

Gonçalves, Judite; Martins, Pedro S.

**Working Paper**

## The effect of self-employment on health: Instrumental variables analysis of longitudinal social security data

GLO Discussion Paper, No. 245

**Provided in Cooperation with:**  
Global Labor Organization (GLO)

*Suggested Citation:* Gonçalves, Judite; Martins, Pedro S. (2018) : The effect of self-employment on health: Instrumental variables analysis of longitudinal social security data, GLO Discussion Paper, No. 245, Global Labor Organization (GLO), Maastricht

This Version is available at:  
<http://hdl.handle.net/10419/181920>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# The effect of self-employment on health: Instrumental variables analysis of longitudinal social security data

Judite Gonçalves<sup>a,b</sup>, Pedro S. Martins<sup>a,b,c,d</sup>

<sup>a</sup>*School of Business and Management, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

<sup>b</sup>*Nova School of Business and Economics, Universidade Nova de Lisboa, Campus de Carcavelos, Rua da Holanda  
1, 2775-405 Carcavelos, Portugal*

<sup>c</sup>*Institute of Labor Economics (IZA), 53072 Bonn, Germany*

<sup>d</sup>*Global Labor Organization*

---

## Abstract

The growth of novel flexible work formats raises a number of questions about their effects upon health and the potential public policy implications. However, answering these questions is hampered by data and identification constraints. This is the first paper that draws on comprehensive longitudinal administrative data to examine the impact of self-employment in terms of health. In addition to variation in work status of each individual over time, we also consider variation driven by a number of novel instrumental variables. We also focus on an objective health outcome—hospital admissions—that is not subject to recall or other biases that may affect previous studies. Our findings, based on a representative sample of over 100,000 individuals followed monthly from 2005 to 2011 in Portugal, indicate that self-employment tends to reduce the likelihood of hospital admission by at least half.

*Keywords:* Self-employment; hospitalization; sick leave; mortality; instrumental variables

*JEL classification:* C26, I18, J24, J31

---

---

*Email addresses:* judite.goncalves@novasbe.pt (Judite Gonçalves), p.martins@qmul.ac.uk (Pedro S. Martins)

## 1. Introduction

The self-employed represent a non-negligible 14% of the labor force in European countries (Eurostat, 2017). Besides, on average 10% of the adult population of the European Union has used online platforms for the provision of some type of labor services at some point in their lives (Pesole et al., 2018). The ‘platform’ or ‘gig’ economy and self-employment in general are closely monitored by most governments, in part because they represent sources of dynamism and tend to be positively associated with economic growth (e.g. Koellinger and Thurik, 2012). The ongoing growth of the platform economy contributes to the growth of the proportion of self-employed, especially among younger workers, and raises a number of questions about public policy. These include for example occupational health and safety risks, social protection, and representation (European Commission, 2017; Garben, 2017; ILO, 2016).

Self-employment in general and the gig economy in particular are characterized by more flexible work formats, distinct from formal employer-employee relationships framed by employment law. Thus, one current and important issue concerns the impact of multiple forms of self-employment on workers’ health. Indeed, key occupational characteristics, namely job control and job demand, vary significantly between self-employment and wage work. At the same time, in many countries self-employment is subject to little or no protection from public policy or regulation, in terms of social security, employment law, or collective bargaining. This highly flexible context, where workers benefit from little or no insurance and are potentially subject to considerable volatility in work levels, can have additional implications for health.

Job control stands for decision authority, e.g. the freedom to decide what work to do, when and at what pace, which reduces work-related stress. Job demand, on the other hand, represents sources of stress at work, such as being assigned a lot of work and/or having little time to carry out specific tasks. This Job Demand-Job Control framework, originally proposed by Karasek (1979) —see also Karasek and Theorell (1990) and Theorell and Karasek (1996)—, suggests that compared to wage work, self-employment is associated with both higher job control and higher job demand (e.g. Prottas and Thompson, 2006; Stephan and Roesler, 2010).

In fact, self-employed individuals are not subject to orders from others higher up the organizational hierarchy, so they have more decision authority and potentially lower work-related stress. Research also shows that the self-employed are more satisfied with their jobs than wage workers because they can be creative and have more autonomy. In other words, the self-employed may often be able to derive utility from the way outcomes are achieved, something termed ‘procedural utility’ (Benz and Frey, 2008; Schneck, 2014).

However, when self-employed, income and assets directly hinge on one’s ability to work and work effort. In addition, greater exposure to unanticipated demand shocks leaves self-employed individuals subject to more volatile workload and income flows. These represent sources of stress. Work-related stress impacts negatively on health and well-being in general and may increase

incidence of disease (e.g. depression, cardiovascular problems), absence from work due to sickness, use of health care services, and mortality (e.g. Bloemen et al., 2018; Browning and Heinesen, 2012; Halpern, 2005; Holmgren et al., 2009). Stress is also associated with unhealthy behavior, such as smoking and drinking, which may also be detrimental for health. Given these two opposite mechanisms —higher job demand and higher job control—, whether self-employment has a positive or detrimental effect on health is a public policy question that can only be answered through empirical evidence of a causal nature.

There are two main empirical challenges to the identification of a causal effect of self-employment on health: reverse causality and individual unobserved heterogeneity. Reverse causality has to do with the possibility that individuals become self-employed or wage workers at least partly for health-related reasons. On the one hand, self-employment may attract individuals that are healthier on average because healthier individuals tend to be more able to focus on business opportunities or may have easier access to financing (e.g. Gielnik et al., 2012). Additionally, income when self-employed tends to be more closely linked to one’s ability to work than when a wage worker, and access to sickness benefits is harder for the self-employed. All these factors suggest a positive (self-)selection of the healthy into self-employment. On the other hand, health problems may constitute a barrier in finding a wage job, particularly if they are visible to the employer, and push individuals who are less healthy into self-employment (e.g. Zissimopoulos and Karoly, 2007).

Furthermore, several individual traits that are difficult to measure may be related to both health and self-employment decisions. Examples include optimism, perseverance, resilience, risk aversion, as well as genetics. Some individuals who are attracted to and persist in self-employment may also have higher capacity to tolerate and manage stress, and may therefore experience lower stress (Baron et al., 2016). This capacity to deal with stressful factors is another example of an individual characteristic related to both health and type of employment. Earlier life circumstances such as childhood health also influence adult health and type of employment (Case et al., 2005; Case and Paxson, 2010). Taken together, these traits and earlier circumstances mean that self-employed individuals and wage workers may have different health profiles —the so-called individual unobserved heterogeneity.

The empirical literature on self-employment and health is growing but still scarce. Most of it is plagued by the endogeneity issues mentioned above, which are difficult to tackle without longitudinal data. A recent study finds significantly lower work-related stress among self-employed individuals without employees compared to wage workers, using longitudinal data from Australia and controlling for individual fixed effects (Hessels et al., 2017). Previous studies on self-employment and stress provide contradictory findings, but most of them are based on cross-sectional data and use descriptive methods (see Hessels et al., 2017, Table 1, for a review).

In the study by Rietveld et al. (2015), self-employed individuals appear healthier than wage

workers. However, while the positive association between self-employment and health holds when the authors control for reverse causality, it vanishes when they control for individual unobserved heterogeneity. This finding suggests a positive selection of the healthy into self-employment. That study considers subjective health measures, including self-reported number of conditions, overall health, and mental health. It uses longitudinal survey data representative of the population 50+ in the US. The results may therefore not be generalizable for a broader working-age population, in particular for younger workers that may be overrepresented amongst ‘gig economy’ jobs.

Another study by Yoon and Bernell (2013) relies on cross-sectional survey data representative of the adult population in the US and adopts an instrumental variable approach. The authors find that self-employment has a positive impact on several health indicators, namely the absence of chronic conditions such as hypertension and diabetes. They find no effects on other health outcomes, including perceived physical health and mental health.

Considering more objective indicators and administrative data, a five-year follow-up study of the total working population in Sweden finds that self-employed individuals have lower average risk of mortality compared to wage workers, even when controlling for several potential confounders (Toivanen et al., 2016). Overall, there is little robust evidence on the causal effect of self-employment on health. Most of the literature does not take endogeneity into account, as longitudinal data or instrumental variables are seldom available.

In this study, we assess the impact of self-employment on hospital admissions, based on a large sample of administrative social security records representative of the active population in Portugal. We observe more than 130,000 self-employed and wage workers over a period of up to 84 months, between January 2005 and December 2011. We contribute to the literature on the relationship between self-employment and health in several ways. First, we tackle explicitly the endogeneity of the decision to become self-employed. Exploiting the longitudinal nature of our data, we control for individual time-invariant unobserved characteristics through individual fixed effects. We also employ instrumental variables, based on exogenous changes in minimum wages and the proportion of self-employed workers in the region. Second, we look at an objective health outcome —hospitalizations—, based on administrative records.<sup>1</sup> Hospital admissions are comparable across individuals and time periods and not subject to recall bias, as may happen with self-reported indicators in survey data. The Portuguese National Health System provides (almost) free care for all, regardless of type of employment. We also investigate the effect of self-employment on mortality, another objective albeit extreme outcome. Third, we consider the whole working population regardless of age, not a sample limited to older workers where self-employment may be less widespread and representative. Fourth, we explore potentially heterogeneous effects across several subgroups, and look at different causes of hospitalization.

