Sporting Activity, Employment Status and Wage

Thierry Kamionka

Sporting Activity, Employment Status and Wage

Thierry Kamionka[*]

This version : July 2021

Abstract

We propose a structural model of participation to sporty activities and labour supply. We jointly model employment, wage and sporting activity using a dynamic model. We estimate for the period going from 1994 to 1999 a dynamic multivariate model with random effects using the German Socioeconomic panel (GSOEP). The error terms of the equations of the model can be correlated. Each of these error terms can be auto correlated allowing shocks on one of the components of the model to have an impact on all the error terms of the model the next periods. Individual effects, one for each equation, can be correlated. The model is estimated using simulated maximum likelihood estimator. The initial conditions problem is taken into account.

Keywords : Sporting activity, Employment, Wage, Heterogeneity, Simulation based estimation, Panel data.

JEL Classification : J21, Z22, J31, C33, C35.

[*] CNRS and CREST, Institut Polytechnique de Paris (kamionka@ensae.fr). Phone : 33 1 70 26 68 12. Address: CREST. 5, avenue Henry Le Chatelier. 91764 Palaiseau Cedex. France.
I would like to thank Investissements d'Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), participants of the Journées de Microéconomie Appliquées (JMA, Casablanca), 25th International Panel Data Conference (IPDC), François Gardes, Louis Lévy-Garboua, an anonymous referee and an associate editor. I would like to thank the German Institute for Economic Research (DIW Berlin) who made available the access to Socioeconomic Panel Study (SOEP).

Abstract

Nous proposons un modèle structurel de décision de pratiquer un sport et d'offre de travail. Nous modélisons conjointement l'emploi, le salaire et la pratique d'un sport à l'aide d'un modèle dynamique. Nous estimons, pour la période allant de 1994 à 1999, un modèle dynamique multivarié avec des effets individuels aléatoires à partir de données de panel allemande (GSOEP). Les termes d'erreur des équations du modèle peuvent être corrélés. Chacun de ces termes d'erreur peut être auto-corrélé permettant à un choc sur l'une des composantes du modèle d'avoir un impact sur tous les termes d'erreur du modèle les periodes suivantes. Les effets individuels, un pour chaque équation, peuvent être corrélés. Le modèle est estimé à l'aide d'un estimateur du maximum de vraisemblance simulé. Le problème des conditions initiales est pris en compte.

Keywords : Pratique d'un sport, Emploi, Salaire, Hétérogénéité, Estimation du maximum de vraisemblance simulé, Donées de Panel.

JEL Classification : J21, Z22, J31, C33, C35.

1 Introduction

Physical activities have positive health effects as they can help to reduce the risk of several non-communicable and age-related diseases (see Reiner, Niermann, Jekauc and Woll, 2013). Using panel data, Lechner (2009) shows that sports activities have positive effects on health. A better health may be associated to a higher productivity. Consequently, sporty people may have higher earnings and wages. This effect of physical activity has been empirically studied in the econometric literature. For instance Rooth (2011) studies the return to leisure sports in the job hiring process. He finds the evidence of a sports premium when he considers the long-run impact of physical fitness on earnings. Lechner (2009) and Lechner and Sari (2014) show that sports activity have positive long-term effects on earning and wage. Lechner and Downward (2013), for England, find evidence of higher earnings opportunities associated to sports participation. Using data for Finland, Hyytinen and Lahtonen (2013) show that physical activity has a positive impact on the long-term income. Cabane and Clark (2015) find a correlation between childhood sports and both managerial responsibilities and autonomy at work when individuals become adults.

Sports participation may have other effects than health effects. For instance, it has been underlined that sporty people may have more non cognitive skills (see Cabane, 2014). Non cognitive skills consist in personality traits like team skills, tenacity, self-confidence, discipline, competitive spirit, perseverance, motivation, self-esteem, forward-looking behavior. Cognitive and non cognitive skills may have an impact on wages, schooling, health, occupational choice, crime behavior and success on the labour market (Cunha and Heckman, 2009). At least some of non cognitive skills are associated to the practice of sports. For instance, for preparation of a sport competition you need to train yourself for month or years in order to be likely to obtain success later (you are forward looking and persevering). The non cognitive skills associated to the practice of a given sport are observable by firms after hiring. These personal traits can be used in employment for instance when employees work within a team. Almost no empirical studies consider the impact of sport participation on the transition into employment (see Cabane, 2014, for a study of unemployment duration and Rooth, 2011, for the impact of sport skill on recall rate). No empirical study documents the employment dynamics according to the sport participation decision.

An other important effect of sports practice consists in the fact that athletes may have access to a large and heterogeneous social network. Recently, Kramarz and Skans (2014) using a Swedish population-wide linked employee-employer data set, show that strong social ties - family ties - are an important factor for selection in the firms where the young workers find their first job. They show that these strong social ties can favor both the access to jobs and labour market outcomes a few years after entry. Singularly, workers who find their entry jobs using strong social ties experience a better wage growth and are more likely to remain in their first job. People practicing sports have a larger social network and can have access to more information. The size of the network may vary according to sports and, consequently, the effects of these social ties constructed during sport practice may be heterogeneous on the labour market outcomes. These effects may differ according to education level of job seekers and according to economic conjuncture. The size of the network associated to the practice of sport may have an impact on the quality of the employer-employee match and, finally, on the conditional probability of separation.

In this paper we model jointly sport participation, labour market transitions and wages. The model incorporates unobserved heterogeneity terms that allow to take into account self selection in the practice of sports. The individual random effects can be correlated across the equations of the model for a single individual. The idiosyncratic terms of the model can be correlated. The error terms can be auto correlated. The model is estimated using panel data from the german SOEP for the period going from 1994 to 1999. The initial conditions are taken into account. These characteristics allow to circumvent the limits of the previous studies realized on micro economic data. As the context is intrinsically multi spells we model the past selection mechanism using an adequate treatment of the initial conditions. The model we use is dynamic and allows to take into account true and spurious state dependence. As the model we use is dynamic we can distinguish different situations in term of sport participation and position on the labor market. For instance, the behavior of an unemployed sporty and an unemployed not practicing sport can be easily and fruitfully distinguished using the past realization of the process. It is also possible to isolate the behavior of employed workers according to sport participation. The impact of the sport participation on the reemployment wage is estimated. We consider a correlated random effect specification that allows to test the null assumption that random effects - and error terms of the model - are uncorrelated with sportiness.

In order to shed light on the economic mechanisms at play behind the empirical results we consider a two periods economic model. The model is such that individual can consume goods, leisure, health care and practice sport under uncertainty. The individual can determine the level of effort he devotes to its job. The demand for health care may depend on the health state. In the first place, the uncertainty comes both from the probability to remain employed and to be promoted. Moreover, the uncertainty comes from the conditional distribution of health on the second period. We examine a context such that individuals practicing sport can modify the conditional distribution of health states. Indeed, as the physical activity can have a positive effect on health, sport practice can reduce the occurrence of ill health.

The main contribution of this paper is threefold. First, we use a reduced form model to study the sport participation decision and labour market transitions. We consider a non linear and dynamic panel data model. The econometric modeling allows to consider the interaction between past position on the labour market and decision to participate in sport. It is the only study that allows to obtain empirical evidence of the impact of sport participation on the transition from non employment to employment and on job stability. Second, we consider the impact of the sport decision on the reemployment wage. There is no empirical evidence in the literature on this relation. It is important to determine whether the specific characteristics of sporty individuals are rewarded by employers. We analyze empirically the effect of sport participation on the annual wage. Doing so we contribute on an aspect of the subject such that empirical evidence is scarce. Third, we propose a structural economic model for the participation to sport activities and labour supply. We consider a utility maximizing consumer and a two period utility function. The consumer maximizes its utility under the budget constraints and a health production function. The individual considers an a priori on the distribution of health states. This distribution depends on the choice to participate or not to sporting activities. Sport participation decision is the result of the comparison of expectation of the indirect utility with respect to the distribution of health.

The paper plan is the following one. The econometric model is presented in section 2. The data set we use is described in section 3. The estimation results are presented in section 4. A structural economic model is considered in section 5. The last section concludes.

2 Econometric Analysis

Let us consider a dynamic model for sporting activity (s), employment (e) and wage (w). Let us consider a sample of n individuals. Hereafter, let t denote the index of time and let i denote the index of the individual in the sample. Let x_{jit}, $j \in E = \{s, e, w\}$, denote a vector of individual characteristics that can include marital status, citizenship, education level, age and gender. β_j is a vector of parameters associated to observed heterogeneity in equation j, $j \in E$. δ_j, $j \in E$, are vectors of parameters associated to past realizations of the endogenous variables ($\delta_{jk} \in \mathbb{R}$).

The latent dependent variable y^*_{jit} is given by

$$y^*_{jit} = x'_{jit}\,\beta_j + z_j(y_{it-1}, y_{it})'\,\delta_j + r_{jit}, \tag{1}$$

for any $j \in \{s, e\}$, $t = 1, \ldots, T$ and $i = 1, \ldots, n$. $y_{it} = (y_{sit}, y_{eit}, y_{wit})' \in \mathbb{R}^3$

is the realization of the dependent variables at time t and for individual i. $z_j(.)$ is a vector depending on the realizations of the lagged values of the dependent variables. We assume that $z_j(.)$ can depend on the current value of y_{it}.

For individual i at time t, the decision j, $j \in \{s, e\}$, is a binary variable and can be written as

$$y_{jit} = \mathbb{1}[y^*_{jit} > 0], \quad (2)$$

where $\mathbb{1}[.]$ is an indicator of the event between brackets which is equal to 1 if the event occurs and zero otherwise.

The log of the wage at time t is

$$y_{wit} = x'_{wit}\,\beta_w + z_w(y_{it-1})'\,\delta_w + r_{wit}, \quad (3)$$

where δ_w and β_w are vectors of parameters. $z_w(.)$ is a vector depending on the lagged realization of the dependent variable y_{it}.

