DISCUSSION PAPER SERIES

Worker Flows and Wage Dynamics: Estimating Wage Growth without Composition Effects

Raquel Carrasco
J. Ignacio García-Pérez
Juan F. Jimeno

DISCUSSION PAPER SERIES

Worker Flows and Wage Dynamics: Estimating Wage Growth without Composition Effects

Raquel Carrasco
Universidad Carlos III

J. Ignacio García-Pérez
Universidad Pablo de Olavide and FEDEA

Juan F. Jimeno
Banco de España, Universidad de Alcalá, CEPR and IZA

ABSTRACT

Worker Flows and Wage Dynamics: Estimating Wage Growth without Composition Effects[*]

Wage dynamics is closely intertwined with job flows. However, composition effects associated to the different sizes and characteristics of workers entering/ exiting into/from employment that may blur the "true" underlying wage growth, are not typically accounted for. In this paper, we take these composition effects into consideration and compute wage growth in Spain during the 2006-2018 period after netting out the consequences of employment dynamics. Our results show that the "true" underlying wage growth in the Spanish economy during recessions (expansions) was, on average, significantly lower (higher) that the observed with raw data. This may help to explain some macro puzzles, such as the "vanishing" Phillips curve.

Corresponding author:
Juan F. Jimeno
Banco de España
Calle Alcalá 48
28014 Madrid
España
E-mail: juan.jimeno@bde.es

1 Introduction

There is a long literature in Labor Economics on the cyclical properties of wages. A conventional result is that, adequately measured wage growth is (slightly) pro-cyclical (Abraham and Haltiwanger, 1995. Daly and Hobijn, 2017). During recessions the prevalence of downwards wage rigidities puts a floor on wage growth (Izquierdo et al., 2017) while in recoveries labor demand increases and unemployment decreases, which push wages upwards.

However, along the last business cycle (comprising the financial crisis of the last decade and the recovery afterwards), these cyclical patterns have shown some peculiarities. First, given the intensity and nature of the crisis, wage growth remained relatively higher than expected by standard wage equations; whereas in the recovery wages seems to have grown significantly less than expected from its historical relationship with expected inflation and unemployment. Although missing wage deflation during the crisis can be explained by the above mentioned rigidities, missing wage inflation in the recovery is somehow a puzzle that remains to be fully explained.

Several explanations seem most relevant at explaining this puzzle. One is that missing wage deflation during the crisis is the reason for missing wage inflation in the recovery: wages failed to adjust fully to the intensity of the crisis and, hence, it would take a longer recovery for wages to start growing again at "normal paces". Another, partly related explanation, is that worker and job flows were significantly different than in the past. Moreover, some structural trends associated to technological changes and population ageing made employment dynamics less prone to wage increases.

It has been argued that job-to-job transitions is one main determinant of wage growth (Fatih et al. 2019). Job-to-Job flows are specially frequent in early stages of a worker career, and, moreover, they are typically associated to wage gains significantly higher also at earlier stages of the working life. Insofar as new cohorts entering into the labor market are relatively smaller,[1] it should be expected that there ought to be a significant decrease of the weight of workers with Job-to-Job flows associated with high wage increases.

While the implications of composition effects for wage growth are well-known and widely researched, not all of them are taken into consideration jointly when measuring wage growth. Using matched employer-employee data sets to accounting for the implications of worker and firm heterogeneity is nowadays standard practice (see, for instance, Carneiro, Guimarães and Portugal, 2012). And as already mentioned, appealing at job flows as a main source of wage fluctuations (e.g., Fatih et al., 2017) is nowadays a popular hypothesis. What is less frequent is to jointly accounting for worker and firm heterogeneity *and* employment dynamics (differences between workers staying in the same jobs, those entering and exiting from employment, and those with transitions to other jobs, with and without unemployment spells during the transition).

In this paper, we compute wage growth in Spain during the 2006-2018 period after netting out composition effects due to both employment dynamics and worker and firm heterogeneity. For that purpose, we use data from the Continuous Sample of Working Lives (*Muestra Continua de Vidas Laborales*, MCVL), which is a micro-

---

[1]According to the European Labor Force Survey, the size of the cohort between 15-24 years old has decreased with respect to the workforce from 9.8% to 4.8% between 2003 and 2019 in Spain. For the UE-15 these figures are 9.1% and 7.1% respectively.

economic data set based on administrative records provided by the Spanish Social Security Administration. First, we decompose the overall observed quarterly wage changes in three factors: (i) Composition effects due to the changes in the relative weights of workers with different job transitions (stayers, movers, entrants, and exits from the labor force), (ii) Composition effects due to changes in the worker and firm characteristics of each group, and (iii) "Pure" wage growth net of composition effects. We use non-parametric and parametric methods to estimate the returns to individual characteristics for each group of workers. Specifically, we estimate wage equations by period and group by ordinary least squares (OLS) and by using a correction model which accounts for unobserved heterogeneity driving potential endogenous selection into each group.[2]

Our results show that the "true" underlying wage growth in the Spanish economy during this period was more pro-cyclical than observed with raw data. The "pure" component of wage growth was, on average, about 1 pp lower than registered wage growth in the crisis. On the contrary, during the recovery, "pure" wage growth was about 1 pp higher than observed from aggregate data. Observed aggregate wage growth fails to account for underlying wage pressures mainly because it ignores changes in the gap between wages of stayers, movers, new entrants and exits, and in the relative employment weights of these groups of workers, which both have marked cyclical patterns. Taking them into account in the computation of wage growth helps to explain, for instance, the puzzle of the missing wage deflation/inflation during the last business cycle.

The structure of the paper is standard. First, we provide some background for the data and their descriptive statistics (An Appendix contains main variables by quartiles of the wage distribution). Then we discuss how to account for employment dynamics and composition effects from individual and job characteristics in the computation of wage growth, and present and discuss the main results. Finally, we provide estimations of a simple Phillips curve that show that the alternative measure of wage growth displays a closer association with labor market slack. The last section contains final remarks.

2 Data

2.1 Sample

The Continuous Sample of Working Lives (*Muestra Continua de Vidas Laborales*, MCVL) is a microeconomic data set based on administrative records provided by the Spanish Social Security Administration. Each wave contains a random sample of 4% of all the individuals registered with the Social Security system (either by being working or on an unemployment contributory scheme or by having received other contributory benefits, such as permanent disability, old-age pensions, etc.) during at least one day in the year the sample is selected. There is some sample selection, especially

---

[2]A related issue regarding measured wage growth has been raised by Morris, Rich, and Tracy (2020). They show that *gowth in an average wage* (which weights each wage observation by work hours) is less clclically sensitive than *average wage growth* (which equally weights each wage observation).

among women, immigrants or young workers, since those individuals without any contact to Social Security in such a year are not included. Hence, in order to minimize the potential selection effects, we combine all the waves, from 2006 to 2018. That is, our sample includes everybody that had a relationship of at least one day during this period with the Social Security administration.

For each individual we observe her entire employment history, including the exact duration of employment, unemployment, disability, and retirement pension spells.[3] For each spell we observe several variables that describe the characteristics of either the job or the unemployment/pension benefits. There is also information on personal characteristics such as age, gender, nationality, and education attainment levels.

