa 2019).

[21] The ebook edited by Bentolila and Jansen (2016) offers a cross-country analysis of long-term unemployment after the Great Recession as well as policy proposals. The chapter by Bentolila, García-Pérez, and Jansen (2016) focuses on Spain and highlight among other things the incidence of the very long-term unemployment (VLTU) (over two years).

25% and 40%, respectively). We analyse those unemployed workers that have two completed spells within each period.

Figure 13: Unemployment duration variance decomposition over the business cycle

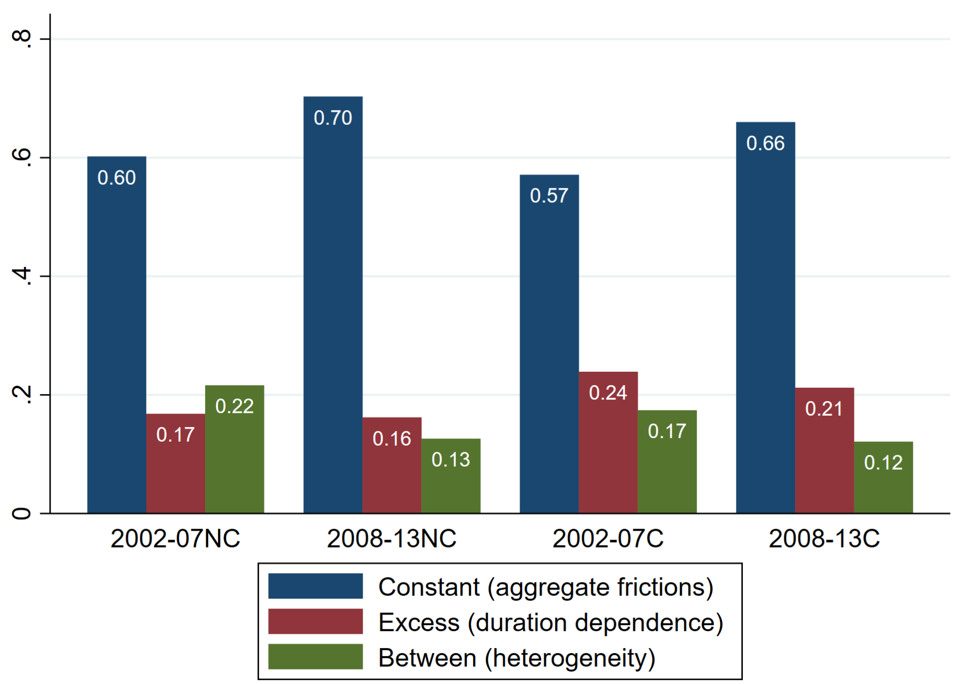


Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$. Data source: MCVL.

Figure 13 shows the results of the decomposition for each period. First, the share of the *constant* component increases its relative importance in the recession. Recall that we interpreted this component as the aggregate frictions in the labour market, or all the different reasons that induce a positive waiting time for workers to find a job. If the labour market is more congested or labour market tightness decreases, we would expect the constant component to go up in recessions, as we find in figure 13. Second, duration dependence is very similar in both periods. This is consistent with the findings of traditional duration models. For instance, Bover, Arellano, and Bentolila (2002) find that the shape of the hazard rate as a function of unemployment duration is not very different for different GDP growth rates in Spain. And Turon (2003) finds that duration dependence is not different over the business cycle in the UK. Third, the heterogeneity component is larger in expansions than recessions. As more people flow into unemployment in recessions, the composition of the unemployment pool becomes more homogeneous.

This result might be surprising given that the unemployment pool composition could be different in recessions and expansions. In our case, we find that some observable characteristics, specially those related to previous status, are somehow different in the two periods, while some other demographic characteristics are remarkably similar in the two periods (see Table E.2 in Appendix E). Overall, our findings are consistent with recessions in the US having more highly attached workers in the pool of unemployment (see Elsby, Hobijn, and Şahin 2015 or Mueller 2017). That is, in periods of economic expansion only few people are found unemployed: young workers trying to find a new job and older, long-term unemployed.

Figure 14: Unemployment duration variance decomposition by education over the business cycle



Notes: Unemployment duration in logs. Decomposition after controlling for spell number (equation (5)). *NC* refers to non-college education and *C* refers to college education. Full results in Appendix B. *Constant* refers to the share of $\hat{\sigma}_c^2$, *Excess* refers to the share of $\hat{\sigma}_e^2$ and *Between* refers to the share of $\hat{\sigma}_b^2$. Data source: MCVL.

We now turn to the decomposition over the business cycle by education groups. Figure 14 displays the results. For non-college workers, the results are similar to the aggregate results mentioned above. However, interestingly, for college workers duration dependence is noticeably higher in the expansion compared to the recession. As before, Table 4 shows that all groups that have higher share of duration dependence have higher mean and standard deviation of the individual variances.

