Small Business Economics

Effects of self-employment on hospitalizations: instrumental variables analysis of social security data

--Manuscript Draft--

Manuscript Number:	SBEJ-D-19-00522R2
Full Title:	Effects of self-employment on hospitalizations: instrumental variables analysis of social security data
Article Type:	Original Research
Keywords:	Self-employment; hospitalization; mortality; small businesses
Corresponding Author:	Judite Gonçalves Universidade Nova de Lisboa Nova School of Business and Economics Carcavelos, PORTUGAL
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidade Nova de Lisboa Nova School of Business and Economics
Corresponding Author's Secondary Institution:	
First Author:	Judite Gonçalves
First Author Secondary Information:	
Order of Authors:	Judite Gonçalves
	Pedro S. Martins
Order of Authors Secondary Information:	
Funding Information:	European Union (VS/2016/0340) — Prof. Pedro S. Martins
Abstract:	The importance of self-employment and small businesses raises questions about their health effects and public policy implications, which can only be addressed with suitable data. We explore the relationship between self-employment and health by drawing on comprehensive longitudinal administrative data to explore variation in individual work status and by applying novel instrumental variables. We focus on an objective outcome —hospital admissions— that is not subject to recall or other biases that may affect previous studies. Our main findings, based on a sample of about 6,500 individuals followed monthly from 2005 to 2011 and who switch between self-employment and wage work along that period, suggest that self-employment has a positive effect on health as it reduces the likelihood of hospital admission by at least half.

Title Page (With all author contact info)

Effects of self-employment on hospitalizations: instrumental variables analysis of social security data

Judite Gonçalves*
Nova School of Business and Economics, Universidade Nova de Lisboa, Campus de Carcavelos, Rua da Holanda 1, 2775-405 Carcavelos, Portugal

AND

Pedro S. Martins
School of Business and Management, Queen Mary University of London, Mile End Road, London E1 4NS, UK
Nova School of Business and Economics, Universidade Nova de Lisboa, Campus de Carcavelos, Rua da Holanda 1, 2775-405 Carcavelos, Portugal
Institute of Labor Economics (IZA), 53072 Bonn, Germany
Global Labor Organization

Acknowledgments
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. The authors also thank funding from the European Union (CoBExt action, grant VS/2016/0340). This study reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information that the study contains.

*Correspondence to: Nova School of Business and Economics, Universidade Nova de Lisboa, Campus de Carcavelos, Rua da Holanda 1, 2775-405 Carcavelos, Portugal. E-mails: judite.goncalves@novasbe.pt, judite_goncalves@msn.com. Phone: +351 913 374 597.

Effects of self-employment on hospitalizations: Instrumental variables analysis of social security data

Abstract

The importance of self-employment and small businesses raises questions about their health effects and public policy implications, which can only be addressed with suitable data. We explore the relationship between self-employment and health by drawing on comprehensive longitudinal administrative data to explore variation in individual work status and by applying novel instrumental variables. We focus on an objective outcome —hospital admissions— that is not subject to recall or other biases that may affect previous studies. Our main findings, based on a sample of about 6,500 individuals followed monthly from 2005 to 2011 and who switch between self-employment and wage work along that period, suggest that self-employment has a positive effect on health as it reduces the likelihood of hospital admission by at least half.

Keywords: Self-employment; hospitalization; mortality; small businesses
JEL classification: C26, I18, J24, L26

1 Introduction

The self-employed represent nearly 16% of employment in the European Union (Eurostat, 2017). Moreover, as many as 10% of the adult population of the EU has used online platforms for the provision of labor services at some point in their lives [Pesole et al., 2018]. The ongoing growth of the 'platform' economy contributes to the expansion of the proportion of self-employed, especially among younger workers, and raises a number of public policy questions regarding, for example, occupational health and safety risks, social protection, and representation (European Commission, 2017; Garben, 2017; ILO, 2016). Indeed, platform economy jobs —and self-employment more generally as well as some types of small businesses— are characterized by more flexible work formats, distinct from formal employer-employee relationships framed by employment law, and typically have more limited access to social protection.[1]

In the current context of such novel forms of self-employment, one important issue concerns the impact of self-employment on workers' health —the subject of this study. Occupational characteristics, namely job control and job demand, vary significantly between self-employment and wage work. 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 considerable amount of work and/or having little time to carry out specific tasks. This Job Demand-Job Control framework, proposed by Karasek [1979], 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, an interaction termed 'job strain' in the literature (Prottas and Thompson, 2006; Stephan and Roesler, 2010).[2]

In fact, self-employed individuals are not subject to orders from other workers 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

[1] Note that the Covid19 crisis and its aftermath may contribute to the growth of self-employment, as wage employment opportunities in the labour market will decrease. Additionally, the Covid19 crisis may lead to a larger share of wage employment conducted under remote work formats given their social/physical distancing properties. Such remote work formats are typically more common amongst the self-employed, which may lead to some blurring of the differentiation between wage work and self-employment.

[2] See e.g. Ingre [2017] for a discussion of the job strain model with respect to the appropriateness of the interaction between the job demand and job control dimensions in Karasek's model.

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, a process sometimes referred to as 'procedural utility' (Benz and Frey, 2008; Schneck, 2014).

However, when self-employed, labour income and assets directly hinge on one's ability to work and work effort in each period. In addition, greater exposure to unanticipated demand shocks leaves self-employed individuals subject to more volatile workload and income flows. Social support at work may also be more limited given the smaller number of co-workers around (Blanch, 2016, discusses the Demand-Control-Support model). All these variables represent sources of stress. Given these two opposite mechanisms —higher job demand and higher job control—, it is unclear whether we should expect self-employed individuals to suffer from more or less work-related stress, compared to wage workers.

The medical literature identifies stress as an important cause of disease, e.g. cardiovascular problems and digestive disorders (Mayer, 2000; Steptoe and Kivimäki, 2012). Overall, stress impacts negatively on health and well-being, and in addition to increasing incidence of disease, it may increase absence from work due to sickness and use of health care services (e.g. Browning and Heinesen, 2012; Halpern, 2005; Holmgren et al., 2009). 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. Stress is also associated with unhealthy behavior, such as smoking and drinking.

The typical occupations of self-employed and wage workers may differ in terms of risk of workplace accidents and other occupational hazards. At the same time, in many countries self-employment is subject to little or no social protection, in terms of coverage by occupational safety regulation, social security, employment law, or collective bargaining, potentially representing additional negative implications for health. On the other hand, the greater flexibility regarding regulation may also represent additional work opportunities compared to wage work. Overall, whether self-employment has a positive or detrimental effect on health is a public policy question that can only be answered with 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 [Torrès and Thurik, 2019]. 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 to 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 [Bujacz et al., 2019]. 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 along dimensions not observable in the data.

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 to a broader working-age population.

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. Nikolova [2018], using German longitudinal survey data and a difference-in-differences strategy, finds that switching from wage work to self-employment leads to both physical and mental health gains.

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 who own limited liability companies (but not sole proprietors) have lower average risk of mortality than wage workers [Toivanen et al., 2016]. Similarly, Toivanen et al. [2018] find that limited liability company owners have lower rates of hospitalization for myocardial infarction than wage workers, and no different hospitalization rates for stroke. The authors unveil relevant heterogeneous effects not only by enterprise legal type of self-employed individuals but also by industry.

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. Furthermore, it is important to distinguish the effect that is due to differences in the intrinsic characteristics of self-employment and wage work, namely job control and job demand, from institutional factors such as different access to social security benefits. This may be difficult with survey data and self-reported health indicators. Separating-out the effect that is due to differences in the typical occupations of self-employed and wage workers, which are

associated with different exposure to occupational hazards, would also be of interest.

The main research question in this study is "What is the impact of self-employment on the likelihood of hospital admission?" We answer this question based on a large sample of administrative social security records representative of the working-age population in Portugal, that includes almost 130,000 self-employed and wage workers followed between January 2005 and December 2011. We focus on a subsample of about 6,500 individuals who switch between self-employment and wage work along that 84-month period. We contribute to the literature in several ways. First, we tackle explicitly the endogeneity of the decision to become self-employed by controlling for individual fixed effects and employing instrumental variables. Second, looking at hospitalizations allows us to separate-out institutional factors, because access to hospital care and social security benefits when hospitalized are unrelated with type of employment, and most hospitalizations correspond to unplanned or unavoidable acute events. Administrative records of hospital admissions are also comparable across individuals and time periods and not subject to recall bias, an advantage over self-reported indicators in survey data. Third, to explore to which extent the effect may be due to differences in the typical occupations of self-employed and wage workers, we look at diagnoses underlying hospitalizations. Fourth, we consider the whole working population regardless of age, and explore potentially heterogeneous effects across demographic subgroups. Lastly, we also investigate the effects of self-employment on the length of hospitalization and mortality.

Hospital admissions are also a relevant outcome for policy. They represent roughly 40% of health expenditure in Portugal.[3] A significant 7% of sickness leave episodes correspond to hospital admissions (own calculations for the years 2005-2011). In 2011, sickness leave episodes cost Social Security 454 million euros;[4] 7% of that represents almost 32 million euros. This adds to the costs for the health system and other societal costs more difficult to quantify, including productivity and well-being losses.

The remaining of this paper is as follows: the next section lays down the background for the study, Section 3 presents our data and empirical strategy, Section 4 presents the results, and in

[3] https://www.pordata.pt/Portugal/Despesa+corrente+em+cuidados+de+saúde+total+e+por+tipo+de+prestador-2958

[4] https://www.pordata.pt/Portugal/Despesa+da+Segurança+Social+com+alguns+subs%C3%ADdios+à+população+activa-116

Section 5 we discuss our findings.

2 Institutional background

In 2016, about 17% of employment in Portugal corresponded to self-employment or own-account workers. More than one fourth of those had employees. The proportion of own-account workers differs across groups. It is lower in the capital region than in other regions, among women, among younger age groups, and among more educated groups. By industry, we find the largest proportions of own-account workers in agriculture and other primary sector activities (71.5%), real estate (36%), consulting, scientific, and technical activities (29.5%), construction (27.4%), retail (21.3%), hospitality services (20.3%), and artistic and sports activities (19.6%). From the "Self-employment" module of the Labor Force Survey (LFS), conducted in the second quarter of 2017, we also know that more than 60% of own-account workers decide their work schedule. They also report much higher autonomy over their tasks than wage workers. This is in line with the hypothesis of higher job control. While only less than 20% of own-account workers report no difficulties with their work over the previous 12 months, 16% report periods without work, and 14% claim that clients do not pay or pay late. This may suggest that the self-employed are subject to higher job demand. Own-account workers report lower levels of satisfaction at work than wage workers, although this is driven by the low satisfaction levels of those who have employees. Virtually no own-account workers report that they would prefer to be wage workers [Torres and Raposo, 2018].

In this study, we adopt the Portuguese Social Security definition of self-employment or 's independent workers' (*trabalhadores independentes*), which does not include own-account workers with employees. Family and informal workers, which are captured in the LFS, do not appear in our data, as they do not pay social security contributions. This explains the lower proportion of self-employment in our data, described below, compared to the proportion of own-account workers in the LFS. For example, agriculture and other primary sector activities, which have by far the largest proportion of own-account workers in the LFS, will have limited expression in our data for those reasons.

In Portugal, statutory sick leave covers both the 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 a sickness episode until a sickness benefit starts to be paid (i.e. waiting or 'elimination' period). However, for the self-employed, this waiting period is much longer, at thirty days (ten days from 2018 onwards).

Due to the different waiting periods, social security records include sickness episodes that last four days or more in the case of wage workers, but 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 other things equal, the sickness spells of the self-employed that are administratively recorded are, on average, much more selected and severe.

These different waiting periods can 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. In some cases, collective bargaining provisions, determined by unions and firms or employer associations, may even lead to the payment (by the firms) of the first three days of absence as well. As these provisions apply to wage workers but not to the self-employed, the former may engage more often in moral hazard: 'cheat' by going on sick leave when they are not really sick.

In stark contrast, there is no waiting period for either self-employed or wage workers in the case of hospitalization. Furthermore, benefits are the same for both types of workers. Besides, due to the specific, acute nature of hospitalizations, these 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 outcome and hospital admissions should be strictly comparable between wage workers and self-employed individuals.

As to the amount of the support, 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^{st} to the 365^{th} day, and 75% from the 365^{th} day onwards. 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 31st to the 90th day. Sickness benefits are granted for a maximum of 1,095 days for wage workers and 365 days for self-employed individuals (Law-decrees 28/2004, 133/2012 and 146/2005).

