مجلّه‌ی پژوهش‌های آماری ایران

J. Statist. Res. Iran **17** (2020): 157–170

دوره‌ی ۱۷، شماره‌ی ۱، بهار و تابستان ۱۳۹۹، صص ۱۵۷–۱۷۰

DOI: 10.52547/jsri.17.1.157

DOR: 20.1001.1.25385771.2020.17.1.8.9

A Bayesian Nominal Regression Model with Random Effects for Analysing Tehran Labor Force Survey Data

Masumeh Ahmadzadeh and Taban Baghfalaki*

Tarbiat Modares University

Received: 2022/18/02 Approved: 2022/27/08

Abstract. Large survey data are often accompanied by sampling weights that reflect the inequality probabilities for selecting samples in complex sampling. Sampling weights act as an expansion factor that, by scaling the subjects, turns the sample into a representative of the community. The quasi-maximum likelihood method is one of the approaches for considering sampling weights in the frequentist framework. To obtain it the ordinary log-likelihood is replaced by the weighted log-likelihood. There is a Bayesian framework as a counterpart to quasi-maximum likelihood method is called Bayesian pseudo posterior estimator. This method is the usual Bayesian approach by replacing likelihood with quasi-likelihood function. Another approach for considering sampling weights called the Bayesian weighted estimator. This method is in fact a data augmentation method in which a quasi-representative sample is generated by sampling instead of the observed data using normalized sampling weights. In this paper, these two approaches are used for parameter estimation of a nominal regression model with random effects. The proposed method is applied to small area estimates for the Tehran labor force survey in 2018.

Keywords. Bayesian approach, labor force survey, nominal data, random effects, sampling weights, small area estimation.

* Corresponding author

MSC 2010: 62F15, 62P99.

1 Introduction

One of the most important statistical plans that has a potential role in the economic planning of countries is the labor force survey plan. This plan is implemented quarterly with the aim of estimating labor force indicators such as the unemployment and employment rates throughout the country, and in each province. Unemployment rate is the most important and influential issue in many sectors of society and is thus an issue of primary interest for society in general and in particular for local, regional and central governments.

There are many studies for estimating the unemployment rates. Unemployment rates for US states were estimated using the Hierarchical Bayesian method (Datta et al., 1999). Fabrizi (2002) discussed the estimation of unemployment rate for the Italian labor force in the small area using the Bayesian paradigm. You and Rao (2002) and You (2008) propose unmatched models for estimating under-count rates in the Canadian Census of population. More recently, Aitkin (2008) and Rao and Wu (2010) incorporate sampling weights into pseudo Bayesian methods for a multinomial empirical likelihood leading to Dirichlet posterior distribution. They provide Bayesian interval estimates for the population mean that are asymptotically valid in a frequentist framework. Margolis and Okatenko (2008) proposed a job search model for unemployed workers with the Bayesian approach.

Since the unemployment status has three levels: unemployment, employment and inactive. For estimating the unemployment rate, the inactive population will be removed from the study and in fact unemployment and employment are considered as two competing events. In this paper, we consider these three levels of labor force survey. A multinomial logit model is considered for modeling the mentioned categorical variable. Also, this study is considered for the counties of Tehran province. For considering association among the people in each county, a random effects model is also considered. Two Bayesian paradigms including Bayesian pseudo posterior and Bayesian weighted estimators are considered for statistical inference. The proposed methods are used for analysing Tehran labor force survey in 2018.

The paper is organized as follows. The data of Tehran labor force survey in 2018 is introduced in Section 2. Section 3 includes a general description of nominal regression model with random effects and description of the parameters estimation with Bayesian pseudo posterior estimator and Bayesian weighted estimator. We apply the model to analyse the data in section 4. The last section includes some conclusions.

