Simultaneous causality between health status and employment status within the population aged 30-50 in France*

Thomas Barnay† François Legendre‡

november 2011

Economic literature clearly establishes the link between socio-economic status, good health and a high level of education. Health status also appears to be a determining factor in an individual's present and future preferences (Disney et al., 2006). The relationship between health status and employment status is the subject of numerous research studies and can be apprehended from the principle of double causality: self-selection bias (healthy worker effect) and causal effect (reverse causality) (Currie and Madrian, 1999). These two non-contradictory and potentially simultaneous working hypotheses are studied in this study.

1 Problematics

1.1 Healthy worker effect

The healthy worker effect hypothesis states that a deteriorated health status can have an impact on an individual's professional integration by reducing individual productivity and thereby earning capacity on the labour market, or by increasing medical expenditures thereby reducing savings and personal patrimony. Poor health status is one of the factors contributing to changes in individuals' preferences for leisure and labour market participation. Numerous studies reveal that self-selection bias resulting from health-related attitudes can have an impact on socio-economic status and more particularly, employment status. This literature gives rise to numerous methodological debates as biased self-reports and health status measurements can bias error results. Self-reporting

*This study was funded through a call for projects of French Ministry of Health and Labour (DREES-DARES).

†Érudite (Univ. Paris-Est) and Tepp (FR CNRS 3126), Barnay@u-pec.fr.

‡Érudite (Univ. Paris-Est) and Tepp (FR CNRS 3126), F.Legendre@u-pec.fr.

can be influenced by individuals' social characteristics, their position on the labour market or health status resulting in socio-cultural biases and endogeneity (Bound, 1991). When the data allows it, it appears necessary to compile objective measurements (such as mortality) with subjective measurements since the exclusive use of self-reported health status measurements could introduce statistical artefacts into the relationship.

Studies investigating the financial determinants of early retirement from the labour market are abundant (Gruber and Wise, 1999; Blöndal and Scarpetta, 1998). Using European data on the population aged 50 and over, Blanchet and Debrand (2007) however show that financial incentives although measured in an imperfect way, have a significant impact at individual level but that nevertheless seems less marked than that of the non-financial variables'. In an exhaustive review of the literature, Lindeboom (2006) clearly establishes the relationship between a deteriorated health status and early retirement from the labour market (Bound et al., 1999; Kerkhofs et al., 1999; Campolieti, 2002). Most of these studies, however, focus exclusively on the population aged 50 and over and identify health status as a determining factor in the low employment rate of older workers. Studies involving this age group are rendered more complex by , possible age discrimination in the employment of older workers and the different schemes permitting early retirement from the labour market. Financial incentives, retirement opportunities or ex-post justification (Barnay, 2009) can bias the relationship between health and occupational status. Using French data, Barnay and Briard (2007) analysed the retirement behaviours of French Social Security pension scheme beneficiaries born in 1940 who had claimed their pension rights despite an insufficient insurance period. The authors show that early retirement, on a reduced-rate pension or because of incapacity or invalidity, is related to a deteriorated health status, and equally highlight the existence of cumulative effects and their interactions with labour market experiences, especially at the end of an individual's working life. Barnay and Jeger (2007) furthermore show that the relative lack of claims for invalidity pensions in France can be explained by the higher financial incentives offered by alternative schemes. In effect, the wage replacement rate offered by invalidity pensions (less than 50% without complementary insurance coverage) is often inferior to unemployment benefits replacement rates at the end of a working career (from 65 to 75%). Barnay (2008) finally advances the hypothesis that exemption from seeking employment, progressively being abolished since the summer of 2008, could benefit individuals in poor health looking for early retirement possibilities through the various systems available.

Besides the numerous specific studies dealing with the effects of health on the employment of older workers, it would appear that health status has no effect on professional mobility. Health problems do not increase the risk of decline in professional mobility (Van de Mheen et al., 1999). Its effect on income or salary, however, remains ambiguous. Currie and Madrian (1999), for example, conclude that health has no effect on salary whereas Lechene and Magnac (1994) demonstrate that health has a transitory effect on salary explained by stagnation in human capital accumulation. Similarly, Rioux (2001) shows that in France, RMI beneficiaries in better health seek employment more efficiently. The effect of health on the accumulation of personal assets is thus potential even though strongly related to ageing (Smith, 1998).

1.2 The reverse causality effect

The reverse causality hypothesis suggests that professional integration determines health status via a multitude of potential intermediary factors such as risk behaviours, access to healthcare, material life or early life conditions, and their cumulative effects.

Risk behaviours (tobacco, alcohol or drug consumption, sedentariness. . .) are determining factors in the early appearance of chronic illness. These are more widespread in disadvantaged environments (Stronks et al., 1997; Smith, 1999). Marmot (2000) thus shows that a third of the variance in mortality rates due to coronary disease can be attributed to tobacco consumption, high blood pressure, sedentariness or even an individual's size.

Access to healthcare could equally play a role in the causality relationship between employment status and health status. In France, the Statutory Health Insurance scheme covers the quasi totality of the population, only 6% remaining without complementary health insurance (CHI) coverage. Among the individuals without CHI coverage, 46% self-report financial reasons. Furthermore, coverage levels among CHI beneficiaries are extremely heterogeneous. The more generous company-provided CHI reinforces social health inequalities between the economically active and inactive populations but also within the working population: 76% of executives benefit from employer-provided CHI against only 58% of unskilled workers and 53% of commercial sector employees (Perronnin et al., 2011). The assumption here is that favourable socio-economic conditions result in access to better quality healthcare through access to better information and a higher degree of preventive behaviour.

Furthermore, a better professional integration improves income level and thereby increases the opportunities of investing in one's health through sport, a varied, healthy diet, better housing conditions and even better working conditions. By improving one's material living conditions, socio-economic status thus leads to a better health status. The literature has shown that an improvement in living conditions favours a better health status whether defined in terms of income (Ecob and Davey Smith, 1999; Soobader and Le Clere, 1999; Fiscella and Franks, 2000; Deaton and Paxton, 2001), employment status (Or, 2000; Mesrine, 2000), working conditions (Dyèvre and Léger, 1999; Askenazy, 2000; Dumartin, 2000), education level (Blampin and Eneau, 1999; van Rossum et al., 2000; Everson et al., 2002), socio-demographic factors (Grundy and Holt, 2000) or living environment and neighbourhood (Bosma et al., 2001). A contrario, the persistence of massive unemployment and recurrent forms of inactivity has a deteriorating effect on health status. In France, the annual risk of death for an unemployed male is three times higher than for an employed male of the same age (Mesrine, 2000). Numerous studies show a causality relationship between unemployment and poor health (Chan and Stevens, 2001, Kalwij and Vermeulen, 2008; Bockerman and Ilmakunnas, 2009; Sullivan and von Wachter, 2009; Eliason and Storrie, 2009).

Several studies attach importance to the cumulative effects of events experienced throughout an individual's life cycle. It has been established that an individual's socio-economic situation during childhood, or even in utero, (according to the latency hypothesis or latency model Barker, 1999; Wadsworth, 1999), and labour market trajectory (Wadsworth, 1986; Currie and Hyson, 1999; Case et al., 2002; Currie and Stabile, 2003; Case et al., 2005) are important determinants of health status in adulthood. Lindeboom et al. (2006) for example, show that childhood environment influences the probability of experiencing occupational accidents and disability. A large body of literature equally acknowledges the role played by the social determinants of health (Wilkinson, 1992; Marmot and Wilkinson, 1999; Berkman and Kawachi, 2000). Psychosocial factors such as the feeling of hierarchical dominance, the loss of autonomy or deprivation increases the probability of exposure to stress and psychological distress. An individual's socio-professional situation will thus favour (or not) the emergence of psychosocial problems that will lead to a deterioration in health status.

1.3 Hence the necessity of considering the causality of the relationship

The relationship between health and employment status is thus extremely difficult to grasp in that it is characterized by endogeneity, simultaneous causality and measure. Although the existence of a health-related self-selection effect on labour market integration has been established, no econometric study in the health economics domain has clearly established the role played by employment status on health. Certain confounding factors can simultaneously influence both health status and socio-economic status such as a present-biased preference or the degree of aversion to risk.

In order to resolve some of these difficulties, Tessier and Wolff (2005) carried out the most comprehensive study to date using cross-sectional French data. It analyses the population aged 25 to 55 using a discreet time simultaneous equation model. The authors clearly establish the impact of health on labour market participation from an individual's entry into employment. On the other hand, professional activity does not appear to have an impact on health which could indicate a specification problem in the econometric model used.

Our aim in this work is to understand the simultaneous causality relationship between health status and employment status in the French population aged 30-59 using a similar model enriched with SIP data. In order to achieve this, several theoretical hypotheses will be tested using a gender-based approach.

One of the most striking features of the post-war economy was the growth in women's labour force participation (Maruani, 2004) although for 30% of them it essentially consisted in unemployment and part-time work (Arnault 2005). Women are faced with more constraints throughout their professional careers than men. The female image' is more likely to be modelled around the existence of family constraints than the male image' (Barnay, 2005). Finally, the call for projects specifically mentions the female population having had to cumulate family and professional responsibilities' as a specific domain for analysis. In this case, is the attempt to measure the health-employment status relationship in the female population relevant? This question is all the more legitimate in that several studies highlight the absence of a relationship between economic inactivity (the housewife' status) and health status in women aged over 50 (Barnay, 2008). Yet, women self-report more health problems than men of the same age (Strauss et al., 1998). Furthermore, women aged over 50 more frequently report activity limitations and illnesses than men. It is likely that for the younger generations, women's activity behaviours are closer to those of men. Some factors however remain specific to women's labour participation: the tax treatment of a household's second income, childminder schemes and child and parent support services, family benefits, parental or maternity leave, different factors that can potentially constrain women's activity behaviours.

Furthermore, the early retirement decision does not generally stem from individual behaviour (Chiappori, 1992). Yet the classical dynamic models of the type Rust and Phelan (1997) and option value models (Blanchet and Mahieu, 2001) assume that within a couple, retirement decisions are taken independently. The unit of reference is thus the individual rather than the couple. It however seems likely that, financial means permitting, the preference for "leisure" increases in value if the spouse has already retired from the labour market.

