Does Self-Employment Pay?
The Role of Unemployment and Earnings Risk *

Joaquin Garcia-Cabo† Rocio Madera‡

Abstract

This paper documents the role of unemployment and earnings risk in reconciling evidence in payoff differentials between self-employment and paid-employment. Using Spanish administrative data, we characterize the distribution and dynamics of earnings and document lower and less dispersed earnings in self-employment. We consider alternative hypotheses and highlight the role of lower unemployment risk in self-employment. We decompose earnings risk dynamics by estimating a life-cycle earnings process. Indeed, the self-employed experience lower returns but also face lower volatility and persistence of shocks throughout their life-cycle. Our results challenge the conventional view that self-employment necessarily entails higher risk and highlight that accounting for differences in labor earnings risk is important to reconcile the payoff differentials between self-employment and paid-employment.

Keywords: self-employment, segmented labor markets, earnings risk, income process

*We would like to thank Laura Hospido and Israel Arroyo for their help with the data, and Phil Cornell, Nicolas Savignon, and Tyler Powell for excellent research assistance. We thank our editor Francesco Zanetti and two anonymous referees for their great contributions towards improving this paper. We are also thankful to Ellen McGrattan, Kevin Donovan, Daniel Millimet, and various seminar and conference participants for their insightful comments. Most of the material in this paper was previously circulated under the title "The self-employment option in rigid labor markets: an empirical investigation". This research uses anonymized administrative data from the Muestra Continua de Vidas Laborales con Datos Fiscales (MCVL) obtained under permission of Spain's Dirección General de Ordenación de la Seguridad Social. Researchers interested in this dataset will need to obtain access to the MCVL data by applying to Spain's Dirección General de Ordenación de la Seguridad Social. The data used in this paper was acquired prior to the author(s) current employment at the Federal Reserve Board. This data was obtained from the Spanish Social Security while affiliated with University of Minnesota and Heller-Hurwicz Economic Institute and we have obtained all necessary permissions in relation to this paper. The views expressed are those of the authors and not necessarily those of the Federal Reserve Board or the Federal Reserve System. **Current version**: February 12, 2024.

†joaquin.garcia-caboherrero@frb.gov; Federal Reserve Board and Heller-Hurwicz Economic Institute.

‡romadera@smu.edu; Southern Methodist University and CESifo.

1 Introduction

This paper documents the role of unemployment and earnings risk in reconciling evidence in payoff differentials between self-employment and paid-employment. The entrepreneurship literature has documented that the returns to self-employment experience are lower than those for similar salaried workers, even when returning to paid-employment. This finding appears counterintuitive, considering the higher risk traditionally associated with self-employment. Various explanations have been proposed for this differential, including negative selection into self-employment (Evans and Leighton, 1989), the existence of non-pecuniary benefits from self-employment (Hamilton (2000)), and the value of experimentation (Dillon and Stanton, 2013). However, the quantitative nature of self-employment risk remains an open question in the literature (Moskowitz and Vissing-Jørgensen (2002)). This paper characterizes the dynamics of labor earnings risk for the self- and paid-employed.

Using data from Spain, we present new evidence indicating lower unemployment and earnings risks in self-employment relative to salaried alternatives. Our contribution, in the spirit of Smith (1776) and Dillon (2018), is applied to the self-employment choice, and states that occupational payoffs also capture all the inherent risks of the occupation. Our analysis illustrates that accounting for unemployment risk, beyond the conventional notion of entrepreneurial risk, could mitigate the overall risk faced by the self-employed individuals when labor markets exhibit high turnover, as it is the case in the presence of dual or segmented labor markets.

The Spanish labor market, known for its high unemployment rates and segmented job opportunities, serves as an ideal laboratory. Efforts to increase market flexibility while maintaining strict employment protections have resulted in a segmentation of paid employment opportunities: 30 percent of salaried workers are employed under unstable, fixed-term, *temporary* contracts, while 70 percent of workers are employed under *open-ended*, permanent contracts with high dismissal costs. These high mandatory dismissal costs for those already employed have made firms more prone to hiring under temporary unprotected contracts whenever possible, giving rise to the so-called dual labor markets in terms of job security. This segmentation, particularly prevalent among younger workers, disproportionately exposes them to unemployment risk, since most of these contracts end up in unemployment rather than converted into permanent positions (Güell and Petrongolo, 2007). During the Great Recession, Spanish unemployment spiked above 25 percent, and youth unemployment surpassed 50 percent. In the context of high unemployment and scarce stable employment opportunities, these rigidities and search frictions translate into lower job finding rates and high turnover for new salaried workers. At the same time, self-employment remains one of the highest in Europe at about 18 percent of the labor force.

We use the complete labor histories of a 4% sample of Spanish social security affiliates (approximately 1.2 million individuals), observing workers transitioning across labor market states and their earnings. Our dataset, known as *Muestra Continua de Vidas Laborales* (MCVL) has three key advantages: (1) the administrative nature, alleviating measurement error; (2) the large sample size, reducing the limitations of survey data; and (3) the longitudinal design, which allows us to follow the working histories of all individuals over the last three decades, including two different recessions, a significant labor market reform, and the highest growth decade in Spain's recent history. Most importantly, the richness of the dataset in labor market outcomes, firm characteristics, and demographics allows us to control for observed characteristics and address unobserved heterogeneity. Moreover, and in order to address possible misreporting to the Social Security of the self-employed, we complement our sample from the MCVL with the Survey of Household Finances (EFF, for short of Encuesta Financiera de las Familias, in Spanish) to compare self-employment income measured in survey and administrative sources.

In the first part of this paper, we document stylized facts on labor earnings and life-cycle dynamics across employment states. We first study the cross-sectional distribution of labor earnings for workers in paid employment with different job protection levels and for the self-employed. Most self-employment consists of small business owners who stay in business for longer tenures than all types of salaried workers. We then proceed as follows: First, we look at the cross-sectional earnings distribution and find that (median) earnings in self-employment are lower relative to stable employment but higher than unstable jobs across datasets. However, dispersion in self-employment earnings depends on the definition of earnings used, being lower in the Social Security records. Next, we classify workers into three categories based on their main employment status before age 40: mainly self-employed, mainly temporary, and mainly permanent. According to these groups, we look at workers' age-earnings profiles, finding that those predominantly self-employed when young experience lower earnings growth, even when returning to salaried employment. Finally, we exploit the longitudinal dimension of our data to study the dynamic earnings returns to labor market experience in paid employment and self-employment, paying particular attention to workers' contract histories. In line with the previous literature, we find that workers in self-employment exhibit lower market returns compared to similarly experienced workers in paid employment, even to those in unstable fixed-term employment. Throughout the analysis, we show these findings are robust when considering different sub-samples by skill heterogeneity.

In the second half of the paper, we relate the findings above with some theoretical explanations. In particular, we consider measurement issues, negative selection into self-employment, search frictions and job shopping in segmented labor markets, and lower lifetime labor earnings risk in self-employment. We test these hypotheses against the data to shed light on the mechanisms driving the lower earnings and market returns in self-employment. We do not find support for a job ladder in self-employment, as turnover is low and earnings growth limited, as opposed to paid employment, where the higher unemployment risk drives job shopping, resulting in earnings growth. We also do not see evidence that experimentation is a main driver of selection into self-employment or that workers face little penalty when returning to paid employment (see, for instance, Daly (2015) or Manso (2016) for comprehensive literature reviews). In our dataset, we observe self-employment spells to be longer on average than those in salaried employment, and returns to experience in self-employment after returning to salaried work are low. In our empirical analysis, we introduce individual fixed effects that control for unobserved heterogeneity, and we also explore the role of skill composition to control for observed heterogeneity. Nonetheless, we recognize the inherent limitations of our study in resolving negative selection and in quantifying non-pecuniary advantages.

Finally, we estimate a rich statistical model of lifetime earnings dynamics, grouping workers based on their youth employment status. Key to our mechanism, we find that the self-employed face lower permanent and transitory changes to their income over their lifetime, especially compared to fixed-term workers. The explanation is simple and was long ago introduced by Adam Smith (Smith, 1776) as a compensating earnings differential: When job turnover is high, self-employed workers are willing to accept a lower compensation to insure against unemployment risk. In this sense, self-employment becomes a lower-risk and long-lasting employment option. Without ruling out the existence of non-pecuniary benefits or intangible returns such as *sweat equity* (Bhandari and McGrattan, 2020), we conclude that lower earnings risk in self-employment compared to some salaried alternatives is a relevant channel when reconciling the self-employment option in labor markets with high turnover.

Overall, the findings in this paper suggest that accounting for unemployment and earnings risk is important to correctly account for occupational risk, especially in self-employment. This channel, while already documented in the literature at the industry or occupation level, has yet to be given specific attention to the case of self-employment. For instance, Cubas and Silos (2017) find a positive correlation at the industry level between earnings and occupation risk and average wages. Similarly, Dillon (2018) finds an essential role of earnings and employment risk in sorting workers across occupations. Our empirical analysis hinges on how earnings risk dynamics alter workers' outside options. Humphries (2018) documents the heterogeneity in skills and capital when selecting into self-employment as the driver for the dispersion in self-employment earnings rather than entrepreneurial risk. In line with this latter idea, we document that when most self-employed are not entrepreneurs but small business owners, and while we do not exclude the existence of entrepreneurial risk, heterogeneity in earnings risk matters. As a result, self-employment duration increases as the role of insurance against unemployment risk emerges. The results we present will further allow us to discipline macroeconomic models by using the second moment of the earnings distribution and its decomposition into permanent and transitory components without relying on large taste shocks or switching parameters to reconcile the option of becoming self-employed.

This analysis is also valuable for disciplining the design of active labor market policies aimed at reducing unemployment and promoting self-employment. Many recent papers analyze the role of self-employment promotion policies and their welfare consequences (Humphries, 2018; Hincapié, 2020), some of them in a high unemployment risk context (Hombert et al., 2014; Herreño and Ocampo, 2023). We present evidence that in the context of segmented markets regarding job protection, selection into self-employment is influenced by the lifetime earnings risk in all outside options, including riskier salaried alternatives. Hence, reassessing labor policies to consider earnings risk dynamics is crucial for policymakers. Finally, the results of this analysis are necessary to understand trends even in labor markets that are traditionally more flexible, such as in the United States. The rise of the gig economy, as well as differential unemployment rates across industries, have segmented the job opportunities workers face, affecting unemployment and earnings risk.

The rest of the paper is organized as follows. Section 2 summarizes the main features of the data. Section 3 studies earnings and returns in self-employment compared to two segmented salaried markets. Section 4 compares theoretical explanations to shed light on the mechanisms driving the facts presented in Section 3. Section 5 concludes.

2 Data on Full Labor Histories And Income

The Spanish Social Security Administration (SSA) Data. We use the Spanish SSA's *Muestra Continua de Vidas Laborales* (MCVL) data set. It consists of a 4% representative sample of Spanish individuals affiliated with the Spanish SSA in 2013-2016, whether employed, unemployed, or retired. Workers are added yearly to maintain the sample's representativeness, as workers who end their affiliation leave the database. The sample size is about 1.2 million individuals per year. The underlying source is the actual legal contracts signed between firms and workers. We observe the workers' demographics and daily job histories from the first affiliation day until 2016. Detailed job information is available back to 1967, and reliable monthly earnings data (based on social security contributions) back to 1980.

Sample Selection. We focus on prime-age workers—25 to 55 years old—to avoid capturing atypical behavior at the beginning or end of the career. In the interest of data quality, our preferred time period of analysis for earnings dynamics is 1990 to 2016. However, we use all available information from workers' labor histories to obtain their past trajectories.

Main Variables. For each working spell, relevant variables are the contract start and end date, the type of contract, occupation, salaries, and the cause of dismissal, as well as the firm's location, size, sector, and legal status. Of particular importance is the differentiation of the self-employed by their relation to other self-employed within the same firm, which allows us to identify *dependent* self-employed, although they represent less than 1% of our sample. For non-employment spells, we observe the associated unemployment benefits and pension amount, as well as the retirement date of the worker. Given the importance in our analysis, we discuss the income variables in detail in the next subsection. All nominal amounts are deflated using the Spanish CPI with base year 2016 provided by the National Institute of Statistics (INE). The Social Security matches demographic and residence information from the Census in the dataset.

Spell information in the dataset is detailed at a daily frequency and of high quality due to the administrative origin. These characteristics regarding job spells allow us to precisely identify self-employment spells from paid- and non-employment using the type of *regime*, or legal contract code, alleviating measurement error from self-reporting status. We refer to Appendix C for further details.