---

<sup>1</sup>See Dobkin et al. (Forthcoming) for a recent study that considers hospitalizations to capture health shocks and their economic consequences for individuals.

Also related to our study is research closer to Labor Economics that assesses the relationship between employment protection and absenteeism. The literature tends to find that stronger employment protection (against dismissal) leads to higher absenteeism. For example, Riphahn (2004) finds that when public sector workers in Germany reach tenure and become virtually ‘un-dismissable’, the annual number of days of (self-reported) sickness absence increases significantly. The author identifies the effect based on survey data and a difference-in-differences strategy, using private sector employees as the control group. Similarly, the average number of days of sick leave per week significantly increases immediately after the probation period ends for white-collar workers of a large Italian bank (Ichino and Riphahn, 2005). The authors interpret these findings as a form of moral hazard, whereby workers reduce effort when it becomes harder for their employer to fire them. In contrast, using longitudinal survey data and an instrumental variable strategy, a recent study finds that perceived employment insecurity has negative consequences for mental health and general well-being, which in turn may contribute to more sickness absence episodes (Cottini and Ghinetti, 2017).

Self-employment may be seen as lower-protection employment, as typically it is not subject to any form of employment law. We contribute to the literature on employment protection and absenteeism by comparing the effect of self-employment on the likelihood of standard sick leave to its effect on the likelihood of hospitalization. In contrast to standard sick leave, hospitalizations are not subject to moral hazard, given their acute nature. Thus, the difference in the effects of self-employment on the two outcomes —standard sick leave and hospitalizations— can be regarded as evidence of the extent of moral hazard created by employment protection.

The remaining of this paper is as follows: the next section lays down the background for the study. In Section 3 we present our data and empirical strategy. In Section 4 we present our results for different outcomes and specifications, as well as a number of robustness checks. Finally, in Section 5 we discuss our findings.

## **2. Background**

Our study is based on data from Portugal, where statutory sick leave covers both self-employed and wage workers. As in many European countries, to deter moral hazard, wage workers face a three-day gap from the onset of the sickness episode until the social security benefit starts to be paid (i.e. waiting or ‘elimination’ period). For the self-employed, this waiting period is much longer: thirty days. In 2018, the waiting period for the self-employed was reduced from thirty to ten days. This policy decision highlights the pertinence of this paper.

Due to the different waiting periods, social security records include sickness episodes that last four days or more in the case of wage workers, and at least thirty-one days in the case of the self-employed. The first three/thirty days are not eligible for sickness benefits. Thus, all things equal, on average the sickness events of the self-employed that are administratively recorded are

much more selected and severe.

The different waiting periods also entail different incentives for wage workers and self-employed individuals. Wage workers face much lower opportunity costs from reporting sick to work, i.e. fewer days without income from work. In some cases, collective bargaining provisions, determined by unions and firms or employer associations, may even lead to the payment of the first three days of absence as well. Those provisions apply to wage workers but not to the self-employed. Thus, wage workers may engage more often in moral hazard: ‘cheat’ by going on sick leave when they are not really sick.

In striking contrast, there is no waiting period for either self-employed or wage workers in the case of hospitalization. Benefits are the same for both types of workers. Besides, due to the specific, acute nature of hospitalizations, they are less likely to be timed deliberately by individuals and therefore less likely to be artificial episodes of sickness. In sum, compared to standard—i.e. non-hospitalization—sickness episodes, hospitalizations are a significantly more objective health outcome and hospital admissions should be strictly comparable between wage workers and self-employed individuals. This is why this study focuses on the effect of self-employment on hospital admissions. In addition, the Portuguese National Health System is almost free at the point of use.

For nearly the entire period under analysis here (Sep 2005-Dec 2011), the replacement rate of the Portuguese sickness benefit was equal to 65% of forgone wages for the first 90 days of sick leave, 70% from the 91<sup>st</sup> to the 365<sup>th</sup> day, and 75% from the 365<sup>th</sup> day onwards.<sup>2</sup>

### 3. Data and methods

#### 3.1. Social security data

We use data made available by the Social Security agency (IISS). The dataset is a 1% random sample of all individuals with social security records over the period 2005-2011. Those records include salaries from (self-)employment, as well as sickness, unemployment, maternity, and other social security benefits (see Martins, 2016 for more details about this dataset).

For this study, we consider individuals on a monthly basis, from January 2005 to December 2011. We use information on whether they are wage workers or self-employed, as well as whether they have a sickness episode that leads to a sickness benefit being paid in a specific month. The data allow us to distinguish between standard sickness spells, implying absence from work, and acute cases that lead to hospitalization, as the two cases are treated differently by social security

---

<sup>2</sup>During the first eight months of 2005, the replacement rate was 55% of forgone wages for the first 30 days of sick leave, and 60% from the 31<sup>st</sup> to the 90<sup>th</sup> day. Sickness benefits are granted for a maximum of 1,095 days for wage workers and 365 days for self-employed individuals (Decree-law 28/2004 of Feb 4, Decree-law 133/2012 of Jun 27, and Decree-law 146/2005 of Aug 26).

(see Section 2). Additional individual information available in the dataset includes gender, age, nationality (Portuguese or foreign), and income from work.

The dataset includes 130,000+ different individuals, of which 10,000+ are self-employed at some point over the period 2005-2011. In addition, more than 6,000 individuals switch at least once over this period between self-employment and wage work. The numbers of switches to and from self-employment over time are displayed in Figure A1, in the Appendix. In total, over this 84 month-long period, there are almost 7 million individual-month observations.

Descriptive statistics by type of employment are shown in Table 1. The self-employed account for 4.25% of the person-month observations, or almost 300,000 observations. Based on these representative data, we can say that the average monthly rate of hospitalization is 0.10% among the self-employed, compared to 0.18% —nearly twice as much— among wage workers. The self-employed also have a lower monthly rate of standard sick leave than wage workers. Here, the gap is much larger —0.48% compared to 2.37%—, i.e. standard sick leave is nearly five times more likely among wage workers. There are also differences in the average lengths of hospitalization and standard sick leave. The differences observed over time are shown in Figures A2-A5, in the Appendix. Lastly, we find that the proportion of women is slightly lower among the self-employed than among wage workers (49% versus 53%), that the self-employed are on average older (43 versus 37 years old), and that the proportion of foreigners is also slightly lower among the self-employed (10% versus 12%).

Table 1: Descriptive statistics by type of employment

	Self-employed	Wage workers	Total
Any hospitalization in the following month	0.10%	0.18%	0.18%
Average number of days of hospitalization	12.52 (9.37) [1;90]	10.97 (7.35) [1;90]	11.00 (7.41) [1;90]
Any standard sickness episode in the following month	0.48%	2.37%	2.29%
Average number of days of sick leave	18.44 12.68 [1; 305]	9.36 8.14 [1; 73]	9.44 8.23 [1; 305]
Female	48.73%	52.98%	52.80%
Age	42.80 (11.16) [18; 65]	37.11 (10.44) [18; 65]	37.35 (10.54) [18; 65]
Foreign	9.93%	11.95%	11.86%
Observations (individual-month)	283,809 4.25%	6,396,721 95.75%	6,680,530 100.00%

Continuous variables: standard deviations in parentheses and minimum and maximum values in brackets. Descriptive statistics for standard sickness spells are based on a slightly larger sample, as there are more standard sickness spells than hospitalizations.



### 3.2. Diagnosis-related groups data

Information on the cause of hospitalization, used in complementary analyses, comes from the national diagnosis-related groups (DRG) dataset, managed by the Central Administration of the Portuguese Health System (ACSS). This dataset includes individual-level information on all inpatient and outpatient admissions at public hospitals in Portugal, since 1993.

We cannot match directly individuals across the two datasets. We matched hospitalizations in the social security and DRG databases using the individuals' gender, year of birth, place of residence (municipality), and day of admission. Only exact matches based on these variables were considered. Multiple hospitalizations with the same characteristics were discarded. This resulted in slightly more than 6,000 hospital admissions matched between the two datasets, about half of the total number of admissions in the social security database. The remaining hospitalizations correspond to multiple observations with the same values of the matching variables, which were discarded, and hospitalizations in private hospitals, which are not recorded in the DRG dataset.

Using the DRG classification of hospital admissions, we grouped the causes of hospitalization into eight categories. Table 2 shows the rate of hospitalization by cause and type of employment in the previous month. Regardless of the cause, the hospitalization rate is always lower among the self-employed than among wage workers. We also find that after the 'all others' category, diseases of the digestive and respiratory systems represent the main causes of hospitalization.