Several American studies have attempted to measure complementarities in spouses' preferences for leisure (Hurd, 1990; Blau, 1998; An et al., 1999). Gustman and Steinmeier (2000) and Hurd (1990) show that a correlation in spouses' retirement dates effectively exists and conclude that three factors explain this correlation: shared tastes, a wealth effect related to the marginal benefits of working an extra year, and complementary preferences for "leisure".

An et al. (1999) studied the joint retirement of married couples and demonstrated that living as a couple plays a determining role in the retirement decision whereas pension level and marginal gains from working longer appear to have little influence. Blau (1998) investigates the labour force transitions of older married couples and concludes that preferences for leisure' within a couple are interdependent. Sédillot and Walraët (2002) evaluated the interdependence of retirement decisions based on French data using a modelling method similar to that used by Gustman and Steinmeier (2000). They determine a household model in which spouses share their resources (the joint utility function is the sum of the utility function of each spouse) and benefit from economies of scale. The individual decision to retire (preference (or not) for leisure') depends on the spouse's activity status. For example, the probability of a married woman's retirement decision is three times higher if the husband has already retired than if he is still working. This French study confirms the results of the Loprest et al. (1995) study demonstrating that male and female behaviours are different and that the impact of martial status on retirement decisions differs according to gender. It is thus clear that women, through their investment in the family (turned towards caring for children, spouse and dependent parents) are required to make choices that will have an impact on their labour supply. Moreover, do family policies constitute an obstacle to continuing a professional activity? We would also like to differentiate between age effects and generation effects. Potential age effects can be numerous; one can intuitively deduct, for example, that the presence of children at the beginning of a woman's professional career will considerably constrain her labour supply. It is then likely that the presence of a spouse will have an impact on her labour supply, and finally, after 45 years old, informal care for a dependent parent is likely to have an impact on both health status and employment status. Generation effects can have an impact on fertility behaviour (delaying the age of first maternity, number of children...) and the growth of women's labour supply in relation to sector-based trends on the labour market (notably era' effects).

2 Methodology

2.1 The model

The aim of this study is to simultaneously measure the effects of health-related self-selection on employment status and the reverse causality effect within the population aged 30-59 in France (a total of 8,667 individuals of which 50.6% women). A specific modelling method was elaborated to take both simultaneity bias and the ordinal nature of the variables into account: a simultaneous-equation bi-ordered probit model presented as follows:

$$\begin{cases} y^{\star}_{1\,i} = a_1 y^{\star}_{2\,i} + b_1 x_{1\,i} + c_1 + u_{1\,i} \\ \text{if} \quad \lambda_{1\,j-1} \leq y^{\star}_{1\,i} < \lambda_{1\,j} \quad \text{then} \quad y_{1\,i} = j \quad j = 1, \ldots, K_1 \\ y^{\star}_{2\,i} = a_2 y^{\star}_{1\,i} + b_2 x_{2\,i} + c_2 + u_{2\,i} \\ \text{if} \quad \lambda_{2\,j-1} \leq y^{\star}_{2\,i} < \lambda_{2\,j} \quad \text{then} \quad y_{2\,i} = j \quad j = 1, \ldots, K_2 \end{cases}$$

The first equation explains the latent variable of professional integration, $y^{\star}_{1\,i}$, dependant on the latent variable of health status, $y^{\star}_{2\,i}$, and socio-economic variables written $x_{1\,i}$. Here

	% weighted
1. Inactivity	13.6
2. Unemployment	8.0
3. Employment	78.4

Table 1: Actual professional situation — population aged 30–59

it involves estimating a multilevel ordered polytomous probit model. The second equation explains the latent variable for health status, $y^{\star}_{2\,i}$; dependent on the latent variable for professional integration, $y^{\star}_{1\,i}$, and socio-economic variables written $x_{2\,i}$. Several health status measurement tools will be tested.

2.2 The variables of interest

Actual professional situation $y_{1\,i}$ is estimated from three ordered modalities: economic inactivity, unemployment and employment. The population retained in the sample is of working age and is consequently characterised by a high employment rate of 78% (cf. table 1).

Taking into account the complexity of measuring health status, we opted for the three health status measurements $y_{2\,i}$, constituting the Mini European Health Module. As is traditionally the case in this type of analysis, we use three health indicators susceptible of indicating the plurality of health statuses:

1. The self-perceived health indicator corresponds to the question standardised by the World Health Organisation Regional Office for Europe: “How is your health in general: very good, good, fair, poor or very poor?” 75% of the population aged 30-59 self-report good or very good health.

2. The prevalence of chronic diseases is determined from responses to the question: “Do you currently suffer from one or more chronic conditions?” Almost a third of the population suffers from a chronic disease.

3. Activity limitations are determined from responses to the question: “During the last six months, have you been limited in activities which people normally carry out due to a health problem?” 15% of the population report activity limitations.

2.3 Identification conditions and explanatory variables

Estimating a discreet choice, simultaneous equation model requires that identification conditions be defined in view of the elements of the hypotheses advanced by the literature. It is important to test the robustness of these conditions susceptible of influencing results. We retained the following explanatory factors for employment status and health status.

2.3.1 Explanatory variables for the professional insertion model

The choice of a population aged 30 and over avoids taking into account preliminary training periods considered as being periods of inactivity. Fixing the upper bound at 59 years old avoids the added complexities of analysing the conditions under which labour supply is maintained after the age of 50, due to the various opportunities for early retirement and dedicated schemes, resulting in a radical drop in the older worker employment rate (68.6% against 84.3% for individuals aged 40 to 49). The second phase involves analysing the population using the following stratification: 30-39, 40-49 years old and 50-59 years old since the explanatory variables for professional integration appear to differ significantly according to these age groups. Living as a couple, or the number of children in the household, also influence employment status through the complementarity of preferences for leisure and the financial burdens associated with family life. Employment rate furthermore appears to be impacted when the number of children reaches 3 and over (70.3% against 82.8% when there are only two children).

2.3.2 Explanatory variables for the health status model

Following the review of the literature, we introduce proxies for the intermediary factors influencing the relationship between employment status and health status. We first of all used an exogenous health status estimator allowing the temporal dynamics of health deterioration and individual health histories to be taken into account. This estimator is the sum of diseases and symptoms, disabilities and accidents taken from an individual's employment history but excluding any self-reported relationship with professional integration. The exogenous health status estimator (EHSE) is composed of four ordered modes: 1 (at least three health problems), 2 (two health problems), 3 (one health problem) and 4 (no health problems). We observe a significant correlation with current health status whatever the measurement retained. 92% of individuals who have not experienced health problems during the course of their professional life do not report activity limitations against 55.6% of individuals counting at least two health problems. Social determinants play an early role in an individual's constitution and rythm the deterioration of health capital. In order to measure the impact of childhood conditions, we added a dummy variable that indicates whether an individual experienced healthrelated events during childhood (disability; long-term illness; serious health problems within the respondent's family father, mother, other; physical, psychological, sexual abuse): 25% of individuals were concerned. This characteristic has a significant discriminatory effect on the prevalence of chronic illness in adulthood (at the date of the survey) since 40% of individuals who experienced these difficulties declared suffering from a chronic illness (against 25% for the others). As expected, high household income (sum of current monthly resources taking all types of income into account) is a factor that contributes

	weighted %	Share in employment
Self reported health		
Very bad	0.9	33.5
Bad	3.9	43.4
Fair	20.1	69.4
Good	48.5	82.3
Very good	26.7	84.8
Chronic illness		
No	70.3	82.1
Yes	29.7	69.9
Activity limitations		
No	85.9	82.0
Yes	14.1	57.0
Age		
30-39	33.5	81.8
40-49	34.5	84.3
50-59	32.0	68.6
Sex		
Women	50.6	72.2
Men	49.4	84.8
Safety professional biography		
0%-25% (long employment period)	17.2	46.9
25%-50% (long employment period)	12.6	74.2
50%-75% (long employment period)	20.8	81.5
75%-100% (long employment period)	49.5	89.2
Education level		
No diploma	9.4	62.6
Grade level	46.4	76.7
Bachelor level	16.9	79.5
Graduate level	27.3	86.2
Stability of professional biography		
one period (max)	16.5	88.9
2 periods	28.2	82.3
3 periods	22.3	77.5
at least 4 periods	33.0	70.7
Children		
None	15.7	78.9
1	19.8	80.7
2	38.0	82.8
3 and more	26.5	70.3
Couple living		
No	21.0	73.5
Yes	79.0	79.8

Source: SIP Survey.
Champ: 30-59 years old, N = 8 867

Table 2: Occupational status equation - explained variables

	weighted %	Self reported health good or very good	No chronic diseases	No activity limitations
Occupational Status				
Inactivity	13.6	57.3	57.2	69.1
Unemployment	8.0	61.1	61.0	76.6
Employment	78.5	79.7	73.6	89.8
Age				
30-39	33.5	84.9	78.8	91.7
40-49	34.5	76.0	71.6	86.0
50-59	32.0	64.2	60.1	79.9
Sex				
Women	50.6	73.0	68.4	85.4
Men	49.4	77.4	72.3	86.4
Exogenous health status indicator				
2 health problems and more	5.4	26.4	7.6	55.6
2 health problems	5.6	43.4	19.8	58.8
1 health problem	11.8	60.9	30.7	75.9
no health problem	77.2	83.1	84.5	91.6
Events in childhood				
No	74.7	78.8	74.1	88.6
Yes	25.4	64.7	59.2	77.9
Household earnings				
Less than 1 200	11.6	55.6	60.8	74.4
From 1 200 to 2 000	24.1	68.5	69.2	83.9
From 2 000 to 3 000	31.6	77.7	70.4	86.7
More than 3 000	32.8	84.6	74.6	90.7
Depression				
No	78.7	81.3	73.3	89.0
Yes	21.3	52.5	59.6	74.8

Champ: 30-59 years old, N = 8 867

Table 3: Health Status equation - explained variables

to protecting health capital. In households declaring monthly incomes of over 3,000, 85% of individuals self-report good or very good health against only 56% in households earning less than 1200. Finally, the question concerning depression permits capturing the effect of psychosocial factors on health status. Here it involves determining whether the respondent has felt depressed or down over the last two weeks ("During the last two weeks have you felt particularly sad, down or depressed most of the time during the day and almost every day?").