We exclude employees who are not enrolled in the general regime of the Social Security Administration. Furthermore, in order to minimize the effects of outliers in wages, we dropped 2 percent of the observations corresponding to the top and bottom tails of the wage distribution in each year. Finally we also exclude temporary layoffs from the estimation sample (that is, those working as *fijos discontinuos*), who have a permanent contract but just work some months per year in the firm, while during the rest of the year they earn unemployment benefits. Given this sample selection restrictions, we end up with a sample of 21,605,468 quarterly observations for our dependent variable which is defined in terms of real daily wages. This measure of earnings is computed as the ratio between the Social Security monthly contribution base and the days worked in one particular month each quarter (January, April, July, October), adjusted also by part-time employment. Earnings are deflated using the 2016 general price index.

2.2 Characterizing employment dynamics

We classify working-age population into six groups: stayers, voluntary movers, involuntary movers, entrants, and exits from the labor force (to retirement and to long-term unemployment). Specifically, we consider:

1. $stayers_{t-1,t}$: Workers who remain employed at the same job at $t-1$ and t.

2. $vol\_movers_{t-1,t}$: Workers who remain employed at $t-1$ and t but having changed jobs without a significant unemployment spell (less than three months unemployed).

3. $invol\_movers_{t-1,t}$: Workers who remain employed at $t-1$ and t but having changed jobs after a significant unemployment spell (more than three months unemployed).

4. $entrants_{t-1,t}$: Workers who are employed at t but not at $t-1$.

---

[3]As explained in, for example, Garcia-Perez and Rebollo (2009), there exists a potential issue of measurement error in the number of employment and unemployment spells and their duration given that firms can offer contracts for very short periods but subsequently recall workers. When this occurs, we treat the employment spell as a continuous period, despite the short interruption that appears in between. More specifically, we unify successive registers when they correspond to the same worker in the same firm with the same type of contract, and when the interruption lasts for less than 15 days.

5. $exits\_ret_{t-1,t}$: Workers who are employed at $t-1$ and retired at t. These are defined as those workers receiving a wage at $t-1$ and a retirement pension at t.

6. $exits\_u_{t-1,t}$: Workers who are employed at $t-1$ and unemployed at t. These are defined as those workers receiving a wage at $t-1$ and without a employment contract at t (with or without unemployment benefits at t).

Thus, total employment at (quarter) t, N_t, can be computed as:

$$N_t = stayers_{t-1,t} + vol\_movers_{t-1,t} + invol\_movers_{t-1,t} + entrants_{t-1,t}, \quad (1)$$

and also as:

$$N_t = stayers_{t,t+1} + vol\_movers_{t,t+1} + invol\_movers_{t,t+1} + exits\_ret_{t,t+1} + exits\_u_{t,t+1}. \tag{2}$$

Tables 1 and 2 show the proportion of workers in each group per year and some individual characteristics to be used in the computation of composition effects, respectively. Table 1 shows that, according to distribution 1, more than 91% of the working population are stayers in the whole period , while the proportion of movers and entrants is around 4.4%. As to workers' characteristics, Table 2 shows that 10% of the stayers are older than 55, while for entrants this figure is only 4.1%. Also, there is a higher percentage of college graduates among the stayers than among the entrants (21.7% versus 17.4%), although is among the voluntary movers where we can find the highest proportion of college graduates. We also find significant differences among groups in terms of sector of activity and immigrant status.

Table 1: Proportion of workers by type and year

Year	Worker type			
	Stayers	Vol. Movers	Invol. Movers	Entrants
2005	0.9070	0.0321	0.0234	0.0375
2006	0.8968	0.0322	0.0278	0.0433
2007	0.8932	0.0332	0.0295	0.0440
2008	0.9092	0.0294	0.0222	0.0391
2009	0.9262	0.0219	0.0127	0.0392
2010	0.9243	0.0207	0.0124	0.0427
2011	0.9217	0.0224	0.0130	0.0428
2012	0.9307	0.0194	0.0110	0.0389
2013	0.9248	0.0227	0.0100	0.0425
2014	0.9175	0.0236	0.0113	0.0475
2015	0.9103	0.0265	0.0139	0.0493
2016	0.9066	0.0272	0.0163	0.0500
2017	0.9010	0.0291	0.0187	0.0512
2018	0.9045	0.0269	0.0189	0.0498
% Obs.	91.23	1.74	2.63	4.40

Table 2: Sample means of main variables by worker type

Variable	Worker type				Eq. test (p-value)
	Stayers	Volunt. Movers	Involunt. Movers	Entrants	
Age					
16-24	0.0615	0.1525	0.1479	0.1647	0.000
25-54	0.8380	0.8112	0.8038	0.7938	0.000
55-65	0.1004	0.0364	0.0483	0.0415	0.000
Education					
Primary	0.1306	0.1336	0.1593	0.1685	0.000
Secondary	0.6524	0.6288	0.6460	0.6574	0.000
College	0.2170	0.2377	0.1947	0.1741	0.000
Sector					
Industry	0.1849	0.1200	0.1187	0.1000	0.000
Construction	0.0825	0.1276	0.1570	0.1332	0.000
Serv. High added value	0.3393	0.3053	0.2982	0.2921	0.000
Serv. Low added value	0.3933	0.4471	0.4260	0.4747	0.000
Gender					
Men	0.5582	0.5887	0.5775	0.5802	0.000
Immigrant Status					
Immigrant	0.0568	0.1253	0.0989	0.1191	0.000
N^o of Obs.	19,710,171	375,480	568,324	951,493	

Eq. test: χ^2 test for mean equality across worker types.

Figure 1 shows employment shares, (log) daily wages, and annual wage growth for each group (left column is for employment defined as in equation 1 whereas the right column is for employment defined as in equation 2). The relative employment shares of stayers, movers and entrants show a clear cyclical pattern, with movers and entrants increasing in the recovery phase (2014-2018), while the relative weight of stayers increased during the recession (2008-2013). Exits to unemployment are also counter-cyclical, while exits to retirement show no cyclical pattern and a slightly increasing trend, which, however, is less acute than expected due to population ageing because average retirement age in Spain during this period increases.

As for wage levels, the differences across groups are also non-surprising. Stayers, voluntary movers and exits to retirement have significantly higher wages than involuntary movers, entrants and exits to unemployment. As a result, although the cyclical patterns of wage growth for each group are similar, there are noticeable differences in wage growth rates across groups, both in the recession and in the recovery phases, with stayers and exits to retirement displaying less fluctuations than the wage growth of movers and entrants. In the Appendix we also provide weights and wages of the different worker types by quartiles of the wage distribution. The noticeable differences

of the weights and wages along the wage distribution suggest that any analysis of the wage evolution should take into account employment dynamics.