Table 2: Rate of hospitalization by cause and type of employment in the previous month

	Self-employed	Wage workers	Total
Diseases of the nervous system	0.0035%	0.0042%	0.0042%
Diseases of the respiratory system	0.0046%	0.0110%	0.0107%
Diseases of the circulatory system	0.0046%	0.0074%	0.0073%
Diseases of the digestive system	0.0134%	0.0192%	0.0190%
Diseases of the musculoskeletal system	0.0042%	0.0065%	0.0064%
Pregnancy and childbirth	0.0014%	0.0053%	0.0052%
Mental disorders	0.0000%	0.0018%	0.0017%
All others	0.0169%	0.0313%	0.0307%

### 3.3. Identification and empirical strategy

To determine the effect of self-employment on the likelihood of hospitalization, we estimate several specifications of a linear probability model, as in the equation below:

$$Pr[\text{hosp}_{i,t} = 1] = \beta_0 + \beta_1 \text{self-employed}_{i,t-1} + \gamma X_{i,t} + \tau_t + \mu_i + \varepsilon_{i,t} \quad (1)$$

The dependent variable,  $\text{hosp}_{i,t}$ , is binary and indicates whether individual  $i$  is hospitalized in month  $t$  or not. We start by regressing  $\text{hosp}_{i,t}$  on a constant and the one-month lag of the self-employment indicator,  $\text{self-employed}_{i,t-1}$ , with  $\beta_0$  and  $\beta_1$  the associated coefficients and  $\varepsilon_{i,t}$  the usual error term (Model 1 – SLR).

We find that the choice of lagging the self-employment indicator one month is inconsequential. Using all lags up to the third or the twelfth, for example, gives estimated total effects of self-employment that differ only slightly in magnitude, having the same sign and statistical significance. This shows the relative stability of the self-employment indicator, as individuals rarely change type of work more than once over the seven-year period considered here. In this study, we are interested in the overall effect of self-employment and not in the time dynamics. That overall effect can be captured by (any) single lagged indicator, given the high correlation between adjacent lags. Furthermore, using more than one lag would make us lose many more observations. Given the chosen specification,  $\beta_1$  gives the effect of being self-employed, as opposed to being a wage worker, on the likelihood of being hospitalized in the following month.

Next, we control for gender, age, and nationality, included in  $X_{i,t}$ , with  $\gamma$  the associated vector of coefficients. We also include 82 time fixed effects, denoted by  $\tau_t$  (84 months minus Jan 2005, due to the lag, and Feb 2005, which is the reference month; Model 2 – MLR).

The following specification controls also for time-invariant individual unobserved heterogeneity through the inclusion of individual fixed effects,  $\mu_i$  (Model 3 – FE). It allows us to compare the likelihood of hospitalization for the same individual over the up to 84 month-long period covered by our data. In this specification, identification comes from those more than 6,000 individuals who switch at least once between wage work and self-employment over the period of analysis.

As discussed in the Introduction, another potential source of endogeneity is reverse causality from health to type of work. There may also be unobserved individual characteristics that vary over time and thus are not captured by the individual fixed effects. To tackle these issues, in addition to the individual fixed effects, we employ an instrumental variable strategy (Models 4-6 – IV-FE). We consider two instruments. The first is the difference between the individual’s income and the statutory minimum wage, calculated as follows:

$$iv1_{i,y(t)} = \ln(\text{avginc}_{i,y(t-1)}) - \ln(\text{minwage}_{y(t)}) \quad (2)$$

This instrument,  $iv1_{i,y(t)}$ , varies across individuals,  $i$ , and years,  $y(t)$ , but not within years. In Portugal as in many countries, minimum wages increase once per year, in January. The first term in (2),  $\text{avginc}_{i,y(t-1)}$ , is the average of individual  $i$ ’s income from work over the twelve months of the previous year. The second term,  $\text{minwage}_{y(t)}$ , is the national statutory minimum wage in the current year.

The motivation for this instrument follows from parallel ongoing work, in which we find that a larger difference between the new statutory minimum wage and an individual’s previous wage is associated with a higher probability of moving from wage work to self-employment (Gonçalves and Martins, 2018a). When the minimum wage increases from one year to the next, in January, some individuals whose wages must increase in order to comply with the new minimum wage may be dismissed or have their (fixed-term) employment contracts discontinued. These individuals may move to self-employment, which is not subject to minimum wages or any form of employment law,

including collective bargaining. In some cases, they may even continue to work for the same firm, but as independent contractors or self-employed workers, rather than as employees.

During 2005-2011, the statutory minimum wage in Portugal increased substantially, in both nominal and real terms, providing useful variation for our analysis.<sup>3</sup> We expect instrument one to be relevant in explaining the likelihood of self-employment. More precisely, given the specification in (2), we expect to find a negative relationship between instrument one and the likelihood of self-employment. In other words, the higher the wage of the individual compared to the value of the new minimum wage, the less likely that she will become self-employed. The difference in logs accounts for a non-linear relationship and limits the influence of outliers.

Our second instrument,  $\text{iv2}_{j,y(t)}$ , is the natural logarithm of the proportion of self-employed workers in municipality  $j$  in year  $y(t)$ . There are more than 300 municipalities in Portugal, which gives significant variation. The proportion of workers in a given municipality who are self-employed gives some indication of the structure of the labor market in that area. It has been used to instrument self-employment decisions in previous literature, e.g. Noseleit (2014). We expect that the larger that proportion, the higher the likelihood that any individual  $i$  residing in municipality  $j$  is self-employed.

Besides being relevant, the instruments must be validly excluded from the main equation; i.e. they must impact health solely through their impact on the likelihood of self-employment. Intuitively, we believe both instruments are valid, especially since we are conditioning on time and individual fixed effects. With two instruments and one potentially endogenous variable, we are able to test statistically the validity of the overidentifying restriction.

We present the results of IV estimations using both instruments as well as using each instrument separately, allowing for their comparison (Models 4-6). Given that the endogeneous variable,  $\text{self-employed}_{i,t-1}$ , is lagged, we also use the lags of the instruments. When using the first instrument, this leads to losing the following twelve months of observations, until Feb 2006.

All models with individual fixed effects (Models 3-6) are estimated using the *reghdfe* package in *Stata* (Correia, 2016). This command allows us to estimate models with large numbers of both time and individual fixed effects in a computationally efficient manner, including when using IV.

In complementary analyses, we assess whether the overall effect of self-employment on the likelihood of hospitalization is driven by specific causes or diseases as presented in Table 2. To do this, we repeat the estimation considering only each particular cause of hospitalization: diseases of the nervous system, diseases of the respiratory system, diseases of the circulatory system, diseases of the digestive system, diseases of the musculoskeletal system, pregnancy and childbirth, mental

---

<sup>3</sup>The values of the statutory minimum wage are 374.70€ (2005), 385.90€ (2006), 403.00€ (2007), 426.00€ (2008), 450.00€ (2009), 475.00€ (2010), and 485.00€ (2011). Each year, the new minimum wage comes into effect in January. Centeno et al. (2011), using similar data but focusing on the employment impact of minimum wage increases, finds evidence of negative effects consistent with our approach here.

disorders, and all other causes. In these analyses, hospitalizations with unknown cause —i.e. not matched with DRG data— were discarded.

We also assess whether the effect of self-employment on the likelihood of hospitalization differs according to gender, age, nationality, income level, type of switch, and duration of self-employment. To do this, we estimate separate regressions for men and women, under and over 35 year olds, natives and foreigners, individuals with average monthly income above or below the sample median, individuals who switch from wage work to self-employment or vice-versa, and individuals who are self-employed for more or less than 24 months over the period of analysis.

Finally, we estimate the same model specifications to assess the effects of self-employment on the length of hospitalization and likelihood of standard sick leave. Regarding the former, the analysis is based only on individual-month observations in which there is a hospitalization, a much smaller sample. The natural logarithm of the number of hospitalization days is used as the dependent variable, to account for the skewness in the distribution of hospitalization days.

#### *3.4. Analysis of mortality rates*

Besides standard sickness spells and hospitalizations, the dataset allows us to investigate mortality. To explore the effect of self-employment on mortality, we aggregate the data to the person-year level, as we know the year but not the month in which the individual passes away. We create a binary dependent variable that takes value one if individual  $i$  passes in year  $t + 1$  and zero otherwise. The self-employment indicator takes value one if the individual is self-employed during the whole year  $t$ . We estimate the same model specifications as described in the previous section, adjusted for the annual frequency considered here. All explanatory variables are measured in year  $t$ . We exclude observations for the year in which the person passes away, as we do not know in which month in the year that event takes place.