These different, simple statistical relationships must be tested so as to determine their resistance/robustness in an analysis, all other things being equal.

3 Econometric results

3.1 Estimation of the simultaneous equation model among the general population

A first series of estimations tested the causality relationship between health status and occupational status among the general population. In order to test the robustness of the subjective measure of health status, we carried out four estimations successively taking into account:

- perceived health status;
- the prevalence of chronic disease(s);
- the prevalence of activity limitations;
- perceived health after correction for chronic illness.

Whatever the health measurement, a better health status improves professional integration and employability. An improvement in self-perceived health results in better professional integration (coefficient equal to +0,5) and the presence of illness or activity limitations has a negative impact on employment status (respectively −0,2 and −0,6). The health-related self-selection hypothesis is thus corroborated, and significantly so, after controlling for socio-economic variables (age, gender, number of children, nature of the professional careers) and taking into account simultaneity biases. The other factors explaining employment status have an impact in unexpected areas: the diminishing degree of employability by age, and the virtuous circle effect of educational level for example. The variables characterising an individual's employment history deliver coherent messages, notably that a stable professional trajectory (amount of time spent in long-term employment) is protective of employment status quality at the time of the survey.

When we broach reverse causality or the impact of professional integration on current health status, the analysis leads to contrasted results. After correction for intermediary factors (income, childhood conditions, health status estimator, psychosocial factors), a better professional integration remains protective of individuals' perceived health status (estimated coefficient of 0.2) and reduces the probability of activity limitations. This causality relationship is nevertheless more statistically fragile than the preceding one. Intermediary factors play the role predicted by literature. The feeling of depression has a negative effect on the way individuals assess their perceived health status and reinforces the risk of suffering from activity limitations. Whatever the health measurement, health problems experienced during childhood tend to explain current health status.

On the contrary, when health status is approached from the prevalence of chronic illness, the results are not quite as expected. Initially surprising among the general population, a better professional integration appears to favour the emergence of chronic illness reinforcing the potentially pathogenic nature of professional activity.

3.2 Estimation of the simultaneous equation model according to gender

As gender naturally plays a significant role, especially on the labour market, we then carried out estimations for the male and female populations so as to consolidate results in view of a stratified analysis.

The health selection hypothesis is strengthened: for both sexes, a good health status facilitates professional integration. The determinants of professional integration are however markedly different between men and women. A higher number of children or living as a couple, for example, penalizes women's labour market integration whereas the relationship is reversed for men. The proportion of long-term employment periods during the professional trajectory constitutes a highly determining factor in men and women's professional integration. Nevertheless, the role played by the professional stability indicator (number of long-term employment periods) in the professional integration of men and women is diametrically opposed. A limited number of long-term employment periods favours professional integration for the male population but disadvantages women. When one examines the second health equation, we observe that, for both men and women, professional integration causes more chronic illnesses, all other things being equal, and confirms the hypothesis previously advanced. For the health equation, intermediary factors play a similar role for men and women both in terms of intensity and the significance of the relationship.

3.3 Estimation of the simultaneous equation model according to age (30-44 versus 45-59 years old)

Are these results resistant to an analysis by age range? Among the younger age group (30-44 years old), if we examine the professional integration perceived health model we observe that both the selection effect and virtuous circle effect of work on health are confirmed. On the other hand, the explanatory model differs for two variables. The advancement of age, integrated here as a continuum, facilitates professional integration for the 30-44 age group whereas it penalizes the 45 -59 age group. The presence of children limits the insertion of younger workers, notably due to maternity leave, but facilitates that of older workers. The determinants of perceived health status are virtually the same with the slight difference that advancement of age has a negative impact on self-perceived health between the ages of 30 and 45 but is not significant after 45 years old. The deleterious effect of professional integration on the prevalence of chronic illness equally resists in the analysis by age range.

Conclusion

In this paper, we try to establish from a model of equations simultaneous and latent variable model a full explanation of the relationship between health and occupational status

The results lead to corroborate three assumptions:

1. healthy worker effect;
2. reverse causality;
3. negative causal effect of occupational integration on health.

It seems that the results are sensitive to the measure of health used. The prevalence of disease appears to have a special status. In particular, further investigations should be conducted to explore the specificity of relationship we found between workforce development and the prevalence of chronic diseases.

The introduction of variables in the model controlling working conditions or measuring, more accurately, the nature of diseases by differentiating, for example those belonging (or not) of occupational diseases should help to better understand this phenomenon. In addition, a business approach should also provide input to characterise the causal relationship.

References

[1] P. Adams, M. D. Hurd, D. L. McFadden, A. Merrill, and T. Ribeiro. Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, 112(1):3–56, 2003.

[2] C. Afsa and P. Givord. Le rôle des conditions de travail dans les absences pour maladie: le cas des horaires irréguliers. *Économie et Prévision*, 187:83–104, 2009.

[3] J. H. Albert and S. Chib. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

[4] M. Y. An, B. J. Christensen, and N. Datta Gupta. A bivariate duration model of the joint retirementdecisions of married couples. Technical report, Centre for Labour Market and Social Research, 1999.

[5] S. Arnault. Le sous-emploi concerne 1,2 million de personnes. *Insee Première*, (1046), 2005.

[6] P. Askenazy. Lean Production and Workplace Health. Technical report, CEPREMAP, 2000.

[7] A. I. Balsa and T. G. McGuire. Statistical discrimination in health care. *Journal of Health Economics*, 20(6):881–907, 2001.

[8] D. J. P. Barker. Fetal origins of coronary heart disease. *British Medical Journal*, 311(6998):171–174, 1995.

[9] T. Barnay. Santé déclarée et cessation d'activité. *Revue Française d'Économie*, 20(2):73–106, 2005.

[10] T. Barnay. Chômage et invalidité après 50 ans : deux dispositifs alternatifs de sortie de l'emploi pour les seniors en mauvaise santé? *Économie et Statistiques*, (411):47–63, 2008.

[11] T. Barnay. In which ways do unhealthy people older than 50 exit the labour market in France? *The European Journal of Health Economics*, 11(2):127–140, 2009.

[12] T. Barnay and K. Briard. Carrière incomplète et départ en retraite: une estimation de l'incidence de l'état de santé à partir de données individuelles. *Revue économique*, 60(2):345–364, 2009.

[13] T. Barnay and F. Jeger. Quels dispositifs de cessation d'activité pour les personnes en mauvaise santé? *Études et Résultats*, (492), 2006.

[14] A. Biswas and K. Das. A Bayesian analysis of bivariate ordinal data: Wisconsin epidemiologic study of diabetic retinopathy revisited. *Statistics in Medicine*, 21(4):549–559, 2002.

[15] D. Blanchet and R. Mahieu. Une analyse microéconométrique des comportements de retrait d'activité. *Revue d'Économie Politique*, 111(3):9–31, 2001.

[16] N. Blanpain and D. Eneau. État de santé et accès aux soins des allocataires du rmi. *Insee Première*, (655), 1999.

[17] D. M. Blau. Labor force dynamics of older married couples. *Journal of Labor Economics*, 16(3):595–629, 1998.

[18] S. Blöndal and S. Scarpetta. The retirement decision in OECD countries. *OECD Economics Department Working Papers*, 1999.

[19] P. Böckerman and P. Ilmakunnas. Unemployment and self-assessed health: evidence from panel data. *Health Economics*, 18(2):161–179, 2009.

[20] H. Bosma, H. Dike van de Mheen, G. J. J. M. Borsboom, and J. P. Mackenbach. Neighborhood socioeconomic status and all-cause mortality. *American Journal of Epidemiology*, 153(4):363–371, 2001.

[21] J. Bound. Self-Reported Versus Objective Measures of Health in Retirement Models. *The Journal of Human Resources*, 26(1):106–138, 1991.

[22] J. Bound, M. Schoenbaum, T. R. Stinebrickner, and T. Waidmann. The dynamic effects of health on the labor force transitions of older workers. *Labour Economics*, 6(2):179–202, 1999.

[23] M. Campolieti. Disability and the labor force participation of older men in Canada. *Labour Economics*, 9(3):405–432, 2002.

[24] A. Case and A. Deaton. Broken down by work and sex: how our health declines. In D. A. Wise, editor, *Analyses in the Economics of Aging*, pages 185–205. University of Chicago Press, 2005.

[25] A. Case, D. Lubotsky, and C. Paxson. Economic status and health in childhood: The origins of the gradient. Technical report, National Bureau of Economic Research Cambridge, Mass., USA, 2001.

[26] S. Chan and A. H. Stevens. Job loss and employment patterns of older workers. *Journal of Labor Economics*, 19(2):484–521, 2001.

[27] P. Chiaporri. Collective Labour Supply and Welfare. *Journal of Political Economy*, 100(6):437–67, 1992.

[28] J. Currie and R. Hyson. Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. Technical report, National Bureau of Economic Research Cambridge, Mass., USA, 1999.

[29] J. Currie and B. C. Madrian. Health, health insurance and the labor market. *Handbook of labor economics*, 3:3309–3416, 1999.

[30] J. Currie and B. C. Madrian. Health, health insurance and the labor market. *Handbook of Labor Economics*, 3-C:3309–3416, 1999.

[31] J. Currie and M. Stabile. Socioeconomic status and health: why is the relationship stronger for older children? Technical report, National Bureau of Economic Research Cambridge, Mass., USA, 2002.

[32] A. Deaton and C. Paxson. *Themes in the Economics of Ageing*, chapter Mortality, education, income, and inequality among American cohorts. University of Chicago Press, 2001.

[33] F. Derriennic, M.-J. Saurel-Cubizolles, and C. Monfort. Santé, conditions de travail et cessation d'activité des salariés âgés. *Travail et Emploi*, (96):37–53, 2003.

[34] R. Disney, C. Emmerson, and M. Wakefield. Ill health and retirement in Britain: A panel data-based analysis. *Journal of Health Economics*, 25(4):621–649, 2006.

[35] S. Dumartin. Trois quarts des français se considèrent en bonne santé. *Insee Première*, (702), 2000.

