# Choosing the Field of Study in French Post-Secondary Education: Do Expected Earnings Matter?

$X^{èmes}$ Journées de Méthodologie Statistique de l'INSEE, Paris

Arnaud Maurel (CREST-ENSAE, PSE, Paris, and IZA, Bonn)
Joint with M.Beffy and D.Fougère (CREST, Paris and IZA, Bonn)

25 mars 2009

# Outline

# Outline

# Main issues

- How do students choose the major of their post-secondary studies?
- How much this choice depends on individual preferences, ability and labor market productivity?

Focus on :

- the impact of expected earnings on this choice
- when the length of study is uncertain

# Main issues

- How do students choose the major of their post-secondary studies?
- How much this choice depends on individual preferences, ability and labor market productivity?

Focus on :

- the impact of expected earnings on this choice
- when the length of study is uncertain

# Existing literature

- Rational expectations framework
  - Seminal papers : Willis and Rosen (1979), Berger (1988)
  - Dynamic structural discrete choice models : Keane and Wolpin (1997,2001), Cameron and Heckman (1998,2001), Eckstein and Wolpin (1999), Belzil and Hansen (2002). Lee(2005) within a general equilibrium setting
    Sequential schooling decision model by Arcidiacono (2004,2005)
- Relaxing the rational expectations assumption
  - *Myopic* expectations framework : Freeman (1971,1975), Manski (1993). Recent paper by Boudarbat and Montmarquette (2007)
  - Buchinsky and Leslie (2000) : assuming different forecasting behaviors (*myopic*, rational and adaptive) has a significant impact on the predictions in terms of schooling choices

# Existing literature

- Rational expectations framework
  - Seminal papers : Willis and Rosen (1979), Berger (1988)
  - Dynamic structural discrete choice models : Keane and Wolpin (1997,2001), Cameron and Heckman (1998,2001), Eckstein and Wolpin (1999), Belzil and Hansen (2002). Lee(2005) within a general equilibrium setting
    Sequential schooling decision model by Arcidiacono (2004,2005)
- Relaxing the rational expectations assumption
  - *Myopic* expectations framework : Freeman (1971,1975), Manski (1993). Recent paper by Boudarbat and Montmarquette (2007)
  - Buchinsky and Leslie (2000) : assuming different forecasting behaviors (*myopic*, rational and adaptive) has a significant impact on the predictions in terms of schooling choices

# Our approach

We examine how the choice of the major is affected by expected earnings, when the final level of education is uncertain

- We estimate a semi-structural three-equation model : choice of the major, realization of the level of education and earnings

- Main contributions of our paper :
  - Simulation of the effects of an exogenous variation of the earnings distribution on the proportions of students who choose a given major
  - In a framework in which the length of study as well as future earnings are uncertain to the individual
  - Identification strategy exploiting variations in the relative returns to each major across the French business cycle (92-98)

# Our approach

We examine how the choice of the major is affected by expected earnings, when the final level of education is uncertain

- We estimate a semi-structural three-equation model : choice of the major, realization of the level of education and earnings

- Main contributions of our paper :
  - Simulation of the effects of an exogenous variation of the earnings distribution on the proportions of students who choose a given major
  - In a framework in which the length of study as well as future earnings are uncertain to the individual
  - Identification strategy exploiting variations in the relative returns to each major across the French business cycle (92-98)

# Outline

# A sequential model

- Stage 1 : After graduating from high school, individuals choose their post-secondary major
- Stage 2: Students keep on studying in the field chosen in stage 1, until they reach a given level of education
- Stage 3: All individuals leave the educational system and enter the labor market

The three outcomes are correlated through observables and unobservables

# A sequential model

- Stage 1 : After graduating from high school, individuals choose their post-secondary major
- Stage 2: Students keep on studying in the field chosen in stage 1, until they reach a given level of education
- Stage 3: All individuals leave the educational system and enter the labor market

The three outcomes are correlated through observables and unobservables

# A sequential model

- Stage 1 : After graduating from high school, individuals choose their post-secondary major
- Stage 2: Students keep on studying in the field chosen in stage 1, until they reach a given level of education
- Stage 3: All individuals leave the educational system and enter the labor market

The three outcomes are correlated through observables and unobservables

# A sequential model

- Stage 1 : After graduating from high school, individuals choose their post-secondary major
- Stage 2: Students keep on studying in the field chosen in stage 1, until they reach a given level of education
- Stage 3: All individuals leave the educational system and enter the labor market