#### *3.5. Robustness checks*

It is possible that when individuals are less healthy, they end up in a non-employment situation, such as very long sick leave, disability benefits, or unemployment. In months where individuals are not employed, there aren't records of social security contributions, which means that the individuals do not appear in the dataset. This may cause selection bias. To investigate this, we repeat the analyses on two balanced panels and compare results with the main ones. The first balanced panel includes only individuals observed in all 84 months between Jan 2005 and Dec 2011 (18,504 individuals). In the second balanced panel there are all 133,000 individuals in the dataset, including in months where they are not employed. In this analysis, there is a non-employment indicator in addition to the self-employment indicator.

Lastly, we repeat the main analysis using quarterly rather than monthly data and compare the magnitudes of the estimates. Aggregating the data in this way reduces total sample size to about one third.

## 4. Results

### 4.1. Effects on hospital admissions

Table 3 shows the estimated effect of self-employment on the likelihood of hospitalization. The unadjusted average monthly difference in the likelihood of hospitalization between wage workers and self-employed individuals is  $-0.00081$  ( $p < 0.01$ ; Model 1). In relative terms, this difference is considerable: the self-employed are 46% less likely than wage workers to have a hospitalization in any given month ( $-0.00081/0.00182$ ). This figure also emerges from the comparison of the monthly hospitalization rates presented in Table 1: 0.10% for the self-employed versus 0.18% for wage workers.

Controlling for gender, age, nationality, and time fixed effects makes the association between self-employment and likelihood of hospitalization even more negative (Model 2). Results also indicate that female, older, and native workers have higher rates of hospitalization. Moreover, the estimated coefficient on the self-employment indicator in the model with individual fixed effects is  $-0.00079$  ( $p < 0.01$ ; Model 3), almost identical to that in Model 1. This suggests that individual time-invariant unobserved characteristics associated with both self-employment and likelihood of hospitalization are not driving the results.

Results of the specification tests in the IV models are shown at the bottom of Table 3 and indicate that the instruments are both strong and validly excluded from the main equation. The instruments have large associated F statistics and have the expected signs in the first stage regressions (Table A1 in the Appendix).<sup>4</sup> They also pass the overidentification test, with a small, insignificant Hansen J statistic. Lastly, results of the endogeneity test are not conclusive, leading to the rejection of exogeneity of the self-employment indicator when using instrument one (Models 4 and 5), but not when using only instrument two (Model 6).

Overall, our findings from the IV models are consistent in sign with those of the previous models, indicating a negative effect of self-employment on the likelihood of hospitalization. This suggests that self-employment has a positive impact on health: the negative association between self-employment and likelihood of hospitalization is not (only) a matter of the healthy self-selecting into self-employment. Moreover, the IV results indicate that there is a positive bias when endogeneity is not fully addressed. If we assume that there is an omitted variable—individual health—that varies over time and is therefore not entirely captured by the individual fixed effects, then likely it is negatively correlated with self-employment. That negative correlation, in conjunction with a negative impact of health on the likelihood of hospitalization, gives a positive bias of the self-employment coefficient. Thus, one possible explanation for the positive bias is that it may be more difficult for less healthy individuals to find wage employment, so that they tend to

---

<sup>4</sup>The standard errors are clustered at the individual level. Clustering at the municipality level, which is the level of observation of the second instrument, gives only slightly smaller F statistics.

become self-employed.

The IV coefficients are much larger in absolute terms, in some cases by a factor of 20. To put this result further into perspective, the IV coefficient in Model 4 is less than seven times the average of the individual fixed effects. The larger estimated coefficients when using IV may have to do with the fact that IV estimation gives a local average treatment effect (LATE), the effect for individuals who earn about the minimum wage and are affected by minimum wage increases, and for individuals who live in municipalities where the rate of self-employment changes and impacts their individual likelihood to become self-employed (the compliers).

Table 3: Effect of self-employment on the likelihood of hospitalization

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
Self-employed (t-1)	-0.00081*** (0.00007)	-0.00104*** (0.00007)	-0.00079*** (0.00014)	-0.01686*** (0.00281)	-0.01827*** (0.00293)	-0.00467 (0.00548)
Female	–	0.00056*** (0.00004)	–	–	–	–
26-35 years old	–	0.00027*** (0.00005)	-0.00020** (0.00009)	-0.00012 (0.00012)	-0.00012 (0.00011)	-0.00020** (0.00009)
36-45 years old	–	0.00091*** (0.00006)	-0.00017 (0.00014)	0.00005 (0.00017)	0.00006 (0.00017)	-0.00015 (0.00015)
46-55 years old	–	0.00136*** (0.00007)	-0.00004 (0.00021)	0.00037 (0.00026)	0.00039 (0.00026)	-0.00003 (0.00022)
56-65 years old	–	0.00158*** (0.00010)	0.00047 (0.00031)	0.00092** (0.00037)	0.00093** (0.00037)	0.00045 (0.00031)
Foreign	–	-0.00011** (0.00006)	–	–	–	–
Time FE	–	Yes	Yes	Yes	Yes	Yes
Individual FE	–	–	Yes	Yes	Yes	Yes
Constant	0.00182*** (0.00002)	0.00055*** (0.00014)	0.00191*** (0.00011)	–	–	–
Instrument strength	–	–	–	F=383.33***	F=559.49***	F=467.00***
Instrument validity	–	–	–	$\chi^2=0.875$	–	–
Endogeneity	–	–	–	$\chi^2=36.943***$	$\chi^2=38.363***$	$\chi^2=0.501$
Observations	6,680,530	6,680,530	6,678,743	5,231,898	5,277,454	6,583,929

Clustered standard errors in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

See Table A1 in the Appendix for the first stage results.

We find no evidence that a lower likelihood of hospitalization among self-employed individuals comes at the expense of longer lengths of hospital stays, which would suggest that self-employed individuals delay going to the hospital until they are more severely sick (Table A2 in the Appendix). Nevertheless, these results should be interpreted with caution, because in this analysis, statistical power is significantly reduced. First, only observations of hospitalizations are used. Second, many individuals that have only one hospitalization over the entire period of analysis are discarded from

the estimation when individual fixed effects are included. The instruments are also not valid in this case.

Next, we explore potentially heterogeneous effects of self-employment for different subgroups (Table 4). First, we find that the negative effect of self-employment on the likelihood of hospitalization is stronger for women than for men, and for workers of foreign nationalities than for Portuguese workers, regardless of the specification considered. According to the results of IV models, the effect is also stronger for older workers than for younger ones.

Regarding income level, the estimated coefficient on self-employment is more negative for high income individuals if endogeneity is not taken into account. In models controlling for individual fixed effects and instrumental variables, the negative effect of self-employment on the likelihood of hospitalization appears stronger for low income workers.

There is also some evidence of a more negative effect of self-employment on the likelihood of hospitalization for individuals who switch from wage work to self-employment than for those who switch from self-employment to wage work. This suggests that, on average, switching to self-employment is more beneficial for individuals' health than it is detrimental to switch to wage work.

Lastly, according to the results of the model that controls only for individual fixed effects (Model 3), being self-employed for a longer period of time seems to have a protective effect on health, as the coefficient is more negative for individuals who are self-employed for at least 24 months during the period of analysis. However, IV results give more negative coefficients for the short-term self-employed.

In our last investigation of hospital admissions, we consider separately each group of causes of hospitalization, using information from the DRG database. As mentioned before, the numbers of admissions in the dataset are roughly cut in half, as it was not possible to find exact matches in the DRG database for all hospitalizations in the social security database. In other words, we drop all hospitalizations which underlying medical cause we cannot identify, as well as hospitalizations for other causes than the ones considered in each specification. Results are reported in Table 5. Overall, we find that the self-employed have lower monthly rates of hospitalization for any cause (Model 1), something that was already visible from the summary statistics in Table 2. Controlling for endogeneity, there are still significant differences between self-employed and wage workers in the rates of hospitalization due to diseases of the circulatory and digestive systems, and well as for admissions related to pregnancy and childbirth. The latter result suggests a potential overestimation of the effect found for all hospitalizations, where the analysis includes hospitalizations related to pregnancy and childbirth that have nothing to do with health problems. Still, we find negative effects for other causes of hospitalization, for men, and for older workers, which underlines the robustness of our main qualitative findings.