[36] R. Ecob and G. Davey Smith. Income and health: what is the nature of the relationship? *Social Science & Medicine*, 48(5):693–705, 1999.

[37] M. Eliason and D. Storrie. Does Job Loss Shorten Life? *Journal of Human Resources*, 44(2):227–302, 2009.

[38] S. A. Everson, S. C. Maty, J. W. Lynch, and G. A. Kaplan. Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes. *Journal of psychosomatic research*, 53(4):891–895, 2002.

[39] K. Fiscella and P. Franks. Individual income, income inequality, health, and mortality: What are the relationships? *Health Services Research*, 35(1 Pt 2):307, 2000.

[40] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi. *GNU Scientific Library Reference Manual - Third Edition*. Network Theory Ltd., 3rd revised edition, 2009.

[41] W. H. Greene. *Econometric Analysis Second Edition*. Macmillan, New York, 1993.

[42] W. H. Greene and D. A. Hensher. Modeling Ordered Choices: A Primer and Recent Developments. Working Papers 08-26, New York University, Leonard N. Stern School of Business, Department of Economics, 2008.

[43] E. Grundy and G. Holt. Adult life experiences and health in early old age in Great Britain. *Social Science & Medicine*, 51(7):1061–1074, 2000.

[44] A. L. Gustman and T. L. Steinmeier. Retirement in dual-career families: a structural model. *Journal of Labor Economics*, 18(3):503–545, 2000.

[45] P. Haan and M. Myck. Dynamics of health and labor market risks. *Journal of Health Economics*, 28(6):1116–1125, 2009.

[46] M. D. Hurd. *Issues in the economics of Aging*, chapter The joint retirement decision of husbands and wives, pages 231–254. University of Chicago Press, 1990.

[47] F. Jusot, M. Khlat, T. Rochereau, and S. Sermet. Une mauvaise santé augmente fortement les risques de perte d'emploi. In *Données Sociales La Société française*, pages 533–542. Paris, 2006.

[48] A. Kalwij and F. Vermeulen. Health and labour force participation of older people in Europe: what do objective health indicators add to the analysis? *Health Economics*, 17(5):619–638, 2008.

[49] M. Kerkhofs, M. Lindeboom, and J. Theeuwes. Retirement, financial incentives and health. *Labour Economics*, 6(2):203–227, 1999.

[50] L. Lechene and M. T. *Trajectoires sociales et inégalités: Recherche sur les conditions de vie, MIRE et INSEE*, chapter Analyse des déterminants des salaires, pages 221–243. 1994.

[51] M. Lindeboom. Health and work of older workers. In A. Jones, editor, *The Elgar Companion to Health Economics*, pages 26–35. Edward Elgar Publishing, 2006.

[52] M. G. Marmot, R. G. Wilkinson, and E. Brunner. *Social determinants of health*. Oxford University Press Oxford, 2006.

[53] M. Maruani. Activité, précarité, chômage: toujours plus? *Revue de l'OFCE*, (3):95–115, 2004.

[54] F. Meallia and D. B. Rubin. Assumptions allowing the estimation of direct causal effects. *Journal of Econometrics*, 112(1):79–87, 2003.

[55] A. Mesrine. La surmortalité des chômeurs: un effet catalyseur du chômage? *Économie et Statistique*, (334):33–48, 2000.

[56] A.-F. Molinié. Interroger les salariés sur leur passé professionnel: le sens des discordances. *Revue d'épidémiologie et de Santé Publique*, 51(6):589–605, 2003.

[57] Z. Or. Determinants of health outcomes in industrialised countries: a pooled, cross-country, time-series analysis. *OECD Economic Studies*, pages 53–78, 2000.

[58] M. Perronnin, A. Pierre, and R. T. La complémentaire santé en france en 2008 : une large diffusion mais des inégalités daccès. *Questions d'économie de la santé Irdes*, (161), 2011.

[59] L. Rioux. Recherche d'emploi et insertion professionnelle des allocataires du rmi. *Économie et statistique*, 346(1):13–32, 2001.

[60] C. P. Robert. Simulation of truncated normal variables. *Statistics and Computing*, 5(2):121–125, 1995.

[61] C. P. Robert. *Méthodes de Monte Carlo par chaînes de Markov*. Economica, Paris, 1996.

[62] M. Salm. Does job loss cause ill health? *Health Economics*, 18(9):1075–1089, 2009.

[63] A. Skalli and J. M. étienne. Health Status and Socioeconomic Inequalities: A Review of the French Empirical Literature. In A. Skalli, E. Johansson, and I. Theodossiou, editors, *Are the Healthier Wealthier or the Wealthier Healthier? The European Evidence*. Taloustieto Oy, Helsinki, 2006.

[64] J. P. Smith. Healthy bodies and thick wallets: The dual relation between health and economic status. *The Journal of Economic Perspectives*, 13(2):145–166, 1999.

[65] J. P. Smith and R. Kington. Demographic and economic correlates of health in old age. *Demography*, 34(1):159–170, 1997.

[66] M. J. Soobader and F. B. LeClere. Aggregation and the measurement of income inequality: effects on morbidity. *Social Science & Medicine*, 48(6):733–744, 1999.

[67] S. Stern. Measuring the Effect of Disability on Labor Force Participation. *The Journal of Human Resources*, 24(3):361–395, 1989.

[68] J. Strauss and D. Thomas. Health, nutrition, and economic development. *Journal of economic literature*, 36(2):766–817, 1998.

[69] K. Stronks, H. D. van de Mheen, C. W. N. Looman, and J. P. Mackenbach. Cultural, Material, and Psychosocial Correlates of the Socioeconomic Gradient in Smoking Behavior among Adults. *Preventive Medicine*, 26(5):754–766, 1997.

[70] R. Sturm and C. R. Gresenz. Relations of income inequality and family income to chronic medical conditions and mental health disorders: national survey. *British Medical Journal*, 324(7328):20, 2002.

[71] D. Sullivan and T. Wachter von. Job Displacement and Mortality: An Analysis Using Administrative Data. *Quarterly Journal of Economics*, 124(3):1265–1306, 2009.

[72] B. Sédillot and E. Walraët. La cessation d'activité au sein des couples: y at-il interdépendance des choix? *Économie et Statistique*, (357–358):79–98, 2002.

[73] P. Tessier and F.-C. Wolff. Offre de travail et santé en France. *Économie et Prévision*, (168):17–41, 2005.

[74] H. Van de Mheen, K. Stronks, C. Schrijvers, and J. Mackenbach. The influence of adult ill health on occupational class mobility and mobility out of and into employment in the netherlands. *Social science & medicine*, 49(4):509–518, 1999.

[75] C. van Rossum, H. van de Mheen, M. Breteler, D. E. Grobbee, and J. P. Mackenbach. Socioeconomic differences in stroke among Dutch elderly women: the Rotterdam Study. *Stroke*, 30(2):357, 1999.

[76] M. E. J. Wadsworth. *Class and health: research and longitudinal data*, chapter Serious illness in childhood and its association with later-life achievement. Routledge, 1986.

[77] R. Wilkinson. Income distribution and life expectancy. *British medical journal*, 304(6820):165, 1992.

A The SBOP and its' estimation

This appendix details the method used in this study: the Gibbs sampling method to estimate a simultaneous equation model in which the variables to be explained are ordered qualitative variables. It thus involves an ordered, simultaneous, bi-probit model which we name SBOP (Simultaneous bi ordered probit).

In the first part of this appendix, we specify the two points of departure utilised: the Gibbs sampler and the estimation of the ordered probit model using the Gibbs sampler. In the second part, we present the method we elaborated.

A.1 The points of departure

A.1.1 The Gibbs sampler

The Gibbs sampler is used to sample a random vector $\underline{\theta}$, partitioned into p elements such that $\underline{\theta}' = (\underline{\theta}'_1, \cdots, \underline{\theta}'_p)$. This method is notably applicable when it is easy to sample from the conditional distribution $f(\underline{\theta}_i \mid \underline{\theta}_j, j \neq i)$. In effect, the Gibbs sampler comes down to iterating the following algorithm from an initial value $\underline{\theta}^{(0)}$:

$$\left\{\begin{array}{rl} 1. & \text{Generate } \underline{\theta}_1^{(t+1)} \sim f(\underline{\theta}_1 \mid \underline{\theta}_2^{(t)}, \cdots, \underline{\theta}_p^{(t)}) \\ 2. & \text{Generate } \underline{\theta}_2^{(t+1)} \sim f(\underline{\theta}_2 \mid \underline{\theta}_1^{(t+1)}, \underline{\theta}_3^{(t+1)}, \cdots, \underline{\theta}_p^{(t)}) \\ \cdots & \\ j. & \text{Generate } \underline{\theta}_j^{(t+1)} \sim f(\underline{\theta}_j \mid \underline{\theta}_1^{(t+1)}, \cdots, \underline{\theta}_{j-1}^{(t+1)}, \underline{\theta}_{j+1}^{(t)}, \cdots, \underline{\theta}_p^{(t)}) \\ \cdots & \\ p. & \text{Generate } \underline{\theta}_p^{(t+1)} \sim f(\underline{\theta}_p \mid \underline{\theta}_1^{(t+1)}, \cdots, \underline{\theta}_{p-1}^{(t+1)}) \end{array}\right.$$

One shows that, for $t \to \infty$, the joint distribution of $\underline{\theta}^{(t)}$ tends towards the joint distribution of $\underline{\theta}$. After an initial "burn-in" period, that is to say for $t \geq t_0$, it will thus be possible to consider that $\underline{\theta}^{(t)}$ is effectively a simulated value of $\underline{\theta}$.

A priori the method is not particularly appealing in that, on the one hand, it requires numerous iterations during the initial "burn-in" period without obtaining any results and, on the other, because the "burn-in" period has to be reiterated each time one wants to obtain a simulated value for the random vector $\underline{\theta}$.

