Identifying dynamic discrete choice models:
An application to school-leaving in France

Thierry Magnac, INRA-LEA and CREST
David Thesmar, INSEE and CREST

First version: March 1998
This version: March 1999

Abstract

In this paper, we analyze the non parametric identification of dynamic discrete choice models without consumption smoothing using individual data on discrete choices. We posit that preferences are not restricted and that agents have private information about their tastes at each period that no other agents, including econometricians, can observe. Agents however are supposed to coordinate their expectations of future random tastes on a (common knowledge) distribution function. We first prove that if agents expect that random tastes are uncorrelated over time, data that they observe are consistent with this assumption provided that the expected d.f. of future random shocks is the true marginal d.f. Second, we prove that, for econometricians, the (functional) degree of underidentification is large. The common knowledge d.f. of future random tastes and the discount rate cannot be identified. Fixing those to arbitrary constants, we show that the present value of choosing an alternative can be decomposed into its value in the current period and into its future value. The identification proof also suggests an estimation procedure in two stages which amounts to running two sets of (non parametric or parametric) regressions. We apply this procedure to data on school leaving in France for the 1963-1973 cohorts. In that particular application, we then tighten the structural restrictions by using an income maximizing model and we use additional data on future life-cycle incomes in order to estimate private schooling costs. We finally return to the general model and explore the assumption where agents have more information than econometricians about other agents. It generates unobserved correlated heterogeneity over time. We show that the (functional) degree of underidentification of structural objects is even larger .

Keywords: Dynamic discrete choice, non parametric identification, education.

JEL classification: D91, C25, C14, I21

1. Introduction[1][2]

Over the last decade, there has been a huge increase in applications using structural models of dynamic discrete choice (Eckstein and Wolpin, 1989) and estimating structural parameters driving investments in human capital (Keane and Wolpin, 1996, Belzil and Hansen, 1997), labor market histories (Keane and Wolpin, 1996), retirement decisions, fertility and so on. These methods enhance our capacity to simulate dynamic consequences of economic policies. It comes at the cost of having to use in combination ML estimation methods with methods solving Bellman equations by backwards induction. These procedures are computer intensive, to say the least, and their degree of complexity is large (Rust, 1995). In three seminal papers, Hotz and Miller, 1993, Hotz, Miller, Sanders and Smith, 1994, and Altug and Miller, 1998, showed that ML methods and backwards induction could be dispensed of, at least at the estimation stage, by noting that Bellman equations could be used as moment conditions. Hotz and Miller's framework however seems to be less general than applications using ML methods and backwards induction (Keane and Wolpin, 1996) or applications that use a static model (Cameron and Heckman, 1998) which can accomodate correlated unobserved heterogeneity over time.

Nevertheless, we do not know much about the identification of dynamic discrete choice models except that they are generically not identified (Rust, 1994). All methods could be misleading if some assumptions are incorrect. That is why, in this paper, we first analyse the non parametric identification of dynamic discrete choice models without consumption smoothing using individual data on discrete choices and we pay particular attention to the difference in identification between models without and with correlated unobserved heterogeneity. We posit that agents' preferences over alternatives are completely unrestricted. We follow up Manski (1993) to model agents' expectations. Agents are first supposed to have

[1] *Corresponding author:* Thierry Magnac, INRA-LEA, 48, Boulevard Jourdan, 75014 Paris, France. *Email:* Thierry.Magnac@ens.fr

[2] We would like to acknowledge helpful discussions with A.Kirman, F.Kramarz, V.Patilea and participants in seminars at CREST, UCLondon, LSE-Sticerd, ESEM98, UCLouvain, *Investigaciones Economicas* meetings, Toulouse, Cergy-Pontoise, Orléans and Marseille. The usual disclaimer applies.

private information about their tastes at each period that no other agents, including econometricians, can observe. However, we assume that they coordinate their expectations over future random tastes on a (common knowledge) distribution function, since they cannot learn about it. We first prove that if agents coordinate on a distribution of random tastes that is uncorrelated over time, data that they observe are consistent with this assumption: neither the agents, nor the econometrician can test the common knowledge assumption using data on choice histories. In this sense, expectations are consistent. The absence of correlated unobserved heterogeneity over time along with that of coordination on this belief is thus an untestable assumption (if no other restriction is assumed). Second, we prove that, under this assumption of absence of unobserved heterogeneity for econometricians, the (functional) degree of underidentification is large. The common knowledge d.f. of future random tastes and the discount rate cannot be identified. It is only by fixing those functional parameters to arbitrary constants that we are able to show that the present value of choosing an alternative can be decomposed into its value in the current period and into its future value.

Given arbitrary assumptions on the unidentified objects, the identification proof also suggests a very simple estimation procedure in two stages in the spirit of Manski (1991) and Hotz and Miller (1993). The idea is that the structural Bellman equations do not only depend on deep structural parameters like instantaneous utility functions that are the object of interest but also on future probabilities of choices. First, we recover these probabilities of choice using standard non parametric regressions. As in Hotz and Miller (1993), these probabilities of choice allow us to infer the value functions. Secondly, using the Bellman Equation as a moment condition, we simply run a second non parametric regression, to obtain estimates of the instantaneous utility functions in each alternative. Our estimation procedure is close to, but simpler than Hotz and Miller's since they write Bellman equations in terms of utility functions while we write them in terms of value functions.

This methodology is applied to data on school leaving of French students belonging to the 1963-1973 cohort using two (flexible) parametric stages. The microeconomic model is a standard human capital investment model where students trade off the costs of acquiring an extra year of education against its returns. The interest of the empirical application stems from the huge increase in the education level of young generations in France over the years 1985 to 1993. As it seems empirically plausible in the case of France where there are very few reentries into school or college after leaving, we assume that work is an absorbing state and we study the dynamic choice of staying or leaving the education system. Our identification procedure permits to distinguish current returns, the value of an additional year at school, and future returns, the option value that staying at school brings up because of the possibility of staying longer (Comay, Melnik and Pollatschek, 1973). We show that increases in the education level of young generations were mainly due to increases in the staying rate at the end of high school and beginning of college over this period. Particularly at the end of the period, net returns to staying at school at the beginning of college increased and triggered an increase in the option value of staying at school at the end of high school. In that particular application, we then tighten the structural restrictions by using an income maximizing model and by using another source of information on earnings. These additional restrictions allows us to recover private schooling costs if they are supposed to be deterministic. We estimate life cycle earnings during the 80s and 90s and we show that education returns in earnings were stable over the whole period. Unsurprisingly therefore, increases in the staying rates at school are explained by decreases in schooling costs. We show how these costs can be computed.

We finally return to the general model and explore the assumption where agents have more information than econometricians about other agents. It generates unobserved correlated heterogeneity over time. We show that the (functional) degree of underidentification of structural objects is even larger. Thus, further research should first aim at studying identifying restrictions in the model without unobserved heterogeneity.

Section 2 sets up the theoretical model and provides the main proposition about identification. Section 3 is devoted to the empirical application of these results to school leaving behaviour of the young generations in France. Section 4 provides the extension of these results when we use additional information on life cycle earnings. Section 5 tackles the case with correlated unobserved heterogeneity.

2. The theoretical set-up

We here consider a structural dynamic discrete choice model. We study the non parametric identification of this model using individual data on sequences of choices of alternatives. We know that this type of structural models is generically not identified (Rust, 1994) unless some additional structural restrictions are adopted. What is not known, as far as we are aware of, is the exact degree of underidentification of these models and that is what we study in this section. We first set up the model following up a structural framework that uses results derived by Manski (1992, 1993) and Rust (1994). Second, in order for the paper to be self-contained we present the result derived by Hotz and Miller (1993). We then turn to the main identification result of the paper. We conclude by presenting the estimation strategy of structural objects when identifying restrictions are imposed.

2.1. The structural model

We first state assumptions on preferences and then turn to assumptions on the expectations of the decision maker.

2.1.1. Preferences

We consider a decision maker whose intertemporal utility is additively time separable and whose instantaneous utility is defined over a discrete set of alternatives $i = 1, ., K$. There are no restrictions on instantaneous utilities such as monotonicity or concavity restrictions (Matzkin, 1996). The observable – to the econometrician – state variables that she considers are denoted H. Time is supposed to be one of the state variables in order to simplify the presentation. Furthermore, if H' is a possible subsequent history to H, we assume that it contains "more" information than H. It contains H, all the observable decisions taken by the agent during the current period and new information. Therefore if H' is supposed to vary into a set denoted S, H varies in the same set provided it is completed by a null future history.

There are also state variables that are observed in the current period by the agent but not by the econometrician. As usual, these unobserved variables are modelled as random shocks ε_i affecting utilities in state i, such that:

$$u_i(H, \tilde{\varepsilon}) = u_i^*(H) + \varepsilon_i$$

where $\tilde{\varepsilon} = (\varepsilon_1, ., \varepsilon_K)$. Stochastic assumptions on random shocks are developed below. The timing in a period is the following. At the beginning of the period, the agent is endowed with an history H. First, random values for shocks $\widetilde{\varepsilon}$ are drawn. Second, knowing these shocks and past history H, the agent chooses an alternative, $d = i$. Third, nature picks up at random a future history H' conditional on $H, \widetilde{\varepsilon}$ and d.

Hence, there are two sources of uncertainty: the first one is the shock affecting instantaneous utility, and the second one comes from the transition process from state $H, \widetilde{\varepsilon}$ and decision $d = i$ to state $H', \widetilde{\varepsilon}'$. We will consider that this latter process is exogeneous to the model (see Heckman, Lochner and Taber, 1998 for a different assumption). It influences the decision but the agent takes it as given. In the case of education choices, this transition process could be interpreted as a measure of the educational system's toughness at making students repeat classes.

The state variable therefore consists in $(H, \tilde{\varepsilon})$. Bellman equations relating current value functions in each alternative and future value functions are:

$$v_i(H, \tilde{\varepsilon}) = u_i(H, \tilde{\varepsilon}) + \beta E(\max_j v_j(H', \tilde{\varepsilon}') \mid I) \tag{2.1}$$

where I is the information set of the agent and where β is the discount factor. As time is included in H, the value functions v_i need not be indexed by time. We now state assumptions about the information set, I, assumptions that give structural content to (2.1).

2.1.2. Expectations

So as to solve Bellman equations, the decision maker is endowed with an information set I. We first exhibit an example borrowed from Manski (1993) that proves that it is not sufficient to suppose that the information set consists in history H, chosen alternative d and random shock $\widetilde{\varepsilon}$ only. Namely, in order to be able to compute the dynamic term in Bellman equations, the decision maker should have priors on the transition probabilities:

$$\Pr(H', \widetilde{\varepsilon}' | H, d = k, \widetilde{\varepsilon})$$

for all $\widetilde{\varepsilon}', \widetilde{\varepsilon}, k, H, H'$. Following Manski (1993), the knowledge of this transition process could be infered from the observation of "older selfs". An "older self" is the same agent or another agent sharing the same $(\widetilde{\varepsilon}, k, H)$ with the agent but for whom $(\widetilde{\varepsilon}', H')$ is observed[3]. If the agent observe all older selfs , she can exactly infer:

$$\Pr(H', \widetilde{\varepsilon}' | H, d = i, \widetilde{\varepsilon})$$

As $d = i$ is a deterministic function of $(H, \widetilde{\varepsilon})$ however, the agent will never be able to observe out of equilibrium transition processes

$$\Pr(H', \widetilde{\varepsilon}' | H, d = k, \widetilde{\varepsilon})$$

for $k \neq i$. In this case, Bellman equations have no structural content if no prior is assumed, since the decision maker will not be able to compute the maximand out of equilibrium.

This is for this reason that we have to assume additional structure on the distribution function of $\widetilde{\varepsilon}', H'$ conditional on H, d and $\widetilde{\varepsilon}$. This d.f. belongs to the information set I as in Manski (1992). We derive these assumptions from assumptions on what the agent can observe or not:

Assumption A (Agents' Information Sets):

i/. *Random shocks $\widetilde{\varepsilon}$ are observable to the agent only.*

ii/. *State variables H are observed by everybody.*

iii/. *The expected distribution function of future histories, H', conditional on present and past variables observed by the agent, is equal to the true distribution function of future histories, H', conditional on observable history only, and denoted $P(H' \mid H, d)$. This assumption is common knowledge.*

iv/. *The expected distribution function of random shocks, $\widetilde{\varepsilon}'$, conditional on history, H', present and past random shocks observed by the agent, is equal to the*

[3] If stationarityis assumed, an older self can be another agent. If not, perfect information is assumed.

true distribution function of random shocks conditional on history only and denoted $G(. \mid H')$. This assumption is common knowledge. This distribution function is absolutely continuous with respect to the Lebesgue measure.

Assumption A.i states that agents do not observe more than the econometrician about the other agents. Combined with assumption A.ii, it implies that agents are only able, using observable data, to compute distribution functions of observables, *i.e.* the transition probability function $\Pr(H' \mid H, d = k)$ and choice probabilities $\Pr(d \mid H)$ from the observation of agents ("older selfs"). This is why assumptions A.iii,iv are needed. They detail how expectations in (2.1) can be computed. Since $\widetilde{\varepsilon}$ (or past random shocks) is observable to the agent only, the distribution functions $P(H' \mid H, d = k, \widetilde{\varepsilon})$ and $G(\widetilde{\varepsilon}' | H', \widetilde{\varepsilon})$ cannot be learnt about from the observation of "older selfs". This is why we need to ensure through A.iii,iv that agents exogenously coordinate upon some *prior* distribution functions that are here assumed independent of past random shocks**.** We shall show in the sequel that A.iii,iv. are untestable restrictions. Finally, absolute continuity in A.iv is not necessary but makes the problem smoother.