Table 4: Heterogeneous effects of self-employment on the likelihood of hospitalization

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
<b>By gender</b>						
Women	-0.00096***	-0.00120***	-0.00115***	-0.02369***	-0.02564***	-0.00931
(N=3,527,337)	(0.00010)	(0.00010)	(0.00021)	(0.00471)	(0.00503)	(0.00854)
Men	-0.00062***	-0.00086***	-0.00043**	-0.01091***	-0.01195***	-0.00025
(N=3,153,193)	(0.00009)	(0.00009)	(0.00019)	(0.00330)	(0.00334)	(0.00699)
<b>By age</b>						
18-35 years old	-0.00099***	-0.00102***	-0.00087***	-0.01455***	-0.01827***	0.00374
(N=3,292,930)	(0.00007)	(0.00007)	(0.00017)	(0.00351)	(0.00371)	(0.00745)
36-65 years old	-0.00098***	-0.00104***	-0.00060**	-0.02503***	-0.02567***	-0.00833
(N=3,387,600)	(0.00009)	(0.00009)	(0.00025)	(0.00593)	(0.00640)	(0.00888)
<b>By nationality</b>						
Portuguese	-0.00076***	-0.00100***	-0.00079***	-0.01546***	-0.01691***	-0.00493
(N=5,888,266)	(0.00007)	(0.00007)	(0.00015)	(0.00296)	(0.00310)	(0.00561)
Foreign nationality	-0.00127***	-0.00136***	-0.00085**	-0.02719***	-0.02776***	-0.00089
(N=792,264)	(0.00014)	(0.00014)	(0.00036)	(0.00862)	(0.00876)	(0.02335)
<b>By income level</b>						
High income	-0.00119***	-0.00120***	-0.00065***	-0.00985***	-0.00966***	-0.02144*
(N=3,383,531)	(0.00014)	(0.00014)	(0.00017)	(0.00251)	(0.00252)	(0.01260)
Low income	-0.00077***	-0.00096***	-0.00090***	-0.01479**	-0.06261***	0.00316
(N=3,296,995)	(0.00008)	(0.00008)	(0.00021)	(0.00705)	(0.01482)	(0.00551)
<b>By type of switch</b>						
Wage work to self-employment	-0.00074***	-0.00102***	-0.00099***	-0.02994***	-0.03167***	-0.01039
(N=6,556,783)	(0.00008)	(0.00008)	(0.00018)	(0.00533)	(0.00551)	(0.01415)
Self-employment to wage work	-0.00076***	-0.00102***	-0.00080***	-0.02415***	-0.02682***	-0.00639
(N=6,578,910)	(0.00007)	(0.00008)	(0.00018)	(0.00411)	(0.00436)	(0.00701)
<b>By duration of self-employment</b>						
Short-term self-employment	-0.00118***	-0.00116***	-0.00069***	-0.03703***	-0.03940***	-0.00802
(N=6,408,338)	(0.00011)	(0.00011)	(0.00016)	(0.00655)	(0.00679)	(0.01125)
Long-term self-employment	-0.00074***	-0.00102***	-0.00091***	-0.02772***	-0.03080***	-0.00778
(N=6,472,350)	(0.00008)	(0.00008)	(0.00024)	(0.00486)	(0.00516)	(0.01043)

Numbers of observations lower in FE and IV-FE models. Clustered standard errors in parentheses.

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

#### 4.2. Effect on standard sick leave

On average, the self-employed have a 1.89 percentage point, or 80% (1.89/2.37) lower likelihood of sick leave than wage workers ( $p < 0.01$ ; Model 1, Table A3 in the Appendix). The coefficient on the self-employment indicator remains practically the same when controlling for demographics and time fixed effects (Model 2). Controlling for individual time-invariant characteristics results in a smaller coefficient in absolute terms,  $-0.012$  compared to  $-0.019$ , that is also significantly negative ( $p < 0.01$ ; Model 3). IV results must be interpreted with caution, as the instruments do



Table 5: Effect of self-employment on the likelihood of hospitalization for different causes

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
Diseases of the nervous system (N=6,668,910)	-0.00001 (0.00001)	-0.00002 (0.00001)	-0.00004** (0.00002)	-0.00032 (0.00042)	-0.00035 (0.00043)	0.00071 (0.00075)
Diseases of the respiratory system (N=6,669,344)	-0.00006*** (0.00001)	-0.00007*** (0.00001)	-0.00007** (0.00003)	-0.00082 (0.00063)	-0.00086 (0.00065)	-0.00053 (0.00140)
Diseases of the circulatory system (N=6,669,114)	-0.00003** (0.00001)	-0.00005*** (0.00001)	-0.00001 (0.00002)	-0.00072* (0.00042)	-0.00095** (0.00041)	0.00063 (0.00108)
Diseases of the digestive system (N=6,669,895)	-0.00006** (0.00002)	-0.00009*** (0.00002)	0.00004 (0.00004)	-0.00225*** (0.00086)	-0.00237*** (0.00086)	0.00064 (0.00204)
Diseases of the musculoskeletal system (N=6,669,058)	-0.00002* (0.00001)	-0.00003*** (0.00001)	-0.00011** (0.00005)	0.00003 (0.00058)	0.00027 (0.00058)	-0.00081 (0.00124)
Pregnancy and childbirth (N=6,668,973)	-0.00004*** (0.00001)	-0.00002*** (0.00001)	-0.00007* (0.00004)	-0.00128*** (0.00046)	-0.00144*** (0.00048)	-0.00038 (0.00093)
Mental disorders (N=6,668,742)	-0.00002*** (0.00000)	-0.00002*** (0.00000)	-0.00000 (0.00000)	0.00007 (0.00029)	0.00002 (0.00034)	0.00074 (0.00050)
All others (N=6,670,675)	-0.00014*** (0.00003)	-0.00018*** (0.00003)	-0.00010** (0.00005)	-0.00148 (0.00113)	-0.00121 (0.00114)	0.00029 (0.00229)

Numbers of observations lower in FE and IV-FE models. Clustered standard errors in parentheses.

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

not pass the overidentification test, but also suggest a negative effect of self-employment on the likelihood of sick leave.

As explained in Section 2, we cannot say that self-employment is good for health based on this result. The association between self-employment and likelihood of sick leave is partly due to the different waiting periods that self-employed individuals and wage workers face until they can claim sickness benefits. The different waiting periods imply that we cannot observe sick leave spells of self-employed individuals that last for less than 31 days. Moreover, it is likely that there is a higher proportion of fraudulent sickness spells among wage workers for moral hazard reasons. These facts are consistent with the negative associations we find. To explore this further, we consider only sickness spells that last more than 30 days. We still obtain negative and significant estimated coefficients on the self-employment indicator (Table A4).

In sum, given the institutional context and data available, we cannot fully disentangle the effect of self-employment on the likelihood of sick leave that is due to health differences between the self-employed and wage workers from the confounding effect of the heterogeneous waiting periods. Nonetheless, the negative associations, which hold when restricting the analysis to long sickness spells, are in line with the negative effects for hospitalizations.

### 4.3. Effect on mortality

Table 6 presents the effect of self-employment on the likelihood of mortality in the following year. We find that the unadjusted average difference in the likelihood of mortality between wage workers and self-employed individuals is statistically zero ( $p > 0.1$ ; Model 1). When adjusted for age, gender, nationality, and year fixed effects, that difference becomes statistically significant and equal to  $-0.00102$ , i.e. self-employed individuals are 0.10 percentage points (88%) less likely to die in the following year than wage workers ( $p < 0.01$ ; Model 2). We also find evidence that mortality is lower for women, not statistically different between native and foreign workers, and (unsurprisingly) increases with age. When controlling also for individual fixed effects, the association between self-employment and likelihood of mortality is no longer statistically significant (Model 3), and IV Models 4 and 5, but not 6, yield a positive and significant coefficient.

In sum, given the differences in sign and precision of the coefficients across specifications, we do not find conclusive evidence on the effect of self-employment on mortality. This may be because data are aggregated to an yearly frequency or because mortality is such a rare and extreme outcome, which means that there is little variation to identify precisely an effect of self-employment.

Table 6: Effect of self-employment on the likelihood of mortality in year t+1

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
Self-employed	-0.00022 (0.00025)	-0.00102*** (0.00026)	-0.00095 (0.00060)	0.01931* (0.01012)	0.02820** (0.01169)	-0.00837 (0.01546)
Female	–	-0.00118*** (0.00009)	–	–	–	–
26-35 years old	–	0.00016** (0.00007)	-0.00100*** (0.00019)	-0.00102*** (0.00024)	-0.00105*** (0.00024)	-0.00101*** (0.00019)
36-45 years old	–	0.00068*** (0.00010)	-0.00190*** (0.00029)	-0.00193*** (0.00035)	-0.00198*** (0.00035)	-0.00189*** (0.00030)
46-55 years old	–	0.00209*** (0.00017)	-0.00048 (0.00056)	-0.00043 (0.00069)	-0.00035 (0.00070)	-0.00060 (0.00056)
56-65 years old	–	0.00425*** (0.00035)	0.00277*** (0.00093)	0.00329*** (0.00115)	0.00337*** (0.00115)	0.00272*** (0.00094)
Foreign	–	-0.00008 (0.00012)	–	–	–	–
Year FE	–	Yes	Yes	Yes	Yes	Yes
Individual FE	–	–	Yes	Yes	Yes	Yes
Constant	0.00116*** (0.00004)	0.00061*** (0.00011)	0.00180*** (0.00024)	–	–	–
Instrument strength	–	–	–	F=244.12***	F=236.68***	F=410.43***
Instrument validity	–	–	–	$\chi^2=1.148$	–	–
Endogeneity	–	–	–	$\chi^2=4.709**$	$\chi^2=6.555**$	$\chi^2=0.231$
Observations	600,495	600,495	590,249	426,339	430,283	580,943

Clustered standard errors in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

#### 4.4. Robustness checks

Results of robustness checks are reported in Table A5 in the Appendix. Depending on the specification, estimations on balanced panels give coefficients that are smaller or larger in magnitude, compared to the ones from the main analysis, but equally negative, large, and strongly significant. This suggests that our main results do not suffer from selection bias. Using quarterly data gives negative and strongly significant coefficients, which are roughly three times as large as the coefficients in the main analysis, as expected.