Reiteration of the "burn-in" phase, however, is unnecessary in that, to obtain several simulated values for $\underline{\theta}$ and obtain an n-sample, it suffices to retain $\underline{\theta}^{(t_0)}$, $\underline{\theta}^{(t_0+k)}$, $\underline{\theta}^{(t_0+2k)}$, $\cdots$, $\underline{\theta}^{(t_0+(n-1)k)}$ where k is the step used to recuperate a simulation in the Markov chain made up from $\underline{\theta}^{(t_0)}$, $\underline{\theta}^{(t_0+1)}$, $\cdots$, $\underline{\theta}^{(t_0+(n-1)k)}$. In effect, $\underline{\theta}^{(t)}$ and $\underline{\theta}^{(t+1)}$ are strongly autocorrelated by construction and one can suppose that $\underline{\theta}^{(t)}$ and $\underline{\theta}^{(t+k)}$ are approximately independent. It is for this reason that the Gibbs sampler is similar to the method known as the Markov Chain-Monte Carlo (MCMC).

A.1.2 The Gibbs sampler and estimation of the ordered probit model

Albert and Chib (1993) proposed estimating the ordered probit model using the Gibbs sampler. The ordered probit modelling method is designed to explain a qualitative variable the modalities of which are ordered. Suppose, for example, that the variable in question takes the values 1, 2 and 3 to code the responses "a little", "very much" and "passionately" to a given question, these values determine three groups of observations. This model assumes the presence of a latent variable and a threshold level that, when crossed by the latent variable, explains the discreet values taken by the variable to be explained.

Formally, by writing the variable latent $y_i^\star$, the modalities of the variable to be explained would be generated as follows.

$$\begin{cases} \text{if} \quad y_i^\star < \lambda_1 & \text{then} \quad y_i = 1 \\ \text{if} \quad \lambda_1 \leq y_i^\star < \lambda_2 & \text{then} \quad y_i = 2 \\ \text{if} \quad \lambda_2 \leq y_i^\star & \text{then} \quad y_i = 3 \end{cases}$$

There is one threshold less than there are groups. Taking $\lambda_0 = -\infty$ and $\lambda_3 = +\infty$, one can write

$$\text{if} \quad \lambda_{j-1} \leq y_i^\star < \lambda_j \quad \text{then} \quad y_i = j \quad j = 1,\ldots,K$$

where K is the number of groups.

Then, a linear model is specified to retrace the evolution of the latent variable:

$$y_i^\star = a_1 x_{i\,1} + a_2 x_{i\,2} + \cdots + a_p x_{i\,P} + u_i \quad i = 1,\ldots,N$$

where $x_{i\,j}$ $j = 1,\ldots,P$ is the value of the explanatory variable j for the observation i, where a_j $j = 1,\ldots,P$ is the unknown coefficient to be estimated and where u_i is the error term. With matrix notation, this linear model becomes:

$$\underline{y}^\star = X\underline{a} + \underline{u}$$

where X is the explanatory variables matrix and $\underline{a}$ the vector of the coefficients to be estimated. It is advisable to normalize the model without which certain parameters would be unidentifiable. Threshold levels and threshold ranges must be determined. In effect, if the constant is included in the model, we see that the threshold levels remain undetermined. Similarly, the threshold points constitute the rungs of a ladder and one could, for example sets λ_1 to 0 and λ_3 to 10. Usual practice takes $\lambda_1 = 0$ and $\mathrm{V}(u_i) = 1 \; \forall \, i$.

Albert and Chib (1993) propose, on the one hand, using a Bayesian approach taking the coefficients and thresholds as being random variables for which the joint distribution should be obtained and, on the other, a data "augmentation" taking the elements of the latent variable as non-observed random variables to be included in the estimation process. They thus suggest applying the Gibbs sampling algorithm by partitioning the parameters vector $\underline{\theta}$ into $P + (K-1) + N$ blocks. The first block, of size P, is destined to simulate the coefficients (the a_j); the following $K-1$ blocks are destined to simulate the thresholds (the λ_j) and the last N blocks are destined to simulate the elements of the latent variable (the $y_i^\star$).

Albert and Chib (1993) finally show that if one retains a diffuse prior on the coefficients and normality of the error term, the joint conditional distribution of the coefficients is written a follows:

$$\underline{a} \mid \underline{\lambda}, \underline{y}^\star \sim \mathcal{N}_P(\underline{\hat{a}}, \widehat{\Omega}) \tag{1}$$

where $\underline{\hat{a}} = (X'X)^{-1}X'\underline{y}$ and where $\widehat{\Omega} = (X'X)^{-1}$.

Similarly, by taking a diffuse-prior threshold structure, the conditional distribution for one of these is written as follows:

$$\lambda_j \mid \underline{\lambda}_{\{-j\}}, \underline{y}^\star, \underline{a} \sim \mathcal{U}(\phi_j^l, \phi_j^r) \quad j = 1, \ldots, K-1 \tag{2}$$

where $\underline{\lambda}'_{\{-j\}} = (\lambda_0, \ldots, \lambda_{j-1}, \lambda_{j+1}, \lambda_K)'$, where $\mathcal{U}(\cdot,\cdot)$ denotes the uniform distribution and with

$$\begin{cases} \phi_j^l = \max(\lambda_{j-1}, \max\{y_i^\star \mid y_i = j\}) \\ \phi_j^r = \min(\lambda_{j+1}, \min\{y_i^\star \mid y_i = j+1\}) \end{cases} \qquad j = 1, \ldots, K-1$$

Eventually, the conditional distribution for an element of the latent variable is written as follows:

$$y_i^\star \mid y_i, \underline{a}, \underline{\lambda} \sim \mathcal{N}(a_1 x_{i\,1} + a_2 x_{i\,2} + \cdots + a_p x_{i\,p}, 1) \quad \text{truncated on} \quad (\lambda_{y_i}, \lambda_{y_i+1}) \tag{3}$$

It is thus easy to apply the Gibbs sampler by respectively simulating the conditional distributions given by the equations 1, 2, and 3. Of course, for the moment this is nothing but a theoretical curiosity since it is much easier to estimate an ordered probit by using the maximum likelihood estimator from the start. Likelihood estimation is not a complex process and maximising likelihood is relatively easy using a standard numerical algorithm. However, the method suggested by Albert and Chib (1993) proves useful when a more complex model is required.

A.2 A simultaneous equation bi-ordered probit model

The analysis of causality relationships between two ordered qualitative variables is usually carried out using a set of dummy variables for each of the two variables. If $\underline{y}_1$ and $\underline{y}_2$ denote the two variables and supposing they each take the values 1, 2 and 3, the following model would be used:

$$\begin{cases} y^\star_{1\,i} = a'_{1\,1}1(y_{2\,i}{=}1) + a'_{1\,3}1(y_{2\,i}{=}3) + b_1 x_{1\,i} + c_1 + u_{1\,i} \\ \text{if} \quad \lambda_{1\,j-1} \leq y^\star_{1\,i} < \lambda_{1\,j} \quad \text{then} \quad y_{1\,i} = j \quad j = 1,2,3 \\ y^\star_{2\,i} = a'_{2\,1}1(y_{1\,i}{=}1) + a'_{2\,3}1(y_{1\,i}{=}3) + b_2 x_{2\,i} + c_2 + u_{2\,i} \\ \text{if} \quad \lambda_{2\,j-1} \leq y^\star_{2\,i} < \lambda_{2\,j} \quad \text{then} \quad y_{2\,i} = j \quad j = 1,2,3 \end{cases}$$

where $1(y_{2\,i}{=}1)$ is the formula for the dummy variable equal to 1 when $y_{2\,i}$ takes the value 1, where $\underline{x}_1$ is the explanatory variable in the first equation and $\underline{x}_2$ in the second equation. To alleviate the formulae, we assume that only one explanatory variable intervenes (other than the constant) in each of the two equations. It should then be noted that it is impossible to introduce the complete set of indicator variables, for example, in the first equation, $1(y_{2\,i}{=}1)$, $1(y_{2\,i}{=}2)$ and $1(y_{2\,i}{=}3)$: one must retain a modal reference. Finally, as in the previous case, the model must be normalized. Let $\lambda_{1\,0} = \lambda_{2\,0} = 0$, and $\mathrm{V}(u_{1\,i}) = \mathrm{V}(u_{2\,i}) = 1$.

The causality relationship between the variables to be explained are expressed by the coefficients $a'_{1\,1}$, $a'_{1\,3}$, $a'_{2\,1}$, and $a'_{2\,3}$. Simultaneity can also result from a common component of the two error terms $u_{1\,i}$ and $u_{2\,i}$. We thus assume that the variance-covariance matrix of the random vector $\underline{u}'_i = (u_{1\,i}, u_{2\,i})'$ is :

$$\mathrm{V}(\underline{u}_i) = \mathrm{V}\begin{pmatrix} u_{1\,i} \\ u_{2\,i} \end{pmatrix} = \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad \forall\, i$$

This specification does not fully take into account the outcomes of this model. In effect, we suppose the existence of two latent variables, $\underline{y}^\star_1$ and $\underline{y}^\star_2$, but we do not introduce cross coefficients between the two latent variables into the system. We propose using the following model:

$$\begin{cases} y^\star_{1\,i} = a_1 y^\star_{2\,i} + b_1 x_{1\,i} + c_1 + u_{1\,i} \\ \text{if} \quad \lambda_{1\,j-1} \leq y^\star_{1\,i} < \lambda_{1\,j} \quad \text{then} \quad y_{1\,i} = j \quad j = 1, \ldots, K_1 \\ y^\star_{2\,i} = a_2 y^\star_{1\,i} + b_2 x_{2\,i} + c_2 + u_{2\,i} \\ \text{if} \quad \lambda_{2\,j-1} \leq y^\star_{2\,i} < \lambda_{2\,j} \quad \text{then} \quad y_{2\,i} = j \quad j = 1, \ldots, K_2 \end{cases}$$

A.2.1 The maximum likelihood estimator conditioned on the latent variables

Using the standard simultaneous equation formulas, the matrix structure can be written as follows:

$$Y^\star \Gamma + X B = U$$

with

$$Y^\star = \left(\underline{y}^\star_1 \mid \underline{y}^\star_2\right), \;\; X = \left(\underline{x}_1 \mid \underline{\ell}_N \mid \underline{x}_2 \mid \underline{\ell}_N\right), \;\; U = \left(\underline{u}_1 \mid \underline{u}_2\right)$$

where $\underline{\ell}_N$ is the vector size $N{\times}1$ for which the elements are equal to 1 to represent the constant and, finally, with

$$\Gamma = \begin{pmatrix} 1 & -a_2 \\ -a_1 & 1 \end{pmatrix}, \;\; B = \begin{pmatrix} -b_1 & -c_1 & 0 & 0 \\ 0 & 0 & -b_2 & -c_2 \end{pmatrix}.$$