2 Motivation: Tehran Labor Force Survey in 2018

The motivating data discussed in this paper were collected from a labor force survey (LFS) that is a quarterly survey of households for measuring the economically active population conducted by the Statistical Centre of Iran in 2018. The LFS is a probabilistic sample of household units that produces quarterly labor force and related estimates for all members of private settled households whose place of usual residence was located in Tehran Province at the time of the enumeration. The quarterly sample is a multistage stratified sample of 12682 units and they are collected from 16 counties of Tehran. The LFS interview is divided into two parts: 1) household's number and their demographic information, 2) their labor force information. In the second part, information is obtained for each member of the household: with 10 years and older. One of the primary goals of labor force information is to classify persons in one of the three categories: employed (34.92%), unemployed (5.06%), and inactive (60.02%). The Horwitz-Thompson estimation (HTE) of unemployment, employment and inactive rates for each county of Tehran can be found in Table 1. Also, the explanatory variables with Sampling weights includes personal characteristics such as gender, age, current marital status, education status and the number of household members. Details of categories of explanatory variables and their percentages for the categorical and the summary statistics for the continuous explanatory variables are described in Tables 2 and 3, respectively.

3 Material and Methods

3.1 Nominal Regression Model with Random Effect

Let there be n areas in the study and each area has $J_i,\ i = 1, \cdots, n$ units. Also, let $\boldsymbol{y}_{ij}$ be the response variable for the j^{th} unit of the i^{th} area such that $\boldsymbol{y}_{ij} = (y_{ij1}, \cdots, y_{ijR})'$ and $\sum_{r=1}^{R} y_{ijr} = 1$. In addition, we consider $w_{ij}, i = 1, \cdots, n, j = 1, \cdots, J_i$, as sampling weight for the j^{th} unit of the i^{th}

Table 1. Horwitz-Thompson estimation (HTE) of unemployment, employment and inactive rates.

County	Unemployment	Employment	Inactive
Tehran	0.052	0.350	0.598
Damavand	0.025	0.363	0.612
Ray	0.048	0.371	0.581
Shemiranat	0.032	0.298	0.670
Varamin	0.054	0.362	0.584
Shahriar	0.049	0.352	0.599
Eslamshahr	0.033	0.334	0.634
Robat Karim	0.058	0.339	0.613
Pakdasht	0.051	0.358	0.592
Firuzkuh	0.010	0.380	0.610
Qods	0.075	0.292	0.633
Malard	0.063	0.350	0.587
Pishva	0.010	0.359	0.631
Baharestan	0.045	0.350	0.611
Pardis	0.026	0.387	0.587
Qarchak	0.044	0.362	0.593

Table 2. Different levels of the categorical explanatory variables along with their percentages by considering sampling weights.

Explanatory variable	Categories	Percentage
Gender	Female	50
	Male	50
Marital status	Married	60.71
	Widow or divorced	7.07
	Single	32.22
Education status	Any level of education	92.69
	No any level of education	7.31

Table 3. A summary of continuous explanatory variables along with their percentages.

Explanatory variable	Min	Max	Mean	Standard deviation
Age	10	96	38.17	17.14
Number of household members	1	9	3.17	1.2

area. These weights are designed to reflect unequal probabilities of response and selection inherent in complex survey sampling methods. We consider the following model for analysing the data:

$$\boldsymbol{y}_{ij} \sim Mn(1, \boldsymbol{\pi}_{ij}), \boldsymbol{\pi}_{ij} = (\pi_{ij1}, \ldots, \pi_{ijR})', i = 1, \cdots, n, j = 1, \cdots, J_i, \quad (1)$$

where $Mn(1, \boldsymbol{\pi})$ is used to denote a multinomial distribution, also, for $i = 1, \cdots, n, j = 1, \cdots, J_i$,

$$\begin{aligned} \pi_{ijr} &= P(Y_{ijr} = 1|b_{ir}) = \tfrac{\exp(v_{ijr})}{1+\exp(v_{ijr})}, r = 1, \cdots, R-1, \\ \pi_{ijR} &= P(Y_{ijR} = 1|b_{iR}) = \tfrac{1}{1+\exp(v_{ijr})}, \end{aligned} \quad (2)$$

and $v_{ijr} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_r + \boldsymbol{z}'_{ij}\boldsymbol{b}_{ir}$, $\boldsymbol{x}_{ij}$ is $p \times 1$ vector of the explanatory variables, $\boldsymbol{\beta}_r$ is a $p-$dimensional vector of the regression coefficients, $\boldsymbol{z}_{ij}$ is $q-$dimensional vector of the explanatory variables for the random effects, $\boldsymbol{b}_{ir}$ is a $q-$dimensional vector of the random effects such that $\boldsymbol{b}_{ir} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_r)$. Equation (1) can be written as the following multinomial logit model:

$$log(\pi_{ijr}/\pi_{ijR}) = v_{ijr} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_r + \boldsymbol{z}'_{ij}\boldsymbol{b}_{ir},\ r = 1, \cdots, R-1.$$

