Redistributional impact of the National Health Insurance System in France:

A microsimulation approach

European Workshop on Dynamic Microsimulation Modelling

Valrie Albouy (INSEE)

Laurent Davezies (INSEE-CREST-IRDES)

Thierry Debrand (IRDES)

Brussels, 4-5 March 2010

Plan

Context and economic problems

In France, public health insurance can be seen as an insurance over the lifecycle. On the one hand workers have to pay contributions and on the other hand health expenditures are partially or totally reimbursed.

Transfer between "good health period" and "bad health period" for each individual ⇒ Intertemporal aspect

Transfer between "lucky" and "unlucky" ⇒ Insurance

Contributions not fixed by medical risk evaluation, but by legal rules ⇒ Redistribution

What are the problems ?

Difficulties :

- Multiplicity of determinants of health expenditures
- Rules for contributions (ex : several contributions, mainly on wages, but also on pensions, unemployment benefits...)
- Rules of reimbursement (ex : threshold of household income for no co-payment, long-term diseases...)
- Differences of mortality rates between social groups

$\Rightarrow$ Not easy to understand if the legal rules are favorable to a particular group.

Notation

- Age of death of individual i : A_i
- Contribution of individual i at age a : C_{ia}
- Health care utilization for individual i at age a : D_{ia}
- Health expenditures of individual i at age a : M_{ia}
- Reimbursement of individual i at age a : R_{ia}, $0 \leq R_{ia} \leq M_{ia}$
- Net transfer benefit of individual i at age a : $B_{ia} = R_{ia} - C_{ia}$
- Net present value of transfer for individual i with a discount account r : $NPV_i = \sum_{a=0}^{A_i} \frac{R_{ia} - C_{ia}}{(1+r)^a}$

How to describe transfers over the lifecycle ?

A (too) simple framework : 2 groups G_1 and G_2, and imagine a comparison between the present system and a fictive one, contribution based on risk in *absence of information asymmetry*.

- Evaluate transfers between groups : $\mathbb{E}(NPV_i|G_1)$ and $\mathbb{E}(NPV_i|G_2)$, difference between contribution and a "pure" risk-based insurance premium for the class of risk G_1 or G_2
  $\Rightarrow$ Measure of the redistribution
- Evaluate dispersion of transfers within groups : $\mathbb{V}(NPV_i|G)$, dispersion of the risk in the group G, measure of risk covered within the class of risk G_1 or G_2
  $\Rightarrow$ Measure of the insurance

Questions : who are the "winners" and the "losers" in the system ? Are the transfers between groups smaller than dispersion of transfers within groups ?

"Cross section" and "time-series" dimension

- Terms $\mathbb{E}\left(NPV_i|G_1\right)$ can be estimated only with cross-sectional observations :

$$\mathbb{E}\left(NPV_i|G\right) = \sum_{a=0}^{+\infty} \frac{\mathbb{E}\left(B_{ia}\mathbb{1}_{\{a\leq A_i\}}|G\right)}{(1+r)^a}$$

$$= \sum_{a=0}^{+\infty} \frac{\mathbb{E}\left(R_{ia}\mathbb{1}_{\{a\leq A_i\}}|G\right) - \mathbb{E}\left(C_{ia}\mathbb{1}_{\{a\leq A_i\}}|G\right)}{(1+r)^a}$$

- Terms $\mathbb{V}\left(NPV_i|G_1\right)$ need longitudinal observations to be estimated :

$$\mathbb{V}\left(NPV_i|G\right) = \sum_{a=0}^{+\infty}\sum_{a'=0}^{+\infty} \frac{\mathbb{C}ov\left(B_{ia}\mathbb{1}_{\{a\leq A_i\}}, B_{ia'}\mathbb{1}_{\{a'\leq A_i\}}|G\right)}{(1+r)^{a+a'}} =$$

$$\sum_{a=0}^{+\infty}\sum_{a'=0}^{+\infty} \frac{\left\{\begin{array}{c} \mathbb{C}ov\left(R_{ia}\mathbb{1}_{\{a\leq A_i\}}, R_{ia'}\mathbb{1}_{\{a'\leq A_i\}}|G\right) + \mathbb{C}ov\left(C_{ia}\mathbb{1}_{\{a\leq A_i\}}, C_{ia'}\mathbb{1}_{\{a'\leq A_i\}}|G\right) \\ -\mathbb{C}ov\left(R_{ia}\mathbb{1}_{\{a\leq A_i\}}, C_{ia'}\mathbb{1}_{\{a'\leq A_i\}}|G\right) - \mathbb{C}ov\left(C_{ia}\mathbb{1}_{\{a\leq A_i\}}, R_{ia'}\mathbb{1}_{\{a'\leq A_i\}}|G\right) \end{array}\right\}}{(1+r)^{a+a'}}$$

Why microsimulation ?