The three outcomes are correlated through observables and unobservables

# Unobserved individual heterogeneity

- $R$ types of individuals, with $\Pi_r$ denoting the proportion of type $r$ in the population of students. $R = 3$ in the estimation (Heckman and Singer,1984)
- Three type-specific intercept terms in preferences (major choice), ability (length of study) and productivity (earnings equation)

# Unobserved individual heterogeneity

- $R$ types of individuals, with $\Pi_r$ denoting the proportion of type $r$ in the population of students. $R = 3$ in the estimation (Heckman and Singer,1984)
- Three type-specific intercept terms in preferences (major choice), ability (length of study) and productivity (earnings equation)

# Stage 1 : choice of the field of study

- Major chosen from a set of $J$ fields within a random utility framework
- For a student of type $r$, the value associated with field $j$ is :

$$V_j^r = v_{0j}^r + v_{1j}^r, \text{ for } j = 1, ...M$$

where

$$v_{0j}^r = \alpha_{(1,j)}^r + X_1' \beta_1^j + u_j$$

and

$$v_{1j}^r = \alpha \sum_{k \in \{0,1,...,L\}} \Pr(K = k \mid r, J = j).E\left(V_{e(j,k)}^r \mid r, J = j, K = k\right)$$

- The chosen major is :

$$j_r^* = \arg \max_{j \in \{1,...,M\}} V_j^r$$

# Stage 1 : choice of the field of study

- Major chosen from a set of $J$ fields within a random utility framework
- For a student of type $r$, the value associated with field $j$ is :

$$V_j^r = v_{0j}^r + v_{1j}^r, \text{ for } j = 1, ...M$$

where

$$v_{0j}^r = \alpha_{(1,j)}^r + X_1' \beta_1^j + u_j$$

and

$$v_{1j}^r = \alpha \sum_{k \in \{0,1,...,L\}} \Pr(K = k \mid r, J = j).E\left(V_{e(j,k)}^r \mid r, J = j, K = k\right)$$

- The chosen major is :

$$j_r^* = \arg \max_{j \in \{1,...,M\}} V_j^r$$

# Stage 2 : determination of the educational level

- $L+1$ levels in each major
- The realized level $k_j^*$ is such as :

$$\forall k \in \{0, 1, ..., L\}$$
$$k_j^* = k \quad \Leftrightarrow \quad s_k < \widetilde{k}_j^r \leq s_{k+1}$$

- where $\widetilde{k}_j^r$ is the individual propensity to reach a given level of education, $\{s_1, \ldots, s_L\}$ are thresholds to be estimated

with :

$$\begin{aligned} \widetilde{k}_j^r &= \alpha_2^r + X_{2,j}'\beta_2 \\ &+ v \end{aligned}$$

# Stage 2 : determination of the educational level

- $L+1$ levels in each major
- The realized level $k_j^*$ is such as :

$$\forall k \in \{0, 1, ..., L\}$$
$$k_j^* = k \quad \Leftrightarrow \quad s_k < \widetilde{k}_j^r \leq s_{k+1}$$

- where $\widetilde{k}_j^r$ is the individual propensity to reach a given level of education, $\{s_1, \ldots, s_L\}$ are thresholds to be estimated

with :

$$\begin{aligned} \widetilde{k}_j^r &= \alpha_2^r + X_{2,j}'\beta_2 \\ &+ v \end{aligned}$$

# Stage 3 : earnings

- Individuals with post secondary education $(j^*, k^*)$ enter the labor market
- Earnings equation :

$$U_{j,k}^{w,r} = \alpha_w^r + X_{w,j,k}\beta_w + \epsilon_w$$

- Where the earnings is set equal to:

$$U_{j,k}^{w,r} = \ln \frac{\sum_{s=1}^{N_e} w_s l_s^e + \sum_{s=1}^{N_u} b_s l_s^u}{\sum_{s=1}^{N_e} l_s^e + \sum_{s=1}^{N_u} l_s^u}$$

# Stage 3 : earnings

- Individuals with post secondary education $(j^*, k^*)$ enter the labor market
- Earnings equation :

$$U_{j,k}^{w,r} = \alpha_w^r + X_{w,j,k}\beta_w + \epsilon_w$$

- Where the earnings is set equal to:

$$U_{j,k}^{w,r} = \ln \frac{\sum_{s=1}^{N_e} w_s l_s^e + \sum_{s=1}^{N_u} b_s l_s^u}{\sum_{s=1}^{N_e} l_s^e + \sum_{s=1}^{N_u} l_s^u}$$

# Estimation

For a given type $r$, the model consists of three equations with normally distributed residuals which are assumed to be independent across equations:

