DISCUSSION PAPER SERIES

DP13610
(v. 2)

UNEMPLOYMENT DURATION VARIANCE DECOMPOSITION A LA ABS: EVIDENCE FROM SPAIN

Maia Güell and Cristina Lafuente

LABOUR ECONOMICS

ISSN 0265-8003

UNEMPLOYMENT DURATION VARIANCE DECOMPOSITION A LA ABS: EVIDENCE FROM SPAIN

Maia Güell and Cristina Lafuente

Discussion Paper DP13610
First Published 22 March 2019
This Revision 15 April 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

UNEMPLOYMENT DURATION VARIANCE DECOMPOSITION A LA ABS: EVIDENCE FROM SPAIN

Abstract

In a recent paper, Alvarez, Borovickova, and Shimer (2014) revisit the analysis of the determinants of unemployment duration by proposing a new method (the ABS method hereafter) that directly estimates the importance of each component and implementing it using precise information on unemployment spells from social security administrative data for Austria. In this paper, we apply the ABS method to social security administrative data for Spain with the objective of comparing these two very different labor markets as well as Spain along the business cycle. Administrative data have many advantages compared to Labor Force Survey data, but the incomplete nature of the data needs to be addressed in order to use the data for unemployment analysis (e.g., unemployed workers that runout of unemployment insurance have no labor market status in the data). The degree and nature of such incompleteness are country-specific and are particularly important in Spain. Following Lafuente (2018), we approach the matter of data incompleteness in a systematic way by using information from the Spanish LFS data as well as institutional information. We hope that our approach will provide a useful way to apply the ABS method in other countries. Our findings are as follows: (i) The aggregate component is clearly the most important one, followed by heterogeneity and duration dependence, which are roughly comparable. (ii) The relative importance of each component and, in particular, duration dependence is quite similar in Austria and Spain, especially when minimizing the effect of fixed-term contracts in Spain. Similarly, we do not find big differences in the relative contribution of the different components along the business cycle in Spain. (iii) These comparisons suggest that statistical discrimination due to dynamic sample selection does not seem to be the main driver of duration dependence.

JEL Classification: E24, J64

Keywords: Unemployment Duration, administrative social security data, Duration Dependence

Maia Güell - maia.guell@gmail.com
The University of Edinburgh, FEDEA, IZA and CEPR

Cristina Lafuente - cristina.lafuente@ed.ac.uk
University of Edinburgh

Acknowledgements
We are very grateful to Carlos Carrillo-Tudela, Mike Elsby, Nacho Garcia-Perez, Rafael Lopes de Melo, John Moore, Laura Pilossoph, Jose V. Rodriguez Mora, Rob Shimer, 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, and CUNEF. We are grateful to Joan Gieseke for editorial assistance.

Unemployment Duration Variance Decomposition à la ABS: Evidence from Spain*

Maia Güell[†] Cristina Lafuente[‡]

This version: April 2020

Abstract

In a recent paper, Alvarez, Borovičková, and Shimer (2014) revisit the analysis of the determinants of unemployment duration by proposing a new method (the ABS method hereafter) that directly estimates the importance of each component and implementing it using precise information on unemployment spells from social security administrative data for Austria. In this paper, we apply the ABS method to social security administrative data for Spain with the objective of comparing these two very different labor markets as well as Spain along the business cycle. Administrative data have many advantages compared to Labor Force Survey data, but unemployment spells without benefit entitlement are underreported. The degree and nature of such underreporting are country-specific and are particularly important in Spain. We propose a strategy to account for such spells by using information from the Spanish LFS data as well as institutional information. We hope that our approach will provide a useful way to apply the ABS method in other countries with similar data. Our findings are as follows: (i) The aggregate component is clearly the most important one, followed by heterogeneity and duration dependence, which are roughly comparable. (ii) The relative importance of each component and, in particular, duration dependence is quite similar in Austria and Spain, especially when minimizing the effect of fixed-term contracts in Spain. Similarly, we do not find big differences in the relative contribution of the different components along the business cycle in Spain. (iii) These comparisons suggest that statistical discrimination due to dynamic sample selection does not seem to be the main driver of duration dependence.

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

*We are very grateful to 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, and CUNEF. 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
[‡]European University Institute. Email: Cristina.Lafuente@eui.eu

1 Introduction

Existing studies of the determinants of unemployment duration traditionally estimate duration models using self-reported duration information from Labor Force Survey data. In a recent paper, Alvarez, Borovičková, and Shimer (2014) revisit this question by proposing a new method and implementing it using precise information on unemployment spells from social security administrative data for Austria. In this paper, we apply this method (the ABS method hereafter) to social security administrative data for Spain. Our goal is to compare the decomposition of the duration of unemployment of these two very different labor markets using comparable administrative data as well as to compare such decomposition along the business cycle for Spain. These comparisons will allow us to shed some light on which theories can explain duration dependence.

Traditionally, unemployment duration has been studied through the lens of duration models, which generally assume a parametric structure of the hazard function, distinguishing between duration dependence and unobserved heterogeneity parameters.[1] These methods require long enough panel data (i.e., large t) as well as some functional form assumptions. The ABS method is a more statistical approach that directly estimates the relative contribution of each component to the variance of any potential duration variables, unemployment duration in particular. This 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 (i.e., small t) per individual will suffice. In particular, the method can be estimated with just two completed spells of a given labor market status per individual. The nature of social security administrative data makes these data a natural fit for this method.

The contribution of this paper is twofold. The first one is methodological. Administrative social security data offer several advantages with respect to standard Labor Force Survey data. Administrative data have precise information on spell duration, unlike self-reported information from survey data. But social security administrative data were not designed to study unemployment nor unemployment duration: these data are 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). Unemployment spells without

[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.

benefit entitlement will be underreported. However, the fact that we observe the complete history of individual spells will allow us to identify them and propose a strategy to account for them. Dealing with underreported spells is important because estimates are sensitive to the implementation of treatment. One approach is to focus the analysis on non-employment rather than unemployment as followed by Alvarez, Borovičková, and Shimer (2014). However, for high unemployment countries, it is worthwhile to distinguish between unemployment and non-employment. Thus, in this paper, we take another approach in an attempt to distinguish unemployment from other situations by using complementary information. Specifically, we approach these underreported spells in a systematic way by using information from the Spanish Labor Force Survey as well as institutional information, as in Lafuente (2020). The degree and nature of such underreporting—which also interacts with labor market institutions—are country-specific. The complexity and dysfunctionality of the Spanish labor market is such that we hope that our proposed approach for the Spanish case will provide guidance that is general enough to be useful in applying this method to other countries with similar data.

The second contribution of the paper is to use our comparison exercises to shed some light on which theories can explain duration dependence. Indeed, we compare Spain, a high-unemployment-rate country with dual labor market dynamics, to Austria, which is characterized by lower unemployment rates and a more stable labor market. The Spanish case is also particularly relevant in this context because of the high quality of the Spanish social security data, which feature large n as well as large t. This also allows us to study how the composition of the duration of unemployment changes along the business cycle in Spain. To our surprise, we find that Spain and Austria, two very different labor markets, do not actually differ much in the relative importance of each component of the variance of the duration of unemployment, in particular with respect to the importance of duration dependence. In a similar vein, we find that within Spain, the decomposition is very similar along the business cycle. These comparisons will allow us to discard theories behind duration dependence that vary with the level of unemployment.

Substantial work has been written on unemployment duration in Spain.[2] Our contribution to this literature is to provide a measure of the relative importance of each of the determinants

[2]See for instance, Bover, Arellano, and Bentolila (2002), Bentolila, García-Pérez, and Jansen (2016), and Rebollo-Sanz and Rodríguez-Planas (2020).

of unemployment duration. Also, as we will explain, the treatment of some of the underreported spells in the administrative data has an economic interpretation in terms of the incidence of such temporary contracts. This approach will allow us to analyse to some extent the role of such contracts in the decomposition of the variance of unemployment duration.

The rest of the paper is organized as follows. Section 2 explains the ABS method in more detail. Section 3 presents the Spanish social security administrative data and our proposed approach to dealing with underreporting. Section 4 provides a general context for comparing the labor market in Austria and Spain. Section 5 explains our results. Section 6 discusses the existing theories of duration dependence and how our results relate to them. Finally, Section 7 concludes.

2 The ABS method

In this section, we explain the ABS method proposed in more detail. The starting point is to assume that 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.

The population mean and variance are given, respectively, by

$$\overline{\mu}_y = \frac{1}{n}\sum_{i=1}^{n}\int y dF_i(y) \quad \text{and} \quad \overline{\sigma}_y^2 = \frac{1}{n}\sum_{i=1}^{n}\int (y-\overline{\mu}_y)^2 dF_i(y), \tag{1}$$

where n is population size.

The population variance can be decomposed into two components that in the ABS method are labelled as the *within* and the *between* components. Let's start with the latter.

The between component represents the variance that comes from heterogeneity in the distribution functions $F_i(y)$ across individuals. That is,

$$\overline{\sigma}_b^2 = \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \bar{\mu}_y)^2. \tag{2}$$

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. In this case, the total variance in unemployment duration would be attributable to individual heterogeneity captured by equation (2). Note that this heterogeneity captures both observed and unobserved characteristics of the individual.