To estimate the statistical distribution of NPV, the simplest case is to observe a long panel data with reimbursements and contributions.

- For past generations, these datasets are not available in France
- For present and future generations, future reimbursements and contributions can not be observed

Why microsimulation ?

We have :

- On the one hand, ESPS-EPAS is a "short" panel (6 years, $\approx$ 7100 individuals) with M_{it} observed and socio-demographic information X_{it} $\Rightarrow$ statistic law of expenditures $(M_{it})_{t=1...6}|(X_{it})_{t=1...6}$ can be identified
- On the other hand, microsimulation model DESTINIE of INSEE, gives us good simulation of socio-demographic and contributions trajectories X_{it} over the complete lifecycle. So, we can simulate trajectories of health expenditures over the complete lifecycle and analyze the system. And next, reimbursements and NPV.

Plan

Estimation of expenditures with ESPS-EPAS panel data

- 2 types of consumption :
  - Ambulatory: 90 % of utilisation $D_{it} = 1 \Rightarrow M_{it} > 0$
  - In-patient: 15% of utilisation $D_{it} = 1 \Rightarrow M_{it} > 0$
- Distribution characteristics :
  - Existence of non-consumers ($D_{it} = 0$ vs $D_{it} = 1$)
  - Strong unobserved individual heterogeneity (many things are not explained by covariates X)
  - Thick tail of distribution (few people have huge expenditures and many have small)
  - Strong temporal correlation (due to correlation of observed and unobserved heterogeneity in the time-series dimension)
  - Heteroskedasticity

Estimation of expenditures

$$D_{it} = \mathbb{1}_{\{D_{it}^{\star}>0\}} = \mathbb{1}_{\{D_{it-1}\gamma_D + ln(M_{it-1})\gamma_{ln(M)} + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}}$$

$$ln(M_{it}) = \left(D_{it-1}\beta_D + ln(M_{it-1})\beta_{ln(M)} + X_{it}\beta + v_i + \eta_{it}\right) D_{it}$$

- Thick tails of distribution ⇒ We work in "log": $ln(M_{it})$ ⇒ heteroscedasticity of η_{it} change the average amount of simulation because the exponential retransformation is not linear ⇒ modeling of heterosckedasticity.
- Three parts could explain the correlation in the time-series dimension : the observed heterogeneity $X_{it}\gamma$ and $X_{it}\beta$, the unobserved heterogeneity u_i and v_i and a state dependence $D_{it-1}\gamma_D + ln(M_{it-1})\gamma_{ln(M)}$ and $D_{it-1}\beta_D + ln(M_{it-1})\beta_{ln(M)}$

Estimation of expenditures

$$D_{it} = \mathbb{1}_{\{D^{\star}_{it}>0\}} = \mathbb{1}_{\{D_{it-1}\gamma_D + ln(M_{it-1})\gamma_{ln(M)} + X_{it}\gamma + u_i + \varepsilon_{it} > 0\}}$$

$$ln(M_{it}) = \left(D_{it-1}\beta_D + ln(M_{it-1})\beta_{ln(M)} + X_{it}\beta + v_i + \eta_{it}\right) D_{it}$$

- Impossible to assume that heterogeneity u_i and v_i is independent of lag variables D_{it-1}, M_{it-1}, need to treat this problem for estimation and simulation (WP-20 IRDES).
- Given X, number of use and average expenditure per use are correlated $\Rightarrow u_i \not\perp v_i$ or $\varepsilon_{it} \not\perp \eta_{it}$, $\Rightarrow$ Take this correlation into account in estimation (avoid bias selection) and in simulation
- 4 estimations : ambulatory and inpatient by gender