### 5. Discussion and conclusion

It is probably as challenging as it is important to determine whether self-employment is good or detrimental for health. The potential self-selection of the healthy into or out of self-employment is difficult to rule out empirically. However, separating the effect of self-employment on health from that selection effect is crucial to inform policy decisions. Furthermore, as the ‘gig economy’ grows around the world, leading to increasing shares of the workforce in self-employment, causal evidence about the health implications of that type of work become more pressing.

In this study, we provide causal evidence on the impact of self-employment on health by relying on a large sample of social security records representative of the working population in Portugal. We take advantage of the longitudinal nature of the data and track more than 130,000 individuals over up to 84 months. In addition, we employ an instrumental variable strategy to rule out endogeneity. We also focus on hospital admissions, which is a direct result of individuals’ bad health and can be considered an objective health outcome.

We find that self-employed individuals have 46% lower likelihood of hospitalization than wage workers, an effect that becomes stronger when taking into account the endogeneity of self-employment decisions. This is in contrast with the results of Rietveld et al. (2015), who find a negative association between self-employment and health that is fully explained by a selection effect. The different results between the two studies may be due to the type of health measures and samples used. While we focus on administrative records of hospitalizations and consider the whole working population, Rietveld et al. (2015) draw on survey-based subjective health measures and focus on the 50+ population.

Our results for hospital admissions indicate that the beneficial effects of higher job control when self-employed exceed the detrimental effects of higher job demands. When we look at different causes of hospitalization, we find that the overall negative effect of self-employment on the likelihood of hospitalization is partly driven by diseases of the circulatory and digestive systems. The medical literature identifies stress as an important cause of cardiovascular diseases and digestive disorders (e.g. Mayer, 2000; Steptoe and Kivimäki, 2012). Our classification of causes of hospitalization is broad and may include diseases unrelated with stress. Nonetheless, our results do not contradict the interpretation that self-employed individuals seem to suffer from lower stress

than wage workers. Bloemen et al. (2018) also find that the probable mechanism driving the effect of job loss on mortality is stress, through acute diseases of the circulatory system. Regarding the negative effect of self-employment on admissions related to pregnancy and childbirth, this result is consistent with the results in Gonçalves and Martins, 2018b, where we focus on fertility issues.

We also find that the likelihood of standard sick leave is lower among self-employed individuals than among wage workers, even when restricting the analysis to longer sickness spells, of more than 30 days. However, this analysis may be confounded by other factors. Although consistent with the results for hospitalizations, we refrain from drawing definitive conclusions about an effect on standard sick leave. Similarly, we do not have enough evidence to conclude on an effect of self-employment on mortality.

Our findings also contribute to a related literature that indicates that employment protection legislation can lead to absenteeism. Our results from the analysis of hospitalizations suggest that the relationship between employment protection and absenteeism may not be entirely explained by moral hazard as implied previously (e.g. Riphahn, 2004). Moreover, our findings are in contrast with the study by Cottini and Ghinetti (2017), who find that perceived employment insecurity has negative consequences upon mental health and general well-being.

Our results may also reflect changes in the type of work *per se* when individuals switch to/from self-employment. Unfortunately, we cannot explore this issue with the available data, as we do not know the industry/occupation of self-employed individuals. Nevertheless, it seems plausible that in most cases individuals do not transition to drastically different jobs when switching between self-employment and wage work —e.g. from blue-collar to white-collar jobs. Even if individuals do change industry when switching to/from self-employment, most often they are likely to remain in relatively similar occupations. Still, the inability to investigate the potentially different effects of self-employment by industry remains a limitation of this study and a topic that deserves to be explored in future research. For instance, manufacturing workers —typically wage workers— may be more prone to injuries at work. We believe this is not driving our results: when we break down hospitalizations by cause, our results seem to be driven mainly by diseases of the digestive and circulatory systems and not diseases of the musculoskeletal system.

It is also important to note that in Portugal, unlike for example in the US, the National Health System ensures that anyone who needs it can get hospital care for free —i.e. an individual will not avoid going to the hospital because he or she does not have private (own or employer-purchased) insurance. A related premiss is that the self-employed, having more limited access to sickness benefits, could delay appropriate care until they are seriously sick and have to be hospitalized. This would imply that self-employment leads to higher rate of hospitalization, which is the opposite of what we find.

Although hospital admissions and mortality may correspond to acute or extreme sickness events, they are also highly objective and comparable outcomes, at least in our context in which we use

administrative data. Nevertheless, future research should also explore other health outcomes, again using data and methods that allow the identification of a causal effect of self-employment.

In conclusion, this study provides causal evidence of a positive effect of self-employment on health. Such impact may be at least partly explained by greater control by the individual over different aspects of the working life associated with this form of employment. One important dimension of the ongoing debate about the ‘future of work’ is precisely how to increase protection for workers under flexible contracts, such as those that increasingly emerge under the ‘gig economy’ (e.g. Garben 2017, European Commission 2017). This may involve multiple policy dimensions such as social security, employment law and collective bargaining. Our results indicate that, despite the existing concerns, at least as far as health is concerned there may be important social gains from more flexible work formats.

### **Acknowledgements**

We thank financial support from the European Union (grant VS/2016/0340) and access to data provided by the Portuguese Social Security and Ministry of Health. We appreciate comments from Claudio Lucifora, Elena Cottini, Georgios Kavetsos, João Pereira dos Santos, Kevin Denny, Pedro Pita Barros, Rudolf Winter-Ebmer, Twitter followers, and participants at a Nova Healthcare Initiative seminar (Nova School of Business and Economics), a CoBExt meeting (Queen Mary University of London), and at the European Health Economics Association 2018 conference.

### **References**

- Baron, R. A., Franklin, R. J., Hmieleski, K. M. 2016. Why entrepreneurs often experience low, not high, levels of stress: the joint effects of selection and psychological capital. *Journal of Management*, 42, 742–768.
- Benz, M., Frey, B. S. 2008. Being independent is a great thing: Subjective evaluations of self-employment and hierarchy. *Economica*, 75, 362–383.
- Bloemen, H., Hochguertel, S., Zweerink, J. 2018. Job loss, firm-level heterogeneity and mortality: Evidence from administrative data. *Journal of Health Economics*, 59, 78–90.
- Browning, M., Heinesen, E. 2012. Effect of job loss due to plant closure on mortality and hospitalization. *Journal of Health Economics*, 31, 599–616.
- Case, A., Fertig, A., Paxson, C. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24, 365–389.
- Case, A., Paxson, C. 2010. Causes and consequences of early-life health. *Demography*, 47, S65–S85.
- Centeno, M., Duarte, C., Novo, A. A. 2011. The impact of the minimum wage on low-wage earners. *Banco de Portugal Economic Bulletin*, 107–121.

- Correia, S. 2016. A Feasible Estimator for Linear Models with Multi-Way Fixed Effects. Working paper.
- Cottini, E., Ghinetti, P. 2017. Employment insecurity and employees' health in Denmark. *Health Economics*, 1–14.
- Dobkin, C., Finkelstein, A., Kluender, R., Notowidigdo, M. The economic consequences of hospital admissions. *American Economic Review* (Forthcoming).
- European Commission. 2017. Proposal for a directive of the European Parliament and of the Council on transparent and predictable working conditions in the European Union.
- Eurostat. 2017. Employment by sex, age, and professional status.
- Garben, S. 2017. Protecting workers in the online platform economy: an overview of regulatory and policy developments in the EU. Luxembourg, European Agency for Safety and Health at Work.
- Gielnik, M. M., Zacher, H., Frese, M. 2012. Focus on opportunities as a mediator of the relationship between business owners' age and venture growth. *Journal of Business Venturing*, 27, 127–142.
- Gonçalves, J., Martins, P. S. 2018a. A multidimensional analysis of the effects of minimum wages.
- Gonçalves, J., Martins, P. S. 2018b. Does self-employment impact fertility? Evidence from longitudinal social security data.
- Halpern, D. F. 2005. How time-flexible work policies can reduce stress, improve health, and save money. *Stress and Health*, 21, 157–168.
- Hessels, J., Rietveld, C. A., Zwan, P. V. D. 2017. Self-employment and work-related stress: the mediating role of job control and job demand. *Journal of Business Venturing*, 32, 178–196.
- Holmgren, K., Dahlin-Ivanoff, S., Björkelund, C., Hensing, G. 2009. The prevalence of work-related stress, and its association with self-perceived health and sick-leave, in a population of employed Swedish women. *BMC Public Health*, 9, 1–10.
- Ichino, A., Riphahn, R. T. 2005. The effect of employment protection on worker effort: absenteeism during and after probation. *Journal of the European Economic Association*, 3, 120–143.
- ILO. 2016. Non-standard employment around the world: Understanding challenges, shaping prospects. Geneva, International Labour Office, 374.
- Karasek, R. A. 1979. Job demands, job decision latitude, and mental strain: implications for job redesign. *Administrative Science Quarterly*, 24, 285–308.
- Karasek, R. A., Theorell, T. 1990. *Healthy work: stress, productivity, and the reconstruction of working life*. New York, Basic Books.