2.1 Measuring the Earnings of the Salaried and the Self-Employed

We define income throughout the paper as social security labor earnings. Specifically, monthly contribution bases to the SSA.[1] These bases determine Social Security taxes, future disability, unemployment insurance, and pension amounts for all workers, including the self-employed.[2] For the salaried workers, monthly labor earnings correspond to average monthly income, excluding overtime pay. Defining labor earnings is slightly more complicated for the self-employed: In some cases, the law fixes the contribution basis. In others, the worker chooses the basis by projecting their income. As a result, and given that some of the self-employment labor income has a self-reported component, we discuss the possible extent of misreporting in the next paragraph. Both salaried and self-employment income are subject to maximum and minimum caps established in the legislation.[3]

In light of how the SSA records income, the reader might be concerned that this data does not accurately reflect labor income for the self-employed. We next discuss misreporting in the data in the form of over-and under-reporting, and we also point out the existence of incentives to do selective misreporting to the social security, in particular smoothing out labor income over the year. While we will conduct robustness exercises to control for some of these issues, we will also acknowledge the limitations of our analysis through the text.

First, regarding under-reporting—with tax evasion as the primary objective—notice that the link between the contribution bases and generosity of future transfer lowers the incentives of workers to under-report. Second, regarding over-reporting, the link between contribution bases and taxes reduces incentives for self-employed workers to overstate their income.

[1] As argued in many papers, for a large share of the workers, base contributions to social security are a good proxy for total salaries (Garcia-Perez, 2008; Cuadrado, Hernández de Cos and Izquierdo, 2011; de la Roca, 2014).

[2] Both salaried and self-employed workers are entitled to similar amounts and duration of unemployment insurance and retirement pensions. We refer to Appendix C for an extensive description on the institutional background.

[3] These caps lead to some censoring from above and below in the data, especially above the 90th percentile, as shown in Bonhomme and Hospido (2016). However, the 80th percentile and the 50/10 ratio are always observed. Given this and the focus of this paper on job stability, which the spell information captures accurately, we believe our results are not sensitive to these limits. We refer to Bonhomme and Hospido (2016) for further analysis.

Perhaps as important as the issues mentioned above is selective misreporting or the incentives from the self-employed to smooth labor income reporting.[4] Potential issues deriving from smoothing reported income include smaller variation in income data (in the cross-section and at the individual level on the panel), which could artificially smooth the data and our results.

To inspect the existence of these data limitations and assess the robustness of our results, we complement our Full Labor Histories sample from the MCVL with the *Survey of Household Finances* (EFF, for short, in Spanish) to shed some light on the comparison of self-employment income measured in survey and administrative sources. The EFF is a triennial survey focused on measuring income and wealth. We use waves 2008, 2011, 2014, and 2017.[5] This survey contains questions specific about total self-employment income (including the division between labor and asset components), salaried income, contract information, and tenure into employment spells. To be the most comparable with our SSA definition, we obtain labor income information from module 6, which focuses on labor earnings and information regarding workers' labor arrangements.[6] The design of EFF resembles the United States's SCF in that it oversamples the wealthiest. In contrast to the SCF, the EFF has a larger panel component, keeping over half of the households between waves. While that allows us to observe some employment and earnings dynamics, the focus is more cross-sectional, resulting in a longitudinal sample that is very limited in size, length, and frequency. Other limitations include a much smaller sample size in the cross-section, annual rather than monthly information, and survey measurement error (Feng and Hu, 2013), and some studies have pointed out the existence of misreporting even in survey data in Spain (Martinez-Lopez, 2013) and in the United States (Bhandari et al., 2020).

Subsequently, we use the social security measure as our benchmark definition, given the benefits of MCVL's long longitudinal design and high-quality spell data. However, we will contrast our exercises with survey data. We believe that highlighting differences and similarities in results

[4]We acknowledge an anonymous referee pointing out this issue, and we have borrowed their language

[5]We omit the first and last waves. The design was revisited in the second wave, 2008, making variables more homogeneous and comparable across survey years. The last wave is 2020.

[6]For salaried workers, while the natural object of study would be annual earnings, we do not have a split by contract type (permanent or temporary). As a result, we use question 6.13 to obtain information regarding the contract they held at the time of the interview and 6.14 to obtain monthly labor earnings. For the self-employed, we use question 6.102 on the labor earnings of the business at the monthly level. We compute the annual amounts by multiplying them by 12 months and convert them to real earnings using the 2016 CPI deflator. We assess the robustness of this measure by using the annual income information in the survey regarding the prior year (questions 6.64 and 6.72), and we obtain comparable measures for salaried and self-employed. Moreover, we use this prior year's annual measure for longitudinal comparisons between salaried and self-employed. Workers report in question 6.81 which of the two occupations (salaried or self-employed) they have spent most of their careers.

across databases is a contribution of this paper, adding to the literature on measurement, in this case, for the Spanish labor market. Moreover, to alleviate some of the issues mentioned earlier, our analysis will mostly rely on annual income and provide robustness checks based on other spell variables throughout the paper, such as tenure and spell transitions. Finally, whenever possible, we include evidence using salaries of the paid-employed after a self-employment spell, less exposed to these caveats.

2.2 Descriptive Statistics

We present an overview of the Spanish labor market, focusing on workers' characteristics, in Table 1, which highlights the richness of the dataset. The self-employed are, on average, older, earn less, and have been at their current firm for a longer time compared to workers in salaried employment. They tend to have lower educational attainment and are predominantly male.[7] In terms of the legal nature of the firms they own, they are constituted mainly as sole proprietors or cooperative members.

2.3 Measuring *Main* Employment Status in Early-Career

The analysis in this paper uses both cross-sectional variation and the panel dimension. To perform our cross-sectional analysis, we rely on contract information from the SSA to classify workers as salaried- permanent or fixed-term/temporary- self-employed, and non-employed (either receiving unemployment benefits or not having an attachment with the SSA). The panel analysis requires some additional assumptions as workers transition between contracts, and it is necessary to create bins of workers grouped by specific characteristics to estimate returns to experience or life-cycle income profiles. For this purpose, we follow a similar approach as in Cabrales et al. (2020), categorizing workers based on the occupation where they spent most of their time before age 40 (age 40 is included). We construct these groups as follows: first, we compute a worker's total number of days worked in a given year. Second, we compute the share of days out of the total in each occupation in the year. Finally, we sum the number of days worked a year under each occupation before age 40. This procedure gives us a ranking of the four occupations. We classify workers as "predominantly permanent" if the share of days worked under a permanent

[7] These are common characteristics of the self-employed across countries and datasets. See, for instance, Millán Tapia (2012); Humphries (2018) for European evidence, and Manso (2016); Hincapié (2020) for the United States.

	All	E(P)	E(T)	SE	NE
Demographics (%)					
Female	45.51	41.02	46.52	31.21	56.40
College	17.78	20.89	20.21	12.18	12.86
Mean					
Age	38.57	38.37	35.75	41.21	37.97
Real Monthly Earnings (Euros)	1625.63	1698.96	1367.52	964.32	324.85
Tenure (years)	5.25	4.61	1.11	7.97	0.52
Job Status (%)	100	47.22	16.88	15.28	20.61
Public		8.42			
Non-Public		91.58			
Fixed-Term /Contract of service			95.74		
Training, Substitution, and other			4.26		
Sole or Liberal Professional				64.74	
Coop				21.61	
Scorp CEO				7.77	
Family Business				4.77	
Dependent and Other				1.11	
Unemp. w/ Benefits					30.38
Rest					69.62

Table 1: Summary Statistics

Note to Table 1: Summary statistics of our sample, including the distribution across job status. *All* denotes the full sample, *E(P)* denotes salaried workers under a permanent contract, *E(T)* denotes salaried workers under a temporary contract, *SE* denotes self-employed workers, and *NE* denotes non-employed workers. *Demographics* contain summary measures that are time invariant and only using one observation per worker. *Mean* refers to the average of variables that change every month. *Job Status* contains the distribution of workers across coarse job status, as well as the detail distribution by subgroups within each job status.

contract is the highest of the four categories. Accordingly, a worker will be classified as "predominantly temporary," "predominantly self-employed," and "predominantly non-employed" if these categories have the highest share of days worked out of total days worked before age 40. Note how we purposely exclude workers with large shares of non-employment: if a worker spent most of their time non-employed, they would not be included in the three main employment buckets. However, the other groups may have some time spent on unemployment. Moreover, this does not make any assumptions about labor market careers after age 40.

Additionally, we explore the distribution of time worked outside their predominant occupation in table A.7 in Appendix A. We find that, when young, all three groups spend about two-thirds or more of their worked days before age 40 in their predominant occupation. There is some variation on days spent non-employed, with predominantly temporary workers when young spending twice the share compared to those predominantly permanent or self-employed. Moreover, we find persistence at later stages in life in young statuses when looking at their careers beyond age 40. Those predominantly permanent and self-employed when young continue to be predominantly in those occupations past age 40, and while we observe some of those predominantly in temporary when young finding stability in permanent jobs later in life, there is still significant incidence of fixed-term employment after 40. In our longitudinal analysis below, we use these main statuses when young to study the earnings dynamics of workers by occupation.

3 Self-Employment Earnings and Returns Compared to Two Segmented Salaried Markets

In this section, we document the distribution of earnings and labor market returns in self-employment when workers face higher unemployment rates and how these compare to the job alternatives in a segmented salaried market regarding job protection. Our results rely on administrative records, our preferred source, but we complement our analysis with survey data. While the literature has shown significant discrepancies between survey and administrative records regarding cross-sectional statistics of business incomes (Bhandari et al., 2020), this analysis aims to control for possible misreporting in administrative records. We will be upfront about the advantages and disadvantages of both datasets. We proceed as follows: First, we look at the cross-sectional earnings distribution and find that earnings in self-employment are lower relative to stable employment but higher than unstable jobs across datasets. However, dispersion in self-employment earnings depends on the definition of earnings used, and it is slightly smaller in the . Second, we generate age-earnings profiles based on the main occupational/employment experience at young ages. We find that the earnings of those predominantly in self-employment when young grow less over the life cycle, even when returning to salaried work. Third, we estimate the returns to experience using the longitudinal dimension of our data to find that these are lower in self-employment compared to permanent and fixed-term salaried jobs. Throughout the analysis, we show these findings are robust when considering different sub-samples by skill heterogeneity.

3.1 Cross-Sectional Evidence: Lower but More Stable Income in Self-Employment

The top panel of Figure 1 depicts the cross-sectional distribution of log-annual earnings[8] for workers in salaried jobs and self-employment. We explicitly separate those in fixed-term *temporary* contracts and those in open-ended *permanent* contracts. Using administrative records, we find: First, median annual earnings in self-employment are lower than those in stable, *permanent* paid employment. Second, median earnings at the annual level for fixed-term workers are below those for the self-employed but higher at the monthly level, signaling more turnover throughout the year due to the temporary nature of the contract. [9] Third, self-employed log earnings exhibit the lowest dispersion in the social-security records, with a standard deviation of 0.54, compared to 0.91 and 1.25 for permanent and temporary workers, respectively, driven by a mass of self-employed workers at the minimum legal contribution bases. As discussed before, this lower dispersion measured using the variance could also result from income smoothing by the self-employed, given their commitment to a certain level of contribution basis throughout the year. [10]

Motivated by the concern that the distributional findings of this paper for the self-employed could be due to income misreporting and smoothing—as a reminder to the reader, they commit to a specific contribution basis to the social security—we reassess these facts by replicating this figure using business labor income from the EFF (Figure A.2, Appendix A). Regardless of whether we

[8]Earnings are aggregated at the annual level and employment status defined as the December recorded status by the SSA for each year. We do not observe hours and hourly wages separately, as most of the earnings risk dynamics literature using social-security records (Guvenen et al., 2021). We thus focus on total labor earnings.

[9]We present the cross-sectional log-earnings distribution at the monthly level in Appendix A.

[10]The minimum basis for full-time self-employed workers has varied little in real terms across years in our sample, from about 750 euros in 1990 to 890 in 2016. Garcia-Cabo and Madera (2019) show that there exists substantial heterogeneity within the self-employed in terms of earnings and educational attainment by industry: those in the energy, IT and finance, public administration, and health industries are, on average, more educated and enjoy higher earnings.

rely on contribution bases from administrative records or we use business income from the EFF, facts 1 and 2 regarding the relative ranking of earnings among the three occupations remain true. It is always the case that the median self-employed earns less than the median permanent worker, in line with standard findings that self-employment pays less than standard-paid employment. It is also true that the earnings of the median fixed-term worker are the lowest among the three. However, administrative data for the self-employed is less dispersed. In particular, and while the standard deviation of self-employment log earnings is extremely close in the EFF compared to the SSA (0.60 vs. 0.54, respectively), salaried workers' log earnings are significantly less dispersed in the EFF for both permanent and temporary workers (the standard deviation is 0.54 and 0.53 for permanent and temporary log-earnings, versus 0.91 and 1.25). Overall, temporary income is slightly less dispersed in the EFF than self-employed income, as opposed to what we observe in the SSA.