Descriptive statistics

Table: Health expenditures

	ambulatory				inpatient			
	$D=1$	$M\|D=1$(in €)			$D=1$	$M\|D=1$ (in €)		
Age		Moy	Q1	Q3		Moy	Q1	Q3
0-29	0,89	536	134	656	0,10	4 729	596	2 758
30-44	0,93	871	209	1 035	0,13	3 795	669	3 478
45-59	0,94	1 455	371	1 697	0,16	5 353	645	4 450
60-74	0,97	2 148	775	2 593	0,23	5 491	674	6 166
75 +	1,00	2 952	1 242	3 490	0,29	10 244	1 169	10 245
Time to death								
Alive	0,93	1 151	231	1 371	0,14	5 058	662	4 243
death in 2 years	0,94	3 620	780	4 498	0,40	12 260	1 256	13 073
death in 1 years	0,96	4 391	952	5 299	0,46	15 833	2 684	17 840

Estimation of expenditures

Covariates X :

- Age (flexible form : linear spline in this version)
- Age of schooling
- Status on the labor market
- Time to death : dummy variables that catch increase of expenditure at the end of the life
- For the young people :
  - Status on the labor market of their father
  - Current level of education

Plan

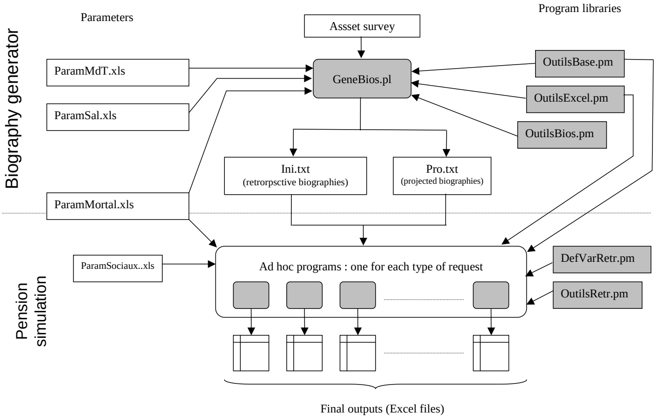
Destinie 2

- Complete rewriting, begin in 2005.
- Keep the main initial choices of Destinie 1 :
  - based on the asset survey of INSEE (2003), survey in cross section with retrospective information
  - closed population
- But new programming structure in two steps, more modular

Destinie 2

- First step (generator of biography) : simulation of biographic and professional trajectories excluding transitions to retirement : the theoretical carriers are fictively prolongated since 70 years old.
- Second step (retirement and pension simulator) : read the result of the first module and add the simulation of choice of retirement and compute pensions.
  - This second step is not a unique and complete program
  - but a library of modules that we use to construct small specific programs of simulation.

Structure de DESTINIE 2



The differential mortality

A important issue in the analyze of transfers induce by the national health system concern the date of death.

- Use of projections of mortality of Insee. However these projections are not differentiate by social groups.
- Use of life tables by social class in 1999, after 30 years old.
- Problem: age of schooling S is the variable of social stratification in Destinie, and not the social class SC.

$$\begin{aligned} P(Death = a|S, T = 1999) &= \sum_{SC} P(Death = a|S, SC, T = 1999) P(SC|S, T = 1999) \\ &= \sum_{SC} P(Death = a|SC, T = 1999) P(SC|S, T = 1999) \quad \text{if} \quad S \perp\!\!\!\perp Death|SC \end{aligned}$$

- Life tables by schooling age for the year T are calculated by changing the mortality rate of year $T-1$ to align mortality on the projection of Insee without changing the relative risks between schooling ages.

Simulation of expenditures

At this step of our work, we have simulate ambulatory and in-patient expenditures under the assumption that "shocks" that affect ambulatory expenditures are independent of "shocks" that affect in-patient expenditures.

This is probably wrong $\Rightarrow$ this bias downward the dispersion of NPV within the groups but not the mean transfer between group.

For the final version, we will estimate and simulate correlated "shocks" between ambulatory and in-patient.

