5 HIERARCHICAL BAYESIAN ESTIMATION OF THE NUMBER OF VISITS TO THE GENERALIST IN 2002/2003 FRENCH HEALTH SURVEY

Marius ŞTEFAN*

Abstract

In our paper we show how to construct a model for one variable in the French Health Survey data set: the number of times an individual visited a generalist in the last twelve months, for which we are interested in estimating the regional means. Then, we test the fit of the model to the data and compare it to other two alternative models. We derive theoretical formulas for the estimates of the twenty-two regional means along with their standard deviations. We compare this to the design-based estimations obtained by INSEE in the case of the five regions with extra sample. We discuss some alternative for future research.

Keywords: small areas, direct and indirect estimations, Markov chains, Gibbs sampling, Metropolis-Hastings algorithm

JEL Classification: C12, C13, C46

1. Introduction

The French Health Survey (FHS) is a large survey (almost 30000 observations) taking place every ten years and collecting information on many health variables (more than 200). The French National Statistical Institute (INSEE) exploits the data by using its own methodology based on the sample design and calculates national estimations for a series of parameters linked to variables from FHS.

Lately, there has been a growing demand for estimations at sub-national levels. For instance, the French regional authorities are interested in estimations at the regional and county level (the French territory is divided into 22 regions, every region being divided into several counties, with a total of 95 counties).

The INSEE methodology is design-based. This means that the randomization comes from the sampling design. As a result, it can be seen that, if used for obtaining regional/county estimates, the resulting estimator called direct estimator uses only the observations coming from that region/county. Because surveys are generally

* Polytechnic University Bucharest, mastefan@gmail.com, mastefan@ulb.ac.be.

designed to ensure precise estimations at the national level, the sizes of the regional/county samples are not sufficient to ensure an adequate level of precision for the direct estimator at sub-national levels. For this reason, the region/county is considered a small area and alternative methods should be used. These methods form the small area estimation theory.

Authorities from five French regions (see below) decided to spend more money to increase their regional sample sizes, so that the regional estimations based on the INSEE methodology have an adequate level of precision. This resulted in approximately 10000 more individuals interviewed in the five regions. Based on this additional sample, INSEE computed and published national estimates plus five regional estimates together with their standard errors for a series of variables in FHS. However, these estimates now more precise due to the increased number of observations are still based on the INSEE methodology. In this paper, we show how to improve regional estimations by using small area estimation methods.

The small area estimation is the new theory trying to improve the classical design-based survey sample theory. The key of the modern small area estimation is the modelling of the variable of interest population values and then basing the estimation on the model. Multivariate estimation when several study variables are modelled simultaneously is also possible, although less frequently used. The model acts like a link between observations coming from different areas of the population. This is why the model-based estimation for a region/county parameter uses the entire national sample and not only the regional/county sample (this is called indirect estimation). Thus, the indirect estimator is more precise by “borrowing strength” from related areas.

INSEE does not have a methodology for small area estimation so *La Direction de la recherche, des études, de l’évaluation et des statistiques (DREES)*, which statistically exploits the data in FHS, financed a research aimed at finding a small area methodology for obtaining estimations at the regional or/and county level for a series of health variables. The results presented in this paper are part of this research.

In Section 2, we will show how to construct a model to estimate the regional means for the study variable R02AM, the number of visits to the generalist during the last year. Two alternatives models are presented and compared to the model used to obtain the estimations. In Section 3, we obtain the theoretical formulas of the estimators and of their standard errors by using full hierarchical Bayesian estimation methods. In Section 4, we test the fit of the three models by a number of discrepancies measures and p-values. In Section 5, using the formulas derived in Section 3, we compute the estimations and their precisions. We show the results only at the regional level and for the variable R02AM. We also estimated the county means of this variable and dealt with other variables such as the number of visits to the specialist in the last year or the body max index. Finally, in the last section we draw some conclusions and specify directions for future work.

2. Construction of the model

We saw earlier that the small area estimation is model-based, so that in this section we will show the steps we followed to construct a model for R02AM.

First, we undertook an extensive exploratory analysis in order to select the variables that explain R02AM (for details see Stefan M. (2006)). Better small area estimation can be obtained when auxiliary variables closely linked to the study variable are available. As a result of the preliminary analysis we retained four auxiliary variables: the Region, indexed by *i*=1,…,22, the Stratum, indexed by *j*=1,..,5 (this is a categorical variable, which is part of the sampling design and has five values depending on the population size of the commune an individual lives in), the Sex, indexed by *s*=1,2 and the Age, indexed by *k*=1,…,8 (we transformed the variable Age into a categorical variable with eight values corresponding to the intervals [0,1], [2,12], [13,23],…[56,67] and [68,104)).

Thus, we concluded that the individuals living in the same region and stratum and having the same sex and age on average should have the same value for the study variable. This is equivalent to cross-classifying the individuals in the population into $22\times 5\times 2\times 8$ cells, with y_{ijskl} being the value of R02AM for an individual *l* in the cell (i, j, s, k). y_{ijskl} should then verify:

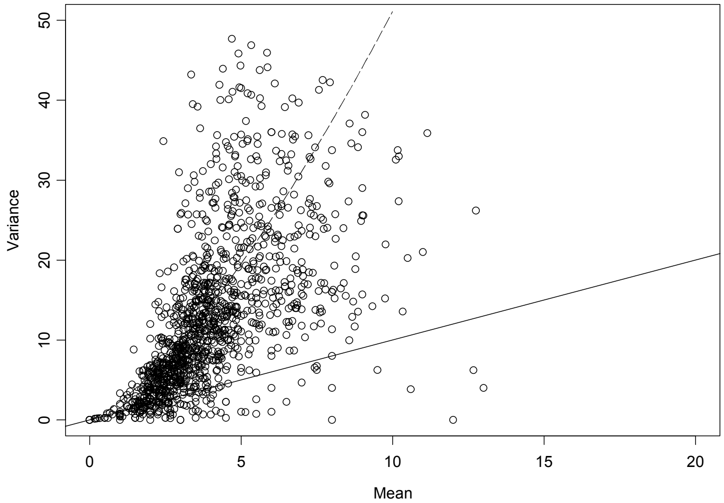
$$E(y_{ijskl}) \approx \mu_{ijsk}$$

with μ_{ijsk} denoting the population mean of the cell (i, j, s, k).

R02AM is a count variable. As a consequence, we will use the Poisson distribution to model it. The Poisson distribution has its mean equal to its variance. We have to test if R02AM verifies this condition. We computed the sample means and variances of the cells (i, j, s, k) for R02AM. Figure 1 represents the cell means versus the cell variances:

Figure 1

Mean versus Variance of R02AM



The solid line is of equation $y = x$. Most of the points are above it, suggesting that the variance is larger than the mean. Such a situation is called over dispersion and is frequently met in practice. The second curve is of equation $y = x + 1.3x^{1.5}$, which is a much closer approximation of the relationship between the variance and the mean. Clearly, the Poisson distribution is not appropriate for R02AM and something has to be done to handle the over dispersion.

There is a large literature about how the overdispersion present in the data can be taken into account. We adopt here one of the ways presented in Congdon P. (2005), pages 155-160, which consists of taking the parameter of the Poisson distribution random. In order to illustrate the effect of ignoring the over dispersion we also considered a model where the Poisson distributions parameters are fixed – this will be model 1 below. Thus, the first line of the hierarchical model will be:

$$y_{ijskl} \mid \nu_{ijsk} \underset{ind}{\sim} \text{Poisson}(\nu_{ijsk})$$

with random ν_{ijsk} having to be modelled.

It is easy to see that a much better choice would be a model where the parameters of the Poisson distribution depend on the individual *l* not only on the cell (i, j, s, k). This means replacing ν_{ijsk} above by ν_{ijskl}, with random ν_{ijskl} still having to be modelled. To illustrate the advantage of taking ν_{ijskl} instead of ν_{ijsk} we will consider model 2 with ν_{ijsk} and model 3 with ν_{ijskl} as parameters of the Poisson distribution:

$$y_{ijskl} \mid \nu_{ijsk} \underset{ind}{\sim} \text{Poisson}(\nu_{ijsk})$$

under model 2 and:

$$y_{ijskl} \mid \nu_{ijskl} \underset{ind}{\sim} \text{Poisson}(\nu_{ijskl})$$

under model 3.