Figure 1
Employment shares, wage levels, and wage growth by employment status

a) Employment shares

b) Wage levels (logs)

c) Wage growth (annual rates)
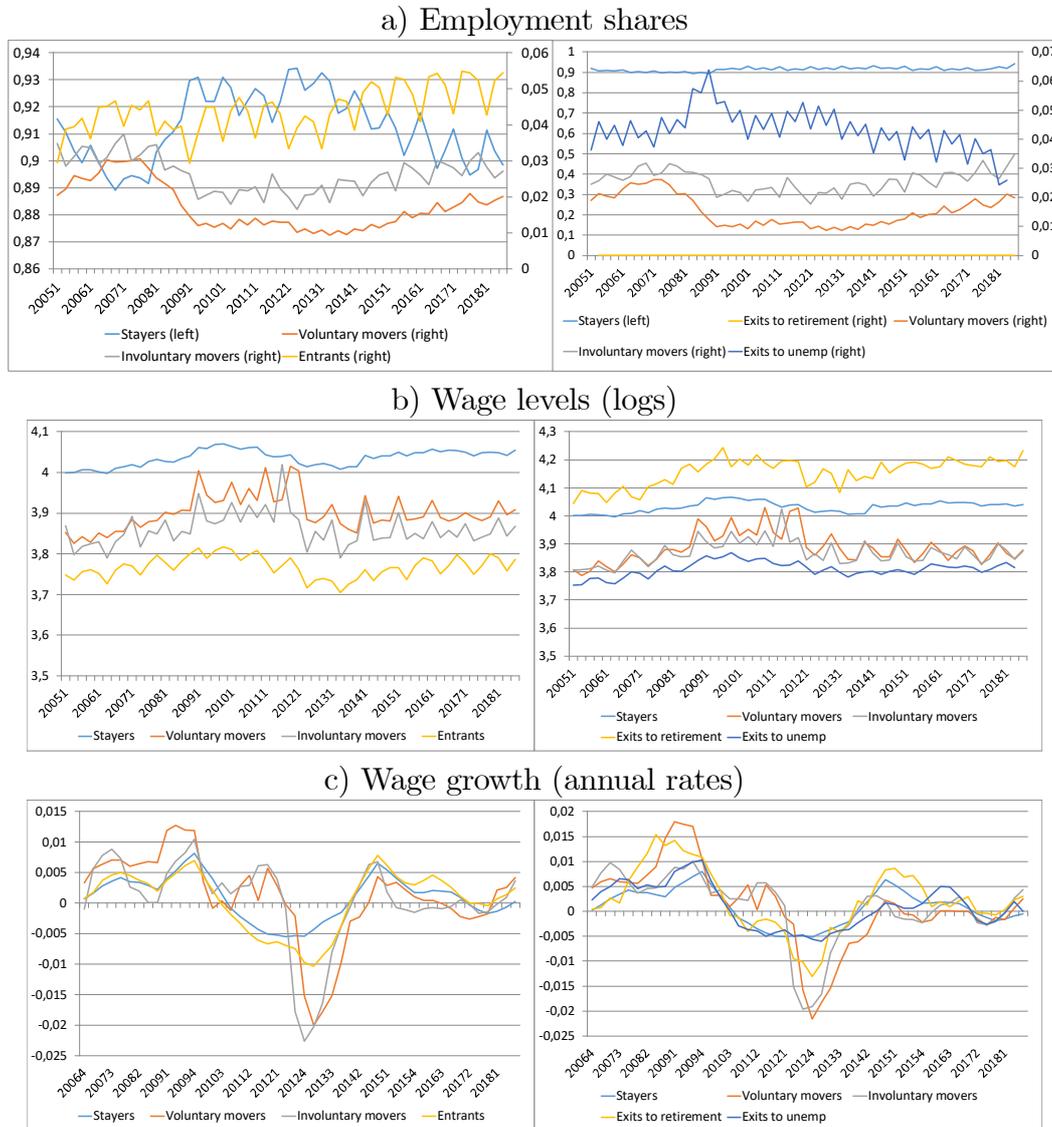
3   Accounting for wage growth

In this Section we perform the decomposition of the average wages to obtain a quarterly wage change net of composition and selection effects. Aggregating over the employment categories defined above,

$$I_{1t} = \{stayers_{t-1,t}, vol\_movers_{t-1,t}, invol\_movers_{t-1,t}, entrants_{t-1,t}\}$$

and

$$I_{2t} = \{stayers_{t,t+1}, vol\_movers_{t,t+1}, invol\_movers_{t,t+1}, exits\_ret_{t,t+1}, exits\_u_{t,t+1}\},$$

the average wage in period t can be computed as follows:

$$\overline{w}_t = \sum_{g \in I_{1t}} \alpha_t^g \overline{w}_t^g = \sum_{g \in I_{2t}} \alpha_t^g \overline{w}_t^g, \tag{3}$$

where $\overline{w}_t^g$ is the average wage for group g in period t and α_t^g is the proportion of workers in the gth group over the total number of workers in period t.

We compute the change from period $t-1$ to period t as follows:

$$\Delta \overline{w}_t = \sum_{g \in I_{1t}} \alpha_t^g \overline{w}_t^g - \sum_{g \in I_{2t-1}} \alpha_{t-1}^g \overline{w}_{t-1}^g. \tag{4}$$

Thus, wage growth is the sum of two components:

i) the change in wages of stayers, voluntary movers and involuntary movers between $t-1$ and t (notice that they are not the same individuals in the two periods, as the weights and wages at $t-1$ are taken from the distribution in I_{2t-1} while the weights and wages at t are taken from the distribution in I_{1t}), and

ii) the contribution of turnover due to the different weights and wages of entrants and exits at t.

Moreover, both components can be breakdown in two terms, one is due to changes in the composition of the labor force in terms of the employment status (composition effect between groups); another is the part of the total change attributed to variations in the wages of the groups, net of compositional group effects. Thus, let α and w denote the weight and average wage of stayers, movers and entrants computed for partition I_{1t} and $\widetilde{\alpha}$ and $\widetilde{w}$ be, respectively, the weights and average wages computed over the partition I_{2t-1}. Thus, wage growth between $t-1$ and t is given by:

$$\Delta \overline{w}_t = \Delta \overline{w}_t^c + \Delta \overline{w}_t^w \tag{5}$$
$$\Delta \overline{w}_t^c = \sum_{g \in S} (\alpha_t^g - \widetilde{\alpha}_{t-1}^g) \widetilde{w}_{t-1}^g + (\alpha_t^e - \widetilde{\alpha}_{t-1}^l) \widetilde{w}_{t-1}^l$$
$$\Delta \overline{w}_t^w = \sum_{g \in S} (w_t^g - \widetilde{w}_{t-1}^g) \alpha_t^g + (w_t^e - \widetilde{w}_{t-1}^l) \alpha_t^e,$$

where $S = \{stayers, vol\_movers, invol\_movers\}$, e stands for entrants, and l stands for exits (leavers).

Additionally, within groups, wage differences, $(w_t^g - \widetilde{w}_{t-1}^g)$ for $g = $ stayers, vol_movers, invol_movers, and $(w_t^e - \widetilde{w}_{t-1}^l)$ are the result of two components: i) differences in worker

and job characteristics, and ii) differences in returns to these characteristics. Thus, let's consider the following wage equations:

$$w_{it}^g = \sum_j \beta_{jt}^g x_{ijt}^g + \epsilon_{it}^g, \ g \in I_{1t}, \tag{6}$$

$$\widetilde{w}_{it}^g = \sum_j \widetilde{\beta}_{it}^g \widetilde{x}_{it} + \widetilde{\epsilon}_{it}^g, \ g \in I_{2t-1}, \tag{7}$$

where i stands for each individual observation and j is a set of worker and jobs characteristics. As standard in the empirical earnings literature, we assume that the data generating process of wages is given by a Mincerian equation (Mincer, 1974). The dependent variable is the logarithm of daily wages of individual i in group g and period t. The vector of explanatory variables account for worker and job characteristics. We include gender, immigrant status, education (primary, secondary and college), and age and age square, sector of activity, and region.