Table 4: Individual variances, business cycle and education

			Non-College		College	
	2002-07	2008-13	2002-07	2008-13	2002-07	2008-13
Mean of σ_i^2	2.21	2.09	2.17	2.07	2.43	2.21
Standard deviation of σ_i^2	3.14	3.11	3.11	3.11	3.34	3.22
N (individuals)	150,741	166,342	129,192	144,227	21,549	22,115

Notes: Duration in log days. Data source: MCVL.

The higher share of duration dependence for college workers during the expansion period is only compatible with a statistical discrimination mechanism. This an intuitive result: from the point of view of the employer, unemployed workers with a college degree do not provide a good signal. Even less in an expansion period. Unemployed workers with a college degree who do not manage to get a job quickly have a harder time finding employment and it becomes harder during an expansion.

In sum, while we cannot disentangle the extent to which the different explanations play a role in duration dependence for non-college workers, our findings are compatible with a discrimination mechanism for college workers.

A relevant paper in this context is Jarosch and Pilossoph (2019), who build a model in which employers endogenously discriminate against the long-term unemployed. This discrimination matters to the extent that a lost interview translates into a lost job. They conclude that employer discrimination is largely due to dynamic selection. That is, duration dependence does not arise because workers' skills deteriorate while they are unemployed. Instead, the belief that good workers leave unemployment quickly due to their being picked up by other firms is what drives the firm's decision not to hire individuals who have been unemployed for a while. In their paper, the selection process happens at the interview stage: firms choose not to invite candidates for an interview where they feel the candidate has little chance of success, due to the cost of interviewing. Our findings support this explanation among workers with a college degree: As a college degree candidate is expected to be hired easily (or at least called for an interview), particularly in a period of economic expansion where there are plenty of jobs available, unemployed college graduates have a hard time convincing employers they are there due to bad luck instead of some lack of unobservable quality. When the job market is slack, duration is not such a good signal of failed interviews. This implies that duration dependence should fall during recessions, as it is the case in our results.

6 Conclusions

In this paper we offered a non-parametric decomposition of the duration of unemployment in Spain, following Alvarez, Borovičková, and Shimer (2014). We have illustrated how this methodology separates the contributions of duration dependence and heterogeneity, without making any functional form assumptions. Estimates require a large number of individuals and a minimum of two completed spells. Social security administrative data are well suited due to its very large panel, high precision, high frequency and even multiple spells per person.

However, being administrative information, the data were not collected with the express purpose of studying unemployment and hence interpretation of the blank spells is required. The most common approaches in the literature are to use only registered unemployment spells or use all non-employment gaps between employment spells. We proposed an alternative way to deal with these blank spells that identifies those who likely correspond to unemployed workers. These are workers who have exhausted UI and those without the right to claim benefits. The complexity and dysfunctionality of the Spanish labor market means that we have a large number of both types of workers to deal with. Moreover, we argued that recalls, which we can also identify in the data, should not be considered unemployment as they exhibit a different behaviour to genuinely unemployed workers who are searching for a job. We illustrate this first by comparing with the classical approach followed by ABS in Austria and then repeating the decomposition using different interpretations of the blank spells.

Our approach exploits the richness of administrative data to overcome some of its limitations. That is, these datasets capture extremely short spells that Labour Force Survey data cannot, but it can also be hard to disentangle if these spells are unemployment, recalls or job-to-job transitions. Instead of either ignoring all of this short spells or considering all of them unemployment, we separate recalls from the rest and use employment insurance rules to determine which spells correspond to workers without the right to claim, which we argue are likely to be unemployed. This more nuanced interpretation can prove useful when applying this method to other countries with similar data.

We then used this approach to decompose the variance of the log duration of unemployment in Spain. We find that the constant component (related to labour market frictions) explains most of the variance (60%), followed by duration dependence (25%) and heterogeneity (15%). The importance of the constant component is robust to performing the decomposition separately for different groups and over the business cycle. We find that the share of the variance explained by duration dependence is larger for women and college graduates. Individuals in these groups have more diverse experiences in unemployment (i.e. more different first and second spells) than in the comparison group. This higher uncertainty over the duration of their spells is interpreted through the lens of the decomposition as a higher incidence of duration dependence.

We also perform the decomposition over the business cycle, for all workers and by education groups. These comparisons allow us to shed some light on which theories can explain duration dependence. We find that the share of the variance explained by duration dependence remains unchanged for non-college workers in recessions and expansions. However, college workers, who have a higher share of duration dependence than non-college workers, display a higher share of duration dependence in expansions. This finding is compatible with a statistical discrimination mechanism due to dynamic sample selection. We argue this makes intuitive sense as employers would expect college graduates to find a job quickly, specially during expansions. Non-college workers do not face the same expectations and thus duration dependence may instead be explained by skill deterioration.