For models 2 and 3 we have to model ν_{ijsk} and ν_{ijskl}, respectively. We will use a Gamma distribution with parameters $\mu_{ijsk}^{1-kappa} / alpha$ and $\mu_{ijsk}^{-kappa} / alpha$, where $alpha > 0$ and $kappa > -1$. The choice of the Gamma parameters was dictated by the data because it can be seen by a well-known formula for conditional means and variances that this choice leads to a relationship between $E(y_{ijskl})$ and $V(y_{ijskl})$, as shown in Figure 1. Clearly, if we put:

$$\nu_{ijsk} \mid \mu_{ijsk}, alpha, kappa \underset{ind}{\sim} \text{Gamma}(\frac{\mu_{ijsk}^{1-kappa}}{alpha}, \frac{\mu_{ijsk}^{-kappa}}{alpha})$$

under model 2 and

$$\nu_{ijskl} \mid \mu_{ijsk}, alpha, kappa \underset{ind}{\sim} \text{Gamma}(\frac{\mu_{ijsk}^{1-kappa}}{alpha}, \frac{\mu_{ijsk}^{-kappa}}{alpha})$$

under model 3, then one will get:

$$E(y_{ijskl}) = E_{\nu_{ijskl}} E(y_{ijskl} \mid \nu_{ijskl}) = E(\nu_{ijskl}) = \mu_{ijsk}$$

and

$$V(y_{ijskl}) = E_{\nu_{ijskl}} V(y_{ijskl} \mid \nu_{ijskl}) + V_{\nu_{ijskl}} E(y_{ijskl} \mid \nu_{ijskl}) = E(\nu_{ijskl}) + V(\nu_{ijskl}) =$$

$$= \mu_{ijsk} + alpha \times \mu_{ijsk}^{1+kappa}$$

which is the type of relationship between $E(y_{ijskl})$ and $V(y_{ijskl})$ indicated in Figure 1. The same proof holds under model 2 when *l* misses.

We are now at a point when we have to model the cell population means μ_{ijsk} under the three models. We will use the log which is the typical link function for the parameter of a Poisson distribution. Given the results of the exploratory analysis we will use the Region, Stratum, Sex and Age as explanatory variables for $\log(\mu_{ijsk})$:

$$\log(\mu_{ijsk}) = \beta_{1i} + \beta_{2j} + \beta_{3s} + \beta_{4k}$$

In order to make sure that some relevant explanatory variables are not overlooked and that the additive relation above is appropriate, we also considered and tested the fit of a model with $\log(\mu_{ijsk}) = \beta_{1i} + \beta_{2j} + \beta_{3s} + \beta_{4k} + \varepsilon_{ijsk}$, where ε_{ijsk} follow a normal distribution. We found that this model including the errors ε_{ijsk} did not fit the data much better that the model 3 without the errors and having much less parameters. Thus, we decided to drop the errors and keep the specification $\log(\mu_{ijsk}) = \beta_{1i} + \beta_{2j} + \beta_{3s} + \beta_{4k}$. Moreover, in order to avoid redundancy we imposed the usual corner constraints on the effects of the four auxiliary variables $\beta_{21} = \beta_{31} = \beta_{41} = 0$.

alpha, *kappa*, and the β are the hyper parameters of the models for which a priori laws has to be specified. We considered uniform laws with the intervals large enough to ensure their non-informative character. Alternative choices would be normal distributions of mean zero and large variances for real parameters, or gamma distribution with both parameters equal and small (0.001 is the usual value) for positive parameters. A sensitivity analysis to the choice of the a priori distributions not shown here was undertaken confirming a well-known fact that when the sample size is large the a priori laws have no or negligible impact on the posterior distributions and, as a consequence, on the inference. Thus, we conclude the hierarchical three models by the specifications:

$alpha \sim \text{Unif}(0,100)$, $kappa \sim \text{Unif}(-1,100)$, $\beta_{1i} \sim \text{Unif}(-10,10)$, $\beta_{2j} \sim \text{Unif}(-10,10)$, $\beta_{3s} \sim \text{Unif}(-10,10)$, $\beta_{4k} \sim \text{Unif}(-10,10)$. Thus, we will have:

Model 1

$$y_{ijskl} \mid \nu_{ijsk} \underset{ind}{\sim} \text{Poisson}(\nu_{ijsk}),$$

$$\log(\nu_{ijsk}) = \beta_{1i} + \beta_{2j} + \beta_{3s} + \beta_{4k} \text{ ,}$$
$$alpha \sim \text{Unif}(0,100) \text{ , } kappa \sim \text{Unif}(-1,100) \text{ ,}$$
$$\beta_{1i} \sim \text{Unif}(-10,10) \text{ , } \beta_{2j} \sim \text{Unif}(-10,10) \text{ , } \beta_{3s} \sim \text{Unif}(-10,10) \text{ ,}$$
$$\beta_{4k} \sim \text{Unif}(-10,10)$$

Model 2

$$y_{ijskl} \mid \nu_{ijsk} \underset{ind}{\sim} \text{Poisson}(\nu_{ijsk}) \text{ ,}$$
$$\nu_{ijsk} \mid \mu_{ijsk}, alpha, kappa \underset{ind}{\sim} \text{Gamma}(\frac{\mu_{ijsk}^{1-kappa}}{alpha}, \frac{\mu_{ijsk}^{-kappa}}{alpha}) \text{ ,}$$
$$\log(\mu_{ijsk}) = \beta_{1i} + \beta_{2j} + \beta_{3s} + \beta_{4k} \text{ ,}$$
$$alpha \sim \text{Unif}(0,100) \text{ , } kappa \sim \text{Unif}(-1,100) \text{ ,}$$
$$\beta_{1i} \sim \text{Unif}(-10,10) \text{ , } \beta_{2j} \sim \text{Unif}(-10,10) \text{ , } \beta_{3s} \sim \text{Unif}(-10,10) \text{ ,}$$
$$\beta_{4k} \sim \text{Unif}(-10,10)$$

and

Model 3

$$y_{ijskl} \mid \nu_{ijskl} \underset{ind}{\sim} \text{Poisson}(\nu_{ijskl}) \text{ ,}$$
$$\nu_{ijskl} \mid \mu_{ijsk}, alpha, kappa \underset{ind}{\sim} \text{Gamma}(\frac{\mu_{ijsk}^{1-kappa}}{alpha}, \frac{\mu_{ijsk}^{-kappa}}{alpha}) \text{ ,}$$
$$\log(\mu_{ijsk}) = \beta_{1i} + \beta_{2j} + \beta_{3s} + \beta_{4k} \text{ ,}$$
$$alpha \sim \text{Unif}(0,100) \text{ , } kappa \sim \text{Unif}(-1,100) \text{ ,}$$
$$\beta_{1i} \sim \text{Unif}(-10,10) \text{ , } \beta_{2j} \sim \text{Unif}(-10,10) \text{ , } \beta_{3s} \sim \text{Unif}(-10,10) \text{ ,}$$
$$\beta_{4k} \sim \text{Unif}(-10,10)$$

Model 3 will be found to have a sufficiently good fit to the data and it will be used to compute the estimations. Models 1 and 2 will only illustrate the effect of not taking into account the over dispersion and of having parameters ν_{ijsk} depending only on the cell and not on the individual.

3. Parameters estimation

In this section, we will derive theoretical formulas for the regional estimators and their standard errors under model 3. The way we do it is similar to Malec *et al*. (1997). The difference is that in their paper they deal with binary variables and they do not use individual parameters like we do under model 3 by choosing ν_{ijskl} instead of ν_{ijsk}. As we will see, despite the increased number of parameters that model 3 has as compared to models 1 or 2, it is worthwhile since model 3 presents a much better fit to the data.