3.2 Bayesian Inference

In this paper, we consider two approaches for Bayesian paradigm: Bayesian pseudo posterior estimator (BPPE) and Bayesian weighted estimator (BWE). In the following, these two approaches will be described.

Bayesian pseudo posterior estimator (BPPE)

For describing this approach, the same as those is done in the usual Bayesian inference, considering the prior distributions for the unknown parameters is necessary. For this purpose, we consider a normal distributional assumption for the regression coefficients ($N_p(\boldsymbol{\beta}_{r0}, \boldsymbol{\psi}_{r0})$) and a inverse Wishart distribution for the covariance matrix of the random effects ($IWishart(\boldsymbol{\Omega}_r, \nu_r)$). Therefore, we have the following hierarchical model:

$$\begin{aligned} &\boldsymbol{y}_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{\beta}_r, \boldsymbol{b}_{ir} \sim Mn(1, \boldsymbol{\pi}_{ij}), \\ &\boldsymbol{b}_{ir}|\boldsymbol{\Sigma}_r \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_r), \\ &\boldsymbol{\beta}_r|\boldsymbol{\beta}_{r_0}, \boldsymbol{\psi}_{r0} \sim N_p(\boldsymbol{\beta}_{r_0}, \boldsymbol{\psi}_{r0}), \\ &\boldsymbol{\Sigma}_r|\boldsymbol{\Omega}_{r_0}, \boldsymbol{\upsilon}_{r0} \sim \mathrm{IWishart_p}(\boldsymbol{\Omega}_{\mathrm{r_0}}, \boldsymbol{\upsilon}_{\mathrm{r0}}). \end{aligned}$$

For obtaining, the pseudo posterior estimator, the pseudo likelihood function is used instead of the likelihood function. Therefore, the pseudo posterior distribution is given by

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y_{ij}}, \boldsymbol{b_{ir}}, w_{ij}) \propto p(\boldsymbol{\theta}, \boldsymbol{b_{ir}}) \prod_{i=1}^{n} \prod_{j=1}^{J_i} p(\boldsymbol{y_{ij}}|\boldsymbol{\theta})^{w_{ij}}.$$

Let define

$$\boldsymbol{\theta} = (\boldsymbol{b_{11}}, \ldots, \boldsymbol{b_{nR-1}}, \boldsymbol{\beta_1}, \ldots, \boldsymbol{\beta_{R-1}}, \boldsymbol{\Sigma_1}, \ldots, \boldsymbol{\Sigma_{R-1}})$$

$$\boldsymbol{D} = \{(\boldsymbol{y_{11}}, \ldots, \boldsymbol{y_{nJ_i}}, \boldsymbol{w_1}, \ldots, \boldsymbol{w_n}, \boldsymbol{x_{11}}, \ldots, \boldsymbol{x_{nJ_i}}, \boldsymbol{z_{11}}, \ldots, \boldsymbol{z_{nJ_i}}), i = 1, \ldots, n\}.$$