The Portuguese National Health Service, financed through taxes, provides general and universal coverage and is almost free at the point of use. In Portugal, secondary and tertiary care (both acute and post-acute care) is mainly provided in hospitals. General practitioners act as gatekeepers in access to hospital care in the public sector; otherwise people can be admitted through the emergency department. Private voluntary health insurance may speed up access to elective hospital treatment and ambulatory consultations, but it has very limited expression in Portugal (<10%) and is not associated with type of work (i.e. self-employment or wage work). Some public and private subsystems provide care to specific groups not relevant for this study (public servants, military, banking sector workers).

In general, access to hospital care in Portugal should be identical for both self-employed and wage workers. The only concern is that self-employed individuals may delay care in order not to lose business, as their income is closely tied to them actually working. (Wage workers could also delay care in order to maintain a good reputation with their employer.) Because we are looking at hospitalizations, which are generally acute, untimed events, this concern is limited. Non-emergency acute interventions are scheduled by the hospital, and because waiting lists are usually long, it is unlikely that individuals pass on the opportunity to receive the care they need when hospitals schedule them, as it may be a long time before a new opportunity arises.

3 Data and methods

3.1 Social security data

We use data from the Portuguese Social Security Information System, made available by the *Instituto de Informática* public agency. The dataset is a random sample such that included individuals represent both a) at least 1% of all individuals who pay Social Security contributions and b) at least 1% of all individuals who receive sickness, maternity, or other benefits from Social

Security, stratified by region and gender.

We observe 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 receive sickness benefits in a specific month due to hospitalization. The data allow us to distinguish sickness benefits due to hospitalization from sickness benefits due to standard (non-acute) sickness spells, as the two cases are treated differently by Social Security (see Section 2). The dataset also includes information on the individuals' gender, age, nationality, place of residence, and income from work, but not their industry or occupation.

We drop individuals below 18 and above 65 years old (mandatory schooling age and statutory retirement age). After deleting also observations with missing information on the key variables, we are left with almost 130,000 individuals, of which about 10,000 are self-employed at some point over the period 2005-2011. In our main analyses, we focus on more than 6,500 individuals who switch at least once between self-employment and wage work over that period (which we refer to as switchers). Over the 84 month-long period, there are more than 300,000 individual-month observations when considering only switchers (almost 7 million individual-month observations in the full sample).

3.2 Identification and empirical strategy

To determine the effect of self-employment on the likelihood of hospitalization, we estimate four different specifications of a linear probability model like the following:[5]

$$\text{hosp}_{i,t} = \beta_0 + \beta_1 \text{self-employed}_{i,t-1} + \gamma X_{i,t} + \tau_t + \mu_i + \varepsilon_{i,t} \tag{1}$$

The binary dependent variable, $\text{hosp}_{i,t}$, indicates whether individual i is hospitalized in month t or not. The variable of main interest is the one-month lag of the self-employment indicator, $\text{self-employed}_{i,t-1}$, which takes value one if individual i is self-employed in month $t-1$.[6]

[5] We opt for the linear probability model given the computational difficulties associated with applying instrumental variables methods to nonlinear panel data models, especially when various large vectors of fixed effects are included. To investigate if the chosen functional form is appropriate, we estimated the logit/panel logit versions of Models 1 and 2 (i.e., with or without individual fixed effects), which provided marginal effects similar to the ones obtained with the linear versions.

[6] Some individuals who receive income from both self-employment and wage work in some months are counted

Using the one-month lag of the self-employment indicator, or all lags up to the third or the twelfth, for example, gives estimated total effects of self-employment with the same sign and level of statistical significance, differing only slightly in magnitude. This shows the stability of the self-employment indicator, as individuals rarely change type of work more than once over the seven-year period considered. 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 lag, given the high correlation between adjacent lags. Furthermore, using more than one lag would result in many more observations being lost. In conclusion, β_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.

The four specifications that we consider are the following:

Model 1 controls for the individual's gender, age group (18-25, 26-35, 36-45, 46-55, or 56-65), nationality (Portuguese or foreign), and place of residence (one of the 18 districts in the mainland or one of the 11 islands),[7] included in $X_{i,t}$. We also include fixed effects for each month in the sample, denoted by τ_t (84 months minus Jan 2005, due to the lag, and Feb 2005, which is the reference month).

Model 2 takes advantage of the longitudinal nature of the dataset and includes also individual fixed effects, denoted by μ_i, to control for time-invariant individual unobserved heterogeneity.

Still, it is possible that endogeneity due to unobserved individual characteristics that vary over time remains, as discussed in the Introduction. To tackle this potential threat, in addition to the individual fixed effects, we employ an instrumental variable strategy. Thus,

Model 3 applies instrumental variables without controlling for individual fixed effects (i.e. instrumental variable estimation of Model 1), and

Model 4 applies instrumental variables controlling for individual fixed effects (i.e. instrumental variable estimation of Model 2).

In sum, Models 1 and 3 treat the data as pooled cross-sections, whereas Models 2 and 4 are fixed effects panel data models; Models 3 and 4 apply an instrumental variable strategy.

We use two instruments. Instrument one is the proportion of self-employed workers in individual

as self-employed. Excluding these observations provides almost identical results.

[7] For simplicity, we only refer to districts throughout the text, i.e. each island is treated as a district.

i's district, excluding her municipality of residence, in the same month (see Online Resource 1 for the division of the Portuguese territory into districts and municipalities). Instrument two is the proportion of self-employed workers of the same gender and age group of individual i in the whole country, also excluding her municipality of residence, in the same month.

The proportion of workers in a given district or gender-age group who are self-employed captures the structure of the labor market in that area or demographic group. For example, there may be a predominant industry in a given district that relies on wage workers, or there may be a new service expanding where young self-employed women abound. In general, we expect that the larger that proportion, the higher the likelihood that any individual i residing in district j or belonging to gender-age group m is self-employed. However, in some cases, low self-employment in the district/demographic group may signal opportunities or conversely, high self-employment may signal a saturated market. That is, some individuals may be defiers, responding in the opposite way to a higher proportion of self-employed workers in the district/demographic group (i.e. violation of the monotonicity assumption). When there are defiers, the two-stage least squares estimator gives a weighted difference between the effect of the treatment among compliers and defiers, which could be misleading. Nevertheless, de Chaisemartin [2017] derives a weaker condition under which the two-stage least squares estimator still provides a local average treatment effect (LATE) for 'surviving compliers'. With binary outcomes, like is our case, that condition holds if defiers' LATE and the two-stage least squares coefficient are both of the same sign, or if defiers' and compliers' LATEs are both of the same sign and the ratio of these two LATEs is lower than the ratio of the shares of compliers and defiers in the population. In this context, it is difficult to assess if that condition is likely to hold, because the effect of self-employment on the likelihood of hospitalization can be positive or negative. Still, we see no reason for the LATEs of compliers and defiers to differ significantly, especially since fixed effects capture individuals' intrinsic characteristics that may explain why they respond differently to the instruments. So, we argue that the condition holds as the ratio of compliers to defiers should exceed the ratio of the two LATEs.[8]

[8]We also explore non-linearities in the first-stage equation by including the squared terms of the instruments. In the case of instrument two, the squared term is not statistically significant. In the case of instrument one, the squared term is negative and statistically significant, but results are virtually unchanged, so we keep the simpler model.

The proportion of self-employed workers in an individual's geographical area has previously been used to instrument self-employment decisions (e.g. Noseleit, 2014). The novelty here is that instead of considering the proportion of self-employed workers in the individual's municipality, we consider only neighboring municipalities excluding the individual's own. This approach to devise instrumental variables has been employed e.g. in Autor et al. [2013] and Nevo [2001]. In both our instruments, the exclusion of the individual's own municipality contributes to eliminate concerns regarding instrument exogeneity. Overall, we believe our instruments are validly excluded from the main equation conditional on the remaining explanatory variables (i.e. they impact hospitalizations solely through their impact on the likelihood of self-employment). For instance in the case of the proportion of self-employed workers in the district (instrument one), the crucial explanatory variables are the district fixed effects. District fixed effects take into account any district characteristics that correlate with both the instrument and the outcome, hospitalizations, as long as those characteristics are constant over time. To explore this issue further, we look at the evolution over time of some district characteristics: a general income index, a general health index, and a firm dimension index, which are composite indices produced by a Portuguese polling firm, Marktest (Online Resource 2). What we observe is that all of those indices are fairly constant over time; therefore such characteristics should be appropriately captured by the district fixed effects. Note also that by comparison, the proportion of self-employed workers in the district exhibits some within-district variation, so the instrument is relevant even when controlling for district fixed effects (Online Resource 2). With two instruments and one potentially endogenous variable, we are able to test statistically the validity of the overidentifying restriction. Given that the endogeneous variable, self-employed$_{i,t-1}$, is lagged, we also use the lags of the instruments.

As mentioned previously, our main analyses focus on the subsample that includes only individuals who switch between wage work and self-employment at least once over the sample period ('switchers'). After all, those are the individuals that are used for identification in the models with individual fixed effects. Moreover, in the instrumental variables model with individual fixed effects, non-switchers are by definition non-compliers, and non-compliers reduce the instruments' statistical power [de Chaisemartin, 2017]. We also present results for all model specifications for

the entire sample, for comparison. Lastly, standard errors are robust to heteroscedasticity and to clustering at the individual level in Models 1 and 2, and at the district level in Models 3 and 4 (because that is the level of observation of instrument one).

3.3 Other analyses

The main time-varying unobserved individual characteristic that may affect both self-employment and the likelihood of hospitalization is health. Unfortunately, we do not have information on health status; only hospitalizations. We construct an indicator variable that takes value one if the individual had any hospitalization in the previous three months, to try to capture any recent (serious) changes in health status. This variable is potentially not enough to fully rule out endogeneity, which is why we resorted to instrumental variables models. Still, as a sensitivity check, we add this variable to Model 2 as a control.

We also compare the effect of self-employment on the likelihood of hospitalization for women versus men, individuals up to 35 versus 36 and more years old, and nationals versus foreigners. To do this, we include interaction terms between the lagged self-employment indicator and the respective demographic dummies. Since we have two instruments, we are able to instrument both the lagged self-employment indicator and the interaction term.

We repeat the main analyses 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.

To shed further light on the types of hospitalizations of self-employed and wage workers, we obtained information on hospitalizations from the national diagnosis-related groups dataset. This allowed us to learn the main diagnosis underlying each hospitalization as well as if it was planned or not, but only for about half of the hospitalizations in the social security dataset that we could match indirectly, as there is not an individual identifier to fully merge the two datasets. These complementary analyses are detailed in the Online Resource 3.

We also apply the model specifications described in the previous section to study the impact of self-employment on the length of hospitalization. First, in a two-part model type of approach, we

restrict the sample to individual-month observations with a hospitalization. We use the natural logarithm of hospitalization days as the dependent variable to account for the skewness in the distribution of hospitalization days. This approach drastically reduces the sample size. We compare the results, qualitatively, to those obtained for the full sample, using the natural logarithm of hospitalization days plus one in order to keep the zeroes.

Our data also allow 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 away in year $t+1$ and zero otherwise, while excluding observations for the year in which the person passes away. We compare results obtained when the self-employment indicator takes value one if the individual is self-employed during at least one, six, or all twelve months of year t. We estimate the same model specifications as described in the previous section, adjusted for the annual frequency considered here. Control variables are measured in year t.

4 Results

We discuss results for the subsample of individuals who switch at least once between self-employment and wage work over time ('switchers'). Results for the full sample are also presented for comparison, in the bottom half of the tables (Panel B).

4.1 Descriptive statistics

Descriptive statistics by type of employment in the previous month are shown in Table 1. Looking at the switchers, the self-employed account for 38.29% of the person-month observations (Panel A). The average monthly rates of hospitalization of self-employed and wage workers are 0.06% and 0.14% respectively. Note that these seemingly very low numbers correspond to monthly, not annual, hospitalization rates.

The average number of days of hospitalization, conditional on there being any, is slightly larger among the self-employed: 12.86 compared to 11.05 days for wage workers. The differences in the rates and lengths of hospitalization over time in both samples are shown in the Online Resource 4.