Let μ_i be the population mean of R02AM for the region *i*. We want to estimate μ_i, which can be written as:

$$\mu_i = \frac{1}{N_i}\sum_j\sum_s\sum_k\sum_l y_{ijskl}$$

where N_i is the population size of the region. In the Bayesian context, a parameter is estimated by its posterior mean and the precision of this estimation will be measured by its posterior variance:

$$\hat{\mu}_i = E(\mu_i \mid \mathbf{y}_{obs}) \text{ and } V(\hat{\mu}_i) = V(\mu_i \mid \mathbf{y}_{obs})$$

where $\mathbf{y}_{obs}$ is the vector of all the observations. The population of a region can be divided in two parts: the observed and the unobserved individuals. Then, the regional mean will be given by:

$$\mu_i = \frac{1}{N_i}[\sum_j\sum_s\sum_k\sum_{l\in obs_i} y_{ijskl} + \sum_j\sum_s\sum_k\sum_{l\in nobs_i} y_{ijskl}]$$

where obs_i and $nobs_i$ are the observed and the unobserved part of the region *i*. By taking the conditional mean and variance of μ_i one will have:

$$\hat{\mu}_i = E(\mu_i \mid \mathbf{y}_{obs}) = \frac{1}{N_i}[\sum_j\sum_s\sum_k\sum_{l\in obs_i} y_{ijskl} + \sum_j\sum_s\sum_k\sum_{l\in nobs_i} E(y_{ijskl} \mid \mathbf{y}_{obs})]$$

and

$$V(\hat{\mu}_i) = V(\mu_i \mid \mathbf{y}_{obs}) = \frac{1}{N_i^2}V(\sum_j\sum_s\sum_k\sum_{l\in nobs_i} y_{ijskl} \mid \mathbf{y}_{obs})$$

It can be shown under model 3 (see Stefan M. (2006) for technical details) that for an individual $l_0 \in nobs_i$ not in the sample:

$$E(\nu_{ijskl} \mid \mathbf{y}_{obs}) = E(\mu_{ijsk} \mid \mathbf{y}_{obs}) \Rightarrow$$

$$\hat{\mu}_i = \frac{1}{N_i}[\sum_j\sum_s\sum_k\sum_{l\in obs_i} y_{ijskl} + E(\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu_{ijsk} \mid \mathbf{y}_{obs})] \tag{1}$$

where N_{ijsk} is the population size of cell (i, j, s, k) and n_{ijsk} is the number of individuals in the sample falling in cell (i, j, s, k). The relation $E(\nu_{ijskl} \mid \mathbf{y}_{obs}) = E(\mu_{ijsk} \mid \mathbf{y}_{obs})$ indicates that an unobserved individual will be replaced by the maximum we know about him or her, which is the mean of the cell he or she belongs to.

We also used model 3 to estimate the national mean μ and the precision of the estimation. To do so it is easy to see that in formula (1) one has to add a sum after the index *i* and replace N_i by the national population size *N*:

$$\hat{\mu} = \frac{1}{N}[\sum_{i}\sum_{j}\sum_{s}\sum_{k}\sum_{l \in obs_i} y_{ijskl} + E(\sum_{i}\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk})\mu_{ijsk} \mid \mathbf{y}_{obs})] \qquad (2)$$

As far as the posterior variance of μ_i is concerned, it can be proved (see Stefan M. (2006) for details) that:

$$V(\mu_i \mid \mathbf{y}_{obs}) = \frac{1}{N_i^2}[E(\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk})\mu_{ijsk} \mid \mathbf{y}_{obs}) +$$

$$+V(\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk})\mu_{ijsk} \mid \mathbf{y}_{obs}) + E(\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk}) alpha \mu_{ijsk}^{1+kappa} \mid \mathbf{y}_{obs})]$$

(3)

With the same modifications as above, the precision of $\hat{\mu}$ can be shown to be:

$$V(\mu \mid \mathbf{y}_{obs}) = \frac{1}{N^2}[E(\sum_{i}\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk})\mu_{ijsk} \mid \mathbf{y}_{obs}) +$$

$$V(\sum_{i}\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk})\mu_{ijsk} \mid \mathbf{y}_{obs}) + E(\sum_{i}\sum_{j}\sum_{s}\sum_{k}(N_{ijsk} - n_{ijsk}) alpha \mu_{ijsk}^{1+kappa} \mid \mathbf{y}_{obs})]$$

(4)

μ_i and μ are not parameters of model 3, so that they cannot be estimated directly by fitting this model. Formulas (1)-(4) are useful because they show how μ_i and μ depend on μ_{ijsk}, $alpha$ and $kappa$, which are parameters of model 3. By fitting model 3 we will get estimators of μ_{ijsk}, $alpha$ and $kappa$ that will be plugged in formulas (1)-(4) to obtain $\hat{\mu}_i = E(\mu_i \mid \mathbf{y}_{obs})$ and $V(\hat{\mu}_i) = V(\mu_i \mid \mathbf{y}_{obs})$.

Because of the multidimensionality of the posterior law of the parameters in model 3 it is impossible to get a closed form of this distribution. To tackle the problem, we will use the Gibbs sampling to obtain estimations of these parameters. The Gibbs sampling generates a Markov chain for each parameter of the model by using its full conditional distribution, which is the distribution of the parameter conditional on all the other parameters of the model and $\mathbf{y}_{obs}$.

A Markov chain for a parameter must be initialised at some value and it can be shown that whatever this value, after a “burn-in” period, the chain converges to its stationary distribution, which is the parameter posterior distribution. This means that after convergence the Markov chain values can be used to estimate different posterior characteristics: mean, variance, quintiles, etc… Thus, it is crucial to be able to detect the point where the Markov chain converged. The second issue addressed below is how many values (or iterations) can be used in order to get with a good approximation the posterior mean or variance.

The Gibbs sampling uses the full conditional distributions to generate the Markov chains. Some conditional distributions can be completely determined. For instance, under model 3 the full conditional distributions of v_{ijskl} are Gamma distributions, which can be sampled easily, but for the rest of the model 3 parameters their full conditional distributions can be determined up to a constant. In order to sample from such a non-standard distribution, we used Metropolis-Hastings and Neal (1997) algorithms (see Stefan M. (2006) for details).

The formulas (1) through (4) show how to use the Markov chains of model 3 parameters to compute $\hat{\mu}_i$ and $\hat{\mu}$ together with their precisions:

$$\hat{\mu}_i = \frac{1}{N_i}[\sum_j\sum_s\sum_k\sum_{l\in obs_i} y_{ijskl} + \frac{1}{G}\sum_g\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk}] \qquad (5)$$

$$V(\hat{\mu}_i) =$$

$$= \frac{1}{N_i^2}\{\frac{1}{G}\sum_g\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk} + \frac{1}{G}\sum_g\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk}) alpha^g \mu^g_{ijsk}{}^{1+kappa^g}$$

$$+\frac{1}{G}\sum_g[\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk}]^2 - [\frac{1}{G}\sum_g\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk}]^2\} \qquad (6)$$

$$\hat{\mu} = \frac{1}{N}[\sum_i\sum_j\sum_s\sum_k\sum_{l\in obs_i} y_{ijskl} + \frac{1}{G}\sum_g\sum_i\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk}] \qquad (7)$$

and

$$V(\hat{\mu}) = \frac{1}{N^2}\{\frac{1}{G}\sum_i\sum_g\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk} +$$

$$+\frac{1}{G}\sum_g\sum_i\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk}) alpha^g \mu^g_{ijsk}{}^{(1+kappa^g)} +$$

$$+\frac{1}{G}\sum_g[\sum_i\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk}]^2 - [\frac{1}{G}\sum_g\sum_i\sum_j\sum_s\sum_k (N_{ijsk} - n_{ijsk})\mu^g_{ijsk}]^2\} \quad (8)$$

where μ^g_{ijsk}, $alpha^g$ and $kappa^g$ $g = 1,...,G$, are the values indexed by *g* of the Markov chains of parameters μ_{ijsk}, *alpha* and *kappa*.

Based on formulas (5) and (7) we can construct Markov chains $\{\mu_i^g\}$ and $\{\mu^g\}$ for μ_i and μ, which are a combination of the Markov chains of μ_{ijsk}, *alpha* and *kappa*. First, these chains are an alternative to formulas (6) and (8) for computing the standard errors by simply taking the chain variance (we did both and obtained the same results). Second, they will be useful in computing the Monte Carlo Standard Errors (see below).

Let us notice in (5)-(8) that apart from the Markov chains μ_{ijsk}^{g}, $alpha^{g}$ and $kappa^{g}$ that can be obtained by fitting model 3 and the known values of the cell sample sizes n_{ijsk}, we also need to know N_{ijsk}, the cell population sizes. Their values are known from the 1999 French Census and were provided by the INSEE.

