

Chapter 6

Entrepreneurship Analysis from a Human Population Surveys' Perspective

José María Millán, Concepción Román and Emilio Congregado

Abstract This paper tries to collect, describe and evaluate all the potential statistical sources—each pursuing different goals—in order to study self-employment in Spain. The improvement of traditional databases together with the recent incorporation of new statistical sources is bolstering the knowledge of today's labour market, self-employment included. Although the available information might be considered accurate for reaching the goals of each source, the information becomes incomplete and even erratic if we intend to analyse entrepreneurial activity by it.

6.1 Introduction

The study of entrepreneurship and its impact on economic activity has always been on the research agenda for economists. However, it has not been one of the most intensely explored topics. This shortcoming is heightened when we observe that most governments and institutions design and implement entrepreneurial support policies that are not sufficiently verified by empirical evidence. In this sense, most indicators assessing the entrepreneurial network have focused on its quantitative composition and, to a great extent, have been biased to the requirements of firm demography studies. Therefore, it's common to observe that most existing indicators are oriented to the quantification of firms, centres or establishments, and to measure their dimensions basically in terms of their number of workers. However, the economic analysis of entrepreneurship must also be taken on other grounds. For instance, the study of entrepreneurial activity requires indicators capturing the number of entrepreneurs in a particular sector (or geographic area) and accounting for the way they carry out their task. In conclusion, it is necessary to go deep into the causes that determine the choice of becoming self-employed as well as the duration and evolution of these business ventures. Therefore, suitable knowledge of relevant statistical sources (i.e. Human Population Surveys) and the continuous process of development, updating and improvement of these sources, constitutes a mandatory

José M Millán
Universidad de Huelva
jose.millan@dege.uhu.es

requirement to test the basic propositions arising from the theoretical analysis of entrepreneurship.

The purpose of this paper is to identify and evaluate the information sources available to study entrepreneurship in Spain. The paper is organized as follows. In Section 2, we review some potential statistical sources. We also briefly present the Human Population Surveys available. In Section 3, we enumerate the exploitation techniques used in the Human Population Surveys. We also summarize the existing literature associated to these techniques and we devote special attention to works using Spanish data. In Section 4, we compare the Human Population Surveys focusing on the advantages and disadvantages of each survey and in the literature that has exploited them. Finally, Section 5 is reserved to conclusions and future perspectives for the Human Population Surveys in Spain.

6.2 Statistical Sources and Human Population Surveys

In regards to the Human Population Surveys, a brief review of the existing statistics shows us the extensive variety of available sources. First, there's the renowned Spanish Labour Force Survey (EPA), which addresses households and has been conducted by the National Statistical Institute (INE) since 1964. The EPA is the basic structural source yielding information about the characteristics of employment in Spain. It also provides information regarding unemployment and the population outside of the labour force. Its quarterly regularity also allows a follow up on the labour market situation.

There are other statistical sources, based on household data, that provide information on the working population (although their primary goals are not to measure employment): the Working and Living Conditions Survey (ECVT 85), the 1991 and 2001 Population Census, the 1991 Sociodemographic Survey, the Household Budget Survey (HBS), the Quality of Life at the Workplace Survey (ECVT) and the European Community Household Panel (ECHP).

The Working and Living Conditions Survey was conducted during the last quarter of 1985 by the Spanish Ministry of Labour. Its main purpose was to assess the informal economy and its importance within the labour market.

The main objective of the Population Census is to recount the entire population comprised in each of the Spanish's administrative units and population settlements. In addition, the Census intends to provide information on the population structure to facilitate the decision-making process' direction.

The Sociodemographic Survey was conducted in 1991 and complemented the same year's Population Census information, mainly regarding the life histories of those surveyed on topics such as educational and social background, migration, etc.

Conducted by the INE since 1958, the Household Budget Survey is a statistical operation with a long-standing tradition. Its main objective is to estimate the weights used in the Consumer Price Index.

Started up by the Ministry of Labour in 1999, the Quality of Life at the Workplace Survey (ECVT) is the unique nation-wide Spanish survey specifically designed to study the quality of life at the workplace.

The European Community Household Panel (ECHP) is a standardized survey for the EU-15 member states. Carried out between 1994 and 2001, it was conceived to study and monitor social cohesion, to study population needs and the impact of socioeconomic policies on households and individuals, and to help design new policies.

The European Statistics on Income and Living Conditions (EU-SILC), is an annual survey conducted to obtain information on household income, poverty and social exclusion. It started in 2004 as an improvement over the ECHP whose content needed to be updated in accordance with the new political demands and the need for faster data generation.

In 2001, the Spanish Central Bank decided to start the Spanish Survey of Household Finances (EFF).¹ The central purpose of this survey is to obtain detailed information concerning the financial position of Spanish households. The EFF is a unique statistical source in Spain that allows the linking of household revenues, assets, liabilities and expenditures.

Another important indicator of entrepreneurial activity is the affiliation to the Social Security system. Today's social security records are designed more for managerial purposes than to conduct population studies. Nevertheless, this situation has changed with a new database issued by the Ministry of Labour, the Continuous Survey of Work Histories, whose information is already available to research Centres.

Finally, there is the Global Entrepreneurship Monitor project (GEM)² which started in 1998. The GEM is an international research programme intending to generate harmonised annual data on entrepreneurship. It first started with 10 countries and currently covers 39 countries including Spain since 2000. The data it uses comes from a survey that aims at capturing the following dimensions: i) entrepreneurial activity; ii) attitudes and perceptions vis-à-vis entrepreneurship; iii) entrepreneurial environment; iv) a series of standardised questionnaires for experts.

6.3 The Use of Human Population Surveys in Empirical Research

Most textbooks approach the entrepreneurial phenomenon from a Business School perspective, setting aside any possible contributions derived from economic theory and empirical research. Nevertheless, the existing surveys on entrepreneurship

¹ See Bover (2004).

² For detailed information on the project see Reynolds, P. *et al.* (2005). On-line information on the International GEM project can be found at <http://www.gemconsortium.org>. On-line information on the Spanish GEM project is available at <http://www.ie.edu/gem>. For on-line information on the project for Andalucía visit <http://www.gem-andalucia.org>.

literature—Audretsch (2002), Blanchflower (2000, 2004a), Parker (2004) or Reize (2004) among others—refer to numerous works that overcome this deficiency. Their main contributions include the following models and techniques: discrete choice models, sample selection models and earnings functions, duration models, Cointegration estimators for time series, and decomposition techniques.

However, not all of these models are fit to use the micro-data offered by the Human Population Surveys (e.g. time series models) nor have they all been implemented with Spanish data (e.g. decomposition techniques).

The discrete choice models have been widely used to identify the factors inducing self-employment (such as the case of an unemployed individual becoming self-employed or an employee switching to self-employment). They may be divided into either binomial or multinomial models. In the former, the individual's decision is simply restricted to being self-employed or not, while the latter involves a wider range of choices. For instance, the individual may face the alternatives of being a paid-employed (employee), a self-employed with employees (employer) or a self-employed without employees (own-account worker).

Along these lines of research, the works by Evans and Leighton (1989), and Blanchflower and Oswald (1998) deserve special attention. In addition, some important works using Spanish data are: Alba-Ramírez (1994), based on the Working and Living Conditions Survey (ECVT 85), Carrasco (1999), based on the Household Budget Continuous Survey (HBCS); Aguado *et al.* (2003), Carrasco and Ejrnæs (2003), and Congregado *et al.* (2005), based on the European Community Household Panel (ECHP); and Congregado *et al.* (2003), based on the HBCS and the ECHP.

Sample selection models try to estimate the probabilities and expected profits associated to self-employed and paid-employed individuals controlling for possible selection bias. Selection bias may arise if self-employed individuals have special characteristics that make them more suitable for self-employment. In this case, controlling for the bias would allow to know whether the same individuals could improve their revenues in an alternative occupation.

Along these lines of research we may highlight the works by Taylor (1996) and Parker *et al.* (2005). The work by García and Montuenga (2004) also deserves special attention: using data from the ECHP, it compares the education returns of self-employed individuals and employees in Spain and Portugal.