Now consider the opposite case, in which all individuals draw from the same distribution function and thus the between variance is zero. In this case, the variance in duration comes from within an individual's draws. That is, the different spells from the same individual are potentially different. The within component is the average individual variance of the outcome y. That is,

$$\overline{\sigma}_w^2 = \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2. \tag{3}$$

Now consider the within component. 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 variance is given by $\overline{\sigma}_y^2 = 1/h^2$. The ABS method labels this as the "constant hazard within variance." This component can be thought as the aggregate component, that is, the macroeconomic component that affects all workers equally. 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." The ABS method labels this as the "excess within variance." Given this, we can decompose the within variance ($\overline{\sigma}_w^2$) into the constant hazard within variance and the excess within variance, which are, respectively, given by

$$\overline{\sigma}_c^2 = \frac{1}{n}\sum_{i=1}^{n} \mu_i^2 \quad \text{and} \quad \overline{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^{n} (\sigma_i^2 - \mu_i^2). \tag{4}$$

As explained in Alvarez, Borovičková, and Shimer (2014), if the hazard rate is non-increasing and non-constant, then $\sigma_i > \mu_i$ and the excess within variance would be positive. This is what one should expect if there is the so-called negative duration dependence.

Note that the focus of the ABS method is different to the traditional duration models. The goal of the ABS method is to estimate the relative importance of these three components. In particular, duration dependence is measured once all heterogeneity (observed and unobserved) has already been accounted.

2.1 Estimation with two spells

Alvarez, Borovičková, and Shimer (2014) show that two completed spells per individual are enough to identify the three different components mentioned above. In particular, they show how to build consistent estimators for all of the above using only two spells.

Let $y_{i,j}$ be the completed spell j for individual i. Suppose that we observe two completed spells for each individual, that is, $y_{i,1}$ and $y_{i,2}$. Then the unbiased estimators for the individual mean and variance are, respectively,

$$\hat{\mu}_i = \frac{y_{i,1} + y_{i,2}}{2},$$

$$\hat{\sigma}_i^2 = \frac{1}{J-1} \sum_{j=1}^{J} (y_{i,j} - \hat{\mu}_i)^2 = \frac{(y_{i,1} - y_{i,2})^2}{2}.$$

And the unbiased estimators for the sample mean and variance are, respectively,

$$\hat{\bar{\mu}}_y = \frac{1}{2n} \sum_{i=1}^{n} (y_{i,1} + y_{i,2}), \tag{5}$$

$$\hat{\bar{\sigma}}_y^2 = \frac{1}{2n-1} \sum_{i=1}^{n} ((y_{i,1} - \hat{\bar{\mu}}_y)^2 + (y_{i,2} - \hat{\bar{\mu}}_y)^2). \tag{6}$$

As $\hat{\sigma}_i^2$ is an unbiased estimator of σ_i^2, then the unbiased estimator of the within variance ($\overline{\sigma}_w^2$) is

$$\overline{\sigma}_w^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\sigma}_i^2. \tag{7}$$

For the between variance component, we need unbiased estimators for the squares of the individual and populations means (μ_i^2 and $\overline{\mu}_y^2$), as shown in equation (2). Taking expectations, these can be shown[3] to be

$$\mathbb{E}(\mu_i) = \hat{\mu}_i^2 - \hat{\sigma}_i^2$$

$$\mathbb{E}(\bar{\mu}_y) = \hat{\bar{\mu}}_y - \frac{1}{2n} \hat{\bar{\sigma}}_y^2.$$

[3]For details, refer to Alvarez, Borovičková, and Shimer (2014), Section 3.2.

So the unbiased estimator of the between component is

$$\overline{\sigma}_b^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_i^2 - \hat{\sigma}_i^2) - (\hat{\overline{\mu}}_y - \frac{1}{2n}\hat{\overline{\sigma}}_y^2). \tag{8}$$

Similarly, we can reformulate the equations in (4) to obtain

$$\hat{\overline{\sigma}}_e^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}_i^2 - \tfrac{1}{2}\hat{\sigma}_i^2) \quad \text{and} \quad \hat{\overline{\sigma}}_c^2 = \frac{1}{n}\sum_{i=1}^{n}(\tfrac{3}{2}\hat{\sigma}_i^2 - \hat{\mu}_i^2). \tag{9}$$

In the particular case in which we decompose the natural logarithm of the duration of unemployment, the constant hazard component is always equal to $\pi^2/6$, while the excess variance estimator will then be

$$\hat{\overline{\sigma}}_e^2 = \hat{\overline{\sigma}}_w^2 - \hat{\overline{\sigma}}_c^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_i^2 - \frac{\pi^2}{6}. \tag{10}$$

Notice that in order to have estimators that are as close as possible to the true components, it is crucial to have a large number of observations (large n) rather than a large number of spells J per individual.

More details on the ABS method can be found in Alvarez, Borovičková, and Shimer (2014).[4]

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 for the years 2004-2013. 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.[5] However these data were not designed to study the labor market; rather, they were designed to have an account of spells of employed workers (who contribute to social security) and spells of unemployed workers who are entitled to unemployment insurance (and are paid by social security). Therefore, the data displays *blank* spells whenever workers do not fall into one of these two categories. However, the fact that we observe the complete

[4]In particular, they relax the assumption that the distribution of the two spells is time-invariant. In this case, regressing duration on the number of spell is enough to control for changes in F_i through time.

[5]Because of the complete working histories, our data include spells dating back to the 1970s, which makes our data comparable to the Austrian data used in Alvarez, Borovičková, and Shimer (2014) that covers workers over the years 1972-2007.

history of individual spells will allow us to identify them and propose a strategy to account for them. In the following subsections, we explain the key data expansions, as in Lafuente (2020), needed for the particular application in this paper. 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.

3.1 Data expansions

Following Lafuente (2020), we will make data expansions in a systematic way by using information from the Spanish Labor Force Survey as well as institutional information.[6] In particular, we will make three expansions to deal with unemployment duration, the number of unemployed workers observed, and the spells being compared.

3.2 LTU expansion

A first source of underreporting in administrative data relates to unemployment duration. Recall that only unemployed workers receiving unemployment insurance are recorded in the data. Once their benefits run out, if they are still unemployed, they appear as *blank* in the data. This implies that unemployment duration in the raw data will be underestimated. We address this underreported duration by implementing what we label as the *long-term unemployment expansion* (LTU expansion), which we develop in this section.

Table 1 illustrates this point with an example of a worker during three consecutive spells. This example is common in our data. The first row displays the actual labor market status of a worker, and the second row displays what appears in the administrative data. In the first spell, the worker is unemployed and receiving unemployment insurance. For this spell, the data and the actual labor force status of the worker coincide. In the second spell, the data display a *blank* while the worker could be either still unemployed without unemployment insurance or out of the labor force. Finally, in the third spell the worker is employed, and this is recorded in the data.

The key here is the *blank* in spell 2. If indeed the worker continued to be unemployed (without insurance), then unless the data are expanded, we would be underestimating unemployment

[6]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 underreporting.

Table 1. Unemployment duration and LTU expansion

	spell 1	spell 2	spell 3
Actual	U with UI	(i) U run out UI, (ii) out LF	E
Data	Registered U	*blank*	E

Notes: U stands for unemployed, UI for unemployment insurance, LF for labor force, and E for employed.

duration (only counting spell 1). It is likely that in this example, the worker indeed continued to be unemployed given that in spell 3 the worker is once again employed. Also, given that we concentrate on the first two completed spells of unemployment observed in the data for any individual, this implies that generally workers are going to be relatively young and that inactivity is a less likely event. Following Lafuente (2020), this is addressed using complementary information in the Labor Force Survey. Using the Spanish LFS from 1987 to 2013, we find that as many as 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 labor force). Accordingly, as a first approximation we replace all the blanks as in Table 1 with unemployment status. Admittedly, this is a broad-brush solution, but it turns out to be very successful in this case. This approach has been implemented by other authors (see, e.g., García-Pérez 2008). This approach, while being broad, has the advantage of having a simple implementation. Moreover, as we will show, it leads to comparable results with respect to the non-employment approach by Alvarez, Borovičková, and Shimer (2014), which gives us confidence that it is not biasing the results in any meaningful way.

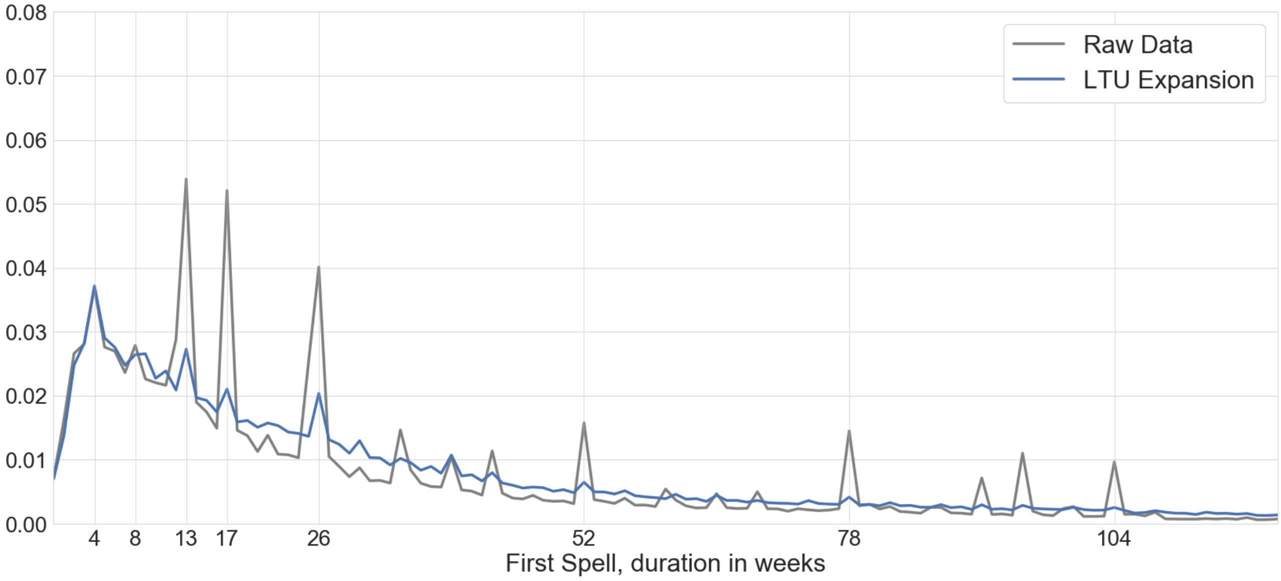
After implementing the LTU expansion, we document the following three facts:

1. Figure 1 shows a histogram of completed unemployment spells with the MCVL.[7] The raw data display pronounced spikes that coincide with the end of unemployment benefits (e.g., at three, four, and six months). This implies underestimating unemployment duration as workers may continue being unemployed without benefits. However, the figure shows that after the LTU expansion, these spikes are greatly reduced.[8]

[7]In this figure, we focus on the duration of the first completed unemployment spell observed for each individual in the data.