- First equation: a MNP model for the choice of major
- Second equation: an Ordered Probit model for the length of study
- Third equation: a log-linear equation for earnings

$\Rightarrow$ Estimation using the sequential version of the EM algorithm introduced by Arcidiacono and Jones (2003)

# Estimation

For a given type $r$, the model consists of three equations with normally distributed residuals which are assumed to be independent across equations:

- First equation: a MNP model for the choice of major
- Second equation: an Ordered Probit model for the length of study
- Third equation: a log-linear equation for earnings

$\Rightarrow$ Estimation using the sequential version of the EM algorithm introduced by Arcidiacono and Jones (2003)

# Outline

# Which data ? - 1

- *Generations 1992 and 1998 (CEREQ)*
- Pooled dataset which contains informations on 40,724 individuals who left the French educational system either in 1992 or in 1998, and who were surveyed five years later
- **Why this data base ?** Information on both educational paths and wages, as well as individual covariates
- **Which subsample ?** Persons graduated from high school (27,389) and who attended university (except medicine and IUT) $\Rightarrow$ **7,346** individuals (3,436 from G92 and 3,910 from G98)

# Which data ? - 1

- *Generations 1992 and 1998 (CEREQ)*
- Pooled dataset which contains informations on 40,724 individuals who left the French educational system either in 1992 or in 1998, and who were surveyed five years later
- **Why this data base ?** Information on both educational paths and wages, as well as individual covariates
- **Which subsample ?** Persons graduated from high school (27,389) and who attended university (except medicine and IUT) $\Rightarrow$ **7,346** individuals (3,436 from G92 and 3,910 from G98)

# Which data ? - 1

- *Generations 1992 and 1998 (CEREQ)*
- Pooled dataset which contains informations on 40,724 individuals who left the French educational system either in 1992 or in 1998, and who were surveyed five years later
- **Why this data base ?** Information on both educational paths and wages, as well as individual covariates
- **Which subsample ?** Persons graduated from high school (27,389) and who attended university (except medicine and IUT) $\Rightarrow$ **7,346** individuals (3,436 from G92 and 3,910 from G98)

# Which data ? - 1

- *Generations 1992 and 1998 (CEREQ)*
- Pooled dataset which contains informations on 40,724 individuals who left the French educational system either in 1992 or in 1998, and who were surveyed five years later
- **Why this data base ?** Information on both educational paths and wages, as well as individual covariates
- **Which subsample ?** Persons graduated from high school (27,389) and who attended university (except medicine and IUT) $\Rightarrow$ **7,346** individuals (3,436 from G92 and 3,910 from G98)

# Which data ? - 2

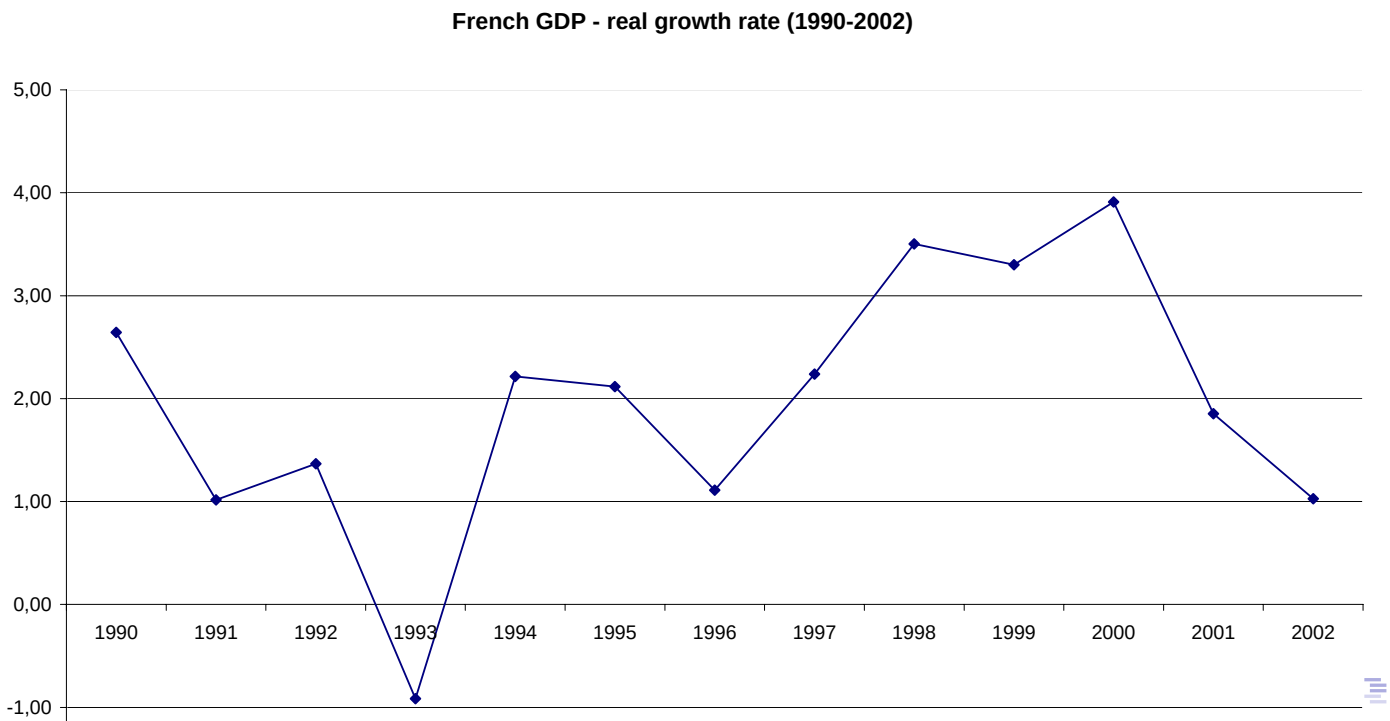
- Three aggregated fields:
  - Sciences
  - Humanities and Social Sciences
  - Law, Economics and Management
- Five levels: from one year after high school to 5 years and more
- The item "one year after HS" = dropout with no university degree