The proportion of women is slightly lower among the self-employed than among wage workers (50% versus 52%), the self-employed are on average slightly older (about 37 versus 36 years old), and the proportion of foreigners is also slightly lower among the self-employed (13% versus 14%). The proportion of self-employed workers in the district (instrument one) is on average 4.69% and varies between 0% and 17.74%. The proportion of self-employed workers in the same gender-age group in the country (instrument two) is on average 4.06% and varies between 0.75% and 17.34%.

4.2 Effects on hospital admissions

Table 2 shows the results of Models 1-4. Starting with the first stage results, we conclude that when the proportion of self-employment in the district or demographic group increases, the individual likelihood of self-employment also increases, as expected. Specifically, when the proportion of self-employed workers in a given district (/demographic group) is one percentage point higher, the likelihood of any individual in that district (/demographic group) becoming self-employed is about 7 (/4.5) percentage points higher, on average (Panel A, Model 4).[9]

Returning to why we focus on the subsample of switchers, as noted e.g. by de Chaisemartin [2017], non-compliers reduce the instruments' statistical power. In the instrumental variables model with individual fixed effects, non-switchers are by definition non-compliers. Judging from the large F- and t-statistics, the instruments appear strong when considering the full sample (Panel B, Model 4). However, looking at the second stage, we can see that the coefficient on the self-employment indicator is implausibly large in absolute terms, and has a huge standard error as well. This suggests that the instruments may actually not be strong enough even though the F- and t-statistics are above conventional thresholds.[10] Therefore, we focus our discussion on the results for the sample of switchers. Note that even in Model 2, which includes individual fixed effects but not instrumental variables, identification of the effect of self-employment also comes from switchers.

[9]These first-stage coefficients are much larger than the ones obtained when using the full sample (Panel B). This is because the switchers sample excludes all individuals that are always wage workers or always self-employed, i.e. individuals that don't respond to changes in the labor market as captured by the instruments. Recall that in the full sample, only 4.24% of the observations are self-employed; in the subsample of switchers, this proportion increases to 38.29% (Table 1). The difference in effect size of the instruments when looking at the full sample vs. the subsample of switchers can be interpreted in relation to these proportions of self-employment in each sample.

[10]In the model without individual fixed effects, instrument one actually has a small t-statistic and the F-statistic is also small (Panel B, Model 3).

As for the instrument validity test, the null hypothesis is not rejected in any case. There is also no evidence of endogeneity. In fact, the coefficients on the self-employment indicator in the instrumental variables models (Models 3 and 4) are very similar to the coefficients in Models 1 and 2, except they are less precisely estimated and not statistically significant (Panel A).[11] In light of this result, unobserved individual characteristics, in particular those that vary over time (e.g. health status), and reversed causality don't seem to pose an issue in our analyses. This is possibly because hospitalization is a fairly objective and a rare/extreme outcome, which doesn't capture health in general but serious (unexpected) manifestations of illness. Furthermore, the estimated coefficient on the self-employment indicator is about the same whether or not individual fixed effects are included (Model 1 versus Model 2), suggesting that self-selection of the healthy into self-employment has no impact on the negative association between self-employment and likelihood of hospitalization. Lastly, controlling for any hospitalization in the previous three months, which is another (*partial*) way to address endogeneity, does not change the estimated coefficients from Model 2.

We find that self-employed individuals are about 0.08 percentage points less likely than wage workers to be hospitalized in any given month. This is the same as the unadjusted difference in hospitalization rates of self-employed and wage workers observed in Table 1. Compared to the average monthly hospitalization rate of 0.14% among wage workers, this means that self-employed individuals are less than half as likely to be hospitalized. Overall, our findings indicate a large negative impact of self-employment on the likelihood of hospitalization that is consistent across models. Results also indicate that female, older, and native workers have higher rates of hospitalization (results available upon request).

4.3 Other results

Looking at potentially heterogeneous effects of self-employment for different subgroups, we find that the negative impact of self-employment on the likelihood of hospitalization is stronger for women than for men. There are no differences between individuals less than or 36+ years old or between nationals and foreigners (Table 3).

[11] Instrumental variables estimation using only instrument one or instrument two produces identical results.

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 (not shown).

Results from our exploration of the types of hospitalizations of self-employed and wage workers, detailed in the Online Resource 3, indicate that self-employment is associated with lower likelihood of hospitalization for any underlying health problem, as well as whether hospitalizations are urgent or planned.

Looking at the natural logarithm of hospitalization days, conditional on there being a hospitalization, we find no significant effects of self-employment. However, this analysis is limited because only observations with a hospitalization are used and many individuals have only one hospitalization over the entire period of analysis. When including the zeroes, by looking at the logarithm of hospitalization days plus one, the estimated coefficients are negative and strongly significant, indicating that self-employment reduces the length of hospitalization by almost 0.2%. However, this analysis is also limited because the choice of adding one to the number of days, in order to keep the zeroes, may influence results. In sum, 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 (results available upon request).

Table 4 presents the effect of self-employment on the likelihood of mortality in the following year. The self-employment indicator takes value one if the individual is self-employed for more than six months in the current year. Similar results are obtained when one month as a self-employed worker is enough to classify an individual as self-employed in year t or when we require individuals to be self-employed during the whole year. The models that (partly) address endogeneity provide negative coefficients for the self-employment indicator (Models 2-4). Although not statistically different from zero, the estimated coefficient from Model 2 indicates that self-employed individuals are about 0.01 percentage points less likely to die in the following year than wage workers. Compared to the average mortality rate of wage workers, this represents a lower likelihood of mortality by about one third. This analysis has limitations, as data are aggregated to a yearly frequency and mortality is such a rare and extreme outcome that there is little variation to

identify precisely an effect of self-employment. Yet, results are in line with our main findings for hospitalizations, suggesting a protective effect of self-employment when it comes to acute events such as hospital admission and death.

5 Discussion and conclusions

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 (and their typically small businesses) 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. Moreover, informing policy is increasingly pressing these days, as new forms of self-employment emerge and the small businesses that they create can have a significant impact on sustainable economic growth. The ongoing Covid19 crisis may also represent a significant push towards self-employment (and wage employment with increased job flexibility, through greater use of remote work) which may have its own additional consequences in terms of health.

Given the motivation above, we seek to provide causal evidence on the impact of self-employment on hospitalizations in this study. We take advantage of the longitudinal nature of our rich data, where we track roughly 6,500 individuals that switch between forms of employment over a period of up to 84 months. On top of that, we also employ an instrumental variable strategy to deal with any remaining endogeneity.

We find that self-employed individuals are 0.08 percentage points (or about half) less likely to be hospitalized in a given month when compared to wage workers. Qualitatively, this result is in line with most available evidence, which tends to find that self-employment is good for health. This includes Toivanen et al. [2016] and Toivanen et al. [2018], who like us look at hospitalizations and mortality. We do not seem to find evidence of endogeneity, contrary to 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 outcomes considered 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.

Hospitalization is a specific, acute outcome and not a measure of health status *per se*. The same can be said of mortality. The Job Demand-Job Control theory is closely linked to work-related stress, yet the most obvious manifestations of stress do not always lead to hospitalization or death (e.g. anxiety, depression). In this regard, we may miss important impacts of self-employment on health, which can be positive or negative. We believe more research is needed on this important topic, looking at different, complementary health outcomes. Nevertheless, as mentioned in the Introduction, stress is an important cause of many health problems, ranging from cardiovascular to respiratory, digestive, and other troubles, which frequently lead to hospitalization (or death). In our analyses of the health problems underlying hospital admissions, we find that self-employed individuals are particularly less likely than wage workers to be hospitalized for troubles of the cardiovascular, respiratory, and digestive systems. Despite the limitations of those analyses, our results do not contradict the interpretation that self-employed individuals seem to suffer from lower stress than wage workers or, in other words, that the beneficial effects of higher job control when self-employed exceed the detrimental effects of higher job demands. Our results are also consistent with the research on 'procedural utility' that finds higher levels of well being among self-employed individuals, something that may be linked with lower stress/better health.

Our results may also reflect changes in the occupations when individuals switch to/from self-employment and small businesses, which may have different exposures to occupational hazards. For instance, manufacturing workers —typically wage workers— may be more prone to injuries at work. We do find that self-employed individuals are significantly less likely than wage workers to be hospitalized for troubles of the musculoskeletal system, which include many work-related episodes. Still, we find equally large or larger differences in hospitalization rates for other types of troubles. Unfortunately, with the available data we cannot explore this issue precisely, as we do not know the industry/occupation of self-employed individuals. The potentially different effects of self-employment by industry remains a topic that deserves to be explored in future research. Toivanen et al. [2016] and Toivanen et al. [2018] already showed promising results in this regard.

We believe that the premiss that self-employed individuals may delay care in order not to

lose business is of limited concern here. Hospitalizations are generally acute, untimed events. Furthermore, non-emergency acute interventions are scheduled by the hospital and long waiting lists deter individuals from passing on a scheduled intervention they need. We find identical relative risk ratios for urgent and planned hospital admissions. Also, if self-employed individuals, having more limited access to sickness benefits, delayed appropriate care until they are seriously sick and have to be hospitalized, we would find that self-employment leads to higher rates of hospitalization, which is the opposite of what we find. As we do not know the diagnoses of all hospital admissions in the data, we cannot exclude admissions related to pregnancy and childbirth, which are unrelated to health status and capture instead fertility decisions. However, while this may partly explain the larger effect of self-employment found for women, it does not explain our findings for men, for whom we also find negative hospitalization effects.

With our approach, we were able to at least partly rule out endogeneity, thanks largely to the rich longitudinal dimension of the data we use. Further research may want to explore additional individual information to investigate potential heterogeneous effects, e.g. by industry or occupation. Further research may also want to consider the case of self-employed individuals with employees, even if this type of self-employment and their small businesses is less common among platform economy jobs.

In conclusion, this study provides evidence of a positive impact of self-employment on health and does so by focusing on an objective outcome —hospital admissions— that is not subject to recall or other biases that may affect previous studies. The positive health effect we document may be at least partly explained by greater control by the individual over different aspects of the working life associated with this form of small businesses.

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 in the platform economy (e.g. Garben 2017, European Commission 2017). This dimension is now even more significant in the context of the Covid19 crisis. This may also involve multiple policy aspects such as social security, employment law and collective bargaining. Our results indicate that, despite the existing concerns, at least as far as significant health events are concerned, there

are important social gains from more flexible work formats. Furthermore, as the platform 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 becomes more pressing.

References

B. D. H. Autor, D. Dorn, and G. H. Hanson. The China syndrome: local labor market effects of import competition in the United States. *American Economic Review*, 103(6):2121–2168, 2013.

R. A. Baron, R. J. Franklin, and K. M. Hmieleski. Why entrepreneurs often experience low, not high, levels of stress: the joint effects of selection and psychological capital. *Journal of Management*, 42(3):742–768, 2016. ISSN 15571211. doi: 10.1177/0149206313495411.

M. Benz and B. S. Frey. Being independent is a great thing: Subjective evaluations of self-employment and hierarchy. *Economica*, 75(298):362–383, 2008. ISSN 00130427. doi: 10.1111/j.1468-0335.2007.00594.x.

A. Blanch. Social support as a mediator between job control and psychological strain. *Social Science & Medicine*, 157(C):148–155, 2016. doi: 10.1016/j.socscimed.2016. URL `https://ideas.repec.org/a/eee/socmed/v157y2016icp148-155.html`.

H. Bloemen, S. Hochguertel, and J. Zweerink. Job loss, firm-level heterogeneity and mortality: Evidence from administrative data. *Journal of Health Economics*, 59:78–90, 2018. ISSN 18791646. doi: 10.1016/j.jhealeco.2018.03.005.

M. Browning and E. Heinesen. Effect of job loss due to plant closure on mortality and hospitalization. *Journal of Health Economics*, 31(4):599–616, 2012. ISSN 01676296. doi: 10.1016/j.jhealeco.2012.03.001.

A. Bujacz, C. Eib, and S. Toivanen. Not all are equal: a latent profile analysis of well-being among the self-employed. *Journal of Happiness Studies*, 2019. ISSN 1573-7780. doi: 10.1007/s10902-019-00147-1. URL `https://doi.org/10.1007/s10902-019-00147-1`.

A. Case and C. Paxson. Causes and consequences of early-life health. *Demography*, 47(S):S65–S85, 2010. ISSN 1533-7790. doi: 10.1353/dem.2010.0007.

A. Case, A. Fertig, and C. Paxson. The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24(2):365–389, 2005. ISSN 01676296. doi: 10.1016/j.jhealeco.2004.09.008.