We inspected the data to understand the sources of these discrepancies for the salaried. Notice that while earnings are capped in the SSA by legal contribution bases, limiting the observed right tail, the SSA database is representative of workers attached in a given month. This selection criterion increases the density of the left tail, as some workers have limited attachment throughout the year. [11] Indeed, we assessed the role of limited labor market attachment, and if we exclude those without continuous employment in a year, the standard deviation of permanent and temporary decreases to 0.73 and 0.93, respectively. The reduction is especially remarkable for workers under temporary contracts, who are more affected by contract turnover and unemployment.

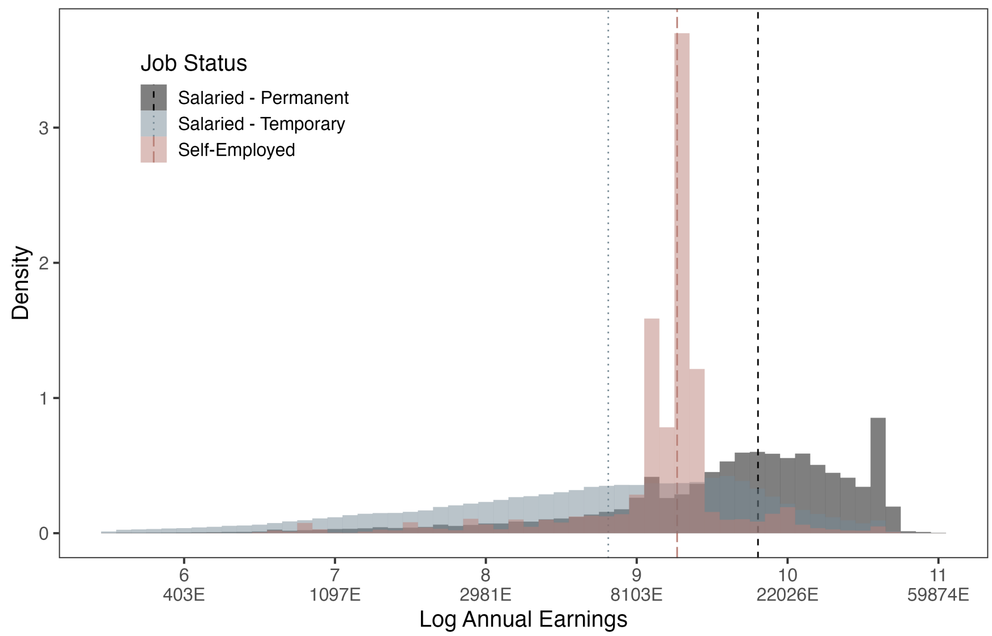
Moreover, we would like to highlight that, while likely reflecting smoothing self-employment income in the social security that it is not as present in survey data, the fact that there exist differences in dispersion in self-employment income across datasets is not novel to this paper: Hamilton 2000 showed longer right tails for the self-employed using SIPP data compared to wage workers, but depending on the definition of self-employment income, one could not find differences relative to the wage distribution. More recently, Bhandari et al. (2020) provide evidence that, for the United States, there indeed exist significant differences in self-employment earnings levels across surveys, with most of them overstating business income. They show that this is true even when adjusting IRS data for misreporting. One of such surveys is the Survey of Consumer Finances, which has the same design as the EFF. Both surveys aim to capture the large concentration of

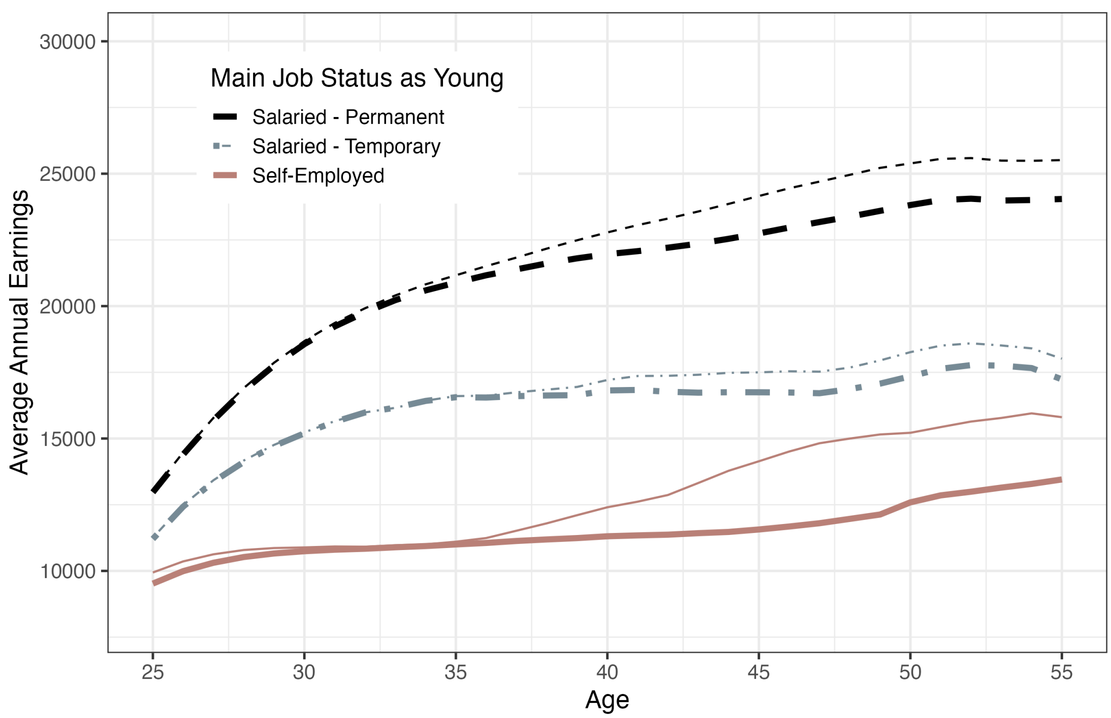
[11] As shown in Bonhomme and Hospido (2016), the top 80 percentile is always observed.

wealth at the top. For that purpose, they undersampled those with low income, which would result in a larger right tail in our sample, even when using population weights. These two factors (smooth reporting in SSA and oversampling of the wealthy in the EFF) likely explain the differences across cross-sectional distributions. However, it is hard, if not impossible, for us to quantify the extent of these given data limitations, which is beyond the scope of this paper.

All in all, we find that all the differences between the EFF and the MCVL arise for both salaried and self-employed. As a result, we do not expect the comparison between the experiences of the self-employed and the salaried to be affected. In the remainder of this analysis, we will continue comparing our findings using survey data. Whenever that is impossible, we do robustness with our data by focusing on workers returning to paid employment after self-employment. We next explore whether there are dynamic benefits to choosing self-employment in terms of higher earnings growth.

Figure 1: Log-Earnings Distribution By Employment Status





Note to Figure 1: **Top panel:** Histogram and Kernel density estimates for the cross-sectional distribution of annual real earnings (in 2016 Euros). Vertical lines denote the median for each distribution of the same color. x-axis includes the corresponding euro amount for ease of interpretation. **Bottom panel:** Main job status as young is defined as predominant status before 40 by number of days. Thick lines are average earnings at each age for all workers of a given age and with the given main status. Thin lines further limit the sample to those that, independently of their status before 40, are salaried after 40.

3.2 Life-Cycle Dynamics: Earnings of the Young Self-Employed Grow Less over Life

We next focus on how earnings evolve as workers age. We first look at average unconditional earnings by age of workers based on their main employment status before age 40 in the bottom panel of Figure 1. Interestingly, this first look at earnings dynamics suggests that self-employment experience is not compensated over time compared to workers predominantly in paid employment before 40 years old, especially those in permanent employment. We demonstrate below that this finding is robust even when conditioning on educational attainment and for workers only on salaried employment at later stages of their working lives.[12]

Workers who spend most of their young lives in self-employment or fixed-term contracts perform worse than workers with highly protected contracts at every age. Moreover, while at age 25, there are small differences between those mainly self-employed before age 40 and those mainly fixed-term, the slope of the earnings profile of the former quickly deteriorates, driven by the flat slope of earnings growth for non-college graduates. To shed light on the differences between fixed-term and self-employment workers, we zoom in on earnings for only those workers who ended up in paid employment after 40 (solid lines). Conditional on having a salaried contract after 40, the earnings gap compresses substantially. However, it only partially closes over the life-cycle for those predominantly in self-employment relative to fixed-term employment before age 40. Motivated by differences in educational attainment by occupation, we additionally estimate these age-earnings profiles separately for two different skill groups: college graduates and non-college graduates, results that can be found in Appendix A.2. When inspecting the different skill subgroups, college grads that were young self-employed have a faster catch up with fixed-term, but in any case, they do not converge to those in permanent contracts at young ages. [13]

Overall, 78% of predominantly self-employed workers before age 40 are still self-employed at age 50, while only 25% of those predominantly fixed-term before age 40 are still under fixed-term employment at age 50. Instead, 57% have obtained a permanent contract, but they do not fare as well as those who started their career in a stable, permanent job. Since worker heterogeneity and selection play a role in these earnings profiles, we decompose earnings growth into returns to the

[12] This result also holds even when excluding workers with zero income at a given age. This result is shown in Appendix A.2.

[13] The limited sample size and short panel dimension of the EFF, and in particular the lack of full labor histories, does not allow for a direct comparison of these findings with survey data.

labor market experience and seniority. This analysis aims to control for observed skills beyond education and unobserved characteristics from workers' labor histories, that even though will not fully resolve selection concerns, will shed light on differences in observed characteristics and the composition of these groups of workers.

3.3 Longitudinal Evidence: Returns to Experience and Seniority are Lower in Self-Employment

We exploit the longitudinal dimension of our dataset and use a prototype model for earnings growth as in (Topel, 1991) to study the sources of the different earnings profiles illustrated earlier. While this simple model has some shortcomings (see Buchinsky et al. (2010) for a literature review), it provides a simple framework to study the source of earnings growth differences across workers. In particular, we use the following reduced-form Mincer equation of earnings determination:

$$y_{i,t} = \mu_i + E_{i,t}\beta_1 + T_{i,t}\beta_2 + X_{i,t}\beta_3 + \xi_{it} \tag{1}$$

The dependent variable in equation 1 is the logarithm of the yearly deflated earnings $y_{i,t}$ for worker i in year t. We include the following regressors: μ_i is a person-specific fixed-effect, $E_{i,t}$ is a vector containing worker i's labor market experience at time t, $T_{i,t}$ denotes seniority at the job, $X_{i,t}$ is a vector of observed characteristics, and ξ_{it} is the error term.

We estimate this equation separately for workers based on their main labor market attachment before age 40: predominantly permanent, temporary, and self-employed. We construct *experience* in the dataset using the difference between the first year of entry of the worker in the labor market, observed as the first affiliation of the worker with the social security, and the current year.[14] *Tenure* is defined as the number of years spent in a given spell at each t. In the estimation, we include a quartic polynomial in experience, a quartic polynomial in tenure, year effects, industry effects, person-specific fixed-effects, and a constant. We cluster standard errors at the individual level.[15]

[14]We look at individual's histories from age 22 or first entry onwards. In our baseline specification, we require at least 365 days of labor history between age 22 and 40 to be in the sample. We have also estimated the model using potential experience, defined as age-22, with our conclusions remaining unchanged.

[15]We exclude from the estimation year-end observations for unemployed workers. Because some older spells lack industry information, we also run the same specification without including a control for industry, finding very similar results.

	Years of experience			
Young Status	**2**	**5**	**10**	**15**
Predominantly temporary before age 40	0.31	0.58	0.74	0.78
Predominantly self-employed before age 40*	0.14	0.28	0.38	0.42
Predominantly self-employed before age 40**	0.21	0.39	0.50	0.56

*all workers, **only those always in salaried work after 40.

Table 2: Estimated cumulative returns to experience

Table 2 summarizes the estimated returns to experience for different groups based on years of labor market experience.[16] We find that those workers predominantly in fixed-term, temporary employment before age 40 exhibit twice as high returns to general labor market experience compared to those predominantly in self-employment since the beginning of their careers. In the aim of controlling for possible misreporting, we also estimate the returns for those young self-employed who return to salaried employment after 40 years old. We also find that, despite higher experience returns compared to the full sample of young self-employed, their returns are about 30 percent lower after 15 years of experience compared to young fixed-term workers, even after controlling for observed and unobserved characteristics.