[8]Card, Chetty, and Weber (2007) also study the relationship between the way in which unemployment spells are measured and the magnitude of the spikes in the administrative data.

Figure 1. Duration in weeks of completed unemployment spells, LTU expansion



2. Figure 2 displays the duration distribution of the unemployment stock in 2013 Q3. The histogram plots the duration of (uncompleted) unemployment spells using the LFS data, the original MCVL, and the MCVL after the LTU expansion. As can be seen, the raw MCVL puts more weight on shorter durations than the LFS, but after the LTU expansion, the LFS and the MCVL are much more comparable.

3. As shown in Lafuente (2020) (Figure 4), the unemployment rate series using the MCVL after this expansion becomes closer to that of the LFS. The reason is that when calculating the quarterly unemployment rate with the MCVL, workers with blanks (as in Table 1) lasting for a full quarter will not be counted as unemployed.

3.3 STU expansion

We now turn to a second source of underreporting in the administrative data, which relates to the number of unemployed workers observed. Unemployed workers without the right of unemployment insurance (typically because of short employment tenures) appear as *blank* in the data. This implies that in the raw data the number of unemployed workers will be underestimated. We address these underreported spells by implementing what we label as the *short-term unemployment expansion* (STU expansion), which we develop in this section.[9]

[9]In Spain, Lafuente (2020) shows that workers who are ineligible to claim unemployment benefits account for most of the gap between LFS unemployment and administrative unemployment. Therefore we focus on this group when implementing the STU expansion. This implies that some workers who are eligible but do not claim

Figure 2. Duration in months of (uncompleted) unemployment in 2013 Q3, LTU expansion

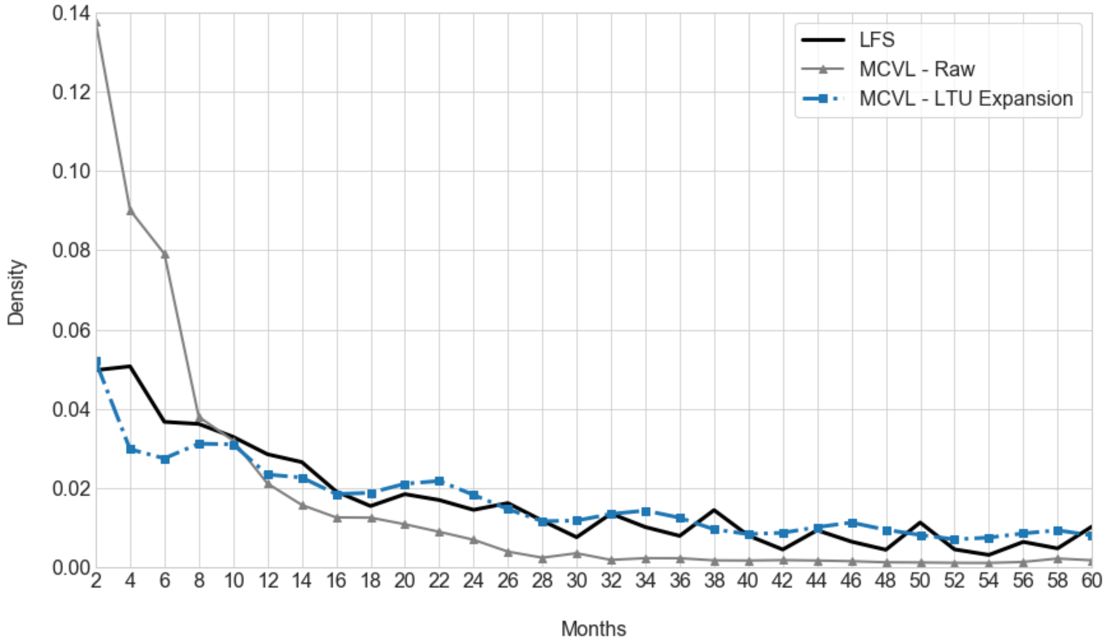


Table 2 illustrates this point with an example of a worker during three consecutive spells. As in Table 1, the first row displays the actual labor market status of a worker, while the second row displays what is recorded in the administrative data. The worker has two employment spells (spell 1 and spell 3) that are registered in the data. In between these employment spells there is a blank recorded in the data (spell 2). This spell could be unemployment without insurance, a job-to-job transition (including recalls to the same firm),[10] or movement out of the labor force. It is crucial to be able to distinguish between these possibilities as much as possible; otherwise, the number of unemployment spells could be underestimated.

In order to distinguish between unemployment without insurance and other possibilities we follow Lafuente (2020) and use the fact that the MCVL has the following key pieces of information: (i) the complete work histories (and thus we can identify who has the right to claim unemployment benefits given the institutional framework), (ii) the identity of the firms when employed (useful for identifying recalls), and (iii) the reason for job separation (which

are treated as out of the labour force transitions. The assumption here is that workers who are not claiming because they are not going to engage in active job search which is a requirement for receiving UI. Although 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 not such studies showing this is an important factor in Spain. See Lafuente (2020) for more details.

[10]Spain has a type of unstable temporary contract in which the worker does not work every day but instead frequently alternates between unemployment and employment in the same firm. This contract would be similar to a zero-hour contract in the UK. For simplicity, we refer to these cases as *recalls*.

Table 2. Number of unemployed and STU expansion

	spell 1	spell 2	spell 3
Actual	E	U no UI, JtoJ, or out of LF	E
Data	E	*blank*	E

Notes: U stands for unemployed, UI for unemployment insurance, JtoJ for job-to-job transitions, LF for labor force, and E for employed.

allows us to identify quits).

We proceed as follows. First, we identify recalls as job-to-job transitions to the same firm in spells 1 and 3 (see Table 2). In this case, spell 2 would not be considered as unemployment.

Second, when the worker does not return to the same employer in spell 3, then spell 2 could be unemployment, job-to-job transitions, or movement out of the labor force. One solution for these types of observations would be to treat them as *non-employment*, as in Alvarez, Borovičková, and Shimer (2014). However, in our case, we would like to distinguish unemployment from other situations. We use legislation on unemployment insurance to establish who is eligible based on employment histories and assign individually whether a spell corresponds to unemployment without insurance. These could be cases of: (i) wage earners with accumulated tenure in past employment that is too short (usually coming from temporary contracts), (ii) self-employed workers, or (iii) workers who have quit their job. In these cases, we replace the blanks in the data at the individual level with unemployment.

In our data, we find that around 81% of the blanks as described in Table 2 do not represent workers who have the right to claim unemployment insurance. In Spain, the bulk of these cases (around 70%) are workers on short contracts who alternate between short employment and short unemployment durations. The remaining 19% of the blanks do have the right to unemployment insurance and are therefore not replaced by unemployment status.

After adding the STU expansion to the previous one (the LTU expansion), we document the following three facts:

1. First, as shown in Lafuente (2020) (Figure 7), the unemployment rate series using the MCVL after this addition becomes closer to the LFS because after the STU expansion, we identify new unemployment spells, which are mostly short spells of unemployment.

2. It follows that we further smooth spikes that coincide with the end of unemployment

benefits, as Figure 3 shows. This figure also shows that compared to the LTU expansion by itself, this data expansion includes more short unemployment spells (less than four weeks). This implies that the hazard function of leaving unemployment is always decreasing for all durations.

