Informality and social protection: Preliminary results from pilot surveys in Bulgaria and Colombia*

Franco Peracchi, Valeria Perotti, Stefano Scarpetta, Gero Carletto, Carmen Pagés, Kinnon Scott

This version: January 18, 2007

Abstract

There is a wide agreement on the fact that a large informal economy leaves many individuals without social protection and reduces government's tax revenue and social security contributions. However, it remains an open question what really drives informality, namely whether workers are simply trapped out of the formal sector or, at least some of them, chose it because it offers better alternatives than a formal job. The policy implications are clearly different in the two cases.

In order to shed light on this important issue, we propose a household survey instrument to assess the links between informality and social protection. It can be implemented either through a stand-alone survey or by adding a specific module to an existing general survey such as the World Bank's Living Standards Measurement Study. After describing the main survey instrument, we present the results of two pilot surveys, carried out in Bulgaria and Colombia, to test the effectiveness of the questionnaire and improve its design.

* Very preliminary. The research on which this paper is based was conducted as part of a joint HDNSP, PREM, DEC project "Understanding the causes and consequences of informal and low quality employment". We thank Pierella Paci, Stefano Paternostro, Marika Santoro and Lucio Scandizzo for insightful discussions.

Contents

1 Introduction

The concept of informality has been studied extensively in the theoretical and empirical literature.[1] In this literature, workers are classified as "informal" if they work in a small business, do not have a written contract or lack affiliation to the mandatory social security system or work in firms who do not comply with regulations. The definition of informality has also evolved, as also demonstrated by the subsequent proposals made by the International Conferences of Labor Statisticians (1970, 1993, 2003). The current proposed definition by the International Labour Office (ILO) shifted the focus from the characteristics of the activity where the worker is engaged toward the characteristics of her job. In particular, "all remunerative work Ű both self-employment and wage employment - that is not recognized, regulated or protected by existing legal or regulatory frameworks and non remunerative work undertaken in an income-producing enterprise (Hussmanns, 2005).

Despite the still ongoing debate on the definition of informality, there is a wide agreement on the fact that a high level of informal activity may leave many workers un-protected from different risks – old age, sickness, unemployment – and reduce government's tax revenues. Further, as stressed by Schneider and Enste (2000), official indicators of economic activity tend to be unreliable in the presence of a large informal sector, causing difficulties for policy makers in setting macro policies.

An important open issue is whether informal work is the only option for people who cannot find a job in the formal sector or is instead chosen – at least by some – because it is considered a better alternative to a formal job. The policy implications are clearly different in the two cases.

These two views of informality are supported by two alternative strands in the theoretical literature. The first regards the existence of an informal sector as the result of disequilibrium in the formal sector. In this class of models, the introduction of wage rigidities, such as a minimum wage above the equilibrium wage, gives rise to an informal sector of small firms where there are no wage rigidities, for example because minimum wage rules cannot be enforced, and unlimited employment opportunities for those who cannot find a job in the formal sector (Rauch, 1991; Fortin *et al.*, 1997). This framework provides a simple explanation for the common observation that small firms tend to be informal.

The second strand of the literature assumes that those who work in the informal sector actually choose to do so. Models of this type are often variants of the original model of self-selection of Roy (1951) but tend to lack a comprehensive definition of informality. Some of them focus on the choice

[1] See, for example, Fields (2005) for a recent review of the theoretical labor market models, and Schneider (2004) for a review of the empirical estimates of the size of the informal sector

of evading taxes but do not consider employment status, while others explain the determinants of entrepreneurship and firm size but do not include tax evasion among the determinants.

Models focusing on tax evasion extend the seminal papers by Allingham and Sandmo (1972) and Srinivasan (1973). The main idea is that tax evasion is risky because there is a positive probability of being detected by the fiscal authorities. The basic model has been extended to include labor supply choices (Baldry, 1979; Pencavel, 1979) and repeated games with the tax authority (Reinganum and Wilde, 1985). One problem with this type of models is that they imply very low compliance rates. To overcome this issue, some models introduce moral costs of evasion (Benjamini and Maital, 1985) or the provision of a public good by the government (Cowell and Gordon, 1988). Tax evasion models are usually static and they do not account for future benefits of compliance, such as old age pension benefits.

As the self-employed are entrepreneurs of small businesses, models of entrepreneurship and firm size are useful if the informal sector is identified with self-employment. Two theoretical models of entrepreneurship are Lucas (1978) and Kihlstrom and Laffont (1979). In the former, entrepreneurial ability varies across workers who are instead identical in terms of labor skills. In the latter, individuals do not differ in managerial or labor ability and the choice to run a business is driven only by attitudes toward risk. Possible obstacles to self-employment are liquidity constraints (Evans and Jovanovic, 1989) and discrimination (Borjas and Bronars, 1989).

In order to shed light on these issues, we propose a household survey instrument to assess the linkages between informality and social protection. The instrument can be implemented either through a stand-alone survey, or by adding a specific module to an existing general survey such as the World Bank's *Living Standards Measurement Study*. The main focus of the proposed survey is to understand the determinants of the joint decision to 1) work in the formal economy; and ii) participate to the social protection system.

A household survey is particularly suited for our purposes because it provides extensive information not only on the individual, but also on other members of her household. This is especially important, as the decision to work in the informal sector is likely to depend on household characteristics, especially the number, and the employment status, of the other members.

After describing the main survey instrument, we present the results of two pilot surveys, carried out in Bulgaria and Colombia, to test the effectiveness of the questionnaire and improve its design.

The remainder of the paper is organized as follows. Section 2 discusses the design of our basic questionnaire on the informal sector. In practice, this basic version has to be adapted to take into

account country-specific institutional settings. We present two adaptations of the questionnaire for Bulgaria and Colombia. Section 3 describes in detail the pilot surveys carried out in these two countries. Section 4 asks how representative are the samples of the two pilots. Section 5 presents the main empirical results of the two pilots. Finally, Section 7 offers our concluding remarks.

2 The basic questionnaire

A survey on informality should pay particular attention to a number of issues that are likely to play a key role in the choice of participating in the labor market, engaging in a self-employment activity or seek a dependent employment, and seeking a job with affiliation to social security. In particular, attitudes toward risk and liquidity constraints are important in self-employment choices, while opinions about the tax and social security system are important to understand whether workers seek formal jobs and whether firms are able/willing to evade tax and/or not report their workers. Impatience and life expectancy are other examples of relevant variables.

Our survey questionnaire focuses on three main sets of variables: (i) employment status (current and past, including subjective information like job satisfaction) and earnings, (ii) social protection (level of information about the system rules, current contributions, and personal attitudes toward the system), and (iii) subjective beliefs about uncertain events, individual attitudes toward risk, and intertemporal substitution.

The general design of the questionnaire follows the World Bank's Living Standards Measurement Study (LSMS). Some questions closely resemble those in existing surveys, such as the the European Community Household Panel (ECHP), the Survey of Health, Ageing and Retirement in Europe (SHARE), the English Longitudinal Study of Ageing (ELSA), the Health and Retirement Study (HRS), the Bank of Italy Survey of Italian Households' Income and Wealth (SHIW), and the Chilean Encuesta de Previsión de Riesgos Sociales (PRIESO). The last part of the questionnaire also takes into account the survey of attitudes toward the welfare state carried out by Boeri, Boersch-Supan and Tabellini (2001).

The basic questionnaire is in English and is organized in five chapters:

1. Demographics and main household characteristics,
2. Employment status and job characteristics,
3. Social protection,

4. Time preferences, subjective beliefs and risk attitudes,

5. Attitudes towards Social Security.

The household head answers Chapter 1, whereas Chapters 2–5 are answered by all household members aged 15+. The full survey includes all five chapters. The information in Chapter 1 and most of the information in Chapter 2 are usually asked in standard household surveys. Thus, only Chapters 3–5 are relevant if the survey is going to be an add-on to an existing survey.

The rest of this section describes the basic questionnaire in detail, whereas Section 3.4 describes its adaptation to the two countries, Bulgaria and Colombia, where pilot surveys have been carried out.

2.1 Demographics and main household characteristics

This part of the questionnaire focuses on key individual characteristics of the members of the household, including gender, age and schooling level (last schooling level attended and last schooling level completed). It also collects information at the household level from the household head. Basic demographic information is available for all household members.

The chapter begins with a membership roster similar to that contained in the questionnaire of the World Bank's LSMS. It continues by asking information on housing, durables and financial wealth, together with the household's total income and savings. Depending on the country, monetary amounts may be asked either net or gross (in some countries people are more familiar with gross income, while in others they are used to refer to net amounts). This part of the questionnaire provides short, but complete background information about the household's living conditions.

2.2 Employment status and job characteristics

This chapter of the questionnaire is divided into three parts: the first (Q1–Q36) is about the jobs (main and secondary) currently held by the respondent, the second is devoted to respondents without a job (Q37-Q41), and the last concerns past employment history (Q42–Q43). The basic version does not specify whether proxy respondents are allowed to answer in place of household members that are temporarily away.

Since respondents can hold more than one job, we begin by asking separate questions about the main job, which is defined as the one demanding more hours among all jobs held. If the individual has secondary jobs, we ask the same questions about the most important of them. Secondary jobs are an important issue in the study of informality. A person could have a formal main job

that offers social security coverage, in fact, and a secondary job that is totally uncovered, but offers higher earnings (or more satisfaction, etc.). The secondary job could be completely different from the formal one in terms of professional content, contractual conditions, time span and career perspectives. From now on, by "second job" we mean the most important secondary job.

Information is collected on both current monthly earnings (earnings in the month before the survey) and annual earnings in the last calendar year. Whether earnings information is asked on a gross or net basis depends on the specificity of a country. Given information on usual weekly hours of work, a measure of current hourly earnings may be constructed by dividing current monthly earnings by monthly hours of work (usual weekly hours times 4.2).

The same set of questions is asked for the main job (questions Q9–Q24) and for the secondary job (questions Q27–Q36). Many more details are asked about the main job. The reference period is the last week as in the World Bank's LSMS. In Q9 (and similarly in Q27 for the second job), employment status is categorized into private employee (distinguishing between paid with regular installments or piecework), civil servant, apprentice, self-employed and unpaid worker in family business. Singling out the latter category is particularly important for assessing informality in developing countries. Q10–Q15 are about job characteristics, namely occupation (Q10), main activity of the business or organization (Q11), number of workers in the business or organization (Q15b), usual weekly hours (Q12), weekly hours last week (Q13), earnings (Q14) and workplace (Q15). Q28–Q33 ask the same questions for the second job.

Job satisfaction can play a role in the decision of holding more than one job, being an informal worker or seeking a new job. Thus, we ask a number of questions concerning the level of job satisfaction of workers, including on the number of hours worked, social security benefits, skills matching, earnings, flexibility, work environment and fringe benefits (Q16), and the overall level of satisfaction with the current main job (Q17). Similar questions are asked for the second job (Q35 and Q36). Employees are asked an additional question (Q19) about the type of work contract (permanent, short-term or no contract, main job only). The following question (Q20) is about the motivations for working as an employee. This question is asked only for the main job. A similar question is asked to the self-employed (Q21).

Because an important feature of informality is the lack of social security coverage, a large part of the questionnaire is devoted to gather information on participation, coverage, and attitudes towards the social protection system. Questions about affiliation are asked in the chapter on employment status, information on the system and benefits is asked in the chapter presented in

Section 2.3, whereas attitudes towards the social protection system are asked in the last chapter of the questionnaire (see Section 2.5).

For each job, the questionnaire asks about the type of social insurance program the worker contributes to (old age, survivors, disability, sickness and maternity, work injury and unemployment insurance).[2] Separate questions are asked for the main job (Q22) and the second job (Q34). Questions Q23 and Q24 ask the percentage of the main job's salary/earning currently paid as contributions for the public old age insurance. Asking about contributions in both jobs is important because the second job may provide supplementary or complementary insurance to the main job. We favor a line of questioning focusing on contributions rather than on entitlements, because the worker may be paying contributions but may not be entitled to any pension because of eligibility conditions. An important feature of our questionnaire is the separate analysis of the various programs included in the "social security package". The fact that they are usually "bundled" could be a reason for workers to be informal, if the willingness to pay for one program is more than offset by the lack of interest in others, or if the overall costs of affiliation is too high but workers would be willing to contribute to only some of the components of the social security system.[3]

If contributions to an insurance program are paid, there is a possibility that other household members could be covered by that insurance. The questionnaire does not address this issue, which may be relevant for some countries, depending on the institutional setting. In addition, asking about health insurance provided by the social security system might be useful in countries where contributing to the program entitles the workers to different and/or additional benefits with respect to the basic medical services.

Questions Q37–Q41 are directed to respondents who are not employed and are looking for a job. They provide information on the type of job that the unemployed are looking for. Questions Q37 and Q38 ask about search activities during the last four weeks and about availability to work within the next two weeks. They are therefore useful to check if respondents meet the ILO criteria for unemployment. We then ask if they would prefer a job in self- or paid employment (Q39) and if they would be ready to accept a job without social security coverage (Q40). Motivations for those who would accept an unprotected job are important to investigate whether informality is a voluntary choice or a forced decision (Q41). Finally, question Q42 offers a very short summary of employment

[2] It asks questions about mandatory contributions through the employer, voluntary contribution through a private provider, both mandatory and voluntary or none.

[3] For example, if a worker wants to have an old age insurance but is not concerned with the possibility of becoming disabled, he may not want to pay contributions for both programs. As a consequence, he may end up contributing to none of them.

status in the past calendar year, on a monthly basis. The choice of the calendar year allows us to have a common reference period for all interviews, also when field work lasts for a long period. The use of monthly employment status for the past year gives information on seasonal variations in the type of job. Employment status is categorized as private employee (distinguishing between paid with regular installments or piecework), civil servant, apprentice, self-employed, unpaid worker in family business, unemployed, retired or "other". Overall earnings in past calendar year are asked in question Q43.

2.3 Social protection

This chapter of the questionnaire aims at understanding the level of information individuals have about the public social protection system – both in terms of eligibility conditions and with respect to the way benefits are calculated. If the employment status is the result of choice, then it is crucial to know what information individuals have about the variables that enter the choice problem.

The chapter begins with a question about the total number of years of contributions to the public social security system (Q44). It then asks if the respondent is currently receiving income from public insurance programs, distinguishing between old age (Q45), survivors (Q48) and disability (Q51). For each program, if the respondent is receiving a pension, she is asked the monthly amount (Q46, Q49, Q52) and the year when she started receiving the benefit (Q47, Q50, Q53). Question Q54 asks if the person is receiving income from a private pension. If this is the case, the monthly amount (Q55) and the first year (Q56) are asked. The subsequent questions ask whether the respondent is receiving income from sickness and maternity (Q57) or public unemployment insurance (Q59). In case of positive answer, the monthly amount is asked (Q58 and Q60).

The level of knowledge of the social security system is assessed through a group of questions asked to all employed individuals. The first question (Q62) asks the respondent what she believes the minimum age to be in order to receive the public old age insurance. To restrict the possible answers, the respondent is asked the age requirements for an individual who contributed for 30 years. The next question (Q63) refers to the same hypothetical situation and asks what the pension benefit would be for that person, retiring at 65 if a man and at 60 if a woman. All amounts and ages in this part of the questionnaire need to be adapted to the local institutional settings. Question Q63 asks about the expected replacement rate, namely the expected pension as a percentage of last income. In question Q64, it is asked by how much the pension would increase if the individual were to work for additional 5 years. In question Q65, the eligibility conditions for public disability insurance are

asked. If the respondent is an employee, she is then asked the eligibility conditions for sickness and maternity benefits (Q67 and Q68), for work injury benefits (Q71) and for unemployment benefits (Q74). For unemployment insurance, an additional question is asked about the amount of the benefit (Q75). For each of these three insurance programs, the respondent is asked if she is currently receiving the benefit or has received it in the last five years (Q69 and Q70, Q72 and Q73, Q76 and Q77).

2.4 Subjective beliefs, time preferences and risk attitudes

Individual beliefs about uncertain events, attitudes towards risk and impatience are important in understanding decisions related to social security coverage. Learning about individual risk attitudes and impatience, and the role they play, is crucial to understanding informal work as a possible object of choice.

Before asking about life expectancy, we ask respondents to self-report their health status. We only ask two questions (Q78 and Q79) because the proposed survey is not focusing on health, but rather on understanding whether personal health conditions may be related to affiliation and attitudes towards the social security system.

We try to uncover subjective beliefs about uncertain events by asking what probabilities are attached to them. There is strong evidence that this approach provides more reliable answers than traditional expectation questions (Hurd and McGerry, 2002; Manski, 2004). To familiarize respondents with this approach, the interviewer is expected to begin with the following example:

> *We have some questions about how likely you think various events might be. When I ask a question I'd like you to give me a number from 0 (absolutely no chance) to 100 (absolutely certain). Let's try an example together and start with the weather. What do you think the chances are that it will be sunny tomorrow? You can say any number from 0 to 100. For example, "90" would mean a 90 percent chance of sunny weather (i.e. sunny weather is very likely).*

The subjective probability of reaching a certain age is elicited by repeatedly questioning the subject for different ages, that can be changed according to the country's demographics (Q80). Question Q81 asks whether the answer is based on information by the media, on family history or on medical records. Questions Q82–Q91 inquire about job and income expectations: standards of living in the next five years (Q82), chances of loss of property, serious illness or accident, physical

incapacity or death within the next 12 months (Q83), chances of receiving/leaving an inheritance (Q84), probability of losing the job for the employees (Q86 and Q87) and of closure of the business/bankruptcy/substantial loss of clients for the self-employed (Q88), beliefs about increase in own income and prices in the next five years (Q89–Q91).