A note about stationarity is in order. If stationarity is assumed, the observation of what the elders did is sufficient. If not, we implicitely assume that no learning process occurs (see Buchinsky and Leslie, 1997 for a different assumption) and that expectations are perfect. Therefore, non stationary macro shocks could be included in this framework provided that their effects and distribution function are identical across agents and common knowledge.

These assumptions help defining the information set of the agent. It consists in any observables in the past and present, history H and past random shocks, in the distribution functions of future observable trajectories $\Pr(H' \mid H, d = k)$ and shocks $\Pr(\tilde{\varepsilon}' \mid H')$ and in the fact that other agents rationally behave as intertemporal optimizers according to (2.1). Private information about current and past own shocks $\tilde{\varepsilon}$ is however of no value when future discounted benefits are to be computed as the only information about the future is given by the knowledge of $\Pr(H' \mid H, d = k)$ and $\Pr(\tilde{\varepsilon}' \mid H')$ because of assumptions A.iii and A. iv.

Thanks to the linearity of the expectation operator, dynamic terms in Bellman equations are obtained by integrating out random effects $\tilde{\varepsilon}$ of "older selfs" that are unobserved. Their distribution is common knowledge and dynamic terms can be written as:

$$E(\max_j v_j(H', \tilde{\varepsilon}') \mid I) = \underset{\tilde{\varepsilon}}{E} \left(E(\max_j v_j(H', \tilde{\varepsilon}') \mid d = i, H, \tilde{\varepsilon}) \mid d = i, H\right)$$

Bellman equations are therefore written as:

$$v_i(H, \tilde{\varepsilon}) = u_i^*(H) + \tilde{\varepsilon} + \beta E(\max_j v_j(H', \tilde{\varepsilon}') \mid d = i, H) \tag{2.2}$$

All objects on the right hand side are known to the agent at time t and she can therefore compute $v_i(.)$. We shall also assume from now on, with no loss of generality, that $u_i^*(H) = E(u_i(H, \widetilde{\varepsilon})|H)$. Thus, we normalize $G(. \mid H)$ such that ε_i is mean independant of H:

$$E(\varepsilon_i \mid H) = 0$$

In conclusion, this framework leads the agent to behave as if $\widetilde{\varepsilon}$ and $(\widetilde{\varepsilon}', H')$ are independent conditional on H and d (Manski, 1993, assumptions 1 and 2, p126). This is often stated as an assumption, called Conditional Independence in Rust (1994) or Aguirregabiria (1997). Remark that this assumption is here the consequence of Assumption A stating that the agents' coordinate on this belief along with the fact that random shocks on preferences are observable to the agent only. We shall show in the next section that this assumption A is untestable and identifying. We do not claim that it is the only possible identifying assumption. We explore in section 5 the case where agents expect that random shocks are correlated over time because all agent collectively believe so and are able to observe all private values of a permanent component of heterogeneity of the other agents that is unobserved by the econometrician. Agents therefore condition their expectations on their own private value. Such a model exhibits dynamic selection as in Cameron and Heckman (1998).

2.2. Definition of identification

We can now turn to the identification problem that econometricians (and maybe agents) face. Value functions (2.2) can be decomposed as current utilities are:

$$v_i(H, \tilde{\varepsilon}) = v_i^*(H) + \varepsilon_i$$

where:

$$v_i^*(H) = u_i^*(H) + \beta E(\max_j(v_j^*(H') + \varepsilon_j') \mid d = i, H) \tag{2.3}$$

that, in the sequel, we shall improperly call Bellman equations. The structural model consists in equation (2.3). The unknown objects that we want to identify consist in the utility functions $u_i(H)$, the discount rate β, the conditional distribution of unobservables $G(\tilde{\varepsilon} \mid H)$ and the law of revision of the state variables, $\Pr(H' \mid H, d)$.

We shall now define the translation between the reduced form and the structural form. The reduced form probabilities of choice $p_i(H)$ are:

$$p_i(H) = \Pr(d = i \mid H)$$

and the structural form choice probabilities are :

$$\forall i; p_i(H) = \Pr(\arg\max_j (v_j(H, \tilde{\varepsilon})) = i \mid H)) \tag{2.4}$$

To identify these structural objects, the data is assumed to provide us with the reduced form probability distribution of observables H and d at n subsequent periods. For simplicity, we concentrate on two periods though the argument is general. Data are summarized by[4]:

$$\Pr(H, d, H', d') = \Pr(d' \mid H') \Pr(H' \mid d, H) \Pr(d \mid H) \Pr(H) \tag{2.5}$$

It is quite obvious that the structural object $\Pr(H' \mid d, H)$ is directly identified and that knowing initial conditions $\Pr(H)$ does not bring, in general, any information about the structural model. Structural restrictions only bear on $\Pr(d \mid H)$ –

[4] By definition of the state variables, H' "contains" H and d and therefore:

$$\Pr(d' \mid H', H, d) = \Pr(d' \mid H')$$

expression which also stands for $\Pr(d' \mid H')$ if $H \in S$. The question of identification that we explore in this paper can then be phrased as follows:

Definition 2.1. *Let the parameter* $b = \{u_1(H), ., u_K(H), \beta, G(\tilde{\varepsilon} \mid H); H \in S\}$. *Define the set of structural restrictions* $\mathcal{S}$ *as:*

$$\mathcal{S} = \{b; \exists v_1(H), ., v_K(H); H \in S \text{ verifying (2.3)}\}$$

b is identified in $B \subset \mathcal{S}$ if:

$$\forall (b, b') \in B^2 \text{ such that } \forall H \in S; \Pr(d \mid H; b) = \Pr(d \mid H; b')$$

$$\Rightarrow b = b'$$

In order to study identification, it is therefore sufficient to use reduced forms $\Pr(d \mid H)$ for any possible history H as summaries of the data.

Finally, we can return to the point that, under assumption A, the conditional independence assumption is untestable. Under assumption A i,A ii, the agent is in a position where she can only observe the d.f. (2.5). There is no way for her to reject assumptions A.iii, A.iv that $\tilde{\varepsilon}$ and $\tilde{\varepsilon}'$ are independent conditional on H' and that H' is independent of $\tilde{\varepsilon}$ conditional on (H, d).

2.3. Mapping choice probabilities onto values of alternatives

We first study equation (2.4) in order to recover the value functions of the alternatives from their choice probabilities. We here follow Hotz and Miller (1993) and state their result. Define $\tilde{p}(H) = (p_1(H), ., p_K(H))$ and $\tilde{v}^*(H) = (v_1^*(H), ., v_K^*(H))$. Then, equation (2.4) can be written:

$$\tilde{p} = \tilde{p}(\tilde{v}^*)$$

It is well known that the level of the value functions cannot be identified. Let K be the reference alternative. Denote:

$$\forall i < K; w_i(H) = v_i^*(H) - v_K^*(H)$$

and:

$$\tilde{w}(H) = (w_1(H), ., w_{K-1}(H))$$

Similarly, renormalize d.f. $G(. \mid H)$ of $(\varepsilon_1, ., \varepsilon_K)$ to be the d.f. of $(\varepsilon_1 - \varepsilon_K, ., \varepsilon_{K-1} - \varepsilon_K)$ which is a slight but unconsequential abuse of notation. Define the mapping $Q : \mathbb{R}^{K-1} \times D \to P$ such that the ith coordinate of Q:

$$Q_i(x, G) = \Pr_G(\forall j; \varepsilon_i + x_i \geq \varepsilon_j + x_j) = E_G 1(\forall j; \varepsilon_i + x_i \geq \varepsilon_j + x_j)$$

where D is the set of distribution functions on $\mathbb{R}^{K-1}$ which are absolutely continuous with respect to the Lebesgue measure (assumption A.iv), and P is the simplex set in R^K (*i.e.* any element $p \in P$ is such that $p_i \geq 0, \sum p_i = 1$).

Then the set of structural restrictions relates values and choice probabilities through the relationship :

$$\tilde{p}(H) = Q(\tilde{w}(H), G(. \mid H))$$

where $G(. \mid H)$ is the distribution of $(\varepsilon_1 - \varepsilon_K, ., \varepsilon_{K-1} - \varepsilon_K)$ conditional on H.

The following result is due to Hotz and Miller (1993):

Proposition 2.2. *Q is invertible with respect to its first argument, i.e. there exists one and only one mapping q from $P \times D$ to R^{K-1} such that:*

$$x = q(Q(x, G), G)$$

Proof. See Hotz and Miller, 1993.

It implies that:

$$\tilde{w}(H) = q(\tilde{p}(H), G(. \mid H))$$

or element by element:

$$\forall i \in \{1, ., K-1\}; \tilde{w}_i(H) = q_i(\tilde{p}(H), G(. \mid H))$$

The inference problem of the agent is therefore coherent. She knows $\tilde{p}(H)$ and $\tilde{w}(H)$ and therefore can infer the distribution function $G(. \mid H)$ which is the common knowledge d.f. of random shocks defined in A.iii. Expectations are consistent with the data. Remark however that any d.f. $G(. \mid H)$ is compatible with this framework. It suffices to say that this d.f. is common knowledge[5].

[5] As there are no restrictions on value functions $v_i(H)$, functions $\tilde{p}(H)$ are compatible by construction with a static random choice model (see Koning and Ridder, 1997, for tests of compatibility in other cases).

The econometrician however does not know $\tilde{w}(H)$ or $G(. \mid H)$. Reduced-form choice probabilities $\tilde{p}(H)$ can be inferred from the data. But there is no general way of identifying $\tilde{w}_i(H)$ without additional restrictions (see Manski, 1988, or Cameron and Heckman, 1998, for the binary case or Matzkin, 1993, for the polychotomous case). It is the same question than in the static choice model. As a matter of fact, any representation of the values of alternatives $\tilde{w}_i(H)$ corresponding to a distribution function $G(. \mid H)$ is compatible with the data. It corresponds to a more fundamental structural question. The decision maker is supposed to know *a priori* the distribution function of unobserved future shocks but any will do. It is not inferred from the data and we, as econometricians, are in the same akward position. At this stage, we shall therefore suppose that we know the distribution function of random shocks that is common knowledge among agents. We shall write:

$$\tilde{w}_i(H) = q_i(\tilde{p}(H), G(. \mid H)) \equiv q_i(\tilde{p}(H))$$

Let us now turn to structural information provided by Bellman equations.

2.4. Mapping values of alternatives onto utility functions

Dynamic terms in Bellman equations can be written as::

$$E(\max v_j(H', \tilde{\varepsilon}') \mid d = i, H) = E(E(\max v_j(H', \tilde{\varepsilon}') \mid H') \mid d = i, H)$$

and:

$$E(\max v_j(H', \tilde{\varepsilon}') \mid H') = E(v_K(H', \tilde{\varepsilon}') \mid H')$$
$$+E(\max(v_j(H', \tilde{\varepsilon}') - v_K(H', \tilde{\varepsilon}')) \mid H')$$

We write:

$$E(\max(v_j(H', \tilde{\varepsilon}') - v_K(H', \tilde{\varepsilon}')) \mid H') = R(\tilde{w}(H'))$$

where function R implicitly depends on the arbitrary distribution function of the random shocks $G(. \mid H')$ and can be written as:

$$R(x) = E_G \max(x_i + \varepsilon_i - \varepsilon_K)$$

where $x \in R^{K-1}$.

Substracting $v_K^*(H)$ from (2.3), Bellman equations can then be written for $i < K$ as:

$$w_i(H) = u_i^*(H) + \beta E(R(\tilde{w}(H')) \mid H, d = i)$$

$$+\beta E(v_K(H', \tilde{\varepsilon}) \mid H, d = i) - v_K^*(H)$$

and for the reference alternative K:

$$0 = u_K^*(H) + \beta E(R(\tilde{w}(H')) \mid H, d = K)$$

$$+\beta E(v_K(H', \tilde{\varepsilon}') \mid H, d = K) - v_K^*(H)$$

As $E(v_K(H', \tilde{\varepsilon}') \mid H, d = i) = E(v_K^*(H') + \varepsilon'_K \mid H, d = i)$, the system of equations can be written using the results derived in the previous section 2.3, $w_i(H) = q_i(\tilde{p}(H))$, as:

$$q_i(\tilde{p}(H)) = u_i^*(H) + \beta E(R(q(\tilde{p}(H')) \mid H, d = i)$$

$$+\beta E(v_K^*(H') \mid H, d = i) - v_K^*(H) \tag{2.6}$$

and:

$$0 = u_K^*(H) + \beta E(R(q(\tilde{p}(H'))) \mid H, d = K)$$

$$+\beta E(v_K^*(H')) \mid H, d = K) - v_K^*(H) \tag{2.7}$$

2.5. Identification in the general case

We sum up the general result in the two-periods case. It can then be applied to longer sequences, considering pairs of periods[6]:

Proposition 2.3. *Consider an arbitrary value for $\beta \in]0, 1[$, an arbitrary conditional distribution function $G(. \mid H) \in D$ and an arbitrary function $v_K^*(H)$. Then, $\forall i$, $u_i^*(H)$ is just identified in the set S_1 of state variables at the first period.*

[6] We do not assume stationary preferences

Proof. Note first that $\tilde{p}(H)$ is known for any $H \in S_1$ and $\tilde{p}(H')$ is known for any $H' \in S_2$, if the latter set comprises state variables at the second period. Choosing an arbitrary family of conditional distribution functions $G(. \mid H) \in D$ implies that functions q and R are known. Hence, by considering an arbitrary value for β, the quantities:

$$g_i(H) = q_i(\tilde{p}(H)) - \beta E(R(q(\tilde{p}(H')) \mid H, d = i)$$

and:

$$g_K(H) = -\beta E(R(q(\tilde{p}(H')) \mid H, d = K)$$

are identified because the law of motion:

$$P(H' \mid H, d)$$

is known which permits to compute the expectation terms.