4. Choice and fit of the model

In this section, we compare the fit of the three models and conclude that model 3 is better than models 1 and 2. Then, we will check if model 3 fits well enough the data in FHS. We remind that model 1 ignores the over dispersion by taking ν_{ijsk} as fixed, while model 2 has random ν_{ijsk} depending on an individual through his or her region, stratum, sex and age only.

We selected several criteria well known in the literature (see Rao J.N.K. (2003) pages 232-237 for an excellent presentation). Given the large number of observations, the computations in this section use the sample without extension having 28259 observations. In this section, for notation facility *i* designates an individual.

One way in which a model fit can be tested is to generate for every individual *i* in the sample new observations $y_{new,i}$ from the posterior predictive density $f(y_i \,|\, \mathbf{y}_{obs})$ and to compare $\mathbf{y}_{new}$ to $\mathbf{y}_{obs}$ by some discrepancy measures. The model having the least posterior mean of one or several discrepancy measures will be chosen. We used three measures of discrepancy appropriate when the study variable is a count variable:

$$T(\mathbf{y}_{new},\mathbf{y}_{obs}) = \sum_i \frac{(y_{new,i}-y_{obs,i})^2}{(y_{new,i}+0.5)},$$

$$d(\mathbf{y}_{new},\mathbf{y}_{obs}) = 2\sum_i [(y_{obs,i}+0.5)\log\frac{(y_{obs,i}+0.5)}{(y_{new,i}+0.5)} - (y_{obs,i}-y_{new,i})],$$

$$D(\mathbf{y}_{new},\mathbf{y}_{obs}) = \sum_i (E(y_{new,i} \,|\, \mathbf{y}_{obs}) - y_{obs,i})^2 + \sum_i V(y_{new,i} \,|\, \mathbf{y}_{obs})$$

$f(y_i \,|\, \mathbf{y}_{obs})$ is the distribution of the number of visits to the generalist of individual *i* after the sample was taken and comparing the values generated by $f(y_i \,|\, \mathbf{y}_{obs})$ to what actually was observed will indicate how well a model fits to the data. It can be shown that a new value $y_{new,i}$ can be sampled from $f(y_i \,|\, \mathbf{y}_{obs})$ as follows: for each individual *i* we have the Markov chain $\{\nu_i^g\}$ obtained by fitting the model. After the burn-in period the values $\{\nu_i^g\}$ come from $f(\nu_i \mid \mathbf{y}_{obs})$; we considered a burn-in period of 2000 iterations and used the next *G*=1000 iterations; for each of the 1000 iterations we generated $y_{new,i}^g$ by sampling $\mathrm{Poisson}(\nu_i^g)$.

The posterior means for the three discrepancy measures above will be computed as follows:

$$E(T \mid \mathbf{y}_{obs}) = \frac{1}{G}\sum_g \sum_i \frac{(y_{new,i}^g - y_{obs,i})^2}{(y_{new,i}^g + 0.5)}\ ,$$

$$E(d \mid \mathbf{y}_{obs}) = 2\frac{1}{G}\sum_g \sum_i [(y_{obs,i} + 0.5)\log\frac{(y_{obs,i} + 0.5)}{(y_{new,i}^g + 0.5)} - (y_{obs,i} - y_{new,i}^g)]\,,$$

$$E(y_{new,i} \mid \mathbf{y}_{obs}) = \frac{1}{G}\sum_g y_{new,i}^g\ ,\ V(y_{new,i} \mid \mathbf{y}_{obs}) = \frac{1}{G}\sum_g (y_{new,i}^g - E(y_{new,i} \mid \mathbf{y}_{obs}))^2$$

from which one will get immediately $E(D \mid \mathbf{y}_{obs})$.

Besides the three measures above, we also considered the Deviance defined as two times the logarithm of the likelihood of $\mathbf{y}_{obs}$:

$$\mathrm{Deviance}(\mathbf{y}_{obs}, \mathbf{v}) = -2 * \sum_i \log(f(y_{obs,i} \mid \nu_i^g))$$

The Deviance can be seen as a measure of discrepancy between $\mathbf{y}_{obs}$ and $\mathbf{v}$. Its posterior mean will be given by:

$$E(\mathrm{Deviance} \mid \mathbf{y}_{obs}) = -2 * \frac{1}{G}\sum_g \sum_i \log(f(y_{obs,i} \mid \nu_i^g))$$

where $f(y_{obs,i} \mid \nu_i^g)$ is the value of the density function of a Poisson distribution of parameter ν_i^g computed at $y_{obs,i}$. The values for the discrepancy measures under the three models are shown in Table 1:

Table 1

Posterior means of the discrepancy measures

Discrepancy Measures	Model 1	Model 2	Model 3
$E(T(\mathbf{y}_{new}, \mathbf{y}_{obs}) \mid \mathbf{y}_{obs})$	204591.7	180845.9	66212.74
$E(d(\mathbf{y}_{new}, \mathbf{y}_{obs}) \mid \mathbf{y}_{obs})$	114884.9	109110.9	48095.09
$D(\mathbf{y}_{new}, \mathbf{y}_{obs})$	647926.3	619880.3	232111.5

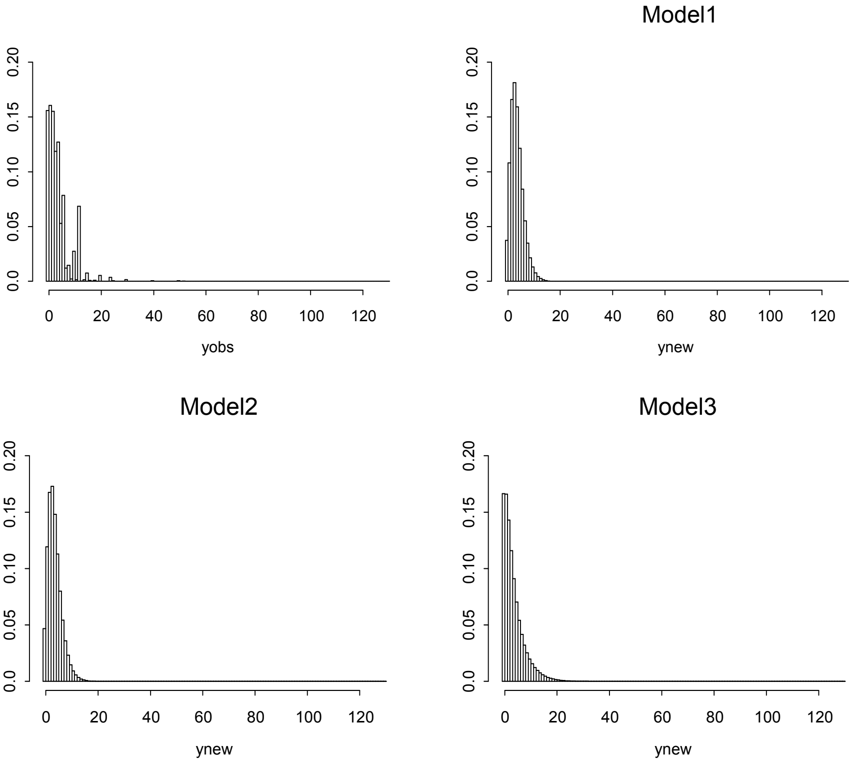
$E(\text{Deviance} \mid \mathbf{y}_{obs})$	174500	168300	103500

Table 1 shows clearly that model 1 ignoring the over dispersion has the poorest fit. Model 2, with random parameters for the Poisson distribution but common for all individuals in the same cell has a somewhat better fit. The model 3 has much more parameters than models 1 and 2. As a consequence, the time needed to estimate model 3 is much longer but as Table 1 demonstrates, model 3 improves heavily on the fit being a good alternative to the first two models.

Another way to compare $\mathbf{y}_{obs}$ and $\mathbf{y}_{new}$ is to examine their empirical distributions under the three models. They are represented in Figure 2:

Figure 2

Distribution of $\mathbf{y}_{obs}$ and Distributions of $\mathbf{y}_{new}$ under Models 1-3



We can again remark the bad fit of models 1 and 2. Under these models, the frequency of some important values of the study variable like 0 and 1 accounting for a large number of individuals in the sample are underestimated. On the contrary, model 3 produces values whose frequencies are close to what actually was observed.