C. de Chaisemartin. Tolerating defiance? Local average treatment effects without monotonicity. *Quantitative Economics*, 8:367–396, 2017. doi: 10.3982/QE601.

European Commission. Proposal for a directive of the European Parliament and of the Council on transparent and predictable working conditions in the European Union, 2017.

Eurostat. Employment by sex, age, and professional status, 2017.

S. Garben. *Protecting workers in the online platform economy: an overview of regulatory and policy developments in the EU*. European Agency for Safety and Health at Work, Luxemburg, 2017. doi: 10.2802/918187.

M. M. Gielnik, H. Zacher, and M. Frese. Focus on opportunities as a mediator of the relationship between business owners' age and venture growth. *Journal of Business Venturing*, 27(1):127–142, 2012. ISSN 08839026. doi: 10.1016/j.jbusvent.2010.05.002.

D. F. Halpern. How time-flexible work policies can reduce stress, improve health, and save money. *Stress and Health*, 21(3):157–168, 2005. ISSN 15323005. doi: 10.1002/smi.1049.

J. Hessels, C. A. Rietveld, and P. van der Zwan. Self-employment and work-related stress: the mediating role of job control and job demand. *Journal of Business Venturing*, 32:178–196, 2017. ISSN 0883-9026. doi: 10.1016/j.jbusvent.2016.10.007.

K. Holmgren, S. Dahlin-Ivanoff, C. Björkelund, and G. Hensing. 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, 2009. ISSN 14712458. doi: 10.1186/1471-2458-9-73.

ILO. *Non-standard employment around the world: Understanding challenges, shaping prospects*. International Labour Office, Geneva, 2016. ISBN 9789221303855.

M. Ingre. *P-hacking in academic research. A critical review of the job strain model and of the association between night work and breast cancer in women.* PhD thesis, Stockholm University, 2017.

R. A. Karasek. Job demands, job decision latitude, and mental strain: implications for job redesign. *Administrative Science Quarterly*, 24(2):285–308, 1979.

R. A. Karasek and T. Theorell. *Healthy work: stress, productivity, and the reconstruction of working life.* Basic Books, New York, 1990.

E. A. Mayer. The neurobiology of stress and gastrointestinal disease. *Gut*, 47(6):861–869, 2000. ISSN 00175749. doi: 10.1136/gut.47.6.861.

A. Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2): 307–342, 2001.

M. Nikolova. Self-employment can be good for your health. *GLO Discussion Paper*, (226), 2018.

F. Noseleit. Female self-employment and children. *Small Business Economics*, 43:549–569, 2014. doi: 10.1007/s11187-014-9570-8.

A. Pesole, M. Urzí Brancati, E. Fernández-Macías, F. Biagi, and I. González Vázquez. *Platform workers in Europe. Evidence from the COLLEEM Survey.* Luxembourg: Publications Office of the European Union, 2018, 2018. ISBN 9789279879968. doi: https://doi.org/10.2760/742789.

D. J. Prottas and C. A. Thompson. Stress, satisfaction, and the work-family interface: a comparison of self-employed business owners, independents, and organizational employees. *Journal of Occupational Health Psychology*, 11(4):366–378, 2006.

C. A. Rietveld, H. Kippersluis, and A. R. Thurik. Self-employment and health: barriers or benefits? *Health Economics*, 24:1302–1313, 2015. doi: 10.1002/hec.

S. Schneck. Why the self-employed are happier: Evidence from 25 European countries. *Journal of Business Research*, 67(6):1043–1048, 2014. ISSN 01482963. doi: 10.1016/j.jbusres.2013.06.008.

U. Stephan and U. Roesler. Health of entrepreneurs versus employees in a national representative sample. *Journal of Occupational and Organizational Psychology*, 83:717–738, 2010. doi: 10.1348/096317909X472067.

A. Steptoe and M. Kivimäki. Stress and cardiovascular disease. *Nature Reviews Cardiology*, 9: 360–370, 2012.

T. Theorell and R. A. Karasek. Current issues relating to psychosocial job strain and cardiovascular disease research. *Journal of Occupational Health Psychology*, 1:9–26, 1996.

S. Toivanen, R. H. Griep, C. Mellner, S. Vinberg, and S. Eloranta. 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, 2016. doi: 10.1136/oemed-2015-103472.

S. Toivanen, R. H. Griep, C. Mellner, M. Nordenmark, S. Vinberg, and S. Eloranta. Hospitalization due to stroke and myocardial infarction in self-employed individuals and small business owners compared with paid employees in Sweden – a 5-year study. *Small Business Economics*, 2018.

O. Torrès and R. Thurik. Small business owners and health. *Small Business Economics*, 53(2): 311–321, 2019.

S. Torres and P. Raposo. O trabalho por conta própria em Portugal. *Instituto Nacional de Estatística*, 2018.

J. Yoon and S. L. Bernell. The effect of self-employment on health, access to care, and health behavior. *Health*, 05(12):2116–2127, 2013. ISSN 1949-4998. doi: 10.4236/health.2013.512289.

J. M. Zissimopoulos and L. A. Karoly. Transitions to self-employment at older ages: the role of wealth, health, health insurance and other factors. *Labour Economics*, 14(2):269–295, 2007. ISSN 09275371. doi: 10.1016/j.labeco.2005.08.002.

Tables

Table 1: Descriptive statistics by type of employment in t-1

	Self-employed (t-1)	Wage workers (t-1)	Total
Panel A: Sample of switchers			
Any hospitalization in month t	0.06%	0.14%	0.11%
Number of days of hospitalization[b]	12.86	11.05	11.45
	(8.39)	(7.26)	(7.55)
	[1;37]	[1; 36]	[1;37]
Female	49.78%	51.90%	51.09%
Age	37.46	35.63	36.33
	(9.71)	(9.20)	(9.44)
	[18; 65]	[18; 65]	[18; 65]
Foreign	12.75%	14.32%	13.71%
Instrument one	4.89%	4.57%	4.69%
	2.20	2.30	2.27
	[0; 17.74]	[0; 17.74]	[0; 17.74]
Instrument two	4.31%	3.90%	4.06%
	2.15	1.85	1.98
	[0.75; 17.34]	[0.75; 16.76]	[0.75; 17.34]
Individuals (n)	6,210	6,341	6,517
Observations (N)	119,769	193,029	312,798
	38.29%	61.71%	100.00%
Panel B: Full sample			
Any hospitalization in month t	0.10%	0.18%	0.18%
Number of days of hospitalization[a]	12.63	10.99	11.03
	(9.36)	(7.35)	(7.41)
	[1;90]	[1; 90]	[1;90]
Female	48.72%	53.00%	52.82%
Age	42.82	37.12	37.36
	(11.16)	(10.44)	(10.54)
	[18; 65]	[18; 65]	[18; 65]
Foreign	9.86%	11.94%	11.85%
Instrument one	5.29%	4.28%	4.32%
	2.18	2.17	2.18
	[0; 17.74]	[0; 17.74]	[0; 17.74]
Instrument two	5.53%	4.22%	4.27%
	2.97	2.29	2.34
	[0.75; 17.34]	[0.75; 17.34]	[0.75; 17.34]
Individuals (n)	10,014	125,207	129,142
Observations (N)	282,072	6,365,225	6,647,297
	4.24%	95.76%	100.00%

Continuous variables: standard deviations in parentheses and minimum and maximum values in brackets.
[a]Total sample size: n=10,001; N=11,811. [b]Total sample size: n=236; N=263.

Table 2: Effect of self-employment on the likelihood of hospitalization

	Model 1	Model 2	Model 3		Model 4	
			1st stage	2nd stage	1st stage	2nd stage
Panel A: Sample of switchers (n=6,517; N=312,798)						
Self-employed (t-1)	-0.07680***	-0.07820***	–	-0.08020	–	-0.09992
	(0.01192)	(0.01432)		(0.09879)		(0.12034)
Instrument one	–	–	0.05390***	–	0.06972***	–
			(0.01263)		(0.01239)	
Instrument two	–	–	0.04637***	–	0.04500**	–
			(0.01474)		(0.01965)	
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE[a]	Yes	–	Yes	Yes	–	–
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	–	Yes	–	–	Yes	Yes
Instrument strength	–	–	F=21.25***		F=22.07***	
Instrument validity	–	–	χ^2=0.273		χ^2=0.759	
Endogeneity	–	–	χ^2=0.005		χ^2=0.003	
Panel B: Full sample (n=129,142; N=6,647,297)						
Self-employed (t-1)	-0.09999***	-0.07936***	–	-2.71471**	–	-2.29952**
	(0.00684)	(0.01427)		(1.18807)		(0.91874)
Instrument one	–	–	0.00226	–	0.00335***	–
			(0.00185)		(0.00060)	
Instrument two	–	–	0.00631**	–	0.00187**	–
			(0.00304)		(0.00087)	
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE[a]	Yes	–	Yes	Yes	–	–
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	–	Yes	–	–	Yes	Yes
Instrument strength	–	–	F=4.12**		F=20.07***	
Instrument validity	–	–	χ^2=0.110		χ^2=1.047	
Endogeneity	–	–	χ^2=4.840**		χ^2=2.848*	

[a]In models with individual fixed effects, the district fixed effects drop due to collinearity.
Standard errors in parentheses, robust to heteroscedasticity and to clustering at the individual level in Models 1 and 2 and at the district level in Models 3 and 4. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
The coefficient of the self-employment indicator was multiplied by 100 to facilitate reading.

Table 3: Heterogeneous effects of self-employment on the likelihood of hospitalization (Model 2)

	By gender	By age	By nationality
Panel A: Sample of switchers (n=6,517; N=312,798)			
Self-employed (t-1)	-0.04287**	-0.08595***	-0.07780***
	(0.01895)	(0.01659)	(0.01554)
Self-employed and female (t-1)	-0.06993**	–	–
	(0.02874)		
Self-employed and 36+ years old (t-1)	–	0.01703	–
		(0.02857)	
Self-employed and foreign (t-1)	–	–	-0.00293
			(0.03937)
Demographic controls	Yes	Yes	Yes
District FE[a]	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Panel B: Full sample (n=129,142; N=6,647,297)			
Self-employed (t-1)	-0.04431**	-0.08677***	-0.07928***
	(0.01866)	(0.01560)	(0.01550)
Self-employed and female (t-1)	-0.06940**	–	–
	(0.02848)		
Self-employed and 36+ years old (t-1)	–	0.01619	–
		(0.02223)	
Self-employed and foreign (t-1)	–	–	-0.00063
			(0.03925)
Demographic controls	Yes	Yes	Yes
District FE[a]	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes

[a]In models with individual fixed effects, the district fixed effects drop due to collinearity.
Standard errors in parentheses, robust to heteroscedasticity and to clustering at the individual level.
*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
The coefficient of the self-employment indicator was multiplied by 100 to facilitate reading.

Table 4: Effect of self-employment on the likelihood of mortality in year t+1

	Model 1	Model 2	Model 3		Model 4	
			1st stage	2nd stage	1st stage	2nd stage
Panel A: Sample of switchers (n=6,219; N=23,702)						
Self-employed (t)	0.02619	-0.01647		-0.16210		-0.71752*
	(0.03548)	(0.05481)		(0.22167)		(0.40849)
Instrument one			0.05845***		0.08344***	
			(0.01275)		(0.01314)	
Instrument two			0.05151***		0.06371**	
			(0.01706)		(0.02738)	
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE[a]	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	–	Yes	–	–	Yes	Yes
Instrument strength	–	–	F=24.78***		F=29.50***	
Instrument validity	–	–	χ^2=1.120		χ^2=1.186	
Endogeneity	–	–	χ^2=2.642		χ^2=2.227	
Panel B: Full sample (n=117,403; N=503,083)						
Self-employed (t)	-0.02415	-0.00844		-6.79791		-8.69471**
	(0.02473)	(0.05445)		(6.91666)		(3.76139)
Instrument one			0.00023		0.00416***	
			(0.00316)		(0.00032)	
Instrument two			0.00611		0.00257**	
			(0.00357)		(0.00119)	
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE[a]	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	–	Yes	–	–	Yes	Yes
Instrument strength	–	–	F=1.56		F=105.57***	
Instrument validity	–	–	χ^2=0.089		χ^2=4.895**	
Endogeneity	–	–	χ^2=2.470		χ^2=0.622	

[a]In models with individual fixed effects, the district fixed effects drop due to collinearity.
Standard errors in parentheses, robust to heteroscedasticity and to clustering at the individual level in Models 1 and 2 and at the district level in Models 3 and 4. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.
The coefficient of the self-employment indicator was multiplied by 100 to facilitate reading.