Bellman equations (2.6) and (2.7) yield:

$$\forall i; g_i(H) = u_i^*(H) + \beta E(v_K^*(H') \mid H, d = i) - v_K^*(H)$$

Consider an arbitrary function $v_K^*(H)$. Then, if:

$$\forall i; f_i(H) = \beta E(v_K^*(H') \mid H, d = i) - v_K^*(H)$$

the Bellman equations yield:

$$\forall i; u_i^*(H) = g_i(H) - f_i(H)$$

Thus $u_i^*(H)$ are identified. As the proof is constructive and as all structural restrictions have been used, $u_i^*(H)$ is just identified ■

This proposition shows that without strong restrictions, the utilities of alternatives are far from identified. The degree of underidentification is large since it consists in β, G and function $v_K^*(H)$. Some implications of proposition 2.3 are well known in the literature. In particular, that the discount rate is not identified is proved by Manski (1993, p129) where it is shown that any structural dynamic model (*i.e.* $\beta \neq 0$) can be written as a static model (*i.e.* $\beta = 0$).

Restrictions have been used in the literature and could be imposed. But, many of them do not help to identify the model. For instance, if K is an absorbing state (Hotz and Miller, 1993, and below), then

$$v_K^*(H) = u_K^*(H) + \beta E(v_K^*(H') \mid H, d = K)$$

which implies that $g_K(H) = 0$. As $g_K(H) = E(R(q(\tilde{p}(H')) \mid H, d = K)$, $g_K(H)$ is a value that is identifiable from the data. Thus, this restriction has no identifying power, but is indeed testable from the data

Another example of restrictions that are commonly used is to impose a terminal date D to the dynamic model, date after which the value function is not alternative dependent and equal to $V(H_D)$ (Keane and Wolpin, 1998, Belzil and Hansen, 1997). It amounts to set:

$$\forall i < K; \forall H_D, p_i(H_D) = 0, p_K(H_D) = 1, v_K^*(H_D) = V(H_D)$$

$p_i(H_D), \forall i \in \{1...K\}$ can be recovered from the data however. Thus, the first two restrictions are redundant with respect to the information brought by the data. They are testable, not identifying restictions. The last one is not identifying either except if $V(H_D)$ is known.

It can be noted that the identification of the value function of the reference alternative is "less" important than the identification of β and $G(. \mid H)$ because some structural objects can still be identified with no assumptions on $v_K^*(H)$. Replace in (2.6) the value $v_K^*(H)$ from (2.7). Then:

$$q_i(\tilde{p}(H)) - \beta E(R(q(\tilde{p}(H')) \mid H, d = i) = u_i^*(H) - u_K^*(H)$$

$$+\beta(E(v_K^*(H') \mid H, d = i) - E(v_K^*(H') \mid H, d = K))$$

The LHS can be identified if β and $G(. \mid H)$ are known. The RHS measures the difference between average values of two sequences of choices: first, choose i now, K tomorrow, and behave optimally afterwards; Second, choose K now and tomorrow and behave optimally afterwards. It is therefore a normalization of the "marginal" value of decision i in the current period. It will prove to be especially useful in the case where the reference state is absorbing (see below).

Concluding, proposition 2.3 precludes the identification of many structural objects. We shall proceed however by fixing β and $G(. \mid H)$ and in some cases, $v_K^*(H)$, and by trying on an empirical example to study the sensitivity of estimation results. We shall therefore need an estimation method that we now present.

2.6. Estimation strategy

The estimation strategy follows the same line than the proof of identification. In the first stage, the estimation method proceeds as in Hotz and Miller (1993). It consists in estimating choice probabilities $\tilde{p}(H)$ by non parametric or parametric methods. Denote $\hat{p}_n(H)$ such an estimate. From these estimates, estimates of value functions can be computed. The second stage consists in using Bellman equations to recover utility functions from value functions. It is closer to what Altug and Miller (1998) do and is simpler than Hotz and Miller (1993). We therefore concentrate on the second stage.

As $v_K^*(H)$ is not identified without additional data or additional restrictions, we can normalize it without loss of generality to:

$$\forall H; v_K^*(H) = 0$$

In this case, the estimating equations are given in the proof of Proposition 2.3:

$$\begin{aligned} \forall i < K; \quad u_i(H) = & \quad q_i(\tilde{p}(H)) - \beta E(R(q(\tilde{p}(H')) \mid H, d = i) \\ u_K(H) = & \quad -\beta E(R(q(\tilde{p}(H')) \mid H, d = K) \end{aligned}$$

Choose now arbitrary values for β and G. If $\tilde{p}(H)$ were known, we could compute the values in the data of $Y_i = q_i(p(H)) - \beta R(q(p(H'))$ and $Y_K = -\beta R(q(p(H'))$ for all observations. Estimating equations are:

$$E(Y_i \mid H, d = i) = u_i(H)$$

and therefore:

$$\forall i; \; Y_i = u_i(H) + \eta_i$$

where $E(\eta_i \mid H, d = i) = 0$. We can therefore estimate non parametrically $u_i(H)$ using the subsample of individuals having chosen $d = i$. There are no structural restrictions across alternatives. The estimating equations can be used independently and the estimation method is efficient.

In fact utility levels Y cannot be computed from the data since $\tilde{p}(H)$ is unknown. What can be computed from the data is the quantity:

$$\hat{Y}_{in} = q_i(\hat{p}_n(H)) - \beta R(q(\hat{p}_n(H'))$$

The empirical model can therefore be rewritten as:

$$\hat{Y}_{in} = u_i(H) + \eta + \hat{w}_n$$

where $\hat{w}_n = Y - \hat{Y}_n$. Results concerning two step estimates with a non parametric first step could therefore be applied (Newey and McFadden, 1994). In this case however, note that the different equations are not independent. It is clear that using parametric or semiparametric methods follows the same line of argument.

3. The empirical application

We apply these results to data on school leaving in France over the period 1979-1993 for young people aged 20 to 30 in 1993. This empirical application is of special interest because the level of education of young generations in France quickly grew over this decade. Less than 30% of a young generation in 1985 was leaving school at the end of high school with a "baccalauréat". This diploma is the focal point in the French education system that is normally passed at 18 years of age. In 1996, more than 60% of a young generation had reached this level (Emin and Fallourd, 1996, Estrade and Minni, 1996).

On the supply side, the main engine for this growth was driven by the strong emphasis given to education policies at the end of the 80s. It was announced in 1985 that the objective was to lead 80% of a generation to baccalauréat in year 2000. Other education policies, particularly relative to vocational education, have been enacted later in that period. Figure 1 reports that the share of education expenditures in GDP increased very quickly from 1989 onwards. In absolute terms, this increase began earlier as an economic boom started in 1987. As estimated by the ministry of Education, average yearly spending per student steadily increased from 1987 onwards (figure 2) especially for high school vocational programs that were enacted during this period ("Baccalauréat professionnel"). As education in

France is almost completely subsidized by the state, tuitions fees are very low by international standards even in private schools. The impressive success of this supply policy at least in terms of numbers can go through three channels that are supposedly under the control of the education authorities. The first channel is through a decrease in direct costs but over the period, private costs were small and stayed the same. They can act also on indirect costs by building new schools that are closer to where the students live for instance. The third channel is to be less selective and let students who would not have been allowed before, to go to upper levels of schooling.

But on the demand side, this increase also depends on the evolution of expected returns to education over the period. At first glance, it does not seem to be the case since wage inequality remained fairly stable over the past two decades (Bayet and Cases, 1996, Wasmer, 1998). It might seem more reasonable to attribute changes in the rates of return to the changing rate of unemployment for young people that was large (around 25%) and increasing over the period. Note that the unemployment rate sharply decreases with the level of education and that partly explains why the rate of youth unemployment is so large as the most educated are not yet active in the labour market at those ages (Elbaum and Marchand, 1993). The argument that an increasing rate of unemployment has a negative effect on the rates of return is not as clear as it might seem however. In the general case, increases in unemployment have ambiguous effects on the rates of return since they affect both the level and the slope of expected incomes as a function of education (Kodde, 1988, Micklewright, Pearson and Smith, 1989).

What we do in the following two sections is to try to disentangle these various effects on costs, selection and returns, as well as to investigate the determinants of education supply and demand separately, in order to explain what might be the main reasons of this increase in education levels over this period.

3.1. An informal look at the data and a brief description of the French educational system

The dataset we use is drawn from a 1993 French survey on education and labour market histories (*FQP,* INSEE*)*. The survey design is based on dwellings. Every member of the household living in each dwelling is interviewed. Information about past schooling history is comprehensively documented from the very first year at school. We focus our study on young men born between 1963 and 1973 since a first attempt revealed that men and women had very different behaviour. This leaves us with a raw sample of around 2000 individuals aged between 20 and 30 in 1993. As information is retrospective, there is no problem of attrition. There is a slight problem of selection though. Age classes are not quite uniformly distributed, decreasing from 12% of the sample for 20 years old, to 6% for 25 years old and increasing afterwards up to 10% for 30 years old. It is due to the survey design based on dwellings which exclude some young people leaving in collective dwellings (including boarding rooms) or doing their military service. We neglect this phenomenon that does not seem to have an impact on the evolution of the proportion of students staying at school (see below) and assume that selection is exogeneous to educational choices[7]. Finally, some educational histories are (right) censored since 22% of the population are still at school or college in 1993. It has some impact on the descriptive statistics that we present but none on the model as it is written as a dynamic discrete choice model in discrete time.

In France, pupils must go to school between the ages of 6 and 16. The choice of leaving school is only given at age 16. Nevertheless, characteristics of the education history before age 16 can determine the future trajectory. We begin by describing what is the educational system before age 16 in order to justify the way we construct our sample. Before level 14, pupils are supposed to experience quite homogeneous education programs. General education is the standard from level 6 to 14[8]. Only 2.2% of our sample went through classes of specialised education for children with handicaps or other serious difficulties. At level 14, different routes open up. Some pupils can continue in general education, some can switch to short-track vocational education and others can go into apprenticeship with some schooling until they are 16. After level 16, schooling becomes more varied.

[7]A possible way of correcting endogeneous selection would consist in using a weighted procedure (Coslett, 1994), where weights would come from the Census of the Population in 1990. This point is left for further research.

[8]For international comparability, we are labelling classes by the age of the pupils who attend these classes and did not repeat a class.

During the period between 16 and 18, which corresponds to the last three years of high school, education programs are split almost equally into general education and long-track vocational education programs . Diplomas at 18 ("Baccalauréat") can be either general or vocational. Other boys finish their short-track vocational degree between 16 and 18. Our final sample excludes pupils who had been through specialised education, through apprenticeship and we selected out individuals who had not reached at 16 years the level that 12 years old pupils normally attend. Almost all pupils who report being in either of these states before 16 leave school at 16. These (quasi) irreversible events before 16 years of age render our model irrelevant. In our sample, we also corrected, when possible, for misreports in variables and dropped observations otherwise. This leaves us with 1519 individuals in our working sample.

Exit rates from education are reported in table 1. They are reported for the complete sample and two subsamples. The first one comprises boys who reached levels less than 15 at 16 years of age and therefore repeated a class at least once. The other group gathers students who did not repeat a class. The exit rate is low – from 3 to 7% – at levels 14 to 16 – increases to 15 or 25% at levels 17 and 18 (end of high school), drops to 7% at the beginning of college. After level 20, the exit rate increases to more than 20%. All observations related to education levels higher than 23 were discarded because of the small size of these populations. Comparing exit rates in the two subsamples provides evidence on the long-run effect of having repeated a class before sixteen years of age. Exit rates are always larger for those who repeated classes. The number of class repetitions is therefore an important explanatory variable to include in the analysis.

Average probabilities of staying in school by cohorts and education levels are reported in figure 3 (with confidence bands). At levels 14 to 16, staying probabilities are so close to one for all cohorts that no differences can be found across cohorts. This is not true at levels 17 and 18 where increases in the staying probabilities over time are statistically significant. At level 17, this probability increases from 65% in the 1963 cohort to 80% in the 1973 cohort while at level 18, it increases from 80% to 90%. Differences across cohorts are less acute at higher levels (except

20) mainly because sample sizes become too small. In conclusion, the increase in the education level of the young generations in the 80s can be mainly attributed to what happened at education levels 17 and 18.

To complete the description of the data, we also graphed the probability of successfully completing an education level (figure 8). These success probabilities as a function of cohorts are fairly stable at levels 15 and 16, are U-shaped at levels 17 and exhibits a significant upward trend at levels higher than 18. These results show that if anything, education authorities became more lax in their selection policy over the period. It confirms one of the possible argument that was given above to explain the increase in education of the younger generations.