Nevertheless, model 3 underestimates the number of persons which visited the generalist 12 times during the last year. In the empirical distribution of $\mathbf{y}_{obs}$ shown in Figure 2, 12 is the value whose frequency of nearly 7% presents a peak. This is contrary to the general tendency of decreasing frequencies as the number of visits increases. This could be explained by two facts: 1) an individual tends to respond "a dozen times" even if he or she visited the generalist ten times or less or thirteen times or more resulting in a larger than normal frequency of value 12; 2) there exists a subpopulation of individuals which visit the generalist once a month.

Whatever the reason, model 3 fails locally by underestimating the frequency of value 12. However, as we will see in the following, measures of overall fit shows that model 3 can be used to make inference about the regional means.

There is no doubt that model 3 fits the data much better than the other two. Now we have to answer the question: Is model 3 good enough? To see how good is model 3 we used the p-values which are linked to a measure of discrepancy between **y** and $\mathbf{v}$. Earlier we used such a measure - which is the Deviance - with the corresponding p-value being the probability that $\mathrm{Deviance}(\mathbf{y}^g_{new}, \mathbf{v}^g)$ exceeds $\mathrm{Deviance}(\mathbf{y}_{obs}, \mathbf{v}^g)$. An estimator of this p-value can be computed as:

$$\hat{p} = \frac{1}{G}\sum_g I[\mathrm{Deviance}(\mathbf{y}^g_{new}, \mathbf{v}^g) \geq \mathrm{Deviance}(\mathbf{y}_{obs}, \mathbf{v}^g)]$$

A value of $\hat{p}$ close to 0.5 indicates a good fit, while extreme values close to 0 or 1 indicate a poor fit and make us reject a model. For the model 3 we obtained $\hat{p} = 0.40$ indicating a good fit, while for models 1 and 2 we obtained 0.

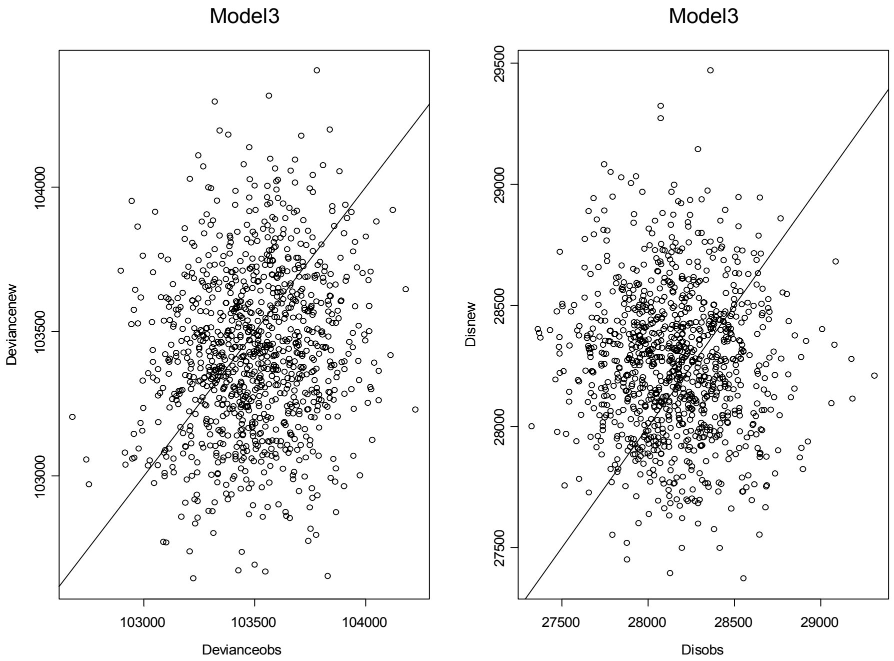
Besides the Deviance we can use another discrepancy measure between **y** and $\mathbf{v}$ appropriate when dealing with a count variable. Its formula and the corresponding p-value are given by:

$$\mathrm{Dis}(\mathbf{y}, \mathbf{v}) = \sum_i \frac{(y_i - v_i)^2}{v_i} \text{ and } \hat{p} = \frac{1}{G}\sum_g I[\mathrm{Dis}(\mathbf{y}^g_{new}, \mathbf{v}^g) \geq \mathrm{Dis}(\mathbf{y}_{obs}, \mathbf{v}^g)]$$

For model 3 we obtained $\hat{p} = 0.58$ and, as for the Deviance, we obtained again 0 under models 1 and 2.

Figure 3

$Deviance(\mathbf{y}_{obs}, \mathbf{v}^g)$ **vs** $Deviance(\mathbf{y}^g_{new}, \mathbf{v}^g)$ **and** $Dis(\mathbf{y}_{obs}, \mathbf{v}^g)$ **vs** $Dis(\mathbf{y}^g_{new}, \mathbf{v}^g)$



In Figure 3 we plotted $\text{Deviance}(\mathbf{y}_{obs}, \mathbf{v})$ against $\text{Deviance}(\mathbf{y}_{new}, \mathbf{v})$ and $\text{Dis}(\mathbf{y}_{obs}, \mathbf{v})$ against $\text{Dis}(\mathbf{y}_{new}, \mathbf{v})$. Under a good model, half of the points fall below the 45° line and the remaining half above the line. The two plots of Figure 3 confirm that model 3 is a fairly well fitted model:

Until now we used $f(y_i \mid \mathbf{y}_{obs})$ to generate new data. An alternative is represented by the cross validation prediction densities denoted by $f(y_i \mid \mathbf{y}_{(i)})$, where $\mathbf{y}_{(i)}$ is the vector of all individuals except *i*. For individual *i*, $f(y_i \mid \mathbf{y}_{(i)})$ suggests what values of y_i are likely when the model is fitted to $\mathbf{y}_{(i)}$. For every individual *i* one can compute the conditional predictive ordinate denoted CPO_i and defined as:

$$\text{CPO}_i = f(y_i \mid \mathbf{y}_{(i)})$$

It can be shown that CPO_i can be estimated by:

$$\text{CPO}_i = \hat{f}(y_i \mid \mathbf{y}_{(i)}) = \frac{1}{\frac{1}{G}\sum_g \frac{1}{f(y_i \mid v_i^g)}}$$

The model to select will be the model having the largest CPO_i, but due to the large number of points the different plots of CPO_i under the three models are useless and we do not show them. We shall use the CPO_i to compute the standardized residuals:

$$r_i = \frac{y_{obs,i} - E(y_i \mid \mathbf{y}_{(i)})}{\sqrt{V(y_i \mid \mathbf{y}_{(i)})}}$$

In order to have the r_i one needs $E(y_i \mid \mathbf{y}_{(i)})$ and $V(y_i \mid \mathbf{y}_{(i)})$. For an arbitrary function $a(y_i)$ we can prove that:

$$\hat{E}(a(y_i) \mid \mathbf{y}_{(i)}) = \hat{f}(y_i \mid \mathbf{y}_{(i)}) \frac{1}{G} \sum_g \frac{b_i(\nu_i^g)}{f(y_i \mid \nu_i^g)}$$

where $b_i(\nu_i^g) = E(a(y_i) \mid \nu_i^g)$. Under model 3 we have:

$$\hat{E}(y_i \mid \mathbf{y}_{(i)}) = \hat{f}(y_i \mid \mathbf{y}_{(i)}) \frac{1}{G} \sum_g \frac{\nu_i^g}{f(y_i \mid \nu_i^g)}$$

$$\hat{V}(y_i \mid \mathbf{y}_{(i)}) = \hat{f}(y_i \mid \mathbf{y}_{(i)}) \frac{1}{G} \sum_g \frac{\nu_i^g + \nu_i^{g2}}{f(y_i \mid \nu_i^g)} - [\hat{f}(y_i \mid \mathbf{y}_{(i)}) \frac{1}{G} \sum_g \frac{\nu_i^g}{f(y_i \mid \nu_i^g)}]^2$$

Let us notice that under model 3, $b_i(\nu_i^g)$ has a closed form, so that the formula above for $\hat{E}(a(y_i) \mid \mathbf{y}_{(i)})$ can be used. Otherwise, one has to sample $f(y_i \mid \mathbf{y}_{(i)})$) and compute $E(a(y_i) \mid \mathbf{y}_{(i)})$ and $V(a(y_i) \mid \mathbf{y}_{(i)})$ as the sample mean and the variance of the $a(y_i)$ values.