In the equation j the error term is

$$r_{jit} = \alpha_{ij} + u_{jit},$$

where α_{ij} is an unobserved heterogeneity component, an individual random effect.

Let u_{jit} denote an error term specific to the equation j, $j \in E$. Let us assume that $u_{jit} \perp\!\!\!\perp x_i$. u_{jit} is assumed to be independent of $u_{j'i't'}$, if $i \neq i'$. We assume that the distribution of the error term u_{jit} has an autoregressive structure. For instance, Hyslop [1999] makes such an assumption in a context of a single equation model. This assumption allows to evaluate the impact of a shock, for a given period, on the behavior of the individual the next period of time. The stochastic processes of error terms u_{jit}, $j \in E$, are assumed to be stationary ($t > 0$).

The random effects α_{ij}, for all $i \in \{1, \ldots, n\}$ and $j \in E$, are independent and identically distributed. The individual effect α_{ij} is assumed to be independent of the observable characteristics x_i and distributed as a normal random variable with mean zero and variance $\sigma^2_{\alpha_j}$, $j \in E$.

We adopt an autoregressive structure for the error terms for the observation periods after the initial time

$$u_{jit} = \rho_j\, u_{jit-1} + \epsilon_{jit}, \quad (4)$$

where $\alpha_{ij} \perp\!\!\!\perp \epsilon_{j'it}$, for all $j, j' \in E$. $\epsilon_{jit} \perp\!\!\!\perp u_{j'it'}$, for all $t' < t$ and $j' \in E$. Let us assume that

$$\epsilon_{jit} \ \sim\ \mathrm{N}(0, \sigma^2_{\epsilon_j}). \quad (5)$$

Let $\rho_{\alpha_j\alpha_k}$ denote the correlation between the random effects α_{ij} and α_{ik} specific, respectively, to equation j and to equation k, $j, k \in E$. Let ρ_{jk} denote the correlation between the idiosyncratic terms ϵ_{jit} and ϵ_{kit}, for all $t = 1, \ldots, T$ and $i \in \{1, \ldots, n\}$.

For two equations of the model, this specification can be considered as a dynamic probit with random effect. The dependent variables corresponding to sport practice and employment are dichotomic. Therefore, we have to normalize the variance of the corresponding residuals

$$\sigma^2_{\alpha_j} + \sigma^2_{u_j} = 1,$$

and

$$\text{var}(r_{ji0}) = \sigma^2_{j0} = 1, \text{ for } j = s, e.$$

We can show that

$$var(u_{jit}) = \sigma^2_{u_j} = \frac{\sigma^2_{\epsilon_j}}{(1 - \rho^2_j)},$$

where $j \in E$, $t > 1$.

As the first observation time does not correspond to the starting time of the process, we cannot consider that the initial state $y_{i0} = (y_{si0}, y_{ei0}, y_{wi0})'$ is independent of the individual effect $\alpha_i = (\alpha_{is}, \alpha_{ie}, \alpha_{iw})'$.

Wooldridge [2005] proposes to consider the distribution of the random effect $\alpha'_i = (\alpha_{is}, \alpha_{ie}, \alpha_{iw})$ conditionally to the state y_{i0} and, possibly, given a set of exogenous explanatory variables.

We are going to assume that the conditional distribution of α_{ij} is a normal distribution:

$$\alpha_{ij} \mid y_{i0}, x_i \sim N(\lambda_{j0} + y_{si0}\,\lambda_{js} + y_{ei0}\,\lambda_{je}, \sigma^2_{\alpha_j}), \tag{6}$$

where λ_{j0}, λ_{js} and λ_{je} are some real parameters to be estimated. In practice, the constant λ_{j0} cannot be identified separately from a constant located in the vector β_j. Then, without lost of generality, we can fix $\lambda_{j0} = 0$ and consider that the vector β_j includes a constant.

The contribution of an individual i to the conditional likelihood function is

$$L_i(\theta) = \int_{A_i} \phi(r; \Omega)\, dr, \tag{7}$$

where $\phi(.;\Omega)$ is the probability density function of a normal distribution with mean zero and variance-covariance matrix Ω. θ is a vector of parameters and $\Omega = var(R \mid x)$ where R is the vector of residuals on the sample (see Kamionka and Lacroix, 2018).

As the contribution to the likelihood function (7) cannot be calculated analytically, then it is simulated. The method consists in replacing a contribution to the likelihood by an average based on random draws r_i^h, $h = 1, \dots, H$ (see Kamionka and Lacroix 2018, Appendix E). These random draws are independent and are realized in such a way that we have no rejection.

The main difficulty, here, is to write the algorithm in the case where the endogenous variables combine qualitative variables like sport practice and a continuous variable like wage (see Chang, 2009).

The simulated maximum likelihood estimator of θ can be then obtained. The SML estimator is consistent and efficient if $\frac{\sqrt{N}}{H} \longrightarrow 0$ when $N \longrightarrow +\infty$ and $H \longrightarrow +\infty$ (cf., for instance, Gouriéroux and Monfort [1997]; Kamionka [1998]; Edon and Kamionka [2008]; Gilbert, Kamionka and Lacroix [2011]; Kamionka and Lacroix [2018]; Kamionka and Vu [2016]).

In practice, the number of draws are fixed to $H = 30$. Several authors have underlined that the SML estimator is close of the consistency for a number of draws H relatively low (about 30). Some specifications were estimated for higher values of H without modifying appreciably the estimation results (cf., also, Kamionka and Lacroix [2008]).

3 Data set

The data we use in this study are part of the German Socio-Economic Panel (SOEP). It is household panel study centered on the study of life course and weel-being (see Wagner et alii, 2007). This survey started in West Germany in 1984 and was designed to be representative of the adult population living in private households but oversample groups of immigrants. At the beginning, the sample size was slightly larger than 12000 adults. After the fall of the Berlin's wall, the East German sample was introduced in June 1990[1]. The sample we use covers the period going from 1984 to 2013 (SOEP version V33.1). All adult members of the households are interviewed (17 years and more). In order to study specific groups of household and to refresh the sample, new sub-samples were added since 1984. In 2003, information on newborn started to be collected.

Soep contains several samples. Sample A consists in residents of Western Germany (collected since 1984). Sample B consists in foreigners in Western Germany (collected since 1984). Sample C was added since 1990 and consists in East Germany households. A sample of immigrants was added since 1995 (sample D). Sample E is a refreshment sample selected in the population of private household to compensate attrition (collected since 1998). The individuals can exit the Soep

[1]Official German unification occured on July 1st, 1990.

because they decide to stop to participate or they die or they move abroad. An individual can enter the Soep, for instance, if she/he moves in an existing Soep household.

The SOEP gather information about all individuals of a given households aged 17 and more. The Soep allows to follow these individuals over time. The members of the household are all surveyed the next years. They are surveyed the next years even if they move from their original household. They are surveyed even after a residential mobility. Soep collects information on education, income, employment, health, housing, life satisfaction.

We use the information from the annual waves from 1994 to 1999. We do not consider other waves because they do not provide information on sporting activity or this information is not provided over a sufficiently long period of observation. We drop the observations with missing values on the main variables. The sample contains 13068 individuals in 1994. All individuals in our sample are observed at least two consecutive years. It is an unbalanced panel.

The percentage of employed individual in 1994 observed on our sample (46.6%, see Table 1) is lower than the one observed in november 2018 for Germany (54.1%)[2]. Participation in sport is equal to 1 if the individual declares to practice this activity every week. 20.72% of individuals declare to participate actively in sport.

We use the international standard classification of education (ISCED 1997) to define the level of education. 3.68% of the individuals (see table 1) have a post-secondary non-tertiary education (Vocational plus Abitur). 19.63% have a First stage or a second stage of tertiary education (higher education or higher vocational).

The sample includes 51.7% of women in 1994 (50.8% of the German population in 2017, World Bank) and 69.56% of the individual in the sample were living in 1994 in the part of Germany corresponding to West Germany (excluding West Berlin).

16.74% of the sample in 1994 have other citizenship (in 2012, 92% of residents in Germany have German citizenship, Federal Statistical Office of Germany). We have already noted that, since 1984, a part of the sample (Sample B) consists in foreigners in Western Germany and that Soep oversample groups of immigrants. 65.84% of the individuals in the sample are married. Only 4.71% of the sample have less than 20 years old. This is due to the fact that only adult members of the households are interviewed.

The proportion of people practicing a sporting activity among employed individuals is larger than the one among unemployed individual (21.7% vs. 19.9 %, see table 2). However, among younger individuals (less than 50 y.o.), the pro-

[2]Tradingeconomics.com, Bundesagentur fur arbeit, Germany.

portion of people practicing a sporting activity is much larger among unemployed individuals (28.2% vs. 23.9%). Among older individuals, we observe the converse phenomenon, the proportion of people practicing a sporting activity is larger among employed individual. Students are frequently practicing a sporting activity regularly. Practicing a sporting activity may be frequently costly (tennis, horse riding, squash, climbing, skiing) and this could explain why the practice of sport is more frequent among employed individuals.

4 Results

The results are presented in table 3. Five specifications are estimated. All specification incorporate the lagged values of employment and sporting activity in the sporting activity equation. The first one (1) incorporates the lagged value of employment in the equation governing employment. The second specification (2) includes the impact of the lagged value of sporting activity is distinguished according to position on the labour market in the employment and wage equations. The third specification (3) includes a correlated random effect includes (CRE) in the employment and wage equations. This specification allows to test the endogeneity of the sporting activity. In 1998 and 1999 participants to German Soep Panel were asked wether or not they smoked tobacco. We use this information as control of sport participation decision (specification (4)). We use the variable "worried about environment" in order to instrument sport participation in an additional specification which includes a correlated random effect (specification (5)).

Let us consider the results that are common to all the specifications. We distinguish the effects of sporting activity on the various outcomes from the effects of other variables.