3.2. Estimating choice probabilities

We are interested in modeling the school leaving behaviour of young people in the cohorts 1963 to 1973. In the present analysis, we assume that when leaving school, work is an absorbing alternative. Very few individuals return to the formal schooling system after having left school. In 1993, schooling temporary drop outs concerned 2% of the population only (source: Labor Force Survey, 1993). The only choice therefore is to stay in the education system or to leave forever. We estimate choice probabilities conditional on covariates that is the first stage of the estimation method developed in section 2.6.

For simplicity and as our data mainly comprise discrete variables, we prefer to use flexible parametric methods. We write:

$$p(H) = F(X\gamma)$$

where $F(.)$ is a logistic distribution function and where X are explanatory variables. Attemps to use probit models instead of logit models gave very similar results to the ones that are presented here. This is true that the distribution of shocks is no longer arbitrary and can be recovered from the data if we impose the additional restriction that its argument is a linear index in the covariates (Manski, 1988). If the functional form is as flexible as possible, we are close to non identification however. This is why we believe that the logit assumption is tenable if the specification of the index is sufficiently general. After having found that schooling leaving decisions are significantly different across education levels, we splitted the sample into subsamples according to education levels. As sample sizes for education levels 22 and 23 were too small and since a standard testing procedure gave evidence that it was not illicit to do so, we grouped them with education level 21 and included dummy variables for levels 22 and 23. We then estimated logit models for all subsamples. The dependent variable is the decision to stay at school an additional year at each education level. As we assumed that random shocks are independent across periods for the same individual (assumption A), there is no selection bias.

As our purpose is to have a precise description of these reduced form choice probabilities, we retain all explanatory variables that were significant or close to significance for at least one education level. Explanatory variables consist in individual and parental characteristics. Descriptive statistics are given in table 2. Among variables describing family background that are not time varying, we have information on: Father's and mother's educational attainment that we regrouped after experimentation in a variable describing the maximal educational attainment of both parents; The number of brothers and sisters; Whether the mother had ever been working; Parental (log) incomes[9]. Individual characteristics, that are not time varying, are the year of birth only. Among time-varying variables describing the educational history, we distinguished two dummy variables describing the number of class repetitions (respectively, one and, two or more, none being the reference). Including variables describing educational histories is not innocuous. As a matter of fact, it implies that the number of class repetitions is not chosen by the pupil or student himself/herself. These variables belong to the state variables which transition process is controlled by nature (see section 2.1) *i.e.* teachers and education authorities. It seems to be a plausible assumption in the case of repetitions.[10]

[9] A word about parental income is in order. As no information is available in the sample about incomes, we used information given on occupations at a very disagregated level that are reported in all surveys of INSEE, to impute parental incomes estimated in another survey, the annual Labor Force Surveys. Details are given in appendix A.1

[10] It would be less plausible for education programs distinguishing for instance between a general or vocational orientation even if teachers play a great rôle in orienting pupils into programs. That is why in this version of the paper we prefered not to include those dummy variables. An interesting generalization of our application however would be a model where not only school leaving is the object of choice but also the type of education programs.

Estimation results are given in table 3. Estimates of cohort coefficients are not reported in this table as a graphical representation is much more palatable (figure 4). The most significant determinant of staying probabilities is always the number of class repetitions as was found in the descriptive analyses above. Repeating a class once, twice or more significantly decreases the probability of staying at all levels except the highest (≥ 21). Repeating a class twice or more has a stronger effect than repeating once. Parents' (log) income has a positive effect on the probability of staying at school but it is not always significant. There is no clear tendency of a decreasing effect of income as reported using U.S.data by Mare (1980) and criticized by Cameron and Heckman (1998) in view of the logit assumption. On the other hand, the effects of parental education are only significant at the lowest levels of education (15-17). They even become negative at higher levels but only one coefficient for one education level is significant. A non working mother does have a significant positive impact on staying but the effect is not always significant. Finally, the number of siblings is only significant and negative for one level of education (18).

These results are globally very similar to results in Mare (1980) and Cameron and Heckman (1998). Income effects of education demands are positive, parental education having an additional positive effect possibly due to direct time investments of parents in the education process. The number of siblings have a negative effect if there is some quantity-quality trade-off. Returning to table 2 which describes average values of covariates at different education levels, we see that it confirms these results. The number of siblings decreases with increasing education level while parental education and income increase. This table summarises the only permanent individual effects along educational histories. In the case where random shocks are not independent – Assumption A is not valid – there is a dynamic selection bias as shown by Cameron and Heckman, 1998. We will return to this point in section 5.

In figure 4, predicted choice probabilities of staying at school are plotted for different cohorts and education levels. To obtain net cohort effects, we fixed the values of all other explanatory variables to their average values in the subsamples corresponding to each education level. Confidence intervals for cohort estimates are computed using the variability of estimates and of these average values. Results are similar to what we obtained in the simple description of figure 3. Staying probabilities are close to 1 for levels 15 and 16. At levels 17, 18 and 20, predicted probabilities reproduce the increasing patterns that we found in figure 3. This shows that the model is flexible enough to capture these empirical regularities and that there are no obvious trends in the covariates that explain increases in the education levels over this period.

3.3. The underlying structural model

The implicit structural model that we use is a human capital model with no consumption smoothing. As leaving school is an absorbing state, the two Bellman equations can be written as:

$$V_s(S,t) = \tag{3.1}$$

$$U_s(Y_s - c(S,t)) + \beta E \max(V_s(S',t+1) + \varepsilon_{st+1}, V_w(S',t+1) + \varepsilon_{wt+1} \mid S, d = s)$$

$$V_w(S,t) = U_w(w(S,t), Y_w) + \beta E(V_w(S,t+1) + \varepsilon_{wt+1} \mid S, d = w)$$

where S is the level of schooling, t time, Y_s and Y_w are exogeneous incomes, V_s and V_w are the valuation functions of both states. $c(S,t)$ is the cost of schooling and $w(S,t)$ is the wage. Current preferences for schooling are described by U_s, for work by U_w. $d = s$ and $d = w$ respectively describe the choices of staying in or leaving school. For simplicity, we did not write explicitely the dependence of these functions on other covariates, Z. Also random shocks ε_{st}, ε_{wt} at period t are not added on either sides of these equations to conform with the transformed Bellman equations (2.3) in section 2.1.

As school leaving is an absorbing state, the framework is slightly different than in section 2.6. The formal result of identification still holds however. To make it clearer, we here prefer to prove it again directly. Let $p(S,t)$ denote the probability of staying in school. Denote $q(.)$ the function that maps choice probabilities onto differences of value functions and $R(.)$ the function that maps differences of value functions onto dynamic rents to staying in school – or the option value – as in sections 2.3 and 2.4:

$$q(p(S,t)) = V_s(S,t) - V_w(S,t)$$

$$R(q(p(S',t+1)) = E[(\max(V_s(S',t+1) + \varepsilon_{st+1} - V_w(S',t+1),0) - \varepsilon_{wt+1} \mid S')$$

Substracting the two Bellman equations in (3.1) yields the estimating equation:

$$D(S,t) = q(p(S,t)) - \beta ER(q(p(S',t)) \mid S, d = s) =$$

$$U_s(Y_s - c(S,t)) - U_w(w(S,t), Y_w)$$

$$+\beta(EV_w(S',t+1) \mid S, d = s) - EV_w(S,t+1) \mid S, d = w) \qquad (3.2)$$

The term $D(S,t)$ is identified provided that the function $q(.)$ and β are arbitrarily fixed as in proposition 2.3. The second Bellman equation describing the accumulation of utilities when working does not bring any information on the structural parameters if U_w is unknown and no additional information is used (see section 4 below). Equation (3.2) tells us that $D(S,t)$ is the sum of the current gain of staying at school plus the discounted expected future gain of staying at school, one and only one additional year. $D(S,t)$ is therefore the net expected returns of staying at school at level S at time t and going out at time $t+1$. In contrast, $q(p(S,t))$ is the net expected returns of staying at school at level S at time t including the option value offered by the possibility of staying at school more than one year (Comay, Melnik and Pollatschek, 1973). This option value is equal to $\beta ER(q(p(S',t)) \mid S, d = s)$.

Estimating $D(S,t)$ follows the same line than in section 2.6. Compute for any individual staying at school at level S at time t, the quantity:

$$Y = q(p(S,t)) - \beta R(q(p(S',t+1))$$

Then the second stage of the estimation consists in writing the model:

$$Y = D(S,t) + \eta$$

where by the Bellman equation $E(\eta \mid S, t, d = s) = 0$. Therefore, this equation can be estimated using the subsample of those individuals who stay at school an additional year after level S.

As before Y cannot be computed from the data since $p(S,t)$ is unknown. What can be computed from the data is the quantity:

$$\hat{Y}_n = q(\hat{p}_n(S,t)) - \beta R(q(\hat{p}_n(S',t+1))$$

where $\hat{p}_n(S,t)$ are first stage estimates of the probability of staying at school. These probabilities were estimated in the previous section (3.2). The empirical model can therefore be rewritten as:

$$\hat{Y}_n = D(S,t) + \eta + \hat{w}_n$$

where $\hat{w}_n = Y - \hat{Y}_n$. This measurement error however goes to zero when the sample size increases and the two-step estimate is consistent.

In order to be more illustrative, as choosing the distribution of random shocks $G(.)$ is arbitrary, we can, without loss of generality, focus on the case where $\varepsilon_s - \varepsilon_w$ is logistically distributed and independent of S and t as was used in the previous section. Thus:

$$p(x) = \frac{1}{1+\exp(-x)}$$

therefore its inverse is:

$$q(p) = \ln p - \ln(1-p)$$

and straightforward computations yield:

$$R(q(p)) = -\ln(1-p)$$

Then:

$$\hat{Y}_n = \ln \hat{p}_n(S,t) - \ln(1-\hat{p}_n(S,t)) + \beta \ln(1-\hat{p}_n(S',t+1))$$

As in the logit case, $\hat{p}_n(S,t) = p(X\hat{\gamma}_n)$, the estimating equation takes the form:

$$\hat{Y}_n = X\hat{\gamma}_n - \beta \log(1+\exp(X'\hat{\gamma}_n)) = Z\delta + \varepsilon$$

when $D(S,t) = Z\delta$. Z comprises variables X but could also comprise interactions of those variables as the estimating equation is non linear. This second stage can be non or semi parametric. With discrete covariates, it amounts to standard ANOVA estimation where standard errors are corrected for the estimation at the first stage.

As the covariates used in our analysis are mainly discrete, we chose to use simple OLS techniques. We allow for heteroskedasticity of unknown form. Correction of standard errors at the second stage is derived in appendix B. The regression estimates give the determinants of the net expected returns of an additional year at school. The scale of this variable – the monetary unit – is not identified. Only signs of the coefficients can therefore be interpreted. We chose a value for β which is equal to 0.95. We experimented with other values and the main results that are presented below are robust to moderate changes in the values of β (0.80 to 0.99).

Results describing determinants of expected net returns of staying at school are given in table 4 and the sensitivity to β of these estimates are given in table 5. We included the same covariates than in the previous analysis of choice probabilities. The corrected standard errors are approximately twice the uncorrected standard errors. Still, coefficients are quite precisely estimated. The number of class repetitions have generally a negative and significant effect on net returns but they are a few exceptions at education levels 17, 18. Parental (log) income has a positive effect, generally significant, at all levels except level 16. Parent's education have the expected positive sign except at levels 18 and 19 and the effect of a non working mother depends very much on the level of education being significant and positive at levels 15, 18 and 20 but significant and negative at levels 17 and 19. The effect of the number of siblings is positive except at level 17.

These results concern the value of just one additional year at school. By comparing these effects to choice probabilities which are a measure of the total value of staying at one education level, we can derive the covariate effects on the option value of continuing school, that is the value of further higher education. Equation (3.2) yields:

$$\beta ER(q(p(S',t)) \mid S, d = s) = q(p(S,t)) - D(S,t)$$

We here briefly describe covariates which effects have opposite signs on $p(S,t)$ and $D(S,t)$. This option value is strongly negatively affected by class repetitions at levels 17 and 18. This is probably due to the fact that the probability of repeating a class is likely to increase with the number of class repetitions. The effect of parental education on the option value is unambiguously positive for levels higher than 18 while the effect of parental income is ambiguous. Parents with higher education can help their children with time investments and/or information to go through worthwhile education programs. Other effects are ambiguous.

Table 5 reports the same estimates than table 4 at level 18 contrasting results for different values for β ranging from 0.8 to 0.99. Most coefficients do not change but there are slight differences for parental income and class repetitions coefficients. Changing the value of β changes the decomposition of the total value of staying at school into the current value and the option value. The higher β is, the larger the option value is and if $\beta = 0$, only the former effect matters. It is clear also that if one is willing to assume that one of the covariates has no effect on the current value of staying an additional year at school, this procedure could give a way to identify β. It seems doubtful that an exclusion restriction on any of our covariates is credible in our case.