Thus, the joint pseudo posterior distribution is given by

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{D}) &\propto p(\boldsymbol{D}|\boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}) \\ &\propto \prod_{i=1}^{n}\prod_{j=1}^{J_i}\left[\prod_{r=1}^{R-1} \frac{\exp(w_{ij} y_{ijr} (\boldsymbol{x'_{ij}\beta_r} + \boldsymbol{z'_{ij}b_{ir}}))}{(1 + \sum_{r=1}^{R-1} \exp{(\boldsymbol{x'_{ij}\beta_r} + \boldsymbol{z'_{ij}b_{ir}})})^{\boldsymbol{w_{ij}}}}\right] \\ &\times \exp\{\frac{-1}{2}(\boldsymbol{b'_{ir}\Sigma_r^{-1}b_{ir}})\} \\ &\times \exp\{\frac{-1}{2}(-(\boldsymbol{\beta_r} - \boldsymbol{\beta_{r0}})'\psi_{\boldsymbol{r}}^{\boldsymbol{-1}}(\boldsymbol{\beta_r} - \boldsymbol{\beta_{r0}}))\} \\ &\times |\boldsymbol{\Sigma_r}|^{\frac{-v+p+2}{2}} \exp\{\frac{-1}{2} tr(\boldsymbol{\Omega'_r\Sigma_r^{-1}\Omega_r})\}. \end{aligned} \tag{3}$$

For performing Bayesian inference, generating samples from the joint pseudo posterior distribution (3) is necessary. As generation of this joint pseudo posterior distribution may not be performed easily, the MCMC approach containing Gibbs sampler along with the Metropolis-Hastings algorithm should be applied. For this purpose, we require all the full conditional distributions. The full conditional distributions of all the unknown parameters are as follows. The full conditional distributions of $\boldsymbol{b_{ir}}$ $(i = 1, \cdots, n,\ r = 1, \cdots, R-1)$

is given by:

$$
\begin{aligned}
p(\boldsymbol{b_{ir}}|\boldsymbol{y_{ij}},\boldsymbol{\beta_r},\boldsymbol{\Sigma_r}) &\propto \prod_{j=1}^{J_i} p(\boldsymbol{y_{ij}}|\boldsymbol{b_{ir}},\boldsymbol{\beta_r},\boldsymbol{\Sigma_r})p(\boldsymbol{b}_{ir}|\boldsymbol{\Sigma_r}) \\
&\propto \prod_{j=1}^{J_i}\left[\prod_{r=1}^{R-1}\frac{\exp(w_{ij}y_{ijr}(\boldsymbol{x'_{ij}\beta_r}+\boldsymbol{z'_{ij}b_{ir}}))}{(1+\sum_{r=1}^{R-1}\exp{(\boldsymbol{x'_{ij}\beta_r}+\boldsymbol{z'_{ij}b_{ir}})})^{w_{ij}}}\right] \qquad (4)\\
&\times \exp\left\{\frac{-1}{2}(-\boldsymbol{b'_{ir}}\boldsymbol{\Sigma_r}^{-1}\boldsymbol{b_{ir}})\right\}.
\end{aligned}
$$

The full conditional distributions of $\boldsymbol{\beta_r}$ $(r=1,\cdots,R-1)$ is given by:

$$
\begin{aligned}
p(\boldsymbol{\beta_r}|\boldsymbol{y_{ij}},\boldsymbol{b_{ir}},\boldsymbol{\Sigma_r}) &\propto \prod_{i=1}^{n}\prod_{j=1}^{J_i} p(\boldsymbol{y_{ij}}|\boldsymbol{\beta_r},\boldsymbol{b_{ir}},\boldsymbol{\Sigma_r})p(\boldsymbol{\beta_r}|\boldsymbol{\beta_{r0}},\boldsymbol{\varphi_r}) \\
&\propto \prod_{i=1}^{n}\prod_{j=1}^{J_i}\left[\prod_{r=1}^{R-1}\frac{\exp(w_{ij}y_{ijr}(\boldsymbol{x'_{ij}\beta_r}+\boldsymbol{z'_{ij}b_{ir}}))}{(1+\sum_{r=1}^{R-1}\exp{(\boldsymbol{x'_{ij}\beta_r}+\boldsymbol{z'_{ij}b_{ir}})})^{w_{ij}}}\right] \qquad (5)\\
&\times \exp\{\frac{-1}{2}(-(\boldsymbol{\beta_r}-\boldsymbol{\beta_{r0}})\boldsymbol{\Sigma_r}^{-1}(\boldsymbol{\beta_r}-\boldsymbol{\beta_{r0}}))\}.
\end{aligned}
$$