Questions Q92–Q105 are more specifically related to the post-retirement period. For our purposes, knowing the sources of expected income is very important, as family support or financial assets can substitute the social security program in providing insurance. Question Q92 asks what is the expected post-retirement pension as a percentage of current labor earnings. Questions Q93–Q104 ask whether the individual is expecting to receive (after retirement) income from pensions of other household members, work, properties, financial assets, friends and relatives or other sources. If some income is expected, the yearly amount is asked. Question Q105 is about the expected change in living standards after retirement.

Questions Q106–Q108 elicit intertemporal preferences and give a rough measure of an individual's discount rate. The method used is the following. Respondents are asked to imagine that they won the prize of a lottery which is worth 1000 in local currency.[4] The prize is certain but can be paid only one year later. There is however the possibility of selling the ticket for 920. If they accept 920, they are not asked other questions on this topic. If they prefer 1000 one year later, they are offered 950. If they still prefer 1000 one year later, they are finally offered 980. From answers to these questions, we can define brackets for the subjective discount rate. Following Harrison *et al.* (2002), the value of the ticket can be defined as $T = x(1 + r)$, where r is a rate of discount. If the respondent prefers to have T after one year, rather than x immediately, her discount rate is smaller than r. By repeating the question for different values of x we can identify four discount levels. The set of questions could be repeated with a different time horizon to capture variations of the discount rate over different time spans. As an alternative approach, it would be possible to ask directly the amount x for which respondents would be indifferent between selling and holding the ticket. These questions are discussed in more detail in Section 5.4, where we present results from the pilot surveys.

Questions Q109–Q111 have a similar structure, dealing with risk attitudes. In this case we assume that the prize won is to be paid immediately. However, the lottery organization is not entirely reliable. There is only a 70 percent probability of being paid. We now define x as the certain amount that is offered to individuals in exchange for the lottery ticket. The random gain

[4] The actual amount would be country specific.

from participation to the lottery, P, is $1000-x$ with 70 percent probability and $-x$ with 30 percent probability. To construct the Arrow-Pratt measure of absolute risk aversion (ARA), we follow Guiso and Paiella (2001). If the respondent is indifferent between receiving x and holding the ticket, her absolute risk aversion is:

$$ARA = 2\frac{E(P)}{E(P^2)} \tag{1}$$

where

$$E(P) = 0.7 \cdot (1000 - x) + 0.3 \cdot (-x), \qquad E(P^2) = 0.7 \cdot (1000 - x)^2 + 0.3 \cdot x^2.$$

By repeating the questions with different levels of x, we are able to identify four brackets for the risk aversion measure. In particular, the first question offers $x = 700$, which is the expected value of the lottery. If the respondent is willing to accept, we then ask if she would accept $x = 600$. In the opposite case, we ask if she would accept $x = 800$. A possible extension is to repeat the questions with different prize values, to see whether risk attitudes change with wealth. An alternative approach would be to ask directly the amount x for which the individual is indifferent between selling the ticket and participating to the lottery. These questions are discussed in more detail in Section 5.4, where we present results from the pilot surveys.

2.5 Attitudes towards social security

This last chapter of the questionnaire asks questions about the worker's personal views on the role of the social protection system, expectations about the future of the pension system, and willingness to pay for public protection programs. Workers who are not affiliated, in fact, may be willing to contribute (perhaps after some rules are changed), while affiliates may have been simply forced to join. Our aim is to understand the effects of personal views and attitudes on strategic behavior in life cycle planning and risk management.

We start with overall questions about the perceived role of the government in providing social security. The first question (Q112) asks the respondent whether the government, or the employer or the worker herself should be responsible for providing social security benefits. Question Q113 is about preferences between a pay-as-you-go (PAYG) and a fully-funded pension system. The next question, Q114, asks what the government should do to maintain the balance between contribution and benefits. Questions Q115 and Q116 are about personal views about social security contributions, i.e. whether the respondent sees them as a way to finance her own social security benefits,

as the duty of a good citizen, as a characteristic of stable jobs with better working conditions, as a way to redistribute wealth or as an investment for the future, etc.

Subjective probabilities of events related to the social security system are asked in questions Q117–Q124. In particular, the respondent is asked about the probability of change in: the type of social security system; contributions and benefits; the eligibility conditions; and the coverage. For contributions and benefits, respondents are asked the subjective probability of an increase or a decrease. For benefits, the probabilities of increase or decrease are asked separately for old age, disability and survivors.

In Q126–Q137 and Q140–Q142, individuals currently affiliated to the social security system are asked whether they would like to change the level of contributions, taking into account the associated changes in benefits. In particular, they are asked if they would like to receive a higher retirement pension even if this would imply higher contributions (Q126). Interviewers should stress that the change in contributions would affect the entire working life. Respondents are then asked whether they would like to reduce contributions even if this would imply a lower pension (Q127). People willing to pay lower contributions are also asked what is their preferred percent reduction in contributions (Q129). They are also asked if they would like to stop contributing (Q128). Question Q130 asks about motivation for low willingness to pay. Questions Q131–Q135 repeat the same questions with respect to workplace accident insurance, while Q140–Q145 are about willingness to pay for the unemployment insurance.

The aim of Q137–Q138 and Q146–Q148 is to explore the willingness of non-contributing workers to become affiliated to the public system. Question Q137 and Q138 ask whether the person has ever paid social security contributions and reasons for not paying. Questions Q146–Q148 ask if she would have accepted to reduce her earnings in order to obtain an unemployment insurance. Different percent reductions in earnings are proposed in the three questions.

Finally, question Q149 tries to elicit preferences between different insurance programs in trade-off situations (i.e. with a budget constraint). This question is important to understand if, for example, "unbundling" the social security package would increase affiliation. Respondents are asked for what public insurance program they would like to increase the benefit, provided that this would imply a reduction in benefits for other programs.

3 The Bulgarian and Colombian pilots

Two pilot surveys were carried out to test the effectiveness of the basic questionnaire, to improve the questionnaire's design, and to identify problem areas and peculiarities of the fieldwork.

3.1 Choice of countries

The two pilot surveys were carried out in Bulgaria and Colombia, during the months of May and June 2006, with a target sample size of 200 households in Bulgaria and 100 households in Colombia.

Appendix B compares a set of indicators for the two countries. In 2004, the level of per capita GDP in terms of purchasing power parity was a little higher in Bulgaria than in Colombia (7840 and 6975 US$ respectively). However, the two economies show marked differences in performance and structural changes in the years preceding the pilot surveys. Bulgaria is a transition economy. Over the 1990s it experienced a systemic change to a market-based economy, with a major drop in GDP/capita in the early phases of the transition and and subsequent modest recovery. Only in the most recent year did the Bulgarian economy recovered fairly rapidly (GDP grew at 4.8 percent in the 2000-2004). Colombia undertook major reforms in the past decade but was also affected by the regional and domestic crises.

Despite these substantial differences, the structure of the two economies is fairly similar: agriculture accounts for around 10% of GDP, industry for around 30% and services for the remaining 60%. Unemployment is around 14% in both countries, but an important difference is the labor force participation rate among people aged 15–64, which is higher in Colombia for both men and women.

The fraction of female employment in industry is higher in Bulgaria than in Colombia, while the opposite is true for employment in services. The fraction of women in agriculture is exactly the same. For men, the fraction of agricultural employment is much higher in Colombia than in Bulgaria, while the opposite is true for employment in industry and roughly the same fraction is employed in services.

The size of the shadow economy as a share of GDP was found to be similar in the two countries in a study by Schneider (2005). The paper reports the share to be increasing from 1990 to 2000 for both countries, from 29.4 to 36.9 percent in Bulgaria and from 33.4 to 39.1 percent in Colombia. In his definition of shadow economy, Schneider includes all activities that are concealed to the authorities in order to avoid compliance with taxes, social security contributions and other government regulations. Another useful indicator of the level of informality is the share of self-employment,

which is much higher in Colombia than in Bulgaria (50.9 versus 14.9 percent). As far as social security, 64.0 percent of labor force is a pension contributor in Bulgaria, while the fraction is only 20.7 percent in Colombia. The tax revenue collected by the government as a fraction of GDP is much higher in Bulgaria than in Colombia (22.3 versus 13.8 percent), whereas the highest marginal tax rate is lower in Bulgaria both at the individual and the corporate level (29.0 and 20.0 percent versus 35.0 and 37.0 percent).

Appendix C compares the key features of the social security systems of the two countries. In Bulgaria, coverage is mandatory for all employed individuals and the system is based on both social insurance and individual account. In Colombia, coverage is mandatory for all workers but there is a choice between social insurance and individual account. The Bulgarian system requires a higher level of contributions towards the old age pension (roughly 25 versus 15 percent of earnings). Further, eligibility conditions are stricter in Bulgaria than in Colombia, both in terms of minimum age (63 versus 60 years for men, 58.5 versus 55 years for women) and in terms of years of contributions (37.0 for men and 33.5 for women versus 20.2 for both men and women in Colombia). The social insurance old age pension benefit for an individual with minimum age and contribution requirements is lower in Bulgaria than in Colombia (roughly 35 percent versus 80 percent). There are less differences with respect to sickness, maternity and work injury benefits. Unemployment benefits in Bulgaria are 60 percent of earnings and are paid for up to 12 months, according to the coverage period. In Colombia, the benefit is equal to 1 monthly wage for each year of employment.

Despite the differences in the main socio-economic indicators and in the characteristics of the fiscal and social security systems, both countries present a large informal sector. The differences in the fiscal and social security systems of the two countries are particularly useful for our purpose, because they can be exploited to identify the effects that the social protection system may have on the decision to participate to the informal sector.

3.2 Prior empirical evidence

Previous studies about informality in Bulgaria and Colombia provide a useful background for our analysis, even if they were largely constrained by the available data. Hassan and Peters (1995) use data from the 1992 Individual Budget of Households survey to discuss the effectiveness of the Bulgarian social safety net. By analyzing the distribution of social security benefits among low, middle and high income groups, they conclude that while both the rich and the poor benefit from the pension scheme, social benefits such as unemployment benefits and child allowances often accrue

to middle- and high-income households.

From a sociological perspective, Chavdarova (2003) analyzes the variety of informal activities in Bulgaria during the 1990s. According to Chavdarova, working in the informal sector has different motivations depending on the number of jobs held. Those who have only one (informal) job are likely to be underpaid and forced by subsistence needs, whereas moonlighters hold a second informal job to avoid regulations. However, there is no empirical evidence to confirm this hypothesis.

Marc and Kudatgobilik (2003) analyze informality in several South-East European countries. The paper is based on poverty assessments by World Bank staff and suggests that in these countries informality is a survival mechanism whose existence is due to the failure of the formal sector. The transition to a market economy led to a loss of public sector jobs and state-guaranteed job security, increasing acceptance of informal and even illegal activities. The authors also stress that the Roma minority is particularly involved in the informal sector, both because of poverty and cultural reasons.

As for Colombia, Magnac (1991) identifies the informal sector with self-employment (excluding employers) and estimates a probit model for participation to this sector using data from a 1980 household survey. Higher educational attainments decrease the probability of holding an informal job, whereas the effect of experience has an inverted U-shape. In the estimated log-wage regression, the coefficient of the informal sector dummy is positive and significant, even after controlling for education and experience. Another result is the larger variance of wages in the informal sector.

An empirical analysis of the determinants of self-employment in Colombia is provided by Destré and Henrard (2004), using data from the 1996 Colombian National Household Survey.[5] The estimates of selectivity-corrected earnings regressions show that selection into self-employment has a negative effect on earnings. Therefore, the authors suggest that self-employment in Colombia might be the result of a forced decision driven by subsistence needs. On the other hand, in a probit model for self-employment, a larger potential earnings differential increases the probability of being self-employed, after controlling for standard socio-demographic characteristics such as age, education, marital status and regional dummies. As in Magnac (1991), the probability of being self-employed decreases with higher educational attainments and increases with age.

The choice between different pension programs is studied by Kleinjans (2003) with specific reference to the Colombian system, where workers can choose between the social insurance and individual account system. Data are from the 1997 Colombian National Survey About Life Quality.

[5] The agricultural sector is not covered by the survey, and the sample is composed by male household heads only.

The paper adopts a random utility approach, where agents choose between a public and a private pension plan by maximizing the expected random utility. The deterministic term of utility is a quadratic function of the rate of return in the chosen pension program. Taking into account the default risk for the public pension and the market risk for the private pension, the parameters of the utility function are estimated by maximum likelihood. Policy simulations suggest that a decrease in the default risk would allow the government to reduce public pension benefits, without causing a large number of workers to switch to the private system.

3.3 Survey organization

The survey were carried out, on behalf of the World Bank, by a consulting company affiliated to Gallup in Bulgaria, and by a team headed by a researcher (Carlos Becerra) in Colombia.

Each country team was responsible for choosing the sampling frame and the sampling design, translating and adapting the survey instrument, selecting and training the interviewers, fielding the survey, preliminary clean-up of the data, and delivery of the final data files.

In Bulgaria, 200 household interviews were carried out in 4 geographical areas: Sofia city (49), Karzhali (51), Varna (50) and Razgrad (50). Households were purposively selected to make sure that the questionnaire was administered to households with different composition and characteristics. In particular, the aim was to represent the most widespread types of employment status. Given the focus on social security, another aim of the sample selection criteria was to include members of the Roma population and households with disabled individuals. In addition, since most of the questions in the survey are directed to employed individuals, households with members of working age were over-represented. The fieldwork team consisted of 1 supervisor and 2 enumerators in each of the 4 areas. All of them were trained and briefed for the survey's objectives on a training session held in Sofia in the period 12-14 May.

In Colombia, interviews were carried out in the capital city of Bogotà and the sample was stratified by income and other living standard into 6 strata, following the "estratificaciòn de Barrios e Manzanas" for the city of Bogotà. Starting from the lowest stratum, the actual proportions in the sample are, respectively 15, 20, 25, 29, 10, and 1 percent.

3.4 Adaptation of the basic questionnaire

The basic questionnaire was adapted to take into account country-specific characteristics, feedback from the survey agencies, and the results of pre-tests aimed at checking the viability of certain questions, especially those having to do with subjective beliefs and attitudes towards the Social

Security system. As a result, the specific formulation of some questions differ. While this provides more evidence on how different formulations of the same question worked in the field, it also limits the comparability of the results for the two countries. In what follows, the questions in the Bulgarian questionnaire are denoted by B*, while those in the Colombian questionnaire are denoted by C*.

In the Bulgarian survey, household head status was self-reported by household members. In most of the cases, the household head was the male with highest earnings in the household.

When a household member was not available for personal interview, a proxy respondent was allowed to answer the majority of the questions in Chapter 2 and Chapter 3. The last two chapters ask about subjective beliefs and opinions, therefore proxy respondents were not asked these questions. The main job was defined as the one demanding more hours. For Colombia, we do not have information on the definition of household head and on the rule for proxy respondents.

There are minor differences between the two pilots in the questionnaire design. For example, different age groups were used as sample selection criteria for different parts of the questionnaire. In the Bulgarian survey, the reference age group was 15+ except for part of the questions in Chapter 4, that were asked only to people aged 40+. In the Colombian survey, the reference group was 15+ except for Chapter 3 and Chapter 4, that were asked to people aged 30+.

Other relatively small differences have to do with the set of questions asked and their precise formulation. In particular, no information on health status was collected in Bulgaria. Earnings are net of taxes and social security contributions in Bulgaria but are gross amounts in Columbia. The questions about life expectancy and retirement age were asked differently in the two surveys. In Colombia people 30+ were asked "till what age do you expect to live" (C83), whereas in Bulgaria people 15+ were asked "what are the chances (any number from 0 to 100, where 0 means unlikely and 100 means certain) that you will live to be age [70|80|90]?" (B89a–c). The formulation of the questions on retirement age (B91a–c and C95) is similar to that of the question on life expectancy. For the subset of questions about employment risks and health risks, property losses, and giving and receiving inheritances, the Colombian sample was split up in roughly two halves and two alternative formulations were adopted: one based on scales ("muy probable", "probable", "poco probable", "muy poco probable") and the other on percentages ("un numero entre 1 *y 100"!*). In Bulgaria, the question on willingness to accept a job without social security coverage was only asked to people looking for a job. In Colombia, instead, it was asked to everybody.

An important piece of information missing in the two surveys is the age when the person first started working.

3.5 Fieldwork

Both pilots were conducted in the Spring of 2006. In Bulgaria, 23 interviews were carried out between May 15 and 19 to pre-test the questionnaire. The pilot survey was carried out between May 27 and June 18. In Colombia, the pilot survey was carried out between May 22 and June 3, and high-income households were interviewed during the weekends.

Interview duration ranges between 30 and 180 minutes and, on average, is equal to little less than 90 minutes. Average interview duration is slightly longer in Bulgaria (84 minutes) than in Colombia (79 minutes), and the minimum and maximum durations are also slightly longer in Bulgaria (30 and 191 minutes respectively) than in Colombia (20 and 185 minutes respectively). In both countries, interview duration varies with the number of household members eligible for interview (aged 15+) and tends to be higher for households where income and educational attainments are low. For these households, completing the questionnaire was often not easy because of the lack of information on many topics and the difficulties in understanding some of the questions. Interviewers spent a substantial amount of their time explaining survey questions in these cases.

As far as interviewers are concerned, it would be useful to have information on characteristics such as gender, age, education and past experience. Unfortunately, the available data does not contain this piece of information.