References

Acemoglu, D. (1995). Public policy in a model of long-term unemployment. *Economica 62*, 161–178.

Alvarez, F., K. Borovičková, and R. Shimer (2014). A Nonparametric Variance Decomposition Using Panel Data. Mimeo University of Chicago.

Alvarez, F. E., K. Borovičková, and R. Shimer (2016). Decomposing duration dependence in a stopping time model. NBER Working Paper No. 22188.

Anderson, P. M. and B. D. Meyer (1997). Unemployment insurance takeup rates and the after-tax value of benefits. *The Quarterly Journal of Economics 112*(3), 913–937.

Bentolila, S., J. I. García-Pérez, and M. Jansen (2016). *Long-term unemployment in Spain.* in Samuel Bentolila and Marcel Jansen (eds.), Long-term unemployment after the Great Recession: Causes and remedies, CEPR Press.

Bentolila, S., J. I. García-Pérez, and M. Jansen (2017). Are the spanish long-term unemployed unemployable? *SERIEs: Journal of the Spanish Economic Association 8*(1), 1–41.

Bentolila, S. and M. Jansen (2016). *Long-Term Unemployment After the Great Recession: Causes and remedies.* CEPR Press, Bentolila and Jansen (eds.).

Blanchard, O. J. and P. Diamond (1994). Ranking, unemployment duration, and wages. *The Review of Economic Studies 61*(3), 417–434.

Blank, R. M. and D. E. Card (1991). Recent trends in insured and uninsured unemployment: is there an explanation? *The Quarterly Journal of Economics 106*(4), 1157–1189.

Bover, O., M. Arellano, and S. Bentolila (2002). Unemployment duration, benefit duration and the business cycle. *The Economic Journal 112*(479), 223–265.

Coles, M. G. and E. Smith (1998). Marketplaces and matching. *International Economic Review 39*(1), 239–254.

Elsby, M., B. Hobijn, and A. Şahin (2015). On the importance of the participation margin for labor market fluctuations. *Journal of Monetary Economics 72*, 64–82.

Fitzenberger, B. and R. A. Wilke (2009). Unemployment durations in West Germany before and after the reform of the unemployment compensation system during the 1980s. *German Economic Review*, 1–31.

Fujita, S. and G. Moscarini (2017). Recall and unemployment. *American Economic Review 107*(12), 3875–3916.

García-Pérez, J. I. (2008). La muestra continua de vidas laborales: una guía de uso para el análisis de transiciones. *Revista de Economía Aplicada XVI(E-1)*, 5–28.

Güell, M. and L. Hu (2006). Estimating the probability of leaving unemployment using uncompleted spells from (repeated) cross-section data. *Journal of Econometrics 1*(133), 307–341.

Güell, M., C. Lafuente, M. Sánchez, and H. Turon (forthcoming). So different yet so alike: micro and macro labour market outcomes in Germany and Spain. *SERIEs: Journal of the Spanish Economic Association.*

Jarosch, G. and L. Pilossoph (2019). Statistical discrimination and duration dependence in the job finding rate. *The Review of Economic Studies 86*(4), 1631–1665.

Krueger, A. B. and A. Mueller (2011). Job Search, Emotional Well-Being and Job Finding in a Period of Mass Unemployment: Evidence from High-Frequency Longitudinal Data. *Brookings Papers on Economic Activity 42*(1), 1–81.

Lafuente, C. (2020). Unemployment in administrative data using survey data as a benchmark. *SERIEs: Journal of the Spanish Economic Association 11*, 115–153.

Lockwood, B. (1991). Information externalities in the labour market and the duration of unemployment. *The Review of Economic Studies 58*(4), 733–753.

Machin, S. and A. Manning (1999). The causes and consequences of long term unemployment in Europe. *Handbook of Labor Economics 3*, 3085–3139.

Mueller, A. I. (2017). Separations, sorting, and cyclical unemployment. *American Economic Review 107*(7), 2081–2107.

Mueller, A. I., J. Spinnewijn, and G. Topa (2019). Job seekers' perceptions and employment prospects: Heterogeneity, duration dependence and bias. NBER Working Paper No. 25294.

OECD (2017). Long-term unemployment rate (indicator). doi:10.1787/76471ad5-en.

Rebollo-Sanz, Y. F. and J. I. García-Pérez (2015). Are unemployment benefits harmful to the stability of working careers? the case of Spain. *SERIEs: Journal of the Spanish Economic Association 6*, 1–41.

Rebollo-Sanz, Y. F. and N. Rodríguez-Planas (2020). When the going gets tough... reducing benefits in the aftermath of the Great Recession. *Journal of Human Resources 55*(1), 119–163.

Schmitt, J. and J. Wadsworth (1993). Unemployment benefit levels and search activity. *Oxford Bulletin of Economics and Statistics 55*(1), 1–24.