We obtain mean wages by estimating previous wage equations by period and group with two alternative methods: i) by ordinary least squares (OLS), and ii) by using a correction model which accounts for unobserved heterogeneity driving potential endogenous selection into each group, as explained below. As an alternative approach we also estimate non-parametrically the mean wage by period and group in each cell weighted by the proportion of workers in that cell over the total number of workers in that group and year.

Then, the difference in mean wages can be expressed as (we skip notation but from now on $\beta's$ are the estimated values)

$$w_t^g - \widetilde{w}_{t-1}^g = \sum_i \sum_j \widetilde{\beta}_{jt-1}^g (x_{ijt}^g - \widetilde{x}_{ijt-1}^g) + \sum_i \sum_j (\beta_{jt}^g - \widetilde{\beta}_{jt-1}^g) x_{ijt}^g \tag{8}$$

$$g \in \{stayers, vol\_movers, invol\_movers\}$$

$$w_t^e - \widetilde{w}_{t-1}^l = \sum_i \sum_j \widetilde{\beta}_{jt-1}^l (x_{ijt}^e - \widetilde{x}_{ijt-1}^l) + \sum_i \sum_j (\beta_{jt}^e - \widetilde{\beta}_{jt-1}^l) x_{ijt}^e.$$

We can express the wage component of the decomposition above as follows:

$$\Delta \overline{w}_t^w = \sum_{g \in S} \alpha_t^g \left[\sum_i \sum_j \widetilde{\beta}_{jt-1}^g (x_{ijt}^g - \widetilde{x}_{ijt-1}^g) + \sum_i \sum_j (\beta_{jt}^g - \widetilde{\beta}_{jt-1}^g) x_{ijt}^g \right] + \tag{9}$$

$$+ \alpha_t^e \left[\sum_i \sum_j \widetilde{\beta}_{jt-1}^l (x_{ijt}^e - \widetilde{x}_{ijt-1}^l) + \sum_i \sum_j (\beta_{jt}^e - \widetilde{\beta}_{jt-1}^l) x_{ijt}^e \right]$$

Hence, we can now further breakdown aggregate wage differences in three components:

$$\Delta \overline{w}_t = \Delta \overline{w}_t^c + \Delta \overline{w}_t^{cw} + \Delta \overline{w}_t^{pw}, \tag{10}$$

where

$$\Delta\overline{w}_t^{cw} = \sum_{g \in S} \alpha_t^g \sum_i \sum_j \widetilde{\beta}_{jt-1}^g (x_{ijt}^g - \widetilde{x}_{ijt-1}^g) + \alpha_t^e \sum_i \sum_j \widetilde{\beta}_{jt-1}^l (x_{ijt}^e - \widetilde{x}_{ijt-1}^l) \qquad (11)$$

$$\Delta\overline{w}_t^{pw} = \sum_{g \in S} \alpha_t^g \sum_i \sum_j (\beta_{jt}^g - \widetilde{\beta}_{jt-1}^g) x_{ijt}^g + \alpha_t^e \sum_i \sum_j (\beta_{jt}^e - \widetilde{\beta}_{jt-1}^l) x_{ijt}^e.$$

The first component in equation 10, $\Delta\overline{w}_t^c$, is the composition effect arising from the changing worker flows (stayers, movers, entrants, leavers). The second component, $\Delta\overline{w}_t^{cw}$, is also a composition effect arising from the different worker and job characteristics of the employment status groups across time. Finally, the third component, $\Delta\overline{w}_t^{pw}$, is the "pure wage effect" caused by the changes in the returns to worker and job characteristics after taking into account the changing weights of the groups with different employment status (stayers, voluntary movers, involuntary movers, entrants, and exits).

3.1 Accounting for self-selection into groups

Obtaining previous decomposition requires the estimation of the β parameters from equations 6 and 7 by period and group. These estimates can be obtained by using OLS. The problem is that if workers are not randomly selected into the groups, OLS coefficients can be contaminated by the effect of unobservable characteristics potentially correlated with wages. Therefore, OLS estimates of the wage equations may be biased because the observed sample of individuals in a given group may not be a random sample of the population and this can also introduce a bias in the decomposition of interest.

To account for these selection issues, we estimate wage equations using a version of the selection correction model by Dubin and McFadden (1984, hereafter DM). In particular, we use the improvement proposed by Bourguignon, Fournier and Gurgand (2007, hereafter BFG). The model is based on the estimation of a multinomial logit model for the group choice in a first stage. The second stage estimates a wage equation for each group and period using OLS with selection correction terms from the first stage to control for the correlation of errors between the two equations.[4]

More formally, our empirical model has two inter-related equations: a discrete group-choice equation (12) and a wage equation (13). That is, for each group g, $g = 1, .., G$,

$$U_{it}^g = Z_{it}^g \gamma_t^g + \tau_{it}^g, \qquad (12)$$

$$w_{it}^g = X_{it}^g \beta_t^g + \varepsilon_{it}^g, \qquad (13)$$

where U_{it}^g is the latent utility that individual i attaches to group g in period t. The utility includes a vector of individual characteristics (Z), and an unobserved stochastic component τ which captures all the variables that are relevant to the individual but unknown to the econometrician. Equation (13) specifies individual log wages as a function of a vector of observable (X) and unobservable characteristics (ε).

---

[4] For the estimation we use the "selmlog" command in STATA.

The outcome variable w_{it}^g is observed if and only if group g is chosen, which happens when:

$$U_{it}^g > \max_{s \neq g}(U_{it}^s). \tag{14}$$

If we assume that the (τ_t^g)'s are independent and identically type I extreme value distributed, as shown by McFadden (1973), we obtain the multinomial logit model with:

$$P_{it}^g = \Pr(U_{it}^g > \max_{s \neq g}(U_{it}^s) \mid Z_{it}^g) = \frac{\exp(Z_{it}^g \gamma_{it}^g)}{\displaystyle\sum_{s=1}^{G} \exp(Z_{it}^s \gamma_{it}^s)}, \tag{15}$$

where P_{it}^g is the probability that category g is chosen in period t by individual i.

Based on expression (15), consistent maximum likelihood estimates of the (γ_t)'s can be easily obtained. The problem is to estimate the parameter vectors β_t^g. If there are unobserved characteristics of the individuals that affect both their group and their wages, then the error terms ε_{it}^g may not be independent of the (τ_{it}^g)'s and the OLS estimates of β_t^g would not be consistent.

We next provide some details on the estimation methods that correct for previous self-selection problem in the wage equations. In particular, Heckman's (1979) two-stage selection model can be generalized to the multinomial logit case and a bias correction term can be based on the conditional mean of ε_{it}^g,

$$E(\varepsilon_{it}^g \mid Z_{it}^1 \gamma_t^1, ..., Z_{it}^G \gamma_t^G) = f_t(P_i^1, ..., P_i^G),$$

where $f(\cdot)$ is a non-linear transformation of the probabilities that an individual is observed in each group, as given in (15). Therefore, one can obtain consistent estimates of β_t^g estimating the following extended model by OLS:

$$w_{it}^g = X_{it}^g \beta_t^g + f_t(P_i^1, ..., P_i^G) + \xi_{it}^g, \tag{16}$$

where ξ_{it}^g is a residual mean-independent of the regressors.

