**Modelling and Estimating Individual and Firm Effects**

**with Count Panel Data**

**Online Appendix**

Jean-François Angers, Denise Desjardins

Georges Dionne and François Guertin

7 December 2017

## Appendix A: CHOICE AND DESCRIPTION OF VARIABLES

The unit of observation is an eligible vehicle with authorization to circulate at least one day in year *t,* and which has been followed up for at least two years. We analyze the accident totals found in SAAQ files. These totals include all the traffic accidents causing bodily injuries and all accidents causing material damage reported by the police in Quebec.

***Dependent variable***

 = the number of accidents in which vehicle *i* of fleet *f* has been involved during year *t.*  can take the values 0, 1, 2, 3, 4 and over.

***Explanatory variables***

We have two types of explanatory variables: those concerning the carrier and those concerning vehicles and drivers.

***Variables concerning the carrier***

* *Size of fleet for year t:* 7 dichotomous variables have been created.

The two-vehicle size is used as the reference category.

* *Sector of economic activity*: 5 dichotomous variables have been created for vehicles transporting goods :

sect\_14 = 1 if the main sector of activity is transporting passengers;

sect\_05 = 1 if the sector of activity is general public trucking;

sect\_06 = 1 if the sector of activity is public bulk trucking;

sect\_07 = 1 if the sector of activity is independent trucking;

sect\_08 = 1 if the sector of activity is a short-term leasing firm.

The “public bulk trucking” sector is used as the reference category.

* Seven (7) variables have been created for vehicles engaged in the *transportation of goods,* to measure the number of convictions per vehicle in the year preceding year *t* for each carrier:
* *Number of violations per vehicle for overweight committed by a carrier in the year preceding year t.* A positive sign is predicted, because more overweight violations should, on average, generate more accidents*.*
* *Number of violations per vehicle for oversize committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations for oversize should, on average, generate more accidents.
* *Number of violations per vehicle for poorly secured loads committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations for poorly secured loads should, on average, generate more accidents.
* *Number of violations per vehicle of Highway Safety Code provisions regarding transportation of hazardous materials committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations of regulations for the transportation of hazardous materials should, on average, generate more accidents.
* *Number of violations per vehicle of hours-of-service regulations committed by a carrier in the year preceding year t:* A positive sign is predicted because more violations of hours-of-service regulations should, on average, generate more accidents.
* *Number of violations per vehicle of Highway Safety Code provisions regarding mechanical inspection committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations of regulations regarding mechanical inspection should, on average, generate more accidents.
* *Number of violations per vehicle, other than those already mentioned, committed by a carrier in the year preceding year t:* A positive sign is predicted, because more violations other than those already mentioned should, on average, generate more accidents.

***Variables concerning vehicles and drivers (a vehicle may have more than one driver)***

* *Vehicle’s number of cylinders****:*** 4 dichotomous variables have been created:

cyl\_0 = 1 if the vehicle’s number of cylinders is not known;

cyl1\_5 = 1 if the vehicle has 1 to 5 cylinders;

cyl6\_7 = 1 if the vehicle has 6 to 7 cylinders;

cyl\_8p = 1 if the vehicle has 8 or more than 10 cylinders.

The group of vehicles with 8 or more than 10 cylinders is used as the reference category.

* *Vehicle’s type of fuel:* 3 dichotomous variables have been created:

diesel = 1 if the vehicle uses diesel as fuel;

fuel = 1 if the vehicle uses gas as fuel;

other = 1 if the vehicle uses another type of fuel.

The group of vehicles using diesel as fuel is considered the reference category.

* *maximum number of axles:* 7 dichotomous variables have been created:

ess\_0 = 1 if the maximum number of axles does not apply to this type of vehicle;

ess\_2 = 1 if the vehicle has two axles and a mass of between 3,000 and 4,000 kg;

ess\_2p = 1 if the vehicle has two axles and a mass higher than 4,000 kg;

ess\_3 = 1 if the vehicle is supported by a maximum of three axles;

ess\_4 = 1 if the vehicle is supported by a maximum of four axles;

ess\_5 = 1 if the vehicle is supported by a maximum of five axles;

ess\_6p = 1 if the vehicle is supported by six or more axles.

The group of vehicles with two axles and a mass of between 3 000 and 4 000 kg is used as the reference category.

* *Vehicle’s type of use:* 3 dichotomous variables for vehicles transporting goods have been created:

compr = 1 if the vehicle is meant for commercial use, including transportation of goods without a CTQ permit;

tbrgn = 1 if the vehicle is meant for transportation of goods but other than in bulk, which requires a CTQ permit;

tbrvr = 1 if the vehicle is meant for transportation of bulk goods.

The group of vehicles transporting bulk goods is used as the reference category.

* Six (6) variables have been created to measure the number of convictions per vehicle accumulated in the year preceding year *t* by one or more drivers:
* *Number of violations for speeding per vehicle, committed in the year preceding year t.* A positive sign is predicted because more speeding violations should, on average, generate more accidents.
* *Number of violations for driving with a suspended license per vehicle, committed in the year preceding year t.* A positive sign is predicted because more driving with a suspended license should, on average, generate more accidents.
* *Number of violations for running a red light per vehicle, committed in the year preceding year t.* A positive sign is predicted because more incidences of running a red light should, on average, generate more accidents*.*
* *Number of violations for failure to obey a stop sign or a signal from a traffic officer per vehicle, committed in the year preceding year t.* A positive sign is predicted because more incidents of failure to respect a stop sign or a signal from a traffic cop should, on average, generate more accidents*.*
* *Number of violations for failure to wear a seat belt per vehicle, committed in the year preceding year t.* A positive sign is predicted because more incidents of failure to wear a seat belt should, on average, generate more accidents*.*
* *Number of violations other than those mentioned per vehicle, committed in the year preceding year t.* A positive sign is predicted because a greater number of violations other than those mentioned should, on average, generate more accidents*.*

**Appendix B: Additional descriptive statistics**

Table B1 shows the distribution of the size of the fleet by year and for a total of 8 years. In 1991, we have 8,650 fleets, this number increases to 11,965 fleets in 1996 and decreases to 10,321 in 1998 for a total of 87,771 fleet-years. Among the 87,771 fleet-years, 46.51% have two vehicles and about 3% have over 20 vehicles.

**Table B1**

Size of fleet distribution (in %) by year

| Size of fleet | % by year | % total |
| --- | --- | --- |
| 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
| 2 | 47.86 | 46.31 | 46.60 | 46.82 | 45.98 | 46.08 | 45.83 | 47.03 | 46.51 |
| 3 | 19.63 | 19.75 | 19.92 | 19.71 | 19.45 | 19.30 | 19.28 | 19.10 | 19.51 |
| 4 to 5 | 15.26 | 15.99 | 15.75 | 15.54 | 16.28 | 16.02 | 16.40 | 16.26 | 15.96 |
| 6 to 9 | 9.16 | 9.40 | 9.19 | 9.46 | 9.62 | 9.88 | 9.62 | 9.27 | 9.47 |
| 10 to 20 | 5.45 | 5.72 | 5.76 | 5.74 | 5.76 | 5.68 | 5.95 | 5.64 | 5.72 |
| 21 to 50 | 1.97 | 2.06 | 2.00 | 1.89 | 2.05 | 2.14 | 2.08 | 2.01 | 2.03 |
| More than 50 | 0.68 | 0.78 | 0.78 | 0.83 | 0.86 | 0.89 | 0.85 | 0.70 | 0.80 |
| Number of fleets | 8,650 | 10,691 | 11,132 | 11,445 | 11,733 | 11,965 | 11,834 | 10,321 | 87,771 |

From Table B2, we observe that 9,963 fleets remain in the same class of fleet size during the eight years of observation (sum of the diagonal of the table), which is 56.80% of 17,542 fleets. There are 2,722 fleets whose size varies between 2 and 3 trucks, and 1,423 fleets whose size varies between 2 trucks to 4-5 trucks.

**Table B2**

Minimum and maximum fleet size distribution during the follow-up, Québec 1991-1998

| Minimum sizeof fleet | Maximum size of fleet | Total |
| --- | --- | --- |
|  2 |  3 | 4 to 5 | 6 to 9 | 10 to 20 | 21 to 50 | + 50 |  N |  % |
| 2 | 7,884 | 2,722 | 1,423 | 365 | 101 | 19 | 2 | 12,561 | 71.35 |
| 3 |  | 790 | 946 | 368 | 85 | 14 | 2 | 2,205 | 12.57 |
| 4 to 5 |  |  | 551 | 644 | 172 | 22 | 1 | 1,390 | 7.92 |
| 6 to 9 |  |  |  | 320 | 394 | 57 | 12 | 783 | 4.46 |
| 10 to 20 |  |  |  |  | 268 | 152 | 22 | 442 | 2.52 |
| 21 to 50 |  |  |  |  |  | 93 | 56 | 149 | 0.85 |
| More than 50 |  |  |  |  |  |  | 57 | 57 | 0.32 |
| Total N | 7,884 | 3,512 | 2,920 | 1,697 | 1,020 | 357 | 152 | 17,542 | 100.00 |
| % | 44.94 | 20.20 | 16.65 | 9.67 | 5.81 | 2.04 | 0.84 | 100.00 |  |

We observe from Table B3 that the average accident rate of trucks per fleet is lowest for the year 1997, followed by 1993, 1998, 1996 and 1994. In the years 1991, 1992 and 1995, the highest average rates of truck accidents per fleet were recorded. These observations are almost stable for different fleet sizes.

**Table B3**

Average truck accidents per fleet according to size of fleet and year.

| Size of fleet | Average truck accident per fleet by year  | Total |
| --- | --- | --- |
| 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
| 2 | 0.2626 | 0.2480 | 0.2219 | 0.2215 | 0.2219 | 0.2155 | 0.1809 | 0.2186 | 0.2224 |
| 3 | 0.4370 | 0.4154 | 0.3811 | 0.4007 | 0.4194 | 0.3712 | 0.3129 | 0.4049 | 0.3909 |
| 4 to 5 | 0.6689 | 0.6864 | 0.6030 | 0.6296 | 0.6408 | 0.5863 | 0.5507 | 0.6490 | 0.6239 |
| 6 à 9 | 1.3914 | 1.2259 | 1.0909 | 1.1311 | 1.1833 | 1.0981 | 1.0018 | 1.1996 | 1.1550 |
| 10 to 20 | 2.6730 | 2.6127 | 2.3744 | 2.5099 | 2.4527 | 2.4824 | 2.0767 | 2.5223 | 2.4497 |
| 21 to 50 | 5.9176 | 5.3818 | 5.0448 | 5.3565 | 5.8875 | 5.2461 | 4.8618 | 5.7681 | 5.4094 |
| More than 50 | 22.6780 | 22.4096 | 21.7701 | 22.0421 | 22.0198 | 21.0935 | 18.4700 | 22.5417 | 21.5014 |
| Average truck accidents per fleet | 0.8575 | 0.8561 | 0.7824 | 0.8157 | 0.8531 | 0.8153 | 0.7106 | 0.8120 | 0.8109 |

We have 111,106 different trucks in the database, nearly three-quarters of which are for commercial use, including transportation of goods. As indicated in Table B4, 17.52% of the trucks are used for transportation of goods other than bulk and 8% for transportation of goods in bulk.

Table B4

Vehicle use distribution.

| Vehicle use | N | %  |
| --- | --- | --- |
| Commercial use, including transport of goods without CTQ permit. (combr) | 82,798 | 74.52 |
| Transport of goods other than in bulk (tbrgn) | 19,470 | 17.52 |
| Transport of goods in bulk (tbrvr) | 8,838 | 7.95 |
| Total | 111,106 | 100.00 |

Table B5 presents the variation of average annual accidents per truck relative to the number of driver’s violations of the Highway Safety Code during the year preceding the accidents. Violations committed by drivers are very powerful in explaining truck accidents during the next year. Indeed, we observe that the year *t* accident rate is an increasing function of previous year violations committed by drivers.