4.1 The effects of sporting activity

For the sporting activity (see panel 1 of table 3), we obtain a positive state dependence for sporting activity: individual practicing sport a given year are more likely to practice sport the next year. The initial conditions are significantly different from zero. Singularly, an individual that is initially practicing sport is more likely to be practicing sport afterwards.

For employment (see panel 2 of table 3), the current practice of sport has no significant impact on the conditional probability to be employed. Among unemployed individuals, athletic people are not more likely to be reemployed the next year. It is consistent with the assumption that to declare a sport practice on a cv has no significant impact on the job offer arrival rate. Indeed, it is no easy for a potential

Table 1: **Sample characteristics**

	First observation 1994
Dependent variables	
Athletic	20.68
Employed	46.59
Education level	
Low educ.	76.69
Abitur	3.68
Higher educ.	19.63
Gender	
Women	51.65
Men	48.35
Urban area	
West Germany	69.56
Baden-Wuerttemberg	13.65
Bavaria	12.51
Berlin	4.23
Brandenburg	4.48
Bremen	0.73
Hamburg	1.14
Hesse	7.06
Mecklenburg-Western Pomerania	3.05
Lower Saxony	7.95
North Rhine-Westphalia	19.13
Rhineland-Palatinate	5.26
Saarland	0.00
Saxony	8.46
Saxony-Anhalt	5.26
Schleswig-Holstein	2.13
Thuringia	4.96
Elements of biography	
Age ≤ 19	4.71
$20 \leq$ Age ≤ 29	21.32
$30 \leq$ Age ≤ 39	21.90
$40 \leq$ Age ≤ 49	16.87
$50 \leq$ Age	35.20
Married	65.84
Citizenship	
German	83.26
Other Citizenship	16.74
Number of individuals	13068

Note : GSOEP 1994-1999. Percentages.

Table 2: **Proportion of sporties in the sample (1994)**

Not Employed	Employed
Age < 50	
28.19%	23.91%
Age $\geq$ 50	
10.35%	13.66%
Sample	
19.89%	21.66%

Note : GSOEP 1994-1999.

employer to verify to what corresponds such an indication (an occasional or regular practice, an intense or superficial practice). It also indicates that if the practice of a sport makes it possible to weave a more extensive social network, these relationships are not strong enough to be useful in the research of a work. In the four last specifications, people who were already employed the previous year and practicing sport (athletic) are more likely to be currently employed compared to individuals who were already employed last year but not practicing a sport. Sporting Activity may be a source of non cognitive capabilities. The importance of such capabilities to determine success in life have been underlined in the literature (see Cunha and Heckman, 2007, Cunha, Heckman, and Schennach, 2008). It corresponds to qualities: team spirit (in some cases), sense of effort, taste of the challenge, search for performance, that sporty individuals are alone in competition (sometimes). These sporting qualities can be useful in work. These qualities of athletes are observed by employers. Individuals who are practicing sport are more likely to be more stable in employment.

For the three last specifications (see panel 2 of table 3), we have considered a correlated random effect (CRE) specification. In such a specification, the average over the period of time of the sporting activity is introduced as explanatory variable in the employment probability. It allows to test the endogeneity of the sporting activity. The impact of the attitude with respect to sport is positive and not significantly different from 0.

For the log of the wage (see panel 3 of table 3), in the four last specifications, among unemployed individuals, sporty people have a lower mean reemployment wage. This characteristics can be explained by a lower reservation wage for sporty individuals. Among individuals who were previously employed last year, people practicing sports have a larger conditional expectation of the log of the wage. Individuals practicing sports regularly have non cognitive skills that can be observed and rewarded by employers. This can be also a wage premium for a lower job to job mobility (existence of a wage tenure contract rewarding large seniority).

In all specifications, people who were initially practicing sport are more likely to have a higher conditional expectation of the wage.

In the three last specifications (3) to (5) we consider a correlated random effect specification (CRE). The impact of the average sporting practice on the conditional expectation of the wage is not statistically significant. As already noted, the impact of the average sporting practice on the conditional employment probability is not statistically significant. This specification allows to test the endogeneity of sport practice. The results provide evidence for the null assumption that random effects are uncorrelated with sportiness. The estimates of the parameters related to sport practice are similar in the specifications (2) to (5) of Table 3.

4.2 The effects of other individual characteristics

For the sporting activity (see panel 1 of table 3), the impact of diploma is increasing with the level of the diploma. Individual with a higher level of diploma have accumulated larger human capital and have larger expected earnings. The increase of sporting activity with the expected income can be the consequence of several effects. It can be the consequence of an income effect (assuming sport is a normal good)[3]. It can be the result of individual preferences (people with higher education have specific attitude towards risk) or a better access to information (sport would help stay healthy). It can be due, for instance, to the fact that higher education institutions promote practice of sport.

Married individuals are less likely to practice sport. Indeed, people living in couple are more likely to have children. In this case it can be more difficult to have a regular practice of sport. This results is similar to the one obtained for women participation to labour market that this less frequent in presence of a young children (see Hyslop 1999). The unemployment rate has a large and positive impact on sporting activity indicating that working and sporting activities are linked because individuals must choose the amount of time they allocate to leisure or work.

The impact of age on the sporting activity is decreasing with age indicating that the health status of individuals can deteriorate with age. Sport practice is more difficult when people are not healthy. Sport performance may decrease with age. Foreigners are less likely to practice sport but the effect is not statistically significant all other things being equal. Sport participation is more likely in the western part of Germany. This can be due to a consequence of the lower number of inhabitants in former GDR lander (152 inhabitants per km2 in east Germany

[3]There is some evidence at the macroeconomic level that sport practice increases with GDP (see Dawson and alii, 2009).

in 1990, 252 inhabitants per km2 in west Germany and 222 in Germany in 1990, see Heilig et al., 1990). Indeed, the proximity of sports facilities is a growing function of inhabitants density. In the two last specifications (4) and (5), Berliners are more likely to participate in sport activities. Smokers are less frequently sporty (specification (4)). Indeed, it is often noted that smoking behaviour is associated to a decreased respiratory capacity.

For employment (see panel 2 of table 3), the impact of diploma is increasing with the level of the diploma. This is consistent with the neoclassic labour supply model. Married individuals are more likely to be employed. This result can be explained for instance by a larger search intensity. Previous position on the labour market has no significant impact on the current sporting activity. The impact of age is increasing then decreasing: this is consistent with the evolution of the observed employment rate with age. For instance, in 2018, the employment rate is lower for younger (47.2%, 15-24 age group) and older (71.5%, 54-64 age group) compared to adults (85%, 25-54 age group)[4]. As expected, the conditional probability to be employed is decreasing with the unemployment rate. Women have a lower probability to be employed that can be explained partly by a lower participation to labour marker. Initial status is informative of unobserved heterogeneity distribution. We observe a diagonal effect for the conditional employment probability: employed individuals a given year are more likely to be employed the next year (see specification (1)). Indeed, the employment status at the previous time is informative of the employability which is correlated with the expected earning stream. When one lives in the part of Germany corresponding to the former West Germany she/he is more likely to be employed. It is a consequence of the difference of unemployment rates in these two parts of Germany.

In the three last specifications, unemployed people the previous year are less likely to be currently employed. The initial conditions are significantly different from zero. Consequently, the initial conditions are informative of unobserved heterogeneity. Singularly, an individual that is initially employed is more likely to be currently employed.

For the log of the wage (see panel 3 of table 3), the impact of diploma is increasing with the level of the diploma. This result is consistent with the presence of positive return to eduction. Foreigners have a lower conditional expectation of log of wage. This result may indicate the presence of a discrimination or a less developed professional network. The conditional expectation of the log of the wage is increasing with age before 50 y.o.: this results is consistent with a positive impact of the return to experience on the labour market. The impact of experience is lower for older individuals. The existence of a non linear impact of experience has been

[4] Source: OECD, 2018-Q4, https://data.oecd.org/emp/employment-rate-by-age-group.htm.

underlined in the literature (Dustmann and Meghir, 2005). Women have lower conditional expectation for wage. Married individuals have a higher conditional expectation of the log of the wage. For male, marriage premium has been already modeled. For instance, Bonilla and Kiraly (2013) propose a theoretical explanation of this premium based on search friction. Individuals living in the western part of Germany and in Berlin have a higher conditional expectation of the wage.

In all specifications, the initial conditions are informative of the unobserved heterogeneity in the equation of the log of wage. For instance, an individual that is initially employed is more likely to have a higher conditional expectation of wage compared to an individual who was not initially employed.

Individuals who were already previously employed have a larger conditional expectation of the log of wage: indeed, these individuals are more likely to beneficiate of the presence of positive returns to seniority.

4.3 Additional results

The error term of the employment and wage equations are positively correlated in all the specifications. The variance of the random effect specific to the employment status is relatively large indicating the presence of an important - unobserved - heterogeneity among workers (see panel 4). The variance of the random effect specific to the log of wage equation is relatively large indicating the presence of an important omitted heterogeneity.

The unobserved heterogeneity components specific to employment and wage are positively correlated in all the specifications. The unobserved random effects specific to the sporting practice and wage are positively and significantly correlated (specification (1) to (3)). The unobserved heterogeneity components of the sporting activity and employment equations are positively correlated. The error term of the sporting activity and employment equations are negatively autocorrelated implying that a shock a given a year is likely to affect the following year negatively. The error term of the wage equation is positively and significantly autocorrelated.

5 Economic Model

In order to understand the econometric results, we consider a two period economic model with the objective of analyzing the relationship between decisions of the individual on the labor market and the practice of sport [5].

[5] The model we propose is an extension of Cameron et alii (1988) to include the practice of sport, promotion, effort and firing.