Stoyan, D. and D. J. Daley (1983). *Comparison Methods for Queues and Other Stochastic Models*. John Wiley Sons.

Tatsiramos, K. and J. C. van Ours (2014). Labor market effects of unemployment insurance design. *Journal of Economic Surveys 28*(3), 284–311.

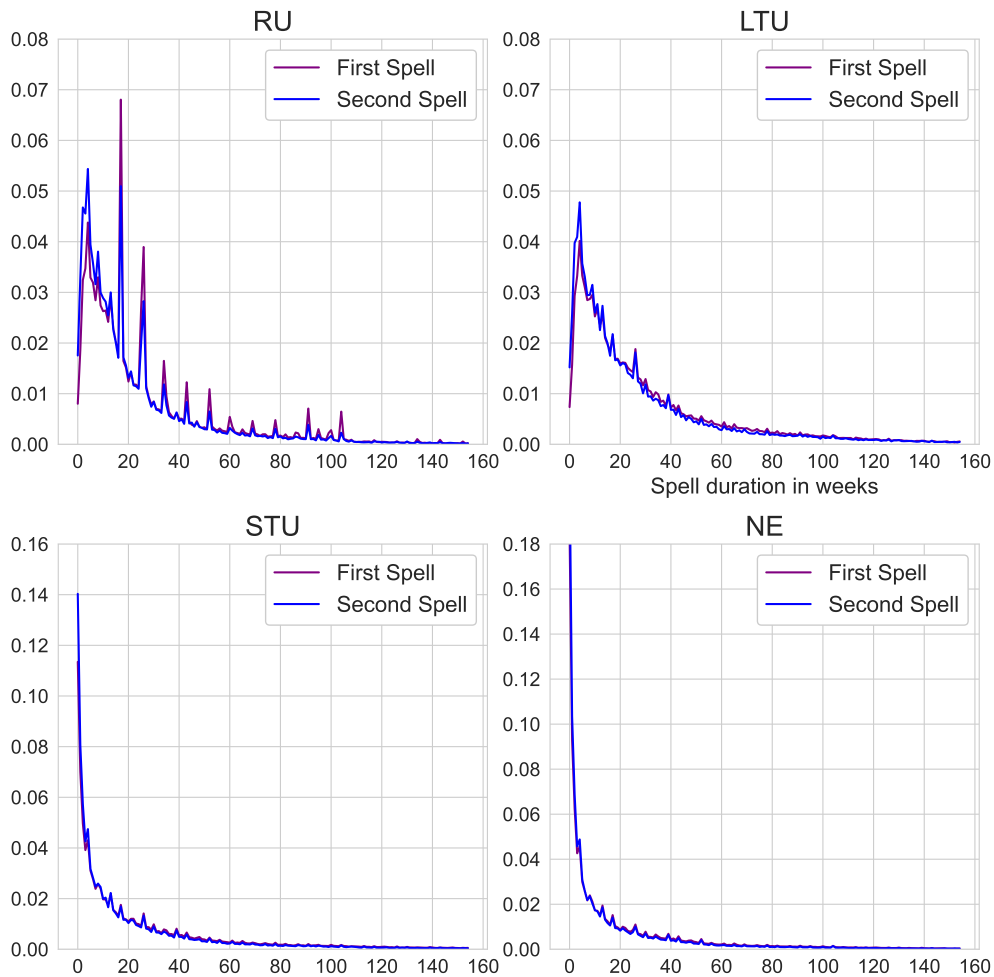
Turon, H. (2003). Inflow Composition, Duration Dependence and Their Impact on The Unemployment Outflow Rate. *Oxford Bulletin of Economics and Statistics 65*(1), 31–47.

White, J. S. (1969). The moments of log-weibull order statistics. *Technometrics 11*(2), 373–386.

A Appendix: Comparison of first and second spells

Figure A.1 displays the histogram for the first spells and the second spells in our data in the RU, LTU, STU and NE interpretations. The RU graph shows that when completely ignoring the blanks, the first and second spells have some differences. While the other three graphs show a similar finding: the first and second spells are very similar.

Figure A.1: Duration of unemployment, first and second spells



Notes: Histograms of first spell and second spells. Number of individuals: 436,634 in the NE sample, 336,228 in the STU sample, 200,357 in the LTU sample and 220,825 in the RU sample. Data source: MCVL.