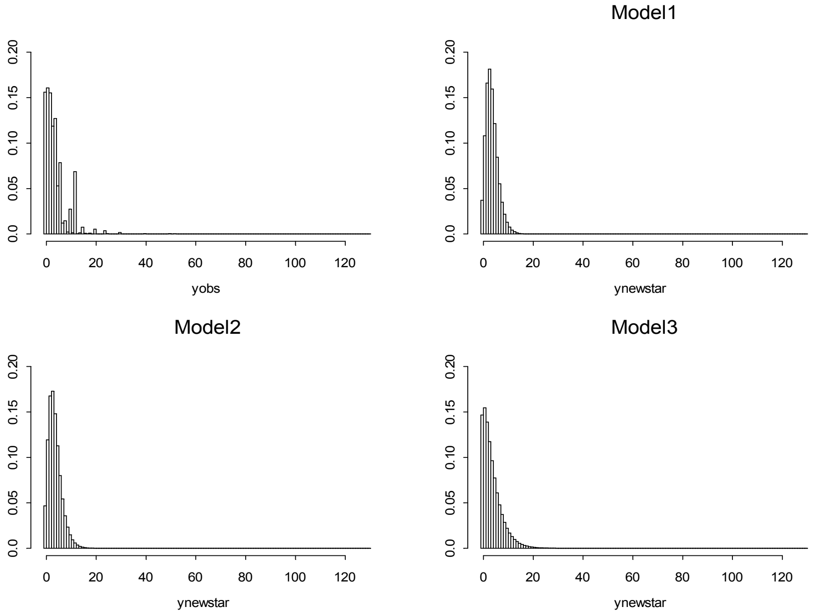
Fortunately, Gelfand (1996) shows how one can sample $f(y_i \mid \mathbf{y}_{(i)})$ without having to rerun the Gibbs sampling for every $\mathbf{y}_{(i)}$ and then sample the corresponding posterior predictive density to generate new values as we did above for $f(y_i \mid \mathbf{y}_{obs})$. Under model 3, the algorithm in Gelfand (1996) is equivalent to: from each vector $\mathbf{v}_i = (\nu_i^g)$ draw a sample with replacement and with probabilities proportional to $1/f(y_{obs,i} \mid \nu_i^g)$ and let $\mathbf{v}_i^* = (\nu_i^{g^*})$, *g*=1,...*G* be the new vector; for each element $\nu_i^{g^*}$ sample a value $y_{new,i}^{g^*}$ from $\mathrm{Poisson}(\nu_i^{g^*})$. The vector $\mathbf{y}_{new,i}^*$ will be composed of *G* values sampled from $f(y_i \mid \mathbf{y}_{(i)})$.

As we did with the empirical distributions of $\mathbf{y}_{new}$, we can plot the empirical distributions of $\mathbf{y}_{new}^*$ under the three models. The histograms are represented in

Figure 4. They are similar to those in Figure 2 and the same observation can be made about value 12 of the study variable.

Figure 4

Distribution of $\mathbf{y}_{obs}$ and Distributions of $\mathbf{y}^{*}_{new}$ under Models 1-3

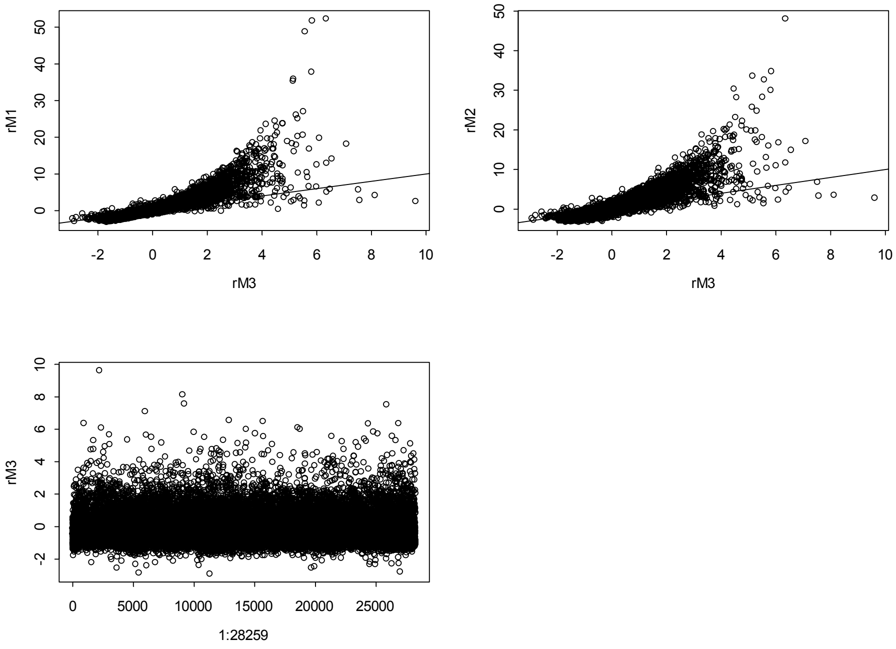


We can now use the values $\mathbf{y}^{*}_{new,i}$ to compute alternatively the residuals r_i or some other measures for which $E(a(y_i)|v_i^g)$ is not available in a closed form. In the first two plots of Figure 5 we represented residuals r_i under Model 3 against those under models 1 and 2, respectively. The solid lines are the 45° lines. One may clearly see that for a large number of individuals r_i are larger under models 1 and 2 than under model 3.

The third plot represents residuals r_i versus *i* under model 3. We can see that most of them are in absolute value less than 2. In fact, the percentage of r_i whose absolute values are higher than 2 is 2.94%. Under model 1 and 2, these percentages are 16.11% and 16.20%, respectively. The mean and standard deviation of r_i under model 3 are -0.10 and 0.90, respectively.

Figure 5

r_i under Model 3 against r_i under Models 1 and 2



5. Numerical computations of the regional estimations

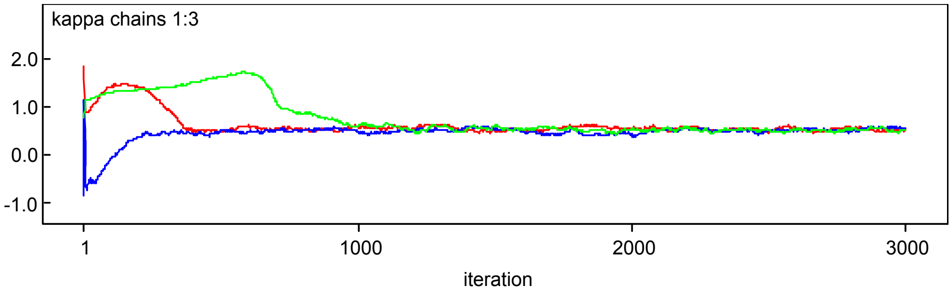
The estimations will use the formulas (5)-(8) where we can see what are the Markov chains that will be used: the Markov chains of parameters *alpha*, *kappa* and μ_{ijsk}. As we already said, we have to make sure that the chains converged.

One way to do it is to run several chains starting from different initial values and to follow the traces of the chains to detect the iteration from which the traces become undistinguishable. Another way is to apply one of the existing convergence diagnostics (see Cowles and Carlin (1996)). In Stefan M. (2006) we did both, but in this paper we will adopt the first.

First, let us notice that monitoring the convergence of the Markov chains of parameters μ_{ijsk} is equivalent to monitoring the chains of all the β parameters. For each parameter we run three chains initialized at different values. In figures 6 and 7 we have the Markov chains traces for parameters *alpha* and *kappa* (to save space we omitted those of β parameters).