Duration models seek to identify the variables affecting the duration of self-employment. These models use either Human Population Surveys (which provide information on the individual's self-employment spell) or firm registries (which provide information on the span between firm entry and exit). We may distinguish between two types of duration models: single-risk models and competing-risk models. Single-risk models are used when a transition can lead to only one destination or exit state (e.g. the transition from unemployment to employment) while competing-risk models allow for multiple destinations (e.g. the transition from unemployment to part-time or full-time employment). Using Human Population Surveys, Bøheim and Taylor (2000) and Falter (2002) make special contributions in these directions.

Works using Spanish data that also deserve to be mentioned are: Carrasco (1999), which uses the HBCS; and Congregado *et al.* (2003), based on data from the ECHP.

Finally, decomposition techniques are used to explain differences between socioeconomic subgroups of the population (e.g. divided by gender or by ethnic background). These techniques allow us to determine whether the differences obtained in variables such as earnings and self-employment survival, are due to special characteristics of each subgroup or to possible discrimination. Some of the main works along these lines are Borjas and Bronars (1989) and Hundley (2001).

6.4 Entrepreneurship and Human Population Surveys

In this section we go further into the characteristics of the Human Population Surveys that make them suitable to study entrepreneurship in Spain. We also point out the exploitation possibilities offered by each survey in relation to the existing empirical approaches to entrepreneurial activity.

6.4.1 Spanish Labour Force Survey

Research on firm demography has intensively explored the indicators related to the number of firms or establishments in a particular geographic area or a specific economic sector. In Spain, studies in this direction have grown substantially since the establishment of the Central Companies Directory (CCD) and the Survey on Labour Juncture (ECL). Developed by the National Statistical Institute, the CCD is a statistical source that was preceded by the Economic Directories Integration Project (EDIP) of late 1989. In a unique information system the CCD consolidates data from all the Spanish firms and its local units that are situated in the national territory. The ECL was established in 1990 and is published quarterly by the Ministry of Labour. It uses data of over 12.000 establishments (all of them with more than 5 workers) gathered from the industrial, building and services sectors (Public Administration and Defence are excluded from the last one).

The statistical sources mentioned above are useful to study corporate entrepreneurial activity but fail to explain entrepreneurship from an individual perspective. To overcome this shortcoming and assess the whole entrepreneurial network (i.e. at individual and corporate levels) we must resort to the Spanish Labour Force Survey (EPA). This survey currently has a rotating panel sample (derived from the 2001 Population Census) of 65.000 households (covering approximately 200.000 persons). It is conducted quarterly and it allows a follow-up on the employment status and the type of occupation for the same persons during six consecutive quarters. The EPA itemizes working individuals into the following categories: employers (self-employed with employees), own-account workers (self-employed without employees), producers' cooperatives members, contributing

family workers, employees and other. Thus, the number of employers and own-account workers constitutes a good proxy to quantify the individual entrepreneurial network.

Nevertheless, we are aware of the predominance of corporate firms in economic activity. In this kind of firms, there exists a clear separation between ownership and control and therefore the entrepreneurial activity may be carried out by corporate officers who are not necessarily firm shareholders. Therefore, in order to assess the whole entrepreneurial network we need indicators accounting for the corporate entrepreneurial activity. Again, the Spanish Labour Force Survey (together with the Population Census) may provide this kind of indicators. When classifying workers by occupation (National Classification of Occupations, CNO-94), there is a category for private and public business executives. The classification of workers by socioeconomic condition includes four categories: directors and heads of agrarian establishments, directors and managers of non-agrarian establishments, directive staff of the public administration and members of the state offices. Finally, when classifying (if applicable) the type of public administration the individual works at, there is a category for public companies and financial institutions. Therefore, the information pertaining these three classifications allows us to identify the private and public business executives. It is up to the researcher's criteria to either consider the hypothesis that no public worker is an entrepreneur, or to assume that those executives working for public companies or public financial institutions are involved in entrepreneurship. Therefore, it is possible to obtain approximate data on the individual and corporate entrepreneurial network, both at the national and regional level. The most significant shortcoming of this source derives from the fact that to a greater level of disaggregation corresponds a greater sample error.

Once we have clarified the way in which the EPA survey allows us to identify the entrepreneurial network in Spain, it is helpful to review the additional information provided in the survey that might be relevant to explain entrepreneurship. In addition to the personal and sociodemographic details of each household member in the sample, the survey provides extensive data on the individual's working situation: current employment (including working time and the economic sector of the activity according to the National Classification of Economic Activities, CNAE-93); job search and unemployment span (allowing to differentiate unemployment from voluntary unemployment); work experience (although referring only to the immediately previous job); educational background, etc.

Surprisingly, the EPA survey has rarely been used to study entrepreneurial activity. To a great extent, this is explained by the fact that the EPA does not include information on individual income and wealth (by contrast, this information is included in the rest of the European Labour Force Surveys). EUROSTAT tried to overcome this shortcoming by conducting pilot surveys in 2004 while expecting to begin the incorporation of the produced data to the EPA in the first quarter of 2005. However, the quality of the information collected did not meet the minimum reliability requirements and the project was put on hold. This is an important drawback since most of the existing literature emphasizes the key role played by liquidity

constraints when deciding to start a business venture³. In this direction, three important works for the Spanish case are Carrasco (1999), and Congregado *et al.* (2003, 2005). Carrasco uses discrete choice models (binomial and multinomial) and data from the Household Budget Continuous Survey (HBCS) to show a positive correlation between family assets and the probability of switching from paid-employment to self-employment. For the multinomial case, i.e. when distinguishing between self-employed with employees (employer) and without employees (own-account worker), he shows that the correlation is positive for both cases although it is greater for the former. Congregado *et al.* (2003) obtain similar results when using data from the HBCS survey. However, when using data from the ECHP survey they find that the probability of switching to the own-account worker state is invariant to the individual's capital and labour income. Nevertheless, in Congregado *et al.* (2005) they include the last two waves of the ECHP and, contrary to their previous work, they obtain results consistent with Carrasco (1999).

Because of its characteristics, the EPA survey can be considered as an appropriate source to be exploited by discrete choice models, sample selection models and some decomposition techniques. In addition, given that it has been conducted quarterly since 1964, it allows us to adjust for the impact of the economic cycle, thus making it possible to test the hypothesis that aggregate economic conditions affect entrepreneurial activity. However, while the ECHP is a fixed-panel (i.e. there is no sample panel renovation), the EPA is a rotating-panel survey where the respondents remain in the sample during at most for six quarters. This fact together with the lack of information on current employment spell, makes it an inadequate survey to study self-employment through duration models. However, the EPA does collect information on the unemployment spell as a discrete variable since the respondent is asked to choose among intervals of different lengths to reflect the duration of his unemployment spell. This information, which may seem unimportant, becomes relevant when estimating unemployment duration models under two possible scenarios: when self-employment is the unique alternative to unemployment (single risk model) and when there are more alternatives (competing risk model). Finally the lack of information on individual income and wealth prevents us from exploiting the EPA through earnings functions techniques. However, Arellano and Meghir (1992) are able to estimate a labour supply function by combining two databases: One using the U.K. Family Expenditure Survey (equivalent to the HBCS in Spain) which has detailed information on individual income. The other using the U.K. Labour

³ See Parker (2002) for a literature survey on this subject. Rees and Shah (1986), Evans and Jovanovic (1989), Evans and Leighton (1989), Dolton and Makepeace (1990), Fujii and Hawley (1991), Holtz-Eakin *et al.* (1994), Blanchflower and Oswald (1998), Clark and Drinkwater (2000), Bernhardt (1994), and Parker (2003) find evidence supporting the existence of liquidity constraints. By contrast, De Wit and Van Winden (1989, 1990, 1991), De Wit (1993), and Grilo and Thurik (2004) do not find evidence supporting their existence. Finally, Gill (1988), and Earle and Sakova (2000) find a negative effect of capital.

Force Survey (equivalent to the EPA) which provides the necessary information on working situation and job search. The fact that information on working time is included in both surveys makes them compatible to estimate the labour supply function.

Concerning the methodological changes experienced by the EPA, we point out the incorporation, since the second quarter of 1987, of a wider and more complex questionnaire that includes new definitions in accordance to the EUROSTAT criteria and the International Labour Organization (ILO)'s recommendations. With this reform, the EPA provides more complete and detailed information on subjects such as: underemployment; unemployment benefits; working time (full-time or part-time) and job contracts (temporary or indefinite). In addition the EPA began to use the National Classification of Economic Activities (CNAE-93) and the National Classification of Occupations (CNO-94; this classification introduced modifications in the socioeconomic condition categories), since the first quarter of 1993 and the second quarter of 1994, respectively. The educational variables used by the EPA (educational level, current studies and area of studies) are codified according to the National Classification of Education (CNED-2000) which substitutes the *ad hoc* classification used before.