We assess the robustness of our results by studying whether differences in skill compositions could drive the differential earnings growth between the young temporary and self-employed workers. In particular, we next re-estimate equation 1 for non-college and college graduates separately. We present the full results in tables A.1 and A.2 in Appendix A.3. We find that controlling for skill does not change our unconditional results: those predominantly in self-employment at young ages get rewarded less in terms of labor market experience compared to young temporary workers. Moreover, and possibly as expected, college grads exhibit higher returns to general labor market experience (and intercepts) for all occupational groups. We perform a final robustness check by estimating 1 using the EFF, despite suffering from several limitations in studying a dynamic model such as this one. In particular, this survey lacks full labor histories that would allow us to construct different occupation groups based on young labor market experience, contains self-reported experience and tenure as opposed to administrative records, and is composed of a

[16]We focus on returns to general labor market experience and present the full specification estimates, including returns to tenure, in Appendix A.3. For the sake of readability, we only compare fixed-term and self-employment returns. We find statistically very similar returns to experience for those predominantly permanent when young compared to those predominantly temporary but much larger estimated constants, which leads to higher career paths.

short and small panel sample. We estimate separately on an unbalanced panel of workers who report being predominantly in salaried employment and those in self-employment, finding qualitatively similar results to the ones presented earlier. However, especially for the self-employed, the estimates are extremely noisy and imprecise, so we would like to remain cautious when using survey data from the EFF for this purpose. We present these results in Appendix Table A.6.[17]

These results are not surprising and reconcile the findings in the literature that self-employment experience does not reward workers in terms of earnings later in life. However, we have abstracted so far from lifetime earnings risk and uncertainty–this exercise only includes workers with positive earnings, abstracting from non-employment episodes with zero earnings–and how self-employment could be a successful option to escape labor market duality, earnings risk, and unemployment when their prevalence is at its highest—i.e., between 20 and 40 years old. We explore potential mechanisms at play in the next section.

4 Potential Explanations

In this section, we relate the findings presented in the previous sections with some theoretical explanations in the literature, testing these hypotheses against the data to shed light on the mechanisms driving these facts.

4.1 Main Hypothesis: Return-(unemployment) risk trade-off

This theory states that workers in riskier jobs should be compensated with higher static and lifetime labor earnings. This idea was first introduced by Smith (1776) in "The Wealth of Nations" as a compensating differentials explanation to wage dispersion. More recently, Cubas and Silos (2017) found a positive correlation at the occupation level between earnings and occupation risk and average wages. Dillon (2018) identifies an important role of earnings and employment risk in sorting workers across occupations: more risk-averse workers sort into occupations that entitle less risk and are willing to give up a sizable amount of lifetime earnings to reduce the uncertainty surrounding their career.

We first test this hypothesis on raw aggregate evidence. We find higher unemployment risk

[17] We have also split the sample between non-college and college grads, finding non-significant results with large standard errors for the self-employed. These results are available upon request.

in paid employment relative to self-employment. This result is remarkably accurate compared to both kinds of labor contracts regarding job protection. In the left panel of Figure 2, we consider monthly transitions between paid employment —differentiating between permanent and fixed-term employment—, self-employment, and unemployment. We show that the self-employed have the lowest probability of entering unemployment among the employed, and also exhibit higher status persistence.[18] This contrasts with the separation rate of fixed-term workers, who enter unemployment at a monthly rate above 5 percent. This simple matrix illustrates the heterogeneity hidden higher unemployment risk faced by the paid-employed in a dual labor market: it is mainly driven by the high turnover of fixed-term employment, whereas stable, permanent employment and self-employment offer higher job stability.

We further illustrate the unemployment risk faced across different employment statuses by examining the tenure distributions of workers. In the right panel of Figure 2, we observe that the self-employed tenure distribution has a higher right mass compared to both types of paid employment. As fixed-term employment has a legal maximum duration of about two years (with some exceptions), most workers in this type of contract enjoy spells that last less than a year. On the other hand, there is a considerable mass of self-employed whose businesses last for more than ten years.[19] Overall, we find, in contrast to studies in labor markets with lower employment duality and unemployment risk, that self-employment is a long-lasting employment option, insuring workers against higher unemployment risk present in paid employment.

We further investigate the role of composition effects in driving these aggregate results. Garcia-Cabo and Madera (2019), in a companion policy paper, calculate survival rates into unemployment from self-employment, confirming the robustness of the aggregate evidence presented in this section after controlling for workers' characteristics.[20] We next turn to an analysis of lifetime earnings profiles to test the hypothesis of whether those predominantly in self-employment face lower id-

[18] While this paper is one of the first to document the transition matrix for both types of salaried employment and self-employment in Spain, previous studies have shown lower separation rates from self-employment into unemployment: Millán Tapia (2012) finds lower hazard rates for the self-employed in European countries. Kredler, Millán Tapia and Visschers (2015) and Hincapié (2020) for the United States, and Herreño and Ocampo (2023) for Mexico also find the self-employed are less likely to fall back into unemployment.

[19] Survival in self-employment in Spain is higher than in the United States: using NLSY79 data, Manso (2016) shows that about 52% of the self-employment spells last less than two years and only around 12 percent of the self-employment spells last more than ten years.

[20] They present the following takeaways: 1) The probability of entering unemployment for the average worker is higher from paid employment than from self-employment. 2) This probability of entering unemployment is higher for women, the young, and the old. 3) Higher-educated workers enjoy longer spells.

From	To E(P)	E(T)	SE	NE
E(P)	98.76	0.29	0.05	0.89
E(T)	2.24	92.22	0.11	5.43
SE	0.1	0.13	99.22	0.55
NE	1.65	4.56	0.63	93.16

Figure 2: Monthly transition probabilities (left) and tenure in years by job status (right)

Note to Figure 2: **Left panel**: Share of workers in status x last month that transitioned to status x' this month, where x are rows and x' are columns. **E(P)** denotes paid employment in a permanent contract, **E(T)** denotes paid employment in a temporary contract, **SE** denotes self-employment, and **NE** denotes unemployment with benefits and non-employment spells. **Right panel**: Distribution of tenures in current contract.

iosyncratic lifetime income risk.

Finally, unemployment risk is not the only source of risk self-employed workers face. The risk associated with the initial investment and the degree to which households are personally liable are important sources of risk that we do not include in our analysis. Specific data on the size of these investments is largely unavailable. Using the EFF, we have instead looked into the size of the businesses, the size of the losses, and to what extent households are likely to be liable for those losses. We calculate these numbers in the 2017 wave[21], which is a year without any significant business-cycle event. The average total value of businesses is 153,979 euros, and only 12.63% of households report having any personal assets involved as collateral in case of business failure. In terms of flows, the average profit is 5,150 euros, while the average total value of losses is only 746. The losses are highly skewed: less than 5 percent of households with businesses report losses. Finally, the average monthly payment related to business debt is 642 euros. To compare to aggregate statistics, the National Study of the Sole Proprietor (Estudio Nacional del Autńomo), reports that only a third of the sole proprietors rely on external financing. We conclude that, while this source of risk seems important ex-ante, it is likely to affect only a small fraction of the self-employed in our context.

[21]The data on the details of individual businesses is available in Part 4 "Negocios y Activos Financieros" of the EFF. We describe the specific content of the questions in Appendix C.2.

4.2 Earnings Risk Dynamics and Entry into Self-Employment over the Life Cycle

To further test our main hypothesis—that self-employment is the least risky option in terms of earnings uncertainty over the life cycle, making it desirable despite the lower returns—we decompose the dynamics of earnings into ex-ante individual heterogeneity, persistent uncertainty, and transitory uncertainty. For this purpose, we estimate a workhorse model of earnings dynamics over the life cycle. Importantly, we allow for the parameters of the statistical model to vary by age, following Karahan and Ozkan (2013).[22] While we do not explicitly model occupational choices, our categorization of workers can be interpreted through the lens of a search model that incorporates on the job search and quits. In such model, workers contact with employers that offer contracts with different job security, can decide to turn down offers and continue unemployed, or can become self-employed. Our paper's focus on the dynamics precisely captures how early-life occupational decisions (as inferred from our definition of main attachment at younger ages) are influenced by the anticipated risk and returns over the life cycle. Next, we outline the specific methodological steps.

In particular, we first calculate the residual earnings after controlling for observed heterogeneity. We then decompose residual earnings into an individual fixed effect that can be interpreted as unobserved ex-ante heterogeneity in workers' ability and a permanent and transitory shock. We can think of the latter part as uncertainty or earnings risk, which is the object of interest. We then compare the persistence and variance of these shocks across groups based on labor status. To exploit the panel structure, we assign a lifetime job status to each person based on the predominant occupation before 40 years old. Formally, we estimate the following equation from earnings data:

$$\tilde{y}_h^i = \alpha^i + z_h^i + \epsilon_h^i$$

$$z_h^i = \rho_{h-1} z_{h-1}^i + \eta_h^i, \tag{2}$$

where $\tilde{y}_h^i$ denote residual earnings for worker i at age h and time t after controlling for observ-

[22] Additionally, and as shown in Appendix B, we allow the parameters to vary by age and educational attainment, as in Blundell, Graber and Mogstad (2015). While the implications derived from the results for the different groups in the pooled sample in this section go through even when allowing for skill heterogeneity in the estimation, the point estimates from the pooled sample are more precise due to the larger sample size. We verify some of the findings by educational attainment as in Blundell, Graber and Mogstad (2015): we observe that college graduates have larger variances for both shocks at the beginning of their working-lives compared to non-college graduates, and both variances decrease over time. On the other hand, the variances of the non-college are U-shaped for the persistent shock for the predominantly temporary– and are higher later in life compared to the college workers.

ables.[23] We decompose the residual $\tilde{y}^i_h$ as the sum of a fixed effect α^i, a persistent z^i_h component, and a temporary component ϵ^i_h. We further assume that the persistent component follows an AR(1) process with auto-correlation parameter ρ and variance σ^2_η and captures long-lasting changes in earnings. The transitory shock has variance σ^2_ϵ and captures measurement error and temporary changes in annual earnings. We refer to Karahan and Ozkan (2013) for further details on the specification and identification of the parameters.[24]

We estimate α, ρ^h, $\sigma^2_{\eta_h}$, and $\sigma^2_{\epsilon_h}$ using Generalized Method of Moments. In particular, we minimize the distance between empirical variances and co-variances from the data by age, $cov(\tilde{y}^i_h, \tilde{y}^i_{h+n})$, and the theoretical counterparts derived from the model summarized by equation 2. We target a non-parametric specification, without imposing a specific functional form in the earnings process. Overall, we have 196 moments to estimate 93 parameters for each group. We present the results in Figure 3, highlighting the importance of considering age variant profiles when studying earnings dynamics.

We observe that, at early ages, shocks are moderately persistent for all groups, as previously found for the United States by Karahan and Ozkan (2013), but especially for the self-employed. In order to interpret persistence estimates, we compare the number of years that a shock received at different ages takes to fade away. If a shock is received at age 30, 64 and 73 percent of its effect dissipates within five years for those mainly in permanent and temporary employment, respectively. The shock is less persistent for those mainly self-employed, as 90 percent of a shock received at age 30 dissipates within five years. Persistence increases as workers age. For instance, if the shock is received at age 40, only about 44 percent fades away after five years for permanent and self-employed workers, and 38 percent for those mainly in temporary employment. Persistence at later ages stabilizes for permanent and self-employed workers but keeps increasing until late in life for those in temporary contracts, reaching its peak close to 50. We next turn our attention to the variances.

In the case of persistent shocks, we observe that the variance declines between ages 25 and 35 and plateaus afterward for all workers, with a small but not significant increase towards the end

[23]Specifically, we run a first stage $\log Y^i_{h,t} = \beta X^i_{h,t} + \tilde{y}^i_h$, where $X^i_{h,t}$ contains a quartic polynomial in age, educational attainment, worker's region fixed-effects, and year dummies. The first stage is run separately for each employment group. We also consider a joint first stage to ensure the differential impact of aggregate shocks is not residualized and affects our results. Results for the common first stage are very similar and are available upon request.