# Which data ? - 2

- Three aggregated fields:
  - Sciences
  - Humanities and Social Sciences
  - Law, Economics and Management
- Five levels: from one year after high school to 5 years and more
- The item "one year after HS" = dropout with no university degree

# Identification - 1

We exploit variations in the relative monetary returns to each major induced by the year of entry into the French labor market: 1992 a period of very weak economic growth and 1998 a period of strong economic growth.



# Identification - 2

Table: Average monthly earnings (constant 1992 Francs) according to the field of study

| | | |
|---|---|---|
| *Generation 1992* | | |
| Sciences | 6,833 | |
| Humanities and Social Sciences | 6,088 | |
| Law, Economics and Management | 6,318 | |
| *Generation 1998* | | |
| Sciences | 7,758 | **+14%** |
| Humanities and Social Sciences | 5,835 | **-4%** |
| Law, Economics and Management | 6,976 | **+10%** |

# Identification - 3

- Whether the individual enters the labor market in 1992 or in 1998 has no direct effect on her choice of major. Every observed characteristics being equal, we assume that **schooling preferences are stable between 1992 and 1998.**
- In the earnings equation *only*, introduction of a dummy variable, which is equal to zero if the individual enters the labor market in 1992 and one if she enters the labor market six years later in 1998, interacted with the field of study. Identification of $\alpha$ exploits these exclusion restrictions.

# Outline

# Estimation and results - 1

Which covariates ?

1. Choice of the field ($1^{st}$ stage): age in $6^{th}$ grade, born abroad, gender, high school major, parental nationality, father's and mother's profession (resp. in 1992 and 1998)

2. Length of study ($2^{nd}$ stage): father's and mother's profession (resp. in 1992 and 1998), born abroad, gender, age in $6^{th}$ grade, high school major, parental nationality, average proportion of college students in the same major and in the same university, year of entry into the labor market (1992 or 1998)

3. Earnings equation: gender, born abroad, parental nationality, Paris dummy, field of study, level of education, field times level, gender times level, **year of entry into the labor market, interacted with field of study** and level of education

2. Length of study ($2^{nd}$ stage): father's and mother's profession (resp. in 1992 and 1998), born abroad, gender, age in $6^{th}$ grade, high school major, parental nationality, average proportion of college students in the same major and in the same university, year of entry into the labor market (1992 or 1998)

3. Earnings equation: gender, born abroad, parental nationality, Paris dummy, field of study, level of education, field times level, gender times level, **year of entry into the labor market, interacted with field of study** and level of education

# Estimation and results - 2

A set of simulations to get an accurate idea of the effect of expected earnings on the choice of the major

- An increase (resp. decrease) of 10% in the expected earnings for a given major (sciences, humanities or law) whatever the length
- The model fit is good
- Main finding: **Expected earnings have a statistically significant but small effect on the choice of the major**

# Estimation and results - 2

A set of simulations to get an accurate idea of the effect of expected earnings on the choice of the major

- An increase (resp. decrease) of 10% in the expected earnings for a given major (sciences, humanities or law) whatever the length
- The model fit is good
- Main finding: **Expected earnings have a statistically significant but small effect on the choice of the major**