Effects of self-employment on hospitalizations: Instrumental variables analysis of social security data

Online Resource 1

Figure 1: Division of the Portuguese territory into districts and municipalities

There are 18 districts in Continental Portugal. Each island is considered as a separate district.

Effects of self-employment on hospitalizations: Instrumental variables analysis of social security data

Online Resource 2

Figure 1: General income index by district

Source: Marktest. Not available for the islands.

Figure 2: General health index by district

Source: Marktest. Not available for the islands.

Figure 3: Firm dimension index

Source: Marktest. Not available for the islands.

Figure 4: Proportion of self-employed workers by district

Source: own calculations.

Effects of self-employment on hospitalizations: Instrumental variables analysis of social security data

Online Resource 3

Complementary analyses: types of hospitalizations of self-employed and wage workers

To explore the types of hospitalizations of self-employed and wage workers, we rely on 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.

There is not a personal identifier that would allow us to directly match individuals in the social security and DRG datasets. We matched hospitalizations in the two registers based on the individuals' gender, year of birth, municipality of residence, and day of admission. This resulted in more than 6,000 of 12,800+ hospital admissions in the social security database exactly matched between the two datasets (i.e. about half). The other half correspond to multiple observations with the same values of the matching variables and hospitalizations in private hospitals, which are not recorded in the DRG dataset.

Using the DRG classification of hospital admissions, we assigned each matched hospital admission to one of ten groups. The description of these groups is shown in Table 1, alongside the description of the top three most common DRG codes in our data.

We estimate a multinomial logit model where the dependent variable has 11 categories described in Table 1 plus a reference category of no hospitalization. The explanatory variable of interest is self-employment in $t-1$. The model controls for individuals' gender, age, nationality, as well as year and month fixed-effects. We use the whole sample.

This analysis is descriptive. Data limitations, such as the numbers of observations in each type of hospitalization-type of work cell being sometimes low, prevent us from fully addressing endogeneity concerns. Even though in our main analyses we did not find evidence of endogeneity,

Table 1: Classification of the hospital admissions and most common DRG codes

Number (%)	**Hospital admissions related to:**	**Most common DRG codes (description)**
308 (2%)	Diseases and troubles of the nervous system	Carpal tunnel decompression, Stroke with infarction, Seizures and/or headache
751 (6%)	Diseases and troubles of the ear, nose, mouth, and throat / Diseases and troubles of the respiratory system	Various procedures in the ear, nose, mouth and/or throat, Procedures in the facial and/or mastoid sinuses, Tonsillectomy and/or adenoidectomy
509 (4%)	Diseases and troubles of the circulatory system	Venous lacheation and phlebotomy, Circulatory disorders other than acute myocardial infarction, with cardiac catheterization, without complex diagnosis, Circulatory disorders with acute myocardial infarction, no major complications, discharged alive, Arrhythmia and/or cardiac conduction disturbances
1,332 (10%)	Diseases and troubles of the digestive system / Diseases and troubles of the hepatobiliary system and pancreas	Procedures for inguinal and/or femoral hernia, Laparoscopic cholecystectomy without choledochal surgery, Appendectomy without complicated main diagnosis
466 (4%)	Diseases and troubles of the musculoskeletal system and connective tissues	Knee procedures, Soft tissue procedures, Local excision and/or removal of an internal fixation device other than the hip and femur
398 (3%)	Diseases and troubles of the skin, subcutaneous cell tissue, and breast	Other procedures on the skin, subcutaneous tissue and/or breast, Perianal and/or pilonidal procedures, Biopsy and/or local excision of the breast by non-malignant disease
237 (2%)	Diseases and troubles of the metabolism, endocrine and nutritional diseases and troubles	Obesity Procedures, Thyroid procedures, Endocrine disorders
680 (5%)	Diseases and troubles of the female genital tract	Procedures in the uterus and/or its attachments for carcinoma in situ and/or non-malignant disease, Tubal, laparoscopic and/or incisional tubal ligation, Endoscopic tubal ligation
366 (3%)	Pregnancy, childbirth and puerperium	Vaginal delivery without complication diagnoses, Other pre-delivery diagnoses, with medical complications, C-section
994 (8%)	Others	Chemotherapy, Radiotherapy, Psychoses
6,855 (53%)	Unknown	–
12,896	–	–

we refrain from any causal interpretations here.

Results are reported in Table 2. All relative risk ratios are below one, indicating that self-employed individuals are less likely to be hospitalized than wage workers, irrespective of the underlying health problem. Four types of hospitalizations are particularly less likely among self-employed than among wage workers, with risk ratios significantly different from one ($p < 0.05$): admissions for (1) Diseases and troubles of the ear, nose, mouth, and throat / Diseases and troubles of the respiratory system, (2) Diseases and troubles of the circulatory system, (3) Diseases and troubles of the digestive system / Diseases and troubles of the hepatobiliary system and pancreas, and (4) Diseases and troubles of the musculoskeletal system and connective tissues.

We also have information on the main diagnosis underlying each matched hospitalization. Diagnosis codes starting with "8" are those related with fractures, dislocations, sprains and

Table 2: Self-employment and the likelihood of different types of hospitalization (n=129,901; N=6,680,530)

Type of hospitalization	Relative risk ratio
No hospitalization	Base outcome
Diseases and troubles of the nervous system	0.656
Diseases and troubles of the ear, nose, mouth, and throat / Diseases and troubles of the respiratory system	0.388***
Diseases and troubles of the circulatory system	0.486**
Diseases and troubles of the digestive system / Diseases and troubles of the hepatobiliary system and pancreas	0.585***
Diseases and troubles of the musculoskeletal system and connective tissues	0.556**
Diseases and troubles of the skin, subcutaneous cell tissue, and breast	0.548*
Diseases and troubles of the metabolism, endocrine and nutritional diseases and troubles	0.389*
Diseases and troubles of the female genital tract	0.901
Pregnancy, childbirth and puerperium	0.397*
Others	0.206***
Unknown	0.483***

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

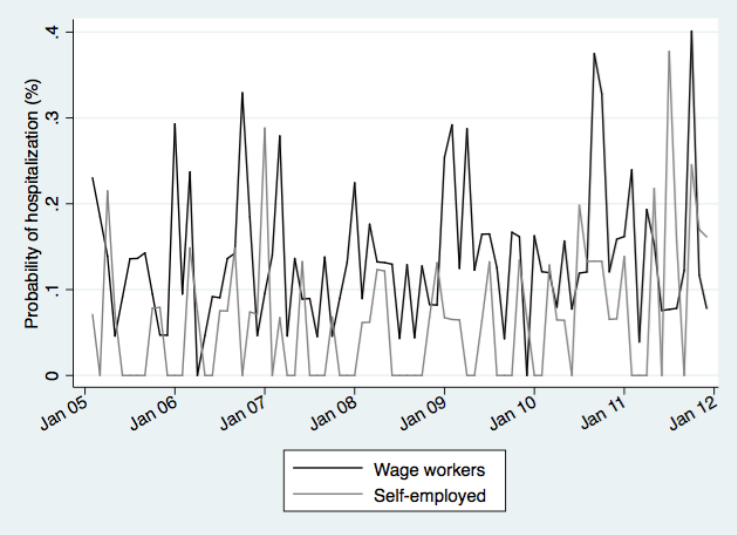
distentions, concussions, lacerations, hemorrhages, head traumas, injuries, lesions, and similar acute events that may be caused by work-related accidents. We find zero admissions with main diagnosis code starting with the number 8 among self-employed individuals.

Lastly, the data allow us to distinguish between urgent and planned hospitalizations. Again, we estimate a multinomial logit model, where the dependent variable has three categories for the type of admission (urgent, planned, unknown), in addition to 'no admission' as the baseline outcome. The estimated relative risk ratios associated with the self-employment indicator are almost identical across outcomes: 0.45 for urgent, 0.52 for planned, and 0.48 for unknown type of admission ($p \approx 0$).

Effects of self-employment on hospitalizations: Instrumental variables analysis of social security data

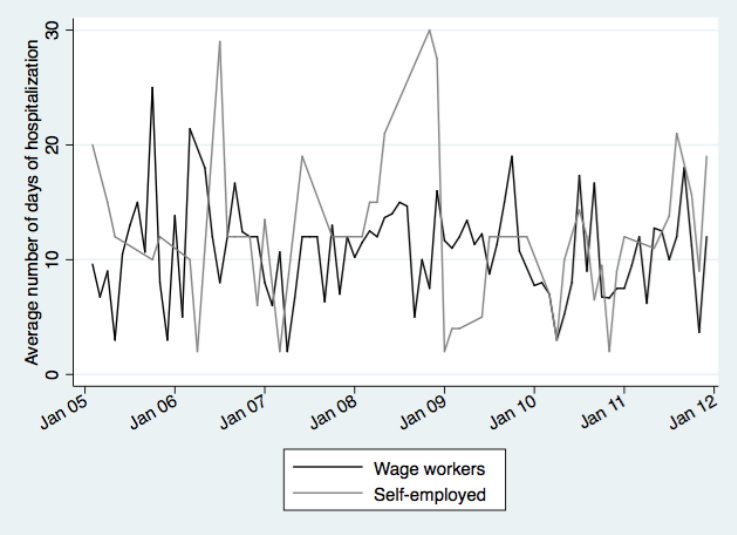
Online Resource 4

Figure 1: Probability of hospitalization in the following month by type of employment, Jan 2005-Dec 2011 (Sample of switchers)



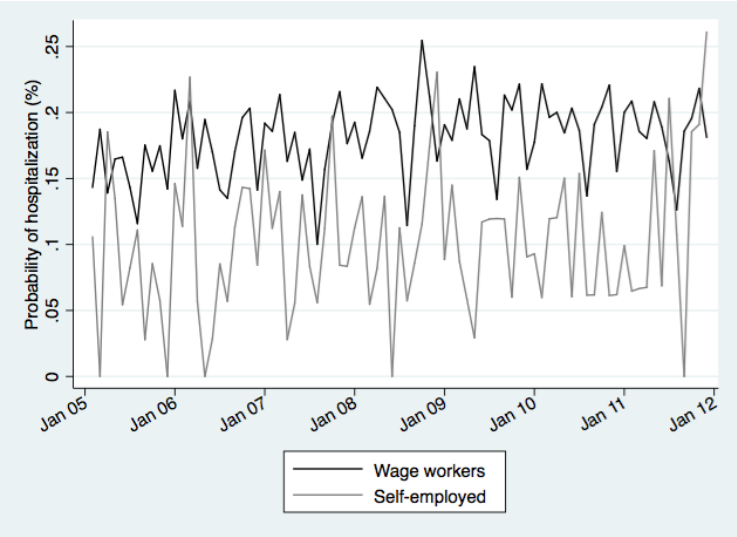
Source: own calculations.

Figure 2: Duration of hospitalization by type of employment, Jan 2005-Dec 2011 (Sample of switchers)



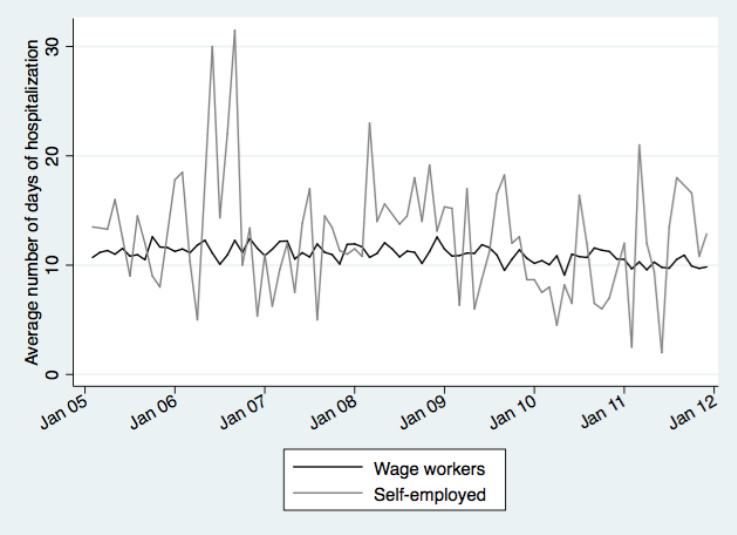
Source: own calculations.