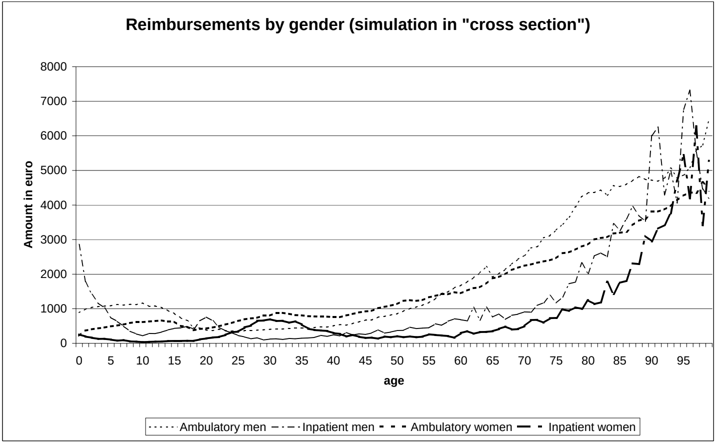
Simulation of reimbursements

Under a threshold ($\approx$ 7 000 €per unity of consumption) of income there is no co-payment. Advantages of Destinie

- Wages, pensions, of individual i
- Family links, household structures to compute unity of consumption
- Wages and pensions of members of household of individual i

Easy to compute the threshold !

Over the threshold : reimbursement depends on many unobserved factors, mainly the type of care. So, we use a statistical flexible regression (on the amount of expenditures and age) to impute a rate of reimbursement.


Reimbursements by gender (simulation in "cross section")
Amount in euro
8000
7000
6000
5000
4000
3000
2000
1000
0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
age
Ambulatory men
Inpatient men
Ambulatory women
Inpatient women

Assumption on economic trends

We simulate socio-demographic and salary trajectories for a cohort born between 1994 and 2003.

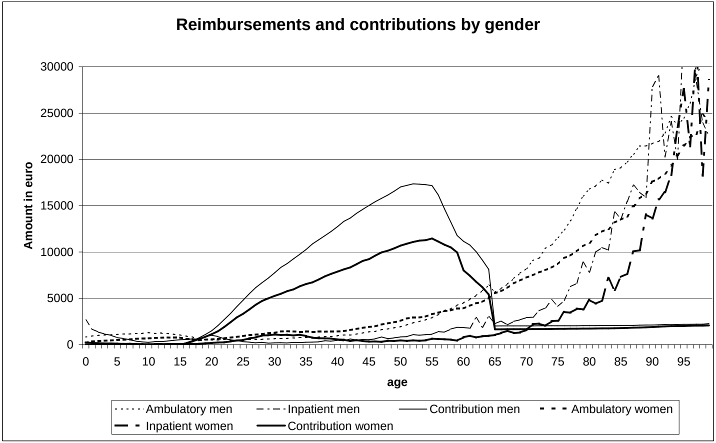
- technical progress : 1.8% that drives evolution of price of medical care (cf. Bac & Cornilleau, 2003)
- productivity growth : 1.8% that drives evolution of real wages and administrative threshold (usual assumption in Destinie model)
- mortality : forecast of National Statistic Institute
- evolution of active population : forecast of National Statistic Institute