Finally, the full conditional distributions of $\boldsymbol{\Sigma_r}$ $(r=1,\cdots,R-1)$ is given by:

$$
\begin{aligned}
p(\boldsymbol{\Sigma_r}|\boldsymbol{y_{ij}},\boldsymbol{\beta_r},\boldsymbol{b_{ir}}) &\propto \prod_{i=1}^{n}\prod_{j=1}^{J_i} p(\boldsymbol{y_{ij}}|\boldsymbol{\Sigma_r},\boldsymbol{\beta_r},\boldsymbol{b_{ir}})p(\boldsymbol{\Sigma}_r|\boldsymbol{\Omega_r},\boldsymbol{v_r}) \\
&\propto \prod_{i=1}^{n}\prod_{j=1}^{J_i}\left[\prod_{r=1}^{R-1}\frac{\exp(w_{ij}y_{ijr}(\boldsymbol{x'_{ij}\beta_r}+\boldsymbol{z'_{ij}b_{ir}}))}{(1+\sum_{r=1}^{R-1}\exp{(\boldsymbol{x'_{ij}\beta_r}+\boldsymbol{z'_{ij}b_{ir}})})^{w_{ij}}}\right] \qquad (6)\\
&\times \exp\{\frac{-1}{2}(\boldsymbol{b'_{ir}\Sigma_r^{-1}b_{ir}})\} \\
&\times |\boldsymbol{\Sigma_r}|^{\frac{-v+p+2}{2}}\exp\{\frac{-1}{2}tr(\boldsymbol{\Omega'_r\Sigma_r^{-1}\Omega_r})\}.
\end{aligned}
$$

Bayesian Weighted Estimator (BWE)

In this subsection, another approach for Bayesian inference by considering sampling weights is described, where is Bayesian weighted estimator (BWE) and it is introduced by Gunawan et al. (2020). Let $\boldsymbol{y}=(y_{11},\cdots,y_{nJ_n})'$ and $\boldsymbol{w}=(w_{11},\cdots,w_{nJ_n})'$. Aslo, let $\boldsymbol{y}^*=(y^*_{11},\cdots,y^*_{nJ_n})'$ be pseudo representative samples. At first, a mechanism for simulating $\boldsymbol{y}^*$ conditional on both the data ($\boldsymbol{y}$) and weights $\boldsymbol{w}$ is considered. This mechanism is denoted $p(\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{w})$. Gunawan et al. (2020) proposed a simulation-based posterior inference, which is based on the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}^*) \propto p(\boldsymbol{y}^*|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $p(\boldsymbol{y}^*|\boldsymbol{\theta})$ is the likelihood of the parametric model of interest. As mentioned by Gunawan et al. (2020), a natural way to handle randomness in the mechanism for simulating $\boldsymbol{y}^*$ is to integrate out over $\boldsymbol{y}^*$, that is

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{w}) &= \int_{\boldsymbol{y}}^{*} p(\boldsymbol{\theta},\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{w})d\boldsymbol{y}^* \\ &= \int_{\boldsymbol{y}}^{*} p(\boldsymbol{\theta}|\boldsymbol{y}^*,\boldsymbol{y},\boldsymbol{w})p(\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{w})d\boldsymbol{y}^* \\ &= \int_{\boldsymbol{y}}^{*} p(\boldsymbol{\theta}|\boldsymbol{y}^*)p(\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{w})d\boldsymbol{y}^*. \end{aligned} \tag{7}$$

Note that the implicit assumption here is that $\boldsymbol{y}$ and $\boldsymbol{w}$ provide no further information about $\boldsymbol{\theta}$ that is not already captured by $\boldsymbol{y}^*$, and, a Monte Carlo sample of $(\boldsymbol{\theta},\boldsymbol{y}^*)$ from $p(\boldsymbol{\theta},\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{w})$ (Gunawan et al., 2020).
BWE depends on two choices: 1) the mechanism for generating a pseudo representative samples, 2) the method used to draw from the posterior of the parameters given $\boldsymbol{y}^*$. We now discuss each of these in the following.