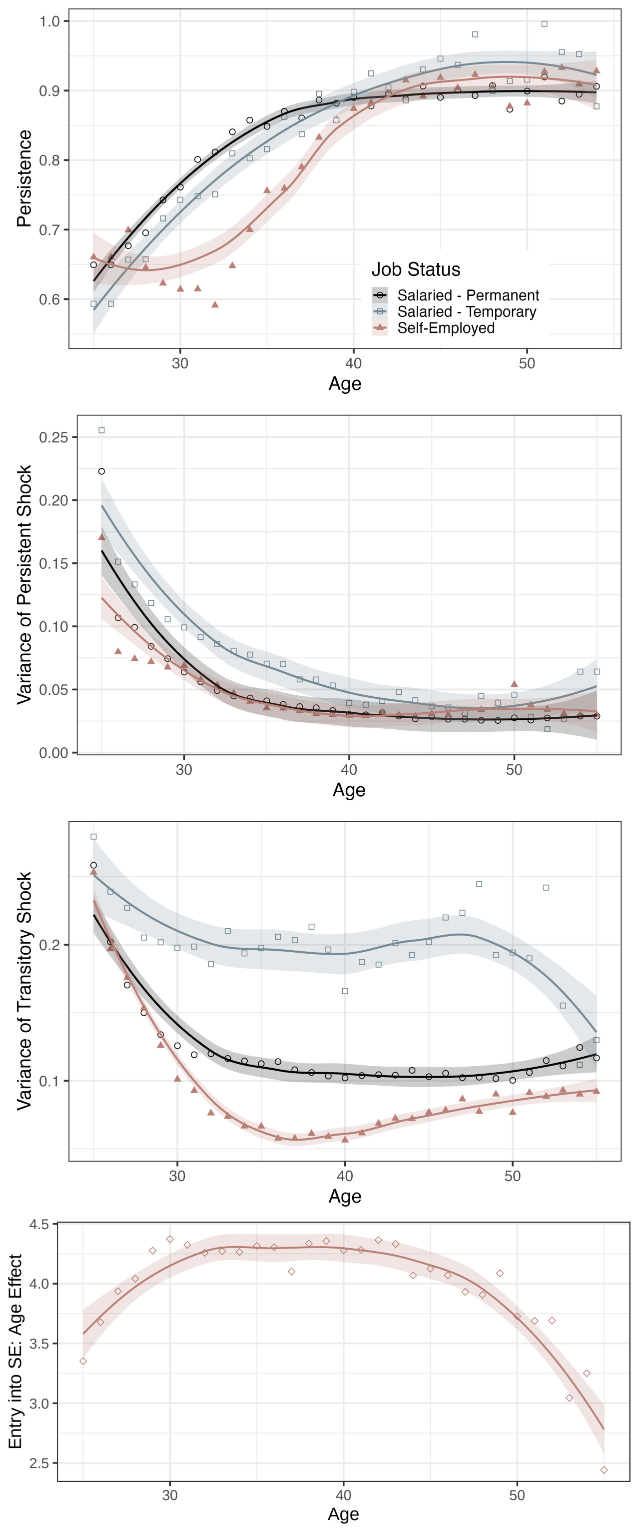
[24]We follow their identification strategy and normalize $\rho_1 = \rho_2$ and $\sigma^2_{\eta_H} = \sigma^2_{\eta_{H-1}}$. In our setting, we abstract from the time-varying loading factors and capture the aggregate dynamics with year dummies.

for those in temporary employment. Overall, this variance is higher over the lifetime for workers employed predominantly in temporary jobs than for the other groups, suggesting more variability across workers in their labor market outcomes.

The variance of temporary shocks is three times as high for workers mainly in temporary contracts compared to those mainly in self-employment. This finding is not surprising, given the high turnover in these jobs (at most two years according to the legislation) compared to the seven years of average tenure in self-employment. Moreover, despite the stability of highly protected jobs, the variance of transitory shocks is also higher than for those in self-employment.

It remains to be shown that there is a link between these uncertainty patterns and entry into self-employment. While a causal study is beyond the objective of this paper, we do include in the fourth panel of Figure 3 the estimated age effects of entering self-employment from any alternative job status. We control for cohort and age effects, following the restricted year effects approach from Deaton and Paxson (1994). We then rescale the fixed effects to the value of the omitted category (age = 26). We can thus interpret the resulting profile as the average entry rate into self-employment, net of cohort and year dynamics. Evans and Leighton (1989) document that entry into self-employment in the United States is flat in age for workers older than 26. More recently, Hincapié (2020) documents that entry occurs well into workers' 30s. Somehow different from the United States, we document that entry is hump-shaped, with the increasing years coinciding with those that were persistent, and the variance of shocks is the lowest compared to salaried options, especially those in unstable salaried options.

Figure 3: Earnings profiles: persistence and variances



Note to Figure 3: Markers denote point estimates in non-parametric regression using dummies. Smoothed profiles (solid lines) are calculated using LOWESS regressions, with bandwidth set to 0.8.

4.3 Alternative Hypotheses

4.3.1 Measurement Issues and Non-Pecuniary Benefits

Many papers document the existence of large non-pecuniary benefits from self-employment (Hamilton, 2000) and of intangible capital such as sweat equity (Bhandari and McGrattan, 2020), to reconcile the observed transitions into self-employment. We do not rule out this possibility, as we cannot identify this channel in our dataset that focuses on the labor side of self-employment. Instead, we focus on mechanisms we can test and leave the interpretation of the results as possibly in addition to non-pecuniary benefits. More broadly, past literature has documented that business owners and the self-employed tend to mis- or under-report their income (Hurst, Li and Pugsley, 2014; Lagakos et al., 2018). Bhandari et al. (2020) document the significant differences in US business income from various surveys relative to tax records. We try to overcome these issues in two ways. First, our data comes from social security records, which improves the quality of survey data. As argued earlier, it might still be the case that the self-employed under-report their income to the SSA, although notice how this would reduce future benefits and pensions. For this purpose, we reassessed the robustness of our results in section 3 using survey data. Second, we further test our results by focusing on salaried income after a self-employment spell. The thin lines in the bottom panel of Figure 1 limit the sample to those that, after the age of 40, only hold salaried jobs. We show that our result that the young-self-employed are stuck on slower earnings growth compared to both salaried markets still holds.

4.3.2 Negative selection in self-employment

Our observed results that the self-employed earn on average less than the salaried could be explained by negative selection into self-employment, both by observed and unobserved characteristics. Evans and Leighton (1989) document that unemployed and workers in the lower tail of the salaried employment distribution are more likely to enter self-employment. More recently, Humphries (2018) shows that non-incorporated self-employed exhibit lower cognitive and non-cognitive ability than incorporated entrepreneurs. We find evidence of negative selection in Table 1. The self-employed have, on average, lower educational attainment than salaried workers and are mainly composed of sole proprietors and cooperative workers, so one could argue that the flatter age-earnings profiles presented in the bottom panel of Figure 1 are driven by the composition of the workforce. In order to further inspect the role of unobserved heterogeneity, the longitudinal analysis in section 3.3 estimates equation (1), controlling for individual fixed-effects and observed characteristics of workers, achieving similar conclusions. Moreover, throughout this paper, we have conditioned our analysis on educational attainment to control for differences in composition across occupational groups that could affect the results, achieving similar conclusions. Nonetheless, we recognize the inherent limitations of our study in addressing negative selection into self-employment, especially when taking into account characteristics that drive a worker into self-employment and are unobserved by the econometrician.

4.3.3 Search frictions and earnings growth through job shopping

An important strand of the literature has studied the contribution of job shopping to earnings and wage growth (understood as hourly or daily earnings). Ex-ante identical workers receive different employment opportunities to bargain over higher wages, generating wage growth and dispersion (see, for instance, Postel-Vinay and Robin (2002); Bagger et al. (2014)). The evidence presented in this paper suggests that we can rule out the existence of job shopping once in self-employment, as we observe low turnover and earnings growth for the self-employed. The innate search frictions in dual labor markets with high firing costs could explain why we do not observe workers in short-term self-employment spells frequently transitioning back to paid employment. In labor markets with relatively lower search frictions, such as the United States, this might not occur. Indeed, the literature has documented for the United States that many workers use self-employment as experimentation, where the acquired experience is later compensated with a labor market premium (Daly, 2015; Manso, 2016).

5 Discussion and Conclusion

This paper documents the role of unemployment and earnings risk in reconciling evidence in payoff differentials between self-employment and paid-employment in labor markets with high unemployment and turnover. We use a large longitudinal dataset from the Spanish social security records to shed light on the return-risk trade-off in life-cycle earnings, both in self-employment and the salaried alternatives. We compare the cross-sectional distributions and life-cycle of earnings for (log) levels and the risk components of earnings. We conclude that, because of riskier forms of paid employment, such as the case of temporary contracts in Spain, earnings risk in self-employment is not necessarily higher than in paid employment, as often assumed in the notion of entrepreneurial risk.

The evidence and estimates presented in this paper can be used to discipline structural models to study policy reforms aiming to reduce unemployment and increase job stability. We show that differential unemployment risk within paid employment becomes a relevant margin when workers decide among employment alternatives. It is essential to account for this channel to avoid overstating the role of entrepreneurial risk and requiring considerable non-pecuniary benefits to reconcile labor market transitions. Reducing unemployment risk can help to alleviate labor market outcomes for those groups of workers that traditionally face high unemployment rates and unstable employment. However, it is necessary to consider the role of negative selection in self-employment, which is not resolved in this paper. However, it is of great research importance: the outcomes described in this paper acknowledge that the median self-employed is different from the typical entrepreneur, and hence, we should not expect them to be engines of growth and employment creation. A welfare analysis, beyond the scope of this paper, is desired to assess the costs and benefits of government intervention.

Finally, while we performed the analysis for Spain, our conclusions are not restricted to Southern Europe: the rise of the gig economy has segmented labor markets into high- and low-earnings volatility jobs worldwide.

References

Bagger, Jesper, Francois Fontaine, Fabien Postel-Vinay and Jean-Marc Robin. 2014. "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics." *American Economic Review* 104(6):1551–96.

Bhandari, Anmol and Ellen R. McGrattan. 2020. "Sweat Equity in U.S. Private Business." *The Quarterly Journal of Economics* .

Bhandari, Anmol, Serdar Birinci, Ellen R. McGrattan and Kurt See. 2020. "What Do Survey Data Tell Us about US Businesses?" *American Economic Review: Insights* 2(4):443–58.

Blundell, Richard, Michael Graber and Magne Mogstad. 2015. "Labor income dynamics and the insurance from taxes, transfers, and the family." *Journal of Public Economics* 127:58–73. The Nordic Model.
URL: *https://www.sciencedirect.com/science/article/pii/S0047272714000930*

Bonhomme, Stéphane and Laura Hospido. 2016. "The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data." *The Economic Journal* 127(603):1244–1278.

Buchinsky, Moshe, Denis Fougère, Francis Kramarz and Rusty Tchernis. 2010. "Interfirm Mobility, Wages and the Returns to Seniority and Experience in the United States." *The Review of Economic Studies* 77(3):972–1001.

Cabrales, Antonio, Maia Güell, Rocio Madera and Analía Viola. 2020. "Income contingent university loans: Policy design and an application to Spain." *Economic Policy* 34(99):479–521.
URL: *https://doi.org/10.1093/epolic/eiz010*

Cuadrado, Pilar, Pablo Hernández de Cos and Mario Izquierdo. 2011. "El ajuste de los salarios frente a las perturbaciones de España." *Economic Bulletin Banco de España* 2.

Cubas, German and Pedro Silos. 2017. "Career choice and the risk premium in the labor market." *Review of Economic Dynamics* 26:1–18.

Daly, Moira. 2015. "The long term returns of attempting self-employment with regular employment as a fall back option." *Labour Economics* 35:26–52.

de la Roca, Jorge. 2014. "Wage cyclicality: Evidence from Spain using social security data." *SERIEs* 5:173–195.

Deaton, Angus and Christina Paxson. 1994. "Intertemporal Choice and Inequality." *Journal of Political Economy* 102(3):437–67.
URL: *http://ideas.repec.org/a/ucp/jpolec/v102y1994i3p437-67.html*

Dillon, Eleanor and Christopher Stanton. 2013. "Self-Employment Dynamics and the Returns to Entrepreneurship." *NBER Working Paper No 23168* pp. 4212–4225.

Dillon, Eleanor W. 2018. "Risk and Return Trade-Offs in Lifetime Earnings." *Journal of Labor Economics* 36(4):981–1021.
URL: *https://doi.org/10.1086/697475*

Evans, David S. and Linda S. Leighton. 1989. "Some Empirical Aspects of Entrepreneurship." *American Economic Review* 79(3):519–535.

Feng, Shuaizhang and Yingyao Hu. 2013. "Misclassification Errors and the Underestimation of the US Unemployment Rate." *American Economic Review* 103(2):1054–70.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.103.2.1054*

Garcia-Cabo, Joaquin and Rocio Madera. 2019. "The Self-Employment Option in Rigid Labor Markets: An Empirical Investigation." *International Finance Discussion Papers, Board of Governors of the Federal Reserve System* (1264).

Garcia-Perez, Jose-Ignacio. 2008. "La muestra continua de vidas laborales: Una guía de uso para el análisis de transiciones." *Revista de Economia Aplicada* 26:5–28.

Güell, Maia and Barbara Petrongolo. 2007. "How binding are legal limits? Transitions from temporary to permanent work in Spain." *Labour Economics* 14:153–183.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan and Jae Song. 2021. "What Do Data on Millions of U.S. Workers Reveal about Lifecycle Earnings Dynamics?" *Econometrica* 89:2303–2339.