Estimates of cohort effects are not reported in table 4 but are directly plotted in figure 5 in the same way than in the previous section. Confidence intervals are computed using the variability of estimates and of the average values of explanatory variables at each education level. We shall concentrate on the interpretation of the evolution of net returns between 17 and 20 as it was the most apparent phenomenon explaining the overall increase in education levels (figures 3 and 5). It is also clear from figure 5 that differences across cohorts at levels 17, 18 and 20, show clearer patterns than at other education levels. At level 17, net returns increased over the first period (cohorts 63-69) and decreased over the second period (except 73). At levels 18 and 20, net returns were quite stable in the first period and brutally increased over the second period. As before, we can decompose time patterns in figure 4 into time patterns of net returns (figure 5) and of the option values of staying. That means that over the first period (cohorts 63-69), increases in the value of staying at school were mainly due to increases in net returns at this level. In the second period, that is the reverse. Net returns at level 17 decreased while the option value of staying increased mainly because of the increasing net returns to education level 18. The option value of staying at 18 can have evolved only if the transition probabilities from one level to the next significantly changed. A quick look at these probabilities in the data confirm this pattern at levels 19 and 20 (figure 8). This overall interpretation is interesting because the structural change occurred between cohorts 1969 and 1970 which at levels 17 and 18, is translated into a change in 1987, the exact date of the change in French educational policies. As, at this date, an economic boom started, it is interesting to study the evolution of education returns over this period.

4. Using other sources of information

The previous analysis permits to decompose the value of staying into a current component and a dynamic component. It stays however at a quite descriptive level. If we want to assess the dynamic effects of education policies, we would like to identify other structural objects and in particular a money metric of utility. We can tackle this problem by using a more tightened structural model of human capital investments which is the standard model of income maximization over the life cycle (for instance Keane and Wolpin, 1996, or Belzil and Hansen, 1997). In this case, we also have to use other sources of information about the value of working. In the income maximizing model, it is supposed to be given by a measure of life-cycle incomes. That is what we do in this section, defining the identifying restrictions in the first part, explaining the construction of life-cycle incomes in the second part and finally giving estimates of the costs of schooling in the last subsection. We show that one additional parameter cannot be identified and that obtaining reasonable estimates of raw costs of schooling hinges on the value of that parameter.

4.1. The income maximization model

Assume from now on that current utilities can be written as:

$$U_s(Y_s - c(S,t) = -c(S,t)$$

$$U_w(w(S,t), Y_w) = w(S,t)$$

In this case, the Bellman equation related to work can be written as:

$$V_w(S,t) = w(S,t) + \beta E(V_w(S,t+1) \mid S, d = w)$$

hence, assuming a terminal or retirement date, T:

$$V_w(S,t) = E(\sum_{\tau=0}^{T-t} \beta^\tau w(S,t+\tau))$$

and the value function of the reference alternative is now identified.

The income process also gives additional identifying power in terms of the shocks affecting current utilities. The shock affecting utilities at work ε_w can be identified from the distribution of wages of those who work as, in our case, they are non selected random samples of the whole population:

$$\varepsilon_w = w(S,t) - Ew(S,t)$$

Observing this process have however no identifying power with respect to the random shock ε_s. An additional identifying restriction is necessary in order to be able to express all the objects of interest in equation (3.1) in the same measurement unit. To fix ideas, consider first the assumption:

$$\varepsilon_s = 0 \tag{4.1}$$

which means that schooling costs are deterministic functions of covariates only. Recall that the distribution function of differences of random shocks $\varepsilon_s - \varepsilon_w$ which was labeled $G(.)$ in the previous analysis was not identified. Restriction (4.1) would allow to use the following method. Estimate non parametrically the distribution function $G(.)$ of $-\varepsilon_w$ using the information on wages. Then compute estimates of functions $q(.)$ and $R(.)$ and reestimate the model of the previous section.

This latter procedure heavily depends on the restriction (4.1). A weaker restriction implied by (4.1) is to assume that:

$$V(\varepsilon_s - \varepsilon_w) = V\varepsilon_w \tag{4.2}$$

In this case, the only additional object that is identified is the standard deviation of the distribution function $G()$ say $\sigma(S,t)$. This scale factor which is not identified using data on school leaving behaviour only is identified using (4.2). This scale factor is also the monetary unit that permits to evaluate returns estimated in the previous section in the same unit than wages. Under these conditions, (3.2) yields:

$$\sigma(S,t)D(S,t) = -c(S,t) - w(S,t)$$

$$+\beta[EV_w(S', t+1) \mid S, d = w) - EV_w(S, t+1) \mid S, d = w)]$$

where $D(S, t)$ are net returns to education as estimated in the previous section. This equation is the simple translation of the accounting equation that net returns evaluated in the same monetary unit than the wage, (σD), are equal to raw returns minus costs. Therefore the estimating equation for schooling costs becomes:

$$c(S, t) = -\sigma(S, t)D(S, t) - w(S, t)$$

$$+\beta[EV_w(S', t+1) \mid S, d = w) - EV_w(S, t+1) \mid S, d = w)] \quad (4.3)$$

$c(S, t)$ is a money metric of the subjective instantaneous cost of staying at school and is identified under the restrictions (4.2), and if $G(.)$ and β are chosen arbitrarily.

4.2. Life cycle incomes and schooling costs

The survey that we used to study educational histories is not well suited to estimating earnings equations in the 1980s or 1990s since only information in 1977, 1985 and 1993 are available (Goux and Maurin, 1995). High frequency changes could be lost. This is why we turn to other more complete surveys on earnings, the Labour Force Surveys which give information over the whole period. For non participants and unemployed persons, we imputed a value 0 for earnings in order to conform with the income maximizing model, neglecting however any unemployment benefits, minimum income or transfers from other members of the family. We also tried to correct for non responses (see appendix A.2). We estimated earnings equation year by year over the period 1982 to 1996:

$$w(S, t) = Z\beta_t + \upsilon_t$$

where Z consists in the level of schooling, dummy variables for age under 25 and a squared polynomial for larger ages and dummy variables for short and long vocational educations. Returns to different education levels are given in figure 6. It clearly appears that apart between 1989 and 1990, these returns are fairly constant over the period. The break in the series in 1990 is, in our opinion, entirely attributable to a change in the way earnings are reported in the surveys (see appendix A.2).

We therefore consider it as a spurious effect. That education returns are constant is not a complete surprise in the case of France (Goux and Maurin, 1994). This is true that Wasmer (1997) find that education returns in wages fell during this period but he does not take into account non participation and unemployment as we do. As the employment probability increases with education, our results indicate that this latter effect compensated the diminishing education returns in earnings over the period. Because of this constancy and the inconsistency in the 1990 change in the survey design, we preferred to impose that education returns are constant over the period by estimating an earnings equation using the main survey FQP in 1993. We therefore implicitly assume that changes in supplies of different levels of education during the period did not affect skills relative prices. It may be because demands for different skills are rather inelastic, that compensating shifts in demands and supplies occurred in that period or that data is plagued with measurement errors. It would render difficult the simulation of economic policies such as increasing tuition costs in a general equilibrium as performed by Heckman, Lochner and Taber, (1998). The estimation of the earnings function and of the function of standard error of earnings ($\sigma(S,t)$) derived using (4.2) is explained in appendix A.2 and estimates are reported there.

If returns do not change, changes in the value of schooling can be only attributed to changes in schooling costs, estimation of which we now turn. The estimation of (4.3) proceeds as before. We plug in (4.3) first stage estimates of the earnings equation, the earnings variance and the choice probabilities. Figure 7 plots the resulting estimates of schooling costs against cohorts by education level. Estimates are given in 1981 French Francs per month. Costs are positive for levels 15, 17 and 18 but are negative at levels 16 and 19. They seem incredibly high and these estimates clearly depend on assumption (4.1) that may not be adapted to this example. Furthermore, except at levels 17 where we find the U-shape that was already apparent in figure 6, no evolution is really significant over the period. It points out the crucial importance of the identifying assumption (4.2) about the variance of the preference random error when at school to perform structural estimation of dynamic discrete choice models. There seems to be nothing in the data that allows us to precisely identify subjective costs of schooling

5. Identification with correlated unobserved heterogeneity

We now briefly return to the question of introducing correlated unobserved heterogeneity such as unobserved permanent tastes or abilities in this kind of model. It will not be a surprise that the underidentification result carries over to this case but it is interesting to develop at least the first steps of the proof. As argued in section 2, assumption A is an untestable identifying assumption. We therefore have to move away from assumption A by assuming, for instance, that there exists a one-dimensional state variable θ which is observed by the agents but not by the econometrician. The utility function is now written as:

$$u_i(H, \tilde{\varepsilon}) = u_i^*(H, \theta) + \varepsilon_i$$

and assumption A is replaced by:

Assumption B (Agent's Information Sets with Heterogeneity):

i/. *Random shocks $\widetilde{\varepsilon}$ are observable to the agent only.*

ii/. *State variables (H, θ) are observed by anyone but θ is unobserved by the econometrician.*

iii/. *The expected distribution function of future histories, H', conditional on present and past variables observed by the agent, is equal to the true distribution function of future histories, H', conditional on history only, and denoted $P(H' \mid H, d, \theta)$. This assumption is common knowledge.*

iv/. *The expected distribution function of random shocks, $\widetilde{\varepsilon}'$, conditional on history, H', θ, present and past random shocks observed by the agent, is equal to the true distribution function of random shocks conditional on history only and denoted $G(. \mid H', \theta)$. This assumption is common knowledge. This distribution function is absolutely continuous with respect to the Lebesgue measure.*

Therefore, the only formal difference with assumption A, is to permit to the agent to condition her expectations on θ: in other words, we are now in a world where there are as many Bellman equations as unobserved types**.** It is straightforward to show that, for the agent, this assumption is consistent with the data by repeating the argument used in section 2 for any value of θ. It is also straighforward that this framework also leads to the unconsequential assumption that $\tilde{\varepsilon}$ and $\tilde{\varepsilon}'$ are independent conditional on (H, θ). It is therefore also untestable. In this case, equation (2.3) can be rewritten as:

$$v_i^*(H, \theta) = u_i^*(H, \theta) + \beta E(\max_j(v_j^*(H', \theta) + \varepsilon_j') \mid d = i, H, \theta) \tag{5.1}$$

The difference from the previous case arises from the presence of dynamic selection biases (Cameron and Heckman, 1998, Taber, 1996). However it is straightforward, because of the convexity of the max-operator among other things, that equation (5.1) is not linear in θ even if $u_i^*(H, \theta)$ is linear in θ. The additivity in θ postulated by Cameron and Heckman cannot be used here.

To make things precise, we will, from now on, restrict the model and show that the degree of underidentification is more acute than before. Suppose that there are only two alternatives, one of which is absorbing, to conform with the model of human capital developed above. Also, correlated unobserved heterogeneity is supposed to enter the utility function of the first state only. Instantaneous utilities are therefore supposed to be given by:

$$u_1(H, \tilde{\varepsilon}) = u_1^*(H, \theta) + \varepsilon_1$$

$$u_2(H, \tilde{\varepsilon}) = u_2^*(H) + \varepsilon_2$$

Moreover, suppose that θ can only take two values 0 and 1, and that the initial distribution $r(\theta \mid H_0) = r(\theta)$ is independent of H_0. Finally, for simplicity, suppose that the distribution functions of future random shocks and of the transition process are independent of θ:

$$G(. \mid H, \theta) = G(. \mid H)$$

$$\Pr(H' \mid H, d, \theta) = \Pr(H' \mid H, d) \tag{5.2}$$

Even under these stringent assumptions, there are still dynamic selection biases since tastes are correlated over time: indeed, knowledge of θ gives access to information about future values that the econometrician does not have. Bellman equations are written as:

$$v_1^*(H,\theta) = u_1^*(H,\theta)+\beta E(R(v_1^*(H',\theta)-v_2^*(H',\theta)) \mid d=1,H)+\beta E(v_2^*(H',\theta) \mid d=1,H) \tag{5.3}$$

$$v_2^*(H,\theta) = u_2^*(H) + \beta E(v_2^*(H',\theta) \mid d=2,H) \tag{5.4}$$

where function $R(.)$ is defined as in section 2 and depends on the d.f. $G(\varepsilon_1 - \varepsilon_2 \mid H)$. As the value function of the second state is not identified we can normalize the problem by setting $v_2^*(H',\theta) = 0$, and as in proposition 2.3, by setting $G(. \mid H)$ and β to arbitrary values. Denoting $q(.)$ the inverse of $G(. \mid H)$, the only structural equation is (5.3) and can be written as:

$$q(p(H,\theta)) = u_1^*(H,\theta) + \beta E(R(q(p(H',\theta) \mid d=1,H) \tag{5.5}$$

where $p(H,\theta) = P(d=1 \mid H,\theta)$. The only object of interest to identify is $u_1^*(H,\theta)$ but the result of proposition 2.3 cannot be applied since $p(H',\theta)$ is unknown. What the data give us is, apart from the transition process $\Pr(H' \mid H, d=1)$ which does not depend on θ by assumption (5.2), is the choice probability $s(H) = p(d=1 \mid H)$ which is the mixture:

$$s(H) = p(H,1)\Pr(\theta=1 \mid H) + p(H,0)\Pr(\theta=0 \mid H) \tag{5.6}$$

It is also straighforward to show that assumption (5.2) implies that if H' is a subsequent history of $(H, d=1)$:

$$\Pr(\theta=1 \mid H') = \frac{\Pr(\theta=1, H' \mid H, d=1)}{\Pr(H' \mid H, d=1)} = \Pr(\theta=1 \mid H, d=1)$$

since $H' \perp \theta \mid H, d$. Therefore:

$$\Pr(\theta=1 \mid H') = p(H,\theta)\frac{\Pr(\theta=1 \mid H)}{s(H)}$$

which is the source of dynamic selection biases (Cameron and Heckman, 1998).