- Koellinger, P. D., Thurik, A. R. 2012. Entrepreneurship and the business cycle. *The Review of Economics and Statistics*, 94, 1143–1156.
- Martins, P. S. 2016. Working to get fired? Regression discontinuity effects of unemployment benefit eligibility on prior employment duration. *IZA Discussion Paper Series*, 10262.
- Mayer, E. A. 2000. The neurobiology of stress and gastrointestinal disease. *Gut*, 47, 861–869.
- Noseleit, F. 2014. Female self-employment and children. *Small Business Economics*, 43, 549–569.
- Pesole, A., Urzì Brancati, M., Fernández-Macías, E., Biagi, F., González Vázquez, I. 2018. Platform workers in Europe. Evidence from the COLLEEM Survey. Luxembourg: Publications Office of the European Union, 2018.
- Prottas, D. J., Thompson, C. A. 2006. Stress, satisfaction, and the work-family interface: a comparison of self-employed business owners, independents, and organizational employees. *Journal of Occupational Health Psychology*, 11, 366–378.
- Rietveld, C. A., Kippersluis, H., Thurik, A. R. 2015. Self-employment and health: barriers or benefits?. *Health Economics*, 24, 1302–1313.
- Riphahn, R. T. 2004. Employment protection and effort among German employees. *Economic Letters*, 85, 353–357.
- Schneck, S. 2014. Why the self-employed are happier: Evidence from 25 European countries. *Journal of Business Research*, 67, 1043–1048.
- Stephan, U., Roesler, U. 2010. Health of entrepreneurs versus employees in a national representative sample. *Journal of Occupational and Organizational Psychology*, 83, 717–738.
- Stephoe, A., Kivimäki, M. 2012. Stress and cardiovascular disease. *Nature Reviews Cardiology*, 9, 360–370.
- Theorell, T., Karasek, R. A. 1996. Current issues relating to psychosocial job strain and cardiovascular disease research. *Journal of Occupational Health Psychology*, 1, 9–26.
- Toivanen, S., Griep, R. H., Mellner, C., Vinberg, S., Eloranta, S. 2016. Mortality differences between self-employed and paid employees: a 5-year follow-up study of the working population in Sweden. *Occupational and Environmental Medicine*, 73, 627–636.
- Yoon, J., Bernell, S. L. 2013. The effect of self-employment on health, access to care, and health behavior. *Health*, 05, 2116–2127.
- Zissimopoulos, J. M., Karoly, L. A. 2007. Transitions to self-employment at older ages: the role of wealth, health, health insurance and other factors. *Labour Economics*, 14, 269–295.

## Appendix

Figure A1: Numbers of switches to and from self-employment, Jan 2005-Dec 2011

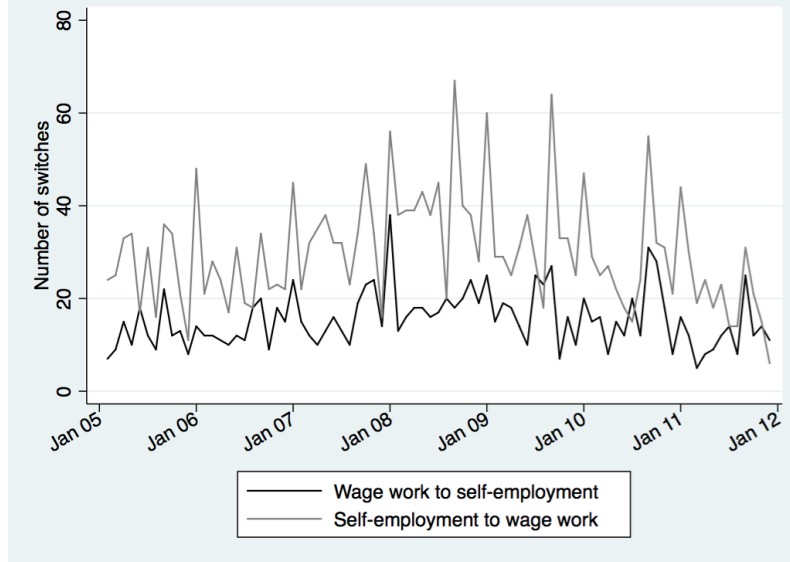


Figure A2: Probability of hospitalization in the following month by type of employment, Jan 2005-Dec 2011

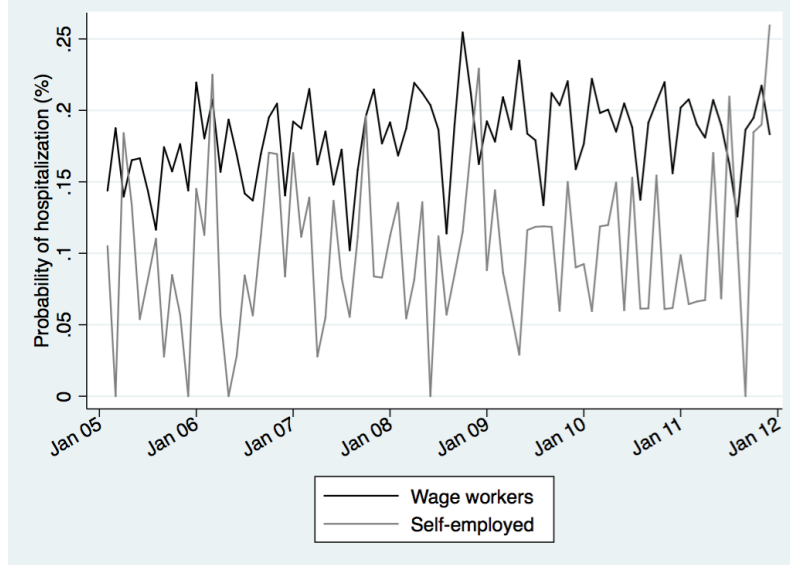




Figure A3: Duration of hospitalization by type of employment, Jan 2005-Dec 2011

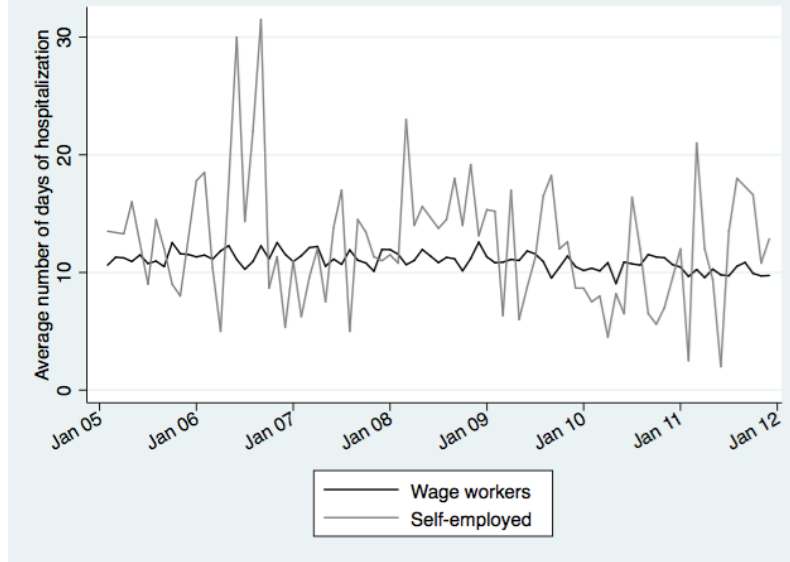


Figure A4: Probability of standard sick leave in the following month by type of employment, Jan 2005-Dec 2011