Hamilton, Barton H. 2000. "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment." *Washington University* 108(3):604–631.

Herreño, Juan and Sergio Ocampo. 2023. "The macroeconomic consequences of subsistence self-employment." *Journal of Monetary Economics* 136:91–106.
URL: *https://www.sciencedirect.com/science/article/pii/S0304393223000193*

Hincapié, Andrés. 2020. "Entrepreneurship Over The Life Cycle: Where Are The Young Entrepreneurs?" *International Economic Review* 61(2):617–681.

Hombert, Johan, David Sraer, Antoinette Schoar and David Thesmar. 2014. "Can Unemployment Insurance Spur Entrepreneurial Activity?" *NBER Working Paper No 20717* .

Humphries, John Eric. 2018. "The Causes and Consequences of Self-Employment over the Life Cycle." *Job Market Paper* .

Hurst, Erik, Geng Li and Ben Pugsley. 2014. "Are Household Surveys Like TaxForms? Evidence from Income Underreporting of the Self Employed." *Review of Economics and Statistics* 96(1):19–33.

Karahan, Fatih and Serdar Ozkan. 2013. "On the persistence of income shocks over the life cycle: Evidence, theory, and implications." *Review of Economic Dynamics* 16(3):452–476.

Kredler, Matthias, Ana Millán Tapia and Ludo Visschers. 2015. "Self-employment, paid-employment and unemployment over the business cycle." *mimeo* .

Lagakos, David, Benjamin Moll, Tommaso Porzio, Nancy Qian and Todd Schoellman. 2018. "Life Cycle Wage Growth across Countries." *Journal of Political Economy* 126(2):797–849.

Manso, Gustavo. 2016. "Experimentation and the Returns to Entrepreneurship." *Review of Financial Studies* 29:2319–2340.

Martinez-Lopez, Diego. 2013. "The underreporting of income by self-employed workers in Spain." *SERIEs* 4:353–371.

Millán Tapia, Ana. 2012. "The choice to become self-employed: Acknowledging frictions." *mimeo* .

Moskowitz, Tobias J. and Annette Vissing-Jørgensen. 2002. "The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle?" *American Economic Review* 92(4):745–778.
URL: *https://www.aeaweb.org/articles?id=10.1257/00028280260344452*

Postel-Vinay, Fabien and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70(6):2295–2350.

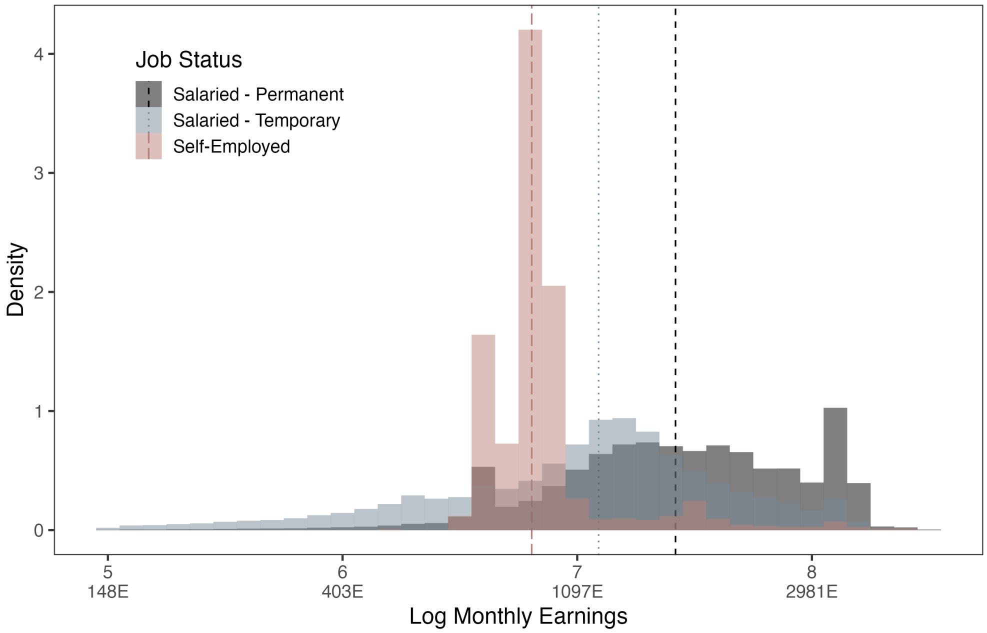
Smith, Adam. 1776. *The Wealth of Nations*.

Topel, R. H. 1991. "Specific Capital, Mobility and Wages: Wages Rise with Job Seniority." *Journal of Political Economy* 99(1):145–176.

A Robustness to Underreporting Self-Employment Earnings

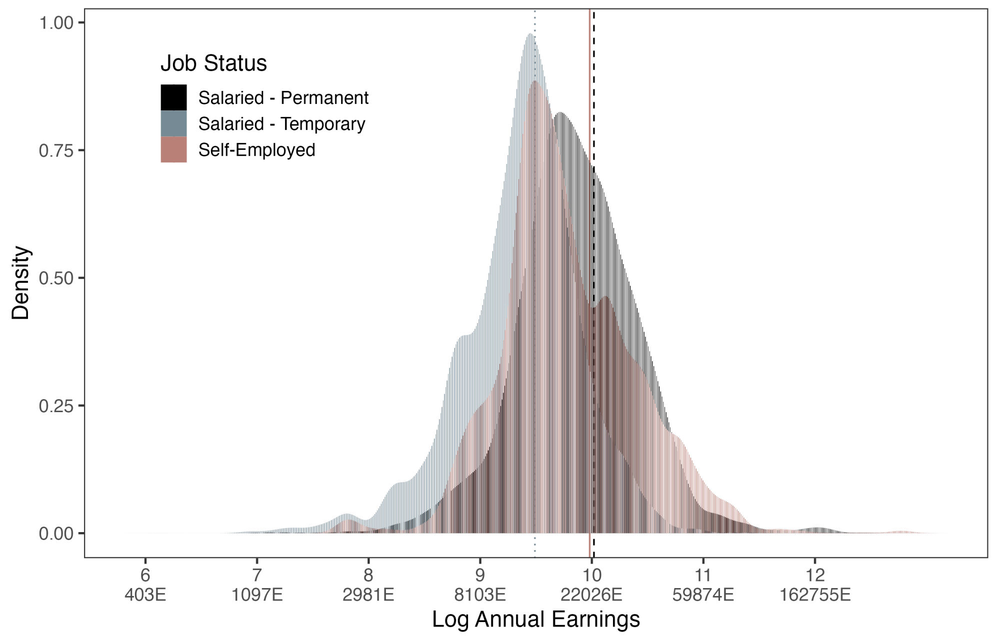
A.1 Comparing self-employment income across datasets

Figure A.1: Monthly log-Earnings Distribution By Employment Status



Note: Histogram and Kernel density estimates for the cross-sectional distribution of annual real earnings (in 2016 Euros). Vertical lines denote the median for each distribution of the same color. x-axis includes the corresponding euro amount for ease of interpretation. Source: MCVL.

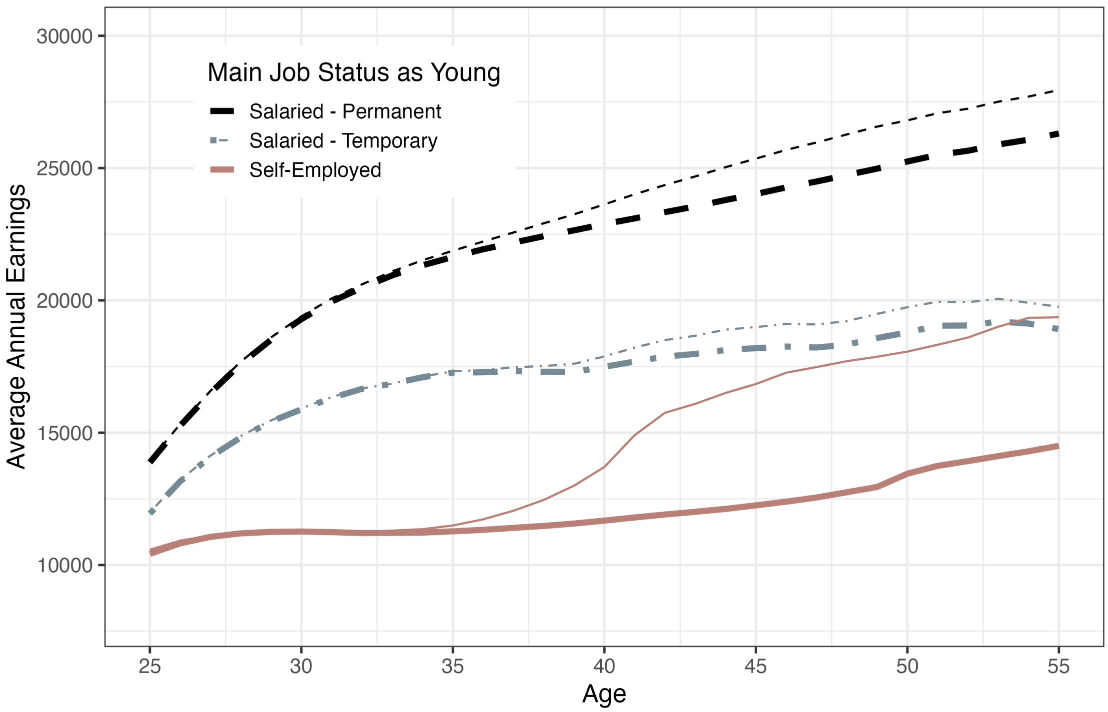
Figure A.2: Annual Log-Earnings Distribution By Employment Status: Survey data



Note: Histogram and Kernel density estimates for the cross-sectional distribution of annual real earnings using survey responses (in 2016 Euros). Vertical lines denote the median for each distribution of the same color. x-axis includes the corresponding euro amount for ease of interpretation. Source: EFF.

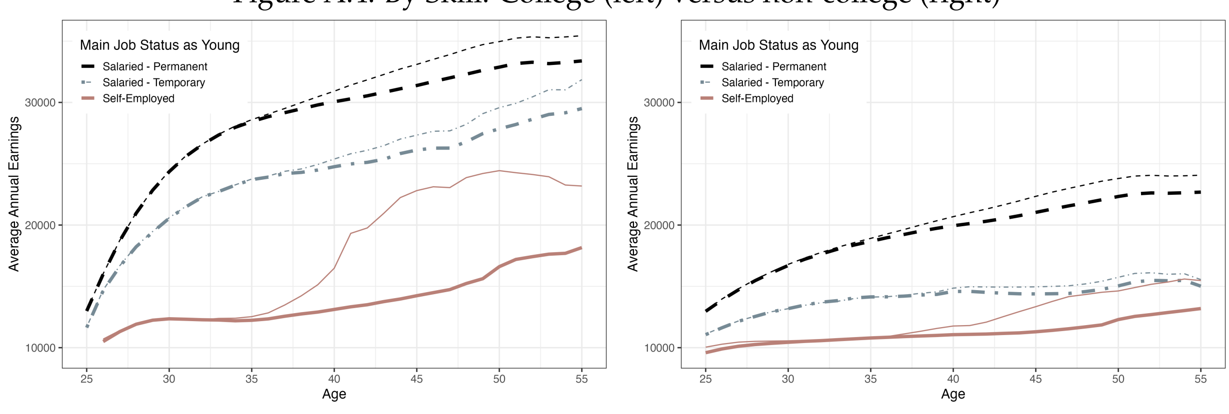
A.2 Excluding Zeros and Skill Heterogeneity

Figure A.3: Excluding Zeros



Note to Figure A.3: Thick lines are average earnings, excluding zeros, at each age for all workers of a given age and with the given main status. Thin lines further limit the sample to those that, independently of their status before 40, are salaried after 40. Main job status as young is defined as predominant status before 40 by number of days.

Figure A.4: By Skill: College (left) versus non-college (right)



Note to Figure A.4: Same as Figure 1 further decomposing by whether workers attained a college degree.

A.3 Labor market returns and skill heterogeneity

	Years of experience			
Young Status	**2**	**5**	**10**	**15**
Predominantly temporary before age 40	0.26	0.49	0.63	0.67
Predominantly self-employed before age 40*	0.13	0.25	0.35	0.39
Predominantly self-employed before age 40**	0.19	0.35	0.44	0.49

*all workers, **only those always in salaried work after 40.

Table A.1: Estimated cumulative returns to experience: Non-college graduates

	Years of experience			
Young Status	**2**	**5**	**10**	**15**
Predominantly temporary before age 40	0.42	0.77	0.95	0.99
Predominantly self-employed before age 40*	0.21	0.39	0.48	0.51
Predominantly self-employed before age 40**	0.29	0.53	0.67	0.80

*all workers, **only those always in salaried work after 40.