**Table B5**

Average truck accidents according to the driver’s violations committed the previous year.

| Violations committed by the driver the previous year  | Year | Total |
| --- | --- | --- |
| 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
| For speeding |  |  |  |  |  |  |  |  |  |
| 0 | 0.1642 | 0.1586 | 0.1432 | 0.1486 | 0.1516 | 0.1435 | 0.1240 | 0.1498 | 0.1472 |
| 1 | 0.2974 | 0.2592 | 0.2640 | 0.2723 | 0.2631 | 0.2523 | 0.2161 | 0.2609 | 0.2556 |
| 2 | 0.2701 | 0.3410 | 0.3045 | 0.4000 | 0.3566 | 0.3249 | 0.3207 | 0.3281 | 0.3337 |
| 3 and more | 0.4194 | 0.5000 | 0.2424 | 0.4651 | 0.5506 | 0.4821 | 0.3973 | 0.4600 | 0.4505 |
| For driving with asuspended license |  |  |  |  |  |  |  |  |  |
| 0 | 0.1696 | 0.1629 | 0.1485 | 0.1547 | 0.1584 | 0.1507 | 0.1321 | 0.1574 | 0.1535 |
| 1 and more | 0.7500 | 0.5217 | 0.3750 | 0.4076 | 0.3549 | 0.3426 | 0.3017 | 0.3265 | 0.3566 |
| For running a red light |  |  |  |  |  |  |  |  |  |
| 0 | 0.1679 | 0.1617 | 0.1473 | 0.1538 | 0.1571 | 0.1491 | 0.1308 | 0.1555 | 0.1521 |
| 1 | 0.2726 | 0.2846 | 0.2764 | 0.2999 | 0.3350 | 0.3135 | 0.2981 | 0.3413 | 0.3036 |
| 2 and more | 0.5294 | 0.6667 | 0.3846 | 0.2727 | 0.6000 | 0.7272 | 0.2308 | 0.2727 | 0.5040 |
| For disobeying stop signs or police signals |  |  |  |  |  |  |  |  |  |
| 0 | 0.1677 | 0.1618 | 0.1474 | 0.1541 | 0.1572 | 0.1498 | 0.1315 | 0.1561 | 0.1524 |
| 1 | 0.3204 | 0.3140 | 0.2797 | 0.2823 | 0.3570 | 0.2931 | 0.2411 | 0.3100 | 0.2993 |
| 2 and more | 0.5000 | 0.2857 | 0.2500 | 0.5833 | 0.2941 | 0.5263 | 0.3125 | 0.5000 | 0.4016 |
| For failing to wear a seat belt |  |  |  |  |  |  |  |  |  |
| 0 | 0.1689 | 0.1626 | 0.1481 | 0.1554 | 0.1588 | 0.1508 | 0.1316 | 0.1576 | 0.1534 |
| 1 | 0.2304 | 0.2246 | 0.2293 | 0.1770 | 0.2376 | 0.2100 | 0.2096 | 0.2124 | 0.2164 |
| 2 and more | 0.4138 | 0.4333 | 0.2571 | 0.2750 | 0.1774 | 0.2653 | 0.3137 | 0.1200 | 0.2741 |
| For overweight |  |  |  |  |  |  |  |  |  |
| 0 | 0.1649 | 0.1583 | 0.1448 | 0.1517 | 0.1544 | 0.1461 | 0.1293 | 0.1540 | 0.1497 |
| 1 | 0.2430 | 0.2764 | 0.2410 | 0.2501 | 0.2432 | 0.2383 | 0.1889 | 0.2631 | 0.2394 |
| 2 and more | 0.3387 | 0.2956 | 0.3364 | 0.2926 | 0.3552 | 0.3026 | 0.2155 | 0.3874 | 0.3065 |
| For oversize |  |  |  |  |  |  |  |  |  |
| 0 | 0.1695 | 0.1632 | 0.1488 | 0.1554 | 0.1596 | 0.1515 | 0.1326 | 0.1577 | 0.1540 |
| 1 and more | 0.2836 | 0.1000 | 0.2917 | 0.2603 | 0.1574 | 0.1545 | 0.2269 | 0.2821 | 0.2119 |
| For poorly secured loads |  |  |  |  |  |  |  |  |  |
| 0 | 0.1688 | 0.1625 | 0.1482 | 0.1550 | 0.1587 | 0.1509 | 0.1323 | 0.1570 | 0.1534 |
| 1 and more | 0.3185 | 0.3198 | 0.2667 | 0.2665 | 0.2656 | 0.2621 | 0.2214 | 0.3778 | 0.2791 |
| For exceeding hours of service |  |  |  |  |  |  |  |  |  |
| 0 | 0.1696 | 0.1632 | 0.1486 | 0.1556 | 0.1592 | 0.1513 | 0.1325 | 0.1575 | 0.1539 |
| 1 and more | 0.5714 | 0.3000 | 0.6333 | 0.1951 | 0.3529 | 0.2743 | 0.3881 | 0.3571 | 0.3496 |
| For failure to undergo mechanical inspection |  |  |  |  |  |  |  |  |  |
| 0 | 0.1691 | 0.1626 | 0.1474 | 0.1546 | 0.1578 | 0.1509 | 0.1321 | 0.1572 | 0.1532 |
| 1 and more | 0.2890 | 0.3180 | 0.2388 | 0.2534 | 0.3024 | 0.2251 | 0.2168 | 0.2768 | 0.2591 |

We note in Table B6 that 78.80% of the 111,106 trucks use diesel as fuel.

Table B6

Type of fuel distribution

| Type of fuel | N | %  |
| --- | --- | --- |
| Diesel | 87,546 | 78.80 |
| Gas | 22,999 | 20.70 |
| Other | 561 | 0.50 |
| Total | 111,106 | 100.00 |

Table B7 illustrates that 21.15% of the 111,106 trucks have six axles or more and 28.57% have two axles and weigh more than 4,000 kg, and Table B8 shows that 64.95% of the 111,106 trucks have 6 to 7 cylinders. Only 1.15% has 5 cylinders or fewer.

Table B7

Number of axles distribution

| Number of axles | N | %  |
| --- | --- | --- |
| 2 axles (3,000 to 4,000 kg) | 15,960 | 14.36 |
| 2 axles (More than 4,000 kg) | 31,747 | 28.57 |
| 3 axles | 21,856 | 19.67 |
| 4 axles | 7,377 | 6.64 |
| 5 axles | 10,666 | 9.60 |
| 6 axles and more | 23,500 | 21.15 |
| Total | 111,106 | 100.00 |

Table B8

Number of cylinders distribution

| Number of cylinders | N | %  |
| --- | --- | --- |
| Unknown | 501 | 0.45 |
| 1 to 5 cylinders | 1,283 | 1.15 |
| 6 to 7 cylinders | 71,159 | 64.05 |
| 8 or more than 10 cylinders | 38,163 | 34.35 |
| Total | 111,106 | 100.00 |

Table B9 indicates that 10.64% of the 111,106 trucks have 8 years of follow-up, which represents 10.64% of the population.

Table B9

Number of years of follow-up of the truck

| Number of years of follow-up | N | %  |
| --- | --- | --- |
| 2 | 30,716 | 27.65 |
| 3 | 23,270 | 20.94 |
| 4 | 17,831 | 16.05 |
| 5 | 11,998 | 10.80 |
| 6 | 9,241 | 8.32 |
| 7 | 6,225 | 5.60 |
| 8 | 11,825 | 10.64 |
| Total | 111,106 | 100.00 |

We note in Table B10 that there are 30,432 trucks for which we have two consecutive years of follow-up, which corresponds to 99.07% (30,432/30,716) of trucks with two observation periods. This percentage varies from 98.43 (3 periods) to 97.65 (7 periods).

**Table B10**

Number of consecutive years of follow-up of the trucks by year of follow-up start, Quebec 1991 to 1997.

| Number of year of follow-up  | Year of follow-up start | Total |
| --- | --- | --- |
| 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 |
| 2 | 8,326 | 2,581 | 2,193 | 2,081 | 2,844 | 2,351 | 10,056 | 30,432 |
| 3 | 6,421 | 2,291 | 1,624 | 1,855 | 1,947 | 8,766 |  | 22,904 |
| 4 | 5,273 | 1,535 | 1,711 | 1,524 | 7,304 |  |  | 17,347 |
| 5 | 3,967 | 1,289 | 1,067 | 5,226 |  |  |  | 11,549 |
| 6 | 3,680 | 818 | 4,441 |  |  |  |  | 8,939 |
| 7 | 2,630 | 3,449 |  |  |  |  |  | 6,079 |
| 8 | 11,825 |  |  |  |  |  |  | 11,825 |
| Total | 42,122 | 11,963 | 11,036 | 10,686 | 12,095 | 11,117 | 10,056 | 109,075 |

## Appendix C1: Poisson and nb2 models Estimation reSults

**Table C1.1: Poisson negative binomial estimates**

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of two trucks or more and trucks with two periods or more), Poisson and NB2 models

| Explanatory variables | Poisson model | NB2 model |
| --- | --- | --- |
| Coefficient | Standard error | Coefficient | Standard error |
| Constant | -3.5846\* | 0.0415 | -3.5895\* | 0.0438 |
| Number of years as carrier at 31 December  | -0.0424\* | 0.0026 | -0.0432\* | 0.0028 |
| Sector of activity in 1998 |  |  |  |  |
|  Other sector | -0.2766\* | 0.0804 | -0.2694\* | 0.0839 |
|  General public trucking | 0.0933\* | 0.0210 | 0.0977\* | 0.0226 |
| Bulk public trucking | Reference group | Reference group |
|  Private trucking | 0.1548\* | 0.0177 | 0.1595\* | 0.0190 |
|  Short-term rental firm | 0.4055\* | 0.0275 | 0.4185\* | 0.0299 |
| Size of fleet |  |  |  |  |
| 2 | Reference group | Reference group |
| 3 | 0.1245\* | 0.0161 | 0.1246\* | 0.0171 |
| 4 to 5 | 0.1900\* | 0.0151 | 0.1926\* | 0.0160 |
| 6 to 9 | 0.2764\* | 0.0148 | 0.2797\* | 0.0158 |
| 10 to 20 | 0.3704\* | 0.0142 | 0.3761\* | 0.0152 |
| 21 to 50 | 0.3698\* | 0.0151 | 0.3782\* | 0.0161 |
| More than 50 | 0.3837\* | 0.0142 | 0.3892\* | 0.0151 |
| Number of days authorized to drive in previous year  | 1.6703\* | 0.0290 | 1.6765\* | 0.0300 |
| Number of violations of trucking standards in year before |  |  |  |  |
|  For overload | 0.1456\* | 0.0104 | 0.1502\* | 0.0117 |
|  For excessive size | 0.1607\*\*\* | 0.0825 | 0.1615\*\*\* | 0.0910 |
|  For poorly secured cargo | 0.2927\* | 0.0329 | 0.2991\* | 0.0380 |
|  For failure to respect service hours | 0.2771\* | 0.0598 | 0.2880\* | 0.0710 |
|  For failure to pass mechanical inspection | 0.2819\* | 0.0280 | 0.2977\* | 0.0316 |
|  For other reasons | 0.2812\* | 0.0699 | 0.2602\* | 0.0807 |
| Type of vehicle use |  |  |  |  |
|  Commercial use including transport of goods without C.T.Q. permit | -0.1167\* | 0.0177 | -0.1249\* | 0.0191 |
|  Transport of other than "bulk" goods | -0.0325 | 0.0203 | -0.0387\*\*\* | 0.0220 |
| Transport of "bulk" goods | Reference group | Reference group |
| Type of fuel |  |  |  |  |
| Diesel | Reference group | Reference group |
|  Gas | -0.3922\* | 0.0124 | -0.3939\* | 0.0130 |
|  Other | -0.3169\* | 0.0684 | -0.3161\* | 0.0713 |
| Number of cylinders |  |  |  |  |
|  1 to 5 cylinders | 0.3536\* | 0.0360 | 0.3527\* | 0.0385 |
|  6 to 7 cylinders | 0.3752\* | 0.0114 | 0.3763\* | 0.0121 |
| 8 or more than 10 cylinders | Reference group | Reference group |
| Number of axles |  |  |  |  |
| 2 axles (3,000 to 4,000 kg) | -0.1603\* | 0.0177 | -0.1616\* | 0.0188 |
| 2 axles (more than 4,000 kg) | -0.1505\* | 0.0122 | -0.1541\* | 0.0132 |
| 3 axles | -0.1156\* | 0.0124 | -0.1203\* | 0.0133 |
| 4 axles | -0.1818\* | 0.0163 | -0.1817\* | 0.0175 |
| 5 axles | -0.2040\* | 0.0145 | -0.2056\* | 0.0156 |
| 6 axles or more | Reference group | Reference group |
| Number of violations with demerit points year before |  |  |  |  |
| For speeding | 0.2961\* | 0.0092 | 0.3098\* | 0.0106 |
|  For driving with suspended license | 0.4895\* | 0.0350 | 0.5590\* | 0.0433 |
|  For running a red light | 0.4549\* | 0.0226 | 0.4723\* | 0.0256 |
|  For ignoring stop sign or traffic officer | 0.4953\* | 0.0244 | 0.5107\* | 0.0277 |
|  For not wearing a seat belt | 0.2295\* | 0.0281 | 0.2386\* | 0.0310 |
| Observation period |  |  |  |  |
| 1991 | 0.0099 | 0.0222 | 0.0142 | 0.0239 |
| 1992 | -0.0225 | 0.0202 | -0.0195 | 0.0217 |
| 1993 | -0.0881\* | 0.0189 | -0.0876\* | 0.0203 |
| 1994 | -0.0228 | 0.0174 | -0.0218 | 0.0187 |
| 1995 | -0.0012 | 0.0163 | -0.0011 | 0.0175 |
| 1996 | -0.0463\* | 0.0157 | -0.0453\* | 0.0168 |
| 1997 | -0.1605\* | 0.0158 | -0.1597\* | 0.0168 |
| 1998 | Reference group | Reference group |
|  |  |  | 0.8135 | 0.0282 |
| Number of observations:  | 456,117 | 456,117 |