Finally, there is a high degree of comparability between the EPA and the Labour Force Surveys conducted in the rest of the European Union. This is so because EUROSTAT provides the criteria to be adopted by the Communitarian Countries in order to homogenize the Labour Force Surveys conducted in each one of them. Moreover, it is important to point out that EUROSTAT carries out an annual survey, the European Union Labour Force Survey (EU-LFS), which includes the second quarter EPA of each year. In addition, two types of modules of survey questions are elaborated: standard modules (comprising questions to be permanently applied in the survey), and focal modules (consisting of questions to be applied in particular quarters).

6.4.1.1 Standard Modules

Facilitating the incorporation of young people into the labour market constitutes one of the most important points in the fight against unemployment in the European Union. To deal with this issue, EUROSTAT decided to elaborate a standard questionnaire on training and education to be applied gradually into the different surveys. In the EPA survey for instance, what was initially a single question concerning the level of completed studies, in 1998 turned into a module comprising questions on current or recent participation in training activities (level, type, duration, etc.) and on completed studies (level, year of completion, etc.). In order to achieve comparability across countries, the module also used the same codes stated in the International Standard Classification of Education (ISCED 1997). In this way, the module harmonisation facilitated the analysis of the transition process from school to the work force in the European Union.

6.4.1.2 Focal Modules

Following the lines established by the EU, since 1999 the EPA survey has incorporated, in its second quarter questionnaires, focal modules related to different topics of the labour market. Some of the topics treated in the focal modules were: labour hazards (1999); the transition process from school to labour market (2000); working relations, conditions and timetables (2001); disabled persons and employment (2002); education/training courses taken in the preceding 12 months (2003).

Due to its importance in the study of entrepreneurship, the 2000 module deserves special attention. The primary objective of this focal module was to determine the relation between the educational background of an individual and their first job obtained after leaving school, as well as the time spent during the job search. It also related the educational level of an individual to the ones of his progenitors. It's important to note that this focal module reinforces the effort carried out by EUROSTAT in the standard module previously discussed.

6.4.2 Working and Living Conditions Survey

The Working and Living Conditions Survey was carried out by the Ministry of Labour during the second half of 1985 to estimate the informal economy and its importance within the labour market. It used a sample of over 60.000 individuals collecting extensive information on the Spanish labour force. Moreover, it allowed us to distinguish between own-account workers and employers and, within this last group, those employing more than 5 workers. Without any doubt, its major drawback to assess current reality is its outdated data. Two important works exploiting this source are Alba-Ramírez (1994) and Gil, Martín and Serrat (1994). The first one uses discrete choice models to study how the unemployment spell affects the transition probability from paid-employment to self-employment. The results are later compared to the ones available for the United States. Furthermore, it estimates earnings functions for self-employed, employees and overall workers. On the other hand, Gil, Martín and Serrat estimate an unemployment duration model in which the possible exit states are self-employment and paid-employment (competing risk model) and they compare it to a simpler model where there is only one exit state (single risk model).

6.4.3 Population Census

A Demographic Census is the statistical project of greatest range that the National Statistics Office of any country must undertake periodically. The denomination, Demographic Census, includes three different censuses: Population, Housing and Building Census. Of these three, the Population Census is, without any doubt, the most important and long-standing. In Spain, the first modern Population Census took place in 1768 and since 1901 it is carried out every 10 years (before 1900,

four population censuses where undertaken so that the 2001 Census is officially the sixteenth Spanish Population Census).

Today, the Population Census is conducted by the INE, and its main objective is the recount of the entire population comprised in the Spanish administrative units and population settlements. In addition, the census intends to provide information on the population structure to facilitate decision-making processes. Thus, by assessing the geographic, demographic, cultural, economic and social characteristics of the inhabitants, the structural image of the population provided serves as a guideline for the design of demographic, economic and social policies.

In accordance with international recommendations and in order to reach comparability with other countries' censuses, the Spanish Population Census covers all the persons whose habitual residence is in the Spanish territory. This includes individuals in exceptional circumstances on Census Day such as diplomatic personnel on official duty abroad and Spanish residents working temporarily abroad.

The 2001 census made considerable improvements over the 1991 census leading to a significant reduction in its workload and costs. Among these improvements were: a more efficient use of the information collected by the Municipal Registers of Inhabitants, the simplification of questions by eliminating marginal answer choices, and the exclusion of certain questions for being either too vague or because they were already included in other surveys conducted by the INE.

Finally, and with respect to the information provided by the census, besides the personal and sociodemographic details of each individual, it includes data on the individual's migration movements, education level, marital status, fecundity status and employment status. Concerning this last one, the census clearly differentiates employers from own-account workers and it includes the individual's occupation, workplace, economic activity and sector of employment. Given the above, the population census provides important information to study the labour force and entrepreneurial activity. However, it also presents some shortcomings: First, it lacks information on the individual's income thus giving rise to the problems already discussed for the EPA survey. Second, the long periods between censuses prevent assessing the situation in a continuous manner. And third, data is collected by self-registration (i.e. the respondent fills in the questionnaire) thus limiting the role of census officers in controlling the information obtained. Because of these shortcomings, duration models and earnings functions cannot be estimated. In addition, discrete choice models and sample selection models must be based on the participation in self-employment (but not on the transition decisions from one state to another because of the static condition of the census). Therefore, it is not surprising that this source has been seldom used to study entrepreneurial activity in Spain.

6.4.4 Sociodemographic Survey

The 1991 Sociodemographic Survey was carried out as a complement to the 1991 population census and it used a sample of over 160.000 individuals. The main

drawback facing this survey is that it has not been updated. Besides the large amount of respondents, its big advantage is its retrospective feature: by means of a single interview it allows the reconstruction of each respondent's history of employment, geographical mobility, educational background, etc. This single-interview method has the advantage of avoiding the problems associated to fixed-panel surveys such as the ECHP, in maintaining the respondent's collaboration during extraordinarily long periods. By contrast, it has the shortcoming of being based on the respondent's memory.

The testimonies left by the respondents are of great significance since they reflect the history of the first 90 years of the 20th Century. In the eldest group, those born before 1911, 55% of the cases began to work before the age of 14 and 80% of the cases before the age of 16. This kind of information is no longer gathered by the surveys on labour activity such as the EPA because of the compulsory schooling covering those ages, which gives an idea of the transcendental change that has taken place in Spain over that period.

6.4.5 Household Budget Surveys

The Household Budget Surveys are among the most long-standing operations in Official Statistics. The first studies on household expenditure for Europe took place in the middle of the 19th century. They began in Spain in 1958 and have been providing information on the consumption expenditure of private households ever since. However, the relevance of this source for our purposes resides in the additional information the survey provides for each household member regarding demographic characteristics, education level, employment status, occupation, and income

The INE has traditionally carried out two types of HBS. The first type is constituted by the Household Budget Basic Surveys conducted in 1958, 1964–65, 1973–74, 1980–81 and 1990–91. Among their various goals, these surveys have provided estimations regarding the level and structure of annual household consumption and the weight structure of expenditure serving to calculate the Consumer's Price Index. The second type is formed by surveys the INE conducted quarterly to estimate household expenditure and its annual variation. These continuous surveys started with the Permanent Consumption Survey (EPC) which was carried out from the second quarter of 1977 until the last quarter of 1983. The EPC was designed as a panel of 2.000 households rotating every four quarters. Later on, from the first quarter of 1985 until the first quarter of 1997, the rotating-panel Household Budget Continuous Survey 1985, was carried out covering 3.200 households each quarter. In this case, the rotation speed was of 1/8 of the sample each quarter.

The coexistence over time of the two types of HBS presented some disadvantages, the most important being related to their costs. In order to optimize resources and in accordance to the European harmonisation recommendations of the Household Budget Surveys, the INE, under the coordination of EUROSTAT, unified the two types of surveys into the Household Budget Continuous Survey 1997 which

started in the first quarter of 1997.⁴ From then until the last quarter of 2005, approximately 8.000 households were interviewed maintaining the quarterly rotating-panel design (with a speed of rotation of 1/8 of the sample each quarter).