Lee's (1983) proposed a selection correction method based only on the own group so it requires the estimation of only one parameter in the correction term. However, the Lee's model places substantial restrictions on the covariance between the error term of the wage equations and the selection indices.[5]

In this paper we follow the approached by BFG (2007). Specifically, they generalize the approach by DM (1984) who propose a selection correction based on a linearity assumption between ε_{it}^g and all the (τ_{it}^s)'s. The number of bias correction terms in each wage equation is then equal to the number of multinomial logit choices. However, the specific form of linearity imposed by DM restricts the class of allowed distributions for ε_{it}^g (the family of Gumbell's distributions). BFG (2007) suggest a variation on DM's assumptions that makes ε_{it}^g linear on a set of normal distributions allowing in

---

[5] Another approach is the one followed by Dahl (2002), who proposed to approximate the correction term by a polynomial function of the probabilities $P_i^1, ..., P_i^G$. Nonetheless, this semi-parametric correction becomes computationally unfeasible to implement in practice as the number of alternatives increases.

particular ε_{it}^g to be also normal.[6]

Substituting the error term of equation (13) by its conditional expectation (see BFG, 2007 for details) plus a residual term, we get this expression:

$$w_{it}^g = X_{it}^g \beta_t^g + \sigma_{gt} \left[\rho_t^g \delta(P_{it}^g) + \sum_{s \neq g} \rho_t^s \frac{P_{it}^s}{P_{it}^s - 1} \delta(P_{it}^s) \right] + v_{it}^g, \qquad (17)$$

where ρ_t^g is the correlation coefficient between τ_{it}^g and ε_{it}^g and $\delta(P_{it}^g)$ is an integral that have no closed form, but that can be computed numerically after the multinomial logit estimation. The number of bias correction terms in each wage equation is equal to the number of multinomial logit groups. To account for the two-stage nature of the procedure, we apply bootstrap to estimate the standard errors of the parameters.

We assume that the model is non-parametrically identified from exclusion of some of the variables in Z from the variables in X. In particular, we consider variables with individual variability such as the number of previous unemployment and employment spells, the number of previous temporary contracts, and the proportion of time employed since the first register with the Social Security administration until current employment.[7][8]

4   Results

Figures 2 and 3 displays the results of the previous accounting exercises that break down (quarterly) wage growth in composition effects from employment status and individual characteristics, and a "pure" wage growth effect, which proxies more closely wage pressure and the cyclicality of wages.

First, Figure 2 shows the decomposition of wage growth as in equation 5, taking into account only employment dynamics, without neither accounting for individual characteristics nor controlling for endogenous selection into employment categories. The magnitude of these composition effects, $\Delta \overline{w}_t^c$ in equation 5, are noticeable (see blue line in the figure), ranging between +1 pp and -1 pp and being strongly countercyclical. As a result, the observed wage growth (see yellow line in the figure) underestimated the "pure" wage growth (see orange line in the figure) at the beginning of the sample until 2008Q3 (before the Great Recession) and in the recovery starting in 2014Q2. On the contrary, between these two sub-periods, observed wage growth grossly overestimated the underlying "pure" wage growth by as much as a 1pp at the drought of the recession. This very simple decomposition shows that aggregate measures of wage growth are very much affected by employment dynamics and that, depending on their usage, may not be a good approximation to the true underlying wage pressure in the economy.

---

[6]BMF (2007) compare Dahl's estimation procedure with Lee (1983) and DM (1984). They conclude that Dahl's and DM are preferable to Lee's method and that their own variant of DM is more robust to various data generating processes, even when the Independence of Irrelevant Alternatives (implicit in the multinomial logit model) is violated.

[7]Bentolila et al. (2017) use also these type of instruments.

[8]Detailed results from OLS and BFG estimations are available from the authors upon request.

**Figure 2. Wage growth and composition effects due to employment dynamics**
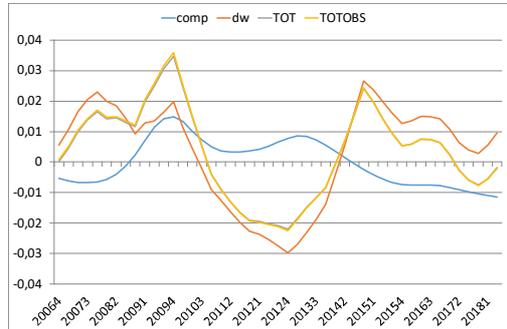
Nonetheless, the decomposition shown in Figure 2 does not account for characteristics or selection effects. Figure 3 shows the decomposition as in equation 14, when composition effects arising from changes in individual characteristics are also taken into account. We perform this decomposition under three alternative specifications: (i) a non-parametric approach (with only cells defined in terms of age, gender and education as individual characteristics are considered), (ii) OLS estimations of the impact of individual characteristics on wages (additionally including sector of activity, region and nationality as covariates), and (iii) and BFG estimations to account for endogenous selection into the groups.[9] First panel in Figure 2 shows the composition effects due to characteristics plus the composition effects due to employment dynamics (terms $\Delta \overline{w}_t^c + \Delta \overline{w}_t^{cw}$ in equation 10), while the second panel shows the "pure" wage growth, net of all previous composition effects (term $\Delta \overline{w}_t^{pw}$ in equation 10).

We draw three main conclusions from these results. First, composition effects have the same cyclical patterns as the one computed from the simple approach not accounting for individual characteristics. Nevertheless, the inclusion of individual characteristics makes the size of composition effects larger since 2009Q4. These composition effects due to worker and job characteristics are different across groups. While ageing and education push wages of stayers upwards (the relative weights of older and more educated workers increase gradually), immigration status and sectors of activities are also relevant for the evolution of wages of the movers, and sectors of activities and gender explain the largest composition effects among entrants-exits (the relative weight of female workers is much larger among entrants than among exits). In any case, the main source of composition effects is employment dynamics, that is, the change in the weights of the (six) categories of employment status considered Secondly, although there are noticeable differences between composition effects estimated by OLS (see orange line in first panel of Figure 3) and BFG (see grey line in first panel of Figure 3), unobserved heterogeneity driving the selection in each one of these six categories does not seem to be quantitatively large. Finally, the cyclicality of the "pure" wage growth is a bit less acute under OLS and BFG estimation (see orange and grey lines in second panel of Figure 3) than under the simple non-parametric approach (as should be expected) but still the same qualitatively result remains: observed wage growth underestimates the "pure" wage growth component in expansions and overestimates

---

[9]The non-parametric approach does not allow to account for as many characteristics as in the parametric ones due to the small number of observations in certain cells.

it in recessions. In our sample the largest differences between observed and "pure" wage growth are found in the first quarter of 2010 (-1,5% to -2,5%) and in the second quarter of 2018 (0.7% to 1.5%).