Figure 6

Markov chains of parameter kappa



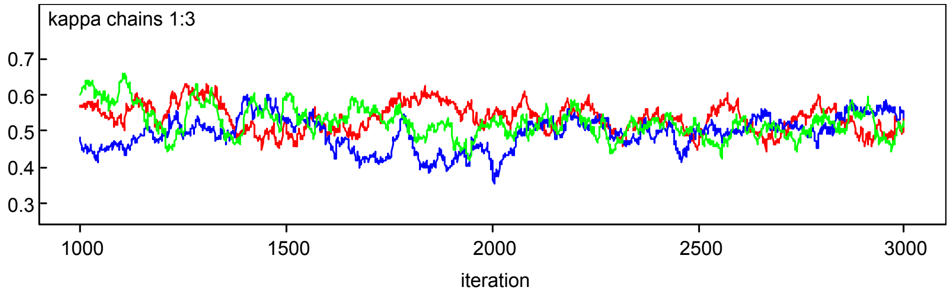
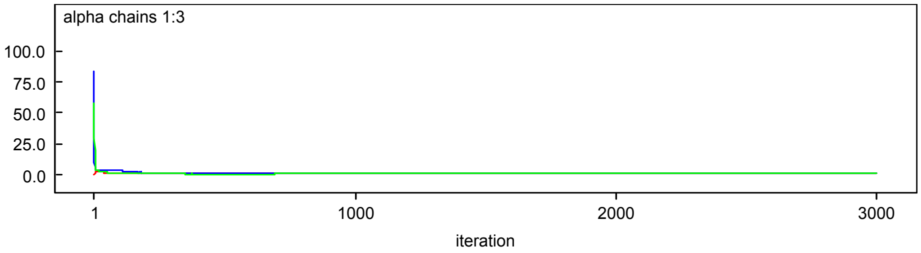
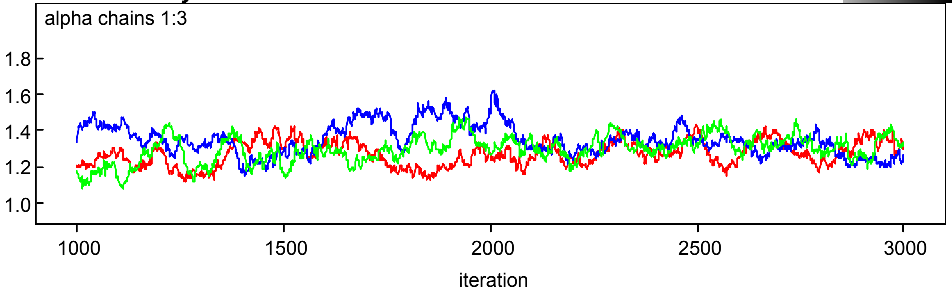


Figure 7

Markov chains of parameter alpha





From figures 6 and 7 one may see that for parameter *alpha* and *kappa* their Markov chains converge from iteration 2000 (the same holds true for the β parameters).

We could have looked directly to Markov chains $\{\mu_i^g\}$ and $\{\mu^g\}$ obtained as a mix. In Stefan M. (2006) we did and noticed a well-known fact that when constructed as a mix the autocorrelation of a Markov chain decreases and, as a consequence, the convergence is much more rapid. For $\{\mu_i^g\}$ and $\{\mu^g\}$ the convergence was attained after 100 iterations.

After discarding the first 2000 iterations we run the three chains for 2000 more iterations and used a total of 6000 iterations to compute the estimations and their standard errors. Table 2 presents our estimations and their precisions together with the six estimations and their precisions published by INSEE for the five regions with extra sample and France.

Table 2

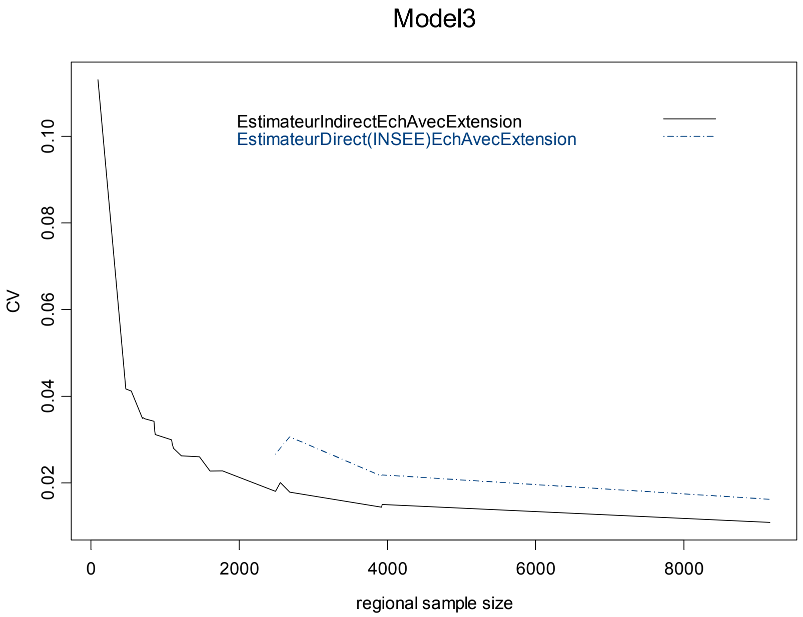
Estimations based on the sample with extension (39900 observations)

Region *** = region with extension**	**Estimation**	**Standard Error**	**Estimation INSEE**	**Standard Error INSEE**
Ile de France*	3.00	0.0330	3	0.0485
Champagne-Ardenne*	4.41	0.0832	4.4	0.1169
Picardie*	4.36	0.0773	4.5	0.1376
Haute-Normandie	4.20	0.1498		
Centre	4.02	0.1145		
Basse-Normandie	4.31	0.1488		
Bourgogne	3.68	0.1248		
Nord Pas de Calais*	5.13	0.0729	5.4	0.1162
Loraine	4.53	0.1176		
Alsace	4.18	0.1331		
Franche Compté	3.85	0.1395		
Pays de la Loire	4.01	0.0916		
Bretagne	3.97	0.1006		
Poitou Charente	4.44	0.1386		
Aquitaine	4.30	0.1003		
Midi-Pyrénées	4.37	0.1247		
Limousin	4.80	0.1957		
Rhône Alpes	3.27	0.0660		
Auvergne	4.15	0.1657		
Languedoc-Roussillon	4.24	0.1212		
PACA*	3.88	0.0607	4	0.0872
Corse	2.71	0.3019		
France Métropolitaine	**3.93**	**0.0207**	**4**	**0.0303**

In Figure 8, the black solid line represents the coefficients of variations of our estimations versus the regional sample sizes. The blue dotted line represents the coefficients of variations of the estimations for the five regions with extra sample provided by INSEE. The small area estimation methodology that we developed clearly represents an alternative to the classical methodology based on the sample design given the fact that we obtained coefficients of variation half those provided by INSEE.

Figure 8

Coefficient of Variation versus Size of the Regional Sample



We have mentioned that when dealing with Markov chains apart the convergence one has to examine the length or the number of iterations to be used in computing the posterior quantities of interest. Clearly, the length of the chain will determine the precision with which the chain will approximate a posterior quantity but this is not the only factor. There is also the autocorrelation which characterizes a Markov chain by definition: larger the autocorrelation more iterations are needed.

The precision of the approximation is called Monte Carlo Standard Error (MCSE) and the formulas when approximating a posterior mean and a posterior standard error are given below (*G* is the number of iterations used and $\hat{\rho}_1$ is the Markov chain coefficient of autocorrelation of order 1):

$$\mathrm{MCSE}(\hat{\mu}) = \frac{\hat{\sigma}}{\sqrt{G}}\sqrt{\frac{1+\hat{\rho}_1}{1-\hat{\rho}_1}}\ ,\ \mathrm{MCSE}(\hat{\sigma}) = \frac{\hat{\sigma}}{\sqrt{2G}}\sqrt{\frac{1+\hat{\rho}_1^2}{1-\hat{\rho}_1^2}} \quad (9)$$

The above formulas are valid only when the Markov chain resembles an autoregressive process of order 1. They will be applied on the Markov chains $\{\mu_i^g\}$ and $\{\mu^g\}$. Thus, we verified that these Markov chains are AR(1) by plotting the partial autocorrelation functions and remarking that all the partial autocorrelation coefficients starting from order 2 are negligible. We omit the plots (details can be found in Stefan M. (2006)).