In response to the users' new demands and in accordance to the international recommendations by EUROSTAT, in 2006 a new survey was initiated, the HBCS 2006. Its main purpose is to ensure the maximum quality of the information provided on annual household consumption expenditure and its annual variations. The methodological design was simplified thus resulting in a considerable reduction in the respondent households' workload. The annual sample is designed to cover approximately 24.000 households, half of which are renewed each year. Every household in the sample will collaborate during 14 days in each of the two consecutive years. As a considerable improvement over the HBCS 1997, the information processing will take place in the provincial deputations of the INE where specific errors and inconsistencies will be controlled. This temporal and physical proximity to the respondent households will improve the quality of the provided information.

Having in mind the study of entrepreneurial activity, we will now proceed to point out the strengths and weaknesses of the HBCS' different versions. The main limitation of the HBCS 1997 with respect to its predecessor is the information it provided on household income. The HBCS 1985 data base offered detailed information on each household member's income and its source from the first quarter of 1985 until the first quarter of 1997, which allowed precise estimations of earnings to be carried out. On the contrary, the HBCS 1997 information on income referred to the household unit which makes it very difficult to impute the respective income to each one of its members. This shortcoming hinders the testing for the presence of liquidity constraints in any given discrete choice model. Therefore, the new European Statistics on Income and Living Conditions (EU-SILC) to be discussed later, constitutes the appropriate source to study household income in detail. Moreover, the information provided by the HBCS 1985 allows us to identify the employment status of the spouse of the head-of-household as employer, own-account worker, employee or unemployed; while the HBCS 1997 only distinguishes the spouses working from those not working. By contrast, the HBCS 1997 information provided on the head-of-household far surpasses its predecessor's. The HBCS 1985 information concerning the head-of-household's main activity was very limited, failing to identify its economic sector even at the most aggregated level (i.e. agriculture, industry and services). Since there are significant variances in the characteristics and behaviours of individuals working in different economic sectors, the impossibility of controlling effects in the activity's sector will induce a bias in the results of any econometric exercise on occupational choice. The HBCS 1997 clearly overcomes all these shortcomings by providing detailed

⁴ With the incorporation of the 2001 Population Census data, the survey's results were revised starting in the first quarter of 1998. However, since the last two quarters of 1997 were not revised, the microdata files corresponding to these quarters are not comparable with the rest. For this reason, quarterly files are only available beginning in the first quarter of 1998.

information on the head-of-household's economic occupation and its economic sector, following the National Classification of Occupations (CNO-94) and the National Classification of Economic Activities (CNAE-93) with all the international comparability advantages that these classifications imply. The CNO-94 even allows the identification of the corporate entrepreneurial network, i.e. those corporate officers undertaking entrepreneurship. Additionally, and unlike the HBCS 1985, the HBCS 1997 includes head-of-household information on working hours and contract type, and it identifies whether his occupation belongs to the private or public sector.

Both surveys (HBCS 1985 and 1997) have the advantage of covering a considerable time span and, along with their quarterly periodicity, they allow for controlling the effects of the economic cycle on the entrepreneurial activity and, in particular, on the individual's decisions. However, it's a pity that the information on education refers only to the head-of-household thus biasing any possible results concerning the effects of education on entrepreneurship to this group of individuals. Finally, while the HBCS may be considered appropriate to assess entrepreneurial activity through discrete choice models, sample selection models and some decomposition techniques, it is inadequate in estimating self-employment duration models because its rotating panel feature allows for a two-year monitoring at maximum.

Regarding the changes introduced in the HBCS 2006, first, it is important to notice that these types of surveys have now gone beyond the pure economic scope, and have begun pursuing important social and socio-economic objectives as well. Switching from quarterly to annual periodicity certainly has a positive effect in cost reduction but it also represents an important disadvantage when studying the entrepreneurial network because the greater span between surveys prevents the implementation of discrete choice models and duration models. This could be overcome by designing the survey's questionnaire so as to accurately retrieve the information pertaining to the in-between surveys periods.

As we discussed earlier, some important works using the HBCS are Carrasco (1999) and Congregado *et al.* (2003). Carrasco uses the HBCS 1985 covering the period 1979–1990 to study the determinants of the individual's transition from different initial states (unemployed or paid-employed) to different final states (employer, own-account worker) through discrete choice models. In addition, he uses duration models to study the determinants of the duration in self-employment. For this last exercise, the HBCS' rotating panel feature prevents the observation of self-employment periods longer than two years. Moreover, the reduced number of observations forced the author to exclude durations of over three quarters clearly illustrating the limitations of the HBCS source to estimate duration models. Congregado *et al.* (2003) used both the HBCS 1985 and the HBCS covering the 90s decade, to study the same issues as in Carrasco (1999) but also to study the variables affecting the individual's decision to switch from own account worker (self-employed without employees) to employer (self-employed with employees), considering this decision as an indicator of entrepreneurial success.

6.4.6 *Quality of Life at the Workplace Survey*

The Quality of Life at the Workplace Survey (ECVT), started by the Spanish Ministry of Labour in 1999, is a sample survey of Spain's employed population, specifically designed to study the quality of life at the workplace at a national level. The ECVT collects objective information on working conditions at the workplace as well as subjective information concerning the personal perceptions that the employed individuals have of their working conditions and relations.

The survey has a sample size of 6.020 employed individuals (ages 16 and over) living in family dwellings. Its geographic scope covers the Spanish national territory with the exception of Ceuta and Melilla. The information provided by this survey is articulated in different modules covering working status, family arrangement, socioeconomic data and information on the quality of life at the workplace.

The module comprising working status provides information regarding the real situation the individual has at his workplace. This includes the size of the firm, type of contract and working time (part or full-time), years in the company, way in which he searched and obtained the job, secondary occupation (if applicable), etc. Concerning the possibility of assessing entrepreneurial activity, the survey allows the identification of self-employment by distinguishing paid-employed from self-employed individuals (but among this last group, it is not possible to identify those having employees). In addition, the National Classification of Occupations (CNO-94) also provides information on the individual's occupation that allows us to identify the corporate entrepreneurial network and the economic activity by applying the National Classification of Economic Activities (CNAE-93). This information is complemented by data on the individual's work experience such as the age and education level at which he started his first job, past occupations, unemployment situations, subsidies received, etc. There is also relevant data on the individual's family social mobility such as his birthplace and that of his progenitors, the educational and labour characteristics of his family, the current family structure and the working situation of the persons living with him.

The family arrangement module provides information on the individual's conjugal status, relationships, family size and type, among others. The socioeconomic module gathers information on the individual's gender, age, marital status, education level, company size, municipality size, etc. The individual is also asked to reveal his political and religious preferences.

Finally, the module on the quality of life at the workplace refers to the individual's attitudes towards his job and the level of satisfaction with it. It collects information on the job's levels of division, organization and communication; on the safety at work and on collective bargaining. It also gathers the individual's opinions on the level of alienation, participation, integration and autonomy at the workplace, and on his working time and retribution. In addition, the module evaluates the level of occupational training and its relation with job promotion.

Presently, the ECVT and the ECHP are the only nation-wide Spanish surveys providing information on the workers' level of satisfaction with their jobs. The ECHP, being a full panel survey, allows for the studying of the dynamic aspects

in satisfaction levels. The ECVT, despite renewing the sample completely each year thus preventing controlling for the economic cycle, provides extensive information on the jobs' characteristics and is therefore particularly adequate for analysing the sources of job satisfaction. Even though the absence of income data prevents estimating earnings functions, the collected information may be exploited via sample selection models or discrete choice models to estimate the probability of being self-employed (but not the transition probability). Regarding the duration models, despite having information on the job tenure, we may only reconstruct incomplete durations (individuals that we do not observe to leave self-employment) but not complete durations (individuals that we observe leaving self-employment).

Briefly, the ECVT provides subjective information, absent from the majority of statistical sources, that might be a good complement in studying important aspects of job satisfaction in self-employment compared to paid-employment. However, it lacks the suitable structure and the appropriate sample size to be used as a source of reference in the analysis of the entrepreneurial activity.

6.4.7 European Community Household Panel

The European Community Household Panel (ECHP) is an EU-specific full panel survey using harmonized data of 76.500 households covering over 155.000 individuals in the 15 member states (7.200 households and 15.900 individuals for Spain). It constitutes one of the most important statistical instruments for the European Commission and it reinforces EUROSTAT's current statistical infrastructure. Its main objectives are to study and monitor the living and working conditions, social cohesion, population needs, the impact of socioeconomic policies, and to help in designing new policies for the member states.