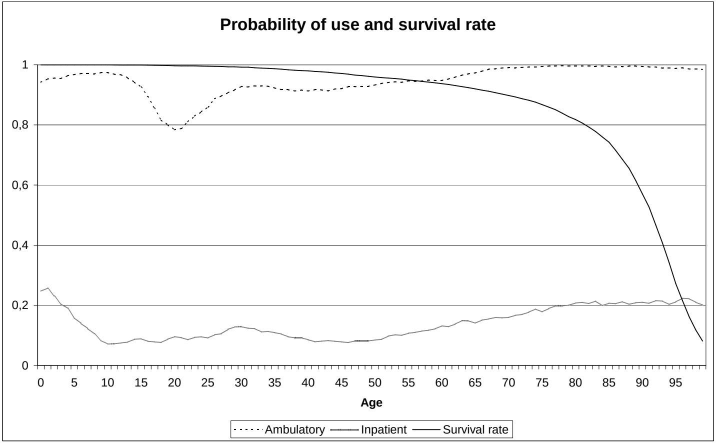
Simulation of contributions

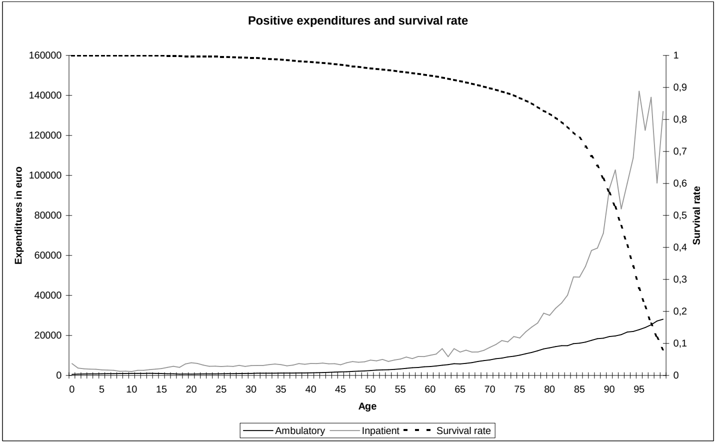
2 types of contributions are taken into account :

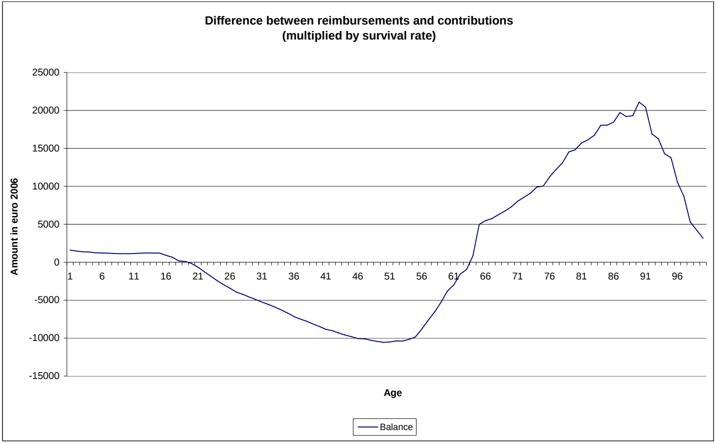
- First type is based on wages and only affects workers : 9,7% of wages for civil servants and 13,9% for the private sector
- Second type is based on incomes, like CSG and CRDS affect workers and retired $\sim$ 6 % on wages and $\sim$ 4 % on pensions (in short !).

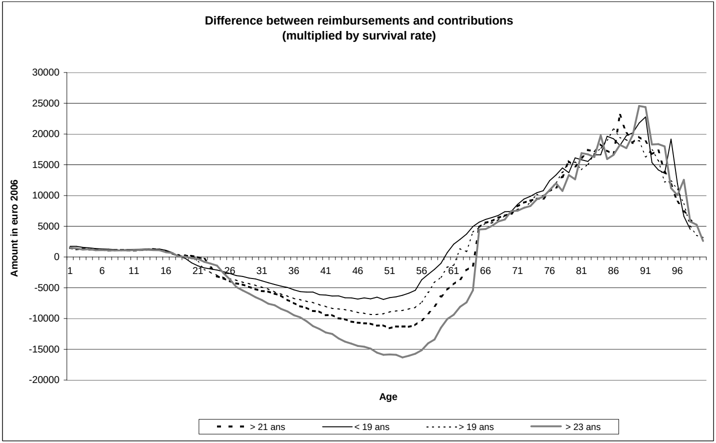
So, contributions are not difficult to determine when we know the status on the labor market, the wages, unemployment benefits and pensions. All these variables are presents in DESTINIE.

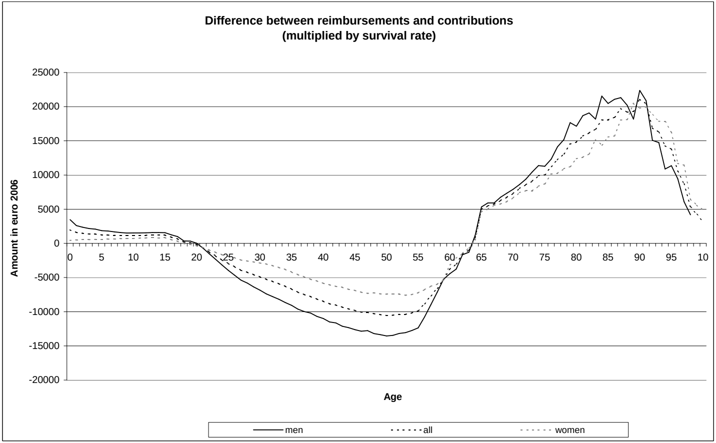

Reimbursements and contributions by gender
30000
25000
20000
15000
10000
5000
0
Amount in euro
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
age
Ambulatory men
Inpatient men
Contribution men
Ambulatory women
Inpatient women
Contribution women