Figure 3. Composition effects (left panel) and "pure" wage growth (right panel)
(under non-parametric approach, OLS and BFG estimations)
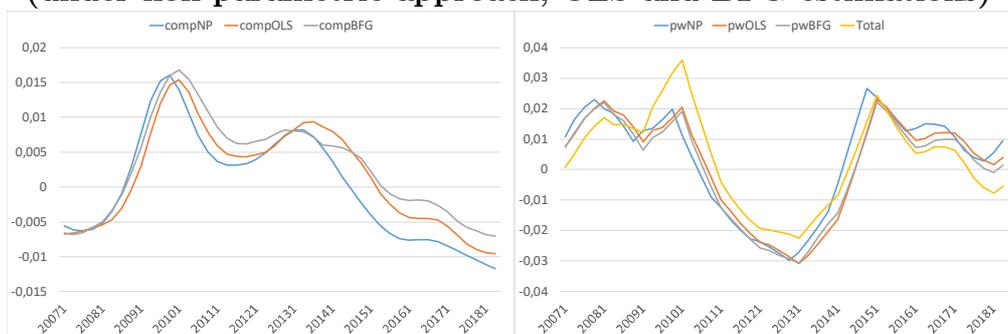
5   A naive test: Simple Phillips curves

Nowadays there is a debate about the "flattening of the Phillips curve", that is, the apparent breakdown of the traditional relationship between inflation and unemployment. One explanation refers to the mis-measurement of the "slack in the labor market" (Hong, Kóczán, Liam, and Nabar, 2018). Another to the lessening of the pass-through of wage inflation to price inflation (Heise, Karahan, and Sahin, 2020). Less attention is given to the possibility that wage growth is also mis-measured. However, our results above suggests that composition effects may blur the "true" wage pressures that should feed into macro analysis.

To address the latter, we perform a very simple test that looks at the correlation of alternative measures of wage growth and labor market slackness (measured both by the unemployment rate and the non-employment rate, that is, 1 minus the employment rate of population 16-64 years of age). The results (see Table 3) are quite convincing. Our preferred measure of "pure" wage growth (obtained under any of the three alternative specifications) show a higher negative correlation with both indicators of labor market slackness than the one obtained when using the observed wage growth. Clearly, the relationship between wage inflation and unemployment (non-employment) seems stronger and more robust when wage growth is measured cleaned of composition and selection effects. On the other hand, this makes more puzzling the vanishing pass-through of wage inflation to price inflations, something which merits more research.

Table 3: Correlation between measures of wage growth and slackness

| | Unemployment rate | | 1 minus Employment rate | |
Wage growth measure	Coefficient	Adj. R^2	Coefficient	Adj. R^2
Observed	-0.123 (3.2)	0.168	-0.190 (3.2)	0.168
Pure (non-parametric)	-0.198 (5.7)	0.410	-0.294 (5.1)	0.356
Pure (OLS)	-0.204 (5.7)	0.414	-0.327 (6.1)	0.446
Pure (BFG)	-0.198 (5.7)	0.410	-0.317 (6.0)	0.437

Note: Sample 2007Q1-2008Q2. t-statistics in brackets.

6   Concluding remarks

How to measure wage growth and which indicator should be used for business cycle analysis is a controversial issue with a long tradition in both Macroeconomics and Labor Economics. Recently, "vanishing deflation" during the Great Recession and "vanishing inflation" during the recovery afterwards have cast some doubts on what is the real measure of wage pressure driving price inflation. Moreover, structural changes in the labor market associated to population ageing and automation are affecting employment dynamics, altering the cyclical patterns of worker flows.

In this paper we have presented accounting exercises that take these changes seriously. With a very simple framework we have shown that changes in worker employment status (flows in and out of employment) affect significantly the different components (composition vs. pure components) of wage growth. Using Spanish data for the recent cycle we conclude that composition effects sustain wage growth during recessions and depress it in expansions. Thus, taking as a measure of wage pressure the "pure" wage growth (cleaning-out composition and selection effects), in the Great Recession wage pressure was lower than the observed wage growth, while it was higher in the recovery afterwards. This finding casts doubt on the flattening of the Phillips curve (in its wage version) but makes it more puzzling the smaller pass-through of wages into prices at the same time that price mark-ups seem to be rising. In any case, the mismeasurement of wage growth could be at the roots of many of the macroeconomic phenomena under discussion nowadays.

7 Appendix

Employment shares, wage levels, and wage growth by employment status along the wage distribution (by quartiles).

Figure A.1 Employment shares, wage levels, and wage growth by employment status.

First Quartile

a) Employment shares

b) Wage levels (logs)

c) Wage growth (annual rates)
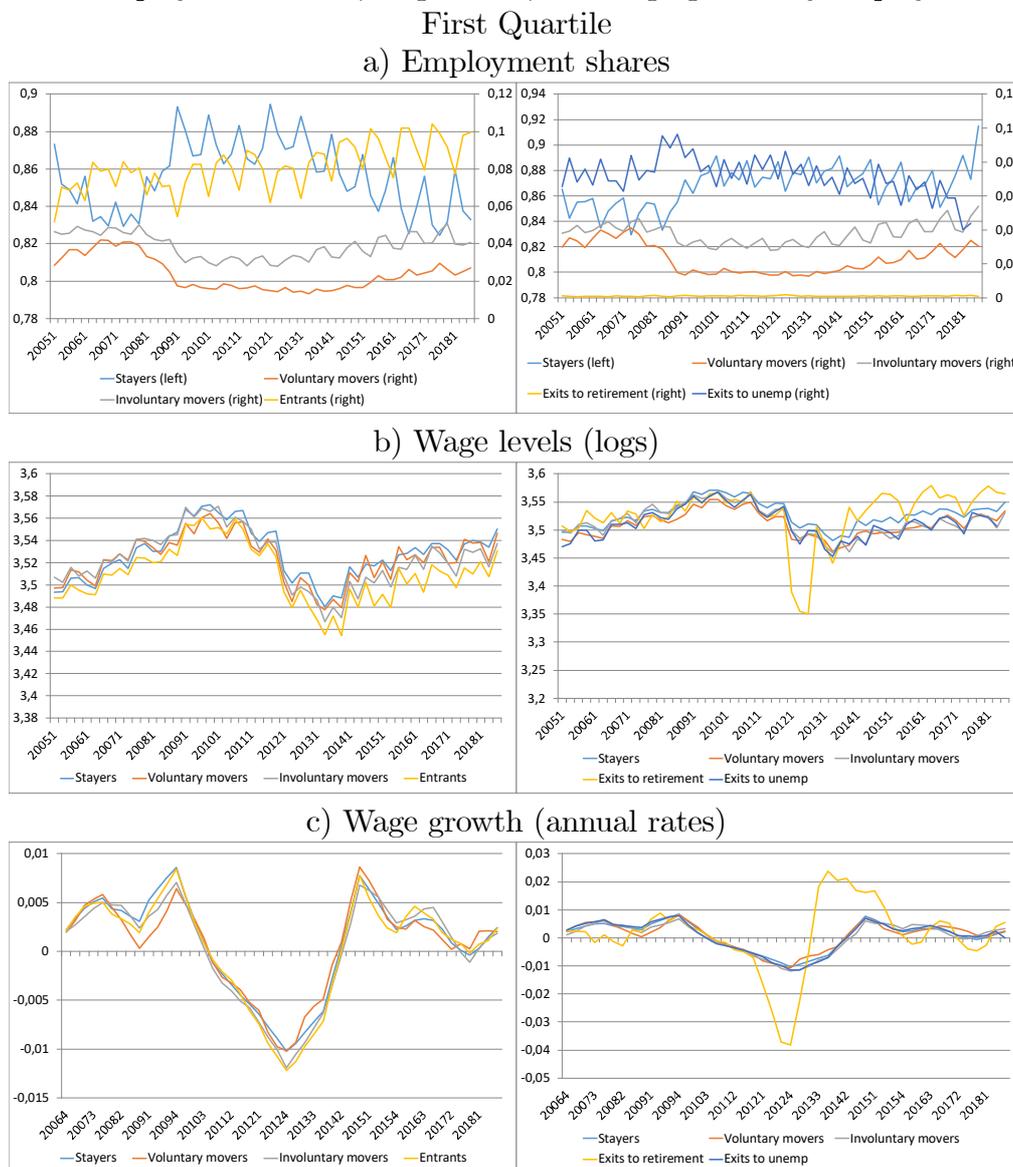
Figure A.2 Employment shares, wage levels, and wage growth by employment status.
Second Quartile