FIGURE 3. Duration in weeks of completed unemployment spells, STU expansion

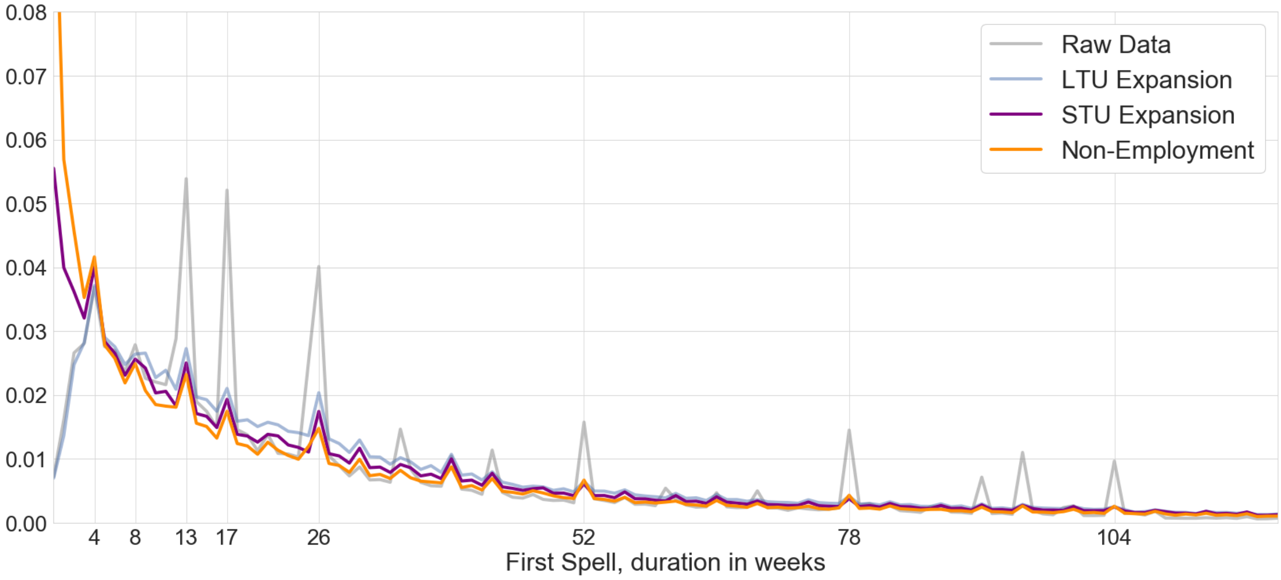


Figure 3 also includes the duration distribution when the blanks in the data are considered non-employment spells, as in Alvarez, Borovičková, and Shimer (2014). As can be seen, the comparison with respect to the raw data is similar to the STU expansion. We will use this non-employment approach when comparing the decompositions for Austria and Spain in Section 5.1.

3.4 Spell number adjustment

In order to implement the ABS method, the first two completed unemployment spells observed for every individual need to selected. This is straightforward in principle, except that some unemployment spells may not be detected in the original data, as explained in the previous subsection. After the STU expansion explained in subsection 3.3, the first two spells observed may not be the same as the first two spells in the original raw data. We address this spell selection issue by implementing what we label as the *spell number adjustment*, which we develop in this section.

Table 3 illustrates this point with an example of a worker during six consecutive spells (in order to include different groups for the first two spells of unemployment). The first row displays the actual labor market status of a worker. This worker has a spell of employment followed by a spell of unemployment without benefits, which is then followed by two sets of employment-unemployment (with benefits) spells. The second row displays the labor market status of a worker according to the raw administrative data. The data capture the actual labor market status except for spell 2, where the data display a blank. Note that the first two spells of unemployment that would be identified would be spells 4 and 6.

TABLE 3. Spell number adjustment

	spell 1	spell 2	spell 3	spell 4	spell 5	spell 6
Actual	E	U no UI	E	U w/UI	E	U w/UI
Data	E	*blank*	E	Registered U (spell #1)	E	Registered U (spell #2)
STU Expansion	E	"U"	E	Registered U (spell #1)	E	Registered U (spell #2)
Spell number adj.	E	U (spell #1)	E	Registered U (spell #2)	E	Registered U

Notes: U stands for unemployed, UI for unemployment insurance, JtoJ for job-to-job transitions, and E for employed.

Suppose that the blank in spell 2 corresponds to unemployment (without benefits). This will be captured after the implementation of the STU expansion, shown in third row. We label it "U" because the STU expansion alone would still consider spells 4 and 6 as the first two spells of unemployment (the first two originally identified).

This is corrected with the spell number adjustment, shown in fourth row. This implies that the first two spells of unemployment would now be spells 2 and 4 rather than spells 4 and 6.

This example has meaningful economic content. The case described corresponds to employment spells of quite short duration with no right to unemployment insurance. In Spain, these cases correspond to spells out of temporary contracts, which are very predominant (35% of the workforce is on temporary contracts). This means that we can somehow approximate the effect of temporary contracts in Spain to the question at hand. Using the first two spells originally identified as in the STU expansion implies ignoring that the employment spell in spell 1 is a temporary contract that leads to unemployment without benefits. Therefore, the STU expansion approximates to some extent the lack of temporary contracts in the economy so that employment would appear as being more stable than it actually is (i.e., some unemployment spells would be ignored in the context of this paper).

3.5 Expansions and number of observations

In this section, we summarize how the three expansions described above affect the number of observations in our sample. Table 4 breaks down the number of individual observations in our original sample (raw data) by their number of registered and non-registered unemployment spells. The six possible cases are described in the first column. The letters 'RU' stand for "registered unemployment" and refer to spells that are recorded as unemployment in the raw administrative data. Conversely, the letters 'NRU' stand for "non-registered unemployment" or gaps between employment spells that we include in the STU expansion.

The second column shows which types of observations are present in the raw data. By definition these are observations with two RU. Some of these might have earlier NRU, which the spell number adjustment will correct for, as the fourth column shows. The third column shows those types of observations that are not included in the original sample but are added after the implementation of the STU expansion. Again, some of these might require spell number adjustment.

Table 4. Number of individuals

Case	Raw data	Added by STU Exp	Spell num adj	Individuals
2 RU and 0 NRU	Yes		No	136,248
2 RU and 1 earlier NRU	Yes		Yes - 1 spell	100,302
2 RU and 2 earlier NRU	Yes		Yes - 2 spells	50,700
0 RU and 2 NRU	No	Yes	No	69,340
1 RU and 1 NRU	No	Yes	No	67,622
1 RU and 2 earlier NRU	No	Yes	Yes - 1 spell	28,948
Total				453,160

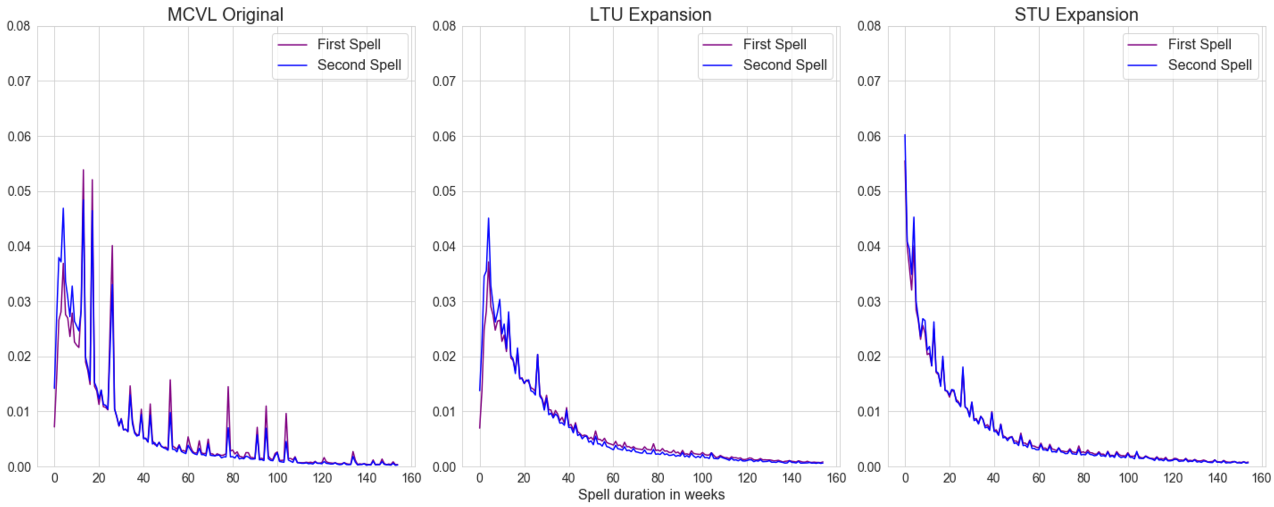
Notes: RU stands for "registered unemployment", and NRU stands for "non-registered unemployment". "2 RU" means that the individual has 2 or more registered unemployment spells.

After all these expansions, we are left with a sample of 1,138,386 individuals. Restricting the sample to the raw data leads to missing 37% of the individuals in our final sample. This highlights the importance of the STU expansion in adequately capturing those who do not have unemployment benefits. On the other hand, 40% of the sample had at least one earlier non-recorded unemployment spell. As explained, this reflects short employment spells (temporary contracts), which translate into unemployment without benefits. The spell number adjustment allows us to capture unemployment spells of a different nature.

3.6 Comparison of first and second spells

Before proceeding with the implementation of the ABS method, we need to check a necessary requirement in order to implement it. As explained, the method uses the first two spells of unemployment observed for each individual. The requirement is that these two spells are not too different. This means that the distribution from which the first spells are drawn is not too different from the distribution from which the second spells are drawn. We check this empirically in our data, as in Alvarez, Borovičková, and Shimer (2014). Figure 4 displays the hazard rates of the first spells (purple lines) and of the second spells (blue lines).

FIGURE 4. Hazard Rate by number of Spell



The panel on the left-hand side displays the hazard rates of the first and second spells using the raw data. As can be seen, the two spells are very similar. As discussed before, the pronounced spikes correspond to unemployment benefit expiration. The panels in the middle and on the right-hand side again display the hazard rates of the first and second spells after the LTU expansion (middle panel) and STU expansion (right panel) have been implemented. The three graphs show the same finding; that is, the two spells are very similar. As can be seen in the last two panels the expansions of the data make the spikes smoother.