Consider now dates 0 and 1. Structural restrictions are given by (5.5) and (5.6):

$$\begin{cases} \theta=0,1 \quad q(p(H_0,\theta)) = & u_1^*(H_0,\theta) + \beta E(R(q(p(H_1,\theta) \mid d=1,H_0) \\ s(H_0) = & p(H_0,1)r(1) + p(H_0,0)r(0) \\ s(H_1) = & p(H_1,1)\dfrac{p(H_0,1)r(1)}{s(H_0)} + p(H_1,0)\dfrac{p(H_0,0)r(0)}{s(H_0)} \end{cases} \tag{5.7}$$

We can now state the (under) identification result:

Proposition 5.1. *Let* $p(H_1,1)$ *and* $p(H_0,1)$ *be two arbitrary functions taking their values between* 0 *and* 1 *such that there exists* $\alpha_0 > 0$ *such that:*

$$\min(\frac{s(H_0)}{p(H_0,1)}, \frac{1-s(H_0)}{1-p(H_0,1)}) \geq \alpha_0 \tag{5.8}$$

$$\frac{s(H_0)}{p(H_0,1)} \min(\frac{s(H_1)}{p(H_1,1)}, \frac{1-s(H_1)}{1-p(H_1,1)}) \geq \alpha_0 \tag{5.9}$$

Consider an arbitrary value $0 < r(1) \leq \alpha_0$. *Then* $u_1^*(H_0,\theta)$ *is just identified.*

Proof. Consider the second equation in (5.7). As $r(0) = 1 - r(1)$, (5.8) and $0 < r(1) \leq \alpha_0$ it implies that $0 \leq p(H_0,0) \leq 1$. Consider the third equation in (5.7). Then (5.9) and $0 < r(1) \leq \alpha_0$ implies that $0 \leq p(H_1,0) \leq 1$. Therefore the second and third equations yield a consistent value for $p(H_0,0)$ and $p(H_1,0)$. Replace now all these quantities in the first equation of (5.7) for $\theta = 0,1$.□

This result means that the behaviour of one of the groups can be considered as almost arbitrary provided that conditions (5.8) and (5.9) are satisfied. This is far from surprising as the model without correlated unobserved heterogeneity is not identified. It is true that some additional restrictions could be imposed on this problem. This is why we tend to think however that it will be more fruitful in future work to look for identifyng restrictions in the case with no unobserved correlated individual effects before moving to the case where there are individual effects.

6. Conclusions

We showed that in a structural dynamic discrete choice model where there are no restrictions on utility functions of the alternatives, an assumption on agents' information sets that implies that random shocks are independent over time is not testable. Under this assumption, we also showed that the discount factor, the distribution function of unobservables, the value function of the reference alternative and utility functions cannot be identified. This is because if the discount factor, the distribution function of unobservables, the value function of the reference alternative are arbitrary, utility functions can be just identified. These results are obtained under very weak assumptions and could be seen as a benchmark case.

There is however some hope that restrictions on utility functions like exclusion restrictions could bring additional information on the structure and we see that question as a promising avenue for research. Adopting a model where there is an additional continuous variable of interest like consumption or effort which is smoothed over the life-cycle could also result in better identification but there is some doubt as exemplified in Blundell, Magnac and Meghir (1997).

If unidentified objects are fixed *a priori,* the identification proof provides a convenient and simple way of estimating utility indices using moment conditions. It also provides, in the case of a model where a state is absorbing, a convenient decomposition of the dynamic value of an alternative into a short-term value and an option value. This property is easily extended to the general case. We also demonstrated on the empirical example on school leaving in France that this framework provides information on the dynamics of school-leaving behaviour (conditional on the belief that assumptions about unidentified objects is founded). If additional information on the value of the reference alternative is used as it seems natural in an model with an absorbing state, we also showed that there is an additional identifying assumption that is untestable. The estimation of private costs of schooling is therefore very fragile.

REFERENCES

Aguirregabiria, V., 1997, "Estimation of Dynamic Programming Models with Censored Decision Variables", *Investigaciones Economicas,* 21:167-208.

Altug, S., and R., Miller, 1998, "The Effect of Work Experience on Female Wage and Labour Supply", *Review of Economic Studies,* 65:45-85.

Bayet, A., and C., Cases, 1996, "Earnings Inequality in France: evolution 1967-1991", INSEE, WP9617.

Belzil, C., and J., Hansen, 1997, "Estimating the Returns to Education from a Non-Stationary Dynamic Programming Model", paper presented at the AEA conference, Maastricht.

Blundell, R.W., T.,Magnac, C.,Meghir, 1997, "Savings and Labour Market Transitions", *Journal of Business Economics and Statistics,* 15:153-164.

Buchinsky, M., and P.,Leslie, 1997, "Educational Attainment and the Changing U.S. Wage Structure", working paper, Yale Univ.

Cameron, S.V., and J.J., Heckman, 1998, "Life-Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy,* 106:262-333.

Conway, Y., A.,Melnik and M.A.,Pollatschek, 1973, "The Option Value of Education and the Optimal Path for Investment in Human Capital", *International Economic Review,* 421-435;

Coslett, S.R., 1994, "Estimation from Endogeneous Stratified Samples" in eds Maddala, Rao and Vinod, *Handbook of Statistics,* North Holland: Amsterdam, 11:1-41.

Eckstein, Z., and K.,Wolpin, 1989, "The Specification and Estimation of Dynamic Discrete Choice Models", *Journal of Human Resources,* 24:562-98.

Elbaum, M. and O.,Marchand, 1994, "Emploi et Chômage des Jeunes dans les Pays Industrialisés: la Spécificité Française", *Travail et Emploi,* 58:111-121.

Emin, J.C., and P., Fallourd, 1996, "L'état du système éducatif français", eds INSEE, *Données Sociales,* INSEE:Paris, 40-47.

Estrade, M.A., and C., Minni, 1996, "La Hausse du Niveau de Formation", *INSEE Première,* 488:55-58.

Goux, D. and E.,Maurin, 1994, "Education, Experience et Salaires: Tendances Récentes et Evolutions de Long Terme", *Economie et Prévision,* 116:155-173.

Heckman, J.J., L.,Lochner and C.,Taber, 1998, "General-Equilibrium Treatment Effects: A Study of Tuition Policy", *American Economic Review,* 88:381-385.

Hotz, V.J., and R.A., Miller, 1993, "Conditional Choice Probabilities and the Estimation of Dynamic Models", *Review of Economic Studies,* 60:497-529.

Hotz, V.,J., R.A., Miller, S., Sanders and J., Smith, 1994, "A Simulation Estimator for Dynamic Models of Discrete Choice", *Review of Economic Studies,* 61:265-289.

Keane, M. and K.,Wolpin, 1996a, "The Career Decisions of Young Men", *Journal of Political Economy,* 105:473-522.

Keane, M. and K.,Wolpin, 1996b, "Eliminating Race Differences in School Attainment and Labor Market Success", WP: U. of Minnesota.

Kodde, D.A., 1988, "Unemployment Expectations and Human Capital Formation, *European Economic Review,* 32:1645-1660.

Koning, R. and G.,Ridder, 1997, "The identification of discrete choice models", working paper, Vrije Universiteit Amsterdam.

Manski, C.F., 1988, "Identification of Binary Response Models", *Journal of the American Statistical Association,* 83:729-738.

Manski, C.F., 1991, "Non Parametric Estimation of Expectations in the Analysis of Discrete Choice under Uncertainty", in Barnett, Powell and Tauchen, eds *Non Parametric and Semi Parametric Methods in Econometrics and Statistics,* Cambridge University Press: New York, 259-275.

Manski, C.F., 1992, "Adolescent Econometricians: How do Youth Infer teh Returns to Schooling?", in eds C.T.,Clotfelter and M.,Rotschild, *Studies of Supply and Demand in Higher Education,* NBER and The Univesity of Chicago Press: Chicago.

Manski, C.F., 1993, "Dynamic Choices in Social Settings", *Journal of Econometrics,* 58:121-136.

Mare, R.D., 1980, "Social Background and Social Continuation Decisions", *Journal of the American Statistical Association,* 75:295-305.

Matzkin, R., 1993, "Nonparametric Identification and Estimation of Polychotomous Choice Models", *Journal of Econometrics,* 58:137-168.

Matzkin, R., 1994, "Restrictions of Economic Theory in Nonparametric Methods", *Handbook of Econometrics,* vol.4: 2524-2554

Micklewright, J., M.Pearson, S.,Smith, 1989, "Unemployment and Early School Leaving", IFS London.

Newey, W.K., and, D., McFadden, 1994, "Large Sample Estimation and Hypothesis Testing", in eds Engle and McFadden, *Handbook of Econometrics,* North Holland:Amsterdam, IV: 2113-241.

Rust, **J.,** 1994, "Structural Estimation of Markov Decision Processes" in eds Engle and McFadden, *Handbook of Econometrics,* North Holland:Amsterdam, IV: 3081-143.

Rust, J., 1995, **"**Numerical Dynamic Programming in Economics", in eds Amman, Kendrick and Rust, *Handbook of Computational Economics,* North Holland: Amsterdam.

Taber, C.R., 1996, "Semiparametric Identification and Heterogeneity in Discrete Choice Dynamic Programming Models", WP Northwestern U.

Wasmer, E., 1998, "Some Labor Market Consequences of Demographic Trends: US and France 1964-97, CEP, LSE.

APPENDICES

A. Data appendix

A.1. The construction of parents' incomes

No information on parental incomes is available in the survey (FQP) recording educational histories. Detailed occupation of the parents are however reported. The so called variable "Catégorie Socio-Professionnelle" can be used as a predictor of incomes because it is constructed on the basis of the type of technical skills, education and incomes in each occupation. We use Labour Force Surveys between 1983 and 1993 to estimate incomes of both parents if they have ever been working using education levels, age and this occupational variable. The partial correlation of incomes and this variable is always very significant all over the period. Instead of using parental incomes at each date and given the stability of earnings over the period (see section 4.2), we prefered to compute an estimate of permanent income by a minimum distance procedure imposing the equality of coefficients over the whole period. We then computed in our working sample, the predicted parental (log) incomes using these estimates.

A.2. The construction of life cycle earnings

Labor Force surveys between 1982 and 1997 are representative of French households. They record monthly earnings of wage workers in a slightly different way before 1990 and after 1990. In the 80s, it is not known whether earnings include bonuses or not. They are reported by bands (about 20 bands between 500FF to 30000FF monthly). In the 90s, earnings include bonuses and a precise value is reported. In some case bands are still used. In order to have a constant framework over the whole period, we also grouped earnings by bands in the second period. Results however still exhibited a spurious change in 1990. When estimating earnings equations as presented in the text, we used the usual endogeneous sampling correction (Coslett, 1994) in order to correct for non responses (including self employed workers) assuming that there is no selection issue. Those individuals who do not respond the wage question are supposed to be an exogenously selected sample of those who work. We weighted observations by the probabilities of responding to the wage question.

We use a slightly different framework to estimate the earnings function and the standard error earnings function using FQP in 1993. We use pseudo maximum likelihood methods and corrected for endogenous sampling. The variance-covariance of estimates are computed using the $J^{-1}IJ^{-1}$ expression. Results are presented in the following table:

	Wage Equation		
	Estimate	Standard Error	Student
Intercept	-5874,5	(915,88)	-6,41
Level=14	-1175,17	(276,21)	-4,25
Level=15	-846,13	(152,08)	-5,56
Level=17	-273,85	(339,32)	-0,81
Level=18	852,45	(197,47)	4,32
Level=19	3155,86	(1721,11)	1,83
Level=20	2537,68	(588,00)	4,32
Level=21	3178,25	(744,39)	4,27
Level=22	3171,65	(4532,69)	0,7
Level=23	3099,99	(732,26)	4,23
Age/10	4389,85	(503,06)	8,73
(Age/10) squared	-357,2	(68,38)	-5,22

	Standard Error of the Wage Equation		
	Estimate	Standard Error	Student
Intercept	3,91	(0,99)	3,96
Level=14	-0,1	(0,16)	-0,63
Level=15	0,21	(0,28)	0,75
Level=17	-0,37	(0,11)	-3,38
Level=18	0,01	(0,03)	0,56
Level=19	1,14	(0,41)	2,79
Level=20	0,28	(0,14)	1,99
Level=21	0,52	(0,25)	2,08
Level=22	0,37	(0,56)	0,66
Level=23	0,29	(0,12)	2,52
Age/10	1,85	(0,53)	3,51
(Age/10) squared	-0,19	(0,06)	-2,93

Note: Pseudo Log Likelihood=-8,01

Table A1:Earnings and Standard Errors of Earning (FQP, 93)

B. Asymptotic Standard Errors of the Two Step Estimates

Our aim here is to get an expression of the asymptotic variance covariance matrix of the second stage estimates. This step consists in regressing an estimate of instantaneous utility on variables Z. The computed dependent variable is, in the Logit case:

$$\hat{Y}_n = X\hat{\gamma}_n - \beta \ln(1 + e^{X'\hat{\gamma}_n})$$

where (X, X') concern the same individuals. We seek δ such that:

$$S = Z\delta + \eta \tag{B.1}$$

Thus, a natural two step estimator of γ is given by:

$$\hat{\delta}_n = (Z'Z)^{-1}Z'\hat{Y}_n \tag{B.2}$$

$\hat{\gamma}_n$ converges towards γ and a straightforward Taylor expansion yields:

$$Z'\hat{Y}_n \# Z'Y + Z'\frac{\partial Y}{\partial \gamma}(\hat{\gamma}_n - \gamma) \tag{B.3}$$

where:

$$Z'\frac{\partial Y}{\partial \gamma} \# Z'X(1 - \frac{1}{1 + e^{-X'\hat{\gamma}_n}})$$

Plugging (B.3) and (B.1) in (B.2) gives us:

$$\hat{\delta}_n - \delta = (Z'Z)^{-1}Z'\eta + (Z'Z)^{-1}Z'\frac{\partial Y}{\partial \gamma}(\hat{\gamma}_n - \gamma)$$

It is easy to show that η and $(\hat{\gamma}_n - \gamma)$ are asymptotically independent. By definition, η is a function of X' and X and $(\hat{\gamma}_n - \gamma)$ is constructed such that it is asymptotically independent of X and X'. Thus, denoting V_γ the asymptotic variance-covariance matrix of $\hat{\gamma}_n$, the variance covariance matrix of $\hat{\delta}_n$ can be written as:

$$\begin{aligned} V\hat{\delta}_n &= (Ez_i'z_i)^{-1}E(z_i'\eta_i^2 z_i)(Ez_i'z_i)^{-1} + \\ & \quad (Ez_i'z_i)^{-1}E(z_i'\frac{\partial y_i}{\partial \gamma}V_\gamma \frac{\partial y_i}{\partial \gamma'}z_i)(Ez_i'z_i)^{-1} \end{aligned}$$

where the first term is the White standard error, robust to heteroskedasticity and the second term accounts for the fact that γ is estimated in the first step.