3.6 Sample size

The initial number of records in the Bulgarian file is 512, from 200 households. After dropping people aged less than 15 (59 observations dropped) and people aged 15 and older who stopped answering at the question on employment status (29 observations dropped), we are left with a sample of 428 individuals of working age from 200 households. Among these, 125 (29.2 percent) were not available for personal interview and proxy respondents answered for them. As a consequence, for questions on subjective beliefs and personal opinions (Chapter 4 and Chapter 5 of the questionnaire), the sample size is 303. The fraction of people not available for personal interview is particularly high for age group 15–19 (80 percent).

The initial number of records in the Colombian file is 280, from 102 households. Of these, however, 26 records contain no suitable information. After dropping people aged less than 15 (69 observations dropped) and people for whom only basic demographic information is available (8 observations dropped), we are left with a sample of 177 individuals of working age from 79 households.

Table 1 shows the distribution of the working age population by age group and gender in the two samples. Notice that women always represent more than half of the sample, but their percentage is lower in the Bulgarian sample (54.4 percent) than in the Colombian sample (57.6 percent).

3.7 General problems with the survey

This section briefly lists some general problems encountered during the fieldwork in the two countries.

Although interviewers report that most survey questions have generally been well understood in Bulgaria, even by respondents with low educational level, a number of problems have to be kept in mind when analyzing the data from the Bulgarian pilot.

First, many respondents have a low level of awareness about the social security system and its specific details. Apparently, many respondents became aware of these details (like their own social security status, the criteria for receiving old age pension, or maternity/work injury/sickness insurances, etc.) during the interview. This group of respondents consist largely of low educated people (including the Roma population), teen-agers and adults without working experience. Some of them give tentative answers based on the knowledge or the estimates generated at the time of the interview, others just answer "Don't know" (DK) or provide no answer.

Second, some people are unwilling to answer questions concerning income from first/second job, total annual income and working time. This is particularly true for respondents employed in the informal sector, who tend to conceal part of their real income. Due to the uncertainty (job loss or control by the tax/social security administration) of the informal employment, they prefer to hide their income or even not to share their full employment status (second, third jobs).

Third, interviewers face strong social-psychological barriers when eliciting expectations about life expectations and other risky events (sickness, accidents). For many respondents these are "taboo" topics, and answering them is interpreted as "challenging the fate". This group of respondents consists largely of people who are low educated, strongly religious, elderly, sick or disabled.

Fourth, the design of the questionnaire ignores a few marginal categories who face difficulties to identify themselves in the indicators and the logic of the questionnaire. These categories include some forms of self-employed (for example seasonal/agricultural workers, people employed in tobacco growing), working pensioners and working students, people employed in family business, and people registered as unemployed but working occasionally.

Fifth, employment status is difficult to define for some types of respondents, such as people

employed in family business and caretakers for other persons.

Many of these problems also arise with the Colombian pilot. Specific problems include the quality of the Spanish translation of the basic questionnaire and the fact that the target for each question was not clear from the format of the paper questionnaire. For example, in Chapter 5, question C131 (about past contributions) is asked to people aged 15+, while the follow-up question on the reasons for not contributing is asked to people aged 30+ only. As a consequence, when answers are missing, it is sometimes difficult to distinguish between nonresponse and non-applicability.

4 How representative are the two samples?

This section analyzes how representative the two samples are. The analysis is restricted to the sample of people of working age, namely those aged 15 and older (428 observations from Bulgaria and 177 observations from Colombia). We report results from the standard questions on basic characteristics (age, gender and educational attainments), labor force status, employment status, and earnings. Results are discussed after a short description of nonresponse rates.

The main conclusion is that, although the distribution of the two samples by gender, age and education is somewhat different from the corresponding population distribution, the basic statistical relationships between labor market outcomes (activity rates, employment rates, unemployment rates and earnings) and personal characteristics are remarkably close to known results from other samples or other countries. This suggests that our samples may be used not only to check the quality of the field operations and the questionnaire, but also to draw tentative conclusions about at least some of the substantive relationships that our sample is designed to help analyze.

4.1 Distribution by age

Table 2 shows the age structure of the two samples by gender, and compares it with the population age structure provided by the U.N. Population Division for 2005.

Both samples over-represent people in the central age group (aged 30–54) and under-represent the oldest people (aged 65+). The Colombian sample also under-represents the younger people (aged 15–29). Overall, however, the structure of the two samples by age and gender broadly agrees with the population structure.

4.2 Distribution by educational attainments

Table 3 presents the breakdown of the two samples by age group, gender and completed education. We recoded educational attainments into three mutually exclusive categories: "tertiary education", "only secondary education" and "at most primary education".

For the Bulgarian sample, the category "tertiary education" includes people with university degrees (bachelor, master or postgraduate), the category "only secondary education" includes people with (vocational or general) upper-secondary education completed or with "semi-higher" education, while the category "at most primary education" includes people with lower educational attainments.

For the Colombian sample, the category "tertiary education" includes people with at least 5 years of graduate education or with postgraduate education, the category "secondary education" includes people with secondary education completed, "escuela tecnica o tecnologica" or less than 5 years of university education, while the category "primary education" includes people with lower educational attainments.

The Bulgarian sample contains only one missing value on education. The person for whom information is missing was not available for personal interview. The Colombian sample contains no missing values for education.

Relative to the Bulgarian sample, the Colombian sample contains a larger fraction of people with primary education (25.4 versus 19.0 percent), a smaller fraction of people with secondary education only (45.7 versus 54.3 percent), and a slightly larger fraction of people with tertiary education (28.8 versus 26.7 percent). Interestingly, whereas in the Bulgarian sample educational attainments vary little with gender and age, in the Colombian sample they are higher for men than for women and tend to fall with age.

4.3 Labor force status

Table 4 shows the breakdown of the two samples by age group, gender and labor force status (employed, unemployed, and out of the labor force).[6] From the table, one can compute and immediately compare activity rates, employment rates and unemployment rates by main socio-demographic group for the two samples.

The aggregate activity rate is a little higher in the Bulgarian than in the Colombian sample (78.9 versus 74.6 percent), but there are also noticeable differences by gender and age group between

[6] In the Colombian sample there is no missing value for labor force status, whereas in the Bulgarian sample there is only one.

the two samples.

In the Bulgarian sample the activity rate is higher for women (82 percent) than for men (75.3 percent), while in the Colombian sample it is just the opposite (85 percent for men and 66.7 percent for women). Thus, not only the gender differential differs in sign between the two samples (negative in Bulgaria and positive in Colombia), but it also differs in magnitude (much smaller in Bulgaria than in Colombia).

Turning to the age profile of activity rates, this has an inverted U-shape in the Bulgarian sample, with a peak of 98.1 percent for the 45–54 age group, but is monotonically declining in the Colombian sample starting from the 20–29 age group. As a result, while the activity rate is much lower in the Bulgarian than the Colombian sample for the age group 20–29 (81.4 versus 88.6 percent), it is much higher for the age group 55–64 (82.8 versus 76.2 percent).

The aggregate employment rate is about the same in the two samples (70.3 percent in Bulgaria and 71.8 percent in Colombia), but in the Bulgarian sample the employment rate is higher for women (73 percent) than for men (67.2 percent), while in the Colombian sample it is just the opposite (80 percent for men and 65.7 percent for women). Again, not only the differential between men and women is negative in Bulgaria and positive in Colombia, but its size is much smaller in Bulgaria than in Colombia.

Turning to the age profile of employment rates, this has an inverted U-shape in the Bulgarian sample, with a peak of 88.9 percent for the 45–54 age group, but is monotonically declining in the Colombian sample starting from the 20–29 age group. As a result, while the Bulgarian sample shows lower employment rates for people aged 20–29 than the Colombian sample (64.3 percent versus 82.9 percent), it shows higher employment rates for people aged 55–64 (81 percent versus 76.2 percent).

The aggregate unemployment rate is much higher in the Bulgarian sample (10.7 percent) than in the Colombian sample (3.8 percent). Further, while in the Bulgarian sample the unemployment rate is about the same for men and women (10.3 versus 11 percent), in the Colombian sample it is higher for men than for women (6.3 versus 1.5 percent).

Turning to the age profiles of unemployment rates, they tend to decline monotonically with age in both countries. The difference in unemployment rates are particularly striking for the 20–29 age group: 21.1 percent in the Bulgarian sample but only 6.5 percent in the Colombian sample. Also notice that there are no unemployed people in the Colombian sample after age 54.

Table 5 summarizes the variability of activity rates, employment rates and unemployment rates

across different socio-demographic groups in the two samples by fitting logit models to the individual data. The outcome variable is a 0-1 indicator representing whether a person is active (or employed or unemployed). The predictors are indicators for the gender, age group and educational attainments of a person (primary education only, or lower education, and tertiary education, or higher education). The intercept of the model represents the log-odds of being active (or employed or unemployed) for a men aged 30–44, with secondary education only. At the bottom of the table, χ^2 denotes the value of the statistics for testing the joint significance of age and education respectively.

While confirming our previous results on the role of age and gender, the table adds information on the role of schooling attainments. For both samples, lower education is associated with lower activity rates, lower employment rates and higher unemployment rates, whereas higher education is associated with higher activity rates, higher employment rates and (at least in Bulgaria) lower unemployment rates. Notice that the pseudo R^2 of our logistic regressions are far from small, especially for activity and employment, indicating that gender, age and educational attainments alone are enough to predict labor force status for 30–40 percent of the cases in the Bulgarian sample and for 20–30 percent of the cases in the Colombian sample.

4.4 Employment status

For people who are employed, we distinguish between formal and informal employment by using the information on the characteristics of the main job and the presence and nature of second jobs.

We consider several possible definitions of informal employment by distinguishing between: (i) self-employed and dependent employees, (ii) workers who pay and who do not pay social security contributions, and (iii) workers who have and who do not have an informal second job. Because almost all second jobs are informal, that is, either self-employment or jobs for which social security contributions are not paid (38 out of 40 in the Bulgarian sample, and 9 out of 11 in the Colombian sample), for simplicity in what follows we do not distinguish between formal and informal second jobs. For employees, we further distinguish between: (iv) workers with or without a (written) contract, and (v) workers employed in a small (less than 6 workers) or in a larger firm (6 employees or more).

Nonresponse rates for questions on job characteristics are very low. However, some respondents are not informed about all job features and sometimes answer "Don't know" (DK). For Bulgaria, there was no nonresponse for questions on employment status and dual job holding. When asked

about contributions to a pension fund, all 301 employed individuals provided an answer, but 8 of them (2.7 percent) answered DK. Information on the type of contract is missing for 23 out of 243 employees (9.5 percent) because they were not asked this question, due to misleading skip patterns (see also the beginning of section 4.3). Information on the size of the business is missing for 7 out of 243 employees (2.9 percent), and 37 respondents answered DK (15.7 percent of respondents).

In the Colombian survey, there was no nonresponse for questions on employment status and firm size. When asked about contributions to a pension fund, only 1 employed person did not provide an answer. Information on the type of contract is missing for only 1 employee. With respect to dual job holding, employed individuals were asked how many jobs they had at the beginning of Chapter 1. For this question, there is 1 nonrespondent out of 127. After questions on the main job, respondents were again asked if they had another job or if they did any paid work outside the main job. For these questions, there was no nonresponse.

In the Bulgarian sample, the fraction of workers who are self-employed is 19.3 percent (20.6 percent for men and 18.2 percent for women), the fraction of workers who report not paying contributions is 15.4 percent (14.2 percent for men and 16.3 percent for women), while the fraction of workers who report having a second job is 13.2 percent (10.6 percent for men and 15.3 percent for women). On the other hand, the fraction of employees who do not have a contract is only 4.5 percent (7.1 percent for men and 2.5 percent for women), while the fraction of employees working in small firms is 14.6 percent (11.9 percent for men and 16.5 percent for women). The fraction of workers with a secondary job is 13.3 percent (10.7 for men and 15.3 for women).

In the Colombian sample, instead, the importance of "informal" jobs is much higher than in the Bulgarian sample. The fraction of workers who are self-employed is 50.4 percent (41.7 percent for men and 58.2 percent for women), the fraction of workers who report not paying contributions is 46.0 percent (41.7 percent for men and 50.0 percent for women), while the fraction of workers who report having a second job is 8.7 percent (10 percent for men and 7.5 percent for women). On the other hand, the fraction of employees who do not have a contract is 22.6 percent (20.6 percent for men and 25.0 percent for women), while the fraction of employees working in small firms is 19.0 percent (14.3 percent for men and 25.0 percent for women). There are three questions on dual job holding. The number of workers reporting a secondary job in any of the questions is 11 out of 127 (8.7 percent, 10.0 percent for men and 7.5 percent for women). However, 4 of these 11 individuals do not give consistent answers to all the questions on second job holding.

Table 7 summarizes the variability of the first three indicators of informality, namely being self-

employed (`self`), do not paying social security contributions (`nocontrib`), and having a second job (`second`), across different socio-demographic groups in the two samples by fitting logit models to the individual data. The outcome variable is a 0-1 indicator of informality. The predictors are indicators for the gender, age group and educational attainments of a person. The intercept of the model represents the log-odds of being informal (according to the four different definitions) for a men, aged 30–44, with secondary education only.

Self-employment and not paying social security contributions are more likely among people with lower education and less likely among people with higher education, while the opposite is true for having a second job. While the role of gender is less clear-cut, informality appears to be more frequent among younger and older people and less frequent among workers in the central age groups.

Table 8 summarizes instead the variability of the two indicators of informality for employees, namely not having a contract (`nocontract`) and working in a small firm (`smallf`). Also in this case, informality is more likely among the lower educated employees and among employees who are younger and older, whereas gender is not statistically significant.

If compared to the 2004 country indicators in Appendix B, self-employment rates are somewhat higher in our sample for Bulgaria, while they are approximately the same for Colombia. With respect to the size of the informal sector, the survey provides multiple indicators that are not directly comparable to the single measure provided by Schneider (2005). However, while the informal sector size estimated by Schneider is similar for the two countries, indicators derived from the pilot surveys point towards a much larger informal sector in Colombia than in Bulgaria.

4.5 Earnings

Information on hourly earnings was derived from questions on monthly earnings and usual weekly hours of work. In Bulgaria, usual weekly hours were missing for only 2 out of 301 workers (0.7 percent). Monthly earnings were missing for 32 workers (10.6 percent). As a result, hourly earnings were missing for 33 workers (11.0 percent). In Colombia, usual weekly hours were missing for 1 out of 127 workers (0.8 percent). Only 6 workers did not report monthly earnings and they are all self-employed (9.4 percent of self-employed workers). As a result, hourly earnings were missing for 8 workers (6.3 percent).

A simple way of describing the variability of earnings between workers with different observable characteristics is to estimate a Mincerian log-earnings equation for each of the two countries. The

outcome variable is the logarithm of current hourly earnings (defined as current monthly earnings divided by 4.2 times usual weekly hours of work), whereas the predictors are indicators for the gender, age and educational attainments of a worker. To avoid problems arising from extreme values of earnings, the log-earnings equation has been estimated by least absolute deviations rather than ordinary least squares. The results then best describe the behavior of the median of log-earnings.

Table 9 presents, for each country, the results obtained from three models: (i) a basic model, fitted to all workers aged less than 65, with only gender, a quadratic term in age and schooling indicators as predictors, (ii) the basic model, fitted to all workers aged less than 65, with additional indicators for contributing to old age insurance (OAI) and for being self-employed , and (iii) the basic model, fitted only to the subset of employees aged less than 65. with additional indicators for working with a contract and for being an employee of a small firm For all three models, the constant term corresponds to the median hourly earnings of a male worker, aged 40, with secondary education only.

Not surprisingly, in both the Bulgarian and the Colombian samples, hourly earnings tend to be higher for men than for women, and tend to increase with the age and the schooling attainments of a person. The main differences between the two samples are that the gender gap and the educational premia tend to be smaller in Bulgaria than in Colombia. Further, while the coefficient on the self-employment indicator is positive in Bulgaria and negative (and insignificant) in Colombia, the coefficient on the indicator for contributing to OAI is negative in Bulgaria and positive in Colombia. Finally, while in Bulgaria the coefficients on the indicators for working with a contract and for being an employee of a small firm are both negative and statistically significant, in Colombia they are both positive but statistically insignificant. The results for our Colombian sample are in line with those obtained by Magnac (1991) and Attanasio *et al.* (2003) using different waves of the "Encuesta de Hogares".

5 Main empirical results

This section focuses on the areas that represent the main novelty of the survey, namely those concerning job satisfaction, job search, subjective beliefs about uncertain events and attitudes towards the social security system. For these questions, a more detailed analysis of nonresponse patterns is provided. First the number of missing answers is reported, relative to potential respondents. In cases where nonresponse rates were higher than 10 percent, a logistic regression is estimated

to detect any systematic relationship between nonresponse and main socio-demographic variables (gender, age and educational attainments). The evaluation of nonresponse rates required a careful analysis of skip patterns in the questionnaire, because missing values and "not applicable" had the same code in the available data.

5.1 Job satisfaction

In this section we consider a set of questions relating to overall satisfaction with the main job and, in particular, satisfaction with social security benefits.

In Bulgaria, the question on overall satisfaction with the main job was: "Overall, how satisfied are you with your first/main job?" (B20). Possible answers were "Very satisfied", "Satisfied", "Neither satisfied nor dissatisfied", "Dissatisfied" and "Very dissatisfied". This question was asked personally to all employed individuals (238 observations). In Colombia, the wording of the question was: "En general, que tan satisfecho está Ud. con su trabajo principal?" (C14). Possible answers were "Muy satisfecho", "Satisfecho", "Ni satisfecho ni insatisfecho", "Insatisfecho" and "Muy insatisfecho". The question was also asked to all employed individuals (127 observations).