B Appendix: Detailed results

As in ABS, we decompose the variance of duration in logs, that is, the natural logarithm of the unemployment spell in days. If the duration of unemployment in days follows an exponential distribution, then the duration in log-days follows a log-Weibull or Gumbel distribution. In this case of constant hazard in levels, the variance of the log of spell duration is $\pi^2/6 \approx 1.64$ (see White 1969), and it will larger if the hazard rate is decreasing. Thus, when measuring the logarithm of the duration, the constant hazard within variance is $\pi^2/6$. Then, *the within variance* in equation (3) becomes

$$\begin{aligned} \sigma_w^2 = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 = \\ \frac{1}{N}\sum_{i=1}^{N}\left(\sigma_i^2 - \frac{\pi^2}{6}\right) + \frac{\pi^2}{6}, \end{aligned} \tag{6}$$

where the first term is the *excess within variance* and the second one ($\pi^2/6$) is the *constant hazard within variance*.

Tables B.1 to B.5 report the full results of the different decompositions in terms of the different variance components in equation (5) and the share of the total variance they represent.

Table B.1: Variance decomposition in levels and log duration, different interpretations

	Raw	LTU	STU	NE
$\hat{\mu}_y$ (days)	188.47	251.36	225.21	181.84
$\hat{\sigma}_y$	66,006.31	136,710.30	170,835.90	142,086.80
Duration in levels				
$\hat{\sigma}_c$	42,575.36	81,526.63	71,360.46	48,872.06
	(0.65)	(0.60)	(0.42)	(0.34)
$\hat{\sigma}_e$	16,377.16	36,837.83	78,832.23	77,406.55
	(0.25)	(0.27)	(0.46)	(0.54)
$\hat{\sigma}_b$	7,053.79	18,345.88	20,643.17	15,808.17
	(0.11)	(0.13)	(0.12)	(0.11)
Duration in logs				
$\hat{\sigma}_s$	0.01	0.01	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.00)
$\hat{\sigma}_c$	1.64	1.64	1.64	1.64
	(1.23)	(1.09)	(0.56)	(0.49)
$\hat{\sigma}_e$	-0.60	-0.50	0.77	0.90
	(-0.47)	(-0.34)	(0.25)	(0.26)
$\hat{\sigma}_b$	0.29	0.36	0.52	0.78
	(0.23)	(0.24)	(0.18)	(0.24)
N (spells)	441,650	400,714	672,456	873,268

Notes: Decomposition based on equation (5). Shares of each component in parenthesis.
Data source: MCVL.

Table B.2: Variance decomposition in levels and log duration, with age and gender controls

	Spell Control Only	All Controls
$\hat{\mu}_y$ (days)	225.21	225.21
$\hat{\sigma}_y$	170,835.90	170,835.90
Duration in levels		
$\hat{\sigma}_c$	71,360.46	
	(0.42)	
$\hat{\sigma}_e$	78,832.23	
	(0.46)	
$\hat{\sigma}_b$	20,643.17	
	(0.12)	
Duration in logs		
$\hat{\sigma}_s$	0.02	0.08
	(0.01)	(0.03)
$\hat{\sigma}_c$	1.64	1.64
	(0.56)	(0.56)
$\hat{\sigma}_e$	0.74	0.74
	(0.25)	(0.25)
$\hat{\sigma}_b$	0.54	0.48
	(0.18)	(0.16)
N (spells)	672,456	672,456

Notes: Decomposition based on equation (5). Shares of each component in parenthesis. STU interpretation of blank spells. *All Controls* column includes: spell number dummy, male dummy and college dummy. Data source: MCVL.

Table B.3: Variance decomposition in levels and log duration, by gender and education

	Women	Men	College	Non-College
$\hat{\mu}_y$ (days)	283.31	177.77	197.46	230.11
$\hat{\sigma}_y$	247,415.30	103,309.30	125,761.40	178,642.70
Duration in levels				
$\hat{\sigma}_c$	104,577.40	44,241.58	52,031.05	74,776.85
	(0.42)	(0.43)	(0.41)	(0.42)
$\hat{\sigma}_e$	118,522.90	46,428.03	60,689.89	82,038.80
	(0.48)	(0.45)	(0.48)	(0.46)
$\hat{\sigma}_b$	24,314.96	12,639.70	13,040.45	21,827.08
	(0.10)	(0.12)	(0.10)	(0.12)
Duration in logs				
$\hat{\sigma}_s$	0.02	0.01	0.02	0.02
	(0.01)	(0.00)	(0.01)	(0.01)
$\hat{\sigma}_c$	1.64	1.64	1.64	1.64
	(0.55)	(0.58)	(0.54)	(0.56)
$\hat{\sigma}_e$	0.87	0.69	0.95	0.74
	(0.28)	(0.24)	(0.30)	(0.24)
$\hat{\sigma}_b$	0.45	0.49	0.45	0.53
	(0.16)	(0.18)	(0.16)	(0.19)
N (spells)	302,246	370,210	101,002	571,454

Decomposition based on equation (5). Shares of each component in parenthesis. Notes: STU interpretation of blank spells. Data source: MCVL.