\* significant at 1%; \*\* significant at 5%; \*\*\* significant at 10%

**Table C1.2: Robustness analysis of the estimations**

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of two trucks or more and trucks with two periods or more), Gamma-Dirichlet models (excluding the variable fleet size at left, cylinders in the middle, and observation period at right).

| Explanatory variables | Gamma-Dirichlet model | Gamma-Dirichlet model | Gamma-Dirichlet model |
| --- | --- | --- | --- |
| Fleet size omitted | Cylinders omitted | Period omitted |
| Coefficient | Standard error | Coefficient | Standard error | Coefficient | Standard error |
| Constant | -3.8990\* | 0.0572 | -3.5304\* | 0.0563 | -3.7451\* | 0.0408 |
| Number of years as carrier at 31 December  | -0.0331\* | 0.0044 | -0.0467\* | 0.0044 | -0.0578\* | 0.0019 |
| Sector of activity in 1998 |  |  |  |  |  |  |
|  Other sector | -0.0802 | 0.1173 | -0.1170 | 0.1169 | -0.1505 | 0.1159 |
|  General public trucking | 0.2429\* | 0.0303 | 0.1843\* | 0.0305 | 0.1658\* | 0.0303 |
| Bulk public trucking | Reference group | Reference group | Reference group |
|  Private trucking | 0.2775\* | 0.0257 | 0.2358\* | 0.0258 | 0.2201\* | 0.0256 |
|  Short-term rental firm | 0.6832\* | 0.0482 | 0.5976\* | 0.0488 | 0.5655\* | 0.0482 |
| Size of fleet |  |  |  |  |  |  |
| 2 |  | Reference group | Reference group |
| 3 |  |  | 0.0830\* | 0.0206 | 0.0811\* | 0.0205 |
| 4 to 5 |  |  | 0.1467\* | 0.0206 | 0.1403\* | 0.0205 |
| 6 to 9 |  |  | 0.2237\* | 0.0211 | 0.2184\* | 0.0209 |
| 10 to 20 |  |  | 0.3093\* | 0.0211 | 0.2981\* | 0.0208 |
| 21 to 50 |  |  | 0.3137\* | 0.0225 | 0.3051\* | 0.0222 |
| More than 50 |  |  | 0.3182\* | 0.0219 | 0.3193\* | 0.0215 |
| Number of days authorized to drive in previous year  | 2.0586\* | 0.0299 | 2.0414\* | 0.0299 | 2.0408\* | 0.0297 |
| Number of violations of trucking standards in year before |  |  |  |  |  |  |
|  For overload | 0.0949\* | 0.0115 | 0.1015\* | 0.0114 | 0.0983\* | 0.0114 |
|  For excessive size | 0.1434\*\*\* | 0.0862 | 0.1378 | 0.0859 | 0.1514\*\*\* | 0.0861 |
|  For poorly secured cargo | 0.2011\* | 0.0357 | 0.2048\* | 0.0355 | 0.2227\* | 0.0355 |
|  For failure to respect service hours | 0.1963\* | 0.0667 | 0.2012\* | 0.0664 | 0.2141\* | 0.0663 |
|  For failure to pass mechanical inspection | 0.1685\* | 0.0300 | 0.1807\* | 0.0299 | 0.1994\* | 0.0300 |
|  For other reasons | 0.1709\*\* | 0.0744 | 0.1779\*\* | 0.0743 | 0.1717\*\* | 0.0744 |
| Type of vehicle use |  |  |  |  |  |  |
|  Commercial use including transport of goods without C.T.Q. permit | -0.1802\* | 0.0212 | -0.2139\* | 0.0212 | -0.1901\* | 0.0212 |
|  Transport of other than "bulk" goods | -0.0900\* | 0.0243 | -0.1233\* | 0.0243 | -0.1138\* | 0.0242 |
| Transport of "bulk" goods | Reference group | Reference group | Reference group |
| Type of fuel |  |  |  |  |  |  |
| Diesel | Reference group | Reference group | Reference group |
|  Gas | -0.3993\* | 0.0137 | -0.4945\* | 0.0134 | -0.3976\* | 0.0136 |
|  Other | -0.2933\* | 0.0738 | -0.4617\* | 0.0734 | -0.3058\* | 0.0736 |
| Number of cylinders |  |  |  |  |  |  |
|  1 to 5 cylinders | 0.2171\* | 0.0406 |  |  | 0.2162\* | 0.0402 |
|  6 to 7 cylinders | 0.3865\* | 0.0127 |  |  | 0.3782\* | 0.0126 |
| 8 or more than 10 cylinders | Reference group |  | Reference group |
| Number of axles |  |  |  |  |  |  |
| 2 axles (3,000 to 4,000 kg) | -0.3091\* | 0.0209 | -0.5535\* | 0.0189 | -0.2854\* | 0.0208 |
| 2 axles (> 4,000 kg) | -0.3003\* | 0.0151 | -0.3884\* | 0.0147 | -0.2806\* | 0.0150 |
| 3 axles | -0.1371\* | 0.0150 | -0.1603\* | 0.0149 | -0.1238\* | 0.0149 |
| 4 axles | -0.1302\* | 0.0191 | -0.1551\* | 0.0190 | -0.1315\* | 0.0190 |
| 5 axles | -0.2048\* | 0.0175 | -0.2027\* | 0.0174 | -0.1954\* | 0.0174 |
| 6 axles or more | Reference group | Reference group | Reference group |
| Number of violations with demerit points year before |  |  |  |  |  |  |
| For speeding | 0.1898\* | 0.0103 | 0.1974\* | 0.0103 | 0.1918\* | 0.0103 |
|  For driving with suspended license | 0.3725\* | 0.0423 | 0.3816\* | 0.0421 | 0.4026\* | 0.0421 |
|  For running a red light | 0.3014\* | 0.0239 | 0.3178\* | 0.0239 | 0.3130\* | 0.0239 |
|  For ignoring stop sign or traffic officer | 0.3519\* | 0.0258 | 0.3626\* | 0.0258 | 0.3620\* | 0.0258 |
|  For not wearing a seat belt | 0.1484\* | 0.0295 | 0.1559\* | 0.0294 | 0.1536\* | 0.0295 |
| Observation period |  |  |  |  |  |  |
| 1991 | 0.1575\* | 0.0332 | 0.0442 | 0.0334 |  |  |
| 1992 | 0.1299\* | 0.0293 | 0.0271 | 0.0295 |  |  |
| 1993 | 0.1431\* | 0.0260 | 0.0592\*\* | 0.0261 |  |  |
| 1994 | 0.2348\* | 0.0226 | 0.1689\* | 0.0227 |  |  |
| 1995 | 0.2483\* | 0.0197 | 0.1978\* | 0.0198 |  |  |
| 1996 | 0.1502\* | 0.0175 | 0.1146\* | 0.0175 |  |  |
| 1997 | -0.0590\* | 0.0163 | -0.0813\* | 0.0163 |  |  |
| 1998 | Reference group | Reference group |  |
|   | 2.0152\* | 0.0424 | 1.9876\* | 0.0415 | 2.0077\* | 0.0422 |
|   | 13.3070\* | 0.2580 | 13.0769\* | 0.2556 | 12.6287\* | 0.2503 |
|  | 4.6682\* | 0.3100 | 4.6666\* | 0.3098 | 4.6683\* | 0.3101 |
| Number of observations:  | 456,117 | 456,117 | 456,117 |

\* Significant at 1%; \*\* Significant at 5%; \*\*\* Significant at 10%

**Table C1.3: Estimating sample results for predicted numbers of accidents**

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1997 period (fleet of two trucks or more and trucks with two periods or more), Gamma-Dirichlet models.

| Explanatory variables | Gamma-Dirichlet model |
| --- | --- |
| Coefficient | Standard error |
| Constant | -3.6903\* | 0.0438 |
| Number of years as carrier at 31 December  | -0.0468\* | 0.0051 |
| Sector of activity in 1998 |  |  |
|  Other sector | -0.1260 | 0.1201 |
|  General public trucking | 0.1792\* | 0.0326 |
| Bulk public trucking | Reference group |
|  Private trucking | 0.2336\* | 0.0275 |
|  Short-term rental firm | 0.5916\* | 0.0508 |
| Size of fleet |  |  |
| 2 | Reference group |
| 3 | 0.0774\* | 0.0219 |
| 4 to 5 | 0.1357\* | 0.0219 |
| 6 to 9 | 0.2071\* | 0.0224 |
| 10 to 20 | 0.2871\* | 0.0222 |
| 21 to 50 | 0.2817\* | 0.0237 |
| More than 50 | 0.3070\* | 0.0229 |
| Number of days authorized to drive in previous year  | 2.0100\* | 0.0317 |
| Number of violations of trucking standards in year before |  |  |
|  For overload | 0.0965\* | 0.0120 |
|  For excessive size | 0.1423 | 0.0901 |
|  For poorly secured cargo | 0.2081\* | 0.0377 |
|  For failure to respect service hours | 0.2213\* | 0.0757 |
|  For failure to pass mechanical inspection | 0.1877\* | 0.0315 |
|  For other reasons | 0.1568\*\*\* | 0.0812 |
| Type of vehicle use |  |  |
|  Commercial use including transport of goods without C.T.Q. permit | -0.1936\* | 0.0228 |
|  Transport of other than "bulk" goods | -0.1061\* | 0.0261 |
| Transport of "bulk" goods | Reference group |
| Type of fuel |  |  |
| Diesel | Reference group |
|  Gas | -0.3819\* | 0.0142 |
|  Other | -0.3830\* | 0.0810 |
| Number of cylinders |  |  |
|  1 to 5 cylinders | 0.2319\* | 0.0433 |
|  6 to 7 cylinders | 0.3702\* | 0.0133 |
| 8 or more than 10 cylinders | Reference group |
| Number of axles |  |  |
| 2 axles (3,000 to 4,000 kg) | -0.2738\* | 0.0221 |
| 2 axles (more than 4,000 kg) | -0.2809\* | 0.0160 |
| 3 axles | -0.1303\* | 0.0160 |
| 4 axles | -0.1421\* | 0.0203 |
| 5 axles | -0.1958\* | 0.0186 |
| 6 axles or more | Reference group |
| Number of violations with demerit points year before |  |  |
| For speeding | 0.1961\* | 0.0113 |
|  For driving with suspended license | 0.4088\* | 0.0449 |
|  For running a red light | 0.3041\* | 0.0256 |
|  For ignoring stop sign or traffic officer | 0.3495\* | 0.0277 |
|  For not wearing a seat belt | 0.1684\* | 0.0306 |
| Observation period | -0.0238\* | 0.0053 |
|   | 2.0657\* | 0.0480 |
|   | 11.7490\* | 0.2480 |
|  | 4.7158\* | 0.3412 |
| Number of observations:  | 393,634 |

 \* Significant at 1%; \*\* Significant at 5%; \*\*\* Significant at 10%.