4 The labor markets in Austria and Spain

In the results section, we will compare Austria and Spain. To that end, it is worth drawing attention to the key labor market differences between these two countries.

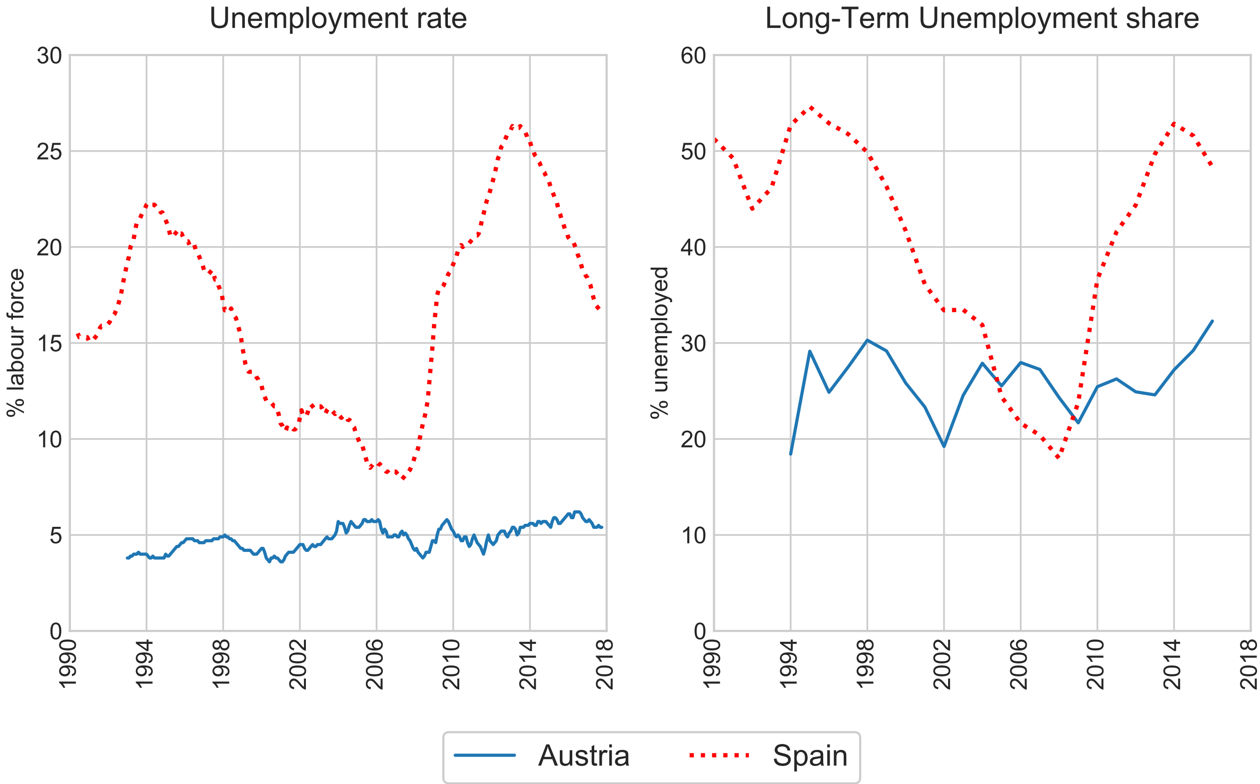
Figure 5 shows the unemployment and long-term unemployment rates for both countries since 1990 (1994 in the case of Austria).[11] The first thing to note is that Spain's unemployment rate is substantially higher than Austria's, and its volatility is higher as well. While for Austria unemployment has 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. ,

A key labor market institution for unemployment duration is unemployment insurance, which is different in both countries. Table 5 compares minimum 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 benefits.

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

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

Figure 5. Unemployment and Long-Term Unemployment, Austria and Spain



Notes: Data from OECD (2017)

Table 5. 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.
Source: Data from the European Commission (`http://ec.europa.eu/social/main.jsp?catId=1101&langId=en&intPageId=4410`).

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.

Other labor market institutions which can shape the composition of inflows into unemployment (such as firing costs and dismissal procedures) are likely to be different in both countries.

We will take this into account when analyzing the results.

5 Results

In this section, we proceed with the estimation of the ABS method using Spanish social security data. As explained before, the data need to be expanded in order to implement the estimation. We first take the ABS approach and we focus on non-employment. We then compare our estimates for Spain with those of Austria from Alvarez, Borovičková, and Shimer (2014) (see subsection 5.1).

We then take our proposed approach by implementing the data expansions discussed in Section 3 for Spain. As we will discuss, some of these data additions have a meaningful economic interpretation, so we will be able to discuss, for example, the impact of temporary contracts on the variance decomposition (see subsection 5.2). Finally, we will compare the estimates for Spain along the business cycle (see subsection 5.3).

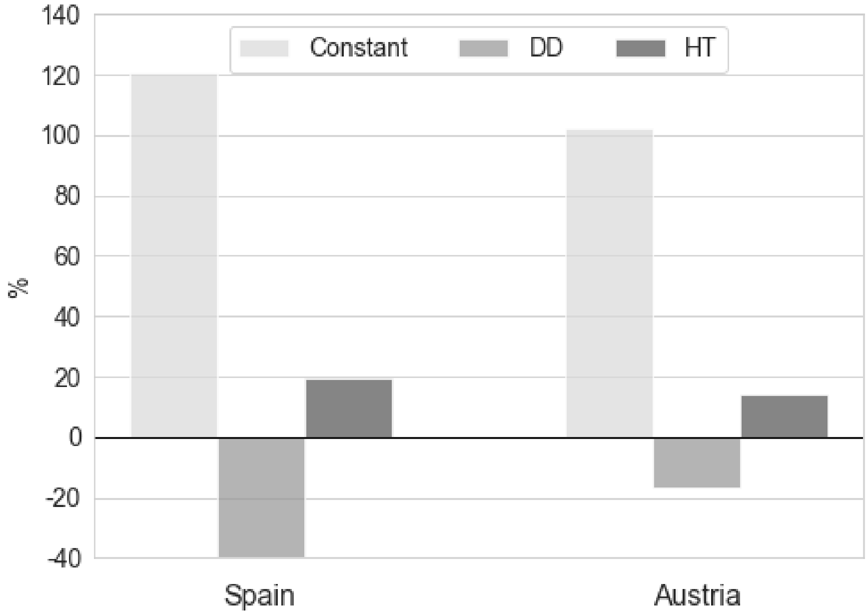
As explained in Section 2, we will distinguish between three different components of the variance in unemployment durations across individuals. These three components are heterogeneity (variation across individuals), duration dependence (variation within individuals due to different elapsed durations), and the aggregate component (constant hazard within variance), consistent with homogeneous workers and no duration effects.

5.1 Comparison between Austria and Spain: non-employment

The goal of this section is to compare Austria and Spain with regard to non-employment. We first report in Figure 6 the variance decomposition of registered unemployment duration (in logs) with the raw data for both countries.

Several aspects from Figure 6 should be highlighted. First, the broad decompositions in Spain and Austria are surprisingly similar. For both countries, the most important and relevant component is the constant or aggregate component, that is, the constant hazard within variance. And its contribution to the overall variance in both countries is very similar. Second, the contribution of the heterogeneity component is also very similar in both countries. This is despite any potential differences in the composition of inflows into unemployment due to different labor market institutions in the two countries. Third, for both countries, the duration

FIGURE 6. Decomposition with raw data

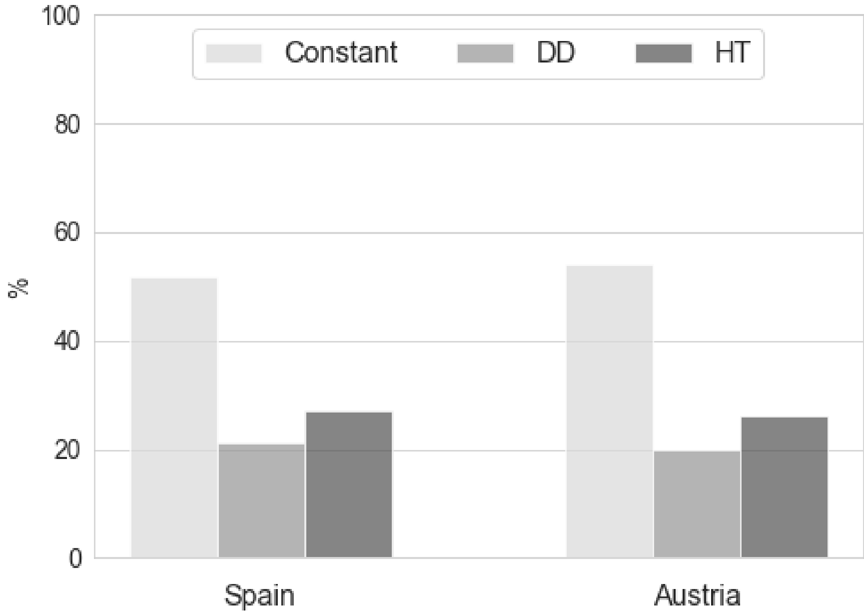


Notes: Registered unemployment duration in logs. Results for Austria taken from Alvarez, Borovičková, and Shimer (2014). Constant stands for constant hazard within variance (aggregate component), DD stands for duration dependence, and HT stands for heterogeneity.

dependence component is actually a negative number. This is an artefact of the raw data. In Austria it is smaller than in Spain because of the lower incidence of long-term unemployment and thus fewer blanks in the raw administrative data of long spells without insurance. In both countries, the raw data display unemployment spells that are shorter than they should be, as they only capture unemployment while receiving benefits. This underestimation of unemployment duration appears as if exit sharply increases around benefit expiration, as shown in Figure 1 (see also Figure 1 in Alvarez, Borovičková, and Shimer (2014) for Austria). This effect is strong enough to generate a positive effect on exit and is present in both Spain and Austria but more pronounced in Spain because of the different functioning of the two labor markets, as explained in Section 4. The longer unemployment benefits and longer unemployment durations can reconcile the differences between the two countries.