This first representation is known as the structural form model. The reduced form model results from solving the system with respect to $Y^\star$:

$$Y^\star = -X B \Gamma^{-1} + U \Gamma^{-1} = X\Pi + W$$

where

$$\Pi = -B\Gamma^{-1}, \;\; W = \left(\underline{w}_1 \mid \underline{w}_2\right) \;\text{ et }\; W = U\Gamma^{-1}$$

One thus obtains $\underline{u}_i = \Gamma^{-1\prime}\underline{w}_i \;\forall\, i$ where $\underline{w}_i' = (w_{1\,i}, w_{2\,i})'$. The variance-covarience matrix of the error terms in the reduced form model is thus:

$$\mathrm{V}\begin{pmatrix} w_{1\,i} \\ w_{2\,i} \end{pmatrix} = \Omega = \Gamma^{-1\prime}\Sigma\Gamma^{-1} \quad \forall\, i$$

This reduced form model permits the model's log likelihood, conditioned on the latent variable, that is to say as if the $Y^\star$ matrix was observable, to be expressed as:

$$\ln(L) = -N\ln(2) + \frac{N}{2}\ln(\Omega) + \frac{N}{2}\mathrm{tr}(\Omega^{-1}S_w)$$

where S_w corresponds to a matrix size 2×2 written $\frac{1}{N}\sum_i w_{j\,i} w_{k\,i}$. It must of course be understood that $w_{1\,i}$ and $w_{2\,i}$ are implicit functions of Γ, B, $\underline{Y}^\star_i$, ant $\underline{x}_i$. After several algebraic manipulations, notably detailed in Greene (1993), the log likelihood can be written as follows :

$$\ln(L) = -N\ln(2) + N\ln|\Gamma| - \frac{N}{2}\ln|\Sigma| + \frac{N}{2}\mathrm{tr}(\Sigma^{-1}S_u)$$

where S_u is a matrix size 2×2 written $\frac{1}{N}\sum_i u_{j\,i}u_{k\,i}$. In the case of our two equation system with single error terms, the log likelihood is simplified as follows:

$$\ln(L) = -N\ln(2) + N\ln(1-a_1a_2) - \frac{N}{2}\ln(1-\rho^2) + \frac{1}{2(1-\rho^2)}\left(\sum_i u_{1\,i}^2 - 2\rho\sum_i u_{1\,i}u_{2\,i} + \sum_i u_{2\,i}^2\right)$$

It would not be very complicated to use a numerical solution algorithm to obtain an estimation of the coefficients. The derivatives of the log likelihood are expressed as follows for the two coefficients of the first equation:

$$\frac{\partial\ln(L)}{\partial a_1} = \frac{1}{1-\rho^2}\left(\sum_i y_{2\,i}^{\star}u_{1\,i} - \rho\sum_i y_{2\,i}^{\star}u_{2\,i}\right) - N\frac{a_2}{1-a_1a_2}$$

$$\frac{\partial\ln(L)}{\partial b_1} = \frac{1}{1-\rho^2}\left(\sum_i x_{1\,i}u_{1\,i} - \rho\sum_i x_{1\,i}u_{2\,i}\right)$$

For the second equation coefficients, the formulae are comparable, mutatis mutandis. For the derivative with respect to ρ, the correlation between the two equations, is not quite as simple:

$$\frac{\partial\ln(L)}{\partial\rho} = \frac{1}{1-\rho^2}\sum_i u_{1\,i}u_{2\,i} - \rho\left(\sum_i u_{1\,i}^2 - 2\rho\sum_i u_{1\,i}u_{2\,i} + \sum_i u_{2\,i}^2\right)\Big/(1-\rho^2) + N\frac{\rho}{1-\rho^2}$$

Our proposal is thus to use the Gibbs sampling algorithm and the asymptotic distribution of the maximum likelihood estimator as the conditional distribution to simulate the coefficients a_1, b_1, c_1, a_2, b_2, c_2, and ρ. In a fairly comparable but much simpler context (a bi-ordered probit model in which the two equations are correlated through the two error terms), Biswas and Das (2002) suggest using the Gibbs sampler.

A.2.2 The Gibbs sampler and estimation of the simultaneous equation bi-ordered probit model

We thus seek to estimate the simultaneous equation system

$$\begin{cases} y^{\star}_{1\,i} = a_1 y^{\star}_{2\,i} + b_1 x_{1\,i} + c_1 + u_{1\,i} \\ \text{if} \quad \lambda_{1\,j-1} \leq y^{\star}_{1\,i} < \lambda_{1\,j} \quad \text{then} \quad y_{1\,i} = j \quad j = 1,\ldots,K_1 \\ y^{\star}_{2\,i} = a_2 y^{\star}_{1\,i} + b_2 x_{2\,i} + c_2 + u_{2\,i} \\ \text{if} \quad \lambda_{2\,j-1} \leq y^{\star}_{2\,i} < \lambda_{2\,j} \quad \text{then} \quad y_{2\,i} = j \quad j = 1,\ldots,K_2 \end{cases}$$

that has the two following particularities. On the one hand, the variables to be explained are ordered qualitative variables, circumvented using latent variables (the variables $\underline{y}^{\star}_1$ and $\underline{y}^{\star}_2$) and thresholds (parameters $\lambda_{1\,j}$ and $\lambda_{2\,j}$) to explain the modalities taken by the variable to be explained. On the other hand, the cross difference between the two equations falls on the latent variable of the opposing equation. The latent variables are non observable and it is thus tempting to use the Gibbs sampler to estimate this system since this method allows one to have these latent variables at one's disposal by simulation.

Let $\underline{\gamma}$ be the vector of the parameters a_1, b_1, c_1, a_2, b_2, c_2, and ρ:

$$\underline{\gamma}' = (a_1, b_1, c_1, a_2, b_2, c_2, \rho)'.$$

We state the following conjecture. It is possible to retain a prior distribution on $\underline{\gamma}$ so that its posterior distribution becomes the asymptotic distribution of the maximum likelihood estimator. This conjecture is particularly reasonable. It is notably verified in a number of simpler models, such as the ordered univariate probit model, by retaining a diffuse prior.

We thus assume that the conditional distribution of the parameter vector is:

$$\underline{\gamma} \mid \underline{y}^{\star}_1, \underline{y}^{\star}_2, \underline{\lambda}_1, \underline{\lambda}_2 \sim \mathcal{N}_7(\underline{\hat{\gamma}}, \widehat{\Phi}) \tag{4}$$

where

$$\underline{\hat{\gamma}} = \arg\max \ln(L) \quad \text{and} \quad \widehat{\Phi} = \left(- \left. \frac{\partial^2 \ln(L)}{\partial \underline{\gamma}\, \partial \underline{\gamma}'} \right|_{\underline{\gamma} = \underline{\hat{\gamma}}} \right)^{-1}$$

We take the asymptotic distribution of the maximum likelihood estimator by estimating the variance-covariance matrix of the asymptotic distribution by the hessian matrix of the log likelihood evaluated at the point that maximises likelihood.

The joint conditional distribution of one of the elements of the two latent variables,

$\underline{y}_i^\star$, is obtained by using the reduced form of the system. This is equal to:

$$\underline{y}_i^\star \mid \underline{y}_{1\,-i}^\star, \underline{y}_{2\,-i}^\star, \underline{\lambda}_1, \underline{\lambda}_2, \underline{\gamma} \sim \mathcal{N}_2(\underline{\mu}_i, \Omega) \quad \text{truncated on} \quad (\underline{\lambda}_{\underline{y}_i}, \underline{\lambda}_{\underline{y}_i+1})$$

where

$$\underline{\mu}_i = \begin{pmatrix} \mu_{1\,i} \\ \mu_{2\,i} \end{pmatrix} = \frac{1}{1-a_1a_2} \begin{pmatrix} b_1x_{1\,i} + c_1 + a_1(b_2x_{2\,i} + c_2) \\ a_2(b_1x_{1\,i} + c_1) + b_2x_{2\,i} + c_2 \end{pmatrix}$$

and

$$\Omega = \Gamma^{-1\prime}\Sigma\Gamma^{-1} = \frac{1}{(1-a_1a_2)^2} \begin{pmatrix} 1 + a_1^2 + 2\,a_1\rho & a_1 + a_2 + \rho + a_1a_2\rho \\ a_1 + a_2 + \rho + a_1a_2\rho & 1 + a_2^2 + 2\,a_2\rho \end{pmatrix}.$$

The signs $\underline{\lambda}_{\underline{y}_i}$ and $\underline{\lambda}_{\underline{y}_i+1}$ designate the four thresholds of observation $\underline{y}_i$; or more precisely:

$$\underline{\lambda}_{\underline{y}_i} = \begin{pmatrix} \lambda_{1\,y_{1\,i}} \\ \lambda_{2\,y_{2\,i}} \end{pmatrix} \qquad \underline{\lambda}_{\underline{y}_i+1} = \begin{pmatrix} \lambda_{1\,y_{1\,i}+1} \\ \lambda_{2\,y_{2\,i}+1} \end{pmatrix}$$

In order to simulate this joint conditional distribution, we once again apply the Gibbs sampler as proposed by Robert (1995). Let us designate by ω_{11}, ω_{12}, and ω_{22} the terms of the variance-covariance matrix Ω:

$$\begin{cases} \omega_{11} = (1 + a_1^2 + 2\,a_1\rho)\big/(1-a_1a_2)^2 \\ \omega_{12} = (a_1 + a_2 + \rho + a_1a_2\rho)\big/(1-a_1a_2)^2 \\ \omega_{22} = (1 + a_2^2 + 2\,a_2\rho)\big/(1-a_1a_2)^2 \end{cases}$$