## Appendix C2: Estimation reSults for fleets of more than four trucks

**Table C2.1**

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of more than four trucks and trucks with two periods or more): Poisson and NB2 models.

| **Explanatory variables** | **Poisson model** | **NB2 model** |
| --- | --- | --- |
| **Coefficient** | **Standard deviation** | **Coefficient** | **Standard deviation** |
| ***Constant*** | -3.5145\* | 0.0495 | -3.5211\* | 0.0524 |
| ***Number of years as carrier at 31 December***  | -0.0372\* | 0.0032 | -0.0377\* | 0.0034 |
| ***Sector of activity in 1998*** |  |  |  |  |
|  Other sector | -0.3248\* | 0.0923 | -0.3180\* | 0.0964 |
|  General public trucking | 0.0913\* | 0.0242 | 0.0964\* | 0.0262 |
| Bulk public trucking | Reference group | Reference group |
|  Private trucking | 0.1714\* | 0.0214 | 0.1776\* | 0.0232 |
|  Short-term rental firm | 0.4264\* | 0.0301 | 0.4416\* | 0.0329 |
| ***Size of fleet*** |  |  |  |  |
| 5 | Reference group | Reference group |
| 6 to 9 | 0.0654\* | 0.0140 | 0.0661\* | 0.0150 |
| 10 to 20 | 0.1622\* | 0.0132 | 0.1649\* | 0.0142 |
| 21 to 50 | 0.1596\* | 0.0142 | 0.1644\* | 0.0153 |
| More than 50 | 0.1705\* | 0.0133 | 0.1720\* | 0.0142 |
| ***Number of days authorized to circulate year before*** | 1.7167\* | 0.0328 | 1.7231\* | 0.0339 |
| ***Number of violations of trucking standards year before*** |  |  |  |  |
|  For overload | 0.1375\* | 0.0119 | 0.1413\* | 0.0135 |
|  For excessive size | 0.1725\*\*\* | 0.0964 | 0.1786\*\*\* | 0.1071 |
|  For poorly secured cargo | 0.2669\* | 0.0374 | 0.2720\* | 0.0433 |
|  For failure to respect service hours | 0.2507\* | 0.0668 | 0.2557\* | 0.0785 |
|  For failure to pass mechanical inspection | 0.2330\* | 0.0327 | 0.2449\* | 0.0374 |
|  For other reasons | 0.3083\* | 0.0758 | 0.2846\* | 0.0885 |
| ***Type of vehicle use*** |  |  |  |  |
|  Commercial use including transport of goods without C.T.Q. permit | -0.0748\* | 0.0210 | -0.0813\* | 0.0229 |
|  Transport of other than "bulk" goods | -0.0065 | 0.0232 | -0.0118 | 0.0253 |
| Transport of "bulk" goods | Reference group | Reference group |
| ***Type of fuel*** |  |  |  |  |
| Diesel | Reference group | Reference group |
|  Gas | -0.3387\* | 0.0140 | -0.3400\* | 0.0148 |
|  Others | -0.2869\* | 0.0735 | -0.2859\* | 0.0769 |
| ***Number of cylinders*** |  |  |  |  |
|  1 to 5 cylinders | 0.3369\* | 0.0424 | 0.3352\* | 0.0454 |
|  6 to 7 cylinders | 0.3725\* | 0.0130 | 0.3732\* | 0.0137 |
| 8 or more than 10 cylinders | Reference group | Reference group |
| ***Number of axles*** |  |  |  |  |
| 2 axles (3,000 to 4,000 kg) | -0.1840\* | 0.0202 | -0.1859\* | 0.0215 |
| 2 axles (more than 4,000 kg) | -0.1308\* | 0.0134 | -0.1344\* | 0.0145 |
| 3 axles | -0.0678\* | 0.0137 | -0.0723\* | 0.0148 |
| 4 axles | -0.1951\* | 0.0178 | -0.1951\* | 0.0191 |
| 5 axles | -0.1850\* | 0.0159 | -0.1864\* | 0.0171 |
| 6 axles or more | Reference group | Reference group |
| ***Number of violations with demerit points year before*** |  |  |  |  |
| For speeding | 0.2819\* | 0.0105 | 0.2930\* | 0.0122 |
|  For driving under suspension | 0.5355\* | 0.0461 | 0.5713\* | 0.0558 |
|  For running a red light | 0.4070\* | 0.0262 | 0.4200\* | 0.0299 |
|  For ignoring stop sign or traffic agent | 0.4735\* | 0.0280 | 0.4843\* | 0.0321 |
|  For not wearing a seat belt | 0.1910\* | 0.0331 | 0.1969\* | 0.0367 |
| ***Observation period*** |  |  |  |  |
| 1991 | 0.0109 | 0.0268 | 0.0146 | 0.0290 |
| 1992 | -0.0221 | 0.0242 | -0.0188 | 0.0262 |
| 1993 | -0.0817\* | 0.0224 | -0.0811\* | 0.0241 |
| 1994 | -0.0147 | 0.0204 | -0.0129 | 0.0220 |
| 1995 | 0.0044 | 0.0188 | 0.0050 | 0.0202 |
| 1996 | -0.0373\*\* | 0.0177 | -0.0355\*\*\* | 0.0191 |
| 1997 | -0.1443\* | 0.0176 | -0.1438\* | 0.0189 |
| 1998 | Reference group | Reference group |
|   |  |  | 0.8032\* | 0.0203 |
| Number of observations:  | 336,772 | 336,772 |

\* significant at 1%; \*\* significant at 5%; \*\*\* significant at 10%

**Table C2.2**

Estimation of the parameters of the distribution of the number of annual truck accidents for the 1991-1998 period (fleet of more than four trucks and trucks with two periods or more): Hausman’s model and Gamma-Dirichlet model.

| **Explanatory variables** | **Hausman’s model** | **Gamma-Dirichlet model** |
| --- | --- | --- |
| **Coefficient** | **Standard deviation** | **Coefficient** | **Standard deviation** |
| ***Constant*** | -0.0290 | 0.0963 | -3.8350\* | 0.0829 |
| ***Number of years as carrier at 31 December***  | -0.0381\* | 0.0038 | -0.0401\* | 0.0068 |
| ***Sector of activity in 1998*** |  |  |  |  |
|  Other sector | -0.3001\* | 0.1068 | -0.1768 | 0.1561 |
|  General public trucking | 0.0988\* | 0.0293 | 0.1442\* | 0.0402 |
| Bulk public trucking | Reference group | Reference group |
|  Private trucking | 0.1761\* | 0.0261 | 0.2470\* | 0.0358 |
|  Short-term rental firm | 0.4730\* | 0.0369 | 0.5967\* | 0.0626 |
| ***Size of fleet*** |  |  |  |  |
| 5 | Reference group | Reference group |
| 6 to 9 | 0.0648\* | 0.0161 | -0.0004 | 0.0918 |
| 10 to 20 | 0.1466\* | 0.0158 | 0.0522\*\* | 0.0219 |
| 21 to 50 | 0.1396\* | 0.0170 | 0.0489\*\* | 0.0244 |
| More than 50 | 0.1373\* | 0.0160 | 0.0515\*\* | 0.0245 |
| ***Number of days authorized to circulate in year before*** | 1.7256\* | 0.0338 | 2.1354\* | 0.0338 |
| ***Number of violations of trucking standards in year before*** |  |  |  |  |
|  For overload | 0.1099\* | 0.0133 | 0.0828\* | 0.0219 |
|  For excessive size | 0.1570 | 0.1030 | 0.1571 | 0.0992 |
|  For poorly secured cargo | 0.2282\* | 0.0411 | 0.1786\* | 0.0397 |
|  For failure to respect service hours | 0.2329\* | 0.0732 | 0.1709\*\* | 0.0728 |
|  For failure to pass mechanical inspection | 0.1807\* | 0.0359 | 0.1141\* | 0.0346 |
|  For other reasons | 0.2982\* | 0.0853 | 0.1800\*\* | 0.0806 |
| ***Type of vehicle use*** |  |  |  |  |
|  Commercial use including transport of goods without C.T.Q. permit | -0.1004\* | 0.0254 | -0.1646\* | 0.0251 |
|  Transport of other than "bulk" goods | -0.0300 | 0.0280 | -0.1009\* | 0.0278 |
| Transport of "bulk" goods | Reference group | Reference group |
| ***Type of fuel*** |  |  |  |  |
| Diesel | Reference group | Reference group |
|  Gas | -0.3521\* | 0.0167 | -0.3509\* | 0.0152 |
|  Others | -0.2782\* | 0.0840 | -0.2652\* | 0.0789 |
| ***Number of cylinders*** |  |  |  |  |
|  1 to 5 cylinders | 0.3366\* | 0.0526 | 0.1441\* | 0.0462 |
|  6 to 7 cylinders | 0.3724\* | 0.0156 | 0.3695\* | 0.0141 |
| 8 or more than 10 cylinders | Reference group | Reference group |
| ***Number of axles*** |  |  |  |  |
| 2 axles (3,000 to 4,000 kg) | -0.1852\* | 0.0242 | -0.3573\* | 0.0237 |
| 2 axles (more than 4,000 kg) | -0.1505\* | 0.0166 | -0.3135\* | 0.0167 |
| 3 axles | -0.1088\* | 0.0170 | -0.0963\* | 0.0167 |
| 4 axles | -0.2013\* | 0.0218 | -0.1174\* | 0.0208 |
| 5 axles | -0.1968\* | 0.0190 | -0.1768\* | 0.0193 |
| 6 axles or more | Reference group | Reference group |
| ***Number of violations with demerit points year before*** |  |  |  |  |
| For speeding | 0.2433\* | 0.0118 | 0.1719\* | 0.0115 |
|  For driving under suspension | 0.4705\* | 0.0519 | 0.3862\* | 0.0495 |
|  For running a red light | 0.3392\* | 0.0286 | 0.2697\* | 0.0272 |
|  For ignoring stop sign or traffic agent | 0.4042\* | 0.0306 | 0.3337\* | 0.0292 |
|  For not wearing a seat belt | 0.1659\* | 0.0357 | 0.1211\* | 0.0341 |
| ***Observation period*** |  |  |  |  |
| 1991 | 0.0231 | 0.0309 | 0.1004\*\* | 0.0496 |
| 1992 | -0.0151 | 0.0276 | 0.0833\*\*\* | 0.0433 |
| 1993 | -0.0761\* | 0.0251 | 0.1252\* | 0.0375 |
| 1994 | -0.0111 | 0.0225 | 0.2314\* | 0.0316 |
| 1995 | 0.0091 | 0.0203 | 0.2498\* | 0.0262 |
| 1996 | -0.0319\*\*\* | 0.0188 | 0.1615\* | 0.0216 |
| 1997 | -0.1409\* | 0.0182 | -0.0464\*\* | 0.0188 |
| 1998 | Reference group | Reference group |
|  | 57.9375\* | 4.0818 |  |  |
|  | 1.8363\* | 0.0420 |  |  |
|   |  |  | 1.9181\* | 0.0416 |
|   |  |  | 22.0245\* | 0.5539 |
|  |  |  | 4.7245\* | 0.3430 |
| Number of observations:  | 336,772 | 336,772 |

\* significant at 1%; \*\* significant at 5%; \*\*\* significant at 10%

## Appendix D: Model estimations, R-Code

#

#Read the dataset

#

library(foreign, pos=15)

Dataset <- read.table("donnee.csv", header=TRUE, sep=",",na.strings="NA", dec=".", strip.white=TRUE)

library(abind, pos=16)

library(e1071, pos=17)

library("BMS")

library("spuRs")

library("MASS")

# **Poisson model estimation**

GLM.2 <- glm(NB\_ATOT ~ AN\_TRANS + SECT\_14 + SECT\_05 + SECT\_07 + SECT\_08 +

 N\_VH3 + N\_VH45 + N\_VH69 + N\_VH20 + N\_VH50 + N\_VH51 +

 DUREE\_AT + NB\_INF1 + NB\_INF2 + NB\_INF3 + NB\_INF6 + NB\_INF7 + NB\_INF89 +

 COMPR + TBRGN + ESSENCE + CARB\_AUT + CYL1\_5 + CYL6\_7 +

 ESS\_02 + ESS\_02P + ESS\_03 + ESS\_04 + ESS\_05 +

 VIT + SANCT + ROUGE + ARRET + CEINTURE +

 an\_91 + an\_92 + an\_93 + an\_94 + an\_95 + an\_96 + an\_97,

 family=poisson(log), data=Dataset)

est=GLM.2$coefficients

# **Negative Binomial model estimation**

GLM.3 <- glm.nb(NB\_ATOT ~ AN\_TRANS + SECT\_14 + SECT\_05 + SECT\_07 + SECT\_08 +

 N\_VH3 + N\_VH45 + N\_VH69 + N\_VH20 + N\_VH50 + N\_VH51 +

 DUREE\_AT + NB\_INF1 + NB\_INF2 + NB\_INF3 + NB\_INF6 + NB\_INF7 + NB\_INF89 +

 COMPR + TBRGN + ESSENCE + CARB\_AUT + CYL1\_5 + CYL6\_7 +

 ESS\_02 + ESS\_02P + ESS\_03 + ESS\_04 + ESS\_05 +

 VIT + SANCT + ROUGE + ARRET + CEINTURE +

 an\_91 + an\_92 + an\_93 + an\_94 + an\_95 + an\_96 + an\_97,

 start=est,init.theta=1, data=Dataset)

estNB=GLM.3$coefficients

# **Negative Binomial model with random effect (Hausman’s model) estimation**

#Create the vector y: number of accident of truck i at time t

y <- as.matrix(cbind(Dataset$NB\_ATOT))

max\_y <- max(y)

s\_y <- sum(y)

#Create the matrix x : Variables concerning the carriers, vehicles and the drivers (a vehicle may have more than one

#driver

x <- as.matrix(cbind(1, Dataset$AN\_TRANS, Dataset$SECT\_14, Dataset$SECT\_05, Dataset$SECT\_07,