Table 3: **Sports Practice, Employment and Wage**

	Wooldridge's approach - Sporting activity				
	(1)	(2)	(3) CRE	(4) CRE	(5) CRE
Constant	-1.6607***	-1.6578***	-1.6583***	-1.6296***	-1.8733***
	(0.0936)	(0.0935)	(0.0936)	(0.0977)	(0.0996)
Woman	0.0086	0.0108	0.0110	-0.0059	0.0004
	(0.0184)	(0.0184)	(0.0189)	(0.0219)	(0.0168)
Married	-0.1275***	-0.1275***	-0.1275***	-0.1370***	-0.1345***
	(0.0190)	(0.0190)	(0.0190)	(0.0199)	(0.0191)
Abitur	0.1027**	0.1004**	0.1011**	0.0978**	0.0934**
	(0.0423)	(0.0423)	(0.0424)	(0.0444)	(0.0422)
High	0.2412***	0.2386***	0.2372***	0.2237***	0.2464***
	(0.0213)	(0.0213)	(0.0214)	(0.0228)	(0.0214)
Foreign	-0.0444	-0.0436	-0.0427	-0.0836	-0.0347
	(0.0486)	(0.0485)	(0.0486)	(0.0530)	(0.0488)
From 20 to 29	0.0255	0.0243	0.0244	0.0244	0.0246
	(0.0223)	(0.0222)	(0.0223)	(0.0232)	(0.0224)
From 40 to 49	-0.1150***	-0.1158***	-0.1164***	-0.1046***	-0.1116***
	(0.0238)	(0.0238)	(0.0239)	(0.0247)	(0.0239)
From 50 and more	-0.3035***	-0.3031***	-0.3018***	-0.3266***	-0.2953***
	(0.0227)	(0.0227)	(0.0228)	(0.0239)	(0.0227)
West	0.1462***	0.1465***	0.1459***	0.2015***	0.1970***
	(0.0192)	(0.0192)	(0.0192)	(0.0218)	(0.0210)
Berlin	0.0548	0.0552	0.0545	0.1219***	0.0940**
	(0.0443)	(0.0443)	(0.0443)	(0.0473)	(0.0452)
Unemploy. rate	5.2887***	5.2011***	5.1892***	5.0618***	5.5541***
	(1.0199)	(1.0200)	(1.0217)	(1.0492)	(1.0314)
Smoker				-0.1793***	
				(0.0206)	
Worried about environment					0.1492***
					(0.0282)
State dependence					
$Employed_{t-1}$	0.0007	0.0090	0.0112	0.0172	0.0099
	(0.0182)	(0.0183)	(0.0198)	(0.0246)	(0.0173)
$Sport_{t-1}$	0.4982***	0.4996***	0.4984***	0.5548***	0.5242***
	(0.0406)	(0.0407)	(0.0408)	(0.0411)	(0.0416)
Initial Conditions					
$Sport_0$	1.1811***	1.1809***	1.1839***	1.1457***	1.1546***
	(0.0303)	(0.0303)	(0.0303)	(0.0307)	(0.0309)

(*) Significant at 10%. (**) Significant at 5%. (***) Significant at 1%.

Table 3: **Sports practice, Employment and Wage**

	Wooldridge's approach - Employment				
	(1)	(2)	(3) CRE	(4) CRE	(5) CRE
Constant	-0.4349*** (0.1107)	1.3792*** (0.1231)	1.3698*** (0.1231)	1.3883*** (0.1286)	1.2768*** (0.1238)
Woman	-0.1248*** (0.0174)	-0.1259*** (0.0173)	-0.1245*** (0.0173)	-0.1281*** (0.0181)	-0.1277*** (0.0177)
Married	0.1022*** (0.0182)	0.1028*** (0.0182)	0.1046*** (0.0182)	0.1018*** (0.0190)	0.1000*** (0.0186)
Abitur	0.2627*** (0.0392)	0.2608*** (0.0393)	0.2604*** (0.0393)	0.2582*** (0.0411)	0.2618*** (0.0400)
High	0.3434*** (0.0202)	0.3439*** (0.0202)	0.3411*** (0.0202)	0.3526*** (0.0211)	0.3523*** (0.0206)
Foreign	-0.1574*** (0.0456)	-0.1590*** (0.0456)	-0.1587*** (0.0456)	-0.1826*** (0.0492)	-0.1818*** (0.0464)
From 20 to 29	0.0759*** (0.0225)	0.0744*** (0.0226)	0.0740*** (0.0226)	0.0713*** (0.0236)	0.0717*** (0.0229)
From 40 to 49	0.1300*** (0.0236)	0.1299*** (0.0237)	0.1315*** (0.0237)	0.1347*** (0.0246)	0.1307*** (0.0241)
From 50 and more	-0.6456*** (0.0212)	-0.6512*** (0.0212)	-0.6470*** (0.0213)	-0.6457*** (0.0221)	-0.6532*** (0.0217)
West	0.0633*** (0.0171)	0.0646*** (0.0171)	0.0623*** (0.0172)	0.0919*** (0.0191)	0.0833*** (0.0188)
Berlin	-0.0330 (0.0403)	-0.0332 (0.0403)	-0.0342 (0.0402)	-0.0355 (0.0430)	-0.0147 (0.0414)
Unemploy. rate	-6.9940*** (1.2272)	-7.2581*** (1.2356)	-7.2808*** (1.2361)	-7.6192*** (1.2770)	-6.8876*** (1.2386)
$Sport_t$	0.0273 (0.0297)	-0.0543 (0.0462)	-0.0377 (0.0640)	-0.0965 (0.1059)	-0.0340 (0.0683)
State dependence					
$Employed_{t-1}$	1.7630*** (0.0367)				
$(Unemployed \times Athletic)_{t-1}$		-1.7285*** (0.0497)	-1.7244*** (0.0500)	-1.6955*** (0.0546)	-1.6704*** (0.0503)
$(Unemployed \times Not\ Sporty)_{t-1}$		-1.7859*** (0.0387)	-1.7843*** (0.0386)	-1.7839*** (0.0395)	-1.7292*** (0.0388)
$(Employed \times Athletic)_{t-1}$		0.0962** (0.0377)	0.1030*** (0.0383)	0.1227*** (0.0462)	0.1000** (0.0395)
Random Effect					
$Employed_0$	0.6547*** (0.0304)	0.6297*** (0.0309)	0.6310*** (0.0309)	0.6351*** (0.0316)	0.6706*** (0.0312)
$\bar{y}_{si}$			0.0105 (0.0636)	0.0252 (0.0844)	-0.0077 (0.0704)

(*) Significant at 10%. (**) Significant at 5%. (***) Significant at 1%.

Table 3: **Sports practice, Employment and Wage (continued)**

	Wooldridge's approach - Wage				
	(1)	(2)	(3) CRE	(4) CRE	(5) CRE
Constant	7.2559***	7.3231***	7.3111***	7.2871***	7.2784***
	(0.0150)	(0.0157)	(0.0177)	(0.0192)	(0.0178)
Woman	-0.3177***	-0.3172***	-0.3145***	-0.3152***	-0.3093***
	(0.0116)	(0.0115)	(0.0115)	(0.0118)	(0.0115)
Married	0.0230***	0.0207**	0.0227**	0.0275***	0.0229**
	(0.0089)	(0.0088)	(0.0089)	(0.0092)	(0.0089)
Abitur	0.2858***	0.2827***	0.2823***	0.2723***	0.2801***
	(0.0209)	(0.0207)	(0.0207)	(0.0212)	(0.0207)
High	0.3993***	0.3950***	0.3930***	0.3975***	0.3915***
	(0.0115)	(0.0114)	(0.0117)	(0.0121)	(0.0117)
Foreign	-0.0892***	-0.0897***	-0.0883***	-0.1000***	-0.0865***
	(0.0264)	(0.0262)	(0.0262)	(0.0270)	(0.0261)
From 20 to 29	0.0769***	0.0789****	0.0782****	0.0796****	0.0790****
	(0.0085)	(0.0084)	(0.0084)	(0.0086)	(0.0084)
From 40 to 49	0.0387***	0.0366***	0.0382***	0.0410***	0.0370***
	(0.0091)	(0.0090)	(0.0091)	(0.0093)	(0.0091)
From 50 and more	-0.0533***	-0.0547***	-0.0524***	-0.0499***	-0.0554***
	(0.0120)	(0.0119)	(0.0122)	(0.0124)	(0.0121)
West	0.1474***	0.1473***	0.1433***	0.1682***	0.1648***
	(0.0106)	(0.0105)	(0.0107)	(0.0117)	(0.0114)
Berlin	0.1414***	0.1418***	0.1400***	0.1644***	0.1700***
	(0.0244)	(0.0242)	(0.0241)	(0.0250)	(0.0246)
State dependence					
$(Unemployed{\times}Athletic)_{t-1}$		-0.1031***	-0.1102***	-0.1146***	-0.1158***
		(0.0090)	(0.0092)	(0.0093)	(0.0093)
$(Unemployed{\times}Not\ Sporty)_{t-1}$		-0.0734***	-0.0726***	-0.0745***	-0.0757***
		(0.0068)	(0.0068)	(0.0069)	(0.0068)
$(Employed{\times}Athletic)_{t-1}$		0.0176***	0.0120**	0.0129**	0.0132**
		(0.0057)	(0.0059)	(0.0059)	(0.0059)
Random Effect					
$Employed_0$	0.8550***	0.7923***	0.7952***	0.7865***	0.7961***
	(0.0125)	(0.0131)	(0.0131)	(0.0135)	(0.0132)
$Sport_0$			0.0699	0.0446	0.0304
			(0.0405)	(0.0485)	(0.0366)
$\bar{y}_{si}$			-0.0293	0.0068	0.0231
			(0.0648)	(0.0790)	(0.0576)

(*) Significant at 10%. (**) Significant at 5%. (***) Significant at 1%.