One has

$$y_{1\,i}^\star \mid y_{2\,i}^\star, \underline{\mu}_i, \Omega, \underline{\lambda}_1 \sim \mathcal{N}_1\left(\mu_{1\,i} + \frac{\omega_{12}}{\omega_{22}}(y_{2\,i}^\star - \mu_{2\,i}), \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}\right) \quad \text{truncated on} \quad (\lambda_{1\,y_{1\,i}}, \lambda_{1\,y_{1\,i}+1}) \tag{5}$$

and

$$y_{2\,i}^\star \mid y_{1\,i}^\star, \underline{\mu}_i, \Omega, \underline{\lambda}_2 \sim \mathcal{N}_1\left(\mu_{2\,i} + \frac{\omega_{12}}{\omega_{11}}(y_{1\,i}^\star - \mu_{1\,i}), \omega_{22} - \frac{\omega_{12}^2}{\omega_{11}}\right) \quad \text{truncated on} \quad (\lambda_{2\,y_{2\,i}}, \lambda_{2\,y_{2\,i}+1}) \tag{6}$$

Finally, the conditional distribution of the thresholds of the first equation is written as follows:

$$\lambda_{1\,j} \mid \underline{\lambda}_{1\,\{-j\}}, \underline{y}_1^\star \sim \mathcal{U}(\phi_{1\,j}^l, \phi_{1\,j}^r) \quad j = 1, \ldots, K_1 - 1 \tag{7}$$

with

$$\begin{cases} \phi_{1\,j}^l = \max(\lambda_{1\,j-1}, \max\{y_{1\,i}^\star \mid y_{1\,i} = j\}) \\ \phi_{1\,j}^r = \min(\lambda_{1\,j+1}, \min\{y_{1\,i}^\star \mid y_{1\,i} = j+1\}) \end{cases} \qquad j = 1, \ldots, K_1 - 1$$

With comparable notations for the second equation thresholds we obtain:

$$\lambda_{2j} \mid \underline{\lambda}_{2\{-j\}}, \underline{y}_2^\star \sim \mathcal{U}(\phi_{2j}^l, \phi_{2j}^r) \quad j = 1, \ldots, K_2-1 \tag{8}$$

We thus use the Gibbs sampling method by simulating the parameters vector $\underline{\gamma}$ (see equation 4), the N elements of the two latents $\underline{y}_i^\star$ (each time subsampling the Gibbs sampler so as to simulate the joint conditionality rule of $y_{1i}^\star$ and $y_{2i}^\star$ by using the equations 5 and 6) and, finally, the K_1-1 thresholds of the first equation and the K_2-1 thresholds of the second equation (see equations 7 and 8).

It was not an easy task to programme this method. We first used the Sas system, notably to be able to reuse the procedure that calculates the maximum likelihood of a bivariate system. The hypotheses are nevertheless unusual in that they assume variances for the two error terms being equal to 1. We finally opted for the C++ language using the GNU Scientific library (the GSL) particularly for the minimising numerical algorithm known as the Broyden-Fletcher-Goldfarb-Shanno method. We programmed the log likelihood derivatives relative to the parameters by using their analytical expression. To evaluate the hessian matrix of second order derivatives, we simply evaluated these derivatives numerically. The pre-burn-in stage is particularly long: the review of the thresholds is slow because the length of the interval from which the new threshold is taken is particularly weak. In addition, the Markov chain is highly auto-correlated, despite the joint simulation of the two elements of the latent variable. We thus take a step of 100 to retain a simulation on this chain. The programme execution time is thus considerable, averaging about ten hours for the data set we used on the relatively rapid PC at our disposal.

	1t equation *OCCUP*	2d equation *SELF_PERC_HEALTH*	Average
Frequency (%)			
1	15	1	
2	7.8	4.3	
3	77	21	
4	—	48	
5	—	25	
Coefficients			
SELF_PERC_HEALTH	0.475 (22)	—	—
OCCUP	—	0.165 (1.8)	—
AGE	−0.269 (−14)	−0.187 (−4.9)	2.1
MALE	0.162 (6)	−0.104 (−3.6)	0.45
INTERCEPT	−0.631 (−5.7)	1.25 (10)	—
NR_CHILDREN	−0.0274 (−3.5)	—	2.8
COUPLE	−0.143 (−5.4)	—	0.72
PROF_SAFETY	0.335 (10)	—	3
EDUCATION	0.0687 (7.1)	—	2.6
CAREER_SPELLS	0.0151 (1.8)	—	2.7
INFANCY	—	−0.192 (−9)	0.26
DEPRESSION	—	−0.517 (−12)	0.22
HEALTH_INDICATOR	—	0.346 (12)	3.6
HOUSEHOLD_INC	—	0.185 (9.3)	2.8
ρ	−0.56 (−7.4)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.343 (7.7)	0.708 (11)	—
λ_3	$+\infty$	1.89 (20)	—
λ_4	—	3.36 (24)	—
λ_5	—	$+\infty$	—

Source : SIP Survey, 8 667 observations.

Table 4: Occupational status versus self perceived health, total population

	1t equation *OCCUP*	2d equation *CHRONIC_DISEASES*	Average
Frequency (%)			
1	15	69	
2	7.8	31	
3	77	—	
Coefficients			
CHRONIC_DISEASES	−0.204 (−22)	—	—
OCCUP	—	0.368 (16)	—
AGE	−0.387 (−27)	0.194 (14)	2.1
MALE	0.145 (6.9)	−0.123 (−5.9)	0.45
INTERCEPT	0.663 (11)	3.84 (47)	—
NR_CHILDREN	−0.0363 (−3.6)	—	2.8
COUPLE	−0.008 71 (−0.42)	—	0.72
PROF_SAFETY	0.438 (43)	—	3
EDUCATION	0.0904 (8.9)	—	2.6
CAREER_SPELLS	−0.0781 (−7.6)	—	2.7
INFANCY	—	0.127 (5.7)	0.26
DEPRESSION	—	−0.005 17 (−0.21)	0.22
HEALTH_INDICATOR	—	−0.885 (−60)	3.6
HOUSEHOLD_INC	—	−0.0335 (−3)	2.8
SELF_PERC_HEALTH	—	−0.496 (−38)	3.9
ρ	−0.2 (−6.9)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.373 (8.8)	$+\infty$	—
λ_3	$+\infty$	—	—

Source : SIP Survey, 8 667 observations.

Table 5: Occupational status versus chronic diseases, total population

	1t equation *OCCUP*	2d equation *ACTIVITY_LIMITATIONS*	Average
Frequency (%)			
1	15	85	
2	7.8	15	
3	77	—	
Coefficients			
ACTIVITY_LIMITATIONS	−0.556	—	—
	(−22)		
OCCUP	—	−0.0522	—
		(−1.9)	
AGE	−0.296	0.186	2.1
	(−13)	(12)	
MALE	0.193	0.122	0.45
	(8.1)	(5)	
INTERCEPT	0.006 65	0.189	—
	(0.089)	(3.7)	
NR_CHILDREN	−0.0321	—	2.8
	(−3.4)		
COUPLE	−0.155	—	0.72
	(−7.1)		
PROF_SAFETY	0.378	—	3
	(22)		
EDUCATION	0.0555	—	2.6
	(5.2)		
CAREER_SPELLS	0.008 81	—	2.7
	(0.96)		
INFANCY	—	0.208	0.26
		(9.3)	
DEPRESSION	—	0.348	0.22
		(15)	
HEALTH_INDICATOR	—	−0.36	3.6
		(−25)	
HOUSEHOLD_INC	—	−0.205	2.8
		(−20)	
ρ	0.41 (7.5)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.366	$+\infty$	—
	(7.8)		
λ_3	$+\infty$	—	—

Source : SIP Survey, 8 667 observations.

Table 6: Occupational status versus activity limitations, total population

	1t equation *OCCUP*	2d equation *SELF_PERC_HEALTH*	Average
Frequency (%)			
1	15	1	
2	7.8	4.3	
3	77	21	
4	—	48	
5	—	25	
Coefficients			
SELF_PERC_HEALTH	0.406	—	—
	(15)		
OCCUP	—	0.183	—
		(1.4)	
AGE	−0.297	−0.157	2.1
	(−14)	(−3.9)	
MALE	0.169	−0.111	0.45
	(7.4)	(−2.6)	
INTERCEPT	−0.377	2.54	—
	(−5.1)	(12)	
NR_CHILDREN	−0.0299	—	2.8
	(−3.6)		
COUPLE	−0.116	—	0.72
	(−4.5)		
PROF_SAFETY	0.347	—	3
	(12)		
EDUCATION	0.0758	—	2.6
	(4.4)		
CAREER_SPELLS	−0.008 55	—	2.7
	(−0.96)		
INFANCY	—	−0.173	0.26
		(−8.2)	
DEPRESSION	—	−0.516	0.22
		(−12)	
HEALTH_INDICATOR	—	0.191	3.6
		(8.7)	
HOUSEHOLD_INC	—	0.178	2.8
		(8.9)	
CHRONIC_DISEASES	—	−0.593	1.3
		(−15)	
ρ	−0.51 (−3.2)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.352	0.692	—
	(7.8)	(10)	
λ_3	$+\infty$	1.9	—
		(18)	
λ_4	—	3.4	—
		(21)	
λ_5	—	$+\infty$	—

Source : SIP Survey, 8 667 observations.

Table 7: Occupational status versus self perceived health with chronic diseases control, total population

	1t equation *OCCUP*	2d equation *SELF_PERC_HEALTH*	Average
Frequency (%)			
1	19	0.88	
2	8.6	4.5	
3	73	22	
4	—	48	
5	—	24	
Coefficients			
SELF_PERC_HEALTH	0.387	—	—
	(16)		
OCCUP	—	0.139	—
		(7.1)	
AGE	−0.225	−0.164	2.1
	(−12)	(−15)	
INTERCEPT	−0.689	1.34	—
	(−7.2)	(15)	
NR_CHILDREN	−0.0763	—	2.9
	(−6.5)		
COUPLE	−0.206	—	0.69
	(−7.5)		
PROF_SAFETY	0.416	—	2.8
	(26)		
EDUCATION	0.073	—	2.7
	(5.9)		
CAREER_SPELLS	0.0698	—	2.7
	(6.7)		
INFANCY	—	−0.22	0.27
		(−13)	
DEPRESSION	—	−0.55	0.27
		(−25)	
HEALTH_INDICATOR	—	0.34	3.6
		(27)	
HOUSEHOLD_INC	—	0.172	2.7
		(20)	
ρ	−0.46 (−12)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.349	0.727	—
	(13)	(9.7)	
λ_3	$+\infty$	1.93	—
		(21)	
λ_4	—	3.4	—
		(28)	
λ_5	—	$+\infty$	—

Source : SIP Survey, 4 748 observations.