For both Bulgaria and Colombia, there were no nonrespondents.

Table 10 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent is overall satisfied or very satisfied with the main job and 0 otherwise. In Bulgaria, 68.5 percent of the sample report to be satisfied (67.4 for men and 69.1 for women). While gender does not seem to be relevant, satisfaction is less likely among young people aged 20–29.

In Colombia, the fraction of satisfied workers is somewhat higher and equal to 76.4 percent (80.0 for men and 73.1 for women). Again, gender is not statistically significant, while satisfaction is less likely among workers of age 45–54.

More specific questions were asked about satisfaction with some features of the main job. In Bulgaria, the question was: "How satisfied are you with respect to the following characteristics of your first/main job? [Working hours per day|Social security benefits|Skills matching|Earnings|Flexibility|Work environment|Fringe benefits]" (B19). Possible answers were fewer than in the overall question: "Satisfied", "Neither satisfied nor dissatisfied" and "Dissatisfied". A "Not applicable" code was also provided in this case. This question was asked personally to all employed individuals.

In Colombia, the wording of the question was: "De las siguientes características de su trabajo principal, qué tan satisfecho está Ud. con [El número de horas trabajadas a la semana| Las presta-

ciones sociales que tiene|Con la aplicación de sus conocimientos o experiencia en su trabajo|Con el pago que recibe en este trabajo|Con el horario laboral actual|Con el ambiente de trabajo|Beneficios o subsidios distintos a los de ley que Ud. recibe]" (C13). Possible answers were different according to the form used. In questionnaire A, available answers were only "Satisfecho", "Ni satisfecho ni insatisfecho", "Insatisfecho" and "No aplica". In questionnaire B, there were two more available answers: "Muy satisfecho" and "Muy insatisfecho". Again, the question was asked to all employed individuals.

Satisfaction with social security benefits is particularly important for the study of informality. We therefore focus on response patterns to this question (B19b for Bulgaria and C13b for Colombia). For both Bulgaria and Colombia, there were no nonrespondents.

Table 11 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent is satisfied with social security benefits provided by the main job and 0 otherwise. In Colombia, we define as satisfied those who answer "Satisfecho" in form A and those who answer either "Muy satisfecho" or "Satisfecho" in form B.

When focusing on social security benefits, the fraction of workers who report to be satisfied with the main job is very different in the two samples.

In Bulgaria, 65.7 percent of the sample report to be satisfied with main job social security benefits (62.0 for men and 68.0 for women). Satisfaction is less likely among lower educated people, while gender and age are not statistically significant.

In Colombia, the fraction of workers satisfied with their main job social security benefits is only 23.4 percent (23.7 for men and 23.1 for women). There is no clear-cut relationship between satisfaction and socio-demographic variables such as gender, age or educational attainments.

5.2 Job search

One of the questions in the survey asks about willingness to accept a job without social security coverage. If the respondent is willing to accept the job, the reason is then asked. This helps to understand whether informal jobs are chosen optimally or as a temporary occupation while looking for another job. Potential respondents were different in the two pilot surveys.

In Bulgaria, the question was: "Would you accept a job that did not offer social security benefits?" (B47). The question was only asked to unemployed people available to answer in person (23 people). In Colombia, the question was asked to both employed and unemployed people, excluding individuals with a second job (123 people). The wording was "Aceptaría Ud. un trabajo

sin prestaciones sociales?" (C44).

In Bulgaria there are no nonrespondents, whereas in Colombia the number of nonrespondents is only 3 (2.4 percent).

In the Bulgarian sample, the fraction of respondents who would accept a job without social security coverage is 65.2 percent (almost the same for men and women).

The fraction of Colombian respondents who would accept a job without social security coverage is only 41.7 percent, and is higher among women (51.6 percent) than among men (31.0 percent). Among the unemployed, however, the fraction is similar to the Bulgarian sample, because 3 out of 5 people are willing to accept the job.

Due to the small number of unemployed individuals in both samples, it is not possible to draw any conclusion on the relationship between willingness to accept an informal job and basic socio-demographic variables.

In Bulgaria, of those willing to accept a job without social security coverage (15 people), almost half (7) said that they would accept it temporarily while looking for a job offering social security coverage, 5 said that social security coverage was not important for them, and 3 gave other reasons. There were no missing values on this question.

In Colombia, of those unemployed willing to accept a job without social security coverage (3 people), 1 said that he would accept it temporarily while looking for a job offering social security coverage, 1 said that social security coverage was not important for him, and 1 gave other reasons. As for the Bulgarian sample, there were no missing values on this question.

5.3 Subjective beliefs and perceptions

In this section we consider a set of questions relating to subjective expectation about life duration, retirement age, replacement rate, quality of life after retirement, future standards of living, and perceived employment and health risks, For each of these topics, we only focus on a few questions that we consider as representative of the response patterns in the two samples.

5.3.1 Life duration

Individual beliefs about future or uncertain events are strictly related to the decision of contributing to social security. In particular, the value of affiliation to an old age pension program can be different depending on the respondent's subjective beliefs about life expectancy. For example, people who expect not to live for a long period after they retire, may be less interested in old age insurance. The way in which these beliefs have been elicited is different in the two pilot surveys.

In Bulgaria, respondents had to attach a probability to survive until age 70, 80 or 90 (B89a–c). The wording of each question was "What are the chances (any number from 0 to 100, where 0 means unlikely and 100 means certain) that you will live to be age [70|80|90]?". The questions were asked personally to all people over 15 years of age and below the target age (284, 297 and 302 individuals for the three questions respectively). In the Colombian survey, instead, people aged 30+ (130 individuals) were simply asked about their life expectancy (C83) ("Hasta qué edad cree Ud. que va a vivir?").

Nonrespondents in Bulgaria were only 8 (all women) for the first question (surviving till age 70), 13 (4.4 percent) for the second question (surviving till age 80), and 14 (4.6 percent) for the third question (surviving till age 90).

In Colombia, nonrespondents were only 2, both women aged between 50 and 60. This question was also answered by some individuals aged less than 30, who were not supposed to be asked. Our analysis of nonresponse and distribution of answers is limited to the reference group 30+.

Table 12 presents some statistics on the distribution of answers to the three survival probability questions in Bulgaria. We define as optimists those who assign a probability higher than 50 percent to a given event. Table 14 presents the estimates of a logit model where the outcome variable is a 0-1 indicator of optimism. Respondents in young and old age groups are more likely to give a high survival probability than middle-aged individuals. The effect of gender is less clear-cut.

Table 13 presents some statistics on the distribution of life expectancy in the Colombian sample. The mean life expectancy for women is equal to 75.7, which is fairly high given the fact that it is close to the current Colombian life expectancy at birth as reported in the 2004 World Development Indicators (see also Appendix B). For men, the sample mean is even higher than the WDI life expectancy at birth (77.6 and 69.6 respectively). It should be noted however that the sample life expectancy is an average of subjective beliefs of individuals of any age 30+.

Respondents are defined as "optimist" if they have a life expectancy higher than the median value (80 years). Table 14 shows that being older than 65 and having lower education increase the probability of a being an optimist. As for the Bulgarian sample, gender is not statistically significant.

5.3.2 Retirement age and replacement rates

The similar sequence of questions was used to elicit respondents' beliefs about age at retirement.

In Bulgaria, respondents had to attach a probability to full-time work after age 50, 60 or 70

(B91a–c). The wording of each question was "What do you think the chances are (any number from 0 to 100) that you will be working full-time after you reach age [50|60|70]?". The questions were asked personally to all people over 15 and below the target age (186, 253 and 284 individuals for the three questions respectively).

In the Colombian survey, instead, the expected age at retirement was asked (C95: "Hasta que edad cree Ud. qué va a trabajar?"). The question was asked to employed individuals aged 30+, excluding unpaid workers in family business (96 individuals).

Nonrespondents in Bulgaria were only 4 for the first question (working full-time after age 50), 7 for the second question (working full-time after age 60), and 6 for the third question (working full-time after age 70).

In Colombia, there is only one nonrespondent. This question was also answered by some individuals aged less than 30, who were not supposed to be asked. Our analysis of nonresponse and distribution of answers is limited to the reference group 30+.

Table 15 presents some statistics on the distribution of answers to the three full-time work probability questions in Bulgaria.

Table 17 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent has a subjective probability higher than 50 percent, and 0 otherwise. Women are less likely to have a high subjective probability of working full-time after age 60 and 70. There is also a negative effect of lower educational attainments and a positive effect for age group 45–54.

Table 16 presents some statistics on the distribution of expected retirement age in the Colombian sample. It is worth noting that the mean expected retirement age, 62 for men and 60.6 for women, is higher than the minimum age required for the social insurance old age pension (60 for men and 55 for women) (see also Appendix C).

Table 17 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent has an expected retirement age higher than the sample median (60), and 0 otherwise. Women and lower educated respondents are less likely to have a high expected retirement age, while the effect of higher education and of age groups 45–54 and 55–64 is positive. Despite the different methods used to elicit subjective beliefs about working life duration, some results are common in the two samples. Being a male, having higher educational attainments and being aged 45–54 have a positive effect either on the subjective probability of being working (Bulgaria) or on the expected retirement age (Colombia).

Another subjective belief that the survey tries to elicit is the expected replacement rate, i.e.

the expected pension as a percentage of last earnings. The wording of the question and the target group was different in the two pilot surveys.

In Bulgaria, the question was: "What percentage of your last (after tax) labor earnings you expect your post-retirement pension to represent?" (B105). This question was asked personally to employed individuals aged 40+ (159 individuals). The status of respondents (working or not working) was assessed in question B99, by checking answers to previous questions B4 and B5. In the Colombian survey, the question was "Qué porcentaje de su último ingreso laboral cree usted que representará su pensión de jubilación?" (C96). Respondents had to choose between the following brackets: "Menos de 30%", "Entre 30% y 40%", "Entre 40% y 50%", "Entre 50% y 60%", "Entre 60% y 70%", "Entre 70% y 80%", "Entre 80% y 90%", "Mas de 90%". The question was asked to employed individuals aged 30+, excluding unpaid workers in family business (96 individuals). Among these, 27 are excluded from nonresponse calculations because their answer is coded as "No aplica".

While nonresponse rates for the other questions were very low for both the Bulgarian and Colombian sample (usually below 5 percent), this question had very different nonresponse rates in the two samples, much higher in Bulgaria than in Colombia. Nonrespondents in Bulgaria were 49 (30.8 percent, 36.8 for men and 27.4 for women). To check the existence of a systematic relationship between nonresponse and socio-demographic variables, a logistic regression model is estimated, with the binary outcome equal to 1 if the answer is missing and 0 otherwise. Being a man increases the probability of nonresponse, as well as having a lower educational attainment and being older than 65. People aged 45–54 are instead more likely to answer the question. On the other hand, in Colombia there is only one nonrespondent.[7]

Table 18 presents some statistics on the distribution of answers to replacement rate question in Bulgaria.

Table 19 presents the estimates of two simple linear regression models where the subjective replacement rate is the dependent variable. In the first model, standard explanatory variables are used, whereas in the second model we use the number of years of contributions instead of age. The intercept of the model is the average subjective replacement rate for a man with secondary educational attainments and 40 years of age or 20 years of contributions. There is no statistically significant relationship except for a negative effect of higher educational attainments, however the coefficient is not significant after controlling for years of contributions.

[7] This question was also answered by some individuals aged less than 30, who were not supposed to be asked. Our analysis of nonresponse and distribution of answers is limited to the reference group 30+.

With respect to the Colombian sample, subjective replacement rates were categorized into intervals. Therefore, Table 19 presents the estimates of interval regression models where the dependent variable is the latent subjective replacement rate. As for Bulgaria, the first model includes age whereas the second model controls for the number of years of contributions. In this sample, the effect of education is reversed with respect to the Bulgarian sample. In both models, the coefficient for lower educational attainments is negative and significant. Further, coefficients for age and years of contributions are positive and significant, although the latter is somewhat smaller. As for the Bulgarian sample, gender is not statistically significant.

5.3.3 Future standards of living

Individuals were asked to report their beliefs about future standards of living in two different questions. The first one asked about the next five years, while the second asked about post-retirement living standards.

Subjective beliefs about standards of living in the next five years were asked in both surveys, although in a slightly different way. In Bulgaria, the question was: "In your opinion, compared to today, your standards of living in 5 years will be:" (B93). Possible answers were "Much improved", "Slightly improved", "Same", "Slightly worse" and "Much worse". This question was asked to all individuals in the sample and available to answer personally (303 individuals).

In the Colombian survey, the question was "En los próximos cinco años su nivel de vida será:" (C85). Possible answers were "Mejor que el actual", "Igual que el actual" and "Peor que el actual". The question was asked to all people aged 30+ (130 individuals).

In Bulgaria there was only 1 nonrespondent, whereas in Colombia there was none.

Table 20 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent is optimist with respect to her quality of life in the next five years and 0 otherwise. For Bulgaria, optimists are those who answer "Much improved" or "Slightly improved". They are 42.4 percent of the sample (49.6 for men and 37.6 for women). Optimism is positively associated with being male and age groups 15–19 and 20–29. People aged 55-64 and 65+ are less likely to be optimists. In Colombia, optimists were a much higher fraction than in the Bulgarian sample: 68.5 percent of respondents (61.8 for men and 73.3 for women). Optimism is less likely among the lower educated and again among people aged 55–64, while gender does not seem to be relevant.

The question on expected quality of life after retirement had a very similar wording in the two pilot surveys. In Bulgaria, the question was: "What do you expect your standards of living will

be after you retire from your main job?" (B114). Possible answers were "Better than today", "Like today", "Worse than today" and DK (DK). This question was asked personally to employed individuals regardless of age (238 individuals). The status of respondents (working or not working) was assessed in question B99 as for the question on expected replacement rate (B105).

In the Colombian survey, the question was "Despúes de retirarse de su trabajo principal, Ud cree que su calidad de vida será:" (C103). Possible answers were "Mejor que la de hoy", "Como la de hoy" and "Peor que la de hoy". A DK answer was not provided. The skip pattern for this question was not clear. To avoid overestimates of nonresponse, we use the same reference group as for question C96 (expected replacement rate), namely employed individuals aged 30+, excluding unpaid workers in family business (96 individuals).

Nonrespondents in Bulgaria were 21 (8.8 percent, 2.2 for men and 12.7 for women) and, among the respondents, 36 answered DK (16.6 percent, 14.9 for men and 17.7 for women). A logistic regression model is estimated, with the binary outcome equal to 1 if the answer is DK and 0 otherwise. People aged 20–29 and with low educational attainments are more likely to give a DK answer. In Colombia, instead, all individuals answered the question and there is no DK answer.

Table 21 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent expects her post-retirement quality of life to be better than at the time of interview, and 0 otherwise. In Bulgaria, optimists were 18.9 percent of respondents (24.1 for men and 15.4 for women). Optimism is associated with male gender and age 20–29. In Colombia, optimists were a much higher fraction than in the Bulgarian sample: 55.2 percent of respondents (55.3 for men and 55.1 for women). There is no clear-cut relationship between optimism and standard socio-demographic variables.

5.3.4 Employment risks

Some questions aimed at eliciting subjective beliefs about employment risks. In particular, one question was about the possibility of loosing one's job in the next 12 months. In Bulgaria, this question was asked personally to employees of any age (186 individuals): "In your opinion, how likely is it that you will lose your job within the next 12 months?" (B101). Possible answers were "Very likely", "Likely", "Unlikely" and "Very unlikely". In Colombia, a similar question (C89) was asked to employed individuals aged 30+, excluding unpaid workers in family business but including the self-employed (96 individuals). The question had a different wording in form A and B. The former asked for a qualitative answer: "Ud. Cree muy probable; probable; poco probable o

improbable que Ud pierda su trabajo dentro de los próximos 12 meses?". Answers were coded as "Muy probable", "Probable", "Poco probable" and "Muy poco probable". For those with form B, the question was: "Cuál es la probabilidad (un número entre 1 y 100) que Ud. pierda su trabajo dentro de los próximos 12 meses?". Unfortunately, in the wording of this question probabilities range from 1 to 100 instead of 0 to 100. Actually, a very large fraction of respondents answered "1". Nonrespondents in Bulgaria were only 2, whereas in Colombia there was none.

Table 22 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent is optimist with respect to employment risks in the next 12 months, and 0 otherwise. For Bulgaria, optimists are those who answer "Very unlikely" or "Unlikely". They are 65.6 percent of the sample (63.8 for men and 66.7 for women). Optimism is negatively associated with older age, while gender does not seem to be relevant.

In Colombia, a similar definition of optimism is used for respondents to form A (answers "Poco probable" and "Muy poco probable"). Optimists were 68.1 percent of form-A respondents (75.0 for men and 63.0 for women). Respondents to form B were defined as optimists if they had a subjective probability lower than 50%. Optimists among this group were 67.3 percent (77.8 for men and 54.5 for women). These results are not directly comparable due to the arbitrary choice of 50% as the probability below which a person is considered as an optimist. For respondents of form B, the median subjective probability is 10%. From Table 22 and for form-B respondents, optimism is less likely among women, while age group variables are not statistically significant. Results from form A are less clear-cut.

5.3.5 Health risks

The survey tries to elicit subjective beliefs about health-related risks for the 12 months following the interview. One of the questions is about the possibility of a serious illness and another one is about the possibility of a serious accident. In both the Bulgarian and the Colombian samples, individual assessments of the two risks are highly positively correlated. However, there are important differences in the way beliefs were elicited in the two samples. The main one is that in Bulgaria people were asked to attach a probability to a certain event, while in Colombia two different forms were used. For respondents using form B, a subjective probability was asked. For those using form A, answers were simply qualitative. This section discusses the main results for the two questions and for each form separately.