The survey was designed in close consultation with the Member States through the denominated National Data Collection Units (NDUs). In most of the countries, the NDUs were formed by the National Statistical Institutes because of the required scientific and technological training.

The preliminary studies were mainly oriented to evaluate the possibility of efficiently using the data provided by pre-existing national and EU surveys with similar information that seemed to fulfil the stated objectives: the Spanish Labour Force Survey (EPA) and the Household Budget Survey (HBS), such as in Spain's case. Regarding the EPA, given its focus on the labour market, it seemed inconvenient to overload its questionnaires and very difficult to adapt it for multiple purposes. This, together with the lack of information on income, invalidated this survey as an instrument for the project. The use of the HBS was also discarded because of the specific requirement of absolute homogeneity across countries in the collected data as well as the necessity of widening it to other research areas.

Therefore, it was necessary to build a new statistical operation, coordinated by EUROSTAT and harmonized for the EU member states, but preserving statistical integration with the other household-addressed surveys (EPA, HBS). It was

also considered that the new survey should go beyond the traditional snapshot of transversal information by allowing to capture longitudinal information, i.e. information pertaining to the same set of households and individuals through different periods in time. After several studies and preparatory conferences in the years 1990-1992, a pilot test was conducted in 1993. In 1994, the first wave of eight (initially only three waves were intended: 1994, 1995 and 1996) was conducted in all of the member states.

The ECHP uses the techniques of a full panel which are ideal for the fulfilment of the stated objectives but entail the difficulties associated to the follow-up of the individuals that remain in the sample during excessively long periods⁵. Given the non-existence of previous household full panels in the majority of the participating countries, the harmonization across countries was produced from the very beginning of the project. The panel follows-up on the same set of households and individuals allowing to study the changes in their lives produced by modifications in socioeconomic policies or in aggregate economic conditions, and to capture their reactions. Taking for granted that the effectiveness of a policy must be evaluated according to the way in which the affected react to it, the ECHP will provide very important indicators on the effects of socioeconomic policies framed in the Common Market.

The target population of the survey is formed by the set of private households and individuals living in principal family dwellings in the EU-15 territory (in Spain it is circumscribed to the whole Spanish territory with the exceptions of Ceuta and Melilla).

The information contained in this source has two different reference points that are complemented in a very interesting way: there are independent blocks of questions referring to households and individuals and there is a relationship file that allows us to easily relate the individuals in each household. Thus, on one hand the household file offers data on the composition and the characteristics of each household as well as detailed information on its income. On the other hand, the personal file comprises twelve sections whose brief data content description is as follows:

- General and demographic information: age, gender, marital status.
- Current employment: main activity, status in employment, type of contract (full-time/part-time), occupation in current job, main activity of the local unit of his company or organisation.
- Unemployment: number of times the person has been unemployed, unemployment spells that have exceeded 12 months.
- Search for a job: type of job search the individual has done, conditions in which he would accept to work.

⁵ These difficulties may be overcome through retrospective studies such as the one conducted in the Sociodemographic Survey of 1991.

- Previous Jobs: whether the individual has worked or owned a business for at least 15 hours per week or not, date and reason for stopping in previous job, comparison between the present job or business with the previous one.
- Calendar of activities: monthly follow-up on the main activity status in the year prior to the interview.
- Income: main source of personal income in the year prior to the survey, current monthly net wage and salary earnings, income received from other sources.
- Education and training: highest level of general or higher education completed, age when the highest level of general or higher education was completed, current studies, and whether the individual has received vocational training paid for or organised by the employer, or not.
- Health: valuation of the individual's health condition, number of times he has consulted a general practitioner or medical specialist, whether he has a state financed health care system or a private medical insurance.
- Social Relations: frequency of relations with the individual's social circle (friends, relatives, neighbours), number of hours spent looking after children or persons (who need special help because of old age, illness or disability).
- Migration: this section analyses factors such as the migration trajectory of the individuals, their current region of residence and their arrival year to the region.
- Satisfaction with various aspects of life: degree of satisfaction with work or main activity, financial situation, housing situation, amount of leisure time.

The study of the labour market is one of the multiple topics that may benefit from the use of the ECHP source. In this sense, some of the important issues in the labour market that may be analysed are the transitions in school-active life (also treated in the standard module of the EPA since 1998), unemployment-employment, within employment, and employment-retirement. This allows the evaluation for instance of the effectiveness of professional training policies as well as the retirement policies. Nevertheless, we should point out that this type of transition analysis sometimes faces an insufficient sample size (e.g. when dealing with infrequent cases). The ECHP is also helpful in illustrating characteristics of the impact of social assistance policies and the role of employment assistance for women. Regarding long-duration unemployment and the return to labour activity thereafter, this source proves to be useful in studying its relation with human capital. As it might be expected, these issues may be redefined in terms of entrepreneurial activity. For the Spanish case, this has been done by the works of Aguado *et al.* (2003), Carrasco and Ejrnaes (2003) and Congregado *et al.* (2003, 2005). In all of them, the available information has been exploited through discrete choice models to study the variables leading an individual to choose to become self-employed. In addition, Congregado *et al.* (2003) develop a self-employment duration model in which, using the first 6 waves of the ECHP and based on the current year's information as well as on retrospective information, they are able to reconstruct self-employment durations of up to 15 years. Because of its characteristics, the panel also allows using sample selection models, decomposition techniques and to estimate earnings functions.

Notwithstanding, the difficulties faced by those who intend to study the entrepreneurial activity through this source deserve some attention. Via the ECHP we may identify self-employed individuals through the employment status stated in the current employment item. However, in order to distinguish among employers, self-employed with employees and own-account workers we must refer to the number of employees as a proxy. But this is obviously an approximation and it lacks the precision one would like to have. There is additional information on the individual's occupation and on the establishment he works at via the International Standard Classification of Occupations (ISCO-88) and the Nomenclature of Economic Activities (NACE-93). In the case of the ISCO-88 it is possible to identify the corporate entrepreneurial network or, in other words, the business executives.

As an additional disadvantage for the use of this source in Spain, the identification of the Spanish geographic zones is made according to the Nomenclature of Territorial Units for Statistics⁶ level 1 (NUTS level 1), dividing its national territory in Northwest, Northeast, Madrid, Centre, East, South and Canary Islands. But this level of aggregation prevents us from studying the impact of the economic cycle because in Spain the economic-cycle indicators are generated at the aggregation level NUTS-2.

There exists nevertheless, a different and wider sample corresponding to the year 2000 that covers 15.600 households in Spain and that allows to use the NUTS level 2 classification. By using this sample it is possible to overcome the aforementioned problem but the households in the sample do not correspond to those of the full panel initiated in 1994. Therefore, the sample must be treated as a transversal cut over time thus allowing us to estimate the probability of being self-employed, but not the probability of transition from one state to another. In addition, the duration models cannot be implemented using this cut either.

Given all of the above, the ECHP constitutes a harmonised source that allows us to compare the characteristics of the entrepreneurial network at an international level. It includes information as relevant as the one referring to income (improving over the EPA information) and as limited as the one pertaining to the degree of satisfaction of the individuals with their jobs (only available in the ECVT). Nevertheless, for some of the information that will be obtained for the whole EU-territory, there are already more appropriate sources than the ECHP as, for example, the EPA or the HBS for the Spanish case. Concerning future perspectives, once the ECHP project was completed, a new instrument replaced it since 2004: the European Statistics on Income and Living Conditions (EU-SILC). We devote our next subsection to it.

⁶ The NUTS is a three-level hierarchical classification in which each Member State is first divided in a number of regions (NUTS level 1), each of which is subdivided into smaller regions (NUTS level 2, corresponding to *Comunidades Autónomas* in Spain, *Regierungsbezirke* in Germany, *Régions* in France, *Regioni* in Italy, etc. . .), which in turn are subdivided again (NUTS level 3, corresponding to *Provincias* in Spain, *Kreise* in Germany, *Départements* in France, *Provincia* in Italy, etc. . .).