Level	Education at 16 >15	Education at 16<16	Complete Sample
14	NA	0,045	0,045
		(0,008)	(0,008)
15	NA	0,043	0,026
		(0,006)	(0,004)
16	0,033	0,092	0,062
	(0,006)	(0,010)	(0,006)
17	0,192	0,300	0,244
	(0,013)	(0,016)	(0,010)
18	0,108	0,209	0,146
	(0,012)	(0,021)	(0,011)
19	0,058	0,103	0,071
	(0,010)	(0,021)	(0,010)
20	0,184	0,393	0,232
	(0,018)	(0,041)	(0,017)
21	0,185	0,209	0,190
	(0,028)	(0,062)	(0,026)
22	0,235	0,269	0,242
	(0,042)	(0,087)	(0,038)
23	0,508	0,333	0,479
	(0,065)	(0,136)	(0,059)
Number of Obs.	793	726	1519

Source: Formation Qualification Professionelle Survey (INSEE, 1993)
Exit rate, standard deviation between parenthesis

Table 1:Exit Rates from Education

LEVEL	14	15	16	17	18	19	20	21	22	23
Number of Siblings	2,50	2,26	2,23	2,06	1,87	1,91	1,87	2,05	2,10	1,97
	(0,08)	(0,05)	(0,05)	(0,05)	(0,05)	(0,06)	(0,07)	(0,12)	(0,16)	(0,20)
Living in Paris	0,17	0,19	0,20	0,21	0,27	0,27	0,30	0,38	0,38	0,41
	(0,02)	(0,01)	(0,01)	(0,01)	(0,01)	(0,02)	(0,02)	(0,04)	(0,05)	(0,08)
Rural Area	0,24	0,22	0,21	0,20	0,14	0,12	0,08	0,06	0,08	0,05
	(0,02)	(0,01)	(0,01)	(0,01)	(0,01)	(0,01)	(0,01)	(0,02)	(0,03)	(0,04)
#Class Repeated	1,27	0,67	0,75	0,92	0,91	0,97	1,06	1,04	1,07	1,16
	(0,02)	(0,02)	(0,02)	(0,02)	(0,03)	(0,03)	(0,04)	(0,06)	(0,09)	(0,14)
Parents’ Education: College	0,20	0,28	0,31	0,34	0,45	0,51	0,57	0,63	0,65	0,73
	(0,02)	(0,01)	(0,01)	(0,01)	(0,02)	(0,02)	(0,02)	(0,04)	(0,05)	(0,07)
Parents’ Education: High School	0,60	0,57	0,55	0,54	0,46	0,41	0,35	0,32	0,32	0,24
	(0,02)	(0,01)	(0,01)	(0,01)	(0,02)	(0,02)	(0,02)	(0,03)	(0,05)	(0,07)
Log(Parental Income)	5,75	5,79	5,81	5,85	5,91	5,93	5,96	5,97	5,99	6,00
	(0,01)	(0,01)	(0,01)	(0,01)	(0,01)	(0,01)	(0,01)	(0,02)	(0,03)	(0,04)
Mother Not Working	0,20	0,18	0,18	0,15	0,14	0,14	0,15	0,14	0,12	0,11
	(0,02)	(0,01)	(0,01)	(0,01)	(0,01)	(0,01)	(0,02)	(0,03)	(0,03)	(0,05)

Source: Formation Qualification Professionnelle survey (INSEE, 1993)
Estimated Mean, Standard Error between parentheses

Table 2: Descriptive Statistics by Levels of Education

Education Level	15	16	17	18	19	20	>20
Intercept	-2,33	1,21	-6,65	-3,54	-2,92	-6,03	2,45
	(0,47)	(0,27)	(0,15)	(0,26)	(0,56)	(0,56)	(0,29)
Non Working Mother	0,48	0,14	0,33	0,63	0,37	0,37	-0,74
	(0,47)	(0,27)	(0,15)	(0,26)	(0,56)	(0,56)	(0,29)
Number of Siblings	-0,12	-0,05	-0,05	-0,14	0,04	0,04	0,01
	(0,47)	(0,27)	(0,15)	(0,26)	(0,56)	(0,56)	(0,29)
Parents Highest Educ.:	1,62	1,45	1,31	-0,42	-0,49	-0,49	0,12
College	(0,47)	(0,27)	(0,15)	(0,26)	(0,56)	(0,56)	(0,29)
Parents Highest Educ.:	1,02	0,83	0,21	-0,99	-0,76	-0,76	-0,25
High School	(0,47)	(0,27)	(0,15)	(0,26)	(0,56)	(0,56)	(0,29)
Parent's Log income	1,00	0,20	1,23	1,12	1,26	1,26	-0,11
	(0,64)	(0,43)	(0,24)	(0,38)	(0,65)	(0,65)	(0,57)
Class repetition=1	-1,14	-0,60	-0,30	-0,76	-1,48	-1,48	0,04
	(0,46)	(0,28)	(0,15)	(0,27)	(0,58)	(0,58)	(0,33)
Class repetition>1	-1,94	-1,36	-0,68	-1,31	-2,07	-2,07	-0,37
	(0,47)	(0,27)	(0,15)	(0,26)	(0,56)	(0,56)	(0,29)

Notes: Logit Estimates. Cohort effects were included and are not reported here.

Table 3: First Step Flexible Estimates.

Education Level	15	16	17	18	19	20	>20
Intercept	-3,08	4,66	-2,55	-0,85	0,85	-5,15	0,40
	(0,55)	(0,41)	(0,43)	(0,92)	(1,05)	(1,13)	(1,07)
Number of Siblings	-0,07	-0,02	0,03	-0,16	0,05	-0,02	-0,01
	(0,01)	(0,01)	(0,01)	(0,03)	(0,03)	(0,02)	(0,02)
Parents' Higher Educ.:	0,32	0,37	0,90	-0,18	-0,56	-0,07	0,03
College	(0,08)	(0,06)	(0,09)	(0,16)	(0,18)	(0,15)	(0,30)
Parents' Higher Educ.:	0,27	0,61	0,49	-0,34	-0,45	-0,27	-0,07
High School	(0,06)	(0,05)	(0,08)	(0,15)	(0,17)	(0,14)	(0,30)
Non Working Mother	0,30	-0,11	-0,16	0,29	-0,53	1,26	-0,24
	(0,07)	(0,06)	(0,06)	(0,12)	(0,13)	(0,13)	(0,15)
Parents' Income	0,73	-0,58	0,20	0,06	0,23	0,98	-0,04
	(0,10)	(0,07)	(0,07)	(0,15)	(0,18)	(0,19)	(0,17)
Class repetition = 1	-0,57	-0,32	0,24	0,43	-0,78	-0,91	0,12
	(0,05)	(0,03)	(0,04)	(0,09)	(0,09)	(0,08)	(0,09)
Class Repetition >1	-0,76	-0,79	0,13	0,17	-1,32	-0,91	-0,18
	(0,06)	(0,04)	(0,04)	(0,09)	(0,09)	(0,07)	(0,08)

Notes: OLS estimates.
Two steps corrected standard errors with unknown heteroskedasticity.
Cohort effects were included and are not reported here.
Two dummy variables for levels 22 and 23 were also included at level>20.
Estimates are respectively -0.23 (0.07) and 0.57 (0.09).

Table 4: Net Returns

Beta	0,8	0,9	0,95	0,99
Intercept	1,27	0,99	0,85	0,73
	(0,80)	(0,88)	(0,92)	(0,95)
Number of Siblings	0,16	0,16	0,16	0,16
	(0,02)	(0,02)	(0,03)	(0,03)
Parents Highest Educ.:	0,22	0,19	0,18	0,17
	(0,14)	(0,15)	(0,16)	(0,16)
Parents Highest Educ.:	0,44	0,37	0,34	0,31
	(0,13)	(0,15)	(0,15)	(0,16)
Non Working Mother	-0,34	-0,31	-0,29	-0,28
	(0,10)	(0,11)	(0,12)	(0,12)
Parents' Income	-0,22	-0,11	-0,06	-0,01
	(0,13)	(0,14)	(0,15)	(0,16)
Class Repetition = 1	-0,24	-0,37	-0,43	-0,48
	(0,07)	(0,08)	(0,09)	(0,09)
Class Repetition > 1	0,06	-0,10	-0,17	-0,24
	(0,08)	(0,08)	(0,09)	(0,09)