Table 3: **Sports practice, Employment and Wage (continued)**

Wooldridge's approach - Residuals $r_{jit} = \alpha_{ij} + u_{jit}$ $u_{jit} = \rho_j\, u_{jit-1} + \epsilon_{jit}$					
	(1)	(2)	(3) CRE	(4) CRE	(5) CRE
			Standard errors of individual effects (α_{ij})		
$\sigma_{\alpha_s} = \frac{\exp(v_s)}{1+\exp(v_s)}$	-0.7966***	-0.7981***	-0.7992***	-0.7471***	-0.7600***
	(0.0445)	(0.0446)	(0.0446)	(0.0470)	(0.0459)
$\sigma_{\alpha_e} = \frac{\exp(v_e)}{1+\exp(v_e)}$	0.2232***	0.2673***	0.2680***	0.2486***	0.1673***
	(0.0647)	(0.0680)	(0.0679)	(0.0685)	(0.0641)
$\sigma_{\alpha_w} = \exp(v_w)$	0.5001***	0.3693***	0.3668***	0.3660***	0.3453***
	(0.0619)	(0.0395)	(0.0395)	(0.0380)	(0.0357)
			Correlations between individual effects (α_{ij})		
$\rho_{\alpha_s\alpha_e} = \tanh(c_{se})$	0.1064***	0.1054***	0.0829**	0.0998**	0.0880**
	(0.0362)	(0.0370)	(0.0392)	(0.0412)	(0.0372)
$\rho_{\alpha_s\alpha_w} = \tanh(c_{sw})$	0.0780***	0.0628***	0.0877*	0.0434	0.0423
	(0.0223)	(0.0204)	(0.0512)	(0.0610)	(0.0452)
$\rho_{\alpha_e\alpha_w} = \tanh(c_{ew})$	2.9861	1.4035***	1.4105***	1.2390***	1.1703***
	(4.2433)	(0.1412)	(0.1427)	(0.1044)	(0.0845)
			Auto-Correlation of error terms (u_{jit})		
$\rho_s = \tanh(d_s)$	-0.1105***	-0.1139***	-0.1130***	-0.1539***	-0.1229***
	(0.0307)	(0.0308)	(0.0309)	(0.0290)	(0.0303)
$\rho_e = \tanh(d_e)$	-0.2352***	-0.2239***	-0.2235***	-0.2188***	-0.2144***
	(0.0167)	(0.0169)	(0.0169)	(0.0173)	(0.0174)
$\rho_w = \tanh(d_w)$	0.9928***	0.8601***	0.8588***	0.8364***	0.8199***
	(0.0342)	(0.0300)	(0.0299)	(0.0299)	(0.0288)
			Correlations between error terms (ϵ_{jit})		
$\rho_{se} = \tanh(f_{se})$	-0.0719***	-0.0205	-0.0266	0.0110	-0.0371
	(0.0202)	(0.0298)	(0.0358)	(0.0646)	(0.0391)
$\rho_{sw} = \tanh(f_{sw})$	-0.0180	-0.0123	-0.0126	0.0042	-0.0153
	(0.0122)	(0.0120)	(0.0122)	(0.0122)	(0.0122)
$\rho_{ew} = \tanh(f_{ew})$	0.1348***	0.1469***	0.1476***	0.1461***	0.1547***
	(0.0101)	(0.0103)	(0.0103)	(0.0106)	(0.0105)
			Standard error of log of wage (u_{wit})		
$\sigma_{u_w} = \exp(f)$	-0.8647***	-0.9762***	-0.9774***	-0.9929***	-1.0073***
	(0.0294)	(0.0245)	(0.0244)	(0.0241)	(0.0231)
Number of obs.	14014	14014	14014	12376	13933

(*) Significant at 10%. (**) Significant at 5%. (***) Significant at 1%.

Let us consider a two period utility function $U(C_0, C_1, L_0, L_1, H(x \mid h, A), e)$, where C denotes consumption, L denotes leisure. H denotes health measured as income equivalent. Let us suppose that the amount of labor the individual offers to a firm is an increasing function of a number of hours $L - L_t$ and of the employee's effort e. The effort is not verifiable and is assumed to be constant. An employer has an imperfect observation of e and decides to dismiss an employee with the probability $1-\tau(e)$ (τ is increasing with e). Indeed, it is the beliefs of the employee about the function $\tau(e)$. Let us assume that z is the income of a dismissed agent and can be considered as the unemployment benefit. An employed individual can be promoted with probability $p(e)$ at the beginning of period 1 ($p(e)$ is an increasing with e). In period 0, wage rate is denoted w_0. $w_1(1)$ is the wage rate in period 1 in absence of promotion and $w_1(2)$ is the wage rate in period 1 if the individual is promoted ($w_1(2) > w_1(1)$). Here the practice of sport is used by the individual as a form of insurance mechanism. The choice to practice sport is a binary decision. We assume that the direct advantages associated to the practice of sport (through training) compensate the direct inconveniences (pain). Let z denote the income in the non employment state.

The utility U and the health production function H are increasing in their arguments. h refers to the health state, s refers to the practice of sport ($s = 0, 1$), the subscripts 0 refers to the current period and the subscripts 1 refers to the future period. The individual has a prior conditional probability measure $\pi(h \mid A, s)$ of the health states. The amount of saving at the end of period 1 $a(s)$ depends of the decision to practice sports ($s = 1$ versus $s = 0$).

The utility function given a given level of health h is

$$U(C_0, C_1, L_0, L_1, H, e) =$$

$$\begin{aligned}
&= \tau(e)\,(1-p(e))\,C_0\,C_1(1)^{a+1}\,L_0^{b+1}\,L_1(1)^{c+1}\,H_1^{\sigma+1}(1+e)^{-(d+1)} \\
&\quad +\tau(e)\,p(e)\,C_0\,C_1(2)^{a+1}\,L_0^{b+1}\,L_1(2)^{c+1}\,H_2^{\sigma+1}(1+e)^{-(d+1)} \\
&\quad +(1-\tau(e))\,C_0\,C_1(3)^{a+1}\,L_0^{b+1}\,L_1(3)^{c+1}\,H_3^{\sigma+1}(1+e)^{-(d+1)}
\end{aligned}$$

where the health production function is

$$H_j = H(x_j \mid h, A) = \prod_k x(j)_k^{\alpha_k(h,A)}$$

and is such that $\alpha_k(h, A) \geq 0$. A are individual characteristics.

The conditional utility consists in three additive terms. The first one corresponds to an individual employed in period 1 who was not promoted. The second one corresponds to an individual employed in period 1 but who was promoted. The last one corresponds to a worker who is unemployed in period 1 (dismissed at the beginning of period 1).

Let $E[U_s]$ denote the expectation of the conditional utility given s ($s = 0, 1$).

The individual chooses the allocation such that

$$\max_{s,C_0,C_1,L_0,L_1,x,e} E[U_s] = \int U(C_0, C_1, L_0, L_1, H(x,h), e)\, d\pi(h \mid A, s) + v_s$$

subject to $Y_{0s} + P + s\, P_s = Y_0 + \omega_0(L - L_0(s))$, $C_0(s) + a(s) = Y_{0s}$, $C_1(s,j) + \tilde{p}' x(s,j) = Y_1 + (1+r)a(s) + \omega_1(j)(L - L_1(s,j))$, for $j = 1, 2$ and $C_1(s,3) + \tilde{p}' x(s,3) = Y_1 + (1+r)a(s) + z$. $a(s)$ represents individual savings, L is the maximum time he can devote to work, P is the insurance premium, P_s is the cost of playing sport and the interest rate is r. Y_0 denotes an exogenous income. x is a vector of health services goods. $\tilde{p}$ is an exogenous vector of price per unit of health care services net of reimbursement.

$a(s)$ is a function of s due to the presence of the cost of practicing sports and because the demand of health care services x depends on s. Moreover, as the utility varies with the health state, the demand of care services depends on h. The practice of sports can have an impact of the distribution of health state and, finally, on the expected utility since the conditional utility depends on h.

Proposition 5.1 *Conditional on the health status h, the state of the market j and sport practice s, the demand equations verify*

$$C_1^*(s,j) = \frac{Y_1 + (1+r)\,a(s) + \omega_1(j)L}{\psi + \frac{1+c}{1+a}}, \text{ where } j = 1, 2,$$

$$C_1^*(s,3) = \frac{Y_1 + (1+r)\,a(s) + z}{\psi}, \quad C_0^*(s) = \frac{\xi_1\, C_1(s,1) + \xi_2\, C_1(s,2) + \xi_3\, C_1(s,3)}{(1+r)(1+a)},$$

$$x_k^*(s,j) = \frac{C_1^*(s,j)\, \alpha_k(h,A)\,(\sigma+1)}{\tilde{p}_k\,(1+a)}, \text{ where } j = 1, 2, 3,$$

where $\psi = 1 + \sum_k \alpha_k(h,A) \frac{1+\sigma}{1+a}$, $\xi_j = \frac{\frac{\kappa_j}{C_1(s,j)}}{\frac{\kappa_1}{C_1(s,1)} + \frac{\kappa_2}{C_1(s,2)} + \frac{\kappa_3}{C_1(s,3)}}$ *(j=1,2,3), and*

$$\kappa_1 = \tau(e)\,(1 - p(e))\, C_1(1)^{a+1}\, L_0^{b+1}\, L_1(1)^{c+1}\, H_1^{\sigma+1} (1{+}e)^{-(d+1)},$$

$$\kappa_2 = \tau(e)\; p(e)\, C_1(2)^{a+1}\, L_0^{b+1}\, L_1(2)^{c+1}\, H_2^{\sigma+1} (1{+}e)^{-(d+1)},$$

$$\kappa_3 = (1 - \tau(e))\, C_1(3)^{a+1}\, L_0^{b+1}\, L_1(3)^{c+1}\, H_3^{\sigma+1} (1{+}e)^{-(d+1)}.$$

Moreover, the demands of leisure are

$$L_0^*(s) = \frac{(1+b)\,C_0^*(s)}{\omega_0},\; L_1^*(s,1) = \frac{(1+c)\,C_1^*(s,1)}{(1+a)\,\omega_1(1)},$$

$$L_1^*(s,2) = \frac{(1+c)\,C_1^*(s,2)}{(1+a)\,\omega_1(2)}.$$

Proof : see appendix 1.