Dataset$SECT\_08, Dataset$N\_VH3, Dataset$N\_VH45, Dataset$N\_VH69, Dataset$N\_VH20, Dataset$N\_VH50, Dataset$N\_VH51, Dataset$DUREE\_AT, Dataset$NB\_INF1, Dataset$NB\_INF2, Dataset$NB\_INF3, Dataset$NB\_INF6, Dataset$NB\_INF7, Dataset$NB\_INF89, Dataset$COMPR, Dataset$TBRGN, Dataset$ESSENCE, Dataset$CARB\_AUT, Dataset$CYL1\_5, Dataset$CYL6\_7, Dataset$ESS\_02, Dataset$ESS\_02P, Dataset$ESS\_03, Dataset$ESS\_04, Dataset$ESS\_05, Dataset$VIT, Dataset$SANCT, Dataset$ROUGE, Dataset$ARRET, Dataset$CEINTURE, Dataset$an\_91, Dataset$an\_92, Dataset$an\_93, Dataset$an\_94, Dataset$an\_95, Dataset$an\_96, Dataset$an\_97))

n <- nrow(x) # Total number of observations

p <- ncol(x) # Number of parameters

p1 <- p+1

p2 <- p+2

nper<- as.matrix(cbind(Dataset$n\_period,Dataset$camion))

nper1<-nper[Dataset$camion == 1,]

n\_period<-cbind(nper1[,1])

#number of trucks

ki <- nrow(n\_period)

#Initial values

r\_beta <- c(est, 57, 1.8)

#Log likelihood function

llf <- function (r\_beta) {

 parp <- r\_beta[1:p]

 a <- r\_beta[p1]

 b <- r\_beta[p2]

 r\_llf <- 0

 nx <- 0

 for (j in 1:ki){

 rl <- 0

 ni <- n\_period[j]

 i\_deb <- nx+1

 i\_fin <- nx+ni

 ri <- i\_deb : i\_fin

 nx <- i\_fin

 yi <- y[ri]

 xi <- x[ri,]

 zi <- xi%\*%parp

 mui <- exp(zi)

 s\_mui <- sum(mui)

 s\_yi <- sum(yi)

 ter\_1 <- lgamma(a+b) + lgamma(a+s\_mui) + lgamma(b+s\_yi)

 ter\_2 <- lgamma(a) + lgamma(b) + lgamma(a+b+s\_yi+s\_mui)

 ter\_3 <- lgamma(mui+yi) - lgamma(mui) - lgamma(yi+1)

 s\_ter\_3=sum(ter\_3)

 rl <- ter\_1-ter\_2+s\_ter\_3

 r\_llf=r\_llf+rl

 }

 return(r\_llf)

}

#Gradient function

llg <- function (r\_beta) {

 parp <- r\_beta[1:p]

 a <- r\_beta[p1]

 b <- r\_beta[p2]

 r\_llg <- matrix(0,1,p2)

 lla <- 0

 llb <- 0

 llp <- matrix(0,1,p)

 nx <- 0

 for (j in 1:ki){

 ter1 <- 0

 ter2 <- 0

 ter3 <- matrix(0,p,1)

 ni <- n\_period[j]

 i\_deb <- nx+1

 i\_fin <- nx+ni

 ri <- i\_deb : i\_fin

 nx <- i\_fin

 yi <- y[ri]

 xi <- x[ri,]

 zi <- xi%\*%parp

 mui <- exp(zi)

 s\_mui <- sum(mui)

 s\_yi <- sum(yi)

 ter1 <- digamma(a+b) +digamma(a+s\_mui) -digamma(a) - digamma(a+b+s\_yi+s\_mui)

 lla = lla +ter1

 ter2 <- digamma(a+b) + digamma(b+s\_yi) - digamma(b) - digamma(a+b+s\_yi+s\_mui)

 llb = llb+ter2

 ter3 <- (t(xi)%\*%mui)\*(digamma(a+s\_mui)) - (t(xi)%\*%mui)\*(digamma(a+b+s\_yi+s\_mui))

 ter3mui <- matrix(0,ni,p)

 for (iii in 1:p) {

 ter3mui[,iii] <- (xi[,iii]\*mui[,1])

 }

 ter3b <- t(ter3mui)%\*%(digamma(mui+yi) - digamma(mui))

 ter3c <- ter3 + ter3b

 llp=llp+t(ter3c)

 }

 r\_llg[1:p]=llp

 r\_llg[p1]=lla

 r\_llg[p2]=llb

 return(r\_llg)

}

#Lower bound of the parameters

Low <- c( "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf",

 "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf",

 "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf",

 "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf", "-inf",

 1e-8, 1e-8)

#Optimization

hausman <- optim(r\_beta, fn=llf, gr=llg, method="L-BFGS-B", lower=Low, control=list(trace=5, fnscale=-1, report=1, maxit=1000), hessian=TRUE)

hausman$convergence # the method converge

#Result

parlistL<-hausman$par

print(parlistL)

#Hessian at the optimal values

hessL<-hausman$hessian

inv\_hessL<- solve(hessL)

#Standard errors

 r\_stder=sqrt(abs(diag(inv\_hessL)))

print(r\_stder)

t\_ratio=parlistL/r\_stder

#P-values

n\_df=n-p2

p\_value=2\*(1-pt(abs(t\_ratio),df=n\_df))

print(p\_value)

# **Poisson fixed effect model estimation**

#

GLMmboot.2 <- glmmboot(NB\_ATOT ~ NB\_INF1 + NB\_INF2 + NB\_INF3 + NB\_INF6 + NB\_INF7 +

NB\_INF89+ VIT + SANCT + ROUGE + ARRET + CEINTURE + N\_VH69 + N\_VH20

+ N\_VH50 + N\_VH51 +an\_91 + an\_92 + an\_93 + an\_94 + an\_95 + an\_96 + an\_97,

family = poisson(log), data=Dataset, cluster=TRNIP,

start.coef = NULL, control = list(epsilon = 1e-08, maxit = 200, trace = FALSE))

summary(GLMmboot.2)

#

#**Gamma-Dirichlet model estimation**#

#Create the vector y: number of accident of truck i at time t

y <- as.matrix(cbind(Dataset$NB\_ATOT))

max\_y <- max(y)

#Create the matrix x : Variables concerning the carriers, vehicles and the drivers (a vehicle may have more than one

#driver

x <- as.matrix(cbind(1, Dataset$AN\_TRANS, Dataset$SECT\_14, Dataset$SECT\_05, Dataset$SECT\_07,

Dataset$SECT\_08, Dataset$N\_VH3, Dataset$N\_VH45, Dataset$N\_VH69, Dataset$N\_VH20, Dataset$N\_VH50, Dataset$N\_VH51, Dataset$DUREE\_AT, Dataset$NB\_INF1, Dataset$NB\_INF2, Dataset$NB\_INF3, Dataset$NB\_INF6, Dataset$NB\_INF7, Dataset$NB\_INF89, Dataset$COMPR, Dataset$TBRGN, Dataset$ESSENCE, Dataset$CARB\_AUT, Dataset$CYL1\_5, Dataset$CYL6\_7, Dataset$ESS\_02, Dataset$ESS\_02P, Dataset$ESS\_03, Dataset$ESS\_04, Dataset$ESS\_05, Dataset$VIT, Dataset$SANCT, Dataset$ROUGE, Dataset$ARRET, Dataset$CEINTURE, Dataset$an\_91, Dataset$an\_92, Dataset$an\_93, Dataset$an\_94, Dataset$an\_95, Dataset$an\_96, Dataset$an\_97))

head(x)

n <- nrow(x) # Total number of observations

p <- ncol(x) # Number of parameters

p1=p+1

p2=p+2

p3=p+3

per\_max <- 8

n\_kappa <- 1

#Create indcam matrix: Equal to 1 if the truck i is present at the year t; 0 otherwise

indc <- as.matrix(cbind(Dataset$ind1, Dataset$ind2, Dataset$ind3, Dataset$ind4, Dataset$ind5, Dataset$ind6,

Dataset$ind7, Dataset$ind8, Dataset$camion))

indc1<-indc[Dataset$camion == 1,]

indcam<-indc1[,1:8]

ncamion <- nrow(indcam) # Total number of trucks

pcam <- ncol(indcam) # Maximum number of observed periods

#Create the vector period: Equal to 1 if the truck i is present at the year 1991 ; ….; Equal to 8 1 if the truck i is present at the year 1998

period <- as.matrix(cbind(Dataset$PERIOD))

nperf <- as.matrix(cbind(Dataset$nper\_f,Dataset$FLOTTE))

nperf1<-nperf[Dataset$FLOTTE == 1,]

nper\_f<-cbind(nperf1[,1]) #Number of periods per firm

nflotte <- nrow(nper\_f) # Total number of firms

taillet<- as.matrix(cbind(Dataset$taille\_t,Dataset$FLOTTE))

taillet1<-taillet[Dataset$FLOTTE == 1,]

tt<-cbind(taillet1[,1]) # Number of trucks per firm

taillec<- as.matrix(cbind(Dataset$taille\_c,Dataset$FLOTTE))

taillec1<-taillec[Dataset$FLOTTE == 1,]

ttc<-cbind(taillec1[,1]) # Number of truck-years per firm

nper<- as.matrix(cbind(Dataset$n\_period,Dataset$camion))

nper1<-nper[Dataset$camion == 1,]

n\_period<-cbind(nper1[,1]) # Number of periods per truck

# We divide the vehicles into two groups (high risk and low risk)

grp <- as.matrix(cbind(Dataset$grp,Dataset$camion))

grp1<-grp[Dataset$camion == 1,]

grpc<-cbind(grp1[,1])

#A vector of starting values

nu<- 2.06

kap<- 12.65

del<- 4.67

parlist<-c(nu,kap,del,betaest)

newparlist<-c(4,8,3,estNB)

#Beginning of the R and C++ interface

library(Rcpp)

library(inline)

foo <- paste(readLines("flotte\_8periode\_1v\_1k\_1d\_ttf\_maxdiff\_chg\_grp.cc"),collapse="\n")

fx <- cxxfunction(signature(),plugin="Rcpp",include=foo)

tclass <- Module("test",getDynLib(fx))

Vraisemblance <- tclass$Vraisemblance

#Initialisation Vraisemblance object

vsemblance <- new(Vraisemblance, ttc, tt, nflotte, p, n, per\_max, y)

vsemblance$init\_x(x)

vsemblance$init\_indcam(ncamion, indcam)

#Optimisation, A quasi-Newton medthod (BFGS)

essaiL<-optim(newparlist,vsemblance$r\_llf,method="BFGS",control=list(trace=5, fnscale=-1, report=1, maxit=1000), hessian=TRUE)

#Optimation results

essaiL$convergence # The method converged

parlistL<-essaiL$par

hessL<-essaiL$hessian

print(parlistL)

inv\_hessL<- solve(hessL) # Invert the Hessian matrix

r\_stder=sqrt(abs(diag(inv\_hessL))) #Standard errors

print(r\_stder)

t\_ratio=parlistL/r\_stder

n\_df=n-p3

p\_value=2\*(1-pt(abs(t\_ratio),df=n\_df))

print(p\_value)

## #C++ files : flotte\_8periode\_1v\_1k\_1d\_ttf\_maxdiff\_chg\_grp.cc

#include <Rcpp.h>

using namespace Rcpp;

#include <iostream>

#include <iomanip>

#include <fstream>

#include <vector>

#include <numeric>

#include <algorithm>

#include <math.h>

#include <values.h>

#include <unistd.h>

#include <stdlib.h>

#include <sys/types.h>

#include <sys/stat.h>

#include <fcntl.h>

#include <strings.h>

#include <gsl/gsl\_sf\_hyperg.h>

#include <gsl/gsl\_sf\_psi.h>

#include <gsl/gsl\_sort\_double.h>

#include <gsl/gsl\_statistics.h>

#include <gsl/gsl\_errno.h>

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

To compile the program we must include this file:

~/.R/Makevars ceci

PKG\_LIBS=-L/home/apps/Logiciels/GSL/1.16/lib -lgsl -lgslcblas $(shell "/home/apps/Logiciels/R/3.2.1-gcc/bin/Rscript" -e "Rcpp:::LdFlags()")

It is also necessary to load these modules

 1) R/3.2.1-gcc 2) GSL/1.16

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

#ifdef \_OPENMP

#include <omp.h>

#endif

typedef std::vector<double> My\_Vector;

class Vraisemblance

{

public:

 Vraisemblance(My\_Vector ttc\_, My\_Vector tt\_, double nflotte\_, double p\_, double n\_, double per\_max\_, My\_Vector y\_);

 void init\_x(NumericVector x\_);

 void init\_indcam(int ncamion, NumericVector indcam\_);

 double r\_llf(My\_Vector& r\_beta);

 double r\_llf\_autrestermes(double delta);

 void chg\_grp\_moyenne(My\_Vector& r\_beta);

 void chg\_grp\_mediane(My\_Vector& r\_beta);

 void chg\_grp\_maxdiff(My\_Vector& r\_beta);

private:

 //variables from R

 int \*ttc; //number of truck per firm

 int \*tt; //number of truck-years per firm

 int p; //number of parameters

 int n; // Total number of observations

 int nb\_alpha; //number of kappa (1/alpha)

 int per\_max; // maximum number of observed periods

 double \*y;

 double \*\*x; //truck characteristics

 double\*\* indcam; //indicate the period (year) the truck is present

 double \*d;

 int nflotte; // total number of firms

 int \*grp; / It indicate in which group the truck is

 //Variables internes

 int \*vect\_nx;

 int \*vect\_nxc;

 int \*nb\_camion\_par\_annee;

 double \*mui; // mui = exp ( x \* beta)

 double \*zi; // for the intermediate calculation of mui;

 double \*s\_yi; // sum of yi for each fleet

 double \*s\_vij; // sum of vij for each fleet

};

Vraisemblance::Vraisemblance( My\_Vector ttc\_, My\_Vector tt\_, double nflotte\_, double p\_, double n\_, double per\_max\_, My\_Vector y\_ ) :

 nflotte(static\_cast<int> (nflotte\_)),

 p(static\_cast<int> (p\_)),

 n(static\_cast<int> (n\_)),

 nb\_alpha(1),

 per\_max(static\_cast<int> (per\_max\_))

{

 gsl\_set\_error\_handler\_off ();

 tt = new int[nflotte];

 ttc = new int[nflotte];

 std::cout << "nflotte " << nflotte << '\n';

 std::cout << "p " << p << '\n';

 std::cout << "n " << n << '\n';

 std::cout << "permax " << per\_max << '\n';

 for (int i=0; i<nflotte; i++)

 {

 tt[i] = static\_cast<int>(tt\_[i]);

 ttc[i] = static\_cast<int>(ttc\_[i]);

 }

 int y\_size = y\_.size();

 y = new double[y\_size];

 for (int i=0; i<y\_size; i++)

 y[i]=y\_[i];

 vect\_nx = new int[nflotte];

 vect\_nx[0] = 0;

 for (int i=1; i<nflotte; i++)

 vect\_nx[i] = vect\_nx[i-1] + tt[i-1];

 vect\_nxc = new int[nflotte];

 vect\_nxc[0] = 0;

 for (int i=1; i<nflotte; i++)

 vect\_nxc[i] = vect\_nxc[i-1] + static\_cast<int>(ttc[i-1]);

 mui = new double[n];

 zi = new double[n];

 s\_yi = new double[nflotte];

 s\_vij = new double[nflotte];

 grp = new int[n];

 for (int f=0; f<nflotte; f++)

 {

 int tt\_f = tt[f];

 int nx\_t=vect\_nx[f];

 s\_yi[f] = std::accumulate(y+nx\_t, y+nx\_t+tt\_f, 0.0);

 }

 nb\_camion\_par\_annee = new int[nflotte];

 for (int i=0; i<nflotte; i++)

 {

 int tt\_f =static\_cast<int>(tt[i]);

 int nx\_t = vect\_nx[i];

 nb\_camion\_par\_annee[i] = tt\_f;

 if (tt\_f==2)

 {

 grp[nx\_t] = 0;

 grp[nx\_t+1] = 1;

 }

 }

}

//conversion of NumericVector in the matrix x

void Vraisemblance::init\_x(NumericVector x\_)

{

 x = new double\*[n];

 for (int i=0; i<n; i++)

 x[i] = new double[p];

 int k=0;

 for (int j=0; j<p; j++)

 for (int i=0; i<n; i++)

 {

 x[i][j] = x\_[k];

 k++;

 }

}

//conversion of NumericVector in the matrix indam

void Vraisemblance::init\_indcam(int ncamion, NumericVector indcam\_)

{

 indcam = new double\*[ncamion];

 for (int i=0; i<ncamion; i++)

 indcam[i] = new double[per\_max];

 int k=0;

 for (int j=0; j<per\_max; j++)

 for (int i=0; i<ncamion; i++)

 {

 indcam[i][j] = indcam\_[k];

 k++;

 }

}

//Function to evaluate

double Vraisemblance::r\_llf(My\_Vector& r\_beta)

{

 double \*v = &r\_beta[0];

 double \*alpha = &r\_beta[1];

 double delta = r\_beta[2];

 int debut\_beta = 1 + nb\_alpha + 1;

 double \*beta = &r\_beta[debut\_beta];

 /\* reading the observations of carrier f \*/

 double r\_ll\_f1=0;

 double r\_ll\_f2=0;

 double r\_ll\_f3=0;

 double r\_ll\_f=0;

 //int size\_x\_col = x[0].size();

 //My\_Vector zi(ni);

 //zi=xi\*par`;

 //zi += x\_ptr[j] \* beta[j];

 for (int f=0; f<nflotte; f++)

 {

 int ni = static\_cast<int>(tt[f]);

 int nx = vect\_nx[f];

 for (int i=0; i<ni; i++)

 {

 zi[nx+i]=0.0;

 for (int k=0; k<p; k++)

 zi[nx+i] += x[nx+i][k] \* beta[k];

 }

 }

#pragma omp parallel for

 //My\_Vector mui(ni);

 //=(di#exp(zi));

 for (int i=0; i<n; i++)

 mui[i] = exp (zi[i]);

#pragma omp parallel for reduction(+:r\_ll\_f1, r\_ll\_f2, r\_ll\_f3) schedule (static, 10)

 for (int f=0; f<nflotte; f++)

 {

 int ttc\_f=static\_cast<int>(ttc[f]);

 int tt\_f =static\_cast<int>(tt[f]);

 int nx\_c = vect\_nxc[f];

 int nx\_t = vect\_nx[f];

 //there is only one alpha

 int ind\_kappa = 0;

 double min\_mui= \*std::min\_element(mui+nx\_t, mui+nx\_t+tt\_f);//min(mui);

 double max\_mui= \*std::max\_element(mui+nx\_t, mui+nx\_t+tt\_f);//max(mui);

 double s\_mui = std::accumulate(mui+nx\_t, mui+nx\_t+tt\_f, 0.0);//sum(mui);

 s\_vij[f]=0;

 double lgamma\_v = 0;

 double present\_v;

 for (int i=0; i<ttc\_f; i++)

 {

 s\_vij[f] += v[0];

 lgamma\_v += lgamma(v[0]);

 }

 double s\_y\_mui = 0;

 for (int i=0; i<tt\_f; i++)

 s\_y\_mui += y[nx\_t+i]\*log(mui[nx\_t+i]);

 {

 double g1=0;

 double g2=0;

 double s\_yi1=0;

 double s\_yi2=0;

 double s\_vi1=0;

 double s\_vi2=0;

 double s\_mui1=0;

 double s\_mui2=0;

 int ind = 0;

 for (int i=0; i<ttc\_f; i++)

 {

 double somme\_y=0;

 double somme\_v=v[0];

 double somme\_mui=0;

 double nb = 0;

 int groupe;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 groupe = grp[nx\_t+ind];

 somme\_y += y[nx\_t+ind];

 somme\_mui += mui[nx\_t+ind];

 nb++;

 ind++;

 }

 if (groupe == 0)

 {

 g1++;

 s\_yi1 += somme\_y;

 s\_vi1 += somme\_v;

 s\_mui1 += somme\_mui/nb;

 }

 else

 {

 g2++;

 s\_yi2 += somme\_y;

 s\_vi2 += somme\_v;

 s\_mui2 += somme\_mui/nb;

 }

 }

 double mui1, mui2;

 if (g1 > 0)

 mui1=s\_mui1/g1;

 else

 mui1=s\_mui2/g2;

 if (g2 > 0)

 mui2=s\_mui2/g2;

 else

 mui2=mui1;

 double par1, par2;

 double ter\_11;

 double par3 = s\_yi[f] + s\_vij[f];

 double par4 = s\_yi[f]+nb\_camion\_par\_annee[f]/alpha[ind\_kappa];

 if (mui1 <= mui2)

 {

 if (g1>0)

 par2=s\_yi1+s\_vi1;

 else

 par2=s\_yi2+s\_vi2;

 par1=((mui2-mui1)/(1/alpha[ind\_kappa]+mui2));

 ter\_11=par4\*log(1+alpha[ind\_kappa]\*mui2);

 }

 else if (mui1 > mui2)

 {

 par2=s\_yi2+s\_vi2;

 par1=((mui1-mui2)/(1/alpha[ind\_kappa]+mui1));

 ter\_11=par4\*log(1+alpha[ind\_kappa]\*mui1);

 }

 double f\_hyp1=gsl\_sf\_hyperg\_2F1(par2,par4,par3,par1);

 double ter\_1=

 s\_yi[f]\*log(alpha[ind\_kappa])

 +lgamma(par4)

 -lgamma(nb\_camion\_par\_annee[f]/alpha[ind\_kappa])

 +lgamma(s\_vij[f])

 +s\_y\_mui

 -lgamma(par3)

 -ter\_11

 +log(f\_hyp1);

 double s\_ter\_3= -lgamma\_v;

 for (int i=0; i<tt\_f; i++)

 s\_ter\_3 -= lgamma(y[nx\_t+i]+1);

 ind = 0;

 for (int i=0; i<ttc\_f; i++)

 {

 double syv = v[0];

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 syv += y[nx\_t+ind];

 ind++;

 }

 s\_ter\_3 += lgamma(syv);

 }

 double rl3 = ter\_1 + s\_ter\_3;

 r\_ll\_f3 += rl3;

 }

 }

 r\_ll\_f = r\_ll\_f1 + r\_ll\_f2 + r\_ll\_f3 + r\_llf\_autrestermes(delta);

 if (!finite(r\_ll\_f))

 r\_ll\_f=-1000000000000.5;

 return (r\_ll\_f);

 //end r\_llf

}

double Vraisemblance::r\_llf\_autrestermes(double delta)

{

 double r\_ll\_f\_temps=0;

#pragma omp parallel for reduction(+:r\_ll\_f\_temps) schedule (static, 10)

 for (int f=0; f<nflotte; f++)

 {

 int ttc\_f=static\_cast<int>(ttc[f]);

 int tt\_f =static\_cast<int>(tt[f]);

 int nx\_c = vect\_nxc[f];

 int nx\_t = vect\_nx[f];

 double lgamma\_d = 0;

 for (int i=0; i<ttc\_f; i++)

 {

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 lgamma\_d += lgamma(delta);

 }

 int ind = 0;

 double lgamma\_sd=0;

 double lgamma\_sy\_plus\_sd=0;

 for (int i=0; i<ttc\_f; i++)

 {

 double somme\_delta = 0;

 double somme\_y = 0;

 for (int j=0; j<per\_max; j++)

 {

 if (indcam[nx\_c+i][j]>0)

 {

 somme\_delta += delta;

 somme\_y += y[nx\_t+ind];

 ind++;

 }

 }

 lgamma\_sd += lgamma(somme\_delta);

 lgamma\_sy\_plus\_sd += lgamma(somme\_y + somme\_delta);

 }

 ind = 0;

 double lgamma\_y\_plus\_d = 0;

 for (int i=0; i<ttc\_f; i++)

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 lgamma\_y\_plus\_d += lgamma(y[nx\_t+ind] + delta);

 ind++;

 }

 r\_ll\_f\_temps += (lgamma\_y\_plus\_d + lgamma\_sd - lgamma\_d - lgamma\_sy\_plus\_sd);

 }

 return (r\_ll\_f\_temps);

}

void Vraisemblance::chg\_grp\_moyenne(My\_Vector& r\_beta)

{

 double \*v = &r\_beta[0];

 double \*alpha = &r\_beta[1];

 double delta = r\_beta[2];

 int debut\_beta = 1 + nb\_alpha + 1;

 double \*beta = &r\_beta[debut\_beta];

 //My\_Vector zi(ni);

 //zi=xi\*par`;

 //zi += x\_ptr[j] \* beta[j];

 for (int f=0; f<nflotte; f++)

 {

 int ni = static\_cast<int>(tt[f]);

 int nx = vect\_nx[f];

 for (int i=0; i<ni; i++)

 {

 zi[nx+i]=0.0;

 for (int k=0; k<p; k++)

 zi[nx+i] += x[nx+i][k] \* beta[k];

 }

 }

#pragma omp parallel for

 //My\_Vector mui(ni);

 //=(di#exp(zi));

 for (int i=0; i<n; i++)

 mui[i] = exp (zi[i]);

 //Compute the mean of mui

 for (int f=0; f<nflotte; f++)