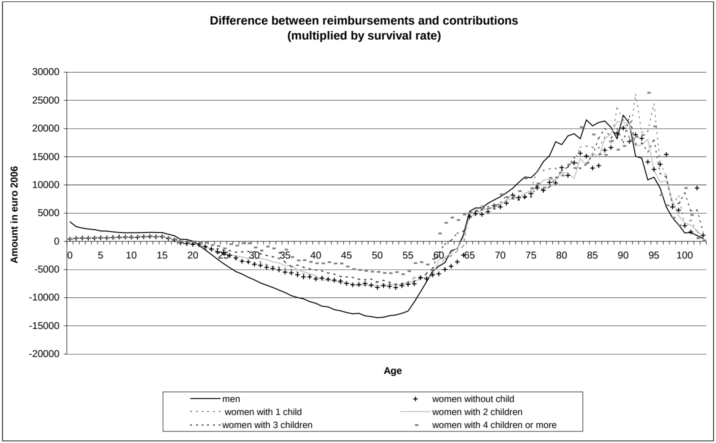

Probability of use and survival rate
1
0,8
0,6
0,4
0,2
0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
Age
Ambulatory
Inpatient
Survival rate


Positive expenditures and survival rate
Expenditures in euro
Survival rate
Age
160000
140000
120000
100000
80000
60000
40000
20000
0
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Ambulatory
Inpatient
Survival rate


Difference between reimbursements and contributions
(multiplied by survival rate)
Amount in euro 2006
25000
20000
15000
10000
5000
0
-5000
-10000
-15000
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
Age
Balance


Difference between reimbursements and contributions
(multiplied by survival rate)
Amount in euro 2006
30000
25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
Age
> 21 ans
< 19 ans
> 19 ans
> 23 ans


Difference between reimbursements and contributions
(multiplied by survival rate)
25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
Amount in euro 2006
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
10
Age
men
all
women


Difference between reimbursements and contributions
(multiplied by survival rate)
Amount in euro 2006
30000
25000
20000
15000
10000
5000
0
-5000
-10000
-15000
-20000
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Age
men
women without child
women with 1 child
women with 2 children
women with 3 children
women with 4 children or more

Remarks on the inference

$\theta_0 = \mathbb{E}\left(NPV|G\right) = f(\beta_0, \gamma_0)$ (or $\theta_0 = \mathbb{V}\left(NPV|G\right) = f(\beta_0, \gamma_0)$)

If $f(x, y)$ could be easily analytically calculated, $\widehat{\theta} = f(\widehat{\beta}, \widehat{\gamma})$ would be a consistent estimator of θ_0 and
$\mathbb{V}(\widehat{\theta}) = \nabla f(\beta_0, \gamma_0)' \mathbb{V}(\widehat{\beta}, \widehat{\gamma}) \nabla f(\beta_0, \gamma_0)$.

Microsimulation is used to approximate f by a random variable having a decreasing variance with the number of simulations, $\widehat{\theta} = \widehat{f}(\widehat{\beta}, \widehat{\gamma})$ consistent and

$$\begin{array}{ccc} \mathbb{V}(\widehat{\theta}) & = & \\ \mathbb{E}[\mathbb{V}(\widehat{\theta}|\widehat{\beta}, \widehat{\gamma})] & + & \mathbb{V}[\mathbb{E}(\widehat{\theta}|\widehat{\beta}, \widehat{\gamma})] \\ \text{"simulation"} & & \text{"estimation"} \\ \text{variance} & & \text{variance} \end{array}$$

First term is close to 0, when the number of simulation is large, but the second is currently not computed (work in progress)

Results (preliminary)

prod. growth: 1,8%; $\nearrow$ tech. prog.: 1,8%; discount factor: 0

TAB: NPV (1994-2003 cohort)

G	N_{sim}	$\widehat{\mathbb{E}}(NPV\|G)$	$\sqrt{\widehat{\mathbb{V}}(NPV\|G)}^{*}$	$\widehat{\mathbb{E}}(\sum_{a=0}^{A} C_{ia}\|G)$
All		187	870	449
Sch. age $\leq$ 19	2127	132	890	-126
Sch. age $\leq$ 21	2161	30	764	-47
Sch. age $\leq$ 23	1962	-32	821	31
Sch. age $>$ 23	1812	-157	982	170

Note : in K € 2006, $1.018^{50} \approx 2.4$, $1.018^{75} \approx 3.8$, $1.018^{100} \approx 6$

* : probably under estimate because unobserved heterogeneity is uncorrelated in the simulation of ambulatory and inpatient expenditures.