Generating a pseudo representative samples For this purpose, One can draw a sample of size $N = \sum J_i$ from the weighted empirical distribution of the data. In our context, that is a discrete distribution with domain $\boldsymbol{y} = (y_{11}, \cdots, y_{nJ_n})'$ and with probabilities $\tilde{\boldsymbol{w}} = (\tilde{w}_{11}, \cdots, \tilde{w}_{nJ_n})'$, where $\tilde{w_{ij}} = \frac{w_{ij}}{N}$ is the normalised weight. Note that when all weights are equal, this is identical to sampling with replacement.

Simulated-based inference After simulating $\boldsymbol{y}^*$ all that remains is to conduct inference as if pseudo representative samples were actual data. For this aim, it is possible to directly draw from $p(\boldsymbol{\theta}|\boldsymbol{y}^*)$ as follows:
For $i = 1, \cdots, N$ do

- draw $\boldsymbol{y}^{*(r)}$ from $p(\boldsymbol{y}^*|\boldsymbol{y},\boldsymbol{w})$.
- draw $\boldsymbol{\theta}^{(r)}$ from $p(\boldsymbol{\theta}|\boldsymbol{y}^*)$.

In this stage a sequential or parallel algorithm can be applied. For more details see Gunawan et al. (2020).

Finally, the usual posterior inference can be carried out on this sample of $\boldsymbol{\theta}$. For instance, all posterior expectations can be approximated by sample means and the credible intervals can be obtained by looking at quantiles of the iterates of $\boldsymbol{\theta}$. s

4 Application

In this section, we analyse the described data of Section 2. For this purpose, let $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, y_{ij3})'$ be the response variable such that y_{ij1}, y_{ij2} and y_{ij3} are indicator variables for employment, unemployment and inactive status, respectively. The aim of this study is to detect significant predictors, therefore, we consider age, education status, current marital status (have any level of education), gender and the number of household members as the explanatory variables using the following linear predictor:

$$\begin{aligned} v_{ijr} = {} & \beta_{0r} + \beta_{1r}\text{Gender}_{ij} + \beta_{2r}\text{Edu}_{ij} \\ & + \beta_{3r}\text{Mar1}_{ij} + \beta_{4r}\text{Mar2}_{ij} + \beta_{5r}\text{Age}_{ij} + \beta_{6r}\text{HSize}_{\text{ij}} + b_{ir}, \end{aligned} \tag{8}$$

Where Age and $HSize$ are used to denote age and the number of household members, respectively. Also,

$$\begin{aligned} &\text{Gender} = \begin{cases} 1 & \text{Female} \\ 0 & \text{Male} \end{cases}, \\ &\text{Edu} = \begin{cases} 1 & \text{Literate} \\ 0 & \text{Illiterate} \end{cases}, \\ &\text{Mar1} = \begin{cases} 1 & \text{Single} \\ 0 & \text{o.w.} \end{cases}, \quad \text{Mar2} = \begin{cases} 1 & \text{Married} \\ 0 & \text{o.w.} \end{cases}. \end{aligned} \tag{9}$$

We assume that $b_{ir} \sim N(0, \sigma_r^2)$. We consider 20000 iterations as MCMC iterations sampler, including 10000 pre-convergence burn-in. Convergence of the chains is checked using Brooks-Gelman-Rubin (BGR) diagnostics (Gelman and Rubin , 1992). Also, another 1000 iterations are considered for implementing the Bayesian weighted estimator. Tables 4 and 6 report parameter estimation, standard deviation of parameters and 95% credible interval for

Bayesian pseudo-posterior method and Bayesian weighted estimator, respectively. The two Bayesian approaches are two strategies for considering the sampling weights and as we expect the results of two approaches are completely close.

Based on the results of these tables, age is a significant regression coefficient such that by keeping other levels of employment status constant, the probability of a person becomes unemployed as he gets older is more than the probability of being employed, and the probability of being employed is more than being inactive. All the regression coefficients of gender are significant, that is, women are more likely to be unemployed and they are more likely to be inactive than employed. All the regression coefficients of education are significant, that is, those who are literate are more likely to be employed and more likely to be unemployed than inactive. Also, all the regression coefficients for singles are significant, i.e., singles are more likely to be unemployed and more likely to be employed than inactive.