a) Employment shares
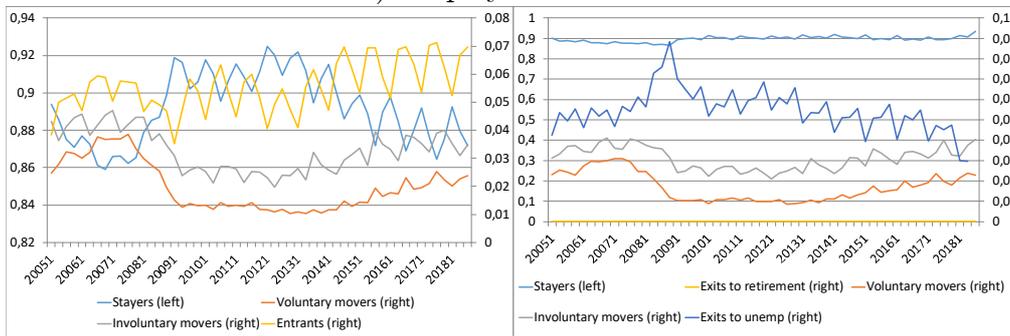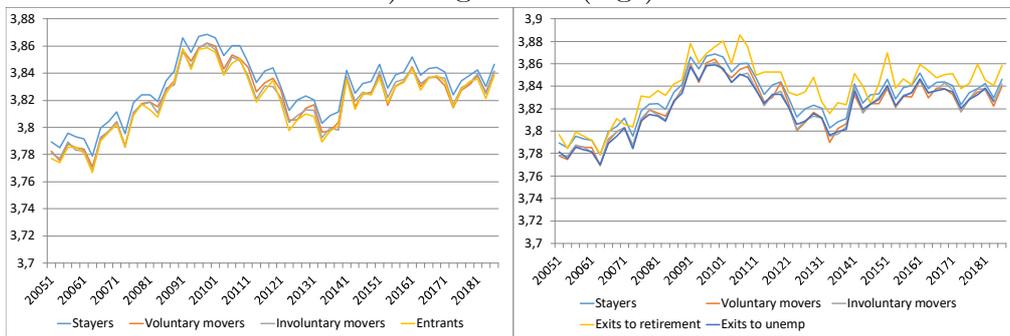

b) Wage levels (logs)

c) Wage growth (annual rates)
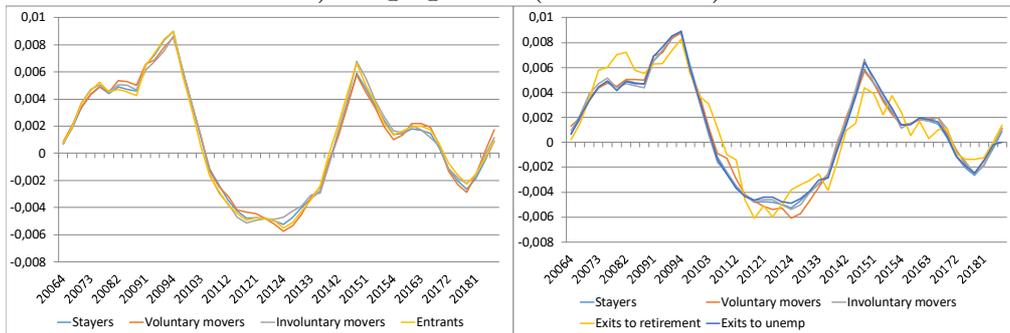
Figure A.3 Employment shares, wage levels, and wage growth by employment status.
Third Quartile

a) Employment shares
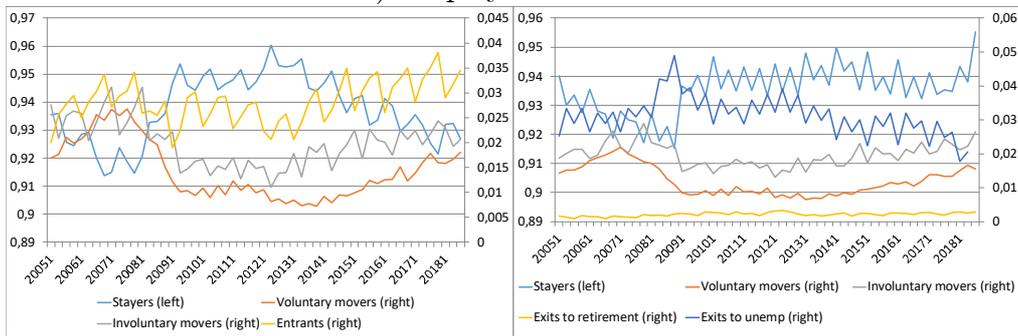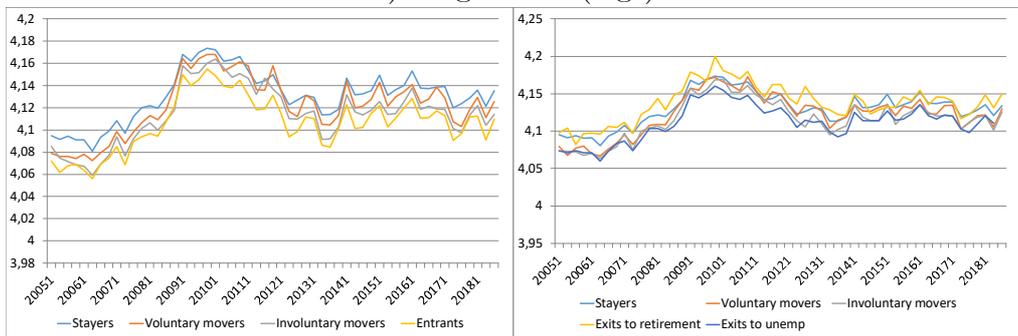
b) Wage levels (logs)