In Bulgaria, both questions on illness and accident risk had the same wording: "Now we would

like to ask you some questions about risks. What are the chances (any number from 0 to 100) that each of the following events could happen to you within the next 12 months? [Serious illness (causing physical incapacity/unable to work for long time)| Serious accident (causing physical incapacity/unable to work for long time)]" (B94b–c). This question was asked to all individuals available to answer personally (303 people).

In the Colombian survey, the questions for individuals with form A were: "Ud. cree muy probable; probable; poco probable o muy poco probable, que en los proximos 12 meses le pueda ocurrir a usted [Enfermedad grave|Accidente grave]" (C86b–c). For those with form B, the questions were: "Cuál es la probabilidad (un número entre 1 y 100) que en los proximos 12 meses le pueda ocurrir a usted [Enfermedad grave|Accidente grave]". Unfortunately, in the wording of this question probabilities range from 1 to 100 instead of 0 to 100. Actually, a very large fraction of respondents answered "1". In both forms, the question was asked to all people aged 30+ (130 people).

Nonrespondents in Bulgaria were only 3 for the illness question, and 5 for the accident question. In Colombia, there were no missing answers to questions in form B (quantitative answers). For form A (qualitative answers), there was only 1 nonrespondent for both the illness and accident questions.

Table 23 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent is optimist with respect to helth risks in the next 12 months, and 0 otherwise. For Bulgaria, optimists are those who have a subjective probability lower than 50%. They are 76.0 percent of the sample (76.0 for both men and women) for the illness question and 79.5 percent of the sample (75.6 for men and 82.1 for women) for the accident question. The median subjective probability for both questions is 10%.

Optimism is positively associated with ages 20–29 and 45–54 for both illness and accident questions. Lower education has a negative and statistically significant effect for optimism about accident risk, while age 65+ has a significant negative effect on optimism about illness risk. Gender is not statistically significant for any of the questions.

In Colombia, respondents to form B were defined as optimists if they had a subjective probability lower than 50%, as for the Bulgarian sample. Optimists among this group were a higher fraction than in Bulgaria: 86.1 percent (82.8 for men and 88.9 for women) for the illness question and 81.5 percent (82.8 for men and 80.6 for women) for the accident question. The median subjective probabilities are 3% for the illness question and 5% for the accident question, both lower than the Bulgarian sample. Colombian respondents to form A were defined as optimists if they answered

"Poco probable" or "Muy poco probable". They were 70.3 percent of respondents (73.1 for men and 68.4 for women) for the illness question and 62.5 percent (61.5 for men and 63.2 for women) for the accident question. These results are not directly comparable to form B, due to the arbitrary choice of 50% as the probability below which a person is considered as an optimist. From Table 23 and for form-A respondents, optimism about both risks is less likely among individuals aged 65+ (as in the Bulgarian sample) and more likely among people with lowest and highest educational attainments, while gender does not seem to be relevant. Results from form B are less clear-cut.

5.4 Impatience and risk aversion

To elicit individual risk aversion and impatience, the survey asked three questions for each of the two topics. As discussed above, respondents were asked to imagine a hypothetical situation where they held the winning ticket of a lottery. In the first three questions (time preferences), the prize is assumed to be certain but payable after one year. In the second group of questions (risk attitudes), the prize would be paid immediately, but there is a positive probability that the lottery organization will not be able to pay for the prize. In both situations, individuals are asked if they would accept to sell the winning ticket in exchange for a lower amount to be paid immediately (time preferences)/for sure (risk aversion). According to the answer, the question is repeated using a different amount to be paid for the ticket. The questioning structure is the same for both pilot surveys, although amounts and skip patterns actually used are different.

5.4.1 Impatience

The impatience question for Bulgaria was: "Imagine you won the first prize of a national lottery, which is worth 1000 leva. The lottery administration is very reliable, so that you would get the money for sure, but only one year from now (suppose there is no inflation). Then a friend of yours asks you to sell him the ticket for 700, which he would pay immediately. What would you prefer to have?" (B115). If respondents preferred to have 700, they were not asked further questions. On the other hand, if they chose 1000, they were asked if they would accept 800 (B116). If they still preferred to have 1000 after one year, they were asked if they would accept 900 (B117). This set of questions was asked to all individuals available for personal interview (303 people). The following table presents the discount rates implied by the answers received.

B115	B116	B117	Discount rate
1000	1000	1000	r < 0.11
1000	1000	900	0.11 < r < 0.25
1000	800	-	0.25 < r < 0.43
700	-	-	0.43 < r

In the Colombian survey, the first question on impatience was: "Imaginese que Ud. se gana el premio mayor de la loteria, que vale 1 millon de pesos, pero la loteria sólo le entregará el dinero dentro de un año. Se sabe que la lotería siempre paga los premios y que el dinero vale lo mismo en un año. Un amigo le propone comprarle el billete de loteria por 920,000 pesos que le pagaria de inmediato. Qué preferiría?" (C104). If respondents chose 920,000, they were not asked any other question on this topic. If they preferred 1,000,000 one year later, they were asked if they would accept 950,000 (C105). Finally, if they still preferred 1,000,000 after one year, they were offered 980,000 (C106). This set of questions was asked to all people aged 30+ (130 people). The following table presents the discount rates implied by the answers received.

C104	C105	C106	Discount rate
1,000,000	1,000,000	1,000,000	r < 0.02
1,000,000	1,000,000	980,000	0.02 < r < 0.05
1,000,000	950,000	-	0.05 < r < 0.09
920,000	-	-	0.09 < r

We define as missing those observations for which the complete set of answers is not available, after taking into account the correct skip patterns. For Bulgaria there were no nonrespondents, whereas in Colombia nonrespondents were only 5.

Table 24 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent is impatient, i.e. if she is willing to accept the lower amount available immediately, and 0 otherwise. Of course, direct comparisons are not possible, because impatience corresponds to a subjective discount rate higher than 43% in Bulgaria and higher than 9% in Colombia. In Bulgaria, 58.4 percent of the sample are impatient (59.5 for men and 57.7 for women), whereas in Colombia, 74.4 percent of respondents were impatient (73.1 for men and 75.3 for women). In both samples, age and gender help predict impatience. On the other hand, the coefficient for lower educational attainments is always statistically significant, although with a different sign in the two countries.

5.4.2 Risk aversion

The risk aversion question for Bulgaria was: "Suppose now that the prize you won is not from a national lottery, but from one which is less reliable, so that there is only a 50% chance to get the

money. However, in case you are lucky you get paid tomorrow. Your friend offers you 500 for this ticket. What would you prefer to have?" (B118). If respondents preferred to have 500, they were asked if they would accept 400 instead (B120). On the other hand, if they chose 1000 at question B119, they were asked if they would accept 600 (B119). This set of questions was asked to all individuals available to answer personally (303 people).

The answers to these questions may be used to obtain a range for the Arrow-Pratt measure of absolute risk aversion (ARA). We use the method presented in Guiso and Paiella (2001). The following table summarizes how individuals are classified according to answers given. Risk neutral individuals have ARA equal to 0, while for risk averse individuals ARA>0 and for risk lovers ARA<0. To obtain the relative measure of risk aversion, ARA should be multiplied by the level of wealth.

B118	B119	B120	Absolute risk aversion				
1000	1000	-			ARA	$<$	-0.0008
1000	600	-	-0.0008	$<$	ARA	$<$	0
500	-	1000	0	$<$	ARA	$<$	0.0008
500	-	400	0.0008	$<$	ARA		

In the Colombian survey, the same structure was used, although with a probability of 70% instead of 50%, and with different amounts. The starting question was "Ahora suponga que el premio ganado, un millon de pesos, no es de una lotería igual de confiable a la anterior y que solo hay un 70% de probabilidad de que le pagen el premio, pero teniendo mucha suerte se la podrían pagar mañana. Un amigo le ofrece 700,000, que preferiría Ud.?" (C107). Unfortunately, the skip patterns were not properly defined for this set of questions. If respondents preferred 700,000, they were not asked any other question on this topic. As a consequence, the measure of risk aversion that we obtain is not as detailed as for Bulgaria, because we cannot separate risk neutral from risk averse individuals. If respondents preferred 1,000,000, they were asked if they would accept 800,000 instead (C108). This set of questions was asked to all people aged 30+ (130 people). The following table summarizes the implied level of absolute risk aversion.

C107	C108	Absolute risk aversion				
1000000	1000000			ARA	$<$	-0.0000009
1000000	800000	-0.0000009	$<$	ARA	$<$	0
700000	-	0	$<$	ARA		

We define as missing those observations for which the complete set of answers is not available, after taking into account the correct skip patterns. In Bulgaria there were no nonrespondents, whereas in Colombia there were only 2.

Table 25 presents the estimates of a logit model where the binary outcome is equal to 1 if the respondent has the lowest level of absolute risk aversion (less than -0.0008 for Bulgaria and less than -0.0000009 for Colombia), and 0 otherwise. Of course, direct comparisons of the two samples is not possible. In Bulgaria, 24.7 percent of the sample consist of risk lovers (25.6 for men and 24.2 for women). Risk loving is negatively associated with age 45–54. Gender is not statistically significant.

In Colombia, a lower fraction of respondents is risk lover (13.3 percent, 9.3 for men and 16.2 for women). Risk loving is more likely among older people (aged 65+). Again, gender does not seem to matter much.

5.5 Participation and attitudes towards Social Security

In this section we consider a set of questions relating to who should be responsible for social security, preferences between a PAYG and a fully-funded pension scheme, reasons for paying social security contributions, expected future changes in the generosity of the social security system, willingness to pay more contributions in order to receive a higher pension, willingness to receive a lower pension in order to pay less contributions, and trade-offs between the various components of the "social security package". As before, for each topic we only focus on a few questions that we consider as representative of the response patterns in the two samples. Unless sated otherwise, these questions were asked to all individuals in the sample.

5.5.1 Personal views on the social security system

Personal views about the social security system were asked somewhat differently in the two countries.

In Bulgaria, opinions about who should be responsible for social security were asked through four different questions: "Who do you think should be responsible for paying the [old age pensions|social pensions, such as disability, survivors, etc.|social insurances (such as for unemployment, work injury, etc.)|medical services]?" (B121-B124). For all questions, the possible answers were "The state", "The employer", "Oneself" or "Other". A DK answer was possible as well.

In Colombia, instead, a single question was asked: "En su opinión, quién debería encargarse de las prestaciones sociales como la pensión de jubilación (vejez), los servicios médicos o de salud, la pensión de invalidez o de supervivencia?" (C110). The possible answers were "El gobierno", "El empleador", "Uno mismo contratando una compañia de seguros" or "Otro".

For Bulgaria, we focus on response patterns for questions B121 (old age pensions) and B122 (disability and survivors pensions). There are no nonrespondents for both questions, and among respondents only 2 answer DK to the first question (old age pensions) and only 1 to the second question (disability and survivors pensions). The number of nonrespondents in the Colombian sample is 7, with no DK.

The fraction of Bulgarian respondents who think that only the state should be responsible for old age pensions is 84.8 percent, and is about the same for men and women. After controlling for gender, age and educational attainments through a logistic regression, this fraction is significantly lower for teen-agers and people with higher education, and significantly higher for older people and people with lower education (Table 26). Gender does not seem to be relevant. Almost all respondents think that only the state should be responsible for disability and survivors pensions (96.7 percent, about the same for men and women). There is no statistically significant relationship between this fraction and main socio-demographic variables.

In Colombia, the fraction of respondents who think that the state should be responsible for social security is 64.1 percent, much lower than in the Bulgarian sample, but again about the same for men and women. After controlling for gender, age and educational attainments through a logistic regression, no systematic relationship emerges except that this fraction is significantly lower for people with higher education.

In Bulgaria, the question about preferences between a pay-as-you-go (PAYG) or a fully-funded pension scheme was: "Suppose there are two possible pension schemes. In the first one, current contributions are used to finance pensions for current retirees, while in the second one contributions add to your own pension. Which would you prefer?" (B125). The possible answers were "Current contributors paying for current pensioners" or "Contributions accumulating to own pension".

In Colombia, the following question was asked: "Suponga Ud. que hay dos planes de pensiones. En el primero, las cotizaciones se usan para financiar las pensiones de los jubilados actuales. Mientras que en el segundo las cotizaciones de cada persona se acumulan para pagar su propia pensión. Cuál plan preferiría Ud.?" (C111). The possible answers were "Cotizaciones para pagar los pensionados actuales" or "Cotizaciones que se acumulan para la propia pensión". A DK answer was also possible.

There were 11 nonrespondents in Bulgaria (3.6 percent) and 8 in Colombia (4.5 percent). Among the respondents, only 1 person in Colombia answered DK.

The fraction of Bulgarian respondents who would prefer a PAYG pension scheme is 22.3 percent,

about the same for men and women. Table 27 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual prefers a PAYG scheme and 0 otherwise. The PAYG scheme is preferred by people aged 65+, while gender is not statistically significant.

In Colombia, the fraction of respondents who would prefer a PAYG pension scheme is slightly lower (18.9 percent), and is again about the same for men and women. Preference for PAYG schemes is positively associated with age 55–64, whereas gender does not seem to be relevant.

5.5.2 Reasons for paying social security

There are many reasons why people may decide to pay social security contributions. One possibility is that the social security system is perceived as a tool to support people in need ("solidarity argument"). Another possibility is that the public pension is considered as a fair return on paid contributions ("economic argument"). Respondents to both surveys were asked to say how much they agreed with a set of statements related to these different motivations.

One statement was about the "solidarity argument". For Bulgaria, the sentence was: "People that are older or poorer need to be helped" (B127g). The possible answers were "Strongly agree", "Agree", "Neither agree nor disagree", "Disagree" or "Strongly disagree". In Colombia, the sentence was instead: "Las personas mayores o las pobres necesitan ser ayudadas" (C113c). The possible answers were "Muy de acuerdo", "De acuerdo", "Ni de acuerdo ni en desacuerdo", "En desacuerdo" or "Muy en desacuerdo". The number of nonrespondents is 4 in Bulgaria and 6 in Colombia.

The fraction of Bulgarian respondents who agree or strongly agree with the "solidarity argument" is 96.3 percent, about the same for men and women. Table 28 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual agrees or strongly agrees with the statement and 0 otherwise. There are no clear-cut relationships between agreement and main socio-demographic variables.

In Colombia, the fraction of respondents who agree or strongly agree is just a little lower (92.4 percent), although with some difference between men (95.8 percent) and women (89.9 percent). Table 28 confirms that agreement is less likely among women and, surprisingly, among older people (age 65+).

The statement about the "economic argument" for paying social security contributions was phrased as follows in Bulgaria: "Paying SS contributions today is a good deal for me because the return is adequate and secure" (B127h). In Colombia, the statement on economic convenience was:

"Pagando las cotizaciones a la SS hoy es un buen arreglo para mi porque la devolución es segura y adecuada" (C113d). For both countries, the possible answers were the same as for the "solidarity argument" question. The number of nonrespondents is 5 in Bulgaria and 6 in Colombia.

The fraction of Bulgarian respondents who agree or strongly agree with the "economic argument" is only 36.9 percent, about the same for men and women. Table 29 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual agrees or strongly agrees with the statement and 0 otherwise. Agreement is less likely among people aged 20–29 and more likely among those aged 65+, while gender is not relevant.

In Colombia, the fraction of respondents who agree or strongly agree is instead quite high and equal to 63.2 percent, with some differences between men and women (56.9 and 67.7 percent respectively). After controlling for gender and age, agreement is associated positively with lower education and negatively with higher education.

An additional question about reasons for not contributing to a pension fund was asked to people who never contributed. In Bulgaria, the question was: "Why haven't you paid (did you stop paying) social security contributions?" (B134). Possible answers were "Benefits are too low compared to the costs", "Salary is too low to afford it", "You don't have any choice (you are forced by the circumstances, the decision of your employer,etc.", "Don't know how to do it", "I have already paid enough" and "Other".[8] Potential respondents were people available for personal interview who never paid contributions to the social security system (25 people). In Colombia, the question was "Por qué razón no cotizò?" (C132). Possible answers were "Prestaciones demasiado bajas en comparación con lo cotizado", "El salario no me alcanza", "Mi empleador no me ha afiliado", "No sabe como hacerlo", and "Otro". Potential respondents were people aged 30+ who never paid contributions to social security (13 people). In both surveys, there were no nonrespondents. In the Bulgarian sample, 16 respondents out of 25 did not pay contributions because they were unemployed, students, or not earning income. Other respondents did not contribute because benefits were too low compared with the costs (4 people), because they did not have any choice (3 people) or because salary was too low to afford it (1 respondent). In Colombia, 13 respondents did not contribute because salary was too low to afford it (6 people), because of the employer (2 people), because they did not know how do it (2 people), or for other reasons (3 people).

[8] The question was only asked to people who never contributed, therefore the wording "(did you stop paying)" and the answer "I have already paid enough" did not apply.

5.5.3 Expected changes in the social security system

Subjective beliefs about future changes in the social security system might affect the decision to affiliate. One of the survey questions asks about expected changes in the generosity of the system in the next 10 years.

In Bulgaria, the question was: "In your opinion, over the next 10 years the public social security system will be more generous, less generous or unchanged?" (B131). The possible answers were "More generous", "Unchanged" and "Less generous". In Colombia, the question was: "Cómo cree Ud que será el sistema público de pensiones en los próximos 10 años?" (C115). The possible answers were the same as in Bulgaria, although with a different ordering: "Más generoso", "Menos generoso" and "Igual que ahora". In the Bulgarian sample nonrespondents were 15 (4.9 percent, 2.5 for men and 6.6 for women), whereas in the Colombian sample they were only 5.