Table 8: Occupational status versus self perceived health, women

	1t equation *OCCUP*	2d equation *CHRONIC_DISEASES*	Average
Frequency (%)			
1	19	68	
2	8.6	32	
3	73	—	
Coefficients			
CHRONIC_DISEASES	−0.135	—	—
	(−11)		
OCCUP	—	0.163	—
		(6.3)	
AGE	−0.305	0.157	2.1
	(−18)	(9.3)	
INTERCEPT	0.337	2.64	—
	(4.2)	(28)	
NR_CHILDREN	−0.089	—	2.9
	(−6.5)		
COUPLE	−0.112	—	0.69
	(−4.1)		
PROF_SAFETY	0.493	—	2.8
	(44)		
EDUCATION	0.0933	—	2.7
	(6.9)		
CAREER_SPELLS	0.0132	—	2.7
	(0.95)		
INFANCY	—	0.208	0.27
		(7.4)	
DEPRESSION	—	0.222	0.27
		(7.7)	
HEALTH_INDICATOR	—	−1.01	3.6
		(−48)	
HOUSEHOLD_INC	—	−0.0519	2.7
		(−3.7)	
ρ	−0.11 (−2.9)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.364	$+\infty$	—
	(12)		
λ_3	$+\infty$	—	—

Source : SIP Survey, 4 748 observations.

Table 9: Occupational status versus chronic diseases, women

	1t equation *OCCUP*	2d equation *SELF_PERC_HEALTH*	Average
Frequency (%)			
1	10	1.1	
2	6.9	4	
3	83	20	
4	—	49	
5	—	26	
Coefficients			
SELF_PERC_HEALTH	0.619 (13)	—	—
OCCUP	—	−0.161 (−0.79)	—
AGE	−0.347 (−11)	−0.354 (−4.9)	2.1
INTERCEPT	−0.184 (−1.2)	1.27 (12)	—
NR_CHILDREN	0.0551 (3.8)	—	2.7
COUPLE	0.009 09 (0.19)	—	0.76
PROF_SAFETY	0.223 (9.9)	—	3.3
EDUCATION	0.002 55 (0.067)	—	2.5
CAREER_SPELLS	−0.0882 (−4.7)	—	2.7
INFANCY	—	−0.21 (−7.6)	0.24
DEPRESSION	—	−0.592 (−10)	0.17
HEALTH_INDICATOR	—	0.468 (9.5)	3.6
HOUSEHOLD_INC	—	0.269 (8.3)	2.9
ρ	−0.39 (−2.2)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.36 (13)	0.728 (9.6)	—
λ_3	$+\infty$	1.89 (18)	—
λ_4	—	3.35 (19)	—
λ_5	—	$+\infty$	—

Source : SIP Survey, 3 919 observations.

Table 10: Occupational status versus self perceived health, men

	1t equation *OCCUP*	2d equation *CHRONIC_DISEASES*	Average
Frequency (%)			
1	10	71	
2	6.9	29	
3	83	—	
Coefficients			
CHRONIC_DISEASES	−0.415	—	—
	(−16)		
OCCUP	—	0.647	—
		(15)	
AGE	−0.481	0.466	2.1
	(−18)	(21)	
INTERCEPT	1.43	1.54	—
	(14)	(13)	
NR_CHILDREN	0.0453	—	2.7
	(3)		
COUPLE	0.176	—	0.76
	(4.8)		
PROF_SAFETY	0.307	—	3.3
	(18)		
EDUCATION	0.0471	—	2.5
	(2.8)		
CAREER_SPELLS	−0.195	—	2.7
	(−13)		
INFANCY	—	0.198	0.24
		(6.5)	
DEPRESSION	—	0.23	0.17
		(6)	
HEALTH_INDICATOR	—	−0.98	3.6
		(−42)	
HOUSEHOLD_INC	—	−0.169	2.9
		(−10)	
ρ	−0.29 (−6)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.388	$+\infty$	—
	(15)		
λ_3	$+\infty$	—	—

Source : SIP Survey, 3 919 observations.

Table 11: Occupational status versus chronic diseases, men

	1t equation *OCCUP*	2d equation *SELF_PERC_HEALTH*	Average
Frequency (%)			
1	8.4	0.4	
2	8.2	2.4	
3	83	16	
4	—	50	
5	—	31	
Coefficients			
SELF_PERC_HEALTH	0.502	—	—
	(17)		
OCCUP	—	0.111	—
		(6.2)	
AGE	0.0205	−0.0293	38
	(8.7)	(−24)	
MALE	0.344	−0.0182	0.45
	(9.6)	(−1.1)	
INTERCEPT	−1.75	1.96	—
	(−19)	(20)	
NR_CHILDREN	−0.121	—	2.6
	(−8.4)		
COUPLE	−0.008 79	—	0.73
	(−0.26)		
PROF_SAFETY	0.398	—	2.8
	(19)		
EDUCATION	0.0827	—	2.8
	(5.2)		
CAREER_SPELLS	0.005 45	—	2.6
	(0.39)		
INFANCY	—	−0.216	0.25
		(−12)	
DEPRESSION	—	−0.485	0.21
		(−22)	
HEALTH_INDICATOR	—	0.371	3.7
		(23)	
HOUSEHOLD_INC	—	0.189	2.8
		(20)	
ρ	−0.54 (−13)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.498	0.613	—
	(17)	(7.2)	
λ_3	$+\infty$	1.79	—
		(16)	
λ_4	—	3.31	—
		(28)	
λ_5	—	$+\infty$	—

Source : SIP Survey, 3 982 observations.

Table 12: Occupational status versus self perceived health, 30-44 old

	1t equation *OCCUP*	2d equation *CHRONIC_DISEASES*	Average
Frequency (%)			
1	8.4	76	
2	8.2	24	
3	83	—	
Coefficients			
CHRONIC_DISEASES	−0.18 (−9.8)	—	—
OCCUP	—	0.311 (12)	—
AGE	0.008 85 (3.4)	−0.000 588 (−0.24)	38
MALE	0.401 (10)	−0.28 (−8.5)	0.45
INTERCEPT	−0.173 (−2.7)	3.09 (49)	—
NR_CHILDREN	−0.125 (−7.4)	—	2.6
COUPLE	0.138 (3.7)	—	0.73
PROF_SAFETY	0.511 (32)	—	2.8
EDUCATION	0.107 (6.7)	—	2.8
CAREER_SPELLS	−0.086 (−5.1)	—	2.6
INFANCY	—	0.183 (5.7)	0.25
DEPRESSION	—	0.172 (5)	0.21
HEALTH_INDICATOR	—	−1.09 (−46)	3.7
HOUSEHOLD_INC	—	−0.0938 (−5.6)	2.8
ρ	−0.16 (−3.9)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.539 (18)	$+\infty$	—
λ_3	$+\infty$	—	—

Source : SIP Survey, 3 982 observations.

Table 13: Occupational status versus chronic diseases, 30-44 old

	1t equation *OCCUP*	2d equation *SELF_PERC_HEALTH*	Average
Frequency (%)			
1	20	1.5	
2	7.5	5.9	
3	72	26	
4	—	47	
5	—	20	
Coefficients			
SELF_PERC_HEALTH	0.446 (17)	—	—
OCCUP	—	0.197 (2.4)	—
AGE	−0.0912 (−34)	0.003 51 (0.44)	52
MALE	0.0897 (3)	−0.184 (−11)	0.45
INTERCEPT	3.27 (27)	0.754 (1.7)	—
NR_CHILDREN	0.0188 (1.7)	—	3
COUPLE	−0.192 (−6.6)	—	0.72
PROF_SAFETY	0.349 (9.7)	—	3.2
EDUCATION	0.0688 (5.3)	—	2.4
CAREER_SPELLS	0.027 (2.7)	—	2.8
INFANCY	—	−0.193 (−8.4)	0.26
DEPRESSION	—	−0.586 (−11)	0.23
HEALTH_INDICATOR	—	0.336 (12)	3.5
HOUSEHOLD_INC	—	0.188 (10)	2.8
ρ	−0.55 (−7)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.29 (12)	0.836 (11)	—
λ_3	$+\infty$	2.07 (18)	—
λ_4	—	3.53 (23)	—
λ_5	—	$+\infty$	—

Source : SIP Survey, 4 685 observations.

Table 14: Occupational status versus self perceived health, 45-59 old

	1t equation *OCCUP*	2d equation *CHRONIC_DISEASES*	Average
Frequency (%)			
1	20	63	
2	7.5	37	
3	72	—	
Coefficients			
CHRONIC_DISEASES	−0.213 (−16)	—	—
OCCUP	—	0.212 (8.1)	—
AGE	−0.0996 (−52)	0.0392 (20)	52
MALE	0.0292 (1.1)	0.0648 (2.5)	0.45
INTERCEPT	4.66 (63)	0.707 (8.5)	—
NR_CHILDREN	0.008 31 (0.62)	—	3
COUPLE	−0.0705 (−2.4)	—	0.72
PROF_SAFETY	0.445 (35)	—	3.2
EDUCATION	0.0875 (6.1)	—	2.4
CAREER_SPELLS	−0.0453 (−3.2)	—	2.8
INFANCY	—	0.223 (7.9)	0.26
DEPRESSION	—	0.274 (9.7)	0.23
HEALTH_INDICATOR	—	−0.931 (−50)	3.5
HOUSEHOLD_INC	—	−0.0703 (−4.9)	2.8
ρ	−0.037 (−0.99)		—
Thresholds			
λ_0	$-\infty$	$-\infty$	—
λ_1	0	0	—
λ_2	0.309 (14)	$+\infty$	—
λ_3	$+\infty$	—	—

Source : SIP Survey, 4 685 observations.

Table 15: Occupational status versus chronic diseases, 45-59 old