The level of effort has an impact on the frequency of realization of the states of the labor market and not, directly, on the structure of the consumption given the realized states of the labor market. This structure of consumption is indirectly determined by the level of effort via the level of savings. However, the demand corresponding to the first period of time incorporates the information relative to the probability of realization of states of the labor market. We are know going to clarify how the distribution of the states of the labor market is linked to the decisions taken in the first period of time.

Let $U(e,z) = U(C_0^*, C_1^*, L_0^*, L_1^*, H^*, e)$ denote the conditional utility given, z, h and s.

Proposition 5.2 *Let us assume the probability to remain employed the next period of time is such that* $\tau(e) = \tau^+$ *for* $e \geq e_0 > 0$ *and* $\tau(e) = \tau^-$ *otherwise (*$\tau^+ > \tau^-$*). The probability to be promoted is such that* $p(e) = p^+$ *if* $e \geq e_0$ *and* $p(e) = p^-$ *otherwise (*$p^+ > p^-$*) .*

Let us consider a hypothetical state of the nature h *and for a given value of* s*. We can have two exclusive situations according to non labor income:*

If $U(e = e_0, z = 0) \leq U(e = 0, z = 0)$*, the individual will not make any effort whatever the value of the unemployment benefit.*

If $U(e = e_0, z = 0) > U(e = 0, z = 0)$*, there exits a threshold* $z_0(h) > 0$ *such that if* $z < z_0(h)$ *then* $e^* = e_0 > 0$ *and if* $z \geq z_0(h)$ *then* $e^* = 0$[6]*.*

Proposition 5.2 states that the level of unemployment benefit has an influence on the level of effort given the probability to be dismissed is sufficiently high when effort is low, non labor income is relatively low and wage is sufficiently large.

Proposition 5.3 *Let us assume that the density of health states is such that* $\pi(h \mid A, s) = \pi(h \mid A)$*.*

[6]If we assume moreover that $\tau^+(1-p^+) > \tau^-(1-p^-)$, then we have $U(e = e_0, z = 0) > U(e = 0, z = 0)$.

Let us consider reference situation such that $p(e) = p$ ($0 < p < 1$), $\tau(e) = \tau$ ($0 < \tau < 1$) and $0 \leq z < \omega_1(1)$. There exists $\bar{\omega}_1(2) > \omega_1(1)$ such that for all $\omega_1(2) \geq \bar{\omega}_1(2) \Longleftrightarrow s^* = 1$.

When they are promoted, sporty individuals can obtain a wage bonus. Proposition 5.3 states that, if the bonus is sufficiently high, this will compensate at least the cost associated to the practice of sport (namely P_s).

The expected utility associated to the choice s of sporting activity is

$$E[U_s] = \int_{\mathcal{H}} U(C_0^*, C_1^*, L_0^*, L_1^*, H^*(h, A), e^*)\, d\pi(h \mid A, s) + v_s$$

$$= EV_s(A, Y_0 + \omega_0 L_0^*(s), P_s) + v_s$$

where EV_s is the indirect conditional expected utility associated to the choice s. v_s is an error term corresponding to unobserved heterogeneity of the individual.

The individual choose to practice sport if $E[U_1]$ is larger than $E[U_0]$.

Additional assumptions : Let us assume the probability to remain employed the next period of time is such that $\tau(e) = \eta_1$ if $e \geq e_0 > 0$ and $\tau(e) = \eta_2$ otherwise ($0 \leq \eta_2 < \eta_1 \leq 1$). The probability to be promoted is such that $p(e) = \eta_3$ if $e < e_0$ and $p(e) = \eta_4$ otherwise ($0 \leq \eta_3 < \eta_4 \leq 1$). In period 1, the wage is $\omega_1(2) = \omega_0 + c_1 + c_2 \mathbb{1}[s = 1]$ if the individual is promoted and $\omega_1(1) = \omega_0 > 0$, where $c_1 > 0$ and $c_2 > 0$.

Consequently, we assume that, unlike effort, the practice of sport do not increase the probability to be promoted. Generally, it is very difficult to assess, a priori, the real content of a sport practice. However, the practice of sport provides an additional wage bonus when the individual is promoted due to the existence of non cognitive skills linked to the practice of sport.

Let $U_s(h, e) = U(e, z) = U(C_0^*, C_1^*, L_0^*, L_1^*, H^*, e)$ denote the conditional utility given, z, h and for s fixed ($s = 0, 1$).

Proposition 5.4 *Let us assume that the density of health states is such that $\pi(h \mid A, s = 0) = 1$ for all $1 \geq h \geq 0$. $\pi(h \mid A, s = 1) = 0$ for all $g \geq h \geq 0$ and $\pi(h \mid A, s = 1) = 1/(1 - g)$ for all $1 \geq h \geq g$. Let us assume that $\mathcal{H} = [0, 1]$. Let us assume that $\alpha_k(h, A)$ is monotone decreasing with h for k and $\alpha_k(1, A) = 0$. Let us assume that $\tau(e) < 1$.*

There exits a threshold $1 > g_0(e) > 0$ such that for all $g > g_0(e)$ then $E_{\pi(h|A,s=1)}(U_1(e, h)) > E_{\pi(h|A,s=0)}(U_0(e, h))$.

Proposition 5.4 states that, for this economy, the individuals practicing sport can reduce the frequency of illnesses in period $t = 1$ (illness context is associated to h low). The model can be generalized to a situation such that s is the intensity of the sport practice that can vary from 0 (no practice) to 1, if we assume that the cost of the practice is increasing with intensity ($F(s, A)P_s$, where P_s is fixed and $F'_s(s, A) > 0$), and g is increasing with s ($g = s$ for instance), where g is the lower bound of the health distribution. $F(s, A)$ is the quantity of training or equipment necessary to achieve the intensity s of the sport practice. Such a model can be consistent with the observation that a significant proportion of the population do not practice sport at all (A is such that for all $s \in \mathcal{H}$, $E_{\pi(h|A,s=1)}(U_s(h, e^*)) < E_{\pi(h|A,s=0)}(U_0(h, e^*))$ and that sporty individuals are heterogeneous in the practice of sport $s^* = s^*(A)$.

Those who are doing sport are more healthy. Being more healthy lower your expenses related to illness ($\alpha_k(h, A)$ decreases with h) and increases your productivity ($c_2 > 0$).

Proposition 5.5 *Let us assume that the density of health states is such that* $\pi(h \mid A, s = 0) = 1$ *for all* $1 \geq h \geq 0$. $\pi(h \mid A, s = 1) = 0$ *for all* $g \geq h \geq 0$ *and* $\pi(h \mid A, s = 1) = 1/(1 - g)$ *for all* $1 \geq h \geq g$. *Let us assume that* $\mathcal{H} = [0, 1]$. *Let us assume that* $\sum_k \alpha_k(h, A)$ *is monotone decreasing with h. Moreover, let us assume that* c_2 *is large enough to achieve* $E_{\pi(h|A,s=1)}[U_1(h, e{=}e_0)] > E_{\pi(h|A,s=1)}[U_1(h, e{=}0)]$[7].

There exits a threshold g_0, $1 > g_0 > 0$, *such that for all* $g > g_0$ *we have*

$$E_{\pi(h|A,s=1)}[U_1(h, e{=}e_0)] > \max\{E_{\pi(h|A,s=0)}[U_0(h, e{=}e_0)]; E_{\pi(h|A,s=0)}[U_0(h, e{=}0)]\} = m.$$

Has physical activity has a positive effect on health, sport practice can modify the distribution of health state. Here it is assumed that sport reduces the probability of occurrence of ill health. Under the assumption that individuals behave maximizing their expected utility, the proposition 5.5 states that, if this reduction is large enough, it is better to practice sport. Conversely, results indicate that health is an important determinant of effort (see appendix 2).

The proposition 5.3 states that, when the individual is promoted, if the part of the bonus specific to athletes is large enough, there is an incentive - ex ante -

[7] For the limit situation such that $p(0) = \tau(0) = 0$ and $p(e_0) = \tau(e_0) = 1$, if c_2 tends to infinity then, for all h, $U_1(h, e_0)$ tends to infinity and $U_1(h, 0)$ is constant.

to practice a sport. This bonus is offered to sporty individuals by firms and can be explained by a higher productivity. Indeed, sporty individuals are more likely to be healthy (Reiner et al., 2013). This result explains that the conditional wage expectation is higher for athletes. Rooth (2011), using Swedish data, finds evidence of a sports premium based on the physical fitness when enlisting at age 18.

In the model, the promotion is driven mainly by the effort of the worker. The promotion is conditional on remaining in the firm at the end of the first period. The firm cannot observe perfectly the effort of the worker. The practice of sport improves or maintains the state of health. It is this state of health which gives entitlement to an additional premium in the event that the individual is promoted. In order to benefit of a larger bonus, sporty individuals have to be promoted and this promotion depend on the level of effort. That this why sporty individuals are more likely to have a larger effort. The proposition 5.5 states that, given the characteristics of the individuals, the expected utility is larger for sporty individuals who make sufficient effort provide that the additional premium is large enough and that the health enhancement is reasonably likely. In such a context, ex post, sporty individuals are more likely to make a relatively large effort on the labour market. That is why athletes are more likely to remain employed.

The effect of the sport practice on health depends on the characteristic of the sport and the intensity of the sport practice. The relation may be nonlinear: too much intensity may have some detrimental effects on the short or the long run. This intensity cannot be observed perfectly. This explain that sporting activity cannot be used by firm easily during the hiring process. In order to do so, characteristics of the sport should be available (individual or team sport, recreational or competitive sport, level of the sport practice). This information is not available in the data set. On the contrary, ex post - at the end of the first period - firm can have some information on the health of the individual and on the productivity, even if this information is not perfect.

The structural model illustrates why sporty individuals are in average better paid - they get a wage premium when they are promoted related to a larger productivity - and are more likely to remain employed as they are more likely to exert a larger effort. These findings and mechanisms are consistent with empirical results obtained on the GSOEP.