Table A.2: Estimated cumulative returns to experience: College graduates

Table A.3: Mincer Equation of Earnings: Full Estimates

	Predominant Status Before 40			
	P	**T**	**SE**	**SE***
Total Experience	0.149	0.187	0.086	0.127
	(0.001)	(0.002)	(0.001)	(0.003)
Total Experience^2/100	-1.143	-1.740	-0.731	-1.215
	(0.009)	(0.024)	(0.015)	(0.035)
Total Experience^3/1000	0.391	0.710	0.284	0.523
	(0.004)	(0.013)	(0.007)	(0.016)
Total Experience^4/10000	-0.047	-0.100	-0.037	-0.075
	(0.001)	(0.002)	(0.001)	(0.002)
Real Tenure	0.143	0.309	0.115	0.231
	(0.001)	(0.006)	(0.001)	(0.011)
Real Tenure^2/100	-1.720	-5.338	-1.496	-4.112
	(0.014)	(0.193)	(0.016)	(0.322)
Real Tenure^3/1000	0.774	3.278	0.685	2.524
	(0.009)	(0.182)	(0.009)	(0.307)
Real Tenure^4/10000	-0.113	-0.631	-0.101	-0.482
	(0.002)	(0.049)	(0.002)	(0.090)
Constant	8.823	8.347	8.553	8.418
	(0.007)	(0.016)	(0.011)	(0.027)
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
N	6122764	931608	1531018	369302

Note to Table A.3: SE* denotes self-employed that transitioned into salaried job and never went back to self-employment after 40. Removing industry fixed effects expands the sample but the coefficients and conclusions do not change in any substantive way.

Table A.4: Mincer Equation of Earnings: Non-college graduates

	Predominant Status Before 40			
	P	**T**	**SE**	**SE***
Total Experience	0.119	0.157	0.0769	0.113
	(0.001)	(0.002)	(0.001)	(0.003)
Total Experience^2/100	-0.861	-1.450	-0.633	-1.086
	(0.010)	(0.027)	(0.0161)	(0.038)
Total Experience^3/1000	0.283	0.593	0.241	0.467
	(0.004)	(0.014)	(0.007)	(0.017)
Total Experience^4/10000	-0.033	-0.083	-0.030	-0.066
	(0.001)	(0.002)	(0.001)	(0.003)
Real Tenure	0.138	0.305	0.115	0.230
	(0.001)	(0.006)	(0.001)	(0.010)
Real Tenure^2/100	-1.626	-5.133	-1.469	-3.966
	(0.009)	(0.187)	(0.016)	(0.291)
Real Tenure^3/1000	0.722	3.097	0.664	2.389
	(0.006)	(0.173)	(0.009)	(0.276)
Real Tenure^4/10000	-0.104	-0.586	-0.0974	-0.448
	(0.001)	(0.046)	(0.0015)	(0.080)
Constant	8.815	8.307	8.546	8.416
	(0.008)	(0.017)	(0.011)	(0.028)
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
N	4783688	676884	1325175	301409

Note to Table A.4: SE* denotes self-employed that transitioned into salaried job and never went back to self-employment after 40.

Table A.5: Mincer Equation of Earnings: College graduates

	Predominant Status Before 40			
	P	**T**	**SE**	**SE***
Total Experience	0.202	0.257	0.131	0.183
	(0.002)	(0.004)	(0.004)	(0.008)
Total Experience^2/100	-1.735	-2.575	-1.330	-1.976
	(0.021)	(0.059)	(0.053)	(0.101)
Total Experience^3/1000	0.635	1.119	0.581	0.974
	(0.010)	(0.034)	(0.026)	(0.055)
Total Experience^4/10000	-0.082	-0.167	-0.083	-0.156
	(0.001)	(0.007)	(0.004)	(0.010)
Real Tenure	0.159	0.336	0.130	0.268
	(0.006)	(0.005)	(0.004)	(0.014)
Real Tenure^2/100	-2.116	-6.662	-2.029	-5.914
	(0.138)	(0.174)	(0.062)	(0.425)
Real Tenure^3/1000	1.016	4.585	1.061	4.284
	(0.093)	(0.174)	(0.038)	(0.436)
Real Tenure^4/10000	-0.156	-0.993	-0.175	-0.962
	(0.019)	(0.051)	(0.008)	(0.136)
Constant	9.093	8.684	8.665	8.513
	(0.024)	(0.048)	(0.043)	(0.080)
Industry FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
N	1286449	239916	189337	63376

Note to Table A.5: SE* denotes self-employed that transitioned into salaried job and never went back to self-employment after 40.

In the EFF

Table A.6: Mincer Equation of Earnings: Full EFF Sample

	E	SE
Total Experience	0.0780	0.0364
	(0.0114)	(0.0352)
Total Experience^2/100	-0.301	0.128
	(0.0942)	(0.239)
Total Experience^3/1000	0.0492	-0.0712
	(0.0292)	(0.0607)
Total Experience^4/10000	-0.00245	0.00726
	(0.00300)	(0.00493)
Real Tenure	0.182	-0.0445
	(0.00868)	(0.0413)
Real Tenure^2/100	-1.588	0.886
	(0.122)	(0.603)
Real Tenure^3/1000	0.607	-0.420
	(0.0568)	(0.287)
Real Tenure^4/10000	-0.0798	0.0606
	(0.00838)	(0.0432)
Constant	8.651	9.405
	(0.0423)	(0.166)
Year FE	Y	Y
N	9375	1714

A.4 Labor Histories by Young Attachment

	Share of Days in Each Kind of Contract				
	P	T	SE	NE	All
By Main Status When *Young* (up to 40)					
P					
When *Young*	**77.76%**	8.92%	2.24%	11.08%	100%
When *Old*	**64.13%**	7.31%	13.58%	14.99%	100%
In Lifetime	**67.19%**	9.46%	9.05%	14.30%	100%
T					
When *Young*	14.41%	**62.70%**	2.18%	20.72%	100%
When *Old*	**39.03%**	32.94%	6.72%	21.31%	100%
In Lifetime	18.07%	**57.64%**	2.86%	21.43%	100%
SE					
When *Young*	9.25%	4.66%	**75.68%**	10.41%	100%
When *Old*	11.09%	4.80%	**74.42%**	9.68%	100%
In Lifetime	10.15%	5.40%	**73.53%**	10.92%	100%
NE					
When *Young*	12.98%	14.49%	4.17%	**68.36%**	100%
When *Old*	30.32%	15.78%	14.13%	**39.76%**	100%
In Lifetime	16.89%	15.85%	6.46%	**60.79%**	100%

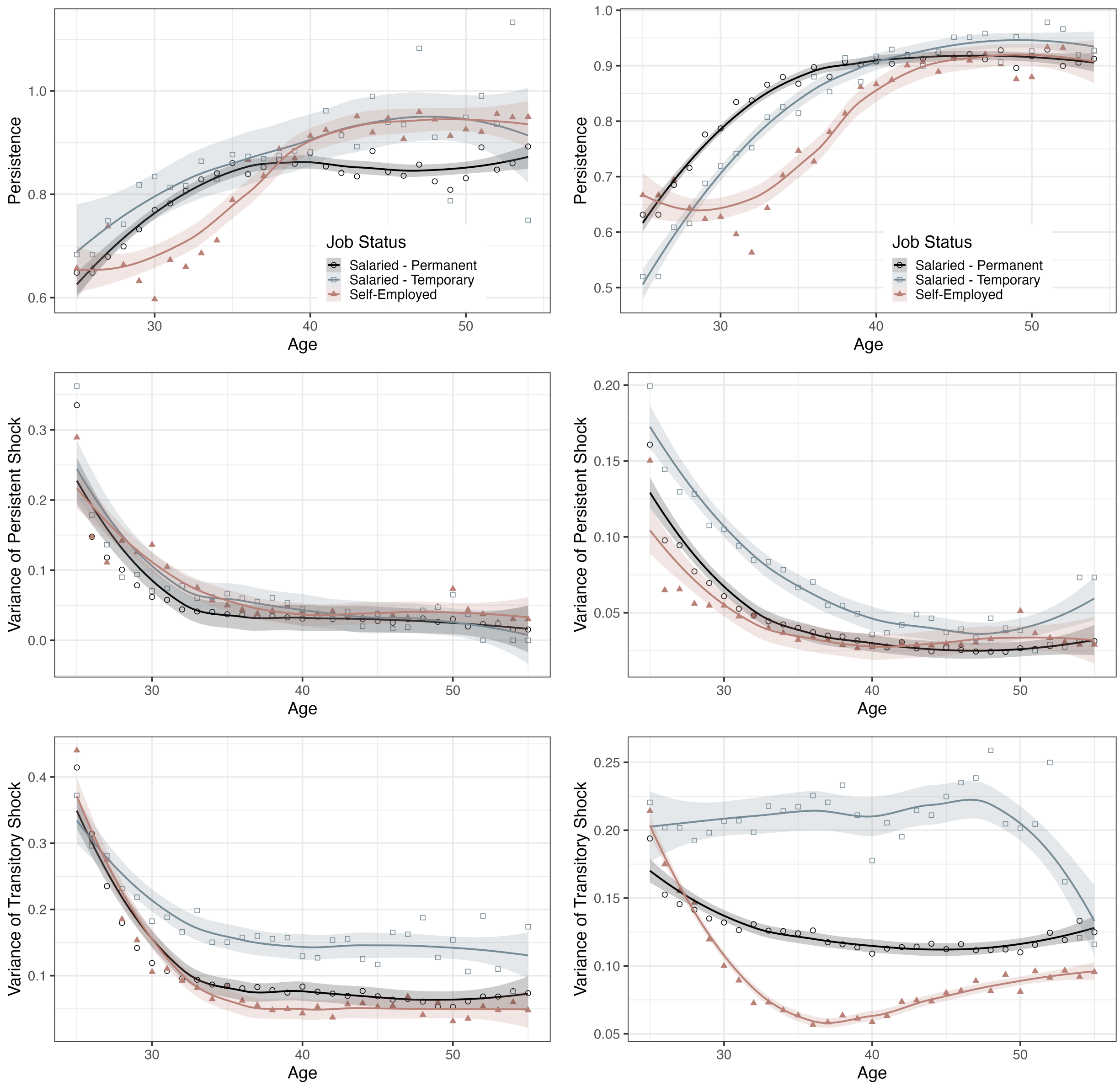
Table A.7: Labor Histories by Young Attachment

Note to Table A.7: By main job status as young, share of days worked in each type of contract or occupation: (P) Salaried - Permanent Contract, (T) Salaried - Temporary Contract, (SE) Self-Employed, (NE) Non-employed. Main jobs is defined as predominant status before 40 by share of days worked by 40, as confirmed by the first row in each main status.

B Earnings Risk Dynamics: Allowing for Skill Heterogeneity

Blundell, Graber and Mogstad (2015) show the existence of substantial mis-specification bias in the estimation of income processes, in particular when assuming age-independent profiles and when not allowing for heterogeneous profiles for different skill groups. Our main specification allows the persistence of the income process and the variance of the permanent and transitory shocks to vary over the life cycle, alleviating some of these concerns. Now, we extend our estimation by allowing income processes to differ by education level. In particular, we classify workers by skill level using education information in the sample. We estimate two different processes: one for those workers with a college degree or higher (*college*) and those without a college degree (*non-college*). Figure B.1 summarizes these results. While for the persistence parameter, we do not observe large differences by allowing for skill heterogeneity (except for a slightly lower persistence for those salaried workers predominantly in permanent contracts at young ages), the variances are somehow different for college and non-college workers. First, college workers have larger variances for both shocks at the beginning of their working lives (25 years old) than non-college workers. However, they both decrease over time, especially for those predominantly Permanent and Self-employed at young ages; they remain low and stable. The variances of the non-college are starkly different relative, which in the case of the persistent shock they are U-shaped for the predominantly temporary– and are higher later in life compared to the college workers. These results are in line with the findings of Blundell, Graber and Mogstad (2015) and are hidden in the pooled sample as college and non-college workers are all mixed in together.