We next take the approach by Alvarez, Borovičková, and Shimer (2014) and focus on non-employment. Figure 7 displays the results for Austria and Spain (Table 6 in Appendix A displays the same figures). As Figure 3 shows (and similarly, Figure 2 in Alvarez, Borovičková, and Shimer (2014) for Austria), focusing on non-employment rather than the raw data captures longer spells of non-employment (which were artificially too short in the raw data) as well as the presence of short spells (i.e., shorter than a few weeks) that did not show up in the raw data. The latter makes the spell duration distribution decline from the beginning. As one would expect, this implies that the duration dependence component is now a positive number, as can be seen in Figure 7. The shares of the within components get redistributed, while as expected, the heterogeneity component remains largely unaffected.

FIGURE 7. Non-employment Decomposition



Notes: Non-employment duration in logs. Results for Austria taken from Alvarez, Borovičková, and Shimer (2014). Constant stands for constant hazard within variance (aggregate component), DD stands for duration dependence, and HT stands for heterogeneity.

The main economic result from Figure 7 is that the decompositions in Austria and Spain are remarkably similar. Second, in both countries, the constant component is the most important one, accounting for roughly half of the total variance, followed by the heterogeneity component, accounting for roughly a bit more than a quarter of the total variance, and finally the duration dependence component, accounting for roughly a bit less than a quarter of the total variance. It is worth noting that the similarities between the two countries are in *relative* terms. But it is not necessarily the case that the absolute variance is the same. In fact, Table 6 in Appendix A shows that it is slightly higher for Spain than for Austria.

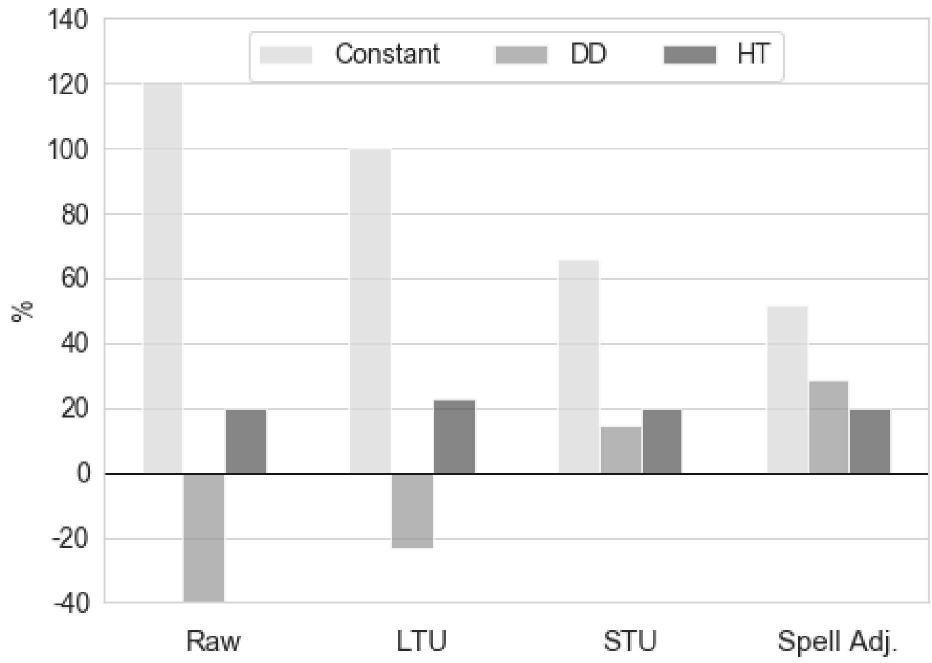
One could think that the similarity in the decompositions in Austria in Spain could be due to the implied selection that the the two-spell-requirement of the ABS method. In Appendix B we show that this is not likely to be case because the distributions of unemployment duration with and without the two-spell-requirement are very similar both in Austria and Spain.

5.2 Spain: Unemployment

In this subsection, we focus on unemployment in Spain using the subsample required by the ABS method, that is, unemployed workers with two completed spells of unemployment. One issue is how representative this subsample is. It is worth noticing that this subsample corresponds to 77% of unemployed workers (more precisely, workers with at least one day of registered unemployment).

We take our proposed approach by implementing the data expansions discussed in Section 3 for Spain, namely, the LTU expansion, the STU expansion, and the spell number adjustment. We will implement one expansion at a time, as some of these data expansions have a meaningful economic interpretation, so by comparing them we will be able to discuss, for example, the impact of temporary contracts on the variance decomposition. Figure 8 displays all the different decompositions with the different data expansions (Table 7 in the Appendix displays the same figures).

FIGURE 8. Decomposition with raw data and data expansions



Notes: Unemployment duration in logs. Constant stands for constant hazard within variance (aggregate component), DD stands for duration dependence, and HT stands for heterogeneity.

LTU expansion

We start by implementing in the Spanish data the LTU expansion explained in Section 3.2. Recall that this expansion refers to the fact that unemployed workers who remain unemployed after exhausting their benefits appear as blank for the remainder of the spell in the raw data. The first two graph sections on left-hand side of Figure 8 display the variance decomposition for the raw data and the data after the LTU expansion. As explained above, the raw data decomposition displays an anomaly of the duration dependence component, which is reduced when incorporating the LTU expansion. However, unlike the non-employment approach, the anomaly is still present. This is because the remaining short-term unemployed workers without benefits, the LTU expansion is unable to capture whereas the non-employment approach does capture them.

Adding the STU expansion

We next add the STU expansion as explained in Section 3.3. Recall that this expansion refers to workers who do not have the right to claim unemployment benefits and typically display short unemployment spells. As the bar sections labelled "STU" in Figure 8 show, the incorporation of this data expansion corrects the anomaly of duration dependence, which is now a positive number. Again, this is because the raw data display a hump shape at very short durations. That is, exit from unemployment is increasing with duration for very short durations, but this pattern disappears when the data incorporate the LTU and STU expansions (see Figure 3). The order of importance of each component is the same as non-employment, the ranking being constant > heterogeneity > duration dependence. However, the actual weights for each component are a bit different with respect to the non-employment decomposition, as shown in Table 7. In particular, the weight on duration dependence is smaller (14% versus 22%). This decrease in the importance of the duration dependence can be explained by the fact that the non-employment decomposition includes recalls, whereas the unemployment decomposition does not. This finding is consistent with Fujita and Moscarini (2017), who argue that duration dependence emerges mostly for those who are eventually recalled to the same firm.

Adding the spell number adjustment

We will next add the spell number adjustment as explained in Section 3.4. This is important in the Spanish context because, as explained before, the absence of the spell number adjustment can be interpreted as if temporary contracts were less prominent in the Spanish labor market (i.e., ignoring short spells of unemployment—in between employment spells—without unemployment benefits).

Results are displayed in the bars labelled "Spell Adj." in Figure 8. Comparing these bars with the bars labelled "STU" we find that the heterogeneity effect remains roughly unchanged, while the constant component gets reduced and duration dependence increases. This finding is in line with previous work that has highlighted that a consequence of temporary contracts is the duality among unemployed workers by which the share of short-term unemployment has increased but those with long durations have more difficulties in leaving unemployment (see, for instance, Güell and Hu 2006). Note that in terms of short unemployment durations, temporary contracts have a similar effect to recalls to the same employer. Therefore, consistent with Fujita and Moscarini (2017), the importance of duration dependence increases when the excess churning of temporary contracts is fully captured in the data.

Note that the increase in the importance of the duration dependence component when adding the spell number adjustment has somehow changed the ranking between the different components. Without such an adjustment, the ranking was constant > heterogeneity > duration dependence as in Austria, whereas with the adjustment, it is constant > duration dependence > heterogeneity. This finding is interesting given that the incidence of temporary contracts in Austria is much lower than in Spain.

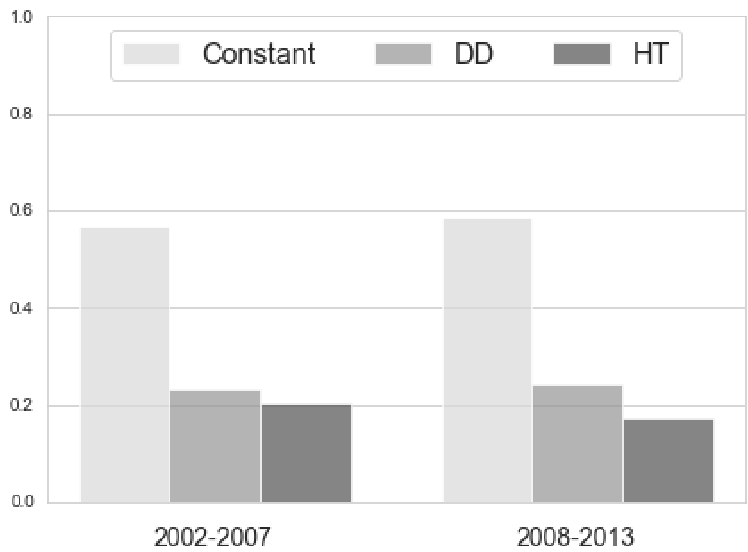
5.3 Spain along the business cycle

An important aspect of the dysfunctional Spanish labor market is the large variation in unemployment rates as well as unemployment durations along the business cycles, as compared to other countries. Moreover, the availability of the complete working histories of a large sample of workers in Spanish social security administrative data allows us to undertake such an exercise. The question is whether the decomposition of the unemployment duration is different along the business cycle.