6 Conclusion

In this paper, we jointly model employment, wage and sporting activity using a dynamic model. The model we use allows to take into account both true and spurious state dependence. Moreover, the specification we use allows to treat selection mechanisms separately by indexing participation to sports by the previous position on the labour market.

The model we use is dynamic and allows to make sport practice and situation on the labour market interact. It is the first model that implement such a modeling strategy in such a context. We show that, ceteris paribus, the current practice of sport has no significant impact on the conditional probability to be employed. This probability is conditioned by past practice of sport. This means that the effects of sports activity take at least a few months. This delay can be explained by the physical mechanisms linking the practice of sport to productivity.

Among unemployed individuals, athletic people are not more likely to be reemployed the next year. Declaring to practice a sport can correspond to very different situations and effective sport involvement is in general not observable. People who were already employed the previous year and practicing sport are more likely to be currently employed compared to individuals who were already employed last year but not practicing a sport. Sporting Activity may be a source of non cognitive capabilities useful in a job. People practicing sports have a larger conditional expectation of the wage.

The economic model we use allows to shed light on the empirical results. The employed worker who practice a sport are more likely to have a large effort (proposition 5.5) and to have a large productivity (see proposition 6.1). As sporty individuals are more likely to have a larger productivity, they get an additional wage premium when they get promoted. This bonus is significantly different form 0. As sporty individuals are more likely to have a larger effort, they are more likely to remain employed. As effort is not observable, only the structural model allows to understand economic mechanisms at play. The structural model is by nature restrictive and only an econometric model allows to be sure of existence and importance of such relations. Finally, both modeling - structural and econometrics - complement each other. The article contains an additional original result : the better the health and the greater the effort (see proposition 6.1). One may interpret this results concluding that workers are not equal on the labour market since they differ by health conditions. However, both empirical and theoretical results indicate that the individual can act to reduce these inequalities through sports practice (see propositions 5.4 and 5.5).

Further research may consists in studying the heterogeneity of the impact of sporting activity according to the category of the sport (alone or within a team, Gymnastics, Athletics). Another research could be to study the relationship between sport participation, social relations and individual performance on the labour market.

References

BONILLA R. and F. KIRALY (2013). – "Marriage wage premium in a search equilibrium", Labour Economics 24 (2013) 107-115.

CABANE C. (2014). – "Unemployment Duration and Sport Participation", *International Journal of Sport Finance*, 9(3), 261–280.

CABANE C. and A. CLARK (2015). – "Chilhood Sporting Activities and Adult Labour-Market Outcomes", *Annals of Economics and Statistics*, 119-120, 123-148.

CAMERON A.C., TRIVEDI P.K., MILNE F. and J. PIGGOTT (1988). – "A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia", *The Review of Economic Studies*, Volume 55, 1, 85-106.

CHANG S.-K. (2009). – "Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women", *Journal of Applied Econometrics*, n/a. doi: 10.1002/jae.1141.

CUNHA F. and J.J. HECKMAN (2007). – "The Technology of Skill Formation", *American Economic Review*, 97(2), 31-47.

CUNHA F. and J.J. HECKMAN (2009). – "The Economics and Psychology of Inequality and Human Development", *Journal of the European Economic Association*, 7(2), 320–364.

CUNHA F., HECKMAN J.J. and S. SCHENNACH (2010). – "Estimating the Technology of Cognitive and Noncognitive Skill Formation" *Econometrica*, vol. 78, no. 3, 883-931.

DOWNWARD P., DAWSON A. and T. DEJONGHE (2009). – " Sport Economics. Theory, Evidence and Policy", *Butterworth-Heinemann eds.*.

DUSTMANN C. and C. MEGHIR (2005). – " Wages, Experience and Seniority", *Review of Economic Studies*, 72, 77-108

EDON C. and T. KAMIONKA (2014). – "Dynamic labor market behavior of married women with endogenous participation, unemployment, working time and wage ", *mimeo Crest*, Paris.

EDON C. and T. KAMIONKA (2008). – "Modélisation dynamique de la participation au marché du travail des femmes en couple", *Annales d'Économie et de Statistique*, 86-2007, 77–108.

GEWEKE J. (1991). – "Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints", *Computing Science and*

Statistics: Proceedings of the twenty-third symposium on the interface. Computing science and statistics, American statistical association, Alexandria, 571–578.

GILBERT L., KAMIONKA T. and G. LACROIX (2011). – "The Impact of Government- Sponsored Training Programs on the Labor Market Transitions of Disadvantaged Men", in *Advances in Econometrics, Theory and Applications*, Miroslav Verbic ed., Intech, ISBN 978-953-307-328-6.

GOURIÉROUX Ch. and MONFORT A. (1997). – "Simulation-Based Econometric Methods", *Core Lecture Series*, Oxford: Oxford University Press.

HAJIVASSILIOU V., MCFADDEN D. and RUUD P. (1992). – "Simulation of the Multivariate Normal Orthant Probabilites : Methods and Programs", *Cowles Foundation Discussion Paper 1021*, Yale University.

HEILIG G., HEILIG G., Buttner T. and W. Lutz (1990). – "Germany's population: Turbulent Past, Uncertain Future", in *Population Bulletin*, Vol. 45, No. 4.

HYSLOP D.R. (1999). – "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women", *Econometrica*, 67, 6, 1255–1294.

KAMIONKA T. (1998). – "SML Estimation in Transition Models", *The Econometrics Journal*, 1, C129–C153.

KAMIONKA T. and G. LACROIX G. (2008). – "Assessing the External Validity of an Experimental Wage Subsidy", *Annales d'Économie et de Statistique*, 91-92, 357–384.

KAMIONKA T. and G. LACROIX (2018). – "Homeownership, Labour Market Transitions and Earnings", *IZA Discussion Papers* N° 11952.

KAMIONKA T. and P. LEVENEUR (2020). – "The Dynamics of Health, Employment and Working Hours", mimeo, *Crest Paris*.

KAMIONKA T. and X. VU NGOC (2016). – "Insertion des jeunes sur le marché du travail, diplôme et quartier d'origine: Une modélisation dynamique", *Revue Économique*, 67.3, 463–494.

KEANE M. (1994). – "A Computationally Practical Simulation Estimator for Panel Data", *Econometrica*, 62, 95–116.

LECHNER M. (2009). – "Long-run Labour Market and Health Effects of Individual Sports Activity", *Journal of Health Economics*, 28, 839–854.

LECHNER M. and P. DOWNWARD (2013). – "Heterogeneous Sports Participation and Labour Market Outcomes in England", *IZA Discussion Papers*, n° 7690.

LECHNER M. and N. SARI (2014). – “Labor Market Effects of Sports and Exercice: Evidence from Canadian Panel Data”, *Labour Economics*, 28, 35(C), 1–15.

ROOTH D.-O. (2011). – “Work Out or Out of Work - The Labor Market Return to Pysical Fitness and Leisure Sports Activities”, *Labour Economics*, 18, 399–409.

WAGNER G.G., FRICK J.R. and J. SCHUPP (2007). – “The German Socio-Economic Panel Study (SOEP) - Scope, Evolution, Enhancements”, 1, SOEP papers on Multidisciplinary Panel Data Reasearch, *DIW Berlin*.

WOOLDRIDGE J.M. (2005). – “Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity”, *Journal of Applied Econometrics*, 20-1, 39–54.

Appendix 1 : Demand equations in the economic model

The utility function is

$$U(C_0, C_1, L_0, L_1, H, e) =$$

$$\begin{aligned} &= \tau(e)\,(1{-}p(e))\,C_0\,C_1(1)^{a+1}\,L_0^{b+1}\,L_1(1)^{c+1}\,H_1^{\sigma+1}(1{+}e)^{-(d+1)} \\ &\quad +\tau(e)\,p(e)\,C_0\,C_1(2)^{a+1}\,L_0^{b+1}\,L_1(2)^{c+1}\,H_2^{\sigma+1}(1{+}e)^{-(d+1)} \\ &\quad +(1{-}\tau(e))\,C_0\,C_1(3)^{a+1}\,L_0^{b+1}\,L_1(3)^{c+1}\,H_3^{\sigma+1}(1{+}e)^{-(d+1)} \end{aligned}$$

where the health production function is

$$H_j = H(x_j \mid h, A) = \prod_k x(j)_k^{\alpha_k(h,A)}$$

where $\alpha_k(h, A) \geq 0$. A are individual characteristics.

Conditional on sport practice s and health state h, the individual chooses the allocation such that

$$\max_{C_0, C_1, L_0, L_1, x, e} U(C_0, C_1, L_0, L_1, H, e)$$

subject to

$$Y_{0s} + P + s\,P_s = Y_0 + \omega_0(L - L_0),$$

$$C_0(s) + a(s) = Y_{0s},$$

$$c_1(s,1) + \tilde{p}' x(s,1) = Y_1 + (1+r)a(s) + \omega_1(1)(L - L_1(s,1))$$

$$c_1(s,2) + \tilde{p}' x(s,2) = Y_1 + (1+r)a(s) + \omega_1(2)(L - L_1(s,2))$$

$$c_1(s,3) + \tilde{p}' x(s,3) = Y_1 + (1+r)a(s) + z$$

In order to obtain the demand equations, we maximize the Lagrangian with respect to the variables. The demand equations, conditional on sport practice s and health state h are

$$C_1^*(s,j) = \frac{Y_1 + (1+r)\,a(s) + \omega_1(j)L}{\psi + \frac{1+c}{1+a}}, \text{ where } j = 1, 2,$$

$$C_1^*(s,3) = \frac{Y_1 + (1+r)\,a(s) + z}{\psi},$$

$$C_0^*(s) = \frac{\xi_1\,C_1(s,1) + \xi_2\,C_1(s,2) + \xi_3\,C_1(s,3)}{(1+r)(1+a)},$$

$$x_k^*(s,j) = \frac{C_1^*(s,j)\,\alpha_k(h,A)\,(\sigma+1)}{\tilde{p}_k\,(1+a)},\text{ where } j=1,2,3,$$

where $\psi = 1 + \sum_k \alpha_k(h,A)\frac{1+\sigma}{1+a}$ and j is the state of the labour market, $\xi_j = \frac{\frac{\kappa_j}{C_1(s,j)}}{\frac{\kappa_1}{C_1(s,1)}+\frac{\kappa_2}{C_1(s,2)}+\frac{\kappa_3}{C_1(s,3)}}$ (j=1,2,3), and

$$\kappa_1 = \tau(e)\,(1-p(e))\,C_1(1)^{a+1}\,L_0^{b+1}\,L_1(1)^{c+1}\,H_1^{\sigma+1}(1+e)^{-(d+1)},$$

$$\kappa_2 = \tau(e)\;p(e)\,C_1(2)^{a+1}\,L_0^{b+1}\,L_1(2)^{c+1}\,H_1^{\sigma+1}(1+e)^{-(d+1)},$$

$$\kappa_3 = (1-\tau(e))\,C_1(3)^{a+1}\,L_0^{b+1}\,L_1(3)^{c+1}\,H_1^{\sigma+1}(1+e)^{-(d+1)}.$$