6.4.8 European Statistics on Income and Living Conditions

The European Statistics on Income and Living Conditions (EU-SILC) is an annual EU-harmonised survey coordinated by EUROSTAT that began in 2004 in Spain. It constitutes an appropriate source for the study of household income, income distribution and social exclusion. Between 1994 and 2001, the European Community Household Panel (ECHP) fulfilled these necessities. However, given that it was necessary to update its content in accordance to the new political demands, and that its functioning needed to be improved (mainly regarding the speed in data production), it was decided that ECHP be replaced by a new instrument with wider coverage; the EU-SILC.

Thus, the EU-SILC was launched in 2003 as a gentleman's agreement between six countries of the EU-15 (Belgium, Denmark, Greece, Ireland, Luxembourg and Austria) and Norway. In 2004, it was re-launched with a more demanding coordination and included most of the EU-15 countries (with the exception of Germany, the Netherlands and the United Kingdom) as well as Estonia, Norway, Iceland, and Turkey. In 2005 the remaining EU-25 member states were incorporated as well as Bulgaria and Rumania. Switzerland will join in 2007.

In Spain, the survey has a rotating panel design where one quarter of the sample is renewed each year allowing the observation of the studied variables' evolution. Around 15.000 households are interviewed and each household remains in the sample during four consecutive years. The households are distributed throughout the entire Spanish territory which allows us to have information at the NUTS-2 level (i.e. *Comunidades Autónomas*) for most of the included variables.

The main objective of the survey is to provide information on income, income distribution and social exclusion in Spain, and to allow comparisons with the other EU countries. It is expected that the survey will include different thematic modules in order to approach relevant social aspects such as social participation and the intergenerational transmission of poverty. More specifically, the EU-SILC is designed to collect data on: household income and economic situation; poverty and social exclusion; employment and labour activity; retirement, retirement benefits and socioeconomic situation of the aged; housing and its related costs; regional Development, and education, health and their impact in the socioeconomic status.

Household income constitutes an essential part in the EU-SILC for the analysis of the living conditions. It includes wage and salary earnings, benefits/losses from self-employment, social assistance, capital and poverty income, private transfers received, children's income, and after-tax income. Some comments regarding this variable should be pointed out. First, as in the ECHP, the income data refers to the year preceding the interview. Second, while the information on wage and salary earnings refers to the individual as well as to the household, the information pertaining to the benefits/losses from self-employment refers only to the household. Third, the survey only collects the monetary component of income (it does not include, for instance, the estimated owner-occupiers dwelling rent, the non-monetary income proceeding from own production or the income in kind). These non-monetary components are expected to be incorporated beginning in 2007. Contrary to the ECHP

where income was provided in net values, the EU-SILC provides gross income values. This allows for a greater degree of income comparability across Member States because it does not depend on the particular tax-scheme or on the Social Security contributions of each country. Given the initial difficulty in providing gross income data, some countries (Spain included) are allowed to provide net income data during the first years of the survey. Nevertheless, a net-gross conversion process was developed and has been applied since 2004 to obtain current gross wage and salary earnings (i.e. before tax deduction and before Social Security contributions).

The information on employment and Labour activity allows the classification of individuals according to their employment status in the following categories: self-employed with employees, self-employed without employees, employee and family worker. In addition, the survey offers information on the individual's type of occupation (in accordance to the International Standard Classification of Occupations ISCO-88), and on the activity of the establishment he/she works at (in accordance to the Nomenclature of Economic Activities NACE-2002). This will allow the identification of the entrepreneurial activity in a strict sense (i.e. the own-account workers with and without employees) but also the identification of the corporate entrepreneurial network (i.e. the business executives).

In addition to information above, the EU-SILC provides demographic data (age, gender, marital status, place of birth...), information on education, but not as detailed as in the ECHP (highest completed education level and age when completion, current studies...), information on current job (type of contract, number or working hours...), professional experience (characteristics of last main job such as employment status, occupation, type of contract...) job search data (again, not as detailed as in the ECHP) and personal information on general health condition and access to medical care.

This new source constitutes an enormous flow of information that will allow to study, through discrete choice models and sample selection models, all the transitions taking place in the Labour market. The abundant information on net and gross income together with the survey's extensive geographic coverage will allow the estimation of earning functions more accurately; allow for studying the influence of different fiscal schemes; and will provide data pertaining to the role that liquidity constraints play in individual decisions. This faster data availability (just one year after being generated) will allow for a faster redressing of the policies implemented thus gaining in efficiency. Because of its rotating panel feature, the EU-SILC will allow to verify with more reliability the role played by economic aggregate conditions in the individual decision making process. By contrast, the same rotating feature constitutes a disadvantage when trying to estimate duration models because the individuals remain in the sample for four years at most.

6.4.9 Spanish Survey of Household Finances

In 2001, the Spanish Central Bank decided to start the Spanish Survey of Household Finances (EFF), following the example set by other countries in which this type of

survey has been conducted for many years. To be more precise, Italy's "Indagine sui bilanci delle famiglie" (IBF)⁷ and, most notably, the United State's "Survey of Consumer Finances" (SCF)⁸ were the ones that inspired the Spanish survey. The first wave, with a sample of 5.143 households, took place at the end of 2002. The second wave took place at the end of 2005, but from then on it is expected to be conducted every two years. The 2005 sample comprises a panel including the households previously interviewed in 2002, as well as a refreshment sample by wealth stratification. The fundamental objective of this survey is to collect detailed information on the patrimonial situation (i.e. real and financial assets' distribution, debt obligations, etc.) and financial decisions of households in Spain. The EFF-questionnaire is divided in nine main sections: 1) Demographic characteristics; 2) Real assets and associated debts; 3) Other debts; 4) Financial assets; 5) Pension plans and insurances; 6) Labour status and labour income (for all household members); 7) Non-labour income; 8) Means of payment; 9) Consumption and savings.

The EFF constitutes a unique Spanish statistical source linking income, assets, debt obligations and expenditures for each family unit. The most important characteristic of this sample might be the incorporation, through a collaboration system between the INE and the Taxation Agency, of an oversampling of households with a higher wealth level. Since a large proportion of the assets are held by a small fraction of the population, a random standard sample would not contain enough observations for most of the relevant analysis. To illustrate the oversample's importance consider the following: According to the aggregate information on tax statements, 40% of the total taxable wealth is held by 0.4% of the households. Therefore, in a random sample of 5.000 households one would expect to find at most 20 of these households while the EFF sample includes over 500 of them. Nevertheless, we should point out that the oversample does not include households from Navarre and the Basque Country because the Taxation Agency does not have personal fiscal data for those two Spanish Regions.

The EFF's questionnaire allows to clearly identify the entrepreneurs in a strict sense but the identification of the corporate entrepreneurial network is less reliable. In a first filter step, own-account workers are distinguished from the rest (i.e. employee, unemployed, retired, etc.). Then, occupied individuals are asked for their type of occupation (there is only one category grouping private business executives and public executives). The individuals are then asked for the economic sector of their activity and, depending on their working status (i.e. own-account worker, employee or unemployed) each is given a corresponding module.

The own-account worker module provides a big amount of information, including whether it is the individual's main occupation or not, weekly hours worked, number of persons working in the business and whether they are household members or not, the company's legal entity (i.e. entrepreneur—natural person-, corporation, limited liability company, other), working status (i.e. liberal professional,

⁷ See Banca d'Italia (2000).

⁸ See Aizcorbe, Kennickell and Moore (2003).

sole proprietor, business owner, partner in family firm, partner in non-family firm), profits/losses in the year prior to the interview, expected profits/losses for current year, profits share, the individual's shareholding of the firm, shareholding of the firm and the firm's market value.

The employee module provides information on weekly hours worked, part/full-time job, type of contract (i.e. indefinite, temporary, without contract, other), gross labour income, number of working years, initial salary when hired by the company (only if the individual is able to remember it), number of workers in the firm and the expectations of remaining in the firm.

Finally, the unemployed module provides information on the unemployment spell, the sources and the total amount of income and a question regarding the wage at which the individual would accept to work. Note that it would be interesting to exploit this information to test the job search theory based on reservation wages.⁹

In addition, the EFF provides information on demographic characteristics (i.e. age, gender, marital status, citizenship, place of birth, etc.), education (area of studies,¹⁰ highest completed education level and age of completion), health condition, parents' main occupation in their lives, real assets and related liabilities (including real state, means of transportation, jewelry, works of art, etc.), financial assets, pensions and insurance, household laboural and non-laboural revenues, laboural history of the household members, household consumption and savings (expenditures, savings destination, debt financing, risk aversion, extraordinary income, future expectations, etc.). Finally, information on the use of different means of payment, phone banking and e-banking is also collected.