Table B.4: Variance decomposition in levels and log duration, over the business cycle

	2002-2007	2008-2013
$\hat{\mu}_y$ (days)	164.80	201.55
$\hat{\sigma}_y$	88,063.33	77,577.36
Duration in levels		
$\hat{\sigma}_c$	39,413.73	45,366.67
	(0.45)	(0.58)
$\hat{\sigma}_e$	36,393.47	27,466.68
	(0.41)	(0.35)
$\hat{\sigma}_b$	12,256.13	4,744.01
	(0.14)	(0.06)
Duration in logs		
$\hat{\sigma}_s$	0.04	0.02
	(0.01)	(0.01)
$\hat{\sigma}_c$	1.64	1.64
	(0.60)	(0.70)
$\hat{\sigma}_e$	0.57	0.44
	(0.18)	(0.17)
$\hat{\sigma}_b$	0.54	0.28
	(0.21)	(0.13)
N (spells)	301,482	332,684

Notes: Decomposition based on equation (5). Shares of each component in parenthesis. STU interpretation of blank spells. Data source: MCVL.

Table B.5: Variance decomposition in levels and log duration, over the business cycle

	Non-College		College	
	2002-2007	2008-2013	2002-2007	2008-2013
$\hat{\mu}_y$ (days)	166.41	205.75	155.15	174.20
$\hat{\sigma}_y$	90,737.88	80,664.35	71,921.82	56,583.48
Duration in levels				
$\hat{\sigma}_c$	40,643.58	47,121.37	32,040.44	33,923.04
	(0.45)	(0.58)	(0.45)	(0.60)
$\hat{\sigma}_e$	37,141.29	28,752.39	31,910.08	19,081.71
	(0.41)	(0.36)	(0.44)	(0.34)
$\hat{\sigma}_b$	12,953.00	4,790.59	7,971.30	3,578.73
	(0.14)	(0.06)	(0.11)	(0.06)
Duration in logs				
$\hat{\sigma}_s$	0.03	0.02	0.05	0.02
	(0.01)	(0.01)	(0.02)	(0.01)
$\hat{\sigma}_c$	1.64	1.64	1.64	1.64
	(0.60)	(0.70)	(0.57)	(0.66)
$\hat{\sigma}_e$	0.53	0.42	0.78	0.56
	(0.17)	(0.16)	(0.24)	(0.21)
$\hat{\sigma}_b$	0.56	0.27	0.46	0.28
	(0.22)	(0.13)	(0.17)	(0.12)
N (spells)	258,384	288,454	43,098	44,230

Notes: Decomposition based on equation (5). Shares of each component in parenthesis. STU interpretation of blank spells. Data source: MCVL.

C Appendix: The labour markets in Austria and Spain

In this appendix we describe the key labor market differences between Austria and Spain, to give more context to the comparisons presented in Figures 4 and 6 in the main text.

In terms of overall unemployment dynamics, figure C.1 shows the unemployment and long-term unemployment rates for both countries since 1990 (1994 in the case of Austria).[22] The first thing to note is that Spain's unemployment rate is substantially higher than Austria's, and its volatility is higher as well. While for Austria unemployment has stayed between 4% and 5%, in Spain unemployment fluctuates between 8% in 2007 and 26% in 2013. The right panel also shows that the share of long-term unemployment is more volatile in Spain, going from over 50% in 1990 to under 20% in 2007. Long-term unemployment is more relevant in Spain, but in the 1995-2008 period it decreased considerably, partly because of the widespread adoption of temporary contracts (see Güell and Hu 2006). Austria's long-term unemployment rate is one of the lowest in the European Union, with levels similar to those in Denmark, Luxembourg, and Sweden. Its share over total unemployment fluctuates between 20% and 30%. Business cycles are clearly marked in Spain, while in Austria fluctuations are small and close to a trend.

Figure C.1: Unemployment and Long-Term Unemployment, Austria and Spain

Data source: OECD (2017).

A key labor market institution relating to unemployment duration is unemployment

[22]The OECD defines long-term unemployment as "people who have been unemployed for 12 months or more."

insurance, which is different in both countries. Table C.1 compares the requirements to claim benefits (in the form of previous contribution to social security), generosity as measured by the replacement rate (percentage of previous wages), and maximum duration of entitlement.

Overall, Austria has a less generous system, with a constant replacement rate of 55% of average net wages over the previous year, while Spain has very generous short-term benefits (70% over the average net wage over the last 3 months), but benefits are lower after 3 months. The main difference comes with time of entitlement duration: in Spain it is proportional to the contribution period (a month for every 3 months of employment) and up to 2 years, while in Austria, for those younger than 40 years of age, it is only 6 months. Workers older than 40 and 50 years of age with enough contribution periods can be granted up to a year of unemployment insurance.