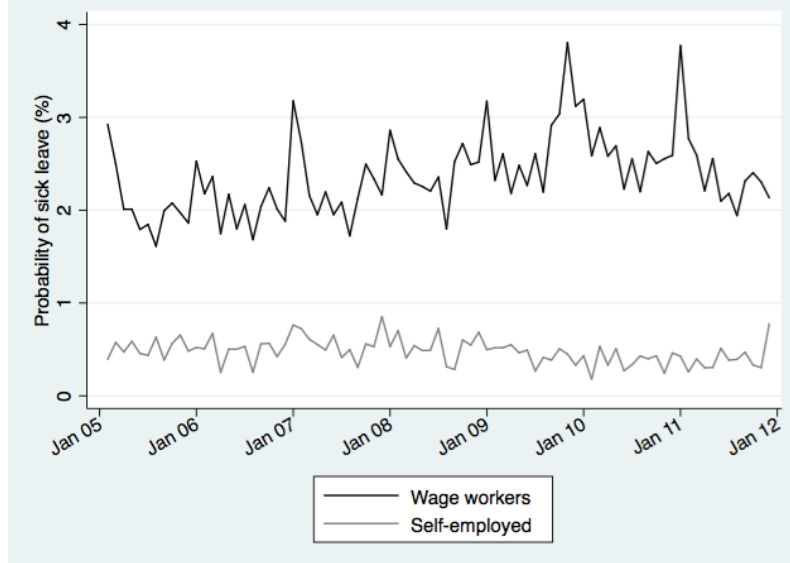


Figure A5: Duration of standard sick leave by type of employment, Jan 2005–Dec 2011

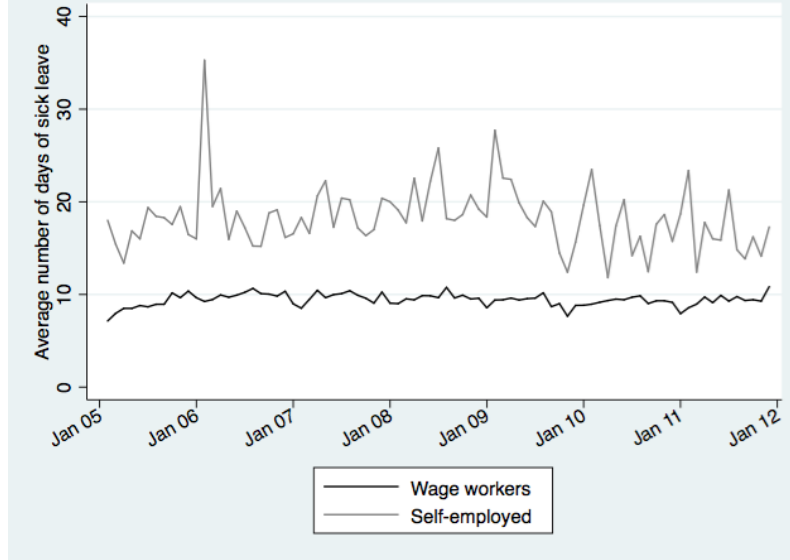


Table A1: First-stage results

	Model 4	Model 5	Model 6
<b>Likelihood of hospitalization</b>			
$\ln(\text{Individual income/Minimum wage})$	-0.03409*** (0.00145)	-0.03401*** (0.00144)	–
$\ln(\text{Proportion of self-employed})$	0.00999*** (0.00056)	–	0.01026*** (0.00048)
F-test	F=383.33***	F=559.49***	F=467.00***
Observations	5,231,898	5,277,454	6,583,929
<b>Length of hospitalization</b>			
$\ln(\text{Individual income/Minimum wage})$	-0.02617 (0.01934)	-0.02717 (0.02157)	–
$\ln(\text{Proportion of self-employed})$	0.02475 (0.02290)	–	0.01102 (0.01065)
F-test	F=1.00	F=1.59	F=1.07
Observations	2,460	2,486	3,124

Clustered standard errors in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table A2: Effect of self-employment on the length of hospitalization

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
Self-employed (t-1)	0.10701** (0.04898)	0.07786 (0.04960)	0.26538 (0.39026)	1.63309 (2.93196)	8.06573 (8.64859)	-6.15287 (8.52222)
Female	–	-0.07668*** (0.01538)	–	–	–	–
26-35 years old	–	-0.05171* (0.02827)	0.17236 (0.17382)	-0.00113 (0.15942)	0.11307 (0.14816)	0.10683 (0.20871)
36-45 years old	–	0.02656 (0.02719)	0.20871 (0.21184)	-0.08104 (0.22055)	0.07074 (0.22031)	0.13610 (0.24930)
46-55 years old	–	0.06267** (0.02849)	0.22283 (0.23702)	0.00783 (0.27034)	0.18657 (0.27796)	0.13445 (0.27556)
56-65 years old	–	0.02103 (0.03806)	0.25177 (0.28106)	-0.18803 (0.34426)	0.05560 (0.36500)	0.11742 (0.32873)
Foreign	–	-0.02032 (0.02940)	–	–	–	–
Time FE	–	Yes	Yes	Yes	Yes	Yes
Individual FE	–	–	Yes	Yes	Yes	Yes
Constant	2.15763*** (0.00781)	2.21152*** (0.07078)	1.94478*** (0.19067)	–	–	–
Instrument strength	–	–	–	F=1.00	F=1.59	F=1.07
Instrument validity	–	–	–	$\chi^2=1.424$	–	–
Endogeneity	–	–	–	$\chi^2=0.138$	$\chi^2=1.635$	$\chi^2=1.644$
Observations	11,901	11,901	3,184	2,460	2,486	3,124

Clustered standard errors in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table A3: Effect of self-employment on the likelihood of standard sick leave

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
Self-employed (t-1)	-0.01888*** (0.00021)	-0.01944*** (0.00022)	-0.01243*** (0.00049)	-0.25866*** (0.01419)	-0.30146*** (0.01669)	0.01301 (0.01942)
Female	-	0.01189*** (0.00020)	-	-	-	-
26-35 years old	-	0.00000 (0.00026)	-0.00006 (0.00040)	-0.00002 (0.00057)	0.00001 (0.00058)	-0.00001 (0.00041)
36-45 years old	-	0.00256*** (0.00030)	-0.00213*** (0.00057)	-0.00148* (0.00077)	-0.00139* (0.00080)	-0.00208*** (0.00058)
46-55 years old	-	0.00530*** (0.00034)	-0.00175** (0.00079)	-0.00063 (0.00104)	-0.00033 (0.00107)	-0.00181** (0.00081)
56-65 years old	-	0.00578*** (0.00043)	0.00087 (0.00107)	0.00102 (0.00140)	0.00118 (0.00144)	0.00091 (0.00108)
Foreign	-	-0.00318*** (0.00029)	-	-	-	-
Time FE	-	Yes	Yes	Yes	Yes	Yes
Individual FE	-	-	Yes	Yes	Yes	Yes
Constant	0.02368*** (0.00011)	0.02131*** (0.00062)	0.02425*** (0.00045)			
Instrument strength	-	-	-	F=383.43***	F=558.48***	F=467.76***
Instrument validity	-	-	-	$\chi^2=82.735$ ***	-	-
Endogeneity	-	-	-	$\chi^2=587.058$ ***	$\chi^2=692.381$ ***	$\chi^2=1.721$
Observations	6,706,903	6,706,903	6,705,126	5,254,049	5,299,381	6,610,453

Clustered standard errors in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table A4: Effect of self-employment on the likelihood of standard sick leave (31+ days)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
All lengths (N=6,706,903)	-0.01888*** (0.00021)	-0.01944*** (0.00022)	-0.01243*** (0.00049)	-0.25866*** (0.01419)	-0.30146*** (0.01669)	0.01301 (0.01942)
Lengths of 31+ days (N=6,554,587)	-0.00007*** (0.00002)	-0.00010*** (0.00002)	-0.00008** (0.00003)	-0.00392*** (0.00090)	-0.00522*** (0.00101)	0.00209 (0.00136)

Numbers of observations lower in FE and IV-FE models. Clustered standard errors in parentheses.

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table A5: Effect of self-employment on the likelihood of hospitalization: robustness checks

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	SLR	MLR	FE	IV-FE	IV-FE	IV-FE
<b>Main results (N=6,680,530)</b>						
Self-employed (t-1)	-0.00081*** (0.00007)	-0.00104*** (0.00007)	-0.00079*** (0.00014)	-0.01686*** (0.00281)	-0.01827*** (0.00293)	-0.00467 (0.00548)
<b>Balanced panel (N=1,535,832)</b>						
Self-employed (t-1)	-0.00055*** (0.00015)	-0.00072*** (0.00015)	-0.00125* (0.00072)	-0.08548*** (0.03063)	-0.08552*** (0.03077)	-0.63975 (1.12757)
<b>Balanced panel, including periods of non employment (N=10,694,899)</b>						
Self-employed (t-1)	-0.00068*** (0.00006)	-0.00081*** (0.00006)	-0.00057*** (0.00008)	-0.01389** (0.00550)	–	–
Not employed (t-1)	-0.00131*** (0.00002)	-0.00130*** (0.00002)	-0.00121*** (0.00003)	-0.00980*** (0.00270)	–	–
<b>Quarterly data (N=2,260,107)</b>						
Self-employed (t-1)	-0.00224*** (0.00020)	-0.00290*** (0.00020)	-0.00230*** (0.00041)	-0.05211*** (0.00808)	-0.05528*** (0.00842)	-0.02206 (0.01607)

Numbers of observations lower in FE and IV-FE models. Clustered standard errors in parentheses.

\* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$