Table 30 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual expects the system to become more generous and 0 otherwise. In Bulgaria, the fraction of optimists is 37.1 percent, higher for men than for women (41.5 percent and 34.1 percent respectively). However, after controlling for standard socio-demographic variables such as age and education, neither gender nor age appear to be statistically significant.

In Colombia, the fraction of optimists is smaller than in Bulgaria (13.9 percent), but again higher for men than for women (17.8 percent and 11.1 percent respectively). The negative relationship between optimism and female gender is confirmed by the estimated logistic regression in Table 30. Further, lower educated people are more likely to be optimists, while age is not relevant.

5.5.4 Willingness to pay and preferences over social security programs

A set of questions in Chapter 5 asks about respondents' willingness to reduce (or to increase) the level of contributions for the old age pension. The question make it clear that a reduction (increase) in contributions would also imply a reduction (increase) in future benefits. Respondents were first asked if they would have liked to pay higher contributions throughout their working life, in order to receive a higher old age pension. Only if the answer was negative, they were asked if they would have liked to reduce contributions throughout their working life, even if that implied a reduction in the old age pension.

Potential respondents were different for the two pilot surveys. In Bulgaria, they were people currently paying social security contributions or who contributed in the past. In Colombia, potential respondents were only workers currently contributing to social security, excluding unpaid workers

in family business.

In Bulgaria, the exact wording for the first question (willing to increase contributions) was: "Would you like to receive a higher retirement pension even if that implies paying a higher contribution throughout all your working life? " (B137). Potential respondents were individuals available for personal interview, who were currently contributing to social security or contributed at least in the past (278 people).

In Colombia, the first question was: "A Ud. le gustaría recibir una pensión de jubilación mayor y mayores beneficios, aunque eso le implique aumentar el valor de las cotizaciones durante toda su vida laboral ?" (C121). Potential respondents were all workers who were currently paying social security contributions (36 people).

In Bulgaria there were no nonrespondents, while in Colombia there was only 1. The number of potential respondents in Colombia is very small due to the low fraction of workers paying contributions to social security.

Table 31 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual is willing to pay higher contributions and 0 otherwise. In Bulgaria, the fraction of people willing to pay more is 56.1 percent, smaller for men than for women (52.2 percent and 58.7 percent respectively). However, after controlling for standard socio-demographic variables such as age and education, gender is not statistically significant in explaining willingness to pay higher contributions. On the other hand, lower educated people are less likely to be willing to pay more. People aged 65+ are instead more likely to be in favour of higher contributions. It should be noted that this question was asked also to people already retired, because potential respondents were all individuals who ever paid contributions.

In Colombia, the fraction of people willing to pay more is much higher than in Bulgaria: 91.4 percent, 100.0 percent for men and 81.2 percent for women. However, results are not directly comparable, because in Colombia only people who were currently paying contributions were asked this question. Retired people are therefore excluded. Due to the small sample size, it is not possible to relate willingness to pay higher contributions to any standard socio-demographic variable.

In both pilot surveys, an opposite question was asked only to those who were not willing to pay higher contributions for the old age pension.

In Bulgaria, the exact wording for the second question (willing to reduce contributions) was: "Would you like to contribute less for your retirement pension throughout your working life even if that implies receiving lower benefits when you retire?" (B138).

In Colombia, the second question was: "A Ud. le gustaría cotizar menos para la pensión de jubilación aunque esto implique una disminución en el valor de la pensión, despúes de su retiro?" (C122).

In Bulgaria there were no nonrespondents out of 122 individuals, whereas in Colombia potential respondents were only 4 and 1 is nonrespondent, male.

Table 31 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual is willing to pay lower contributions and 0 otherwise. In Bulgaria, the fraction of people willing to pay less is 32.8 percent, higher for men than for women (39.6 percent and 27.5 percent respectively). However, after controlling for standard socio-demographic variables such as age and education, gender is not statistically significant in explaining willingness to pay lower contributions. On the other hand, people aged 20–29 are more likely to be willing to reduce contributions. In Colombia, only 1 of the 3 respondents, female, is willing to pay lower contributions.

People who never contributed to social security were asked if they would be willing to contribute. The wording of the question was however different in the two pilots. In Bulgaria, the question was "Would you like to pay social security contributions in order to receive a pension when you retire?" (B135), and it was asked to people available for personal interview who never paid contributions (25 people). In Colombia, the wording took into account the relationship between contributions and benefits: "Hubiera usted cotizado si las contribuciones hubieran sido más bajas, aun cuando los beneficios recibidos también hubieran sido menores?" (C133). Potential respondents were individuals aged 30+ who never contributed to social security (13 people).

In both pilots, there were no nonrespondents. In the Bulgarian sample, the number of people willing to pay contributions is 15 out of 25 if we consider all respondents, and 4 out of 9 if we do not include those who said they did not contribute because they were unemployed, students or with no income. It is worth noting that answers to this question are consistent with reasons for not paying contributions (B134): all respondents who did not pay because "benefits are too low compared to the costs" said they were not willing to contribute, whereas 3 people out of 4 who did not contribute for other reasons ("Don't know how to do it" or "Salary is too low to afford it") said they would be willing to contribute. In the Colombian sample, all 13 respondents were working and 6 of them said they would be willing to contribute. Cross-tabulations with reasons for not paying contributions provide mixed results: for example, those who did not contribute because of the employer's decision said they would not be willing to contribute, whereas those who did

not know how to contribute said they would be willing to pay contributions, and 4 people out of 6 whose "salary was too low to afford it", said they would have affiliated (if contributions were lower).

Respondents were also asked a question about preference among different social security programs. The wording and potential respondents for this question were very different in the two pilot surveys.

In Bulgaria, the question was: "Consider old age, survivors and disability insurance provided by the public social security administration. Suppose that you can only increase the benefits provided by one program but, for budgetary reasons, this increase must be compensated by a decrease in all other benefits. Which of the following benefits would you like to increase?" (B147). The possible answers were "Old age insurance", "Survivors insurance", "Disability insurance", "Sickness and maternity benefits", "Work injury benefits", "Unemployment benefits" or "None".

Potential respondents were people available for personal interview and who contributed to social security at least in the past, either currently working or not (278 individuals).

In Colombia, the question was: "A cuál de los siguientes programas considera Ud. que el gobierno debiera darle más recursos aunque disminuya el presupuesto de otros programas?" (C141). The possible answers were "Pensión de vejez", "Pensión por discapacidad", "Beneficios por enfermedad y maternidad", "Beneficios por accidente de trabajo", "Beneficios por desempleo" or "Ninguno". Potential respondents were employees and apprentices, whether contributing to social security or not (63 individuals).

For Bulgaria, there was only 1 nonrespondent, whereas 1 of the respondents answered DK. In Colombia, nonrespondents were 5, with no DK.

Table 32 presents the estimates of a logistic regression model where the binary outcome is equal to 1 if the individual would like to increase the old age pension benefit and 0 otherwise. In Bulgaria, the fraction of people with a preference for the old age insurance program is 65.0 percent, about the same for men and women. Preference for the old age insurance program is more likely among people aged 55–64 and 65+, and less likely among people aged 20–29. Gender, instead, is not statistically significant.

In Colombia, 58.6 percent of respondents prefers the old age insurance (56.2 percent for men and 61.5 percent for women). There is no clear-cut relationship between preference for the old age insurance program and standard socio-demographic variables.

6 What have we learned?

The results from the pilot surveys provide useful indications in order to improve both the questionnaire and the organization of the survey.

6.1 Improvements to the questionnaire

A few questions should be added on self-reported health status and age when the person first started working. In addition, specific questions might be included in order to identify seasonal workers and students or pensioners for whom work is not the main activity.

Life expectancy questions should be asked to everybody, whereas one should drop the question on what source (media, family history or on medical records) was used to assess life expectancy. It might be useful, however, to ask whether the parents of the respondent are still alive, if so how old they are, and if they are not alive, at what age they died.

All questions related to the social security system should avoid synonyms and use either the term "social security" or "social protection". Amounts used in the risk aversion and impatience questions should be calibrated to obtain useful brackets for the risk aversion parameter and the subjective discount rate. For risk aversion, it should be noted that the obtained measure (ARA) should be multiplied by wealth to obtain the relative risk aversion coefficient.[9] For the subjective discount rate, a possible benchmark is the market interest rate, which should fall in an intermediate bracket.

In some cases, rewording the question could help to reduce difficulties when asking about sensitive issues, such as subjective beliefs about mortality. In the pilot surveys, nonresponse rates were very low even for this question, but problems may be encountered in less developed countries, due to superstition or distrust of strangers. One possibility is to soften the question by avoiding reference to death, and focusing on the probability of being in good health after a certain age. With this formulation, a low subjective probability could imply that the individual does not expect to live enough to receive a pension, but it could also mean that she would need health care and economic support when older. As a consequence, the overall effect on interest in social security programs could be not straightforward.

In countries with a large fraction of low educated individuals, with a large size of the informal sector, or with low media coverage of social security issues, knowledge of the system will be

[9] The relative risk aversion coefficient is often assumed to be between 1 and 2, but some studies provide higher estimates. An example is Barsky *et al.* (1997), where the coefficient is estimated to be between 3 and 4.

infrequent and questions on the social security rules could be affected by nonreponse and a high incidence of "Don't know" answers. In this case, it is preferable to ask a single very simple question, for example about minimum age requirements for the old age pension.

The set of questions on time preferences (Q106–Q108) and risk aversion (Q109–Q111) was asked using simple closed-form answers and follow-up questions. In theory, it would be better to ask a single open-ended question such as "What is the minimum amount for which you would sell the ticket?". This formulation would allow to construct a continuous variable for the subjective discount rate and for the risk aversion parameter. On the other hand, it is possible that the respondent would need more time to provide an answer. For questions on time preferences, it is important to avoid any wording that could introduce the issue of risk. As a consequence, referring to the winning ticket of a lottery might not be the best choice if the lottery administration is not considered reliable.

The part on willingness to pay for social security programs was designed after considering several alternative formulations. In a preliminary version of the questionnaire, willingness to pay was assessed through a payment card method. The respondent was asked the maximum percent increase in contributions she would have paid in order to have a certain percent increase in the pension benefit. The question was then repeated for different percent increases in the benefit. The hypothetical increase in benefits is however very difficult to determine, because one should know the baseline benefit the respondent has in mind. This approach is rather complicated and time-consuming, so a second version of the question was formulated using double-bounded dichotomous choice methods, i.e. a set of yes/no questions with follow-up questions. The wording also specified that the hypothetical increase in contributions would apply to the entire working life of the respondent, to avoid a bias for older respondents. If the actual replacement rate was y, the question asked: "Consider all the contributions you paid for public old age pension and those you will pay in your future working life: Had you had the choice, would you have paid x% higher contributions (for all of your working life) in order to get benefits that are $y(1 + x/100)$% of the salary instead of y%? (Suppose you retire at the minimum age with 30 years of contributions)". Then the respondent was asked if she would otherwise be willing to pay a certain percentage, higher (lower) than x if she answered yes (no). Even with this formulation, the question is too long and complicated for a multi-purpose household survey. A simplified version (for example Q126 for the old-age pension) was therefore used for the pilots, simply asking about willingness to pay higher (lower) contributions throughout one's working life and stressing the positive (negative) consequences implied for future benefits. Of

course, the information that it is possible to get from this question is very different with respect to preliminary versions, and a more quantitative approach might be desirable in a stand-alone survey.

6.2 Improvements of survey organization

The training of interviewers is particularly important. In particular, interviewers should be introduced in details to the current rules of the social security system. In addition, they should be asked to fill in a special module describing the perceived characteristics of the respondents and the interview.

7 Summary and conclusions

Nonresponse rates were generally very low (below 5 percent), even for questions on life duration and subjective probabilities. The only notable exception is the question on expected replacement rates in Bulgaria. Nonresponse or DK answers are always associated with low educational attainments, while the role of age and gender depends on the type of question.

The pilot surveys were carried out in two countries that share some similarities but also have important differences in economic structure and institutional setting. Therefore, the finding of a similar behavior of respondents in the two samples is particularly interesting. For example, in both samples being younger, being older, or having a lower educational attainment is associated with a higher probability of not paying contributions. As for the opinions on the social security system, the same low fraction of respondents (roughly 20 percent) reports to prefer a PAYG to a fully-funded pension system. On the other hand, higher educated people are less likely to say that the government should be responsible for social security. In both samples, the old age pension is the most important social security program for the majority of respondents. The fraction of respondents who think that employment risks are unlikely is about 67 percent in both samples.

As far as health risks are concerned, not surprisingly older workers but also those with lower levels of educational attainments are less optimistic. Finally, there are some similarities in subjective beliefs about retirement. For example, even if Bulgarian respondents provide subjective probabilities of working after a certain age, while Colombian respondents are asked their expected retirement age, the same positive effect of male gender, higher education and age 45–54 is found.

Different response patterns to questions on job satisfaction and opinions on the social security system may reflect differences in the economic structure and the institutional setting of the countries. In the Bulgarian sample, the fraction of respondents who are not paying contributions

towards the old age pension is 15.4 percent, whereas in Colombia it is much higher and equal to 46.0 percent. Therefore, as it might be expected, only 23.4 percent of workers in the Colombian sample report to be satisfied with the social security benefits of their main job, while the fraction is 65.7 percent in the Bulgarian sample. The percentage of workers who report to be overall satisfied with their main job is much more similar in the two samples, and even higher in Colombia than in Bulgaria (76.4 versus 68.5 percent). Another difference is that overall job satisfaction in Bulgaria is less likely among young workers, while in Colombia it is less likely among middle-aged workers. Self-employment rates are much higher in Colombia than in Bulgaria (50.4 versus 19.3 percent). On the other hand, self-employment is associated with higher earnings in the Bulgarian sample, while there is no statistically significant relationship in the Colombian sample.

The fraction of optimists about future quality of life (both 5 years from now and after retirement) and about health risks is higher in Colombia. However, optimism about the future generosity of the social security system is less common in the Colombian than the Bulgarian sample (13.9 versus 37.1 percent). With respect to the old age pension, the average subjective replacement rate is higher in the Colombian than the Bulgarian sample. Only in the former, however, expectations are (positively) related to age and years of contributions.

Finally, most respondents in the Colombian sample report to be willing to pay higher contributions in order to have a higher pension benefit (91.4 percent), whereas the fraction is only 56.1 percent in the Bulgarian sample. On the other hand, while 63.2 percent of the Colombian respondents agree with the "economic argument" for paying contributions ("Paying SS contributions today is a good deal"), the fraction is only 36.9 percent in the Bulgarian sample.

References

Allingham, M.G., and Sandmo, A. (1972), "Income tax evasion: a theoretical analysis", *Journal of Public Economics*, 1, 323–338.

Attanasio, O., Goldberg, P.K., and Pavcnik, N. (2003), "Trade reforms and wage inequality in Colombia", NBER Working Paper No. 9830.

Baldry, J.C. (1979), "Tax evasion and labor supply", *Economics Letters*, 3, 53–56.

Barsky, R.B., Kimball, M.S., Juster, F.T. and Shapiro, M.D. (1997), "Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement Survey", *Quarterly Journal of Economics*, 112, 537–579.

Benjamini, Y., and Maital, S. (1985), "Optimal tax evasion and optimal tax evasion policy", in W. Gartner and A. Wenig (eds.), *The economics of the shadow economy*, Springer, Berlin, 245–264.

Boeri, T., Boersch-Supan, A., and Tabellini, G. (2001), "Would you like to shrink the welfare state? The opinions of European citizens", *Economic Policy*, 32, 9-50.

Borjas, G.J., and Bronars, S.G. (1989), "Consumer discrimination and self-employment", *Journal of Political Economy*, 97, 581–605.

Chavdarova, T. (2003), "Transactions versus work: The informal economy in Bulgaria in the 1990s", in B. Belev (ed.) *The Informal Economy in the EU Accession Countries: Size, Scope, Trends and Challenges to the Process of EU Enlargement*, Center for the Study of Democracy, Sofia, Bulgaria.

Cowell, F.A., and Gordon, J.P.F. (1988), "Unwillingness to pay", *Journal of Public Economics*, 36, 305–321.

Destré, G., and Henrard, V. (2004), "The determinants of occupational choice in Colombia: An empirical analysis", mimeo, TEAM, Université de Paris I and CNRS.

Evans, D.S., and Jovanovic, B. (1989), "An estimated model of entrepreneurial choice under liquidity constraints", *Journal of Political Economy*, 97, 808–27.

Fortin, B., Marceau, N., and Savard, L. (1997), "Taxation, wage controls and the informal sector", *Journal of Public Economics*, 66, 293–312.

Guiso, L., and Paiella, M. (2001), "Risk aversion, wealth and background risk", CEPR Discussion Papers No. 2728.

Harrison, G.W., Lau, M.I., and Williams, M.B. (2002), "Estimating individual discount rates in Denmark: A field experiment", *American Economic Review*, 92(5), 1606–1617.

Hassan, F.M.A., and Peters, R.K. (1995), "Social safety net and the poor during the transition: The case of Bulgaria", Policy Research Working Paper No. 1450, The World Bank, Washington DC.

Hurd, M., and McGarry, K. (2002), "The predictive validity of subjective probabilities of survival", *Economic Journal*, 112, 966–985.

Hussmanns, R. (2005), "Measuring the informal economy: from employment in the informal sector to informal employment", Policy Integration Department Working Papers No. 53, ILO.