c) Wage growth (annual rates)
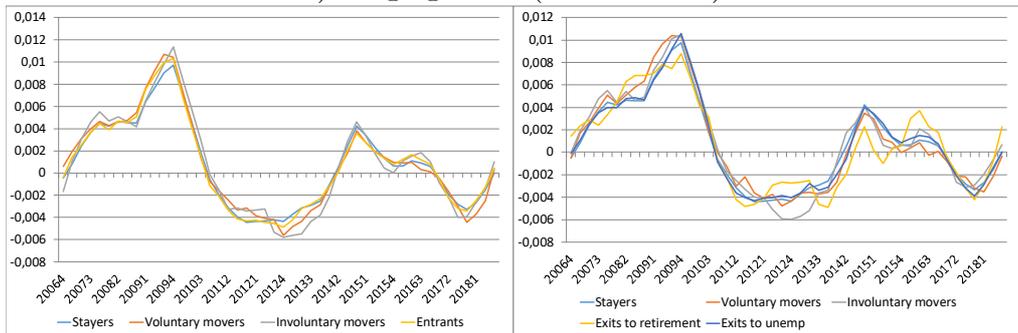
Figure A.4 Employment shares, wage levels, and wage growth by employment status.
Fourth Quartile

a) Employment shares
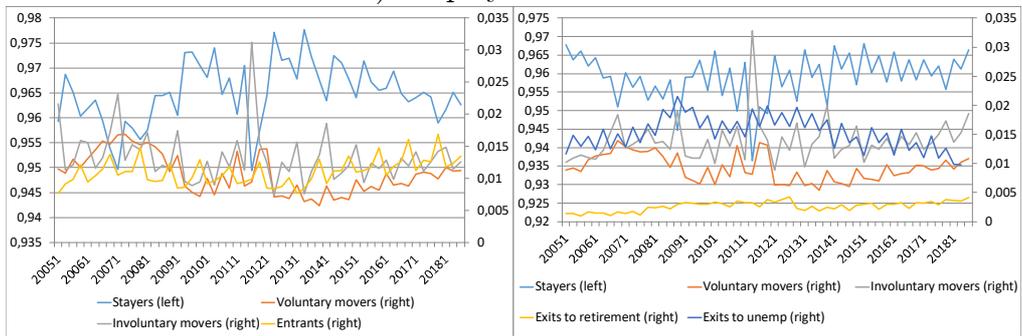
b) Wage levels (logs)
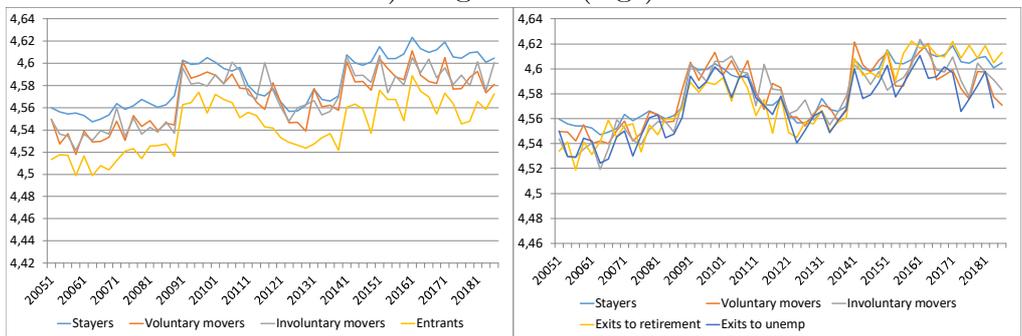
c) Wage growth (annual rates)
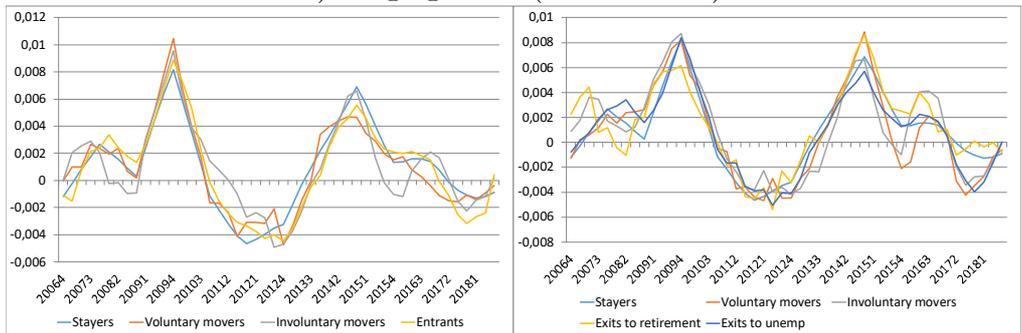
References

[1] Abraham, K.G and J. Haltiwanger (1995): "Real Wages and the Business Cycle", *Journal of Economic Literature*, XXXIII (September). pp. 1215-1264.

[2] Bentolila, S., García-Pérez, J.I. and M. Jansen (2017): "Are the Spanish Long-Term Unemployed Unemployable?", *SERIEs – Journal of the Spanish Economic Association*, 8, pp. 1-41.

[3] Bourguignon, F., Fournier, and D. L. McFadden (2007): "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons", *Journal of Economic Surveys*, 21 (1), 174-205.

[4] Carneiro, A., P. Guimarães, and P. Portugal (2012): "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity", *American Economic Journal: Macroeconomics*, 4 (2): 133-52.

[5] Dahl, G. B. (2002): "Mobility and the returns to education: testing a Roy Model with multiple markets", *Econometrica*, 70, 2367-2420.

[6] Daly, M.C. and B. Hobijn (2017): "Composition and Aggregate Real Wage Growth", *American Economic Review, Papers and Proceedings* vol. 107(5), pages 349-352, May.

[7] Dubin, J. A. and McFadden, D. L. (1984): "An econometric analysis of residential electric appliance holdings and consumption", *Econometrica*, 52, 345-362.

[8] Fatih, K., R. Michaels, B. Pugsley, A. Şahin, and R. Schuh. (2017): "Do Job-to-Job Transitions Drive Wage Fluctuations over the Business Cycle?", *American Economic Review*, 107 (5): 353-57.

[9] García Pérez J.I., and Y. Rebollo (2009): "The use of permanent contracts across Spanish regions: Do regional wage subsidies work?", *Investigaciones Económicas*, XXXIII, 97-130.

[10] Heckman, J. (1979): "Sample selection bias as a specification error", *Econometrica*, 47, 153-161.

[11] Heise, S., F. Karahan, and A. Sahin (2020): "The Missing Inflation Puzzle: The Role of the Wage-Price Pass-Through", NBER wp. 27663.

[12] Izquierdo, M., J.F. Jimeno, T. Kosma, A. Lamo, S. Millard, T.Rõõm, and E. Viviano (2017): "Labour market adjustment in Europe during the crisis:microeconomic evidence from the Wage Dynamics Network survey", ECB Occassional Paper Series, no. 192.

[13] Hong, G.H., Z. Kóczán, W. Lian, and M. Nabar .(2018): "More Slack than Meets the Eye? Recent Wage Dynamics in Advanced Economies", IMF working paper 18/50.

[14] Lee, L. F. (1983): "Generalized econometric models with selectivity", *Econometrica,* 51, 507-512.

[15] Mincer, J.A. (1974): "The Human Capital Earnings Function", NBER Chapters, in: Schooling, Experience, and Earnings, 83-96, National Bureau of Economic Research, Inc.

[16] McFadden, D. L. (1973): "Conditional logit analysis of qualitative choice behavior", In P. Zarembka (ed.), Frontiers in Econometrics. New York: Academic Press.

[17] Morris, M., R. Rich, and J. Tracy (2020): "How Aggregation Matters for Aggregate Wage Growth", *Economic Commentary,* Federal Reserve Bank of Cleveland.