Table C.1: Structure of Unemployment Insurance

	Austria	Spain
Minimum contribution period	12 months in 2 years$^{(a)}$	12 months in 6 years
Replacement rate	55%	70% in the first 3 months, 50% after
Maximum duration	1 year$^{(b)}$	2 years$^{(c)}$

Notes: (a) For young people below 25 years of age, the threshold is 6 months in the last year. (b) Maximum duration only available to those over 50 years of age with sufficient contribution period. For those under 40, maximum duration is 6 months. (c) As in Austria, duration of unemployment benefits is contribution-based. Data from the European Commission. (http://ec.europa.eu/social/main.jsp?catId=1101&langId=en&intPageId=4410).

In both countries, workers have to report to the employment office and regularly meet with an employment advisor to verify that they are actively looking for employment and to show that they have not declined any suitable job offers. In the case of expiration of unemployment insurance, both countries offer unemployment assistance benefits for those who have run out of unemployment insurance. In Austria, these benefits can last up to a year or indefinitely if certain conditions are met. In Spain, benefits can last up to 11 months or up to 18 months if the individual has family responsibilities or if she is older than 55. In both cases, the amount of benefits is a share of the minimum living income calculated by both countries and independent of previous earnings.

D Appendix: Job-to-job transitions

A general caveat of interpreting short blank spells, especially those without UI, is that they could be a job-to-job transition. Still, the frequency of social security data makes this issue potentially less problematic than when using quarterly survey data. We perform a robustness check dropping spells shorter than 2 weeks. Table D.1 reports the results.

Table D.1: Variance decomposition in levels and log duration, days after 2 weeks, different interpretations

	Raw	LTU	STU	NE
$\hat{\mu}_y$	182.91	248.13	254.68	223.08
$\hat{\sigma}_y$	65,692.65	137,572.80	176,887.10	162,337.90
Duration in levels				
$\hat{\sigma}_c$	40,491.46	79,856.00	85,697.81	67,601.41
	(0.62)	(0.58)	(0.48)	(0.42)
$\hat{\sigma}_e$	18,167.22	39,427.99	70,354.35	76,901.37
	(0.28)	(0.29)	(0.40)	(0.47)
$\hat{\sigma}_b$	7,033.96	18,288.75	20,834.91	17,835.13
	(0.11)	(0.13)	(0.12)	(0.11)
Duration in logs				
$\hat{\sigma}_s$	0.01	0.00	0.00	0.01
	(0.01)	(0.00)	(0.00)	(0.00)
$\hat{\sigma}_c$	1.64	1.64	1.64	1.64
	(1.05)	(0.98)	(0.73)	(0.60)
$\hat{\sigma}_e$	-0.38	-0.35	0.25	0.59
	(-0.26)	(-0.21)	(0.11)	(0.21)
$\hat{\sigma}_b$	0.31	0.38	0.34	0.48
	(0.20)	(0.23)	(0.15)	(0.18)
N (spells)	421,054	381,522	554,766	735,266

Notes: Decomposition based on equation (5). Shares of each component in parenthesis.
Data source: MCVL.

In levels, the decompositions for any interpretation remain very much unchanged. As expected, in logs there are some changes as relatively more weight is put in these very short durations. Focusing on the STU interpretation, the heterogeneity component falls somehow, which is compatible with such short duration unemployed workers being different to the rest of unemployment pool. Our sample becomes more homogeneous after removing them. Also, while the within variance is somehow unchanged, duration dependence falls and the component related to aggregate frictions increases. Removing such short durations makes the hazard function less stretched, making duration dependence less prominent.

E Appendix: Descriptive statistics

Table E.1: Descriptive statistics, STU sample

	Mean	Std.	Median
Duration (in days)	225.21	413.32	84
Duration of previous spell (in days)	372.54	733.94	125
Age	32.14	6.76	30
Female	0.45	0.50	0
College or above	0.15	0.36	0
Last spell was TC	0.64	0.48	1
Last spell was PC	0.30	0.46	0
Last spell was SE	0.06	0.24	0
Quit	0.32	0.47	0
N (spells)			672,456

Notes: *TC* stands for 'Temporary Contract', *PC* stands for 'Permanent Contract' and *SE* stands for 'Self-Employed'. Data source: MCVL.

Table E.2: Descriptive statistics, STU sample, over the business cycle

	2002-2007			2008-2013		
	Mean	Std.	Mdn	Mean	Std.	Mdn
Duration (in days)	164.80	296.75	63	201.55	278.53	103
Duration of previous spell (in days)	275.66	538.69	114	324.05	628.00	121
Age	33.31	6.92	32	34.64	7.04	34
Female	0.45	0.50	0	0.44	0.50	0
College or above	0.14	0.35	0	0.13	0.34	0
Last spell was TC	0.72	0.45	1	0.66	0.47	1
Last spell was PC	0.23	0.42	0	0.27	0.44	0
Last spell was SE	0.04	0.20	0	0.06	0.23	0
Quit	0.35	0.48	0	0.20	0.40	0
N (spells)		301,482			332,684	