Kihlstrom, R.E., and Laffont, J.J. (1979), "A general equilibrium entrepreneurial theory of firm formation based on risk aversion", *Journal of Political Economy*, 87, 719–748.

Kleinjans, K. (2003), "The impact of political risk on pension program choice: Evidence from Colombia", mimeo, Department of Economics, University of Pittsburgh.

Lucas, R.E. (1978), "On the size distribution of business firms", *Bell Journal of Economics*, 9, 508–523.

Magnac, T. (1991), "Segmented or competitive labor markets?", *Econometrica*, 59, 165–187.

Manski, C.F. (2004), "Measuring expectations", *Econometrica*, 72, 1329–1376.

Marc, A., and Kudatgobilik, Z. (2003), "Poverty and informality in South East Europe", in B. Belev (ed.) *The Informal Economy in the EU Accession Countries: Size, Scope, Trends and Challenges to the Process of EU Enlargement*, Center for the Study of Democracy, Sofia, Bulgaria.

Pencavel, J.H. (1979), "A note on income tax evasion, labor supply, and nonlinear tax schedules", *Journal of Public Economics*, 12, 115–124.

Rauch, J.E. (1991), "Modelling the informal sector formally", *Journal of Development Economics*, 35, 33–47.

Reinganum, J.F., and Wilde, L.L. (1985), "Income tax compliance in a principal-agent framework", *Journal of Public Economics*, 26, 1–18.

Roy, A. (1951), "Some thoughts on the distribution of earnings", *Oxford Economic Papers*, 3, 135–146.

Schneider, F. (2005), "Shadow economies around the world: what do we really know?", *European Journal of Political Economy*, 21, 598–642.

Schneider, F., and Enste, D.H. (2000), "Shadow economies: Size, causes, and consequences", *Journal of Economic Literature*, 38, 77–114.

Srinivasan, T.N. (1973), "Tax evasion: A model", *Journal of Public Economics*, 2, 339–346.

Table 1: Distribution of the sample by gender and age group.

	Bulgaria			Colombia		
Age group	Men	Women	Total	Men	Women	Total
15–19	20	20	40	4	8	12
20–29	38	32	70	16	19	35
30–44	45	68	113	26	36	62
45–54	48	60	108	15	24	39
55–64	28	30	58	10	11	21
65+	16	23	39	4	4	8
Total	195	233	428	75	102	177

Table 2: Age structure of the sample and population age structure from the U.N. Population Division data for 2005 (UN 2005).

	Sample			UN 2005		
Age group	Men	Women	Total	Men	Women	Total
Bulgaria						
15–19	10.3	8.6	9.3	8.3	7.3	7.8
20–29	19.5	13.7	16.4	18.2	16.0	17.0
30–44	23.1	29.2	26.4	25.4	23.1	24.2
45–54	24.6	25.8	25.2	17.0	16.6	16.8
55–64	14.4	12.9	13.6	14.4	15.1	14.8
65+	8.2	9.9	9.1	16.8	22.0	19.5
Total	100.0	100.0	100.0	100.0	100.0	100.0
Colombia						
15–19	5.3	7.8	6.8	14.3	13.1	13.7
20–29	21.3	18.6	19.8	25.8	24.2	25.0
30–44	34.7	35.3	35.0	31.1	31.2	31.2
45–54	20.0	23.5	22.0	14.1	14.7	14.4
55–64	13.3	10.8	11.9	8.1	8.6	8.4
65+	5.3	3.9	4.5	6.6	8.1	7.4
Total	100.0	100.0	100.0	100.0	100.0	100.0

Table 3: Distribution of the sample by gender, age group and schooling attainments (I: primary, II: secondary, III: tertiary).

	Men				Women				Total			
Age group	I	II	III	Total	I	II	III	Total	I	II	III	Total
Bulgaria												
15–19	14	6	0	20	15	5	0	20	29	11	0	40
20–29	5	22	11	38	3	21	8	32	8	43	19	70
30–44	7	27	11	45	4	32	32	68	11	59	43	113
45–54	6	34	8	48	5	39	16	60	11	73	24	108
55–64	2	12	14	28	3	18	9	30	5	30	23	58
65+	5	8	2	15	12	8	3	23	17	16	5	38
Total	39	109	46	194	42	123	68	233	81	232	114	427
Colombia												
15–19	2	2	0	4	4	4	0	8	6	6	0	12
20–29	1	11	4	16	5	10	4	19	6	21	8	35
30–44	3	12	11	26	6	14	16	36	9	26	27	62
45–54	4	5	6	15	7	11	6	24	11	16	12	39
55–64	1	5	4	10	6	5	0	11	7	10	4	21
65+	3	1	0	4	3	1	0	4	6	2	0	8
Total	14	36	25	75	31	45	26	102	45	81	51	177

Table 4: Distribution of the sample by gender, age group and labor force status.

	Men				Women				Total			
Age group	Empl	Unemp	Out	Total	Empl	Unemp	Out	Total	Empl	Unemp	Out	Total
Bulgaria												
15–19	0	2	17	19	4	2	14	20	4	4	31	39
20–29	27	5	6	38	18	7	7	32	45	12	13	70
30–44	38	4	3	45	60	5	3	68	98	9	6	113
45–54	43	4	1	48	53	6	1	60	96	10	2	108
55–64	20	0	8	28	27	1	2	30	47	1	10	58
65+	3	0	13	16	8	0	15	23	11	0	28	39
Total	131	15	48	194	170	21	42	233	301	36	90	427
Colombia												
15–19	0	0	4	4	1	0	7	8	1	0	11	12
20–29	13	2	1	16	16	0	3	19	29	2	4	35
30–44	25	1	0	26	25	1	10	36	50	2	10	62
45–54	12	1	2	15	18	0	6	24	30	1	8	39
55–64	9	0	1	10	7	0	4	11	16	0	5	21
65+	1	0	3	4	0	0	4	4	1	0	7	8
Total	60	4	11	75	67	1	34	102	127	5	45	177

Table 5: Estimated logistic regressions for labor force status (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Bulgaria			Colombia		
Variables	Active	Empl	Unemp	Active	Empl	Unemp
Female	0.670 **	0.276	0.233	-1.508 **	-0.888 **	-1.619
Age 15–19	-3.849 **	-3.595 **	1.899 **	-4.268 **	-3.833 **	
Age 20–29	-1.281 **	-1.207 **	1.049 **	0.361	0.098	0.660
Age 45–54	1.250	0.353	0.002	-0.224	-0.152	-0.133
Age 55–64	-1.381 **	-0.458	-1.466	-0.520	-0.248	
Age 65+	-3.714 **	-2.599 **		-3.889 **	-3.358 **	
Lower education	-0.214	-0.280	0.468	-0.600	-0.624	0.660
Higher education	1.178 **	1.090 **	-0.944 *	-0.246	-0.263	0.448
Constant	2.275 **	1.450 **	-2.308 **	2.918 **	2.224 **	-2.986 **
No. obs.	427	428	326	177	177	114
Log-likelihood	-130.6	-185.0	-100.9	-71.4	-82.1	-19.0
Pseudo R^2	0.406	0.289	0.109	0.288	0.221	0.074
χ^2 (age)	76.81 **	69.70 **	14.06 **	22.58 **	19.43 **	0.53
χ^2 (educ)	7.09 **	11.00 **	4.95 *	1.36	1.67	0.31

Table 6: Distribution of the sample by gender, age group and type of employment (C denotes contribution to a pension fund, NC denotes lack of contribution to a pension fund).

	Men					Women					Total				
	Self		Empl		Tot	Self		Empl		Tot	Self		Empl		Tot
Age	C	NC	C	NC		C	NC	C	NC		C	NC	C	NC	
Bulgaria															
15–29	0.0	4.0	80.0	16.0	100.0	14.3	9.5	57.1	19.0	100.0	6.5	6.5	69.6	17.4	100.0
30–54	25.3	3.8	63.3	7.6	100.0	10.0	4.5	82.7	2.7	100.0	16.4	4.2	74.6	4.8	100.0
55+	0.0	13.0	82.6	4.3	100.0	5.7	20.0	57.1	17.1	100.0	3.4	17.2	67.2	12.1	100.0
Total	15.7	5.5	70.1	8.7	100.0	9.6	8.4	74.1	7.8	100.0	12.3	7.2	72.4	8.2	100.0
Colombia															
15–29	7.7	23.1	38.5	30.8	100.0	29.4	23.5	29.4	17.6	100.0	20.0	23.3	33.3	23.3	100.0
30–54	8.1	32.4	51.4	8.1	100.0	16.7	38.1	35.7	9.5	100.0	12.7	35.4	43.0	8.9	100.0
55+	30.0	30.0	40.0	0.0	100.0	0.0	85.7	14.3	0.0	100.0	17.6	52.9	29.4	0.0	100.0
Total	11.7	30.0	46.7	11.7	100.0	18.2	39.4	31.8	10.6	100.0	15.1	34.9	38.9	11.1	100.0

Table 7: Estimated logistic regressions for informal employment (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Bulgaria			Colombia		
Variables	Self	Nocontrib	Second	Self	Nocontrib	Second
Female	-0.257	0.143	0.386	0.757 **	0.173	-0.397
Age 15–19	1.000	1.434				
Age 20–29	-1.076 *	1.312 **	-0.491	-0.091	-0.085	-1.341
Age 45–54	-0.093	0.368	-0.038	0.357	0.613	0.152
Age 55–64	-0.072	1.269 **	0.006	1.029	0.089	
Age 65+	0.334	3.500 **	-0.720			
Lower education	0.381	1.929 **	0.291	0.334	1.246 **	-0.019
Higher education	-0.610 *	-0.469	0.420	-0.193	-1.433 **	-0.083
Constant	-1.030 **	-2.813 **	-2.181 **	-0.574	-0.221	-1.770 **
No. obs.	301	293	297	125	124	109
Log-likelihood	-141.6	-100.5	-115.3	-82.0	-72.4	-34.3
Pseudo R^2	0.040	0.201	0.017	0.054	0.152	0.038
χ^2 (age)	4.89	20.91 **	1.11	3.40	1.71	1.73
χ^2 (educ)	4.46	19.89 **	1.32	0.89	17.43 **	0.01

Table 8: Estimated logistic regressions for informal employment of employees (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Bulgaria		Colombia	
Variables	Nocontract	Smallf	Nocontract	Smallf
Female	-1.037	0.810 *	-0.154	0.766
Age 20–29	0.795	1.994 **	2.247 *	-0.012
Age 45–54	-0.739	-0.329	2.195 *	-0.111
Age 55–64	0.667	1.061 *	2.792 *	0.171
Age 65+		0.377		
Lower education	3.172 **	0.826	1.309	0.933
Higher education		-0.056		-1.933 *
Constant	-3.382 **	-2.935 **	-2.754 **	-1.536 **
No. obs.	140	199	39	62
Log-likelihood	-23.1	-73.0	-20.4	-26.0
Pseudo R^2	0.309	0.117	0.179	0.146
χ^2 (age)	2.09	16.53 **	4.34	0.04
χ^2 (educ)	14.16 **	1.24	1.98	4.98 *

Table 9: Estimated log-earnings regressions (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Bulgaria			Colombia		
Variables	All	All	Employees	All	All	Employees
Female	-0.137 **	-0.093 *	-0.145 *	-0.157	-0.234 *	-0.093
Age	0.004	0.007 **	0.005	0.010	0.010 **	0.017 **
Age squared	-0.000	-0.000	0.000	0.000	0.000	0.001
Lower education	-0.360 **	-0.496 **	-0.504 **	-0.544 **	-0.643 **	-0.735 **
Higher education	0.226 **	0.268 **	0.248 **	1.238 **	1.080 **	1.037 **
Self-employed		0.211 **			-0.110	
Not contributing to OAI		0.269 **	0.037		-0.489 **	-0.324
Working without a contract			0.692 **			-0.126
Employee of a small firm			-0.320 **			-0.232
Constant	0.482 **	0.398 **	0.461 **	1.335 **	1.534 **	1.562 **
No. obs.	254	248	165	118	118	62
MAE	0.356	0.344	0.317	0.859	0.823	0.387
Pseudo R^2	0.102	0.130	0.142	0.185	0.219	0.519
χ^2 (age)	1.01	3.85 **	0.79	0.70	2.18	5.81 **
χ^2 (educ)	12.98 **	27.60 **	9.17 **	23.87 **	45.25 **	55.99 **

Table 10: Estimated logistic regressions for overall satisfaction with main job (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	0.025	-0.279
Age 20–29	-0.883 **	-0.535
Age 45–54	-0.210	-1.374 **
Age 55–64	-0.234	0.237
Age 65+	1.439	
Lower education	-0.646	-0.135
Higher education	-0.401	-0.193
Constant	1.211 **	1.909 **
No. obs.	236	125
Log-likelihood	-141.2	-64.3
Pseudo R^2	0.033	0.067
χ^2 (age)	6.67	7.80 *
χ^2 (educ)	2.73	0.15

Table 11: Estimated logistic regressions for satisfaction with social security benefits provided by the main job (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	0.235	-0.081
Age 20–29	-0.484	-0.736
Age 45–54	0.110	0.739
Age 55–64	-0.100	-1.189
Age 65+	-0.837	
Lower education	-1.096 **	-1.899
Higher education	-0.034	-0.866
Constant	0.699 *	-0.484
No. obs.	204	64
Log-likelihood	-127.2	-32.4
Pseudo R^2	0.030	0.071
χ^2 (age)	2.55	2.97
χ^2 (educ)	4.18	3.02

Table 12: Mean and percentiles of subjective survival probabilities by gender and age group, Bulgaria.

	Age 70				Age 80				Age 90			
Age group	Mean	q25	q50	q75	Mean	q25	q50	q75	Mean	q25	q50	q75
Men												
15-19	90.0	80	90	100	55.0	50	55	60	25.0	20	25	30
20-29	66.8	50	65	100	40.0	15	35	60	18.3	0	8	30
30-44	44.2	30	50	50	21.9	0	20	40	6.0	0	0	10
45-54	47.3	10	50	80	27.3	0	20	50	8.4	0	0	20
55-64	53.5	30	50	80	22.4	0	10	50	8.2	0	0	20
65+	70.0	10	100	100	58.2	50	50	100	22.7	0	10	20
Total	52.3	30	50	80	30.7	0	30	50	11.3	0	0	20
Women												
15-19	92.0	100	100	100	54.0	40	50	50	10.0	0	10	20
20-29	61.4	50	60	80	39.2	20	40	50	11.7	0	0	20
30-44	45.5	30	50	60	20.6	6	20	30	3.9	0	0	2
45-54	49.6	30	50	60	27.4	10	20	40	9.5	0	0	10
55-64	53.0	30	50	80	33.1	20	30	50	6.2	0	0	10
65+	71.4	40	80	100	49.3	20	50	60	17.6	0	10	20
Total	52.8	40	50	70	30.5	10	20	45	8.4	0	0	10
Total												
15-19	91.4	80	100	100	54.3	40	50	60	14.3	0	20	20
20-29	63.8	50	60	100	39.5	20	40	50	14.7	0	3	20
30-44	45.0	30	50	50	21.1	0	20	30	4.7	0	0	5
45-54	48.6	20	50	70	27.3	5	20	40	9.0	0	0	10
55-64	53.2	30	50	80	28.8	5	20	50	7.0	0	0	10
65+	71.0	40	90	100	53.2	40	50	60	19.8	0	10	20
Total	52.6	30	50	80	30.6	10	30	50	9.6	0	0	10

Table 13: Mean and percentiles of subjective life expectancy by gender and age group, Colombia.

Agegroup	Mean	q25	q50	q75
Men				
30-44	76.7	70	70	80
45-54	78.7	70	80	85
55-64	73.9	80	80	80
65+	88.8	83	88	95
Total	77.6	70	80	85
Women				
30-44	74.0	70	70	80
45-54	77.2	70	80	80
55-64	75.0	70	75	80
65+	85.0	83	85	88
Total	75.7	70	75	80
Total				
30-44	75.1	70	70	80
45-54	77.8	70	80	80
55-64	74.4	70	80	80
65+	86.9	83	85	90
Total	76.5	70	80	80

Table 14: Estimated logistic regressions for optimism about life duration (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Bulgaria			Colombia
Variables	Age 70	Age 80	Age 90	
Female	-0.106	-0.115	-0.737	-0.577
Age 15–19		2.291 **		
Age 20–29	1.354 **	1.386 **	16.799 **	
Age 45–54	0.638 *	0.705	14.960	-0.184
Age 55–64	0.853 **	0.904		-0.946
Age 65+	1.753 **	2.579 **	17.270 **	2.196 **
Lower education	-0.361	-0.774	-0.376	1.371 **
Higher education	-0.089	0.261	1.010	0.814
Constant	-1.090 **	-2.716 **	-19.228 **	-1.738 **
No. obs.	269	284	238	128
Log-likelihood	-168.3	-106.4	-26.9	-56.5
Pseudo R^2	0.046	0.092	0.230	0.143
χ^2 (age)	14.81 **	19.88 **	241.81 **	8.11 **
χ^2 (educ)	0.72	2.67	2.03	4.81 *

Table 15: Mean and percentiles of subjective full-time work probabilities by gender and age group, Bulgaria.