Table 3

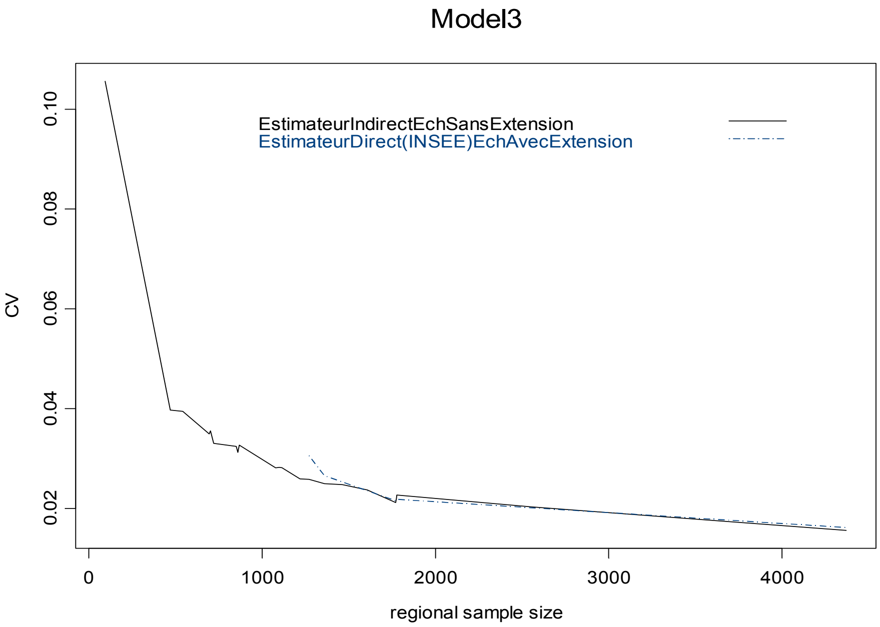
Estimations based on the sample with extension (39900 observations)

Region ***=region with extension**	$\mathrm{MCSE}(\hat{\mu})$	$\mathrm{MCSE}(\hat{\sigma})$	**Quantile 0.025**	**Quantile 0.5**	**Quantile 0.975**
Ile de France*	0.0006	0.0003	2.93	3.00	3.06
Champagne-Ardenne*	0.0014	0.0008	4.25	4.42	4.58
Picardie*	0.0013	0.0007	4.21	4.36	4.51
Haute-Normandie	0.0025	0.0014	3.91	4.20	4.49
Centre	0.0019	0.0011	3.80	4.01	4.24
Basse-Normandie	0.0026	0.0014	4.03	4.32	4.61
Bourgogne	0.0022	0.0012	3.45	3.68	3.93
NordPasdeCalais*	0.0012	0.0007	4.99	5.13	5.27
Loraine	0.0020	0.0011	4.31	4.53	4.76
Alsace	0.0023	0.0013	3.92	4.17	4.44
FrancheCompté	0.0024	0.0013	3.59	3.84	4.12
PaysdelaLoire	0.0016	0.0009	3.83	4.01	4.19
Bretagne	0.0018	0.0010	3.77	3.96	4.17
PoitouCharente	0.0023	0.0013	4.17	4.44	4.72
Aquitaine	0.0017	0.0009	4.10	4.29	4.50
Midi-Pyrénées	0.0022	0.0012	4.14	4.37	4.62
Limousin	0.0034	0.0019	4.42	4.80	5.18
RhôneAlpes	0.0012	0.0006	3.14	3.27	3.39
Auvergne	0.0028	0.0016	3.83	4.15	4.48
Languedoc-Roussillon	0.0021	0.0012	4.01	4.23	4.48
PACA*	0.0010	0.0006	3.76	3.88	4.00
Corse	0.0057	0.0031	2.16	2.69	3.34
FranceMétropolitaine	**0.0003**	**0.0002**	**3.89**	**3.93**	**3.97**

Table 3 presents the Monte Carlo Standard Errors for the regional means estimations $\hat{\mu}_i$ and for theirs standard errors $\hat{\sigma}_i$ computed using 6000 iterations together with estimations for 0.025, 0.5 and 0.975 quintiles with which one can construct confidence intervals. We can see that 6000 iterations are sufficient to approximate closely $\hat{\mu}_i$ and $\hat{\sigma}_i$.

Figure 9

Coefficient of Variation versus Size of the Regional Sample



We repeated the analysis above for the sample without extension. In Figure 9, the solid line represent the coefficient of variation of the 22 estimations based on the sample without extension. The dotted line represents the coefficients of variation of the five estimations published by INSEE but computed from the sample with extension. We ignore if INSEE applied the design-based methodology to estimate the means of the regions which did not benefit from extra sample so, like in Figure 8, the blue line represents only five estimations. For the five regions with extra sample it is clear that with much less observations we obtained almost the same coefficients of variations. Moreover, Figure 9 seems to indicate that smaller the sample size better the small area methodology based estimations even if the latter are computed using fewer observations.

6. Conclusions

The objective of this paper was to construct a model on which to base the inference aimed at obtaining regional estimations for the number of visits to the generalist in the last year. We compared our estimations with the INSEE design-based estimations and found that our estimations are better even computed with one third fewer observations.

We considered one variable R02AM, which is a count variable. In the future, we shall construct models for other variables in FHS (count, binary or continuous). For other count variables the methodology will be essentially the same, but for binary variables the Poisson distribution will be replaced by Bernoulli or Binomial distributions. Over or under dispersion will have to be checked and accounted for properly.

We reported here only regional estimations, but DREES was also interested in county level estimations. In model 3, we replaced the region effect β_{1i} by a county effect β_{1d} which resulted in a model allowing the estimation at the county level and also at the region and national level. In a future paper we shall present these results.

In deriving the formulas (1)-(4) we supposed that the cell sample sizes n_{ijsk} are non-random. In practice, this is generally not true. In the classical survey sampling theory computations using random n_{ijsk} are not feasible, that is why under such circumstances analyses are conditional on the realized sample sizes. In a full hierarchical Bayesian context Oleson *et al.* (2007) proposed a model accounting for random sample sizes and also population sample sizes. Based on their paper, we shall extend our present work.

Survey sampling is generally characterized by non-response and FHS is no exception. If not properly accounted for the non-response, it can lead to biased estimation. In our paper, we supposed that there is complete response. In fact, we removed the 1000 or so individuals that did not provide any value for R02AM and done our analysis on the remaining ones. Nandram *et al*. (2005) and the references therein constitute a large literature to see how the non-response in FHS can be properly dealt with in a full hierarchical Bayesian context.

In testing the fit of model 3, we noticed a failure of the model which was the underestimation of the observed value 12 (overall the model performed well and we concluded that it can be used in estimating the regional means). We explained this either by wrong report by the interviewed individuals or the existence of a subpopulation visiting a generalist every month.

A possible way to remedy this failure could be the use of Bayesian Nonparametric Statistics (BNS). Instead of putting a Gamma distribution on the parameters ν one could consider the set of all the possible distributions of ν and put a distribution on this space (for instance, the Dirichlet process). BNS offer a greater flexibility and adapt better to the data than the classical parametric Bayesian theory.

References

Congdon P. (2005), *Bayesian models for categorical data*, John Wiley&Sons, Chichester-England.

Cowles M.K. and Carlin B.P. (1996), Markov chain Monte Carlo convergence diagnostics:a comparative review, *Journal of the American Statistical Association*, 91: 883-904.

Gelfand A.E. (1996), *Markov Chain Monte Carlo in practice*, edited by Gilks W.R., Richardson S., Spiegelhalter D.J., Chapter 9, 145-158.

Gelman A. and Rubin D.B. (1992), Inference from iterative simulation using multiple sequences. *Statistical science*, 7: 457-472.

Gilks W. (1992), Derivative-free adaptive rejection sampling for Gibbs sampling. In *Bayesian statistics 4* edited pby Bernardo J.M., Berger J.O., David A.P., Smith A.F.M. Oxford University Press, U.K., 641-665.

Malec D., Sedransk J., Moriarity C., LeClere F. (1997), *Journal of the American Statistical Association*, 92: 815-826.

Nandram B., Cox L., Choi W.J., (2005), Bayesian analysis of nonignorable missing categorical data : An application to bone mineral density and family income, *Survey Methodology*, 31(2): 213-225.

Neal R. (1997), Markov chain Monte Carlo methods based on slicing the density function, Technical report 9722, Department of Statistics, University of Toronto.

Oleson J., He C., Sun D., Sheriff S., (2007), Bayesian estimation in small areas when the sampling design strata differ from the study domains, *Survey Methodology*, 33(2): 173-185.

Rao J.N.K. (2003), *Small area estimation*, John Wiley&Sons, Hoboken-New Jersey.

Ştefan M. (2006), *Enquete décennale santé 2002/2003: estimation petits domaines*, Rapport final de recherche, DREES, Paris.