We conclude this subsection by discussing some of the exploitation possibilities of the EFF source. First, the availability of information on income and its sources makes it suitable to estimate earnings functions. Second, despite being a rotating panel, its biannual feature as well as the lack of information on the immediate previous job¹¹ present some difficulties for the implementation of either discrete choice models or sample selection models in the analysis of transitions (nonetheless, it is suitable for participation analysis). Third, because we ignore the sample rotation speed and the continuity of the project (believed to be indefinite) it is hard to evaluate whether it is suitable for the implementation of duration models or not. Therefore, the difficulty in estimating dynamic models prevents the controlling of the economic cycle's impact on individual decisions. Finally, given that entrepreneurs usually have a higher income, the oversampling of higher wealth level households turns this data base into an interesting source for the study of entrepreneurship.

⁹ This theory tells us that the individual searching for a job would accept the job if the wage offered is greater or equal than his reservation wage. Therefore, there is some probability that the individual will not accept to work during some period and will continue to search. This implies that a fraction of the population will remain unemployed which partially explains the unemployment persistence.

¹⁰ Only if possessing a college degree.

¹¹ The working history of the individual is from a general perspective, including the longest held job and whether the individual has been mainly self-employed or not.

6.4.10 Continuous Survey of Work Histories

Another important indicator of the entrepreneurial activity is the affiliation to the Social Security system. Managed by the General Treasury of the Social Security, the information proceeds from the statistical exploitation of the workers' historic files of affiliation to the different social security regimes. Nevertheless, we must clarify that up-to-date samples of this source have only been available to some researchers for conducting very concrete studies¹² on pensions, the influence of temporary help agencies (THAs), rates of employment and unemployment and profits distribution. We cannot forget that social security records are designed more for managerial purposes than to conduct population studies. Therefore, generating suitable data for researchers requires considerable technical work in identifying and documenting the relevant information before extracting it. This difficulty is overcome with the Continuous Survey of Work Histories, a new data base that is already available for research Centres conducting specific projects. The objective is to design a sample supplying the data needed for different kinds of research projects. Naturally, the data is anonymous and necessary precautions have been taken so that the personal information cannot be identified. This continuous survey will be updated every year with new information on the people already in the sample and with a proportion of new individuals who have joined the Social Security during the year. The original sample was taken from all the affiliated persons who were, in 2004, either paying to the social security or receiving contributive or unemployment benefits (irrespective of the length of the unemployment duration). The relevant information available through this source includes age, gender, region of work, whether the individual works for the State or not, type of contract, the company's economic activity (CNAE-93), the type of company (joint-stock, limited liability, etc.), the required qualification for the type of work (which is a good proxy for the education level), dates of activation and withdrawal from the social security system, quitting cause (voluntary, dismissal or retirement), unemployment benefits, and worker's taxable earnings base.

Regarding the advantages of this source, first we must point out that it uses real data rather than data obtained from a survey. Therefore, the information provided on the individual's work history allows for studying mobility in the labour market via discrete choice models and duration models with almost absolute precision and reliability.¹³ The random sample offers 1.1 million anonymous work histories,

¹² See García-Fontes and Hopenhayn (1996), García-Pérez (1997), Jiménez-Martín and Sánchez (2000), Jiménez-Martín and Boldrin (2002), Bover *et al.* (2002), García-Pérez and Muñoz-Bullón (2005); among others.

¹³ In this context, we must remember that a worker has to be part of a social security system all his life whether he is working or not. A worker becomes a member of a social security system on commencing work and even if he ceases to work he will continue to be entitled to social security. If, after ceasing work, a worker resumes working then he will be able to commence work again without the need to re-apply for another social security number because, as we have pointed out, once issued with a social security number ("membership" of the particular social security scheme) it is for life.

representing 4% of the reference population (27.4 million people). The sample will be updated by adding each year 4% of the newly incorporated persons to the social security system. Finally, while the elaboration of the EPA costs 13.5 million euros, the social security sample will have technically zero-cost, due to the fact that the information referring to the individual's work history is already available in the social security records.

Nevertheless, there are some disadvantages that are difficult to overcome. The most important is probably the one related to the individuals' wages. The available information refers to the taxable earnings base which allows to recover total wages in a simple way, except for the cases of very low or very high wages because in these cases payments are established by a minimum and a maximum base. The same problem is presented when the individual is subscribed to the own-account worker special regime, because most individuals in this regime declare according to the minimum base and therefore the information on their perceived income is not accurate. In addition, for partnership companies where the owners are wage earners, these owners appear as employees in the social security records while they appear as own-account workers in other employment statistics (i.e. the number of entrepreneurs is underestimated according to the social security records). Finally and like all the statistics generated from administrative registries, data on affiliation to the social security system is subject to variable legal norms which prevents from obtaining a homogenous series over time.

6.4.11 The Global Entrepreneurship Monitor Project

Finally, we point out the research project Global Entrepreneurship Monitor (GEM) jointly developed by Babson College (Massachusetts) and the London Business School. It was initiated in 1998 intending to generate harmonised annual data on entrepreneurship. It first started with 10 countries and currently covers 39 countries (including Spain since 2000) with a minimum of 2.000 individuals interviewed in each country. Since 2003 there is a regional version of the project, the Regional Entrepreneurship Monitor (REM), which increases and enriches the sample as well as the study's penetration.

Its main objective is to measure entrepreneurial activity in its initial stages for each of the participating countries (this is done in a harmonised way, thus allowing for cross-country comparability). In order to carry out this task, the Total early-stage Entrepreneurial Activity or TEA-index is constructed. This index identifies the starters and owner-managers of new businesses. The starters are those individuals between the ages of 18 and 64 that started a new business in the year prior to the interview and that have not paid more than three payrolls when interviewed. The owner-managers of a new business are those that have not paid salaries or wages for more than 42 months. The sum of the two measures serves to calculate the rates of entrepreneurial activity in each country.

But the GEM not only quantifies entrepreneurial activity, it also compiles information on the economic environment of the businesses and on the influence

of sociological and psychological factors leading to entrepreneurship. Finally, the motivations of potential entrepreneurs are also analysed.

The exploitation possibilities of this source allows for cross-country comparability of entrepreneurial activity rates. However, the definition of entrepreneurial activity that it uses is not comparable with those on which other sources like EUROSTAT or the OECD are based. Also, duration models and earnings functions cannot be estimated because of the lack of panel data, retrospective information, and income data. In addition, discrete choice models and sample selection models can only be based on participation (but not in transitions from one state to another). Because all of the above, it is not surprising that besides the GEM national and regional reports, there are no other studies exploiting this source in Spain.

6.5 Conclusions and Future Perspectives

In this paper we have tried to identify and evaluate the information sources available to study entrepreneurship in Spain. We have seen that the analysis of entrepreneurial activity may be carried out through a wide variety of sources: both in quantitative terms, as in the firm demography studies; and in qualitative terms, through population surveys. We have covered the emergence of new statistical sources that, along with the improvement of the already existing ones, have contributed to enrich the information available to study the labour market.

Nevertheless, in spite of the different organizations' efforts in offering important and substantive amount of information, we see how the suitability of the sources is not fully adapted to the demands of entrepreneurial network analysts: the EPA still does not provide data on income, the ECHP project ended, the ECPF changed from quarterly to annual regularity, the Population Census and the ECVT still do not match the pursued goal, etc. Regarding the new emerging surveys, only the EU-SILC, in spite of its limitations, seems to get close to the ideal: the GEM project will only allow for cross-country comparability of entrepreneurial activity from a particular definition; the Continuous Survey of Work Histories will allow for a better definition of the transitions and the durations but it does not include many entrepreneurs and information is very limited; the EFF will provide enough explanatory information but fails to capture dynamic behaviour. Given that the National Statistical Plan 2005–2008¹⁴ does not even mention the statistical information deficiencies on entrepreneurial activity, future perspectives are therefore not hopeful. Studies on entrepreneurship will still rely more on the skills and imagination of the researchers than on the suitability of the surveys.

Although the available information only allows carrying out a partial analyses of the entrepreneurship phenomenon, not all the blame can be put on data deficiencies.