Table 7 shows estimation of unemployment, employment and inactive rates for Tehran province in labor force data. As the results for the two approaches are close, one obtained by Bayesian pseudo-posterior method is reported to save space and the estimated standard errors for all rates are less than 0.001. It should be mentioned that the reported unemployment rates by the statistical center of Iran is obtained by removing the inactive population, and in fact it is considered a binary response variable with two levels: employment and unemployment. Based on the results of this table, the largest rate of unemployment [=0.075] and the smallest rate of employment [=0.288] is for Quds county. This county includes 2.32% of data.

Note that gardening is highly prevalent in Firoozkooh county due to its suitable geographical location and favorable climate, as well as the existence of rivers and freshwater resources. In addition to livestock farming as well as aquaculture is very common in this county. Also, Pishva county, due to the fertility of its soil, always has a significant role in agriculture, so that most people in this area are engaged in agriculture. On the other hand, due to the fertility of livestock, soil and beekeeping is also very important in this county. Hence, the unemployment rate in these two countise is much lower than other counties as the smallest rate of unemployment [=0.010] belongs to Firuzkoh and Pisheva counties with less than one percent of data (0.28% and 0.71%, respectively). Based on the results of the table Pardis, including 1.3% of data, has the largest employment rate [=0.384]. Finally, Shemiranat with 0.30% of data has the largest rate of inactive states [=0.677] and Ray with 2.55% of data has the smallest rate of inactive states [=0.584].

5 Conclusion

In this paper, a nominal random effect model is applied to analyze the labor force data of Tehran province and then Bayesian weighted and Bayesian pseudo-posterior methods have been compared to estimate their parameters. The explanatory variables affecting the employment rate, unemployment and inactivity have been identified. Also, the employment rate, unemployment and inactivity of people in Tehran province have been estimated. It has been shown that for the data the results of two methods are close. For the applied nominal regression model, the computational time for the Bayesian weighted method was longer than the Bayesian pseudo-posterior method. Although, the first purpose of this paper was analysing the labor force data of Tehran province and estimating different rates of the labor force for the counties of this province, but we had another purpose for suggesting a strategy to choose between the Bayesian weighted and Bayesian pseudo-posterior methods for other statistical models. We propose to use one of the Bayesian methods by checking complexity of model or computation time for future. If one is not able to consider the pseudo likelihood or it is complicated to estimate parameter of it, or if there is available packages for implementing the model, the use of Bayesian weighted method will be proposed but, based of our experience, by considering a small number of iterations the standard error of the parameters will be estimated large and the efficiency of the method is lost. Thus, considering the large number of iterations (we considered 1000) is necessary. In this paper, we have not considered the group structure of the data which can be considered for the future.

Table 4. Bayesian parameter estimates, posterior means (standard errors, S.E.) and 95% credible intervals by Bayesian pseudo-posterior method.

		Parameters	Est.	S.E	2.5%	97.5%
Employed Status		Intercept	0.524	0.005	0.515	0.531
	Gender	Female	−1.007	0.002	−1.010	−1.004
		Male (baseline)				
	Education	Literate	−0.668	0.013	−0.7018	−0.654
		Illiterate (baseline)				
	Marital status	Single	−1.185	0.004	−1.192	−1.179
		Married	0.539	0.003	0.533	0.544
		Divorced or Widow (baseline)				
	Age		0.004	0.002	0.001	0.007
	Family members		−0.053	0.001	−0.054	−0.052
	σ_1^2		9.145	3.648	4.467	18.451
Unemployed Status		Intercept	2.987	0.004	2.977	2.944
	Gender	Female	2.236	0.002	2.232	2.400
		Male (baseline)				
	Education	Literate	−1.243	0.014	−1.258	−1.211
		Illiterate (baseline)				
	Marital status	Single	−0.615	0.004	−0.623	−0.609
		Married	0.897	0.004	0.890	0.903
		Divorced or Widow (baseline)				
	Age		1.109	0.002	1.106	1.112
	Family members		−0.047	0.001	−0.048	−0.046
	σ_2^2		17.282	6.965	8.555	34.650

Table 5. Bayesian parameter estimates, posterior means (standard errors, S.E.) and 95% credible intervals by Bayesian weighted method.