 {

 double somme\_mui = 0;

 int ttc\_f=static\_cast<int>(ttc[f]);

 int tt\_f =static\_cast<int>(tt[f]);

 int nx\_c = vect\_nxc[f];

 int nx\_t = vect\_nx[f];

 for (int cam=0; cam<tt\_f; cam++)

 somme\_mui += mui[nx\_t+cam];

 double moyenne\_mui = somme\_mui / tt\_f;

 //Sum of mui

 int ind1 = 0;

 int ind2 = 0;

 for (int i=0; i<ttc\_f; i++)

 {

 somme\_mui = 0.0;

 int nb\_annee = 0;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 somme\_mui += mui[nx\_t+ind1];

 nb\_annee++;

 ind1++;

 }

 double moyenne\_mui\_cam = somme\_mui / nb\_annee;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 if (moyenne\_mui\_cam <= moyenne\_mui)

 grp[nx\_t+ind2] = 0;

 else

 grp[nx\_t+ind2] = 1;

 ind2++;

 }

 }

 }

}

 //end chg\_grp

void Vraisemblance::chg\_grp\_mediane(My\_Vector& r\_beta)

{

 double \*v = &r\_beta[0];

 double \*alpha = &r\_beta[1];

 double delta = r\_beta[2];

 int debut\_beta = 1 + nb\_alpha + 1;

 double \*beta = &r\_beta[debut\_beta];

 //My\_Vector zi(ni);

 //zi=xi\*par`;

 //zi += x\_ptr[j] \* beta[j];

 for (int f=0; f<nflotte; f++)

 {

 int ni = static\_cast<int>(tt[f]);

 int nx = vect\_nx[f];

 for (int i=0; i<ni; i++)

 {

 zi[nx+i]=0.0;

 for (int k=0; k<p; k++)

 zi[nx+i] += x[nx+i][k] \* beta[k];

 }

 }

#pragma omp parallel for

 //My\_Vector mui(ni);

 //=(di#exp(zi));

 for (int i=0; i<n; i++)

 mui[i] = exp (zi[i]);

 //Compute the mean of mu

 for (int f=0; f<nflotte; f++)

 {

 int ttc\_f=static\_cast<int>(ttc[f]);

 int tt\_f =static\_cast<int>(tt[f]);

 int nx\_c = vect\_nxc[f];

 int nx\_t = vect\_nx[f];

 int ind1 = 0;

 std::vector<double> mediane;

 for (int i=0; i<ttc\_f; i++)

 {

 double somme\_mui = 0.0;

 int nb\_annee=0;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 somme\_mui += mui[nx\_t+ind1];

 nb\_annee++;

 ind1++;

 }

 mediane.push\_back(somme\_mui/nb\_annee);

 }

 std::sort(mediane.begin(), mediane.end());

 double val\_mediane = mediane[mediane.size()/2];

 //calcul des somme des mui sur les camions

 ind1=0;

 int ind2 = 0;

 for (int i=0; i<ttc\_f; i++)

 {

 double somme\_mui = 0.0;

 int nb\_annee=0;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 somme\_mui += mui[nx\_t+ind1];

 ind1++;

 }

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 if (somme\_mui/nb\_annee < val\_mediane)

 grp[nx\_t+ind2] = 0;

 else

 grp[nx\_t+ind2] = 1;

 ind2++;

 }

 }

 }

}

template <class random\_iterator>

class IndexedComparison

{

public:

 IndexedComparison (random\_iterator begin,

 random\_iterator end)

 : p\_begin (begin), p\_end (end) {}

 bool operator () (unsigned int a, unsigned int b) const

 { return \*(p\_begin + a) < \*(p\_begin + b); }

private:

 random\_iterator const p\_begin;

 random\_iterator const p\_end;

};

void Vraisemblance::chg\_grp\_maxdiff(My\_Vector& r\_beta)

{

 double \*v = &r\_beta[0];

 double \*alpha = &r\_beta[1];

 double delta = r\_beta[2];

 int debut\_beta = 1 + nb\_alpha + 1;

 double \*beta = &r\_beta[debut\_beta];

 //My\_Vector zi(ni);

 //zi=xi\*par`;

 //zi += x\_ptr[j] \* beta[j];

 for (int f=0; f<nflotte; f++)

 {

 int ni = static\_cast<int>(tt[f]);

 int nx = vect\_nx[f];

 for (int i=0; i<ni; i++)

 {

 zi[nx+i]=0.0;

 for (int k=0; k<p; k++)

 zi[nx+i] += x[nx+i][k] \* beta[k];

 }

 }

#pragma omp parallel for

 //My\_Vector mui(ni);

 //=(di#exp(zi));

 for (int i=0; i<n; i++)

 mui[i] = exp (zi[i]);

 //Compute the mean of mu

 for (int f=0; f<nflotte; f++)

 {

 int ttc\_f=static\_cast<int>(ttc[f]);

 int tt\_f =static\_cast<int>(tt[f]);

 int nx\_c = vect\_nxc[f];

 int nx\_t = vect\_nx[f];

 int ind1 = 0;

 std::vector<double> moymui;

 for (int i=0; i<ttc\_f; i++)

 {

 double somme\_mui = 0.0;

 int nb\_annee=0;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 somme\_mui += mui[nx\_t+ind1];

 nb\_annee++;

 ind1++;

 }

 moymui.push\_back(somme\_mui/nb\_annee);

 }

 std::vector<unsigned int> indices(ttc\_f);

 for (int i = 0; i < indices.size (); i++)

 indices [i] = i;

 std::sort (indices.begin (), indices.end (),

 IndexedComparison<std::vector<double>::const\_iterator>

 (moymui.begin(), moymui.end()));

 double val\_mui = 0;

 double maxdiff = -1e300;

 for (int i=0; i<ttc\_f-1; i++)

 {

 double diff = moymui[indices[i+1]] - moymui[indices[i]];

 if (diff > maxdiff)

 {

 val\_mui = moymui[indices[i+1]];

 maxdiff = diff;

 }

 }

 float nbCamTot=0;

 float nbCamGrp1=0;

 //Compute the sum of mui on trucks

 ind1=0;

 int ind2 = 0;

 for (int i=0; i<ttc\_f; i++)

 {

 double somme\_mui = 0.0;

 int nb\_annee=0;

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 somme\_mui += mui[nx\_t+ind1];

 nb\_annee++;

 ind1++;

 }

 for (int j=0; j<per\_max; j++)

 if (indcam[nx\_c+i][j]>0)

 {

 nbCamTot += 1.0;

 if (somme\_mui/nb\_annee < val\_mui)

 {

 grp[nx\_t+ind2] = 0;

 nbCamGrp1 += 1.0;

 }

 else

 grp[nx\_t+ind2] = 1;

 ind2++;

 }

 }

 }

}

RCPP\_MODULE(test){

 class\_<Vraisemblance>( "Vraisemblance" )

 //.constructor()

 .constructor< My\_Vector, My\_Vector, double, double, double, double, My\_Vector >()

 .method( "init\_x", &Vraisemblance::init\_x)

 .method( "init\_indcam", &Vraisemblance::init\_indcam)

 .method( "r\_llf", &Vraisemblance::r\_llf )

 .method( "chg\_grp\_moyenne", &Vraisemblance::chg\_grp\_moyenne )

 .method( "chg\_grp\_mediane", &Vraisemblance::chg\_grp\_mediane )

 .method( "chg\_grp\_maxdiff", &Vraisemblance::chg\_grp\_maxdiff )

 ;

}

**Appendix E: Predictive probabilities**

Table E1 presents an example of predictive probabilities calculated for a fleet *f* of three vehicles in 1998. In the estimating sample, the same fleet *f* had five vehicles, so in equation (19),  and . The estimated values of the random effects parameters are equal to , , and  (see Table C1.3). Suppose that fleet *f* will have no accident at time t+1, then  which means that the three vehicles of the fleet *f* in the forecasting sample will have no accident. In applying the formula in (19) with the estimated parameters, the predictive probability of fleet *f* to have no accident at t+1 is equal to 80.2% (1×1.2656×0.8979×0.7839×1×1×1×0.875×1.0290)×100. The calculations are given below.













If now we suppose that fleet *f* will have 1 accident at t+1, then  in table E1. Since fleet *f* has three vehicles, there are 3 possibilities for the fleet to accumulate 1 accident. In applying the formula in (19) we obtain that the predictive probability of fleet *f* to have 1 accident is equal to 12.7% (Table E1). There are 6 possibilities for the fleet *f* of three vehicles to accumulate 2 accidents at t+1 (Table E1). The predictive probability that the fleet *f* will have 2 accidents during the next year is then 1.5%. And so on…

We observe that the predictive probabilities in Table E1 differ from the average predictive probabilities in Table 9 for fleets of 5 trucks. The fleet in this example represents a lower risk than the average fleet of this size: during the last 7 years, the fleet had only one accident by assumption meaning that the implied accident rate for a truck in this fleet is 3% while the mean is 13% for this size of fleet.

 Table E1: Example of predictive probabilities calculated for a fleet *f* of three trucks.

|  |  |  |
| --- | --- | --- |
| Fleet *f* | Estimating sample | Forecasting sample |
|  |  |  |  |
|  | 1 possibility | 3 possibilities | 6 possibilities |
| Truck *i* | Si | Group | Ti |  |  |  |  |  |  |  |  |  |  |  |
| 1 | 0 | 2 | 7 | 0.1304 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 1 | 1 | 0 |
| 2 | 1 | 2 | 7 | 0.0960 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 1 |
| 3 | 0 | 1 | 2 | 0.0633 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| 4 | 0 | 1 | 4 | . | . | . | . | . | . | . | . | . | . | . |
| 5 | 0 | 2 | 2 | . | . | . | . | . | . | . | . | . | . | . |
|  | S0 = 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Predictive probabilities |  |  | 3.6% +3.9%+5.2% | 0.17%+0.17%+0.35%+0.21%+0.29%+0.31% |
|  | 80.2% |  | 12.7% |  |  |  |  | 1.5% |  |  |

  and  are the means of group 1 and group 2 respectively.

  is calculated from the forecasting sample with the estimated coefficients presented in Table C1.3.

**Appendix F: Boostrap replications, R-code**

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Cameron and Trivedi (2013b) propose the following representation of the Hausman test:



where  is the Hausman test statistic,  are the estimated parameters obtained from the fixed effects model and  are the estimated parameters obtained from the random effects model (Gamma-Dirichlet model). To estimate the variance term  we use a panel bootstrap method that resamples over the 5,423 firms of the sample:



where  and  are the estimates obtained from the *bth* bootstrap replication

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**Estimation of the **

library(foreign, pos=15)

#

#Read the data set

Dataset <- read.table("donneePlusDe4.csv", header=TRUE, sep=",",na.strings="NA", dec=".", strip.white=TRUE)

#Read B random samples with replacement, the firm can be selected more than once.