We split our sample into an expansion period and a recession period, namely, 2002-2007 and 2008-2013, respectively.[12] The former period was characterized by years of lower unemployment and a lower share of long-term unemployment (around 10% and 30%, respectively), while the latter period was characterized by years of higher unemployment and higher long-term unemployment (around 25% and 40%, respectively). Within each period, we analyse those unemployed workers that have two completed spells (see Appendix C for details).

We first show the decomposition for the different periods taking the non-employment approach and then repeat the exercise taking the unemployment approach explained in the previous section. Figure 9 displays the decomposition for the periods 2002-2007 and 2008-2013 for non-employment (Table 8 in the Appendix displays the same figures).

FIGURE 9. Spain along the business cycle: non-employment



Notes: Unemployment duration in logs, Spain, fully expanded data. Constant stands for constant hazard within variance (aggregate component), DD stands for duration dependence, and HT stands for heterogeneity.

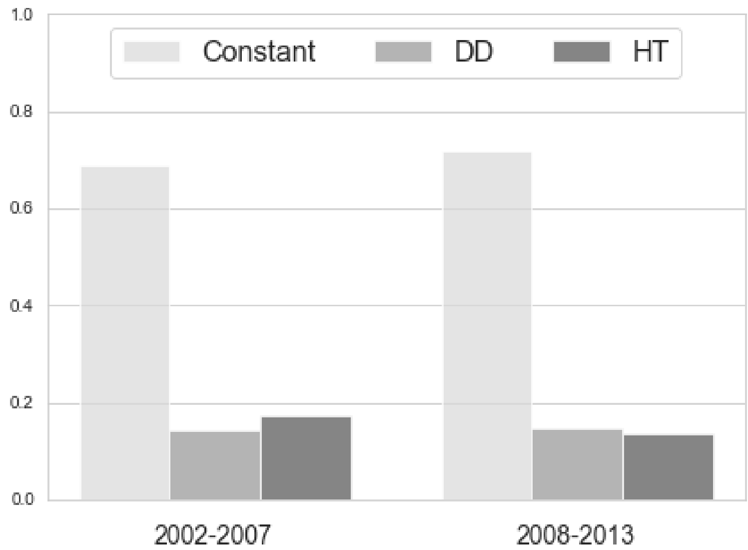
As can be seen in Figure 9, the importance of the three different components is surprisingly similar in the two very different periods considered. Regarding duration dependence, this result echoes previous findings from traditional duration models. For instance, Bover et al. (2002) find that the shape 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

[12]The MCVL panel starts from the year 2004, so it is not recommended to use the data retrospectively. For this application however the extend the sample period 2 years earlier than in the previous analysis. The motivation here is to have a balanced sample with 5 years of recession and 5 years of economic expansion.

not different along the business cycle in the UK.

Figure 9 also shows that the constant component is clearly the most important, followed in similar weights by the duration dependence and heterogeneity components. Note again that the results in Figure 9 are in relative terms, and the absolute figures might not be as similar along the business cycle. In fact, in Table 8 in the Appendix, we note that the heterogeneity component is higher during the expansion period than in the recession period. This could be because in recessions we have more observations with two completed spells, which translates into more cases around the mean.

FIGURE 10. Spain along the business cycle: unemployment



Notes: Unemployment duration in logs, Spain, fully expanded data. Constant stands for constant hazard within variance (aggregate component), DD stands for duration dependence, and HT stands for heterogeneity.

We now repeat the exercise by expanding the data to capture unemployment as explained in subsection 5.2. Figure 10 displays the decomposition for the periods 2002-2007 and 2008-2013 for unemployment (Table 8 in the Appendix displays the same figures). As can be seen, we find the same overall results as before: the decomposition is remarkably similar along the business cycle. Note that the importance of duration dependence is very close in the two periods, while the macroeconomic component is somehow more important in the recession years and the heterogeneity component is somehow less important. This finding is a little different from when we consider non-employment (see Figure 9). We believe this difference is because recalls are included in non-employment (as explained in subsection 3.3). Recalls are less subject to the business cycle (see Fujita and Moscarini 2017), and this could explain these minor differences.

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 are somehow different in the two periods, while some other characteristics are remarkably similar in the two periods (see Table 10 in Appendix C). 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)). However, for our analysis, there could be other factors that could go in opposite direction, such as search effort being countercycical (see Mukoyama, Patterson, and Şahin (2018)).

6 Theories behind duration dependence

In this section, we aim to shed some light on the different possible theories that can explain duration dependence by using our comparison exercises between high-low unemployment countries as well as between high-low unemployment periods within Spain from the previous sections. Different mechanisms could explain the presence of negative duration dependence in the exit rate of unemployment (Machin and Manning 1999): these mechanisms could be 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).[13] Statistical discrimination could be due to dynamic selection of the long-term unemployed as well as the decay in human capital during the unemployment spell.

In general, it is hard to distinguish among these explanations. A reasonable assumption is that the process of skill loss is independent of the state of the economy (e.g., the level of unemployment). Thus, our comparison exercises are not very useful in attempting to discard such an explanation. However, this is not the case for the dynamic selection explanation. Recall that dynamic selection refers to the fact that, ceteris paribus, less employable workers remain unemployed for longer periods. In bad times, there is less hiring overall, and therefore more workers will accumulate over longer durations. Thus, studying the long-term unemployed in bad

[13] 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)).

times does not yield much information. Conversely, studying the long-term unemployed in good times, when fewer workers accumulate over longer durations because of higher hiring rates, is more informative about the quality of the worker. In other words, the dynamic sample selection will be less (more) acute when unemployment is high (low) (see, for instance, Blanchard and Diamond 1994 and Lockwood 1991).

Our comparison exercises between Austria and Spain, as well as Spain alone between various periods along the business cycle (at high and low unemployment points), can shed some light on the possible role of statistical discrimination due to dynamic sample selection. In both comparison exercises, we find that the importance of duration dependence is similar and thus does not change with the level of unemployment. This finding suggests that statistical discrimination due to dynamic selection does not seem to be the main driver of duration dependence.

A relevant paper in this context is Jarosch and Pilossoph (2019), who propose a model in which employers endogenously discriminate against the long-term unemployed. This discrimination matters to the extent that the lost interview translates into a lost job (i.e., if discrimination is due to loss of skills, discrimination most likely would not matter). They conclude that employer discrimination is largely due to dynamic selection, with limited consequences for structural duration dependence. Our results also suggest that duration dependence has less to do with signalling and more to alternative explanations such as loss of skills.

7 Conclusions

In this paper, we have applied the method recently developed by Alvarez, Borovičková, and Shimer (2014) to Spain using administrative social security data. We compare the decomposition of the duration of unemployment in Austria and Spain, two very different labor markets using comparable data. We also compare such a decomposition along the business cycle for Spain. These comparisons allow us to shed some light on which theories can explain duration dependence.

Administrative data have many advantages compared to Labor Force Survey data, but unemployment spells without benefit entitlement are underreported. One approach is to focus on non-employment as in Alvarez, Borovičková, and Shimer (2014). In this paper, we have taken another approach in trying to distinguish unemployment from other situations by using complementary information. Specifically, we deal with underreported spells in a systematic way by using information from the Spanish Labor Force Survey as well as institutional information following Lafuente (2020). This step is crucial because the results are sensitive to the treatment of underreporting. We hope that our approach will provide a useful way to apply the ABS method in other countries.

In a nutshell, we find that the aggregate component is clearly the most important one, followed by heterogeneity and duration dependence, which are roughly comparable. Moreover, the relative decomposition of unemployment duration is very similar for both countries (especially when minimizing the incidence of temporary contracts) as well as along the business cycle within Spain. These comparisons suggest that statistical discrimination due to dynamic sample selection does not seem to be the main driver of duration dependence.

The similarity in the variance decomposition across these two countries and the business cycle may not be an intuitive result at first glance. But upon reflection, we see that these results could be rationalized by a model of the labor market in which the overall levels of unemployment and unemployment duration do not affect the importance of the different components of unemployment duration. We leave this question for future research.

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.

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.

Card, D., R. Chetty, and A. Weber (2007). The spike at benefit exhaustion: Leaving the unemployment system or starting a new job? *American Economic Review 97*(2), 113–118.

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.

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.

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*, forthcoming 2020. Published online on 18 July 2019, doi.org/10.1007/s13209-019-0200-1.

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.

Mukoyama, T., C. Patterson, and A. Şahin (2018, January). Job search behavior over the business cycle. *American Economic Journal: Macroeconomics 10*(1), 190–215.

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

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.

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.