Figure B.1: Earnings profiles: college (left) and non-college (right)



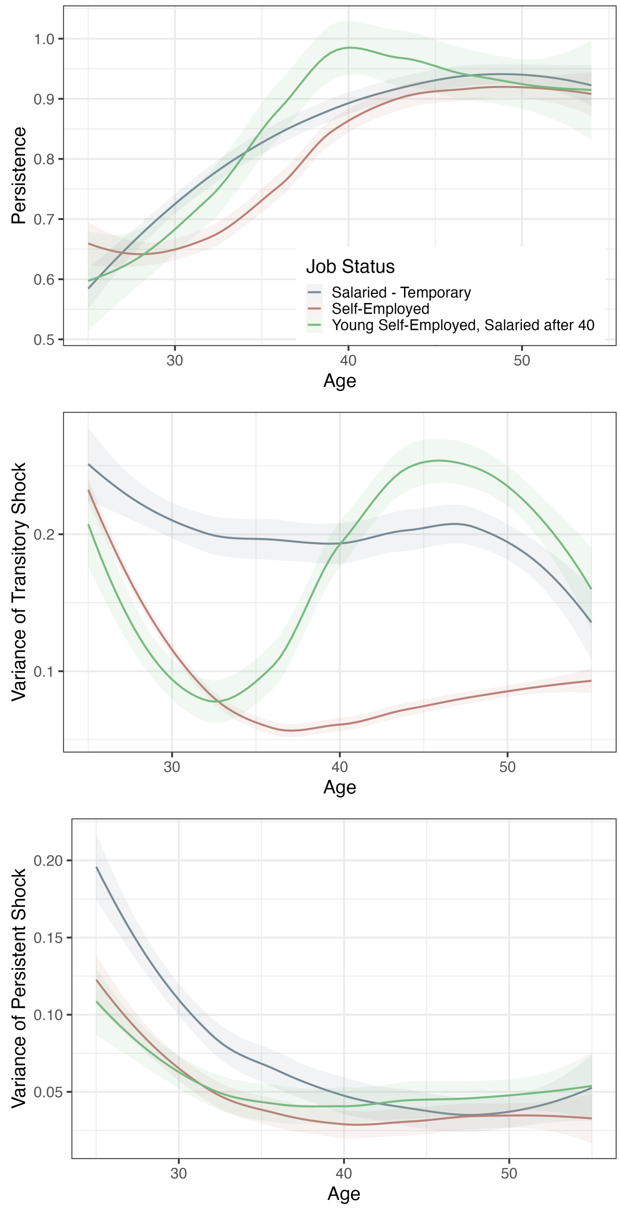
Note to Figure B.1: Markers denote point estimates in non-parametric regression using dummies. Smoothed profiles (solid lines) are calculated using LOWESS regressions, with bandwidth set to 0.8.

B.1 Earnings Risk Dynamics: Returning to salaried work after self-employment

In Section 3, we showed that workers predominantly in self-employment at young ages exhibited lower returns to experience compared to workers predominantly in temporary employment. As one could be concerned about under-reporting of the self-employed in their contribution bases, we also estimated these returns for those young self-employed who only have salaried employment after age 40. Our results suggested that self-employment experience did not have substantial returns even when returning to salaried employment. We now turn to understand why this could be the case.

We estimate the income profile process in Section 4 for the same group of workers: predominantly in self-employment before age 40, that return to salaried employment after. We present the results in Figure B.2. While the variance of the permanent shocks are very similar across groups, with those predominantly in temporary employment when young experiencing larger shocks earlier in life, the variance of transitory shocks is markedly different for those self-employment returning to salaried employment. In particular, at young ages, this variance decreases until their mid-30s, but as they start transitioning to salaried jobs, it increases markedly, and becomes quantitatively similar to those predominantly temporary. This suggests that upon returning to salaried job, they face larger unemployment risk, and more unstable income, one more piece of evidence that self-employment in the context of dual labor markets provides some insurance against unemployment risk.

Figure B.2: Earnings profiles: persistence and variances



Note to Figure B.2: Smoothed profiles (solid lines) are calculated using LOWESS regressions, with bandwidth set to 0.8.

C Data and Institutional Background

The variable description presented below was previously circulated in an extended version of this paper titled "The self-employment option in rigid labor markets: an empirical investigation."

C.1 The Spanish Social Security Administration Data

We use the Spanish Social Security Administration's (SSSA) *Muestra Continua de Vidas Laborales* (MCVL) data set. It consists of a 4% representative sample of Spanish individuals affiliated with the SSA in 2013-2016, whether employed, unemployed, or retired. Workers are added every year to maintain the representativeness of the sample, as workers who end their affiliation are removed. The sample size is about 1.2 million individuals per year. The underlying source is the actual legal contracts signed between firms and workers. We observe all the workers' demographics and daily job histories from the first day of affiliation until the last day of 2016. Detailed job information is available back to 1967, and reliable monthly earnings data (based on social security contributions) is available back to 1980.

Regarding the population and content of the data, the MCVL samples from individuals affiliated at least one day during the reference year of the wave, through a job, contributory pension, or subsidy. This sample selection effectively excludes individuals whose only connection to the SSSA is through public health insurance and those receiving noncontributory subsidies. It also excludes individuals without any connection to the SSSA and public servants included in a special set-to-expire contribution regime (MUFACE). The dataset contains monthly wage data from 1980 with an entry for each job spell the worker has experienced as a salaried or self-employed worker and each non-employment spell involving government benefits. For each working spell, the dataset also reports the start and end date of the contract, the type of contract, and the cause of dismissal, among other relevant variables about the worker's labor history, firm, and job characteristics.[25] For the case of the non-employment spells, we observe the associated unemployment benefits and pension amount.

Sample

We focus on prime-age workers- 25 to 55- to avoid capturing atypical behavior at the beginning or end of the career. Regarding data quality, our preferred analysis period for earnings dynamics is from 1990 to 2016. However, we use all available information from workers' labor histories to obtain their past trajectories. Our baseline sample considers affiliated individuals in all industries.

Definition of main variables

The source of the information in the MCVL is the actual contracts (spells) signed between firms and workers. Spells are defined at the establishment level. The information in the dataset regarding job characteristics is, therefore, very detailed and of high quality. This dataset allows us to analyze many individuals while controlling for their characteristics over time, particularly their labor histories, which can determine the decision to become self-employed. Next, we summarize the variables used in the analysis, including definition, construction, and sources.

[25] These include information regarding a firm's location, size, and sector; particular worker characteristics on the contract (full or part-time, if the worker has a disability), and the worker's professional category, as described in the contract.

Self-employment

In order to identify the self-employment spells in the data, we use the variable *régimen de cotización* (contribution regime). This variable identifies the type of regime (salaried work or self-employment) associated with the spell according to the Social Security administration.[26] A worker could have simultaneous spells since we observe the labor history daily. We rely on the richness of the data and define an entry to self-employment as a worker starting a self-employment spell in a given month who was not self-employed last month, even if a simultaneous spell continues. In case we need to keep a unique monthly observation, we define the main job as the one under which he has the most seniority.[27] This approach reduces the error of attributing a specific job regime to workers with a long-lasting job or entrepreneurial activity but who have a seasonal or temporary source of income from a second activity. Whenever the analysis period is at a lower frequency than monthly, the employment status for each period corresponds to the one held in the last month of the corresponding period. For example, for quarterly analyses, we consider a worker to be self-employed in the first quarter if she was self-employed in March; for yearly analyses, the relevance status is December.

Finally, to identify the self-employed, we use the variable *tipo de relación con otras entidades u autónomos*.

Demographics

We observe the individual's birth date and gender. The dataset also contains information on the highest education level obtained by the worker as reported to the Census and the worker's nationality. Finally, there is information regarding the province and municipality where the worker's address is located at the time of the last data extraction, as well as information on the professional category of the worker at the firm, which is a proxy for occupation.

Contract information

We use information on the worker's spell to control for different types of heterogeneity. In particular, we use the following information regarding the paid-employment or self-employment spell:

- *Average monthly earnings:* The MCVL provides nominal monthly earnings from the contribution basis of the worker to the SSSA. These bases are the amounts over which Social Security taxes are applied and determine future disability, unemployment insurance, and pension amounts for all workers, including the self-employed. These contribution bases are available for all workers and are bottom and top-coded by the SSSA. They correspond to base salaries, that is, they do not include overtime, commissions, or bonuses[28]. The data's definition for monthly labor earnings slightly differs for salaried and self-employed workers. The former corresponds with average monthly income for most of the observations; they do not include overtime pay and lay within the lower and upper limits established in the legislation. For the self-employed, the values in the

[26] Most of the previous literature has relied on self-reported employment status, which creates measurement bias.

[27] We have also considered defining the main job status in the case contract overlap as the job that is the main source of earnings within a month. This criterion does not affect the sample significantly but generates job transitions that do not represent the worker's most stable job over time.

[28] For a large share of the workers, base contributions to social security are a good proxy for total salaries (Garcia-Perez (2008) and Cuadrado, Hernández de Cos and Izquierdo (2011)(2011), de la Roca (2014)).

dataset do not strictly reflect actual compensation, as in some cases, the contribution basis is chosen by the worker, and in others, the law fixes it, also subject to maximum and minimum caps as in the case of salaried workers. We deflate nominal monthly earnings and unemployment benefits using the Spanish CPI with the base year 2016 provided by the National Institute of Statistics (INE).

- *Days worked in a month:* We construct days worked in a given month in a given spell from the administrative start and end dates of the spell and total available days in a given month. When we construct days worked in a year, before 40 years old, after 40 years old, and lifetime days, we sum all the days per spell in full-time equivalent units. If the worker has two simultaneous active spells, this could give rise to days worked in a year above 365 days. When we calculate the share of days under each contract, the denominator is the total days worked at the individual level, so shares will always sum up to 100 percent.

- *Tenure:* We compute tenure as the duration of the contract from the beginning to the end of the spell. We observe the exact date (day, month, and year) when the contract started and ended, as provided from the Social Security Administration, so tenure information is extremely accurate.

- *Contract type:* Two types of contracts with different employment protection coexist in the Spanish labor market: (1) fixed-term or temporary contracts, which offer little or no protection after dismissal and have a finite duration, and (2) permanent contracts for extremely protected jobs with firing costs that could rise to three years' worth of a worker's wages. Because permanent contracts are correlated with job security, we use information in MCVL about the contractual relationship between the workers to control for the role of job security in generating transitions between paid employment and self-employment.

- *Part-time contract:* The MCVL reports the percentage of hours of the relationship concerning a full-time job (100% being a full-time worker), which allows us to distinguish between full- and part-time jobs.

- *Industry:* Associated with each spell, the MCVL contains information about the three-digit level industry classification of the firm, based on the Economic Activity National Classification (CNAE). We classify industries into 12 broad groups to control the industry in which the worker was employed prior to a transition.

Unemployment Benefits

We identify unemployment benefits in the database as payments to the unemployed worker using the variable *"Tipo de relacion laboral"*. This category allows us to identify public unemployment insurance recipiency in duration and amount.[29]

C.2 Survey of Household Finances - EFF

As described in the main text, we use data from Spain's Survey of Household Finances. Users interested in this dataset can access it via the Bank of Spain website: https://app.bde.es/efs_

[29] A drawback of this database is that unless the worker receives unemployment insurance, it is impossible to identify periods of unemployment with no benefits and non-employment separately. However, our sample restrictions try to overcome this problem by considering prime-age workers attached to social security and exhibiting an employment spell before and after the dismissal.

www/home?lang=ES. In this appendix, we describe the main variables we use from the 2017 survey to document the extent to which entrepreneurial risk is an important factor, as described in section 4.

Questions on the Value of Businesses

Regarding assets and business valuation, we use data from module 4, "Negocios y Activos Financieros (Business and Financial Assets)," describing the questions below.

To assess the value of businesses, we use question 4.111 from the questionnaire: "Cual es el valor actual del negocio, una vez descontadas las deudas pendientes de dicho negocio? " (What is the business's current value, once excluded pending debts?)

To document the average profit beyond labor earnings, we use question 4.112: "¿Cuales son los beneficios anuales, antes de impuestos, que le proporciona este negocio a su hogar? (What are the annual profits, before taxes, that this business reports to the household?".

For those households without profits, we use question 4.112b: "¿Cuales son las perdidas anuales, antes de impuestos, que le proporciona este negocio a su hogar? (What are the annual losses, before taxes, that this business reports to the household?)"

Finally, in terms of assets, we use 4.113 to obtain the share of household heads with assets used as collateral in the business: "¿Estan utilizando bienes personales (suyos o de su hogar) como garantia o avalaron algun prestamo para el negocio? (Are you using personal assets (own or household) as collateral or obtaining a business loan?"

Questions on Debts of Businesses

We use information from module 3, "Deudas (Debts)," to assess the debts held by the self-employed.

In particular, for the stock of debts, we use question 3.6: "¿Cual es el importe total pendiente de amortizar? (What is the total outstanding amount?)".

We also tried to assess the flow, and for this purpose, we used responses from question 3.11, "¿Cual es el importe mensual que paga en la actualidad por el prestamo, incluyendo amortizacion e intereses? " (What is the monthly loan payment, including principal and interest?)" only for those who report having contracted a loan for business or professional reasons (value 12) in question 3.3: "¿Con que objetivo se contrajo esta deuda? " (What was the objective of this debt?)".