¹⁴ The National Statistical Plan is the main instrument organizing the statistical activity of the State General Administration. It contains the statistics that are to be carried out during the quadrennial period of reference.

In this sense, part of the problem comes from the fact that the economic analysis of entrepreneurship has not yet reached the degree of development necessary to reveal clear statistical necessities. As a result, there have been erratic uses of sources and indicators depending on the specific approach adopted: individual entrepreneurial network analysis, corporate entrepreneurial network analysis or firm demography studies. As we surpass these obstacles, a homogenization of the demands for this type of data should be created to improve the statistical measurements that would allow capturing the different dimensions in which entrepreneurship affects economic activity.

Acknowledgments The authors would like to thank José Ignacio García Pérez and our colleagues at the Universidad de Huelva, for their helpful comments and advice.

References

- Aguado R, Congregado E, Millán JM (2003) Entrepreneurship, financiación e innovación. *Revista de Economía Industrial* 34(7):125–34
- Aizcorbe A, Kennickell A, Moore KB (2003) Recent Changes in U.S. Family Finances: Evidence from the 1998 and 2001. *Survey of Consumer Finances, Federal Reserve Bulletin*
- Alba-Ramírez A (1994) Self-employment in the midst of unemployment: The case of Spain and the United States. *Applied Economics* 26:189–205
- Arellano M, Meghir C (1992) Female Labour Supply and On-the-Job-Search: An Empirical Model Estimated Using Complementary Data Sets. *Review of Economic Studies* 59(3):537–59
- Audretsch DB (2002) *Entrepreneurship: A Survey of the Literature*. Prepared for the European Commission, Enterprise Directorate General
- Banca d'Italia (2002) Supplements to the Statistical Bulletin. Methodological notes and statistical information. *Italian Household Budgets in 2000*
- Bernhardt I (1994) Competitive advantage in self-employment and paid work. *Canadian Journal of Economics* 27:273–289
- Blanchflower DG (2000) Self-employment in OECD countries. *Labour Economics* 7:471–505
- Blanchflower DG (2004a) Self-employment: more may not be better. NBER Working Paper 10286 (Forthcoming *Swedish Economic Policy Review*)
- Blanchflower DG, Oswald AJ (1998) What makes an entrepreneur. *Journal of Labour Economics* 16:26–60
- Böheim R, Taylor MP (2000) Unemployment Duration and Exit States in Britain. ISER working papers 2000–01, Institute for Social and Economic Research
- Borjas GJ, Bronars S (1989) Consumer discrimination and self-employment. *Journal of Political Economy* 97:581–605
- Bover O, Bentolila S, Arellano M (2002) The Distribution of Earnings in Spain During the 1980s: The Effects of Skill, Unemployment, and Union Power. In: Cohen D, Piketty T, Saint-Paul G (eds) *The Economics of Rising Inequalities*. CEPR, Oxford University Press, London
- Bover O (2004) The Spanish Survey of Household Finances (EFF): Description and Methods of the 2002 Wave. Occasional Paper n 0409, Banco de España
- Carrasco R (1999) Transitions to and from self-employment in Spain: an empirical analysis. *Oxford Bulletin of Economics and Statistics* 61(3):315–41
- Carrasco R, Ejrnæs M (2003) Self-employment in Denmark and Spain: institutions, economic conditions and gender differences. CAM Institute of Economics University of Copenhagen WP 2003–06

- Clark K, Drinkwater S (2000) Pushed out or pulled in? Self-employment among minorities in England and Wales. *Labour Economics* 7:603–628
- Congregado E, Golpe A, Millán JM (2003) Some Empirical Aspects of Self-Employment in Spain during the nineties. Seminario de Análisis Económico, Universidad de Huelva, SAE WP 20
- Congregado E, Golpe A, Millán JM (2005) Determinantes de la oferta de empresarios. In: García J, Pérez J (eds) *Cuestiones Clave de la Economía Española Perspectivas actuales 2004*. Ed. Comares, pp 165–87
- De Wit G (1993) Models of self-employment in a competitive market. *Journal of Economic Surveys* 7:367–97
- De Wit G, van Winden F (1989) An empirical analysis of self-employment in the Netherlands. *Small Business Economics* 1:263–72
- De Wit G, van Winden F (1990) An empirical analysis of self-employment in the Netherlands. *Economic Letters* 32:97–100
- De Wit G, van Winden F (1991) An M-sector, N-group behavioural model of self-employment. *Small Business Economics* 3:49–66
- Dolton PJ, Makepeace GH (1990) Self-employment among graduates. *Bulletin of Economic Research* 42:35–53
- Earle JS, Sakova Z (2000) Business start-ups or disguised unemployment? Evidence on the character of self-employment from transition economies. *Labour Economics* 7:575–601
- Evans DS, Jovanovic B (1989) An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy* 97:808–27
- Evans DS, Leighton LS (1989) Some empirical aspects of entrepreneurship. *American Economic Review* 79 (3):519–35
- Falter JM (2002) Self-Employment entry and duration in Switzerland. Working paper, Employment Center, University of Geneva
- Fujii ET, Hawley CB (1991) Empirical aspects of self-employment. *Economics Letters* 36: 323–29
- García I, Montuenga VM (2004) Education returns of wage earners and self-employed workers: Portugal vs. Spain. *Economics of Education Review* 24: 161–70
- García-Fontes W, Hopenhayn H (1996) Flexibilización y Volatilidad del Empleo. *Moneda y Crédito* 201:205–27
- García-Pérez JI (1997) Las Tasas de Salida del Empleo y el Desempleo en España (1978–1993). *Investigaciones Económicas* 21(1):29–53
- García-Pérez JI, Muñoz-Bullón F (2005) Are Temporary Help Agencies changing mobility patterns in the Spanish labour market? *Spanish Economic Review* 7(1):43–65
- Gil FJ, Martín MJ, Serrat A (1994) Movilidad en el mercado de trabajo en España: un análisis econométrico de duración con riesgos en competencia. *Investigaciones Económicas* 18(3): 517–37
- Gill AM (1988) Choice of employment status and the wages of employees and the self-employed: some further evidence. *Journal of Applied Econometrics* 3:229–34
- Grilo I, Thurik AR (2004) Determinants of Entrepreneurship in Europe. ERIM Report Series Research in Management ERS-2004-106-ORG
- Holtz-Eakin D, Joulfaian D, Rosen RH (1994) Entrepreneurial decisions and liquidity constraints. *Rand Journal of Economics* 25:334–47
- Hundley G (2001) Why women earn less than men in self-employment. *Journal of Labor Research* 22:817–29
- Jiménez-Martín S, Sánchez AR (2000) Incentivos y reglas de jubilación en España. *Cuadernos Económicos de ICE* 65
- Jiménez-Martín S, Boldrin M (2002) Evaluating Spanish Pension Expenditure Under Alternative Reform Scenario. Working Paper ref 652, Universitat Pompeu Fabra (Forthcoming in a Chicago University press volume)
- Parker SC (2002) Do banks ration credit to new enterprises? And should governments intervene? *Scottish Journal of Political Economy* 49:162–95

- Parker SC (2003a) Does tax evasion affect occupational choice? *Oxford Bulletin of Economics and Statistics* 65:379–94
- Parker SC (2004) *The Economics of Self-Employment and Entrepreneurship*. Cambridge University Press, Cambridge
- Parker SC, Belguitar Y, Barmby T (2005) Wage uncertainty and self-employed labour supply. *Economic Journal* 115:190–207
- Rees H, Shah A (1986) An empirical analysis of self-employment in the UK. *Journal of Applied Econometrics* 1:95–108
- Reize F (2004) *Leaving Unemployment for Self-Employment. An Empirical Study*. ZEW Economics Studies 25, Physica-Verlag, Mannheim
- Reynolds P, Bosma N, Autio E, Hunt S, De Bono N, Servais I, López García P, Chin N (2005) *Global Entrepreneurship Monitor: Data Collection Design and Implementation 1998–2003*. *Small Business Economics* 24(3):205–31
- Segarra A, Callejón M (2002) New firms' survival and market turbulence: new evidence from Spain. *Review of Industrial Organization* 20(1):1–14
- Segarra A (dir.) et al (2002) *La creación y la supervivencia de las empresas industriales*. Civitas
- Taylor MP (1996) Earnings, independence or unemployment: why become self-employed? *Oxford Bulletin of Economics and Statistics* 58:253–66