Figure 3: Probability of hospitalization in the following month by type of employment, Jan 2005-Dec 2011 (Full sample)



Source: own calculations.

Figure 4: Duration of hospitalization by type of employment, Jan 2005-Dec 2011 (Full sample)



Source: own calculations.

Non-traditional abstract

Is self-employment generally good for one's health? This is an important question as a large number of people around the world and also in Europe are self-employed, leading on their own their typically small businesses.

Our research question also matters in the context of the Covid19 crisis. This pandemic may contribute to the growth of both self-employment (as the labour market contracts) and of wage employment conducted under remote work formats —which are typically more common amongst the self-employed.

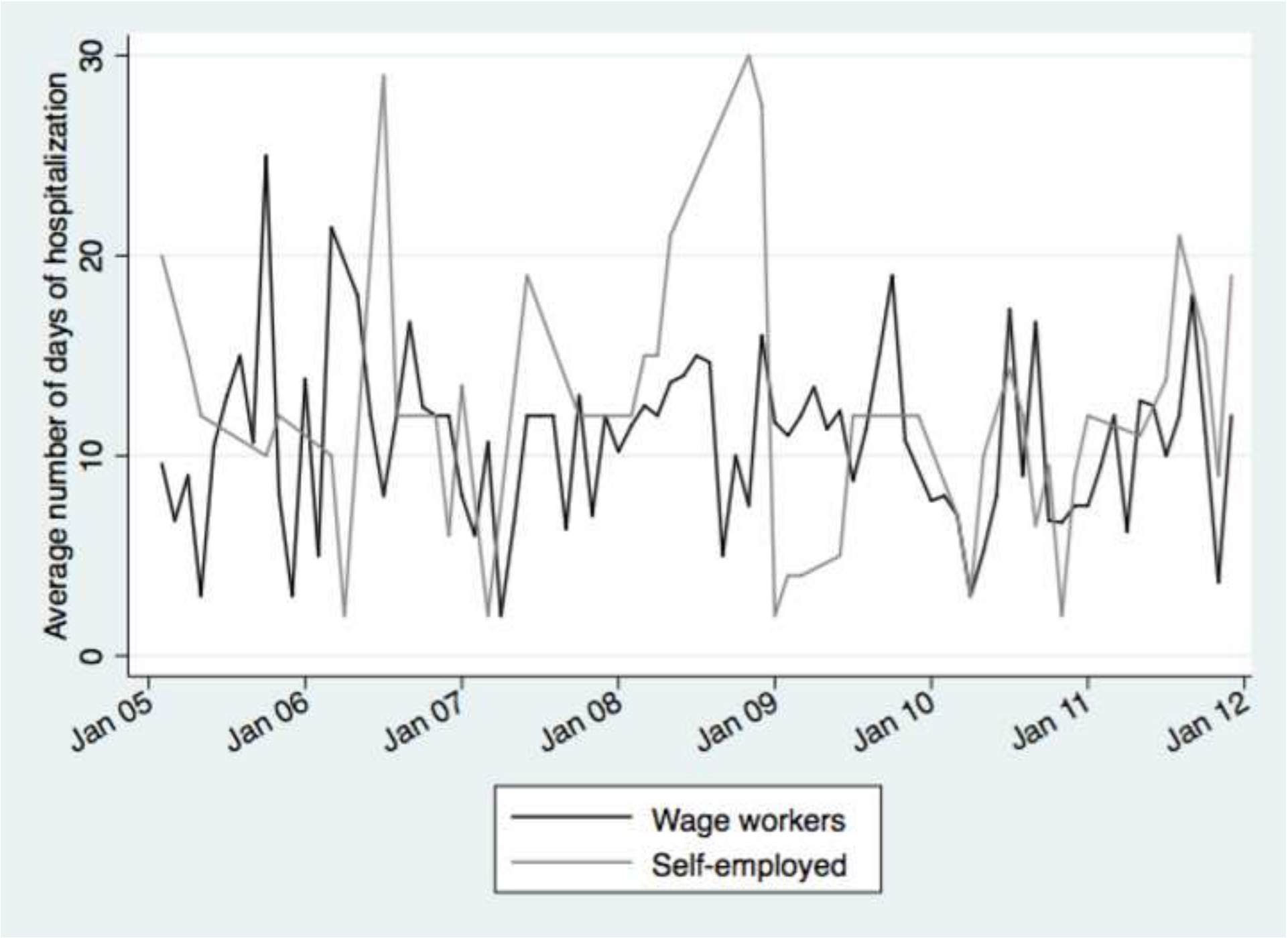
The health outcomes in this study are measured by hospital admissions, which may be regarded as a more rigorous proxy of health than self-reported variables.

The analysis is conducted comparing the same individuals between 2005 and 2011and focusing on those that change between self-employment and wage work. We draw on a sample of individuals in Portugal —a country with similar social protection for employees and the self-employed.

We find that self-employment has a positive effect on health as it reduces the likelihood of hospital admission by at least half. In other words, the benefits of self-employment in terms of greater flexibility appear to trump the costs related to greater uncertainty, as far as health is concerned.