		Parameters	Est.	S.E.	2.5%	97.5%
Employed Status		Intercept	0.613	0.006	0.601	0.625
	Gender	Female	−0.982	0.005	−0.975	−1.001
		Male (baseline)				
	Education	Literate	−0.985	0.004	−0.978	−0.996
		Illiterate (baseline)				
	Marital status	Single	−0.983	0.213	−1.429	−0.570
		Married	0.565	0.004	0.555	0.570
		Divorced or Widow (baseline)				
	Age		0.006	0.003	0.001	0.008
	Family members		−0.061	0.002	−0.072	−0.053
	${\sigma_1}^2$		9.112	3.425	4.389	17.249
Unemployed Status		Intercept	3.111	0.003	3.103	3.118
	Gender	Female	2.312	0.001	2.309	2.313
		Male (baseline)				
	Education	Literate	−1.231	0.012	−1.211	−1.242
		Illiterate (baseline)				
	Marital status	Single	−0.549	0.005	−0.629	−0.541
		Married	0.882	0.005	0.876	0.896
		Divorced or Widow (baseline)				
	Age		1.167	0.003	1.159	1.173
	Family members		−0.055	0.001	−0.049	−0.045
	σ_2^2		16.358	5.985	8.698	32.987

Table 6. Percentage of frequency of samples in Tehran province, taking into account the weight of sampling in the labor force data.

County	Percentage
Tehran	65.60
Dmavand	1.18
Ray	2.55
Shemiranat	0.30
Varamin	2.44
Shahriar	5.49
Eslamshahr	4.153
Pakdasht	2.32
Robat Karim	2.82
Firuzkoh	0.28
Quds	2.32
Malard	2.79
Pisheva	0.71
Baharestan	3.49
Pardis	1.30
Qarchak	2.22

Table 7. Estimation of unemployment, employment and inactive rates for Tehran province in labor force data.

County	Unemployment	Employment	Inactive
Tehran	0.052	0.346	0.602
Dmavand	0.027	0.353	0.620
Ray	0.048	0.369	0.584
Shemiranat	0.031	0.292	0.677
Varamin	0.050	0.350	0.600
Shahriar	0.052	0.353	0.596
Eslamshahr	0.033	0.330	0.637
Pakdasht	0.050	0.352	0.598
Robat Karim	0.058	0.321	0.621
Firuzkoh	0.010	0.370	0.620
Quds	0.075	0.288	0.637
Malard	0.062	0.348	0.590
Pisheva	0.010	0.354	0.635
Baharestan	0.045	0.339	0.616
Pardis	0.020	0.384	0.596
Qarchak	0.045	0.366	0.589

References

Aitkin, M. (2008). Applications of the Bayesian bootstrap in finite population inference. *Journal of Official Statistics*, **24**, 21.

Datta, G.S., Lahiri, P., Maiti, T., and Lu, K.L. (1999). Hierarchical Bayes estimation of unemployment rates for the states of the US. *Journal of the American Statistical Association*, **94**, 1074-1082.

Fabrizi, E. (2002). Hierarchical Bayesian models for the estimation of unemployment rates in small domains of the italian labour force survey. *Statistica*, **62**, 603-618.

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

Gunawan, D., Panagiotelis, A., Griffiths, W., and Chotikapanich, D. (2020). Bayesian weighted inference from surveys. *Australian & New Zealand Journal of Statistics*, **62**, 71-94.

Margolis, D.N., and Okatenko, A. (2008). *Job Search with Bayes Priors.*

Rao, J.N.K., and Wu, C. (2010). Bayesian pseudo-empirical-likelihood intervals for complex surveys. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**, 533-544.

You, Y. (2008). An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology*, **34**, 19.

You, Y., and Rao, J.N.K. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, **30**, 3-15.

Masumeh Ahmadzadeh
Department of Statistics,
Tarbiat Modares University,
Tehran, Iran.
email: *m.ahmadzade69@gmail.com*

Taban Baghfalaki
Department of Statistics,
Tarbiat Modares University,
Tehren, Iran.
email: *t.baghfalaki@modares.ac.ir*