Note: see Table 4

Table 5: Sensitivity to Beta at Education Level 18

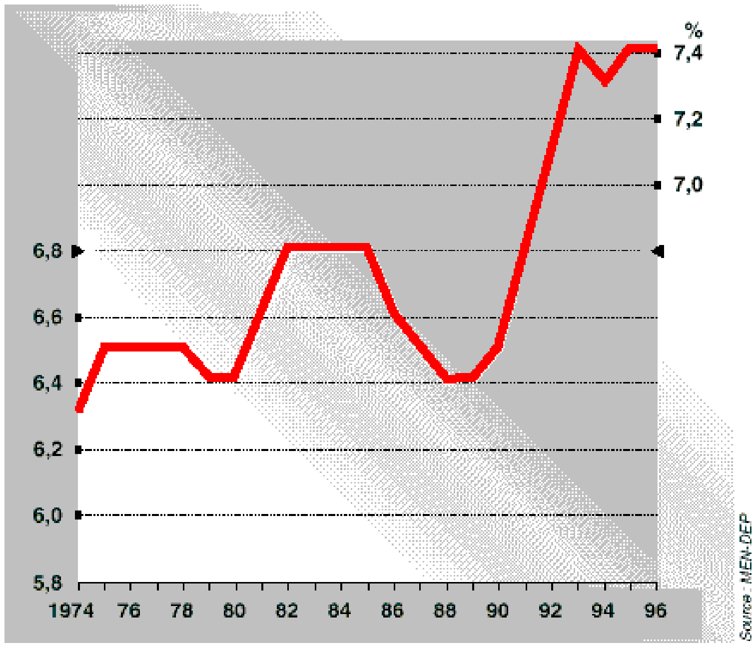

Figure 1: Education Expenditure in GDP

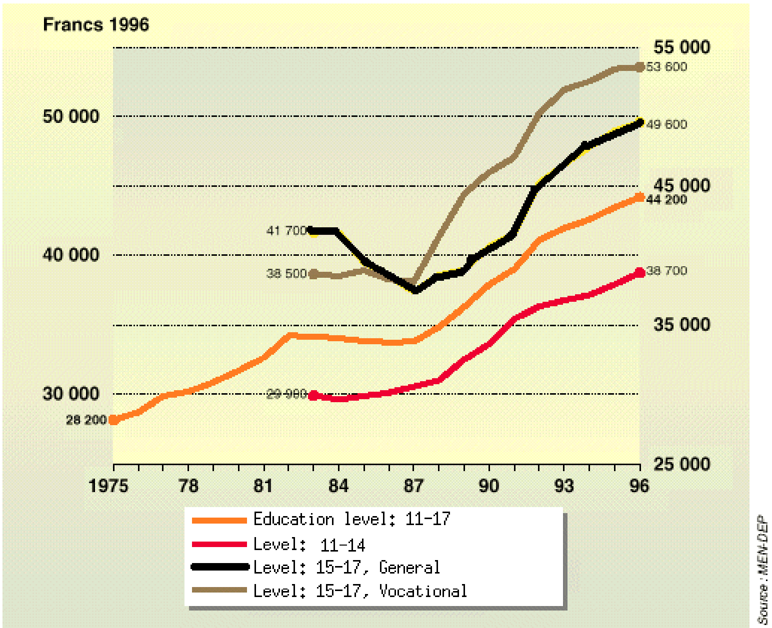

Figure 2:Average Yearly Spending per Student

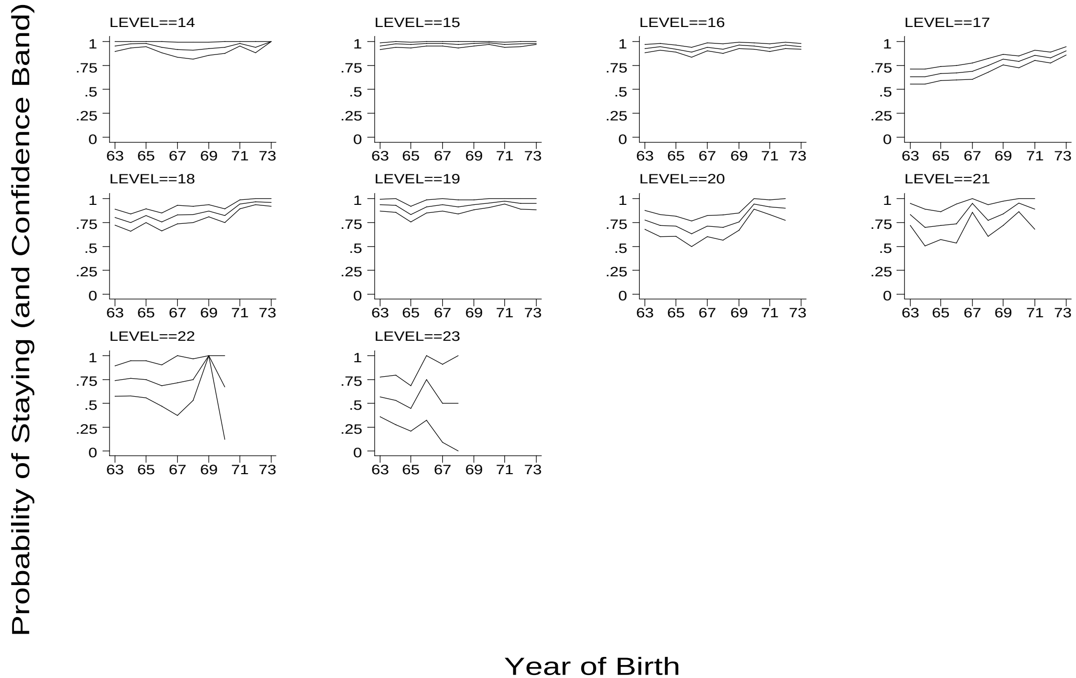

Fig 3: Average Probability of Staying

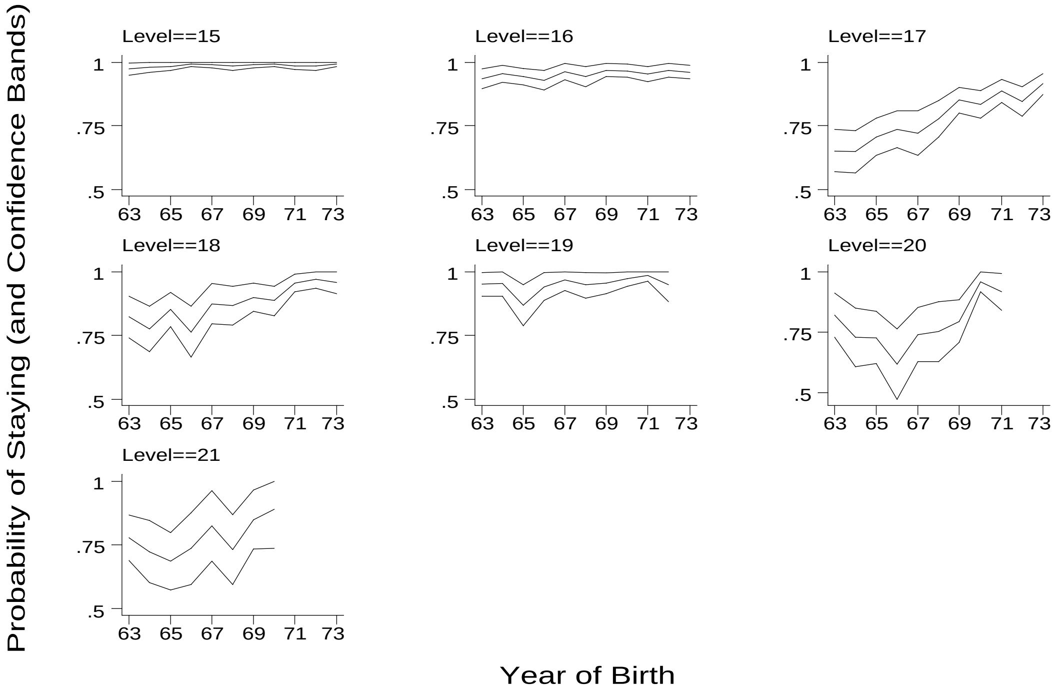

Fig 4: Predicted Probability of Staying

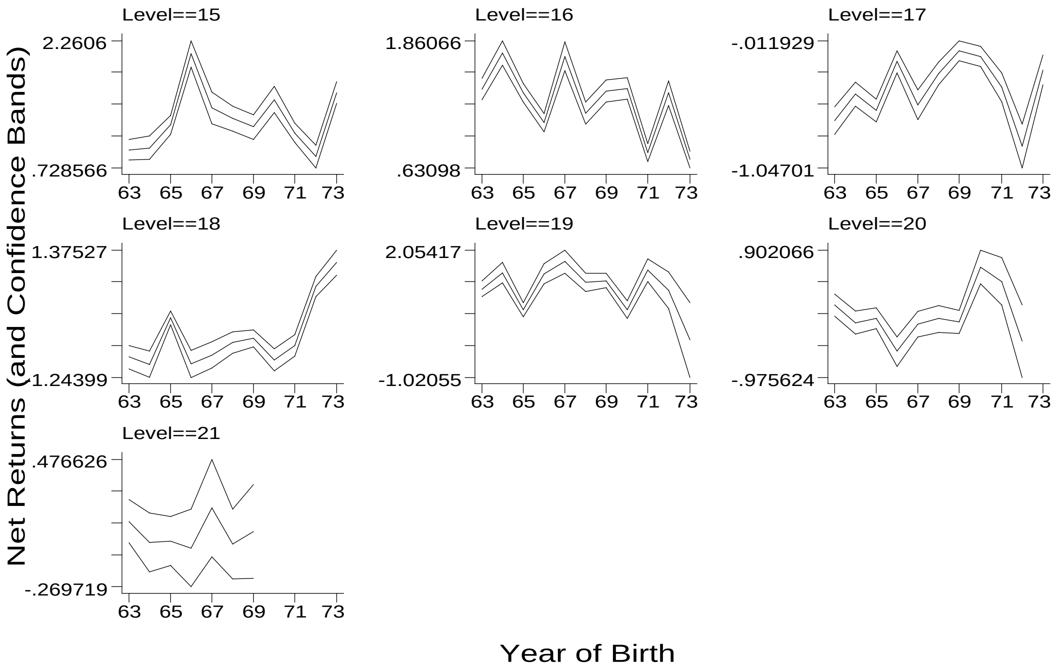

Fig 5: Net Returns by Schooling Level (Beta=0.95)

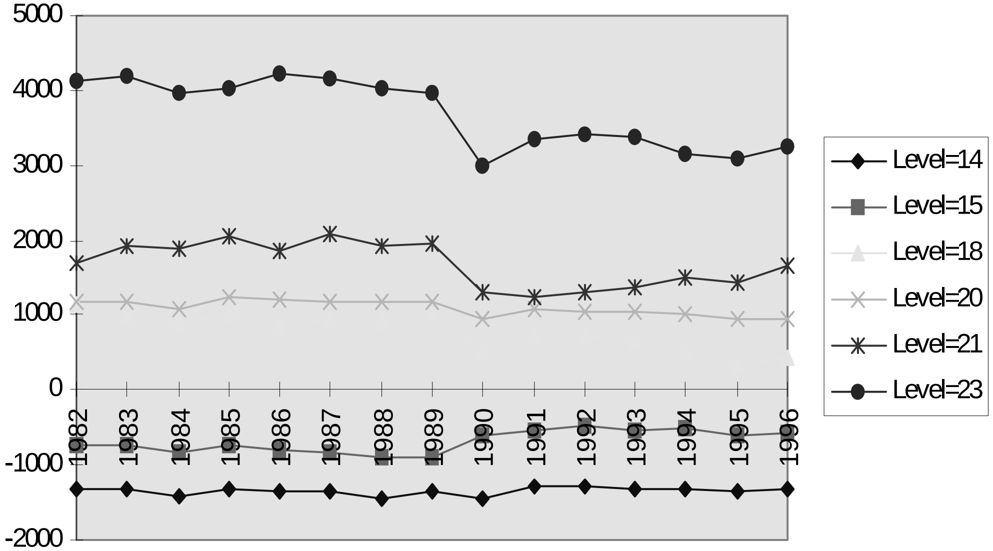

Fig 6: Returns to Education 82-97

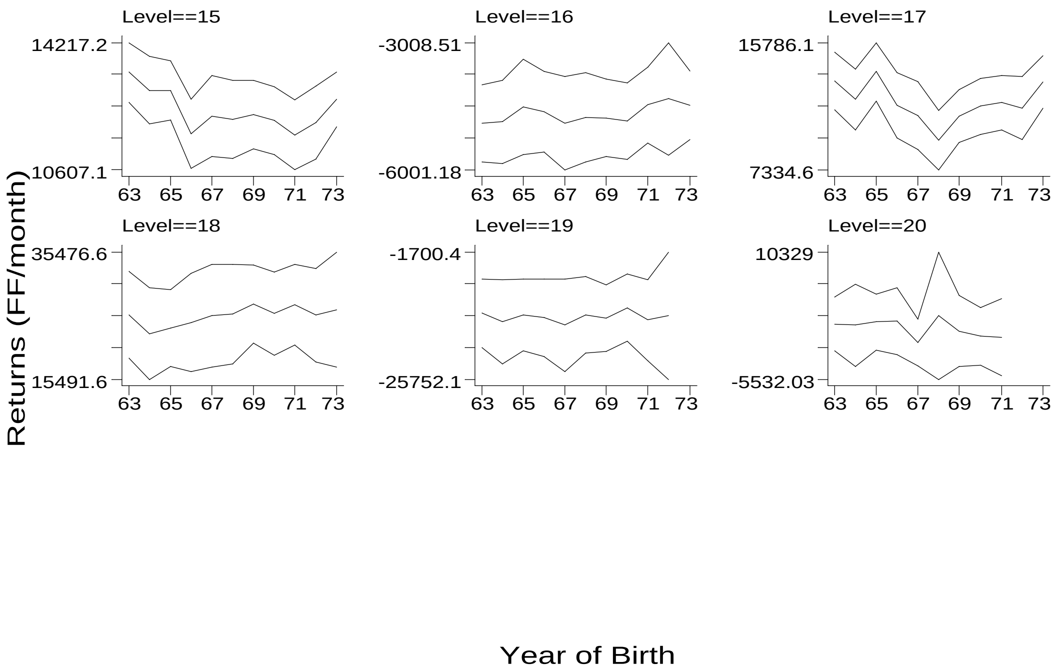

Fig 7:Subjective Returns of Schooling by Level

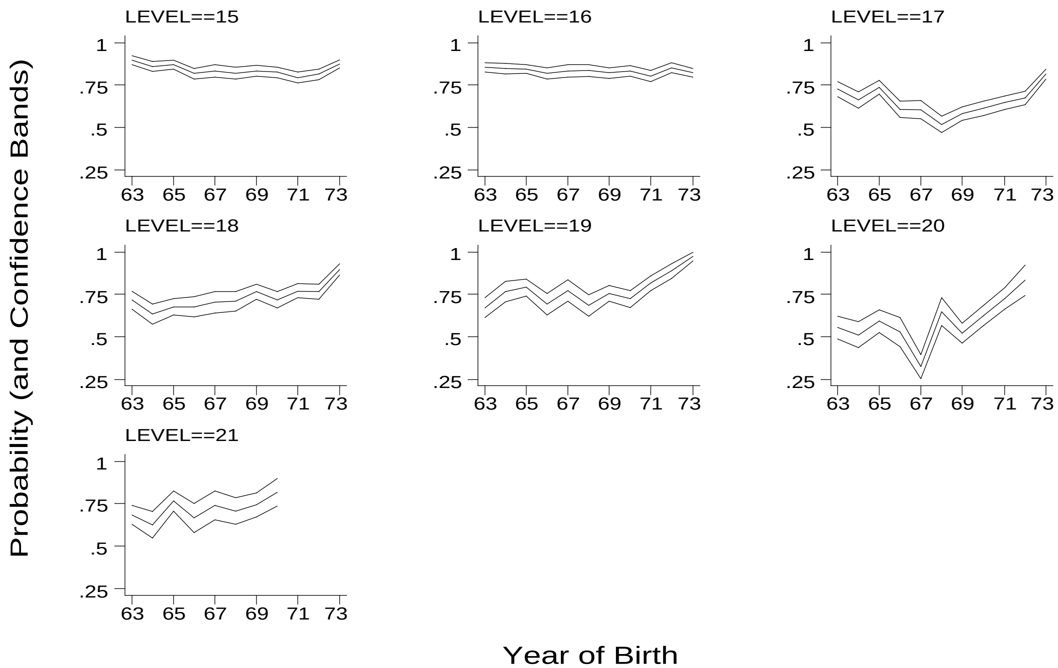

Fig 8: Probability of Success