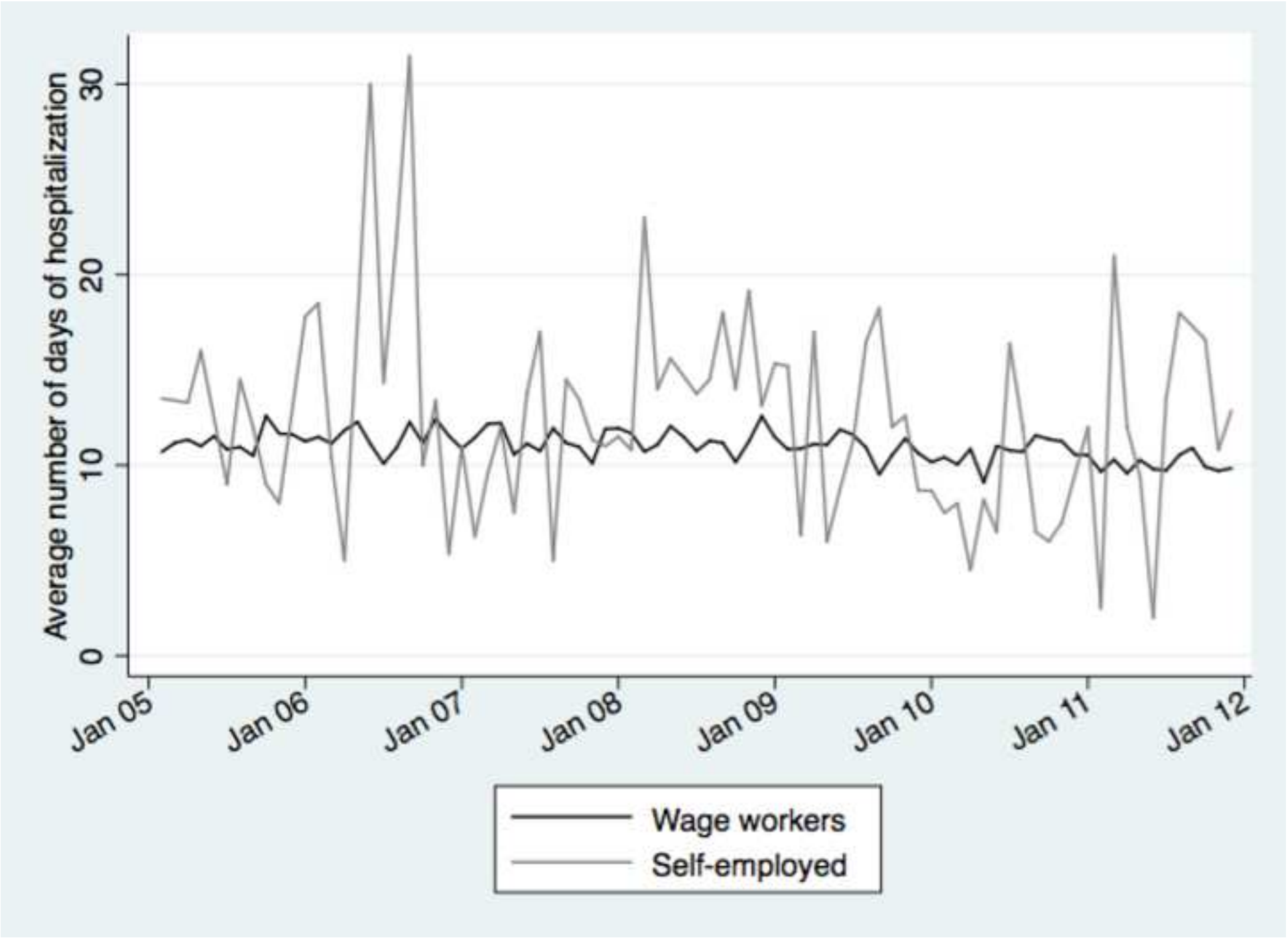
line figure

Click here to access/download;colour figure;duration_hosp_switchers.png



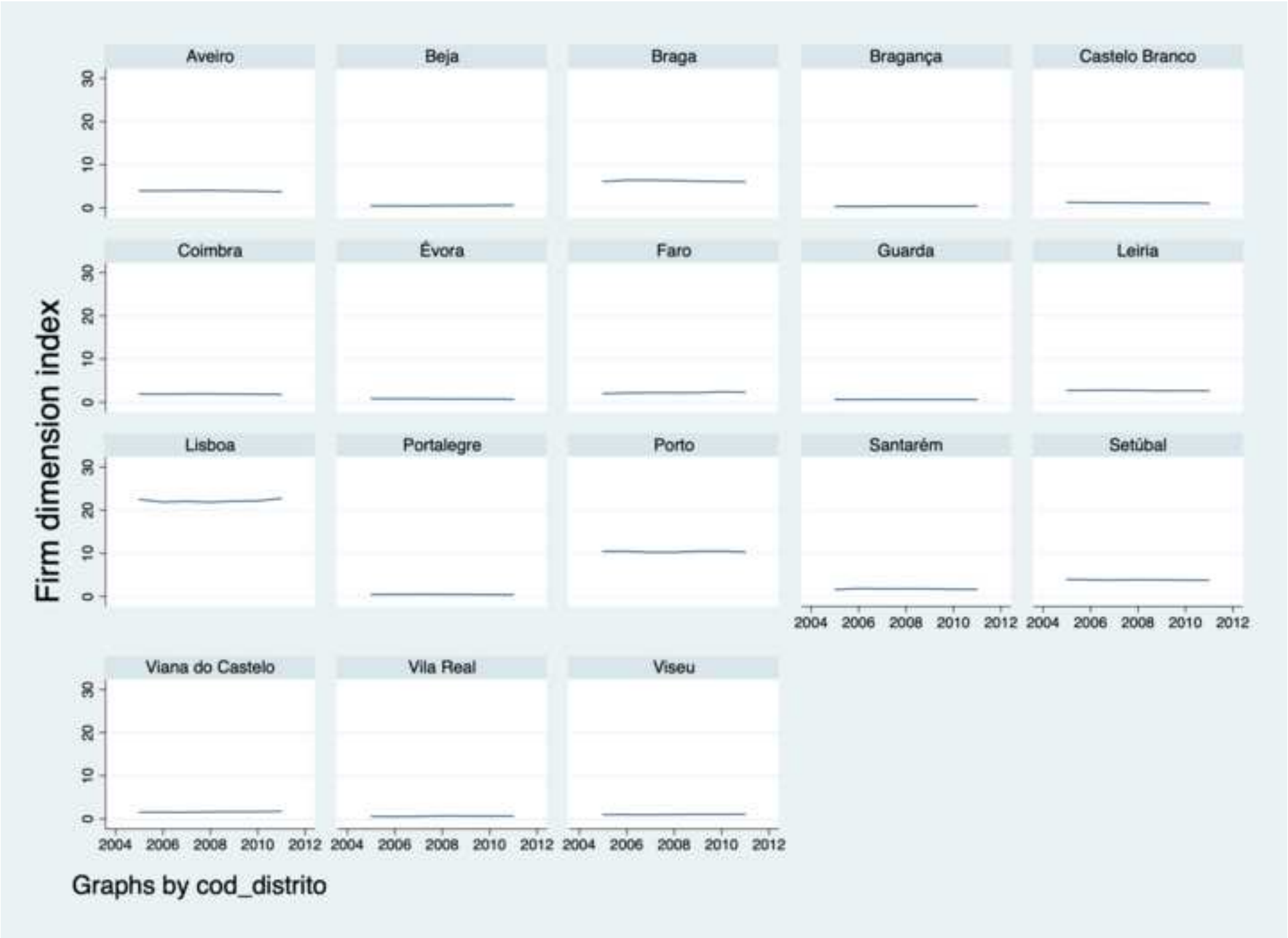
line figure

Click here to access/download;colour figure;duration_hosp.png



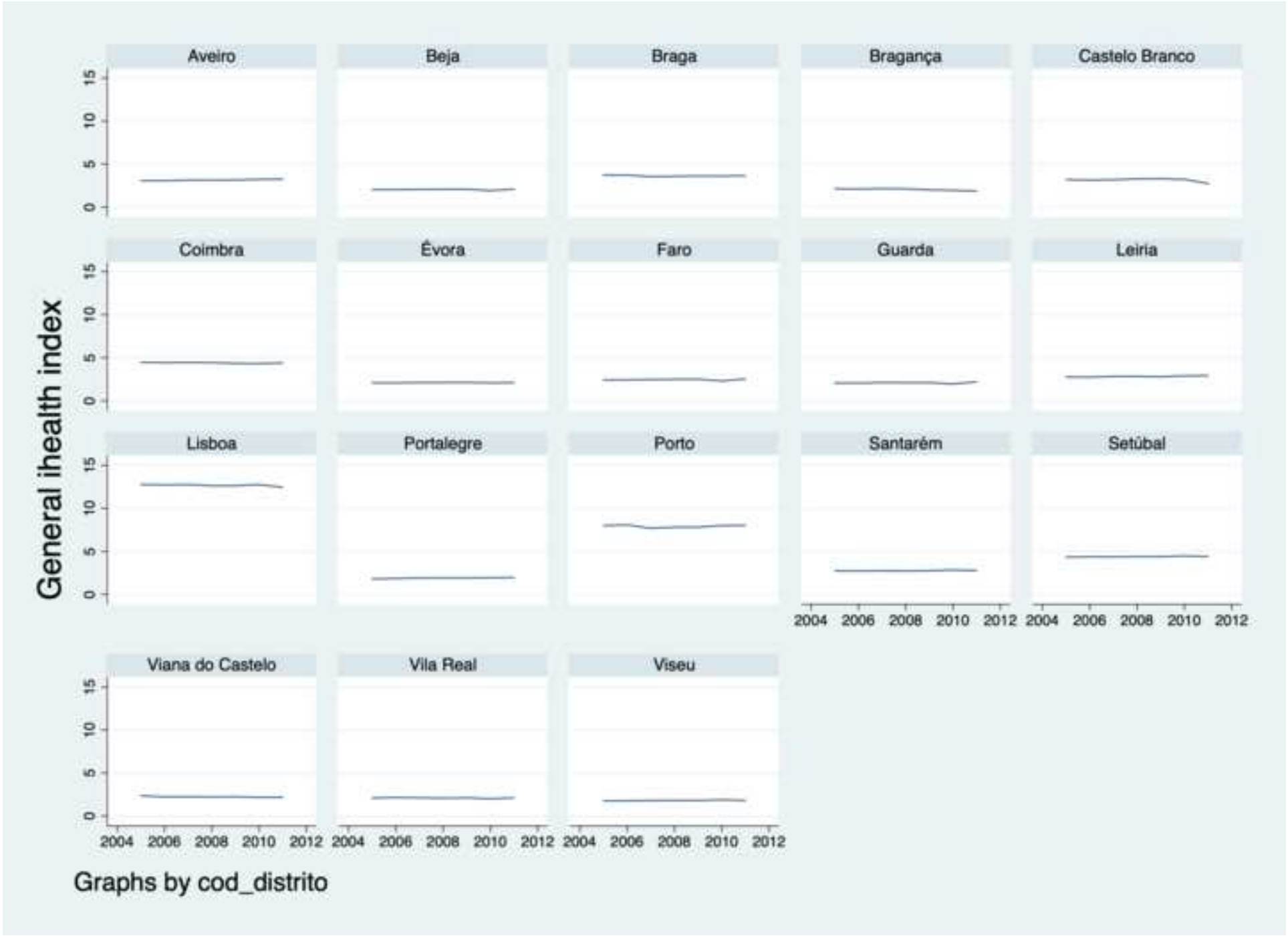
line figure

Click here to access/download;colour figure;firm_dimension_index.png



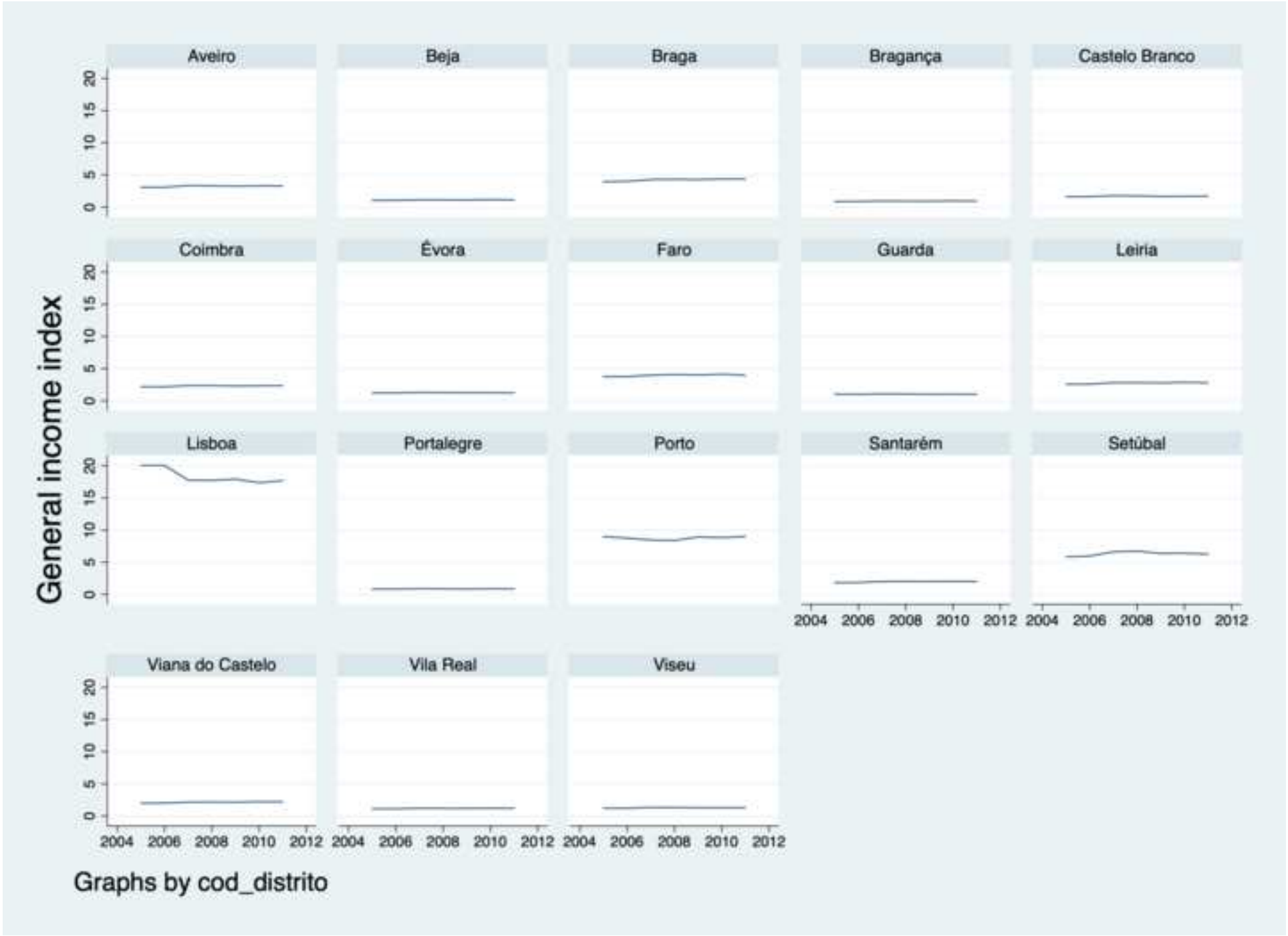
colour figure

Click here to access/download;colour figure;health_index.png



colour figure

Click here to access/download;colour figure;income_index.png



colour figure

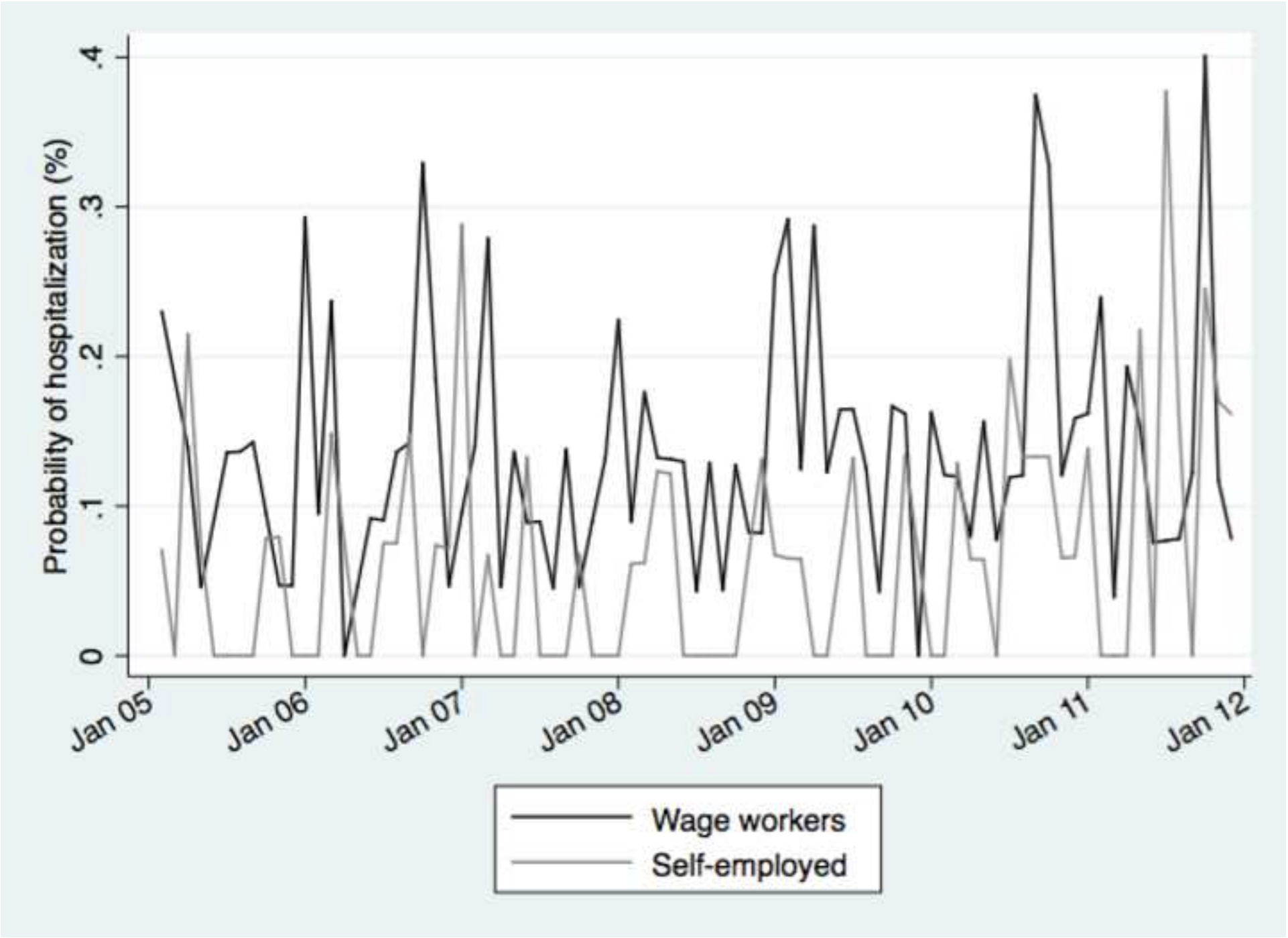
Click here to access/download;colour figure;MapOfPortugal1.png