Moreover, the demands of leisure are

$$L_0^*(s) = \frac{(1+b)\,C_0^*(s)}{\omega_0},\; L_1^*(s,1) = \frac{(1+c)\,C_1^*(s,1)}{(1+a)\,\omega_1(1)},$$

$$L_1^*(s,2) = \frac{(1+c)\,C_1^*(s,2)}{(1+a)\,\omega_1(2)}.$$

Appendix 2 : Firm adjusts the probability of keeping the employee according to his productivity

Proposition 6.1 *Let us assume that the probability to remain employed is $\tau(e,h) = \tau^+(h)$ if $e > e_0$ and $\tau(e,h) = \tau^-(h)$ if $e < e_0$ ($e_0 > 0$). Let us assume moreover $\tau^+(h') > \tau^+(h) > \tau^-(h') = \tau^-(h)$, for all $h' > h$. The probability to be promoted is such that $p(e) = p^+$ if $e \geq e_0$ and $p(e) = p^-$ otherwise ($p^+ > p^-$). Let us consider the case such that $\sigma = -1$ and $\tau^+(h)(1-p^+) > \tau^-(h)(1-p^-)$.*

Then the effort is maximum if and only if the nonemployment income is such that $z < z_0(h)$ where $z_0(h') < z_0(h)$.

Proof: For $e = 0$ the utility is increasing with z but, in such a context, do not depend on the amount of health h. Unlike the utility in the case where $e = e_0 > 0$ which moves in the North West quadrant when h increases. ⋄

Proposition 6.1 states that, if the probability to remain employed is a non decreasing function of health, then the better the health and the greater the effort. In other words, the higher the probability to be healthy, the greater the probability that a person will exert maximum effort. In such a context, sporty individuals are more likely to remain employed because they have a larger productivity and are likely to exert a large effort.

Appendix 3 : Description of the region of integration

The integral is calculated over the set

$$A_i = \{r \in \mathbb{R}^{3T} : r = (r_{s1}, \ldots, r_{sT}, \ldots, r_{w1}, \ldots, r_{wT}) \text{ and } a_{jit} \leq r_{jt} \leq b_{jit}\}$$

The expressions of the boundaries a_{jit} and b_{jit} are fixed, for t=1,. . . , T, as follows:

$$\begin{cases} a_{jit} = -\infty, \text{ if } y_{jit} = 0, \\ b_{jit} = +\infty, \text{ if } y_{jit} = 1, \\ a_{jit} = -x'_{jit}\,\beta_j - z_j(y_{it-1}, y_{it})'\delta_j - y_{si0}\,\lambda_{js} - y_{ei0}\,\lambda_{je}, \text{ if } y_{jit} = 1, \\ b_{jit} = -x'_{jit}\,\beta_j - z_j(y_{it-1}, y_{it})'\delta_j - y_{si0}\,\lambda_{js} - y_{ei0}\,\lambda_{je}, \text{ if } y_{jit} = 0, \end{cases}$$

for $j = s, e$, and $1 \leq t \leq T$.

For wage, as we have to consider a continuous variable and to take into account that wage cannot be observed when the individual is not employed ($y_{wit} = .$), the boundaries are the following ones

$$\begin{cases} a_{wit} = -\infty, \text{ if } y_{wit} = ., \\ b_{wit} = +\infty, \text{ if } y_{wit} = ., \\ a_{wit} = b_{wit} = y_{wit} - z_w(y_{it-1})'\delta_w - y_{si0}\,\lambda_{ws} - y_{ei0}\,\lambda_{we}, \text{ if } y_{wit} \neq ., \end{cases}$$

where $1 \leq t \leq T$.

An alternative method to take into account the initial conditions consists in modeling the first realization of the process (see Kamionka and Lacroix, 2018, Kamionka and Leveneur, 2020). The estimations obtained using both methods are generally similar (see Edon and Kamionka, 2014).

Appendix 4 : Proofs

Proof of proposition 5.2

$a(s) = f_1(C_0) = Y_0 - sP_s - P_j + \omega_0 L - C_0(s) - (b+1)C_0(s)$ is a linear and decreasing function of $C_0(s)$. $C_0 = f_2(a(s)) = \frac{\xi_1 C_1(s,1)+\xi_2 C_1(s,2)+\xi_3 C_1(s,3)}{(1+r)}$ is an increasing and continuous function of $a(s)$.

If the unemployment benefit increases, then the curve representing the function f_2 on a graph (C_0 on the x axis, $a(s)$ on the y axis) is translated in the direction of the East ($C_1(s,3)$ increases, $C_1(s,1)$ and $C_1(s,2)$ are fixed, ξ_3 increases, ξ_1 and ξ_2

decrease). The curve representing the function f_1 is not modified. Consequently, C_0^* increases as z increases.

U is a monotone, continuous and increasing function of C_0^*. And C_0^* is a monotone, continuous and increasing function of z. Indeed, $U = C_0^*(\kappa_1+\kappa_2+\kappa_3)$.

Let us consider the case such that $U(e = e_0, z = 0) > U(e = 0, z = 0)$. For z large $U(e = e_0, z) < U(e = 0, z)$. Consequently, there exists a value of z, namely $z_0 > 0$ such that $U(e = e_0, z_0) = U(e = 0, z_0)$. Let us remark that the individual only considers $e = 0$ and e_0 because others value of the effort are associated with lower utility levels. $\diamond$

Proof of proposition 5.3

If $\omega_1(2)$ increases, then the curve $f_2(a)$ is translated on the right and, consequently, C_0^* increases with $\omega_1(2)$. If the individual considers $s = 1$, the curve $f_1(C_0)$ is translated on the left (C_0 on the x axis, a on the Y axis). There exists $\bar{\omega}_1(2) > \omega_1(1)$ such that $C_0^*(s = 1, \bar{\omega}_1(2)) = C_0^*(s = 0, \omega_1(1))$. Finally, $U^*(s = 1, \omega_1') > U^*(s = 0, \omega_1(1))$ for all $\omega_1' > \bar{\omega}_1(2)$. $\diamond$

Proof of proposition 5.4

Let us remark that $U_0(h, e) > U_1(h, e)$ (for h fixed, there is an additional cost, namely P_s).

Then let $g_0(e)$ denote the value of g such that

$$0 < 1 - g = \frac{\int_g^1 U_1(h, e)\, d\, h}{\int_0^1 U_0(h, e)\, d\, h} < \frac{\int_0^1 U_1(h, e)\, d\, h}{\int_0^1 U_0(h, e)\, d\, h} < 1$$

For $g = g_0(e)$ we have then

$$E_{\pi(h|A,s=1)}(U_1(h, e)) = \frac{\int_g^1 U_1(h, e)\, d\, h}{1 - g_0(e)} = E_{\pi(h|A,s=0)}(U_0(h, e))$$

The derivative of $\frac{\int_g^1 U_1(h,e)\, d\, h}{1-g}$ with respect to g is strictly positive ($0 < g < 1$) if $U_1(h, e)$ is increasing for $g < h < 1$.

Consequently, for $1 > g > g_0(e)$ we have then

$$E_{\pi(h|A,s=1)}(U_1(h, e)) > E_{\pi(h|A,s=0)}(U_0(h, e))$$

For all $g \geq g_0(e)$ it is then optimal to choose $s^* = 1$.

If $U_1(h,e)$ is non monotone everywhere, there is exits $g_0(e)' > g_0(e)$ such that this function is monotone increasing for all $h > g_0(e)'$ (as the limit of the function with respect to h is the infinity). If for all $1 > g > g_0(e)'$, $\frac{\int_g^1 U_1(h,e)\,d\,h}{1-g} > \int_0^1 U_0(h,e)\,d\,h$ and, then, we obtain $s^* = 1$. Otherwise, there exists $g > g_0(e)'$ such that $\frac{\int_g^1 U_1(h,e)\,d\,h}{1-g} = \int_0^1 U_0(h,e)\,d\,h$ (namely $g_0(e)''$). The derivative of $\frac{\int_g^1 U_1(h,e)\,d\,h}{1-g}$ with respect to g is strictly positive for $g_0(e)' < g < 1$. For all $g \geq g_0(e)''$ it is then optimal to choose $s^* = 1$. $\diamond$

Proof of proposition 5.5

There exists g_0, $1 > g_0 > 0$, such that $U_1(h, e{=}e_0)$ is monotone increasing for $h > g_0$. If for all $1 > g > g_0$, $\frac{\int_g^1 U_1(h,e=e_0)\,d\,h}{1-g} > m$ and then $s^* = 1$. Otherwise, there exists $g > g_0$ such that $\frac{\int_g^1 U_1(h,e=1)\,d\,h}{1-g} = m$ (namely g_0'). If g_0' is large enough, the derivative of $\frac{\int_g^1 U_1(h,e)\,d\,h}{1-g}$ with respect to g is strictly positive for $g_0' < g < 1$. For all $g \geq g_0'$ it is then optimal to choose $s^* = 1$. $\diamond$