SampleURS data set include Three variables

 1. Replication number

2. Firm identification

3. Number of hit refers to the number of times a firm is selected

SampleURS <- read.table("SampleUSR.csv", header=TRUE, sep=",",na.strings="NA", dec=".", strip.white=TRUE)

library(abind, pos=16)

library(e1071, pos=17)

library("BMS")

library("spuRs")

library("MASS")

library(Rcpp)

library(inline)

library(splitstackshape)

#Write the 1,000 coefficient estimations in OURRandom file

**outputfile = "OUTRandom"**

cat("replicate", "nu", "kap", "del", "Intercep", "NB\_INF1", "NB\_INF2", "NB\_INF3", "NB\_INF6", "NB\_INF7",

 "NB\_INF89", "VIT", "SANCT", "ROUGE", "ARRET", "CEINTURE", "N\_VH69", "N\_VH20",

"N\_VH50", "N\_VH51", "an\_91", "an\_92", "an\_93", "an\_94", "an\_95", "an\_96", "an\_97", sep="\t",file=outputfile, append=T)

cat("\n",file=outputfile , append=T)

B = 1000

for (repl in 1:B){

sample <- SampleURS[SampleURS$Replicate==iter,]

sample\_1 <- merge (Dataset, sample, by=c("TRNIP"))

sample\_2 <- expandRows(sample\_1, "NumberHits", drop=FALSE)

sample\_2$frac <- as.numeric(row.names(sample\_2))

sample\_2$integ <- trunc(sample\_2$frac)

sample\_2$Hits <- round((sample\_2$frac - sample\_2$integ)\*10,digits=1)

sample\_2$TRNIP\_1 <- as.character(paste(sample\_2$TRNIP, sample\_2$Hits, sep="" ))

attach(sample\_2)

sort.sample\_2 <- sample\_2[order(TRNIP\_1,VEH\_1,AN),]

detach(sample\_2

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Estimate  with sort.sample\_2 dataset

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

parlistL<-essaiL$par

cat(i, parlistL,sep="\t",file=outputfile, append=T)

cat("\n",file=outputfile , append=T)

}

#Write the 1,000 coefficient estimations in OUTFixed file

**outputfile = "OUTFixed"**

cat("replicate ", "NB\_INF1", "NB\_INF2", "NB\_INF3", "NB\_INF6", "NB\_INF7",

 "NB\_INF89", "VIT", "SANCT", "ROUGE", "ARRET", "CEINTURE", "N\_VH69", "N\_VH20",

"N\_VH50", "N\_VH51", "an\_91", "an\_92", "an\_93", "an\_94", "an\_95", "an\_96", "an\_97", sep="\t",file=outputfile, append=T)

cat("\n",file=outputfile , append=T)

B = 1000

for (repl in 1:B){

sample <- SampleURS[SampleURS$Replicate==iter,]

sample\_1 <- merge (Dataset, sample, by=c("TRNIP"))

attach(sample\_1)

sort.sample\_1 <- sample\_1[order(TRNIP\_1,VEH\_1,AN),]

detach(sample\_1

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Estimate  with sort.sample\_1 dataset

If the same firm *f* appears twice in a bootstrap resample *iter* then  needs to treat the fixed effect as being the same for both observations *f.*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

#Poisson model to estimate the initial values of the parameters

GLM <- glm(NB\_ATOT ~ N\_VH69 + N\_VH20 + N\_VH50 + N\_VH51 + VIT + SANCT + ROUGE + ARRET + CEINTURE + an\_91 +

 an\_92 + an\_93 + an\_94 + an\_95 + an\_96 + an\_97 + NB\_INF1 + NB\_INF2 + NB\_INF3 + NB\_INF6 +

 NB\_INF7 + NB\_INF89, family=poisson(log), data=sort.sample\_1)

param=GLM$coefficients

x <- as.matrix(cbind(sort.sample\_1$N\_VH69, sort.sample\_1$N\_VH20, sort.sample\_1$N\_VH50,

 sort.sample\_1$N\_VH51,

 sort.sample\_1$VIT, sort.sample\_1$SANCT, sort.sample\_1$ROUGE, sort.sample\_1$ARRET,

 sort.sample\_1$CEINTURE,

 sort.sample\_1$an\_91, sort.sample\_1$an\_92, sort.sample\_1$an\_93, sort.sample\_1$an\_94,

 sort.sample\_1$an\_95, sort.sample\_1$an\_96, sort.sample\_1$an\_97,

 sort.sample\_1$NB\_INF1,sort.sample\_1$NB\_INF2, sort.sample\_1$NB\_INF3,

 sort.sample\_1$NB\_INF6, sort.sample\_1$NB\_INF7, sort.sample\_1$NB\_INF89))

y= as.matrix(cbind(sort.sample\_1$NB\_ATOT))

max\_y=max(y)

parm=c('n\_vh69', 'n\_vh20', 'n\_vh50', 'n\_vh51', 'vit', 'sanct', 'rouge', 'arret', 'ceinture', 'an\_91', 'an\_92', 'an\_93',

 'an\_94', 'an\_95', 'an\_96', 'an\_97', 'nb\_inf1', 'nb\_inf2', 'nb\_inf3', 'nb\_inf6', 'nb\_inf7', 'nb\_inf89' )

p=ncol(x)

n=nrow(x)

p1=p+1

#t is number of periods per truck,

nper=as.matrix(cbind(sort.sample\_1$n\_period, sort.sample\_1$camion))

nper1=nper[sort.sample\_1$camion==1,]

t=cbind(nper1[,1])

#fl is the number of trucks per fleet,

taillec=as.matrix(cbind(sort.sample\_1$taille\_c, sort.sample\_1$FLOTTE))

taillec1=taillec[sort.sample\_1$FLOTTE==1,]

fl=cbind(taillec1[,1])

#fl\_an is the number of year-trucks per fleet.

taillet=as.matrix(cbind(sort.sample\_1$taille\_t, sort.sample\_1$FLOTTE))

taillet1=taillet[sort.sample\_1$FLOTTE==1,]

fl\_an=cbind(taillet1[,1])

# Number of time the firm had been selected

Hits<- as.matrix(cbind(sort.sample\_1$NumberHits,sort.sample\_1$FLOTTE))

 Hits1<-Hits[sort.sample\_1$FLOTTE == 1,]

NHits<-cbind(Hits1[,1])

#ki is total number of trucks

ki=nrow(t)

# kf is total number of fleets.

kf=nrow(fl)

#Total number of parameters including firm fixed effects.

n\_parm=p+kf

#Initial values of the parameters.

ini\_fix=matrix(0,1,kf)

ini\_p=param[2:p1]

r\_beta= c(ini\_p, ini\_fix)

eps=1e-8

diff=1

x\_sol=matrix(0,nrow=1, ncol=n\_parm)

for(iter in 1:100){

 par=r\_beta[1:p]

 r\_ll\_ff=0

 r\_ll\_gf=matrix(0, nrow=1, ncol=p)

 r\_ll\_hf=matrix(0, nrow=p, ncol=p)

 ster\_the=matrix(0, nrow=kf, ncol=1)

 xpar=matrix(0, nrow=kf, ncol=p)

 nx=0

 for(f in 1:kf){

 ind\_f=f+p

 thetaf=r\_beta[ind\_f]

 #Number of trucks in the fleet f

 nf=fl[f]

 #Number of year-trucks in the fleet f

 snf=fl\_an[f]

 #rf is the vector of subscripts for each year-trucks in the fleet f

 i\_deb=nx+1

 i\_fin=nx+snf

 rf=i\_deb:i\_fin

 #nx brings us to the next fleet

 nx=i\_fin

 yf=y[rf]

 NHitsf <- NHits[f]

 xf=x[rf,]

 zf=xf%\*%par

 muf=exp(zf+thetaf)

 s\_muf=sum(muf)

 # log-likelihood (for the fleet f).

 ter\_t=-muf+yf\*(thetaf+zf)- lgamma(yf+1)

 #r\_ll\_ff is the sum of all fleets

 ster\_t= NHitsf \*sum(ter\_t)

 r\_ll\_ff=r\_ll\_ff+ster\_t

 ter\_1f=matrix(0,snf,p)

 for(i in 1:p){

 ter\_1f[,i]=xf[,i]\*muf[,1]

 }

 hgf= colSums(ter\_1f)

 xparf=hgf/s\_muf

 xpar[f,]=xparf

 ter\_2f=matrix(0, nrow=snf, ncol=p)

 for(ii in 1:snf){

 ter\_2f[ii,]=xf[ii,]-xparf

 }

 ter\_3=t(ter\_2f)%\*%(yf-muf)

 llpf=t(ter\_3)

 r\_ll\_gf=r\_ll\_gf+NHitsf\*llpf

 ter2muf=matrix(0,snf,p)

 for(iii in 1:p){

 ter2muf[,iii]=ter\_2f[,iii]\*muf[,1]

 }

 ter\_p=-t(ter\_2f)%\*%(ter2muf)

 r\_ll\_hf=r\_ll\_hf+NHitsf\*ter\_p

 ter\_the\_f=(yf-muf)

 ster\_the\_f=sum(ter\_the\_f)

 ster\_the[f,]=ster\_the\_f/s\_muf

 }

#end for (f in 1: kf)

 inv\_r\_ll\_hf=solve(r\_ll\_hf)

 delta=inv\_r\_ll\_hf%\*%t(r\_ll\_gf)

 delta\_theta=-ster\_the+xpar%\*%delta

 #We iterate on the solution to make it converge to the final estimators.

 x\_sol[1:p]=t(delta)

 x\_sol[p1:n\_parm]=t(delta\_theta)

 r\_beta=r\_beta-x\_sol

 diff1=which.max(abs(x\_sol))

 diff=max(abs(x\_sol))

 if(diff<=eps) {

 break }

}

#end for(iter in 1:100)

# Optimation results

beta\_sol=r\_beta[1:p]

parlistL<- beta\_sol

cat(i, parlistL,sep="\t",file=outputfile, append=T)

cat("\n",file=outputfile , append=T)

}

# end for (repl in 1:B)

**Hausman test R-code**

random=read.csv(OUTRandom.csv", header=TRUE, sep=";")

fixed=read.csv("OUTFixed", header=TRUE, sep="\t")

N\_VH69\_r=random$N\_VH69

N\_VH69\_f=fixed$N\_VH69

N\_VH20\_r=random$N\_VH20

N\_VH20\_f=fixed$N\_VH20

N\_VH50\_r=random$N\_VH50

N\_VH50\_f=fixed$N\_VH50

N\_VH51\_r=random$N\_VH51

N\_VH51\_f=fixed$N\_VH51

VIT\_r=random$VIT

VIT\_f=fixed$VIT

SANCT\_r=random$SANCT

SANCT\_f=fixed$SANCT

ROUGE\_r=random$ROUGE

ROUGE\_f=fixed$ROUGE

ARRET\_r=random$ARRET

ARRET\_f=fixed$ARRET

CEINTURE\_r=random$CEINTURE

CEINTURE\_f=fixed$CEINTURE

an\_91\_r=random$an\_91

an\_91\_f=fixed$an\_91

an\_92\_r=random$an\_92

an\_92\_f=fixed$an\_92

an\_93\_r=random$an\_93

an\_93\_f=fixed$an\_93

an\_94\_r=random$an\_94

an\_94\_f=fixed$an\_94

an\_95\_r=random$an\_95

an\_95\_f=fixed$an\_95

an\_96\_r=random$an\_96

an\_96\_f=fixed$an\_96

an\_97\_r=random$an\_97

an\_97\_f=fixed$an\_97

NB\_INF1\_r=random$NB\_INF1

NB\_INF1\_f=fixed$NB\_INF1

NB\_INF2\_r=random$NB\_INF2

NB\_INF2\_f=fixed$NB\_INF2

NB\_INF3\_r=random$NB\_INF3

NB\_INF3\_f=fixed$NB\_INF3

NB\_INF6\_r=random$NB\_INF6

NB\_INF6\_f=fixed$NB\_INF6

NB\_INF7\_r=random$NB\_INF7

NB\_INF7\_f=fixed$NB\_INF7

NB\_INF89\_r=random$NB\_INF89

NB\_INF89\_f=fixed$NB\_INF89

n\_vh69=N\_VH69\_f-N\_VH69\_r

n\_vh20=N\_VH20\_f-N\_VH20\_r

n\_vh50=N\_VH50\_f-N\_VH50\_r

n\_vh51=N\_VH51\_f-N\_VH51\_r

vit=VIT\_f-VIT\_r

sanct=SANCT\_f-SANCT\_r

rouge=ROUGE\_f-ROUGE\_r

arret=ARRET\_f-ARRET\_r

ceinture=CEINTURE\_f-CEINTURE\_r

an91=an\_91\_f-an\_91\_r

an92=an\_92\_f-an\_92\_r

an93=an\_93\_f-an\_93\_r

an94=an\_94\_f-an\_94\_r

an95=an\_95\_f-an\_95\_r

an96=an\_96\_f-an\_96\_r

an97=an\_97\_f-an\_97\_r

nb\_inf1=NB\_INF1\_f-NB\_INF1\_r

nb\_inf2=NB\_INF2\_f-NB\_INF2\_r

nb\_inf3=NB\_INF3\_f-NB\_INF3\_r

nb\_inf6=NB\_INF6\_f-NB\_INF6\_r

nb\_inf7=NB\_INF7\_f-NB\_INF7\_r

nb\_inf89=NB\_INF89\_f-NB\_INF89\_r

diff=cbind(n\_vh69, n\_vh20, n\_vh50, n\_vh51, vit, sanct, rouge, arret, ceinture, an91, an92, an93, an94, an95, an96, an97, nb\_inf1, nb\_inf2, nb\_inf3, nb\_inf6, nb\_inf7, nb\_inf89)

#, Table 10 column 2 page 26

FE=c(0.0283,0.0532,0.0347,0.0841,0.2248,0.3857,0.3068,0.3443,0.1219,0.0586,0.0292,-0.0584,-0.0209,-0.0157,

 -0.0608,-0.1914,0.1584,0.2828,0.2321,0.2245,0.1583,0.2331)

#, Table 10 column 4 page 26

RE=c(0.0168,0.0864,0.0829,0.0849,0.2584,0.4245,0.3804,0.4105,0.1651,0.0995,0.0823,0.0877,0.1694,0.1631,

 0.0672,-0.1759,0.2006,0.2675,0.2770,0.2777,0.2012,0.2369)

#

diff\_true=RE-FE

# 

V=cov(diff[1:B,])

V\_1=solve(V)

# Hausman test statistic

TH[=t(diff\_true)%\*%V\_1%\*%diff\_true