colour figure

Click here to access/download;colour figure;MapOfPortugal2.png

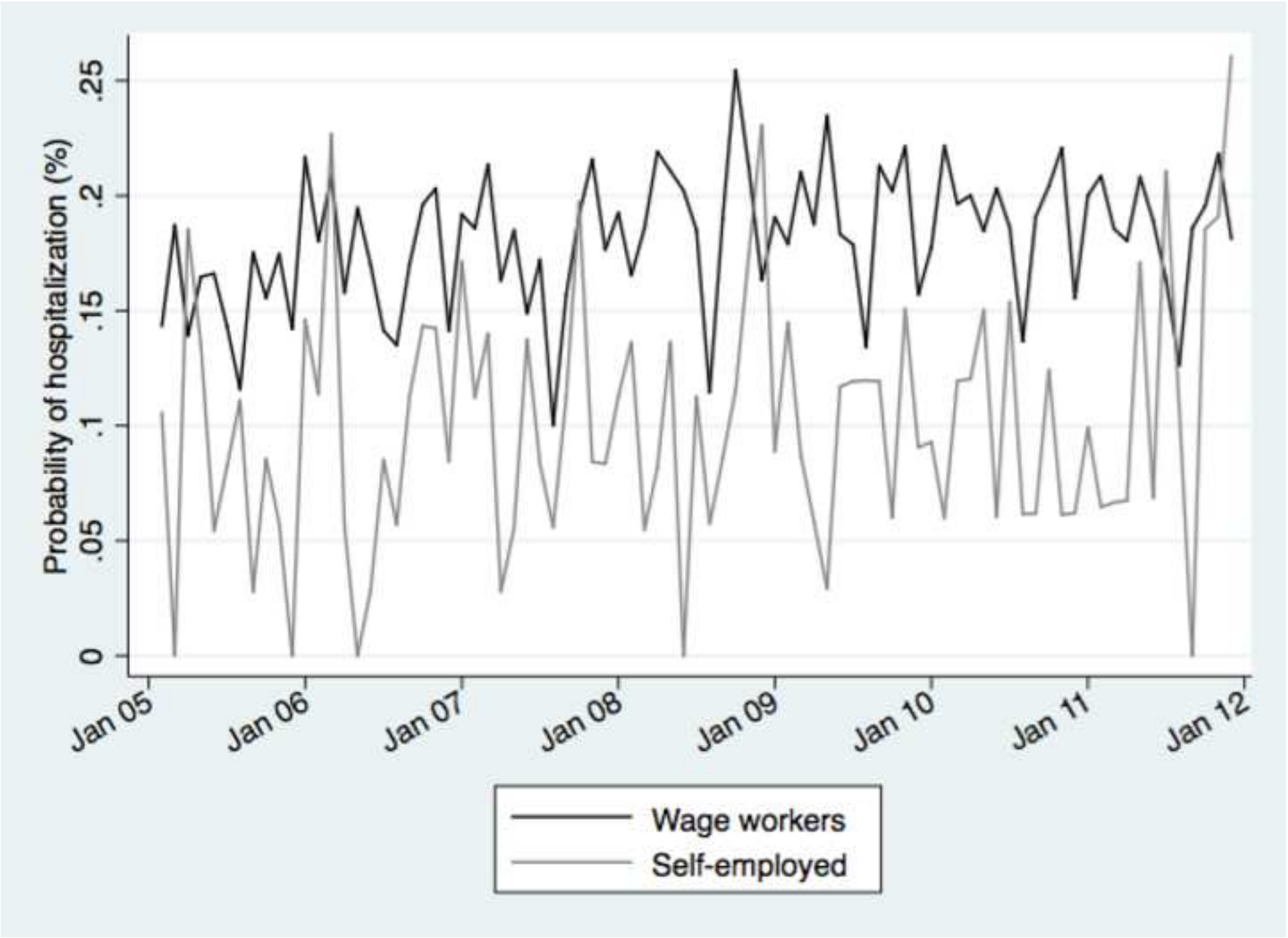
colour figure

Click here to access/download;colour figure;prob_hosp_switchers.png



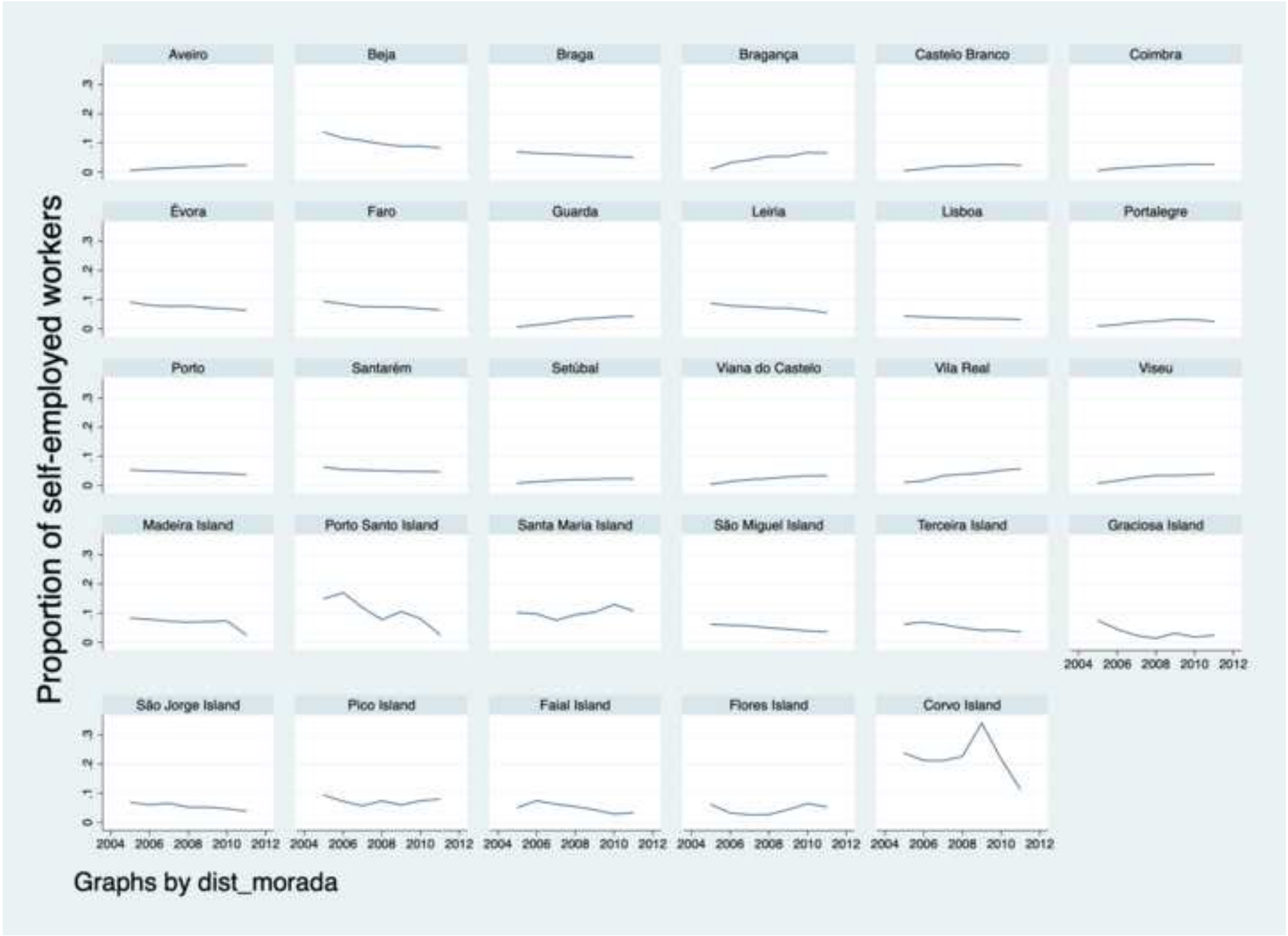
colour figure

Click here to access/download;colour figure;prob_hosp.png



colour figure

Click here to access/download;colour figure;prop_self_employed.png



We would like to thank the Editor and both reviewers for their careful reading of the paper and the valuable comments offered.

We believe that we addressed all comments in a satisfactory manner and that consequently the new version of the paper is a significant improvement on the previous one. We answer each comment (in *italics*) individually below, in blue.

Reviewer #1:

2 Major comments

2.1 Positioning of the paper

There seems to be some inconsistency in what the main analysis is for the findings of this paper (sometimes the switchers, sometimes the full sample). This is quite confusing to me as a reader. Please make this clear throughout the framing of the manuscript.

We tried to make the focus on switchers clearer from the beginning and throughout the whole document:

- Abstract: "Our main findings, based on a sample of about 6,500 individuals followed monthly from 2005 to 2011 and who switch between self-employment and wage work along that period…"
- Introduction (p. 6): "We focus on a subsample of about 6,500 individuals who switch between self-employment and wage work along that 84-month period."
- Data (p. 10): "In our main analyses, we focus on more than 6,500 individuals who switch at least once between self-employment and wage work over that period (which we refer to as switchers)."
- Empirical strategy (p. 13): "As mentioned previously, our main analyses focus on the subsample that includes only individuals who switch between wage work and self-employment at least once over the sample period (`switchers')."
- Results (p. 15): "We discuss results for the subsample of individuals who switch at least once between self-employment and wage work over time (`switchers'). Results for the full sample are also presented for comparison, in the bottom half of the tables (Panel B)."
- We also inverted Panels A and B in all tables, so that now results for switchers appear on the top.

2.2 Selection of results

The authors claim that they keep their main focus on the sample of switchers, yet the results for this sample of the IV-analysis are not significant (even though the effect size estimate seems in line with their previous results). A more elaborate discussion on this is necessary.

We added a potential justification of this finding on p. 17: “In light of this result, unobserved individual characteristics, in particular those that vary over time (e.g. health status), and reversed causality don't seem to pose an issue in our analyses. This is possibly because hospitalization is a fairly objective and a rare/extreme outcome, which doesn't capture health in general but serious (unexpected) manifestations of illness.”

2.3 Strength of the instruments
There reported F-statistic in model 3 Panel A is low (< 10). This is below the rule of thumb of 10, yet the authors do not discuss this.

We do not discuss this in the main text, as we focused the entire results section on the subsample of switchers, to avoid confusing the reader. However, we mention this point in footnote #10. As we discuss on page 16, the instruments may not be strong enough when including non-switchers in the sample, even when they appear strong (large t- and F-statistics in the model with individual fixed effects; Panel B, Model 4). This is why we focus on the switchers.

3 Minor comments

3.1 References

There are some inconsistencies in the references (I only saw this one quickly, but please make sure they are correct): Hessels, J., Rietveld, C. A., and 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. should be Hessels, J., Rietveld, C. A., and van der Zwan, P. (2017). Self-employment and work-related stress: the mediating role of job control and job demand. Journal of Business Venturing, 32:178 196. Please check all references.

This reference has been corrected and all references have been checked.

Reviewer #2:

The reviewer's comment was to drop some of the figures in the paper in order to keep the manuscript shorter. We consequently moved the appendix of the previous version to an online appendix also following the suggestion of the editor (see below).

Editor:

Also I want you to turn your appendix into a separate document which should be well defined as supplementary online material since it will no longer serve as an appendix. This document consists of the current appendix and could also contain some material of the main text which you decide to move out. Refer to it in this way in the main text and not appendix anymore. Of course, the suggestion of referee 2 to move things out of the appendix is not valid any more since it becomes an online supplementary document anyhow. You now have 28 pages of main text. Try to keep it that way, or shorter if can be.

The appendix was converted into supplementary online material. References to the Appendix in the text have been replace with references to "Online Resource X", according to the instructions for authors.

Focusing the results section on the subsample of switchers, following the comment of Reviewer #1, also allowed us to slightly shorten the manuscript.

Personally I do not like your mention of 'novel work formats' in the abstract since nobody know at this point what they refer to. Try other justifications for the entrepreneurship health nexus. The literature is full of them.

We removed the reference to 'novel work formats' in the abstract and used instead 'self-employment'

At SBEJ we experiment with two abstracts. One traditional one and one for non-scientific publicity reasons with a twitterable and headlinable approach. Please, create a separate document for this second abstract which can be longer than our orthodox one. But keep it very short though.

We are happy to participate in this and have written an abstract for non-scientific publicity.

We have also added footnote #1 to the first paragraph of the introduction and a sentence to the last paragraph of the conclusion in mention to the Covid19 crisis and the greater relevance that it brings to the research question of our study.