	Age 50				Age 60				Age 70			
Age group	Mean	q25	q50	q75	Mean	q25	q50	q75	Mean	q25	q50	q75
Men												
15-19	100.0	100	100	100	75.0	50	75	100	20.0	10	20	30
20-29	81.0	65	100	100	60.5	40	55	95	11.5	0	0	20
30-44	84.0	80	95	100	55.7	30	50	70	14.5	0	3	20
45-54	85.2	100	100	100	68.3	50	70	100	14.2	0	0	20
55-64	.	.	.	.	63.3	40	70	100	5.3	0	0	0
65+	.	.	.	.	.	.	.	.	0.0	0	0	0
Total	84.0	80	100	100	62.4	50	65	100	12.1	0	0	20
Women												
15-19	100.0	100	100	100	40.0	30	50	50	0.0	0	0	0
20-29	83.5	80	100	100	48.1	20	50	80	11.0	0	0	20
30-44	80.0	50	100	100	48.6	20	50	80	6.7	0	0	10
45-54	91.3	100	100	100	58.0	30	55	90	6.3	0	0	0
55-64	.	.	.	.	46.7	0	40	100	8.7	0	0	10
65+	.	.	.	.	.	.	.	.	10.0	0	0	10
Total	84.2	80	100	100	51.2	20	50	80	7.5	0	0	5
Total												
15-19	100.0	100	100	100	50.0	30	50	70	5.7	0	0	10
20-29	82.4	80	100	100	53.5	20	50	80	11.2	0	0	20
30-44	81.4	65	100	100	51.1	25	50	75	9.5	0	0	10
45-54	88.4	100	100	100	62.2	40	70	100	9.6	0	0	10
55-64	.	.	.	.	52.9	5	55	100	7.4	0	0	3
65+	.	.	.	.	.	.	.	.	7.5	0	0	5
Total	84.1	80	100	100	55.5	30	50	90	9.3	0	0	10

Table 16: Mean and percentiles of expected retirement age by gender and age group, Colombia.

Agegroup	Mean	q25	q50	q75
Men				
30-44	60.8	58	60	65
45-54	64.1	60	65	70
55-64	60.1	65	65	70
65+	80.0	80	80	80
Total	62.0	60	65	70
Women				
30-44	57.8	50	60	60
45-54	63.4	58	60	65
55-64	64.0	60	61	65
Total	60.6	56	60	61
Total				
30-44	59.3	55	60	60
45-54	63.7	60	60	70
55-64	61.8	60	65	69
65+	80.0	80	80	80
Total	61.3	58	60	65

Table 17: Estimated logistic regressions for optimism about working life duration (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Bulgaria			Colombia
Variables	Age 50	Age 60	Age 70	
Female	-0.296	-0.540 *	-2.176 **	-1.175 **
Age 15–19		0.147		
Age 20–29	-0.152	0.246	0.059	
Age 45–54	0.590	0.584 *	2.394 **	1.681 **
Age 55–64		0.302		2.930 **
Lower education	-1.057 **	-1.609 **		-1.492 *
Higher education	0.041	0.182	2.261 **	1.268 **
Constant	1.633 **	0.093	-4.685 **	-1.219 **
No. obs.	175	246	188	94
Log-likelihood	-84.4	-159.8	-25.1	-47.3
Pseudo R^2	0.036	0.062	0.241	0.243
χ^2 (age)	1.78	3.25	6.06 **	14.46 **
χ^2 (educ)	4.48	11.86 **	6.16 **	10.58 **

Table 18: Mean and percentiles of expected replacement rates by gender and age group, Bulgaria.

Age group	Mean	q25	q50	q75
Men				
30-44	60.0	40	50	70
45-54	55.2	40	50	70
55-64	55.0	45	50	60
65+	60.0	60	60	60
Total	56.0	40	50	70
Women				
30-44	54.4	40	50	60
45-54	65.4	50	60	80
55-64	62.0	40	60	80
65+	50.0	50	50	50
Total	62.0	48	60	70
Total				
30-44	55.7	40	50	70
45-54	61.8	47	55	73
55-64	59.8	40	60	80
65+	55.0	50	55	60
Total	60.0	45	50	70

Table 19: Estimated regressions for expected replacement rate (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent; Pseudo R^2 is McKelvey and Zavoina's R^2).

Variables	Bulgaria		Colombia	
Female	3.642	3.280	1.691	1.679
Age	0.182		0.505 **	
Years of contribution		0.132		0.420 **
Lower education	-8.256	-6.967	-15.303 **	-12.076 **
Higher education	-6.856 *	-6.377	2.454	0.614
Constant	56.200 **	57.068 **	64.598 **	70.696 **
No. obs.	109	107	68	68
Log-likelihood	-473.0	-464.8	-125.8	-126.2
R^2	0.039	0.034		
Pseudo R^2			0.156	0.147
χ^2 (educ)	1.72	1.41	9.11 **	4.8 *

Table 20: Estimated logistic regressions for optimism about next 5 years quality of life (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.659 **	0.613
Age 15–19	2.420 **	
Age 20–29	1.235 **	
Age 45–54	-0.285	0.110
Age 55–64	-0.812 **	-0.909 *
Age 65+	-2.479 **	1.534
Lower education	-0.379	-0.883 *
Higher education	-0.150	-0.354
Constant	0.303	0.853 *
No. obs.	302	130
Log-likelihood	-177.1	-75.7
Pseudo R^2	0.139	0.066
χ^2 (age)	36.12 **	5.91
χ^2 (educ)	0.95	2.96

Table 21: Estimated logistic regressions for optimism about post-retirement quality of life (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.628 *	-0.000
Age 20–29	0.890 *	
Age 45–54	-0.457	-0.199
Age 55–64	-1.161	-1.003
Lower education	-1.185	-0.344
Higher education	-0.207	-0.471
Constant	-0.861 **	0.662
No. obs.	213	95
Log-likelihood	-96.6	-63.7
Pseudo R^2	0.074	0.027
χ^2 (age)	10.31 **	2.71
χ^2 (educ)	2.24	1.03

Table 22: Estimated logistic regressions for optimism about employment risks (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia A	Colombia B
Female	0.122	-0.375	-1.252 *
Age 20–29	-0.643		
Age 45–54	0.106	0.910	0.520
Age 55–64	-0.038		-0.566
Age 65+	-1.390 *		
Lower education	-0.350	-1.130	0.655
Higher education	0.262	-0.326	0.325
Constant	0.703 *	0.706	1.022
No. obs.	184	39	48
Log-likelihood	-114.0	-24.8	-28.1
Pseudo R^2	0.031	0.045	0.080
χ^2 (age)	5.58	1.35	1.25
χ^2 (educ)	1.39	1.15	0.55

Table 23: Estimated logistic regressions for optimism about health risks (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

	Illness			Accident		
Variables	Bulgaria	Colombia A	Colombia B	Bulgaria	Colombia A	Colombia B
Female	-0.001	-0.466	0.200	0.399	0.067	-0.404
Age 15–19	1.088			1.112		
Age 20–29	1.850 **			1.048 **		
Age 45–54	0.880 **	-0.846	-0.284	0.812 **	0.171	-0.743
Age 55–64	-0.069		-0.096	0.719	0.995	-0.728
Age 65+	-0.774 *	-4.046 **		-0.056	-3.346 **	
Lower education	-0.528	2.587 **		-0.855 **	2.691 **	1.247
Higher education	-0.371	1.665 **	-0.999	-0.435	1.270 *	-0.585
Constant	1.030 **	0.466	2.029 **	0.952 **	-0.460	2.079 **
No. obs.	300	53	48	298	64	61
Log-likelihood	-150.3	-28.2	-22.3	-142.5	-34.7	-28.5
Pseudo R^2	0.091	0.183	0.038	0.056	0.180	0.058
χ^2 (age)	21.70 **	5.86 *	0.11	8.93	6.38 *	1.07
χ^2 (educ)	2.34	7.33 **	1.45	4.90 *	7.80 **	2.31

Table 24: Estimated logistic regressions for impatience (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.016	0.274
Age 15–19	-0.168	
Age 20–29	0.219	
Age 45–54	0.198	-0.664
Age 55–64	-0.102	0.284
Age 65+	0.148	0.309
Lower education	0.848 **	-1.169 **
Higher education	-0.127	-0.595
Constant	0.180	1.599 **
No. obs.	303	125
Log-likelihood	-201.6	-67.2
Pseudo R^2	0.020	0.055
χ^2 (age)	1.11	3.12
χ^2 (educ)	5.83 *	4.56

Table 25: Estimated logistic regressions for risk aversion (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.163	0.685
Age 15–19	0.614	
Age 20–29	-0.470	
Age 45–54	-0.639 *	0.769
Age 55–64	0.142	-0.709
Age 65+	-0.976	1.689 *
Lower education	-0.596	0.896
Higher education	-0.034	0.846
Constant	-0.640 **	-3.250 **
No. obs.	303	128
Log-likelihood	-164.2	-45.2
Pseudo R^2	0.032	0.099
χ^2 (age)	8.12	5.04
χ^2 (educ)	1.75	1.98

Table 26: Estimated logistic regressions for government responsibility for social security (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria (OAI)	Bulgaria (SDI)	Colombia
Female	0.216	-0.461	-0.083
Age 15–19	-2.372 **	-1.058	0.220
Age 20–29	0.092		-0.346
Age 45–54	-0.073	0.219	0.003
Age 55–64	-0.071	0.808	0.551
Lower education	1.752 *	0.969	-0.281
Higher education	-0.720 **	1.510	-0.715 *
Constant	1.699 **	2.832 **	0.850 **
No. obs.	272	226	162
Log-likelihood	-116.4	-38.6	-104.4
Pseudo R^2	0.058	0.059	0.027
χ^2 (age)	6.17	1.53	2.15
χ^2 (educ)	8.99 **	2.44	3.34

Table 27: Estimated logistic regressions for preference for PAYG pension scheme (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	0.026	-0.049
Age 15–19	-0.484	0.499
Age 20–29	-0.365	0.553
Age 45–54	0.297	0.248
Age 55–64	0.200	1.262 **
Age 65+	0.818 *	0.801
Lower education	0.337	-0.122
Higher education	0.145	-0.140
Constant	-1.529 **	-1.784 **
No. obs.	292	169
Log-likelihood	-151.6	-79.5
Pseudo R^2	0.021	0.030
χ^2 (age)	5.19	4.56
χ^2 (educ)	0.72	0.11

Table 28: Estimated logistic regressions for agreement with solidarity argument in paying social security contributions (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	0.869	-1.409 *
Age 20–29	-0.635	-0.480
Age 45–54	0.813	0.935
Age 55–64	0.695	-0.318
Age 65+		-4.009 **
Lower education	-0.949	1.849
Higher education	0.091	-1.078
Constant	2.702 **	3.933 **
No. obs.	262	161
Log-likelihood	-41.9	-36.6
Pseudo R^2	0.082	0.189
χ^2 (age)	3.11	10.03 **
χ^2 (educ)	1.79	4.80 *

Table 29: Estimated logistic regressions for agreement with economic argument in paying social security contributions (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.023	0.369
Age 15–19	1.046	-0.532
Age 20–29	-1.383 **	0.267
Age 45–54	-0.020	0.337
Age 55–64	0.216	0.972
Age 65+	0.795 *	-0.526
Lower education	-0.641	0.886 *
Higher education	-0.141	-0.756 **
Constant	-0.367	0.206
No. obs.	298	171
Log-likelihood	-185.9	-103.6
Pseudo R^2	0.052	0.079
χ^2 (age)	15.81 **	4.31
χ^2 (educ)	2.61	9.89 **

Table 30: Estimated logistic regressions for optimism about future generosity of the social security system (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.345	-0.882 *
Age 15–19	1.095	-0.099
Age 20–29	0.258	0.391
Age 45–54	-0.341	0.159
Age 55–64	-0.242	-0.093
Age 65+	0.288	
Lower education	-0.386	1.180 **
Higher education	0.044	-0.884
Constant	-0.243	-1.591 **
No. obs.	288	164
Log-likelihood	-186.3	-62.0
Pseudo R^2	0.019	0.092
χ^2 (age)	5.10	0.60
χ^2 (educ)	1.06	9.04 **

Table 31: Estimated logistic regressions for willingness to pay higher or lower contributions, Bulgaria (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Higher	Lower
Female	0.259	-0.647
Age 20–29	-0.33	1.019 *
Age 45–54	0.221	0.294
Age 55–64	0.086	-0.333
Age 65+	1.551 **	
Lower education	-1.051 **	-0.943
Higher education	0.048	-0.645
Constant	0.013	-0.18
No. obs.	278	115
Log-likelihood	-182.4	-69.6
Pseudo R^2	0.043	0.063
χ^2 (age)	10.78 **	4.27
χ^2 (educ)	6.51 **	3.18

Table 32: Estimated logistic regressions for preference for the old age insurance (* denotes asymptotic p-values between 5 and 10 percent, ** denotes asymptotic p-values below 5 percent).

Variables	Bulgaria	Colombia
Female	-0.124	0.183
Age 20–29	-0.783 *	-0.949
Age 45–54	0.361	0.111
Age 55–64	1.220 **	1.201
Age 65+	2.175 **	
Lower education	-0.434	
Higher education	0.414	-0.414
Constant	0.265	0.243
No. obs.	277	49
Log-likelihood	-163.6	-32.0
Pseudo R^2	0.088	0.058
χ^2 (age)	24.49 **	3.26
χ^2 (educ)	3.53	0.41

A The basic questionnaire

B Main socio-economic indicators for Bulgaria and Colombia

Table 33: Main socio-economic indicators

	Bulgaria	Colombia
Economy		
GDP per capita (current US$)[1]	2351.44	2101.00
GDP per capita (PPP US$)[1]	7840.03	6974.85
GDP growth (average annual growth 1990-2000)	-1.8	2.8
GDP growth (average annual growth 2000-2004)	4.8	2.9
Labor		
Labor force participation rate, female (% ages 15–64)	53.3	65.0
Labor force participation rate, male (% ages 15-64)	63.0	85.3
Unemployment (% of total labor force)	13.7	14.2
Employment in agriculture, female (% of total employment)	8.0	8.0
Employment in industry, female (% of total employment)	29.0	17.0
Employment in services, female (% of total employment)	64.0	75.0
Employment in agriculture, male (% of total employment)	12.0	31.0
Employment in industry, male (% of total employment)	37.0	21.0
Employment in services, male (% of total employment)	51.0	49.0
Shadow economy (% of GDP, average 1990/1991)[2]	29.4	33.4
Shadow economy (% of GDP, average 1994/1995)[2]	33.2	36.2
Shadow economy (% of GDP, average 1999/2000)[2]	36.9	39.1
Self-employment (% of total employment)[3]	14.9	50.9
Tax policies		
Tax revenue collected by central government (% of GDP)	22.3	13.8
Highest marginal tax rate (Individual %)	29.0	35.0
Highest marginal tax rate (Corporate %)	20.0	37.0
Social security		
Pension contributors (% of labor force)	64.0	20.7
Public expenditure on pensions (% of GDP)	8.9	1.1
Average pension (% of per capita income)	39.3	72.2
Contribution towards OASDI, if social insurance only (% of earnings)[4]	23.0	15.5
Contribution towards OASDI, if individual account (% of earnings)[4]	23.0	18.5
Minimum age for OA social insurance pension, men (2006)[4]	63.0	60.0
Minimum age for OA social insurance pension, women (2006)[4]	58.5	55.0
Minimum years of contributions for OA social insurance pension, men (minimum age)[4]	37.0	20.2
Minimum years of contributions for OA social insurance pension, women (minimum age)[4]	33.5	20.2
OA social insurance benefit with minimum age and contribution, men (% of earnings)[4]	37.0	80.0
OA social insurance benefit with minimum age and contribution, women (% of earnings)[4]	33.5	80.0

Source: World Development Indicators 2004; (1) IMF World Economic Outlook 2004; (2) Schneider (2005); (3) ILO LABORSTA 2004; (4) ISSA Social Security Worldwide (2006)

C Key features of the social security systems of Bulgaria and Colombia

The social security system has different features in Bulgaria and in Colombia. In Bulgaria, until 2002, the system was based on social insurance only and coverage was mandatory for all employed individuals. Starting in 2002, a second pillar was introduced. After this reform, an individual account program is also mandatory for workers born after December 31, 1959. In Colombia, insurance is mandatory for all workers, who can choose either the social insurance or individual account system. It is possible to switch program every three years.

In Bulgaria, contributions towards the old age pension are 23.0 percent of earnings for both employees and self-employed. In Colombia, they are 15.5 percent if the social insurance is chosen, and 18.5 percent if the individual account system is chosen.

For old-age pension eligibility in Bulgaria, the minimum age requirement is 63 years for men and 58 years and 6 months for women. The age limit for women is increasing by 6 months every year until 2009, when it will reach age 60. For the social insurance program, the sum of age and years of contributions must be at least 100 for men and 92 for women. The social insurance old-age pension benefit is 1 percent of taxable income for each year of insurance coverage.

In Colombia, the minimum age to receive the old-age pension is 60 for men and 55 for women. The requirement in terms of contributions for the social insurance benefit is 1050 weeks, although it is increasing by 25 weeks each year, up to 1300 in 2015. For the individual account program, the accumulated capital must be sufficient to purchase an annuity greater than 110 percent the minimum wage. The social insurance old-age pension benefit is equal to 55–65 percent of monthly wage, plus 1.5 percent for each 50-week contribution period, up to 80 percent. For the individual account system, the benefit depends on accrued interest.

In Bulgaria, sickness and maternity benefits are paid to eligible workers with at least 6 months of insurance coverage. In Colombia, the coverage must be of at least 4 weeks for sickness benefits and 9 months for maternity benefits.

For work injury benefits, there are no qualifying conditions both in Bulgaria and Colombia.

In Bulgaria, eligibility for the unemployment benefit requires at least 9 months of insurance coverage. The benefit is 60 percent of average earnings in the last 9 months, and duration of the benefit depends on the length of the coverage period, up to 12 months. In Colombia, the unemployment benefit is equal to 1 monthly wage for each year of employment.