A Appendix: Detailed results

Table 6. Variance decomposition with raw data and non-employment

	Raw data		Non-Employment	
	Spain	Austria	Spain	Austria
Total	1.366	1.605	3.185	3.040
Constant				
variance	1.645	1.645	1.645	1.645
share	*1.204*	*1.025*	*0.516*	*0.541*
Duration Dependence				
variance	-0.547	-0.271	0.676	0.604
share	*-0.400*	*-0.169*	*0.212*	*0.199*
Heterogeneity				
variance	0.268	0.231	0.864	0.791
share	*0.196*	*0.144*	*0.271*	*0.260*
N	574,500	3,046,298	1,125,832	4,583,140

Notes: Log duration in days. Results for Austria taken from Alvarez, Borovičková, and Shimer (2014).

TABLE 7. Variance decomposition for Spain with data expansions

	LTU Expansion	STU Expansion	Spell Adjustment
Total variance	1.640	2.506	3.179
Constant			
variance	1.645	1.645	1.645
share	*1.003*	*0.657*	*0.517*
Duration Dependence			
variance	-0.383	0.357	0.910
share	*-0.234*	*0.142*	*0.286*
Heterogeneity			
variance	0.378	0.504	0.624
share	*0.231*	*0.201*	*0.196*
N	574,500	906,320	906,320

Notes: Log duration in days.

TABLE 8. Variance decomposition along the business cycle in Spain

	Non-Employment		Unemployment	
	2002-2007	2008-2013	2002-2007	2008-2013
Total	2.911	2.812	2.396	2.296
Constant				
variance	1.645	1.645	1.645	1.645
share	*0.565*	*0.585*	*0.686*	*0.717*
Duration Dependence				
variance	0.674	0.682	0.341	0.342
share	*0.231*	*0.243*	*0.142*	*0.149*
Heterogeneity				
variance	0.592	0.485	0.410	0.309
share	*0.203*	*0.172*	*0.171*	*0.134*
N	359,100	423,308	252,306	332,758

Notes: Log duration in days. Unemployment refers to fully expanded data.

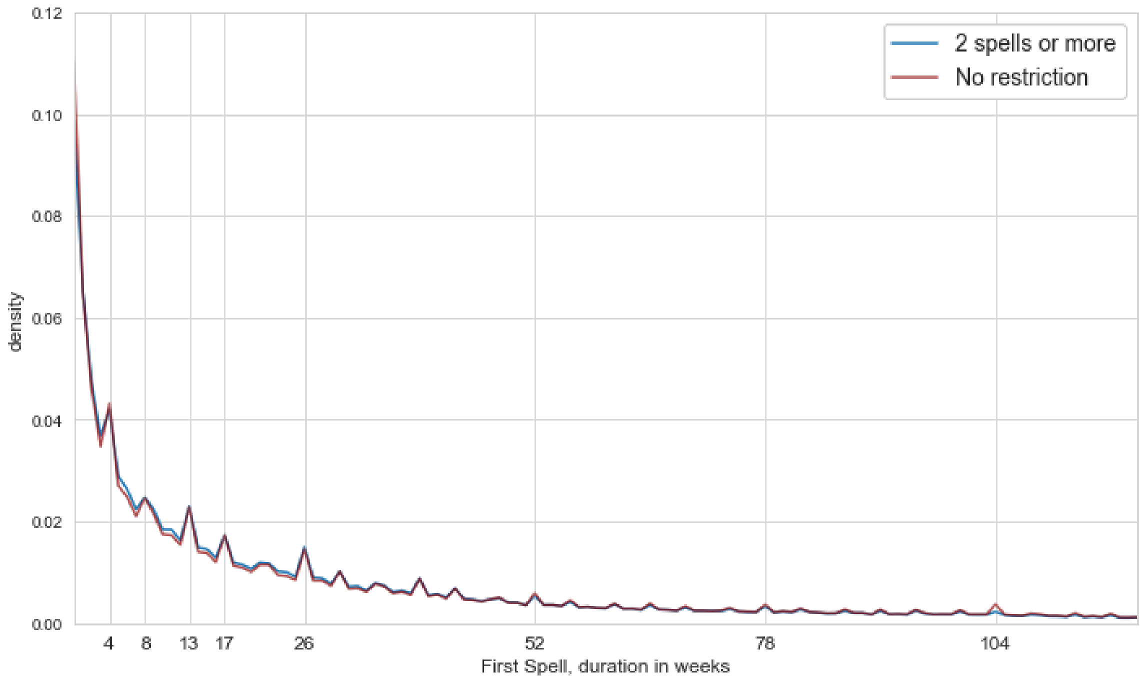
B Appendix: Two-spell-requirement selection

In this appendix we show that the two-spell-requirement of the ABS method is not likely to be driving the similarity in the decompositions between Austria and Spain.

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 11 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 finding suggests that sample selection bias is not likely to be driving our comparative results.

FIGURE 11. Two-Spell-requirement selection



Notes: Histogram of unemployment duration in weeks, Spain, fully expanded data. For individuals with 2 spells or more, the plot shows the duration of the first spell.

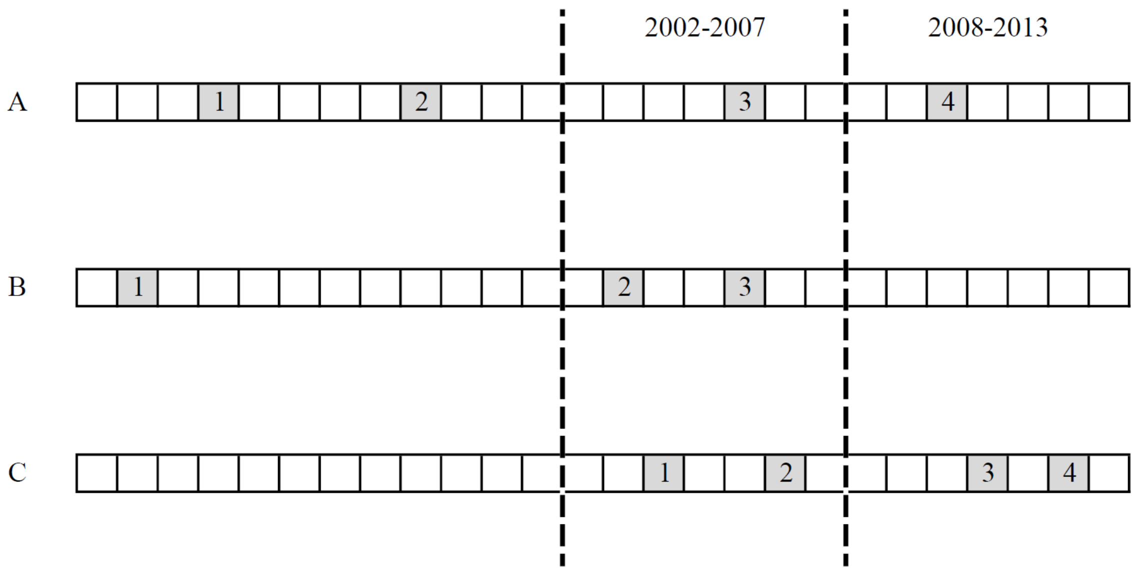
C Appendix: Business cycle samples

In this appendix we provide details of the samples for the business cycle decomposition used in Section 5.3. As mentioned, we split our sample between 2002-2007 (expansion) and 2008-13 (recession). Given the two-spell requirement of the ABS method, a worker would need to have two unemployment spells within each period in order be included in the sample.

Figure 12 illustrates different cases of unemployment spells for the different periods for 3 types of workers (A, B, C). Recall that our sample starts in 2004 and given that we have complete life histories, some spells can be prior to 2004 as in Alvarez, Borovičková, and Shimer (2014).

Given the two-spell requirement, workers with one unemployment spell in one period and another spell in the other period will not appear in our business cycle analysis. This is illustrated by case A. On the other hand, workers that have two spells in a given period (even if these are not their first and second spells) could be included, possibly in both the expansion and the recession periods. This would be cases B and C in the table. In the main analysis in the text, cases such as B and C have been included in the expansion and recession periods, respectively. In this Appendix, we also perform a more restrictive approach by including strictly the first and second unemployment spell of each worker. Table 9 displays the results in this restricted sample ("Strict spell order") and compares it to the sample in the main text ("All spells"). As can be seen, the shares of each of the components for the expansion and recession periods are robust to the different samples.

Figure 12. Sample Selection Example



Notes: Letters refer to worker types and numbers refer to unemployment spells in chronological order for each worker.

Table 9. Variance decomposition along the business cycle — 2 ways

	Strict spell order		All spells	
	2002-2007	2008-2013	2002-2007	2008-2013
Total	2.621	2.181	2.396	2.296
Constant				
variance	1.645	1.645	1.645	1.645
share	*0.628*	*0.754*	*0.686*	*0.717*
Duration Dependence				
variance	0.497	0.313	0.341	0.342
share	*0.189*	*0.144*	*0.142*	*0.149*
Heterogeneity				
variance	0.480	0.222	0.410	0.309
share	*0.183*	*0.102*	*0.171*	*0.134*
N	180,308	178,924	252,306	332,758

Notes: Log duration in days. Unemployment refers to fully expanded data.

We now turn to the composition of the unemployment pool in both periods. Table 10 below provides summary statistics of the main variables available in our data set.

Table 10. Descriptive statistics — Fully Expanded data

	2002-2007	2008-2013
Age	33.54	34.65
	(6.69)	(6.74)
Male	0.55	0.56
	(0.50)	(0.50
Work Experience (years)	6.13	8.02
	(5.53)	(6.14)
Quit	0.31	0.19
	(0.46)	(0.39)
College or higher education	0.139	0.142
	(0.346)	(0.349)
N	252,312	332,756

Notes: Standard deviation in parenthesis