Results (preliminary)

prod. growth: 1,8%; ↗ tech. prog.: 1,8%; discount factor: 0

TAB: NPV (1994-2003 cohort)

G	N_{sim}	$\widehat{\mathbb{E}}(NPV\|G)$	$\sqrt{\widehat{\mathbb{V}}(NPV\|G)}^{*}$	$\widehat{\mathbb{E}}(\sum_{a=0}^{A} C_{ia}\|G)$
All		187	870	449
♂	4118	-52	990	77
♀ no child	689	-1	809	-57
♀ 1 child	582	72	778	-64
♀ 2 child.	1414	27	657	-69
♀ 3 child.	910	86	702	-98
♀ 4+ child.	349	158	703	-158

Note : in K € 2006, $1.018^{50} \approx 2.4$, $1.018^{75} \approx 3.8$, $1.018^{100} \approx 6$

* : probably under estimate because unobserved heterogeneity is uncorrelated in the simulation of ambulatory and inpatient expenditures.

Results (preliminary)

prod. growth: 1,8%; ↗ tech. prog.: 2,5%; discount factor: 0

TAB: NPV (1994-2003 cohort)

G	N_{sim}	$\widehat{\mathbb{E}}(NPV\|G)$	$\sqrt{\widehat{\mathbb{V}}(NPV\|G)}^{*}$	$\widehat{\mathbb{E}}(\sum_{a=0}^{A} C_{ia}\|G)$
All		590	1386	449
Sch. age $\leq$ 19	2127	132	1487	-126
Sch. age $\leq$ 21	2161	20	1217	-47
Sch. age $\leq$ 23	1962	-31	1298	31
Sch. age $>$ 23	1812	-145	1522	170

Note : in K € 2006, $1.018^{75} \approx 3.8$, $1.025^{75} \approx 6.4$

* : probably under-estimate because unobserved heterogeneity is uncorrelated in the simulation of ambulatory and inpatient expenditures.

Results (preliminary)

prod. growth: 1,8%; ↗ tech. prog.: 2,5%; discount factor: 0

TAB: NPV (1994-2003 cohort)

G	N_{sim}	$\widehat{\mathbb{E}}(NPV\|G)$	$\sqrt{\widehat{\mathbb{V}}(NPV\|G)}^{*}$	$\widehat{\mathbb{E}}(\sum_{a=0}^{A} C_{ia}\|G)$
All		590	1386	449
♂	4118	-43	1597	77
♀ no child	689	-24	1230	-57
♀ 1 child.	582	93	1235	-64
♀ 2 child.	1414	7	1007	-69
♀ 3 child.	910	85	1143	-98
♀ 4+ child.	349	151	1065	-158

Note : in K € 2006, $1.018^{75} \approx 3.8$, $1.025^{75} \approx 6.4$

* : probably under estimate because unobserved heterogeneity is uncorrelated in the simulation of ambulatory and inpatient expenditures.

Disclaimers before conclusions

- Work in progress
- Needs of robustness checks
- Estimation of precision due to estimation on the ESPS-EPAS (bootstrap on the ESPS-EPAS)

Conclusions

- Distribution between social groups but represents only a small part of the explained variance of transfers, vertical redistribution more important ($\sim$ 20% of the variance of transfers), the system is principally an insurantial system.
- Redistribution mainly explained by the contribution system (and not by the behaviors of expenditures)
- Insurance mainly explained by generosity of the reimbursement
- Open questions : What about the change of behaviors in case of reform :
  - on the reimbursements $\Rightarrow$ renunciation to heath-care ? unsustainable out-of-pocket ?
  - on the contributions $\Rightarrow$ labor supply ?
  - on the link between health and work