# Estimation and results - 2

A set of simulations to get an accurate idea of the effect of expected earnings on the choice of the major

- An increase (resp. decrease) of 10% in the expected earnings for a given major (sciences, humanities or law) whatever the length
- The model fit is good
- Main finding: **Expected earnings have a statistically significant but small effect on the choice of the major**

# Simulations: a 10% variation in expected earnings for Sciences

| | Observed proportions | Predicted proportions | $(p^S - p^P)$ | $\widehat{\sigma}_{(p^S - p^P)}$ |
|---|---|---|---|---|
| | | 10% *increase in Sciences* | | |
| Sciences | 28.67 | 27.97 | 0.251 | 0.019 |
| HSS | 37.59 | 41.17 | -0.189 | 0.013 |
| LEM | 33.75 | 30.86 | -0.062 | 0.009 |
| | | 10% *decrease in Sciences* | | |
| Sciences | 28.67 | 27.97 | -0.276 | 0.021 |
| HSS | 37.59 | 41.17 | 0.209 | 0.014 |
| LEM | 33.75 | 30.86 | 0.068 | 0.009 |

# Simulations: a 10% variation in expected earnings for Humanities and Social Sciences

| | Observed proportions | Predicted proportions | $(p^S - p^P)$ | $\widehat{\sigma}_{(p^S - p^P)}$ |
|---|---|---|---|---|
| | | 10% *increase in HSS* | | |
| Sciences | 28.67 | 27.97 | -0.189 | 0.013 |
| HSS | 37.59 | 41.17 | 0.526 | 0.048 |
| LEM | 33.75 | 30.86 | -0.336 | 0.038 |
| | | 10% *decrease in HSS* | | |
| Sciences | 28.67 | 27.97 | 0.209 | 0.014 |
| HSS | 37.59 | 41.17 | -0.580 | 0.053 |
| LEM | 33.75 | 30.86 | 0.371 | 0.042 |

# Simulations: a 10% variation in expected earnings for Law, Economics and Management

| | Observed proportions | Predicted proportions | $(p^S - p^P)$ | $\widehat{\sigma}_{(p^S - p^P)}$ |
|---|---|---|---|---|
| | | 10% *increase in LEM* | | |
| Sciences | 28.67 | 27.97 | -0.062 | 0.009 |
| HSS | 37.59 | 41.17 | -0.337 | 0.038 |
| LEM | 33.75 | 30.86 | 0.399 | 0.042 |
| | | 10% *decrease in LEM* | | |
| Sciences | 28.67 | 27.97 | 0.068 | 0.009 |
| HSS | 37.59 | 41.17 | 0.371 | 0.042 |
| LEM | 33.75 | 30.86 | -0.439 | 0.046 |

# Outline

# Conclusion

Low elasticity of post-secondary major choices to expected earnings
$\Rightarrow$ This choice is mainly driven by *consumption value* of schooling, including *preferences* and *abilities*
$\Rightarrow$ Evidence, in line with Carneiro, Hansen and Heckman (2003), that non pecuniary factors are a key determinant of schooling choices, here in terms of post-secondary major

# Conclusion

Low elasticity of post-secondary major choices to expected earnings
$\Rightarrow$ This choice is mainly driven by *consumption value* of schooling, including *preferences* and *abilities*
$\Rightarrow$ Evidence, in line with Carneiro, Hansen and Heckman (2003), that non pecuniary factors are a key determinant of schooling choices, here in terms of post-secondary major

# Descriptive statistics - 1

| | Number | Percent |
|---|---|---|
| *University fields* | | |
| Sciences | 2,106 | 28.67 |
| Humanities and Social Sciences | 2,761 | 37.59 |
| Law, Economics and Management | 2,479 | 33.75 |
| *Post-secondary education level* | | |
| Dropout | 1,762 | 23.99 |
| Two years of college | 732 | 9.97 |
| Licence (BA degree) | 1,400 | 19.06 |
| Maîtrise (MA degree) | 1,486 | 20.23 |
| Post Maîtrise (Graduates) | 1,966 | 26.76 |

# Descriptive statistics - 2

Table: Average monthly earnings (constant 1992 Francs) according to the length and the field of study

| Field | Length | Average monthly earnings |
|---|---|---|
| | Dropout | 4,920 |
| | Two years of college | 5,983 |
| | Licence (BA degree) | 6,181 |
| | Maitrise (MA degree) | 6,739 |
| | Post Maitrise (Graduates) | 8,414 |
| Sciences | | 7,277 |
| HSS | | 5,942 |
| LEM | | 6,666 |