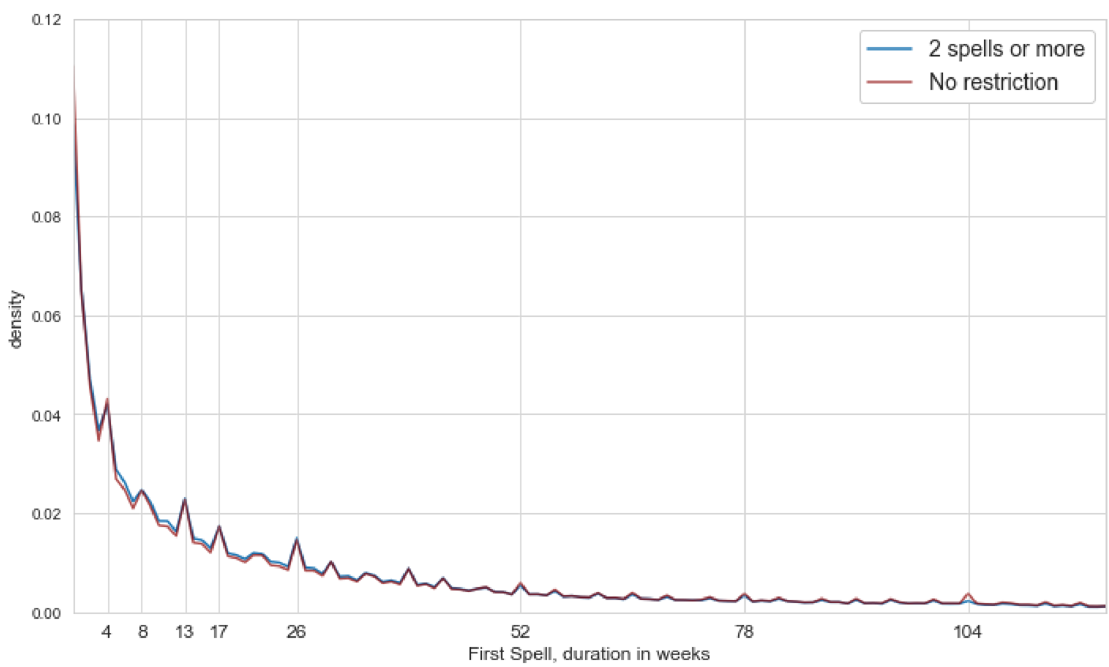
Notes: *TC* stands for 'Temporary Contract', *PC* stands for 'Permanent Contract' and *SE* stands for 'Self-Employed'. Data source: MCVL.

F Two spell requirement selection

The ABS method requires two completed spells of unemployed per individual. This requirement could be relevant in terms of a potential sample selection.

In a companion paper, Alvarez, Borovičková and Shimer show that the distributions of unemployment duration with and without imposing restrictions of the number and duration of spells are very similar for Austria (see Figure 13 in Alvarez, Borovičková, and Shimer 2016). Figure F.1 is the equivalent figure for Spain. That is we compare the distributions of unemployment duration for the restricted sample (i.e., those with two spells of unemployment or more) and the unrestricted sample (i.e., those with one spell of unemployment or more). As the figure shows, we share the same finding as Alvarez, Borovičková, and Shimer (2016). That is, the two distributions are very similar. This suggests that sample selection bias is not likely to be driving our results.

Figure F.1: Two-Spell-requirement selection



Notes: Histogram of first unemployment spells. Number of individuals: 336,228 for the sample with *2 spells or more* and 483,235 for the sample with *No restriction* (one spell or more). Data source: MCVL.

We next check how representative is the sample of unemployed workers with two spells among all unemployed workers. Our STU sample has 336,228 individuals with 2 spells or more. Table F.1 also reports the number of individuals with only one spell of unemployment. It is worth noticing that the subsample of unemployed used in our analysis corresponds to 70% of unemployed workers (more precisely, workers with at least one day of registered unemployment).

Table F.1: Unemployed by number of spells

	Individuals	Share
With 1 spell	147,007	0.30
With 2 spells or more	336,228	0.70
Total	483,235	1.00

Data source: MCVL. Individuals *with 2 spells or more* correspond to the STU sample.

Table F.2: Unemployed workers by number of spells, by education

	Individuals		Share	
	Non-College	College	Non-College	College
With 1 spell	120,992	26,015	0.30	0.34
With 2 spells or more	285,727	50,501	0.70	0.66
Total	406,719	76,516	1.00	1.00

Data source: MCVL. Individuals *with 2 spells or more* correspond to the STU sample.

Table F.2 repeats this accounting for college and non-college workers. Our sample of no-college workers constitutes 70% of the unemployed with a non college degree. While our sample of college graduates, despite the smaller number of observations, they still represent 66% of unemployed workers with a college degree.